https://doi.org/10.29227/IM-2025-01-02-011 Received: 30 Oct 2024, Accepted: 01 May 2025, Published: 01 Sep 2025

Investigation of underground anomaly by application of Convolutional neural network for Ground penetrating radar data analysis

LE Van Anh Cuong[1,2*]

[1] University of Science, Ho Chi Minh City, Vietnam
[2] Vietnam National University, Ho Chi Minh City, Vietnam
[*] *Corresponding email: lvacuong@hcmus.edu.vn*

Abstract: *Ground penetrating radar method is widely known for its effectiveness in investigation of underground civil engineering structures. They could be represented by the high-frequency electromagnetic signals as reflection or diffraction events in the measurement data slices. Meaningful high-frequency electromagnetic wave signals are reflected/scattered from the underground objects. Requirement of fast locating the underground anomalies was an inspiration to apply modern technology of artificial intelligence in the ground penetrating radar data analysis. There is suggested a novel workflow for detecting diffracted signals which uses the convolution neural network for this research paper. The real high-frequency datasets measured in the Ho Chi Minh City area, Vietnam are used as training, testing, and validating datasets for building a convolutional neural network model. The measured data in Nguyen Van Cu Street area, District 5, Ho Chi Minh City, Vietnam is predicted with the network model for the high accuracy result.*

Keywords: *Ground penetrating radar, GPR, Convolutional Neural Network, CNN, High-frequency electromagnetic data, anomaly.*

1. Introduction

In investigation of shallow geology structures and underground civil engineering constructions, especially in big cities, Ground Penetrating Radar (GPR) method has greater advantages than other geophysical ones in abilities of quick high resolution data collection, and artificial noise removing integrated in its equipment.

Ground Penetrating Radar (GPR) data interpretation needs previous data analysis process that helps to enhance signal over noise ratio. The data analysis process includes noise removal, velocity analysis for time to depth conversion and underground object detection using special signature of the electromagnetic events such as strong hyperbolae. Its physics behind is to follow laws of high frequency electromagnetic wave propagation, reflection, diffraction and energy conservation (Tzanis, 2010, Nguyen et al., 2017, Le and Nguyen, 2020, Sandmeier, 2020). Normally, in the research, underground singular objects can be detected through visible diffraction hyperbole recorded in the GPR slices.

High technology such as Artificial Intelligence (AI) can boost GPR interpretation in new successful levels. That is, traditionally analysing a big amount of thousands of kilometres GPR data profiles can cost a lot of time and enormous efforts of experts as geophysicists and geologists. Therefore, many papers focus on how to apply AI into GPR processing and interpreting (Pham and Lefèvre, 2018, Kang et al., 2019, Kang et al., 2020, Wang et al., 2021) to remedy the drawbacks and still achieve better GPR interpretation results.

In the paper the Convolutional Neural Network (CNN), model having two stages has been designed in the following steps:

(i) Processing data for making data collection,

(ii) Establishing CNN model with the prior data collection.

In this research, real GPR data measurements were set up for preparing GPR data collection prior application of Convolutional Neural Networks (CNN). The successful CNN model was trained, tested, validated with three different GPR datasets. Then, one real GPR data example in Nguyen Van Cu Street, District 5, Ho Chi Minh city was processed by the AI model. The suggested workflow was expressed in Fig. 1.

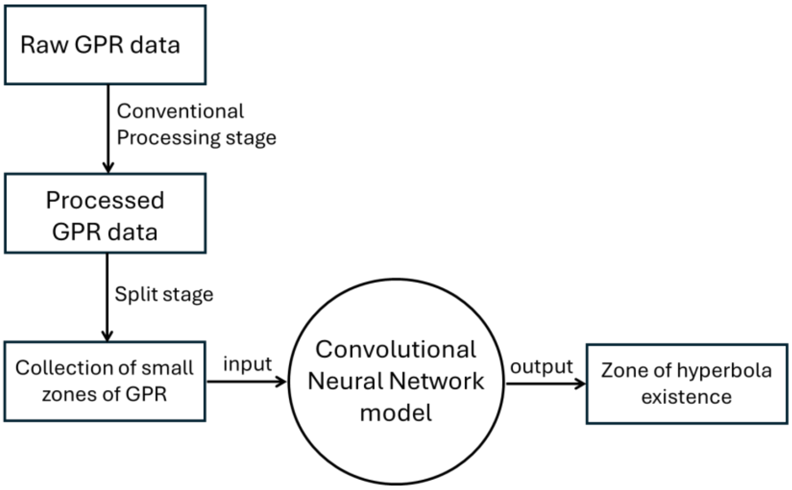


Fig. 1. GPR interpretation workflow using CNN application

The raw GPR data were traditionally automatically processed and split into a collection of small zones. Then, the collection was delivered into the CNN model for having confirmation of having existence of hyperbola. Finally, the GPR processed data can contain information of underground anomaly thanks to the output of the CNN application.

2. Methodology

Building a suitable CNN GPR model needs two important factors as big data collection and a suitable AI configuration. Firstly, the data collection is formed by the processed GPR data. Secondly, the AI configuration is referenced from MATLAB software (MathWorks, 2019f, MathWorks, 2019g) and integrated with its useful options to reduce overfit issue.

