PREDICTING OF URBAN EXPANSION USING CONVOLUTIONAL LSTM NETWORK MODEL: THE CASE OF SEOUL METROPOLITAN AREA, KOREA

J.M. Kim[1], J.S. Park[2], C.Y. Lee[3], S.G. Lee[4*]

[1]Dept. of Urban Planning and Engineering, Hanyang University, Seoul, Korea (jinmankim324@hanyang.ac.kr)
[2]Dept. of Urban Planning and Engineering, Hanyang University, Seoul, Korea (june1264@hanyang.ac.kr)
[3]Dept. of Urban Engineering, Jeonbuk National University, Jeonju, Korea (idealcity@jbnu.ac.kr)
[4*]Dept. of Urban Planning and Engineering, Hanyang University, Seoul, Korea (sugielee@hanyang.ac.kr)

Commission IV, WG IV/9

KEY WORDS: Urban Growth Model, Land Use Change, Machine Learning Model, LSTM Model.

ABSTRACT:

As urbanization progresses, many studies about the analysis and prediction of land-use change and urban sprawl have been conducted recently. As the sprawl phenomenon progresses rapidly, the urban expansion phenomenon became uncontrolled and it has affected negatively on the city's environment and transportation finally. So, it is essential to identify lands likely to be urbanized in the future because it aids in establishing land use plans and policies pre-acting the negative impact of spatially urban expansion the sprawl by determining factors affecting the urban sprawl. Previous studies based on statistical models are limited to identifying determining factors, so the prediction performance is low compared to deep learning. On the other hand, existing studies using machine learning and deep learning overlook selecting specific region-focused variables. Therefore, this study aims to analyze and predict changes in the Seoul Metropolitan Area's sprawl in Korea using the Convolutional Long Short-Term Memory Network (ConvLSTM) with factors at the city scale and neighboring factors at the local scale in the Seoul Metropolitan Area (SMA). ConvLSTM is a type of combination model: combining Recurrent Neural Network(RNN) and Convolutional Neural Network(CNN). This study showed that ConvLSTM with factors at the city and neighboring factors at the local scale predicted the urbanized land. The determinants contain population and roads ratio at the city scale, and neighboring urban lands, distance to the nearest subway stations, slope, and elevation at the local scale. The results reveal that predicted urban lands in 2020 increase over the entire region. In particular, the expected urban lands in 2020 increase by reducing farmlands in the southern part of the SMA. It is consistent with the trend of urbanized lands from 1980 to 2010. In addition, urbanization occurs in areas adjacent to Seoul due to the well-established urban infrastructure. The results of this study can be used as evidence to establish sustainable land use plans and regulations in the future.

1. INTRODUCTION

1.1 Background of the study

Cities are the most typical form of human life, and urbanization has increased worldwide over the past 100 years (Newbold and Scott, 2013). In addition, 55% of the world's population resides in urban areas, and by 2050, this figure is expected to increase to 68% (Yadav and Ghosh, 2021). This rapid urban sprawl reduces open space, biodiversity, etc., and also causes socioeconomic problems such as traffic congestion and poor public health (Bhatta, 2010; Macdonald et al., 2011; Kim et al., 2020). Urban sprawl continues to appear in the metropolitan area of Korea, and most of the development of suburban areas of the metropolitan area consists of high-density apartment complexes focusing on housing site development projects (jin et al., 2013; Jon & Woo, 2019). Therefore, in order to come up with countermeasures against these problems, it is necessary to identify the factors affecting urban expansion. It can also be said that it is an important part of urban research to use this to predict future urban expansion and to prepare a management plan based on it (Ghavami et al., 2017).

Most studies to predict urban expansion used remote sensing (RS) satellite image data. This remote sensing data is used to analyze changes in the physical environment such as ecosystems and land use on the Earth's surface. These RS data are typically used to analyze changes in the ecosystem, natural resources, living space, and physical environment of the Earth's surface caused by human activity (Song et al., 2018; Boulia et al., 2021). However, there is a limitation that such data should be mapped using image processing technology and then used instead of original data. As a result, some land use can cause errors or distortions due to inaccurate image classification. Therefore, in this study, analysis was performed using accurate classification data of land use using land cover map image data.

