



Identification of vacant land in Tegal Regency using cnn algorithm based on goolge earth imagery

Wresti Andriani¹, Fatkhurrohman Fatkhurrohman², Gunawan Gunawan³

^{1,2,3}STMIK YMI TEGAL, Tegal City, Central Java, Indonesia

Article Info

Article history:

Received May 29, 2024

Revised Jun 04, 2024

Accepted Jun 13, 2024

Keywords:

Convolutional neural network;
Google earth imagery;
Image preprocessing;
Image sharing;
Vacant land identification.

ABSTRACT

This research developed a Convolutional Neural Network (CNN) algorithm to identify vacant land in Tegal Regency using imagery from Google Earth. By utilizing labeled imagery datasets, CNN models are optimized to recognize texture characteristics, colors, and distribution patterns of vacant land. Preprocessing and image sharing techniques are applied to improve model quality. The results of this study offer a new methodology in visual data processing for accurate and efficient identification of vacant land, providing a solid basis for more sustainable and efficient land use policies. This research contributes significantly to the scientific literature and field practice, particularly in natural resource management and regional planning.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Fatkhurrohman,
Information System,
STMIK YMI Tegal,
#1 Pendidikan Street, Tegal City, Central Java 52142, Indonesia.
Email : fatkhurrohman150602@gmail.com

Introduction

Effective use of vacant land is important in regional development and sustainable natural resource management (Liu et al., 2020). Areas that have the potential for untapped vacant land require the use of advanced technology to identify and manage land effectively (Bibri, 2020). Identification of vacant land is critical in the context of regional development and sustainable natural resource management because it allows for more efficient spatial planning, avoids indiscriminate and potentially destructive land use, and supports sustainable economic and social growth. Poorly managed land often becomes an illegal landfill, causing soil and water pollution, as well as erosion and loss of topsoil which is important if the land is open and not covered with vegetation. With the development of remote sensing technology and advances in the field of artificial intelligence, there is new potential in the use of satellite imagery for land identification (Zhang, Wu, et al., 2022). Google Earth imagery provides high resolution that allows detailed observation of the texture, color, and distribution patterns of the land that play a crucial role in the identification of vacant land. For CNN is an artificial intelligence architecture that is highly effective in understanding the visual features of complex images, making it particularly suitable for pattern recognition tasks such as the identification of vacant land. With these capabilities, the results of this research can make a significant contribution to real-world applications, such as more efficient land planning, support for conservation and sustainability, and sustainable development.

This approach ensures stable data processing despite potential data access issues (Abbas et al., 2024). Many studies have applied remote sensing technology in land identification, but there are still

challenges in the accuracy and efficiency of vacant land identification (Z. Li et al., 2022). This issue is important for proper mapping in supporting development and land management decisions (Ayhan et al., 2020). Therefore, the purpose of this study is to address the shortcomings of previous studies by using CNNs that can understand the features of satellite imagery in detail and produce more precise land identification (W. Li & Hsu, 2020). The use of CNNs based on Google Earth imagery offers new possibilities in processing visual data for efficient and accurate identification of vacant land (Giuffrida et al., 2021).

This research will be combined with preprocessing and image sharing techniques to improve the quality of the models used (Cai et al., 2021). The novelty lies in CNN's algorithm for identifying vacant land using Google Earth imagery, which has not been widely explored in the literature. The success of this approach could pave the way for further development in the application of artificial intelligence technology for land management and regional development in Indonesia.