2.1. Data preparation

The raw GPR data is gone through an automatic traditional workflow for having processed interpretable data. The raw data were collected using the Detector Duo GPR machine made in Italy (IDS, 2010). The automatic processed workflow consisted of several analysis steps as moving start-time, frequency band filter, and decay gain (Tzanis, 2010, Nguyen et al., 2017, Le and Nguyen, 2020, Sandmeier, 2020). The start-time option is to change start-time value to the defined value of the user. For frequency band filter, the frequency band consists of their frequency min and max values as $^1/_2$ and $^3/_2$ of the main frequency (i.e., 700 MHz), respectively. For the decay gain, the data is compensated for loss energy in which its weak signals can be visible in the deep time. For recognition of an underground anomaly as water supplying or drainage water pipes, hyperbolic shapes in GPR slices were detected (Figures 2 and 3).

https://doi.org/10.29227/IM-2025-01-02-011 Received: 30 Oct 2024, Accepted: 01 May 2025, Published: 01 Sep 2025

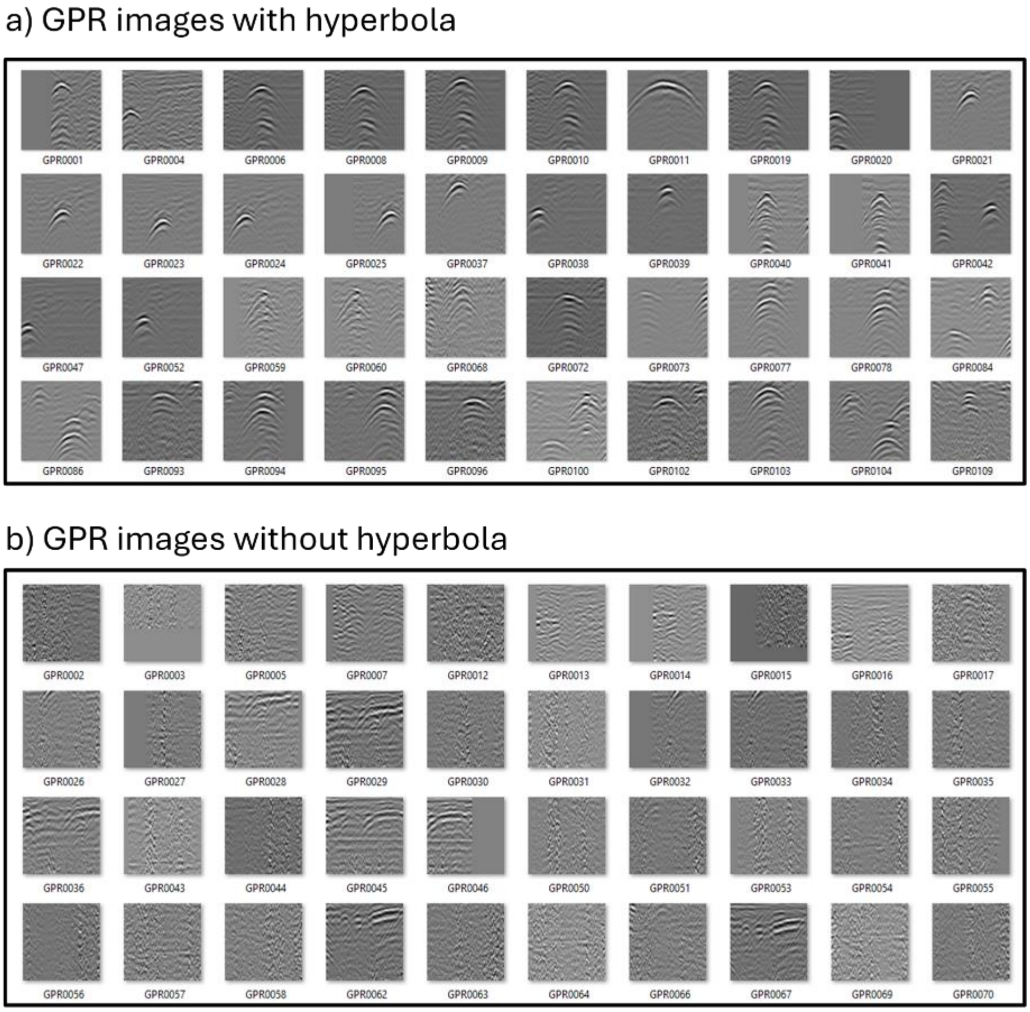


Fig. 2. Some of GPR images used for training. a) GPR images consist of one or many diffraction hyperbolae. b) GPR images do not have diffraction hyperbolae

In data preparation, collection of GPR data images can be separated into two categories as having hyperbolic and non-hyperbolic shapes. For a measured 2D GPR data line, its processed data can be split into different smaller images with their 100x100 pixel sizes. All the images were then rescaled to the [0,1] band which can help the network to understand the data distribution during the training stage.