As an analytical methodology, methods such as cellular automata (CA) and Markov chains (MC) have traditionally been used to predict changes in land use (Wang et al., 2019; Rimal et al., 2018). Since methodologies such as CA and MC presuppose linear assumptions, there is a limit to predicting situations with nonlinear relationships in reality such as urban expansion. Therefore, methodologies using artificial intelligence such as deep learning and machine learning are being developed to solve these problems (Wang et al., 2019). Among them, the Long Short-Term Memory (LSTM) network model is a model that can efficiently predict time series data (Pascanu et al., 2013). However, since the LSTM model is designed to process one-dimensional data, there is a limitation that the land cover map image data should be reduced to one-dimensional and then predicted. Therefore, in this paper, we try to address these limitations by applying the Convolution LSTM (ConvLSTM) model combined with the Convolutional Neural Network (CNN) model. The ConvLSTM network model has the advantage of being able to learn without reducing spatial data (Zhong et al., 2019).

* Corresponding author

Therefore, this paper uses land cover data and the ConvLSTM model to predict urban expansion in the Seoul metropolitan area in Korea. The temporal range of the study is from 1980 to 2010, and finally, the urbanization area in 2030 is predicted to present policy implications for urban expansion in the Korean metropolitan area.

1.2 The purpose of the study

The purpose of this study is to propose a ConvLSTM approach to predict urban expansion in the metropolitan area of Korea. Specifically, the ConvLSTM model was learned using the land cover map data of 1980, 1990, and 2000, and the accuracy of the model was verified using the ground truth image of 2010. Since then, the 1990s, 2000s, and 2010 models have been used to predict the urban expansion of the Korean metropolitan area in 2020 and 2030 and to present policy implications for creating a sustainable city.

2. LITERATURE REVIEW

2.1 Land-use prediction

Traditionally, land use has been predicted using Markov Chain (MC) and cellular automata (CA) algorithms as a way of predicting land use. Sibanda and Ahmed (2021) used CA and MC models to predict wetlands in the Shashi region of Zimbabwe. The analysis was conducted using Landsat satellite data from 1984, 1995, 2005, and 2015, and the analysis claimed the importance of various factors for designing sustainable monitoring and conservation strategies, with the wetland area decreasing by about 53% by 2045. Etemadi et al (2018) simulated Land Cover Change (LCC) in mangrove forests in 2025 using satellite images and predicted it using the CA-Markov model. As a result of the analysis, the mangrove growth area of 21ha was changed to an open water area, and 28ha was expected to be expanded to land.

Hamley et al. (2015) studied changes in Land Use Land Cover (LULC) in some parts of the northwestern desert of Egypt and predicted changes using the Markov-CA integrated approach. The distribution of LULC in the desert for 1988, 1999, and 2011 was mapped, and similar analysis results were derived from LULC in 2011. Saha et al. (2022) predicted the pattern of LULCs in five urban areas of the Silikuri region of India using the Multi Layer Perceptron Neural Network (MLPNN)-based Markov-Chain (MC) model. The MLPNN-MC model has the advantage of outperforming other models in predicting future patterns of LULC due to its ability to represent multiple or full land-use transformations at once, and eventually produced a LUCL prediction map for 2050. As a result of the analysis, it was predicted that the urban area would almost double the current amount by 2050, and it was argued that the transportation network was one of the main factors causing urban expansion.
Dadashpoor & Salarian (2020) predicted land use using Slope, Land Cover, Exclusion, Urban grwoth, Transport and Hillsha elevation, and access to wide-area public transportation such as subways. Therefore, it can be seen that it is essential to predict land use changes by reflecting the variables.