Previous studies estimated land use using the Support Vector Machine algorithm, with results showing an increase in urban built-up areas and higher surface temperatures, and emphasizing the need for urban master plans that focus on urban plantations and natural resource conservation to reduce the impact of the urban heat island phenomenon (SUHI) (Kafy et al., 2021). The next study, using a bibliometric analysis method to define urban vacant land (UVL), showed a UVL probability of 8.46%-8.88% in Guangzhou and 3.17%-5.08% in New York, and an average housing vacancy rate of 11.48% in 65 American cities. UVL is often fragmented and irregular, caused by factors such as soil overdistribution, population decline, and soil speculation (Song et al., 2020). Another study, Deep Convolutional Neural Network Architecture and the latest innovations in its development used a survey method to classify innovations in CNN architecture into seven categories based on various aspects such as spatial exploitation, depth, multipath, width, feature map exploitation, channel enhancement, and the results of the research provide the latest CNN architecture taxonomy, identify key components, challenges faced, and CNN applications in various fields such as computer vision, natural language processing, and speech recognition (Khan et al., 2020). Further research, the use of Convolutional Neural Networks to detect and identify impaired drivers through image analysis with the results showing that this model managed to achieve an average accuracy of 99% in interference recognition (Masood et al., 2020). Related research includes the development of Automatic Bee Identification System (ABIS) for automatic monitoring and identification of insect species using methods using deep convolutional networks with results showing high identification accuracy, with an average accuracy of top-1 of 93.95% and top-5 of 99.61% in ABIS project data, as well as an increase in accuracy in butterfly data with top-1 accuracy of 96.72% and top-5 of 99.99% (Buschbacher et al., 2020).

This study differentiates from previous research by introducing a Convolutional Neural Network (CNN) architecture that is specifically optimized for Google Earth image processing in identifying vacant land. Therefore, this study intends to further explore the potential of Deep Learning architecture, especially CNN, in identifying and mapping urban vacant land more accurately and efficiently (Addesso et al., 2021). The purpose of this research is to develop a model that is able to integrate urban environmental data, provide insights for more effective and sustainable urban planning, and support environmental conservation initiatives (Ahmad et al., 2022).

Method

Research Flow

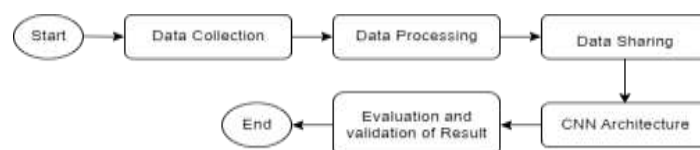


Figure 1. Research flow

The research flow begins with the process of data collection, namely data collection. Every data obtained goes to the Data Processing stage, at this stage the image will undergo resizing and data

sharing for research needs. Furthermore, Citra was used to train network models on the CNN algorithm. Once the model is trained, evaluate its performance using metrics such as accuracy, precision, recall, and F1-score. Finally, perform in-depth validation of the results obtained from the model.

Data Collection



Figure 2. Google eath pro app

Data collection utilizes sniping tools in the Google Earth Pro application, from various locations of vacant and non-vacant land in Tegal Regency. This process involves finding specific coordinates based on predetermined locations, ensuring that each image taken represents a different geographical condition. Each image is then stored at high resolution to ensure that important details from each location are captured clearly, then used as input data to train the Convolutional Neural Network (CNN) model to identify vacant land.



Figure 3. Multiple datasets

First, the desired area is identified and enlarged to obtain the necessary details. Secondly, the Snipping Tool is used to take a screenshot of the area. This method allows researchers to selectively collect images of specific vacant lots, which are then stored as visual data. Each screenshot is taken to ensure that coverage and geographic orientation remain consistent across all images, making comparative analysis between images easier. Snipping Tool can quickly and efficiently collect images without the need for additional software to crop or edit (Jun et al., 2021). The above dataset has not been resized, so it will be changed to create a uniform dataset and train the CNN model.

Data Processing

In this study, the dataset used consisted of two folders, each of which contained images of vacant land and not vacant land. To ensure the quality of the data used, each image is inspected and cleaned of anomalies or irrelevant images. After that, the images are resized to 400x400 pixels to ensure uniformity and facilitate the analysis process.



Figure 4. Data that has been processed

This research develops a function to automate the retrieval of images from predefined directories, utilizing information stored in XML files contained in each category of image directories. The use of the ElementTree library allows an efficient process of parsing XML files for the extraction of necessary data, such as image file locations (Zhang, Huang, et al., 2022). Next, the obtained image is processed using the load_img function of TensorFlow Hard, which allows us to load and convert the image into an array. Image size is adjusted to 400x400 pixels to ensure consistency in neural network model input. During the image capture and processing process, we also implemented a robust error handling mechanism. This function effectively identifies and resolves situations where directories or image files are not found, by logging error messages and continuing the process without stopping the entire operation.