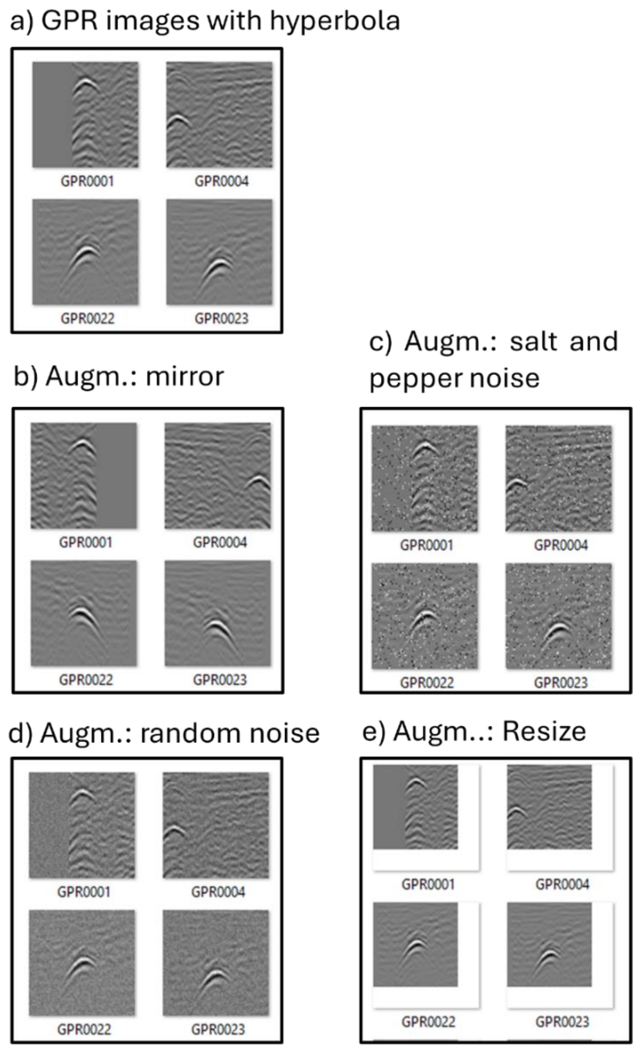


Fig. 3. Augmentation (Augm.) of selected images

A big GPR data collection which can prevent the overfitting issue is crucial to build an effective neural network model (Kang et al., 2020). The data collection for the CNN workflow (Figure 4) includes (i) the images originally extracted the processed GPR data and (ii) the augmented images which are modified from the previous images (Aquino et al., 2017, Hernández-García and König, 2018, Fonseka and Chrysoulas, 2020). The deep neural network models are heavily dependent on big data sets for improving accuracy prediction. By solving the problem of limited datasets, data augmentation can increase number of data input from making new versions of existing data images (Aquino et al., 2017, Hernández-García and König, 2018, Fonseka and Chrysoulas, 2020). For this study, the augmented images also need keep key hyperbolic shapes through four data making approaches as mirror transformation, addition of "salt and pepper" noise, addition of random noise, and resize transformation (Figure 3) (MathWorks, 2019b, MathWorks, 2019c, MathWorks, 2019d, MathWorks, 2019e).

2.2. Training the data

Modelling operator: Neural networks can show existence of underground anomalies through recognizing hyperbolae in GPR images. When a GPR image as an input is transferred through its node layers, the neural networks as the modelling operator can answer whether there is a hyperbola or not. Note that for CNN approach, the GPR images seen as an 2D matrices (i.e., each matrix size is 100 rows x 100 columns) are extracted from 2D field trip GPR datasets.

Inversion: Achieving the parameters, weigh w and bias b, in each layer of the CNN model is to deal with the minimization of the loss function (Perol et al., 2018, Amidi and Amidi, 2018, Li et al., 2021) considering as minimization of the objective function when solving an inversion problem (Siripunvaraporn et al., 2005, Berdichevsky and Dmitriev, 2008, Le et al., 2016, Le et al., 2019). The problem consists of two components for building a loss function as data input and layer configuration. For data input **x**, each GPR image is seen as an 2D matrix (100 rows x 100 columns) input for the inversion. Each image can be extracted from the GPR diagrams from field trip. For the whole training, testing and validation processes, the data has only one channel as grey-white colour.

Layer configuration can consist of many key components as follows:

(i) Convolutional layers (LeCun et al., 2015): Each convolution layer helps to extract significant features from the input data (i.e., image boundaries). An input data is performed through a dot product operation for its whole space domain. Its output can exhibit specific patterns.

(ii) Batch normalization (Ioffe and Szegedy, 2015, MathWorks, 2019a): the process is to normalize the parameters from a layer of activation in training a batch of data examples. It supports to stabilize training, to accelerate convergence, and to enhance generalization.

(iii) The rectified linear unit (reLu) (LeCun et al., 2015): the function supports the network to learn complicated patterns by replacing all negative values with zeroes and keeping positive values.

(iv) Max pooling (LeCun et al., 2015): the analysis step reduces size of the data input by selecting the maximum values representing within a local region of the data.

(v) Fully Connected layer (LeCun et al., 2015, Kang et al., 2019): This step connects all neurons in a layer to all neurons in its next layer.

(vi) drop-out layer (Srivastava et al., 2014, LeCun et al., 2015, Kang et al., 2019): The action can support the network to learn more robust attributes of the data input.

(vii) softmax (Kinga and Adam, 2015, Perol et al., 2018, Kang et al., 2019): The step transforms the output of the final fully connected layer into a probability value of group of desired output in which each data input falls.

The loss function so-called as the cross-entropy loss can verify the difference between our predicted probability and the true class probability distribution for the two categories as hyperbola and no hyperbola (Perol et al., 2018). Optimization of the loss function is to achieve the model weights and bias parameters after many iterations by using ADAM algorithm (Kinga and Adam, 2015, Perol et al., 2018).

The workflow of CNN processing is presented in Fig. 4. A 2D GPR image was gone through the network layers to determine the presence of hyperbolic diffraction.