2.2 Convolutional LSTM Network

LSTM is one of the Recurrent Neural Networks (RNNs) that treats inputs and outputs as a sequence, an algorithm that improves the Vanishing phenomenon, where the learning rate decreases as the layer becomes deeper due to a large number of learning errors (Hochreiter & Schmidhuber, 1997). The LSTM network has a limitation in that spatial information is lost because learning proceeds through vectorized data. Convolutional Neural Network (CNN), on the other hand, is a widely used algorithm to extract features of maps using Convolution operations to learn relationships between input pixels. Convolution LSTM (ConvLSTM) is a method that was developed by combining these two algorithms. ConvLSTM was proposed by Xingjian et al. (2015), and has the advantage of being able to train models without reducing the dimension of space, which was mentioned as a limitation of LSTM.

After semantic segmentation using satellite image data, Boulila et al (2021) calculated the map with three components: 2D convolution, ReLU, and Batch Normalization, and compared the performance of Pix2Pix, Dual GAN, and ConvLSTM for predicting urban expansion. As a result of the analysis, the ConvLSTM model was found to be the best, and the possibility of using transfer learning and deep learning techniques for detecting changes in high-resolution satellite images was presented. You et al (2017) predicted crop yields using remote sensing data, and trained CNN and LSTM using histogram-based dimensionality reduction. This approach results in Root Mean Squared Error (RMSE) being 30% better than conventional methods and 15% better in terms of Mean Absolute Percentage Error (MAPE). Compared to these traditional prediction methods, recent AI-based prediction models are more frequently used due to the flexibility of data that is easy to implement and can handle large amounts of data (Boulila et al., 2021).

In the study of Boulila et al (2021), urban expansion in Saudi Arabia was predicted using satellite images and CNN-LSTM models, and the proposed model was found to have the best performance compared to Pix2Pix and Dual GAN network models. In the study of Guo et al. (2021), air quality prediction was used by learning the spatiotemporal distribution of PM2.5 fine dust using the CNN-LSTM model. Comparison with other machine learning methods using Mean Absolute Error (MAE), Root Mean Squre Error (RMSE), and Mean Absolute Percentage Error (MAPE) showed the best performance. Judging from this, it can be seen that the CNN-LSTM methodology has excellent predictive power in making predictions by learning time series data of two-dimensional images.

2.3 Limitation of previous research

The differences presented in this study are as follows. First, in this study, an urbanization prediction model in the metropolitan area of Korea was constructed using the ConvLSTM model and land cover data. In the process of learning the model, the MultiInput model was combined to consider other influencing factors affecting urban expansion. Therefore, factors affecting urban expansion were considered more than the ConvLSTM model that had been previously carried out, and areas that had not been expanded to the city were sufficiently controlled and used for analysis. Second, this study presents the possibility of utilizing artificial intelligence for predicting urban expansion by utilizing ConvLSTM deep learning techniques that have better predictability and flexibility for data compared to existing analysis methods.

3. RESEARCH METHODOLOGY

3.1 Case study area

This study predicts urban sprawl by utilizing ConvLSTM for the spatial range of Seoul metropolitan area (SMA). The density of SMA in the world ranks as the sixth most densely population in the world and highly urbanized area with traffic, industrial factories(Kim et al., 2017; Sohn and Shim, 2010). The temporal range targets the 10-year cycles in 1980, 1990, 2000, and 2010, where land cover maps exist, and predicts urban expansion in 2020 and 2030.