Data Sharing

The dataset used consists of images of vacant land and used land that has gone through the pre-processing stage. The dataset is divided into two manually, namely 50% for vacant land data and 50% for non-vacant land data. Each category is then further divided into training sets and test sets, 80% for training the model and 20% for testing the model's performance. Manual sharing of data is done to ensure that each category is represented proportionally in the training and testing process.



Figure 5. Data on vacant land and not vacant land

The data is divided into two categories based on images of vacant land and not vacant land, then put into their respective folders according to their categories. Each image is named according to its category to facilitate the rest of the process.

CNN Architecture

The convolutional neural network architecture used in this study consists of 4 convolution layers + relu, 4 maxpooling layer and 4 fully connected layer. Here is the architecture used:

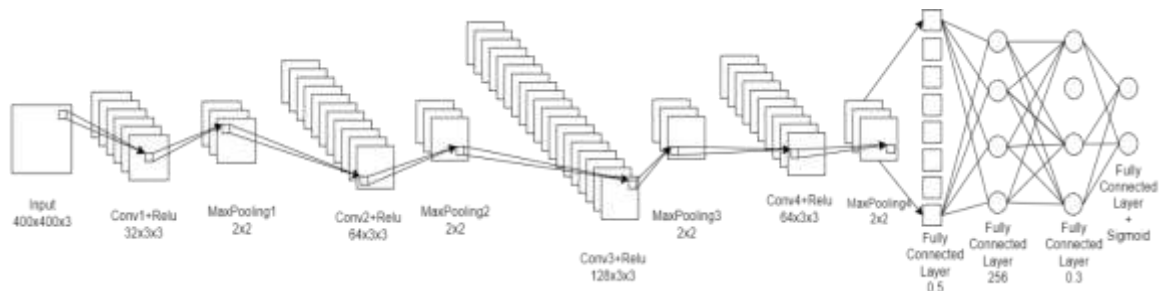


Figure 6. CNN Architecture

First, the 500x500 input image is processed using the first convolution layer which has a 3x3 kernel with 32 filters. After that, the first max pooling operation is performed with a 2x2 kernel to reduce the spatial dimensions of the feature map. Next, a second convolution layer is applied with a 3x3 kernel and 64 filters, followed by a second max pooling which also uses a 2x2 kernel. This process continues to the third convolution layer which has a 3x3 kernel with 128 filters, then followed by the third max pooling with a 2x2 kernel. Finally, the fourth convolution layer uses a 3x3 kernel with 64 filters, followed by a fourth max pooling with a 2x2 kernel. This process aims to gradually extract more complex features from the image and reduce the data size to increase computational efficiency.

Evaluation and Validation of Results

In this study, the Convolutional Neural Network (CNN) model was developed to identify vacant land using a dataset that has been labeled, consisting of two categories of vacant land and non-vacant land. Data preprocessing and separation techniques are applied to improve the quality of the model. After the training phase, the model's performance is evaluated using different data to ensure that the model can generalize its knowledge to new data that was not involved during training. Performance evaluations are conducted using standard metrics such as accuracy, precision, recall, and F1 scores, providing a comprehensive picture of the model's effectiveness in classifying vacant land. This validation process is designed to verify that the model is not only effective on training data but also demonstrates high effectiveness and applicability under diverse conditions.

$$Accuracy = \frac{TP + TN}{Total\ number\ of\ samples} \quad (1)$$

The model shows high accuracy if the value is close to 100%. This shows the effectiveness of the model in classifying the image (Ahsan et al., 2021).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

High precision means that when the model predicts an image as complete, it is most likely that the prediction is accurate (Kamir et al., 2020).