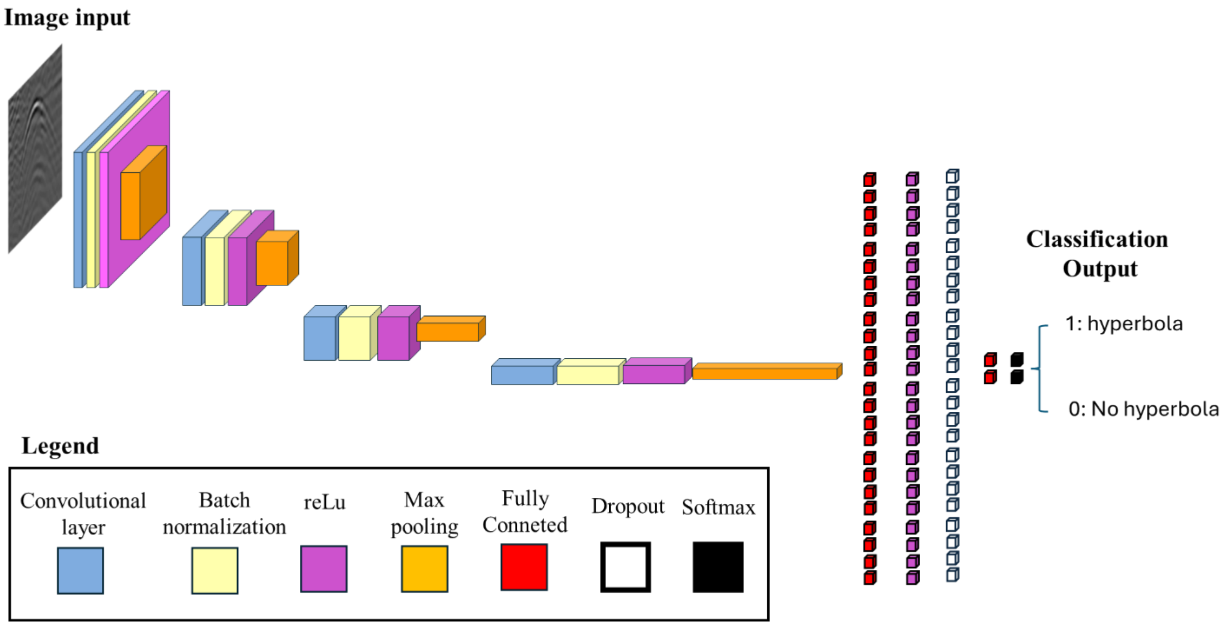


Fig. 4. Workflow of CNN approach for detecting presence of hyperbola in a 2D GPR image

3. Results

3.1. Analysing the modelled CNN result after training

Training the CNN model needs a big data of GPR images. There were used three independent types of collections (i) Training dataset, (ii) Validation dataset, and (iii) Test data. After 10 epochs, both the results of validation and training are higher than 90%, which is suitable for the test data (Figure 5). Moreover, the loss values also reduced after the epochs.