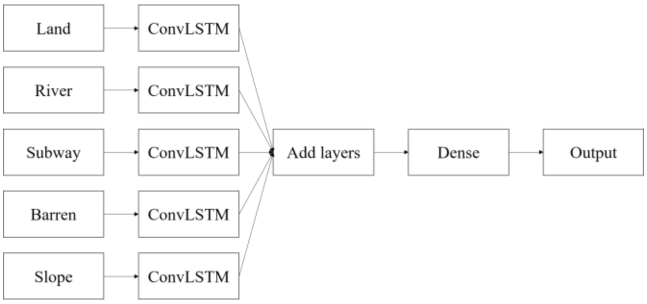


Figure 1. Multi-input ConvLSTM framework

3.2 Data sources

As the utilization data, the Land Coverage (Source: Ministry of Environment, Korea, Republic of) and de (SLEUTH) models, and predicted land use changes in 2040, in northern Iran. As a result of the analysis, most of the development in 2040 was distributed in the central region of the Majandaran region, and other road networks and topographical conditions resulted in poor growth. In this study, it was also argued that road network accessibility, development restriction zones, and building land density were decisive variables of land use change. Including the aforementioned land use prediction studies, the main variables that affect land use change examples mentioned in most previous studies can be seen as road network accessibility, subway station accessibility, altitude, and land use data were used. In the case of urbanization areas of the SMA, there were 2,573 in 1980 with 500x500 grid, 3,769 in 1990, 5,901 in 2000, and 6,596 in 2010.

In the case of urbanization areas, there is a continuously increasing trend. On the other hand, in the case of agricultural areas, it is gradually decreasing to 11,751, 10,590 in 1990, 10,668 in 2000, and 8,476 in 2010. In the case of Forest, it decreased to 21,890 in 1980 and 21,093 in 1990, and then increased to 21,633 in 2000 and 22,244 in 2010. In the case of Forest, the recent trend of increasing its area in each regional government due to the importance of green areas is believed to be the reason. In light of these points, it is judged that it is timely to predict the Urban Sprawl of SMA and present appropriate policy implications through this.

3.3 Methodology

In the case of the used model, ConvLSTM introduced by Shietal (2015) was used. In the case of ConvLSTM, it is known that it is one of the models that considered the Convolution Neural Network and the Recurrent Neural Network together and has the best performance among the homogeneous models. In this study, Multi-input ConvLSTM was configured and used to perform predictions in consideration of independent variables affecting the urban expansion. In the case of the used independent variables, as in Figure 1, there are river, land-use, distance to subway, slope, etc. The previous studies mentioned that the independent variables such as slope, land use, distance to subway affect sprawl phenomenon(Dadashpoor and Salarian, 2020).

4. ANALYSIS RESULTS

4.1 Model accuracy

Comparing the actual raster in the urban area in 2010 with the predicted raster, it can be seen that the predicted values in the northern, southern, and eastern areas are generally less than the actual values (Figure 1). Comparing the actual number with the predicted number, 163 unexpected rasters were found. Using this, the accuracy of the model was calculated, and the calculation equation is as shown in (1).

$$\frac{\text{Predict urbanized area}}{\text{Real urbanized area}} \times 100 \qquad (1)$$