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

High recall indicates that the model is effective in capturing most positive cases (LeBien et al., 2020) v .

$$F1\ Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

F1 scores are important in the context of evaluating classification models, particularly in prioritizing a balance between precision and gain (Maxwell et al., 2021).

Results and Discussion

Table 1. Validation results

Validation	Epoch	Time	Practice accuracy	Validation accuracy	Training Loss	Loss Validation
Validation 1	10	31 seconds	0.67	0.89	0.88	0.87
Validation 2	10	30 seconds	0.71	0.82	0.71	0.67
Validation 3	10	49 seconds	0.73	0.82	0.77	0.77

In Table 1, from these results it can be seen that the highest exercise accuracy value achieved was 0.71, and the highest validation accuracy value achieved was 0.89. The achieved exercise loss value was 0.088, while the validation loss value achieved was 0.87.

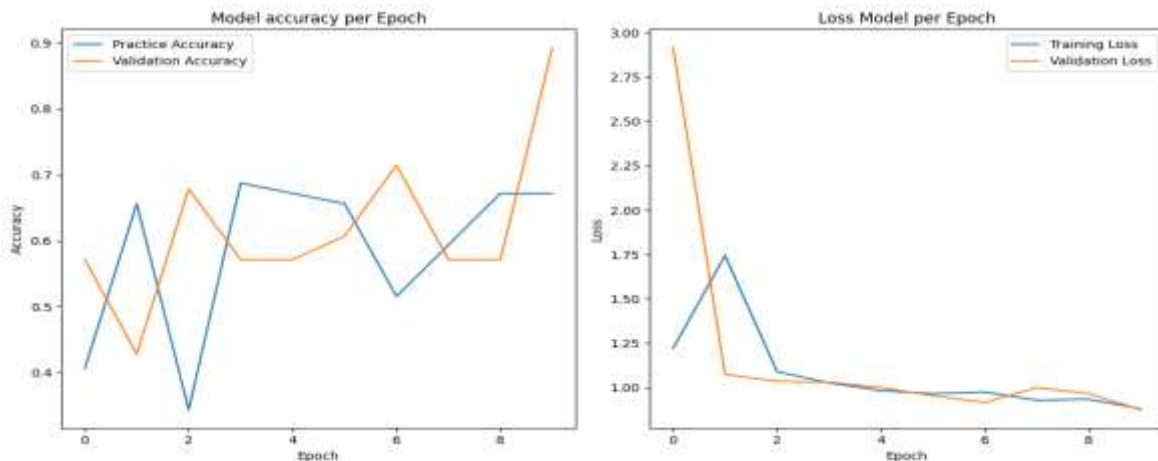


Figure 7. Validation results 1

The results of model accuracy and model loss at epoch 10 in the first validation can be seen in figure 4. From these results, the training accuracy value achieved was 0.67. For validation accuracy, the value achieved is 0.89. The training loss value achieved is 0.88, while the validation loss achieved is 0.87.