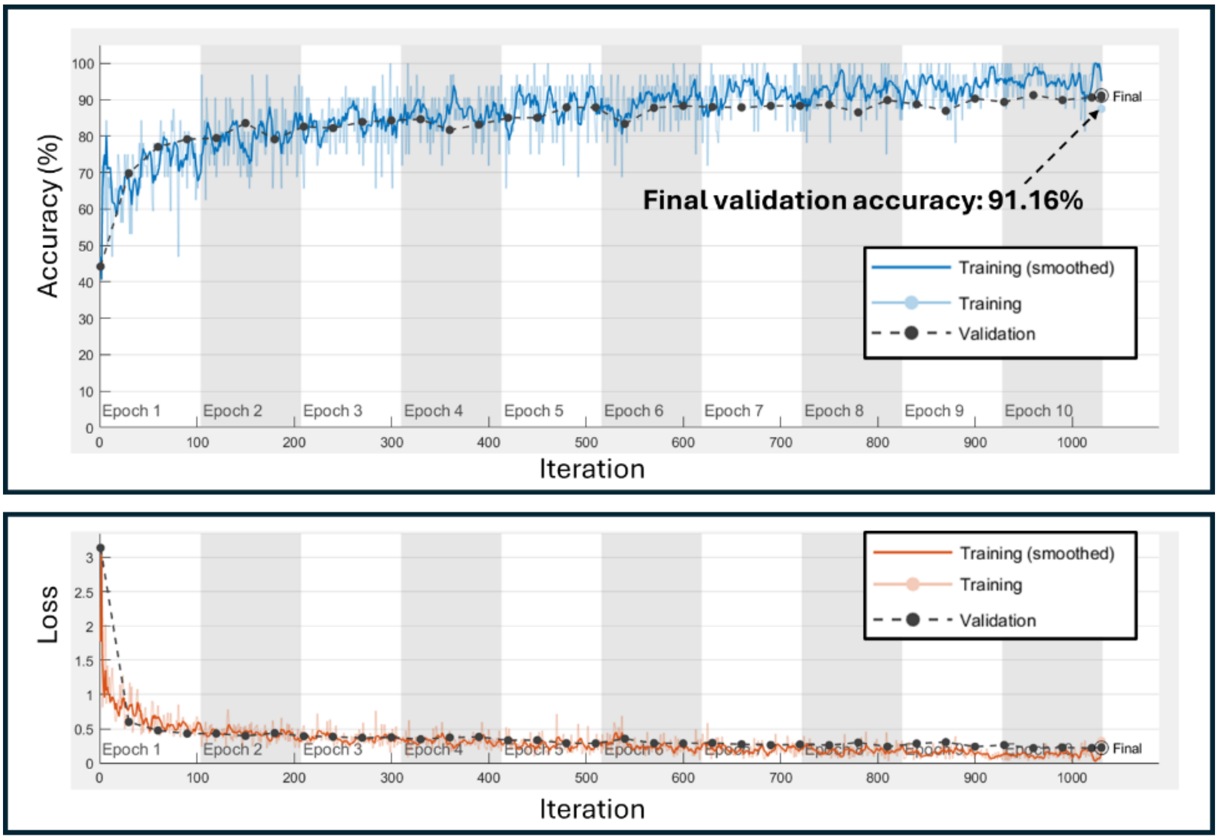
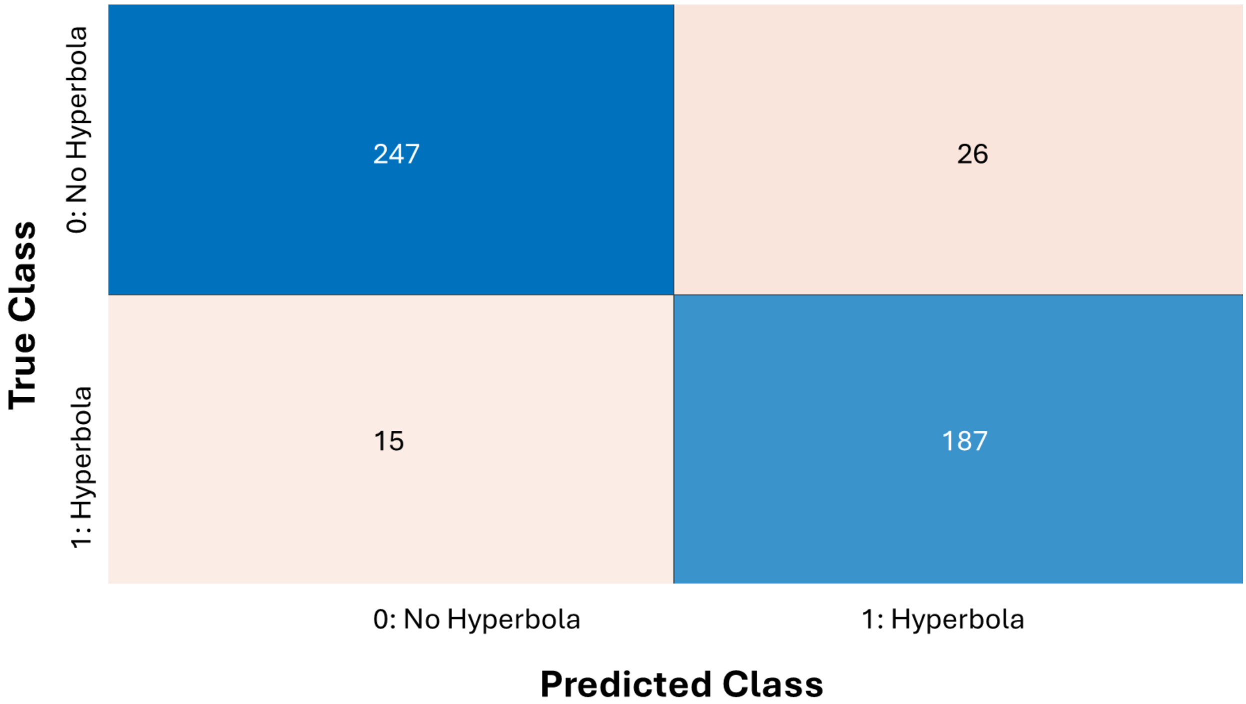


Fig. 5. Training progress

For testing datasets, the CNN model was applied. Overall, 187 correct and 15 wrong answers were simulated in the hyperbola case while 247 correct and 25 wrong answers were predicted in the no- hyperbola case. This reflects that the CNN model can work well (Figure 6).



https://doi.org/10.29227/IM-2025-01-02-011 Received: 30 Oct 2024, Accepted: 01 May 2025, Published: 01 Sep 2025

Fig. 6. Results of predicting the Test dataset from the trained CNN model. Accuracy=91.37%. Precision=87.79%. Recall=92.57%

Three indexes as accuracy, precision, and recall can measure how good the CNN model is for the test (Visa and Salembier, 2014). Then, the accuracy (434/475), precision (187/213), and recall (187/202) values are 91.37%, 87.79%, and 92.57%, respectively. For the precision measure, the CNN model is 87.79% accurate when it declares that an image has existence of hyperbola. Meanwhile, the recall value 92.57% reflects the ration between the number of images having hyperbolic curve correctly predicted over the total number of the images having hyperbolic curve.

Some images as the results of CNN prediction were shown in Figures 7 and 8. In Figure 7, GPR images in the test data collection with and without hyperbolic curves are correctly predicted by the CNN model. Meanwhile, Figure 8 shows the wrong predictions about GPR images.