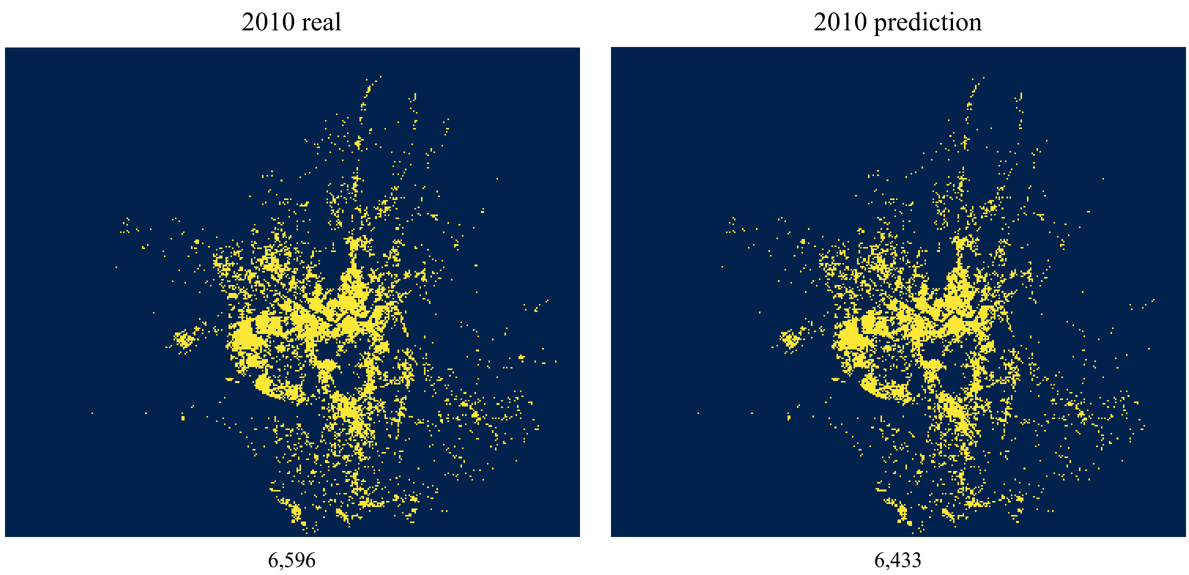


Figure 2. SMA urban region in 2010 (left: real, right: prediction)

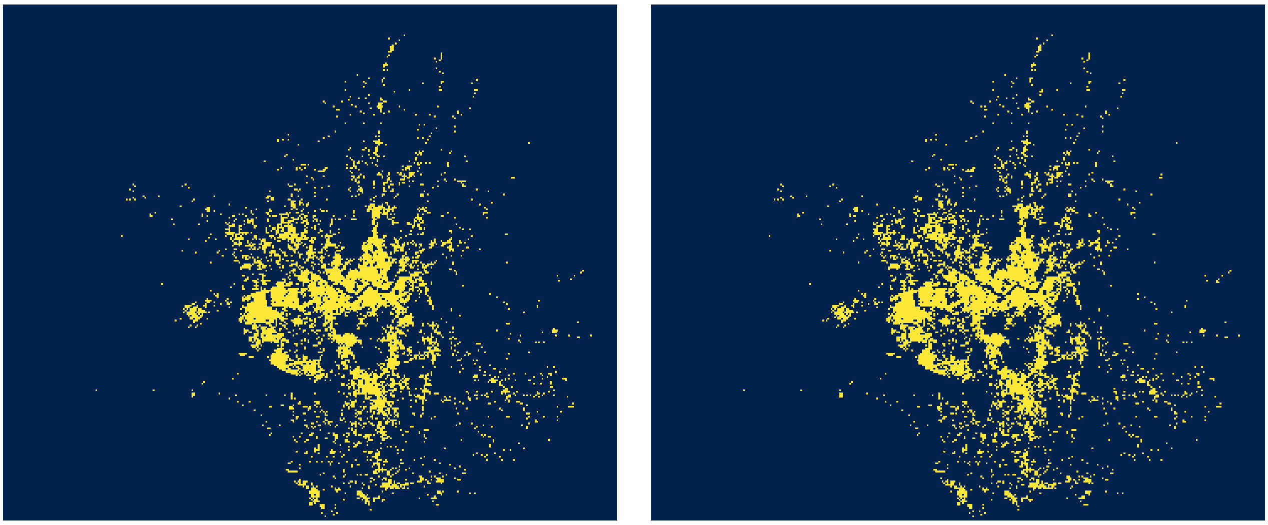
Accuracy was derived by dividing the area correctly predicted in the actual urbanized area. The total number of urban rasters in 2010 was 6,596 and 6,433 were correctly predicted. Finally,this model showed an accuracy of 97.5%.

4.2 Prediction

The values predicted for 2020 and 2030 with this model are as above. In 2020, 281 urbanization rasters increased compared to 2010, and in 2030, 518 urbanization rasters increased compared to 2010. In addition, urbanization raster in 2030 increased by 237 raster compared to 2020, mainly in the southern region(Figure 2). Compared to 2010, it can be seen that there are many urbanized areas centered on the southern region in 2020(Figure 3). In the case of urbanization in 2030, it can be seen that urbanization has progressed mainly in the southern region. In particular, compared to 2020, there are fewer areas where urbanization has progressed in the northern region, but in the southern region, it can be seen that urbanization has continued. Through this, it suggests that a policy is needed to prevent the sprawl phenomenon in the southern part of the metropolitan area in the future.