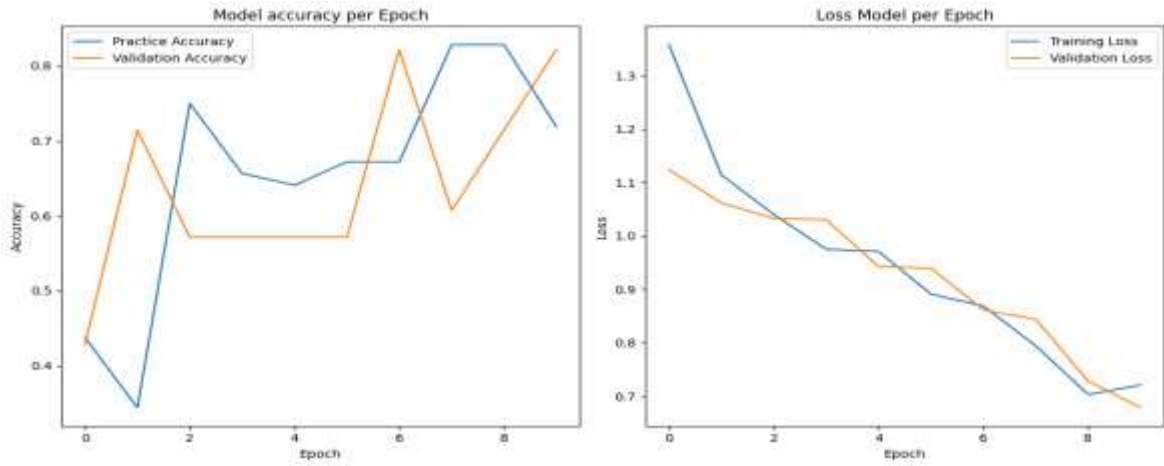


Figure 8. Validation results 2

The results of model accuracy and model loss at epoch 10 in the second validation can be seen in figure 2. From these results, the training accuracy value achieved was 0.71. For validation accuracy, the value achieved is 0.82. The training loss value achieved was 0.71, while the highest validation loss achieved was 0.67.

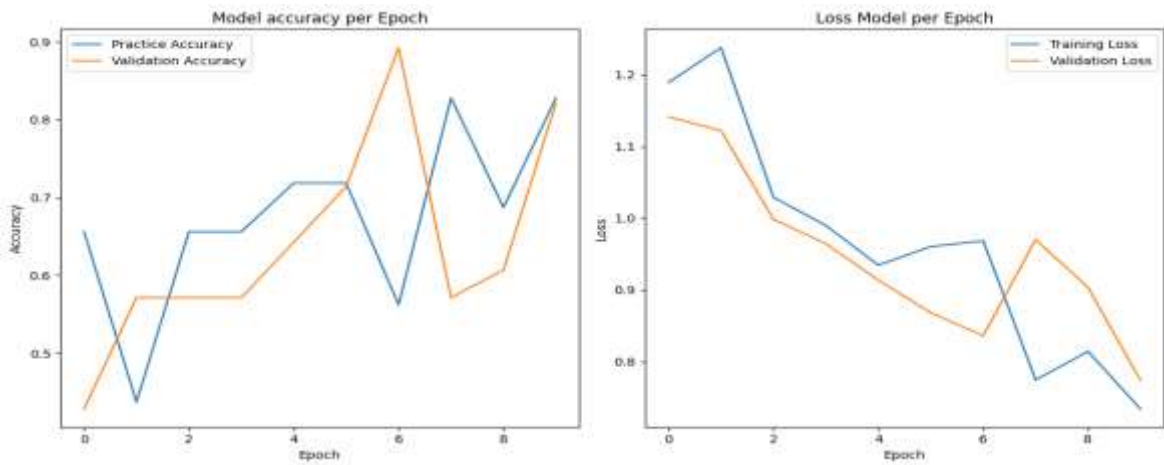


Figure 9. Validation results 3

The results of model accuracy and model loss at epoch 10 in the second validation can be seen in figure 2. From these results, the training accuracy value achieved was 0.73. For validation accuracy, the value achieved is 0.82. The training loss value achieved was 0.82, while the highest validation loss achieved was 0.77.

Table 2. Matrix Results

No.	Matrik	Value
1.	Accuracy	82.14%
2.	Precision	70.59%
3.	Recall	100.00%
4.	F1 Score	82.76%

The matrix results show that the developed CNN model can identify vacant land with high accuracy with an accuracy value of 82.14%, precision of 70.59%, recall of 100%, and F1 score of 82.76%, compared to previous research using the Support Vector Machine (SVM) algorithm. This study

provides more specific results in identifying the characteristics of vacant land based on complex visual features. While the results are encouraging, there is significant room for improvement, especially in addressing areas of land that are more difficult for the system to identify. This development is important to ensure that all types of land, especially those that have visual features like other types of land, can be classified more accurately. The performance of this system not only demonstrates its possible application in the land management and urban planning industries, where a quick and accurate understanding of land types can add value, but also in the context of environmental conservation, where technology can be used to help identify and document different types of land automatically and efficiently.