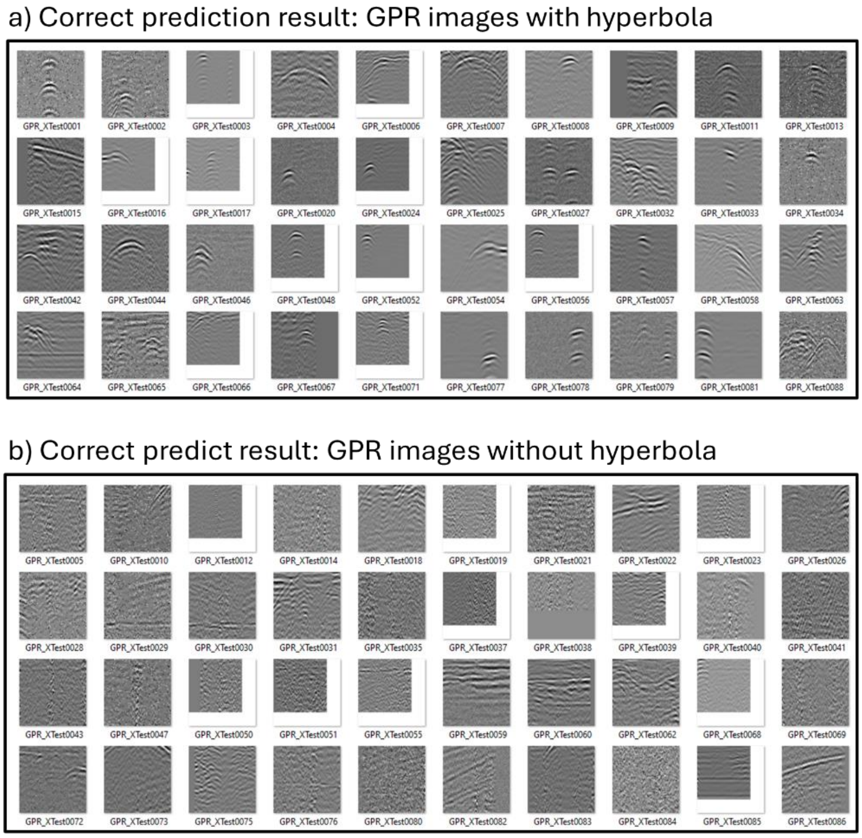


Fig. 7. Illustration of the correct decision of the CNN model

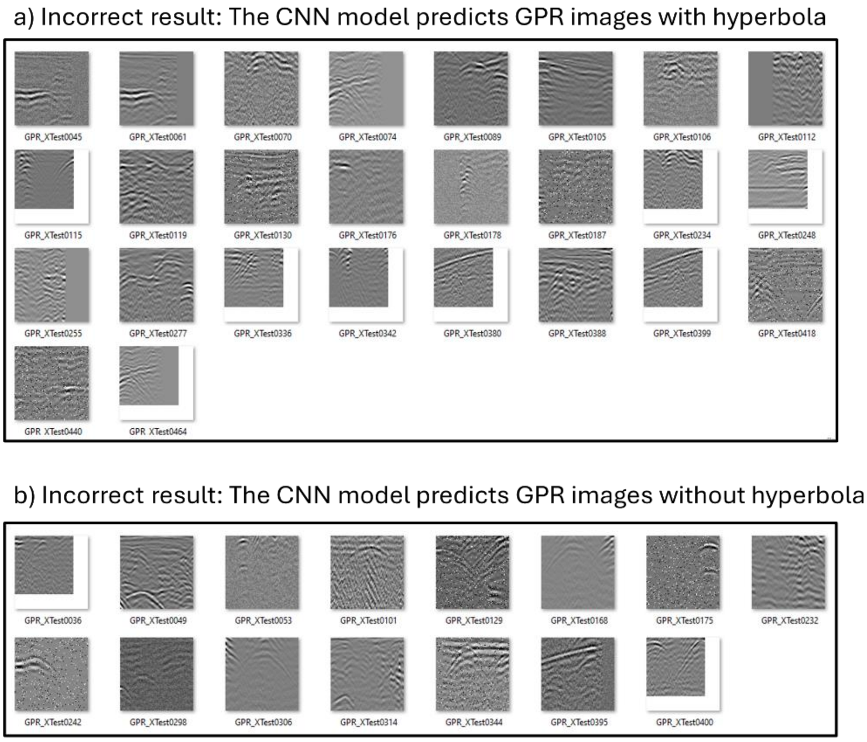


Fig. 8. Illustration of the incorrect decision of the CNN model

3.2. Real data interpretation in Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam

A 2D GPR profile was conducted in Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam using the Detector Duo GPR machine, manufactured in Italy (IDS, 2010). Underground anomalies were investigated using two methods: (i) traditional data processing and application of the CNN model to capture areas with hyperbolic curves (Figure 9).

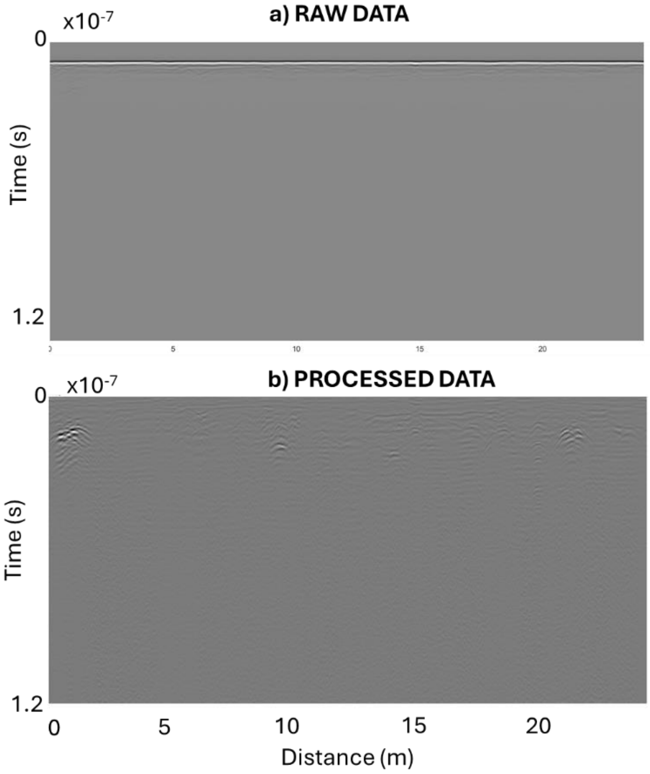


Fig. 9. Raw (a) and processed (b) data in Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam

When the raw GPR data cannot show the meaningful features, conventional processing stages (Methodology, 2.1) were applied to the raw data to achieve its interpretable data (Figure 9). Then, the interpretable data was injected to the CNN model and the predicted result is represented in Figure 10. That is: zones 1, 3, 4, 5, and 6 are correct because the hyperbolic curves are easily seen in the Figure 10. Meanwhile, zone 2 does not show clearly hyperbolic shape existence although there are very weak hyperbolic signals. For positions without hyperbolic curves, the CNN can show very great predictions. Therefore, the CNN model can work well. For improvement of the CNN model, bigger GPR data collection and better configuration of neural network layer can be created for training.