Comparing result with TerrSet2020 which was made by Clark University, the urban pattern is almost same. Especially Southern and Western part of SMA is expanded in 2030 compared to 2020 both of result. With the result of two model, policies need to be suggested to prevent and control the sprawl phenomenon in the SMA, Southern and Western part.

2020 prediction 2030 prediction



6,877 7,144

Figure 3. SMA urban expansion prediction (left: 2020, right: 2030)

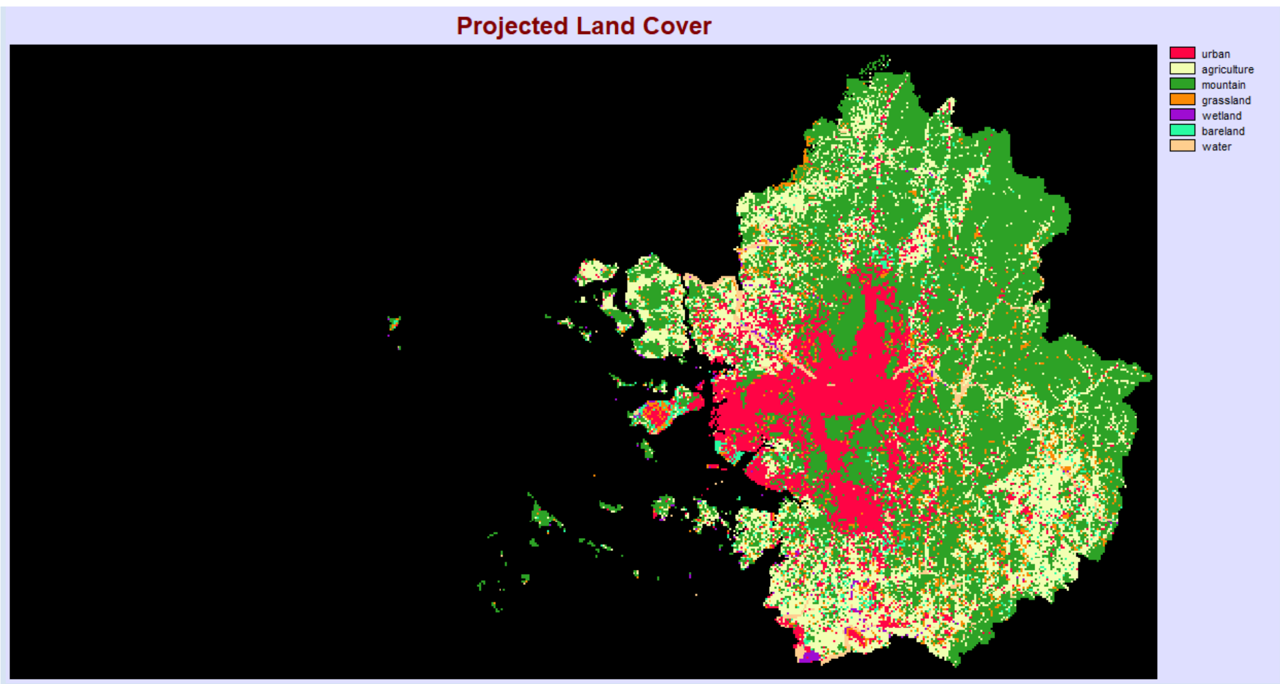


Figure 4. SMA urban expansion in 2020 by terrset 2020 program

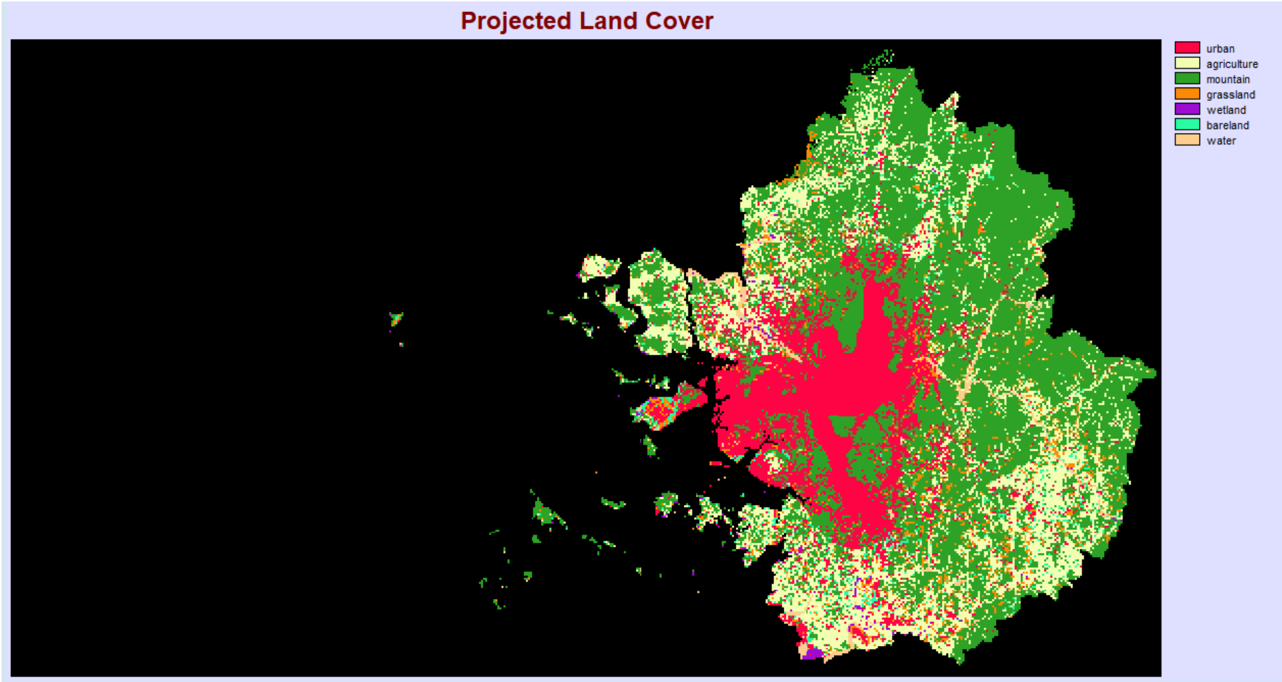


Figure 5. SMA urban expansion in 2030 by terrset 2020 program

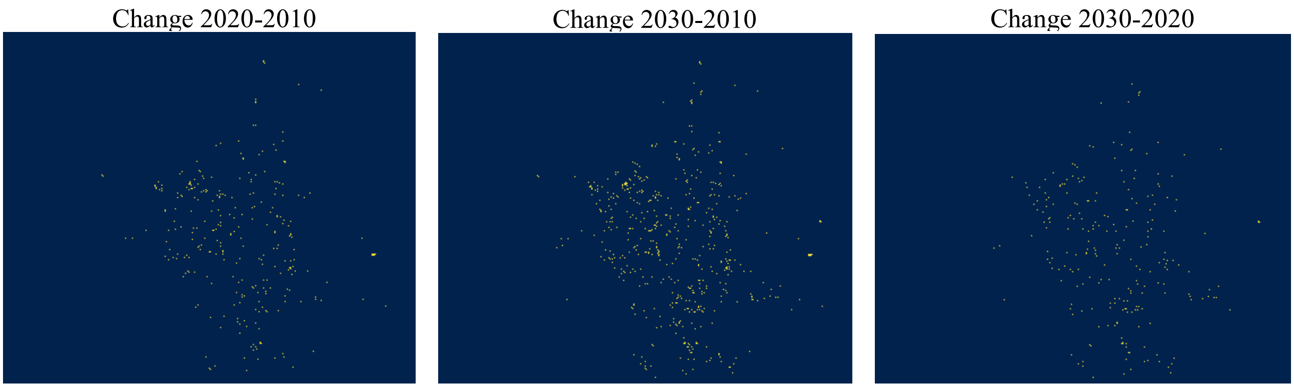


Figure 6. Comparing urban expansion

5. DISCUSSION AND CONCLUSION

The purpose of this study is to propose a ConvLSTM approach to predict urban expansion in the metropolitan area of Korea. As a result of reviewing previous studies, it was confirmed that ConvLSTM was suitable for Urban sprawl prediction, and the accuracy was high in this study. As a result of the analysis, it was generally predicted that urbanization would continue to occur in the overall area of SMA. In particular, in the case of 2030, unlike 2020, the urbanization phenomenon in the northern region is expected to decrease, and urbanization is expected to proceed around the southern region. Through this, this study suggests that there is a need for a policy that can prevent the urban sprawl phenomenon, centering on the southern part of SMA.