Conclusions

The results of this study support decision-making and urban planning in Tegal Regency by developing a Convolutional Neural Network (CNN) algorithm to identify vacant land efficiently and accurately using Google Earth imagery. By applying this technology, the government and developers in Tegal Regency can more easily determine the most suitable location for infrastructure development and urbanization projects. These algorithms also help in optimizing land use, preventing waste of resources, and directing development in a more sustainable direction. Through innovative data preprocessing and labeling techniques, this research has succeeded in achieving high performance in recognizing the visual characteristics of vacant land. Thus, the results of this study make an important contribution to the scientific literature and field practice in natural resource management and urban planning, demonstrating the potential use of CNNs in new contexts. Although successful, this study has limitations on certain geographical coverage and certain types of land. For this reason, future research needs to overcome these limitations by testing models in more diverse locations and improving the model's ability to recognize different types of land. This step will pave the way for further development in artificial intelligence applications for more effective and sustainable land management and regional planning. This research proves the potential of the Convolutional Neural Network (CNN) algorithm in identifying vacant land through Google Earth imagery, which is not only relevant for Indonesia but can also be applied globally. Its implications for more effective natural resource management, more structured urban planning, and adaptation to climate change, show significant contributions in various fields. However, there are several risks and limitations to consider, including potential inaccuracies in satellite data, risk of bias in AI data and models, and data privacy and security concerns. Therefore, further research is urgently needed to improve the generalization of the model to different land types and geographical conditions, as well as to improve the accuracy and reliability of the algorithm so that it can be applied more safely and effectively in land management and regional planning practices.

References

- Abbas, A., Alroobaea, R., Krichen, M., Rubaiee, S., Vimal, S., & Almansour, F. M. (2024). Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things. *Personal and Ubiquitous Computing*, 28(1), 59–72. <https://doi.org/10.1007/s00779-021-01626-0>
- Addesso, P., Restaino, R., & Vivone, G. (2021). An improved version of the generalized Laplacian pyramid algorithm for pansharpening. *Remote Sensing*, 13(17), 1–20. <https://doi.org/10.3390/rs13173386>
- Ahmad, T., Madonski, R., Zhang, D., Huang, C., & Mujeeb, A. (2022). Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renewable and Sustainable Energy Reviews*, 160, 112128. <https://doi.org/10.1016/j.rser.2022.112128>
- Ahsan, M., Based, M. A., Haider, J., & Kowalski, M. (2021). COVID-19 detection from chest X-ray images using feature fusion and deep learning. *Sensors*, 21(4), 1480. <https://doi.org/10.3390/s21041480>
- Ayhan, Ç. K., Taşlı, T. C., Özkök, F., & Tatlı, H. (2020). Land use suitability analysis of rural tourism activities: Yenice, Turkey. *Tourism Management*, 76, 103949. <https://doi.org/10.1016/j.tourman.2019.07.003>
- Bibri, S. E. (2020). Compact urbanism and the synergic potential of its integration with data-driven smart urbanism: An extensive interdisciplinary literature review. *Land Use Policy*, 97, 104703. <https://doi.org/10.1016/j.landusepol.2020.104703>
- Buschbacher, K., Ahrens, D., Espeland, M., & Steinhage, V. (2020). Image-based species identification of wild bees using convolutional neural networks. *Ecological Informatics*, 55, 101017.