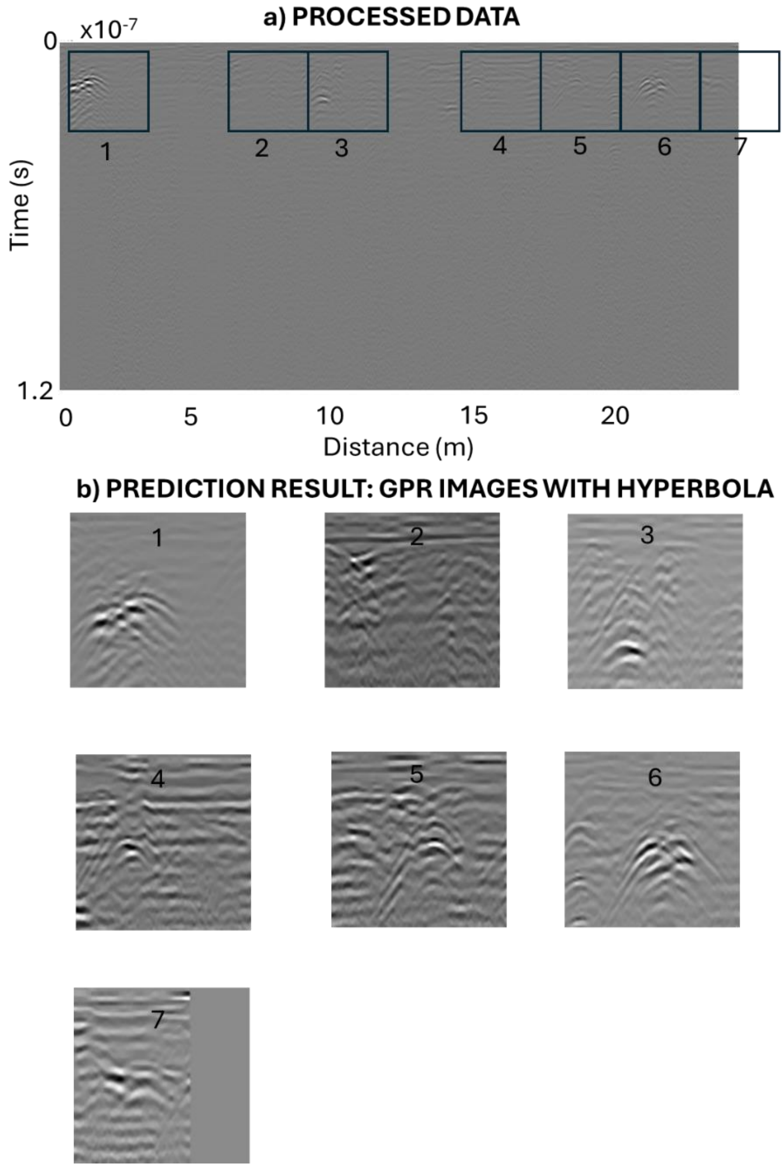


Fig. 10. CNN modelling for the real data in Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam

4. Conclusion

There was researched application of convolutional neural network (CNN) for detecting zones of hyperbolic curves in GPR data. The CNN model was trained from the collection of real data recorded in Southern Vietnam and a suitable configuration of layers. After the training process with high accuracy percentage of three collections of training, validation, and test data, the CNN model result was applied to the real data in Nguyen Van Cu Street, District 5, Vietnam. Zones of hyperbolic curves from the 2D GPR slices can be mostly properly detected although there is one wrong prediction from one zone.

Acknowledgements

This research is funded by Vietnam National University, Ho Chi Minh City (VNU-HCM) under grant number B2024-18-02. The author thanks to dGB Earth Sciences and Curtin University for providing access to software tools. The author would like also thank Thuan Van Nguyen for providing useful information about Ground Penetrating Radar data collections. The author thanks the anonymous reviewers for their helpful feedback, which improved this article.

Declarations

Competing interests: The author declare no competing interests.

Author Contribution

The author made himself the analysis and interpretation of geophysical data.

Data availability statement

The data used in this research can be made available by the corresponding author upon reasonable request.