This study has the following limitations. First, there is a point that variables such as development restriction zones were not considered. This will be considered through data construction in the future to supplement the research. Second, there is a point that data on areas such as U.Sarmy base in SMA could not be established. However, despite these limitations, this study showed the possibility of urban sprawl prediction using artificial intelligence, suggesting that the negative impact of urban sprawl can be minimized.

REFERENCES

Bhatta, B., 2010. Causes and consequences of urban growth and sprawl. In *Analysis of Urban Growth and Sprawl from Remote Sensing Data*, 17-36.

Boulila, W., Ghandorh, H., Khan, M. A., Ahmed, F., Ahmad, J., 2021. A novel CNN-LSTM-based approach to predict urban expansion. Ecological Informatics, 64, 1-13.

Dadashpoor, H., Salarian, F., 2020. Urban sprawl on natural lands: Analyzing and predicting the trend of land use changes and sprawl in Mazandaran city region, Iran, *Environment, Development and Sustainability*, 22(2), 593-614.

Halmy, M. W. A., Gessler, P. E., Hicke, J. A., Salem, B. B., 2015.

Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. *Applied Geography*, 63, 101-112.

Jeon, H. J., Woo, M., 2019. A study on measuring urban sprawl and its policy implications for urban growth management and urban regeneration in Seoul capital region. *Journal of the Korean Regional Science Association*, 35(1), 3-18.

Jin, J. I., Jin, E. A., Lee, W. J., 2013. A study on the impacts of urban sprawl on the commuting pattern. *Journal of Korea Planning Association*, 48(5), 269-283.

Kim, G. H., Choi, H. S., Kim, D. B., Jung, Y. R., Jin, D. Y., 2020. Urban sprawl prediction in 2030 using decision tree. *Journal of the Korea Society of Environmental Restoration Technology*, 23(6), 125-135.

Kim, H. C., Kim, S., Kim, B. U., Jin, C. S., Hong, S., Park, R., Stein, A., 2017. Recent increase of surface particulate matter concentrations in the Seoul Metropolitan Area, Korea. *Scientific Reports*, 7(1), 1-7.

McDonald, R. I., Green, P., Balk, D., Fekete, B. M., Revenga, C., Todd, M., Montgomery, M., 2011. Urban growth, climate change, and freshwater availability. *Proceedings of the National Academy of Sciences*, 108(15), 6312-6317.

Pascanu, R., Gulcehre, C., Cho, K., Bengio, Y., 2013. How to construct deep recurrent neural networks. https://doi.org/10.48550/arXiv.1312.6026.

Rimal, B., Zhang, L., Keshtkar, H., Haack, B. N., Rijal, S., Zhang, P., 2018. Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS International Journal of Geo-Information*, 7(4), 1-21.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., Woo, W. C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 1-9.

Sibanda, S., Ahmed, F., 2021. Modelling historic and future land use/land cover changes and their impact on wetland area in Shashe sub-catchment, Zimbabwe. *Modeling Earth Systems and Environment*, 7(1), 57-70.

Sohn, K., Shim, H., 2010. Factors generating boardings at metro stations in the Seoul metropolitan area. *Cities*, 27(5), 358-368.

Song, X. P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., Townshend, J. R., 2018. Global land change from 1982 to 2016. *Nature*, 560(7720), 639-643.

Yadav, V., Ghosh, S. K., 2021. Assessment and prediction of urban growth for a mega-city using CA-Markov model. *Geocarto International*, 36(17), 1960-1992.