- <https://doi.org/10.1016/j.ecoinf.2019.101017>
- Cai, L. Y., Yang, Q., Hansen, C. B., Nath, V., Ramadass, K., Johnson, G. W., Conrad, B. N., Boyd, B. D., Begnoche, J. P., & Beason-Held, L. L. (2021). PreQual: An automated pipeline for integrated preprocessing and quality assurance of diffusion weighted MRI images. *Magnetic Resonance in Medicine*, *86*(1), 456–470. <https://doi.org/10.1002/mrm.28678>
- Giuffrida, G., Fanucci, L., Meoni, G., Batič, M., Buckley, L., Dunne, A., Van Dijk, C., Esposito, M., Hefele, J., & Vercruyssen, N. (2021). The Φ -Sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–14. <https://doi.org/10.1109/TGRS.2021.3125567>
- Jun, J., Kim, J., Seol, J., Kim, J., & Son, H. Il. (2021). Towards an efficient tomato harvesting robot: 3d perception, manipulation, and end-effector. *IEEE Access*, *9*, 17631–17640. <https://doi.org/10.1109/ACCESS.2021.3052240>
- Kafy, A. Al, Faisal, A. Al, Shuvo, R. M., Naim, M. N. H., Sikdar, M. S., Chowdhury, R. R., Islam, M. A., Sarker, M. H. S., Khan, M. H. H., & Kona, M. A. (2021). Remote sensing approach to simulate the land use/land cover and seasonal land surface temperature change using machine learning algorithms in a fastest-growing megacity of Bangladesh. *Remote Sensing Applications: Society and Environment*, *21*(August 2020), 100463. <https://doi.org/10.1016/j.rsase.2020.100463>
- Kamir, E., Waldner, F., & Hochman, Z. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 124–135. <https://doi.org/10.1016/j.isprsjprs.2019.11.008>
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*, 5455–5516. <https://doi.org/10.1007/s10462-020-09825-6>
- LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J. P., Dodhia, R., Ferres, J. L., & Aide, T. M. (2020). A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. *Ecological Informatics*, *59*, 101113. <https://doi.org/10.1016/j.ecoinf.2020.101113>
- Li, W., & Hsu, C.-Y. (2020). Automated terrain feature identification from remote sensing imagery: a deep learning approach. *International Journal of Geographical Information Science*, *34*(4), 637–660. <https://doi.org/10.1080/13658816.2018.1542697>
- Li, Z., Shen, H., Weng, Q., Zhang, Y., Dou, P., & Zhang, L. (2022). Cloud and cloud shadow detection for optical satellite imagery: Features, algorithms, validation, and prospects. *ISPRS Journal of Photogrammetry and Remote Sensing*, *188*, 89–108. <https://doi.org/10.1016/j.isprsjprs.2022.03.020>
- Liu, J., Jin, X., Xu, W., Gu, Z., Yang, X., Ren, J., Fan, Y., & Zhou, Y. (2020). A new framework of land use efficiency for the coordination among food, economy and ecology in regional development. *Science of the Total Environment*, *710*, 135670. <https://doi.org/10.1016/j.scitotenv.2019.135670>
- Masood, S., Rai, A., Aggarwal, A., Doja, M. N., & Ahmad, M. (2020). Detecting distraction of drivers using convolutional neural network. *Pattern Recognition Letters*, *139*, 79–85. <https://doi.org/10.1016/j.patrec.2017.12.023>
- Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: Literature review. *Remote Sensing*, *13*(13), 2450. <https://doi.org/10.3390/rs13132450>
- Song, X., Wen, M., Shen, Y., Feng, Q., Xiang, J., Zhang, W., Zhao, G., & Wu, Z. (2020). Urban vacant land in growing urbanization: An international review. *Journal of Geographical Sciences*, *30*, 669–687. <https://doi.org/10.1007/s11442-020-1749-0>
- Vargas-Munoz, J. E., Srivastava, S., Tuia, D., & Falcao, A. X. (2020). OpenStreetMap: Challenges and opportunities in machine learning and remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, *9*(1), 184–199. <https://doi.org/10.1109/MGRS.2020.2994107>
- Zhang, B., Huang, S., Zhang, L., Liu, X., Song, G., Qin, J., & Li, M. (2022). Method of railway shunting operation sheet information extraction guided by table header. *IET Intelligent Transport Systems*, *16*(10), 1391–1403. <https://doi.org/10.1049/itr2.12213>
- Zhang, B., Wu, Y., Zhao, B., Chanussot, J., Hong, D., Yao, J., & Gao, L. (2022). Progress and challenges in intelligent remote sensing satellite systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 1814–1822. <https://doi.org/https://doi.org/10.1109/JSTARS.2022.3148139>