Literature - References

1. AMIDI, A. & AMIDI, S. 2018. *Convolutional Neural Networks cheatsheet* [Online]. Available: https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks [Accessed 15 October 2024 2024].
2. AQUINO, N. R., GUTOSKI, M., HATTORI, L. T. & LOPES, H. S. 2017. The effect of data augmentation on the performance of convolutional neural networks. Brazilian Society of Computational Intelligence, 2017 Brazil.
3. BERDICHEVSKY, M. N. & DMITRIEV, V. I. 2008. *Models and methods of Magnetotellurics*, Springer.
4. FONSEKA, D. & CHRYSOULAS, C. 2020. Data augmentation to improve the performance of a convolutional neural network on image classification. 2020 International Conference on Decision Aid Sciences and Application (DASA), 2020. IEEE, 515-518.
5. HERNÁNDEZ-GARCÍA, A. & KÖNIG, P. 2018. Further advantages of data augmentation on convolutional neural networks. Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part I 27, 2018. Springer, 95-103.
6. IDS 2010. DETECTOR DUO SYSTEM – User manual.
7. IOFFE, S. & SZEGEDY, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *In:* BACH, F. & BLEI, D., eds. ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning 2015 Lille France. JMLR.org, 448-456.
8. KANG, M.-S., KIM, N., IM, S. B., LEE, J.-J. & AN, Y.-K. 2019. 3D GPR Image-based UcNet for Enhancing Underground Cavity Detectability. *Remote Sensing,* 11**,** 2545.
9. KANG, M.-S., KIM, N., LEE, J. J. & AN, Y.-K. 2020. Deep learning-based automated underground cavity detection using three-dimensional ground penetrating radar. *Structural Health Monitoring,* 19**,** 173-185.
10. KINGA, D. & ADAM, J. B. 2015. A method for stochastic optimization. International conference on learning representations (ICLR), 2015 San Diego, California;. 6.
11. LE, C. V. A., HARRIS, B. D. & PETHICK, A. M. 2019. New perspectives on Solid Earth Geology from Seismic Texture to Cooperative Inversion. *Scientific Reports,* 9**,** 14737.
12. LE, C. V. A., HARRIS, B. D., PETHICK, A. M., TAKAM TAKOUGANG, E. M. & HOWE, B. 2016. Semiautomatic and Automatic Cooperative Inversion of Seismic and Magnetotelluric Data. *Surveys in Geophysics,* 37**,** 845-896.
13. LE, C. V. A. & NGUYEN, T. V. 2020. Detection of Underground Anomalies Using Analysis of Ground Penetrating Radar Attribute. *Inżynieria Mineralna – Journal of the Polish Mineral Engineering Society,* 1**,** 23-34.
14. LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *Nature,* 521**,** 436-444.
15. LI, Z., LIU, F., YANG, W., PENG, S. & ZHOU, J. 2021. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems,* 33**,** 6999-7019.
16. MATHWORKS. 2019a. *batchnorm: Normalize data across all observations for each channel independently* [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/dlarray.batchnorm.html [Accessed 23/5/2023].
17. MATHWORKS. 2019b. *flipdim Flip array along specified dimension* [Online]. Available: https://www.mathworks.com/help/matlab/ref/flipdim.html [Accessed 15 October 2024].

18. MATHWORKS. 2019c. *imnoise add noise to image* [Online]. Available: https://www.mathworks.com/help/images/ref/imnoise.html [Accessed 15 October 2024].
19. MATHWORKS. 2019d. *imresize resize image* [Online]. Available: https://www.mathworks.com/help/matlab/ref/imresize.html [Accessed 15 October 2024].
20. MATHWORKS. 2019e. *rand Uniformly distributed random numbers* [Online]. Available: https://www.mathworks.com/help/matlab/ref/rand.html [Accessed 15 October 2024].
21. MATHWORKS. 2019f. *Semantic image segmentation using deep learning* [Online]. Available: https://www.mathworks.com/help/vision/ref/semanticseg.html [Accessed 2020].
22. MATHWORKS. 2019g. *Train a neural network* [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/trainnetwork.html [Accessed 2020].
23. NGUYEN, T. V., LE, C. V. A., NGUYEN, V. T., DANG, T. H., VO, T. M. & VO, L. N. L. 2018. Energy Analysis in Semiautomatic and Automatic Velocity Estimation for Ground Penetrating Radar Data in Urban Areas: Case Study in Ho Chi Minh City, Vietnam. *In:* DIEU TIEN BUI, A. N. D., HOANG-BAC BUI, NHAT DUC HOANG, ed. International Conference on Geo-Spatial Technologies and Earth resources GTER2017, 2018 2017 Ha Noi, Vietnam. Springer, 34-51.
24. PEROL, T., GHARBI, M. & DENOLLE, M. 2018. Convolutional neural network for earthquake detection and location. *Science Advances,* 4**,** e1700578.
25. PHAM, M.-T. & LEFÈVRE, S. 2018. Buried object detection from B-scan ground penetrating radar data using Faster-RCNN. IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium, 2018. IEEE, 6804-6807.
26. SANDMEIER, K.-J. 2020. *Reflexw - GPR and seismic processing software* [Online]. Available: https://www.sandmeier-geo.de/reflexw.html [Accessed July 7th, 2020].
27. SIRIPUNVARAPORN, W., EGBERT, G., LENBURY, Y. & UYESHIMA, M. 2005. Three-dimensional magnetotelluric inversion: data-space method. *Physics of the Earth and Planetary Interiors,* 150**,** 3-14.
28. SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I. & SALAKHUTDINOV, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research,* 15**,** 1929-1958.
29. TZANIS, A. 2010. matGPR Release 2: A freeware MATLAB® package for the analysis & interpretation of common and single offset GPR data. *FastTimes,* 15**,** 17-43.
30. VISA, G. P. & SALEMBIER, P. 2014. Precision-recall-classification evaluation framework: Application to depth estimation on single images. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13, 2014. Springer, 648-662.
31. WANG, J., LIU, H., JIANG, P., WANG, Z., SUI, Q. & ZHANG, F. 2021. GPRI2Net: A Deep-Neural-Network-Based Ground Penetrating Radar Data Inversion and Object Identification Framework for Consecutive and Long Survey Lines. *IEEE Transactions on Geoscience and Remote Sensing*.