

Prediction of Stock Value using Pattern Matching Algorithm based on Deep Learning

Yoon-Ho Go, Jin-Keun Hong

Abstract: In this paper we began with finding ways to predict stock value flows of stock using deep learning. The purpose of this paper is to analyze the patterns in stock value and to analyze the relationship from stock values by deep running to predict what patterns will happen next stock value.

In this paper we made the data by dividing the stock value information of the time series for a certain period of time and the pattern of stock value by analyzing these data. It is configured the model to be used for deep learning and learned the patterned time series information using the created model. And then it is predicted the next pattern of stock value.

This paper focused machine learning. It is used of a time-series stock value information to predict the rise and fall of stock value. This paper is about how to analyze and how to predict. On the other hand, we can expect trend of stock value with high probability by analyzing pattern of current chart and anticipating pattern to follow. This is about what the deep-learning machine will analyze and predict for what.

If we analysis the patterns used in this paper more clearly and concisely, and if more learning is carried out, we will be able to make clearer predictions with no noise for future trends. As interest in stock forecasts and machine learning develops fast, performance is expected to improve day by day.

Keywords : stock analysis, prediction, pattern analysis, machine learning, deep learning.

I. INTRODUCTION

Over the past few years, machine learning has made remarkable progress. With the advancement of hardware, the speed of computation and the abundant computer resources have grown enough to handle numerous operations for machine learning. In addition, many researches for machine learning have established many theories, and various artificial intelligence models are emerging, and machine running enjoys the best heyday ever [1-6].

Artificial neural networks, which mimic human neural networks by computer operations, unlike past artificial intelligence, which counts a large number of case and operates according to human-determined processes, has shown a much higher level of results through guidance learning [7-8].

As artificial intelligence develops and attracts attention, libraries for artificial intelligence have emerged that can be used in a variety of programming languages. Artificial

Revised Manuscript Received on July 22, 2019.

First Author Name*, his/her department, Name of the affiliated College or University/Industry, City, Country. Email: xyz1@blueeyesintelligence.org

Second Author Name, department, Name of the affiliated College or University/Industry, City, Country. Email: xyz2@blueeyesintelligence.org

Third Author Name, department, Name of the affiliated College or University/Industry, City, Country. Email: xyz3@blueeyesintelligence.org

intelligence only had be in the area of mathematics, but its development has exploded with the introduction of artificial neural networks such as TensorFlow actively supported by Google, Caffe created by Facebook, and Torch, an open source project library. Libraries made easy to use have led to the universalization of artificial intelligence, while lowering the barriers to entry into AI development [9-10].

As the era of the fourth industry began, artificial intelligence became an important future technology, and it has already begun to be introduced in various industrial fields and services. Especially, the "AlphaGo", an artificial intelligence Go program that was introduced in 2016 by Google's DeepMind, was the biggest topic of the future of artificial intelligence.

At this time, there are people who have looked at artificial intelligence as a new perspective. They are stock analysts. For many years now, stock analysts have been working hard to predict how stock prices will change. They have used a variety of calculations to analyze stock prices. We thought that the various technical analyzes and technical indicators invented by stock analysts were the way to find out what the signal was. And we thought that past stock value data would have implications and influence what stock value data will be in the future. In other words, the "signal" in current chart would suggest the shape of the next chart. So in this paper we applied artificial intelligence algorithm to discover the signal.

II. RELATED RESEARCH

S. Selvin et al. have made artificial intelligence to predict short-term future value rather than share price by combining sliding-window and deep learning[11]. They explained in their study that CNN performed best. Because the stock information is irregular, has a lot of sudden changes, and does not always follow the same cycle.

Ji-Hoon Lee had made a dataset of KOSPI, S & P500, NASDAQ and NIKKEI[12]. It had been made by using a variety of technical indicators (e.g., Bollinger band, STOCHASTIC, moving average, etc.) through start price, the best price, the lowest price, the closing price the trading volume. He have created an indicator dataset and tried to predict using MLP, CNN, LSTM. He taught more than 20 Korean companies. The training set showed an average of 51.11% in MLP, 51.19% in CNN and 52.09% in LSTM.

Dae-Sup Song presented a model for forecasting the KOSPI index using the Deep Learning[13].

He uses Deep Learning, SVM (Support Vector Machine), LS-SVM (Least Square-Support Vector Machine) and ensemble technique (build multiple models to overcome problems.) The purpose of this study is to solve the problem of overfitting that can occur during training. In conclusion, he showed a 67.7% hit ratio in the Deep Learning model using ensemble technique, 63.9% in SVM, and 64.6% in LS-SVM. He concludes that this model can only determine whether the closing price of the next day's KOSPI is up or down. It is difficult to connect to profit using this model because it may be better in fluctuations of huge profit than frequent fluctuations of little profit.

Dong-Young Kim et al have used SNS and news articles through the Internet that include economic information and stock information in modern society in other to the research[14]. They searched for relevant keywords in articles and SNS, and judged positive or negative by using deep learning in articles and SNS, and predicted the rise and fall. Their prediction accuracy was more than 78%. However, they said companies that do not collect enough data were expected to experience poor forecast accuracy.

Xia Ding et al extracted the events from a news and expressed them as dense vectors[15]. They also predict short-term and long-term effects by learning how an event affects stock prices. They said they had a 6% higher hit rate compared to the state-of-the-art benchmark.

III. CHARACTERISTICS OF SOURCE DATA

In this paper, we conducted research on publicly traded stocks listed in the 1 ~ 20th cap of the market value of January 30, 2018 in the Korean stock market. However Samsung Bio Logics' listing period was short, so we could not get enough training set and test set. Therefore it is studied with 21st LG Electronics. The data period is January 2, 2001, and stocks listed after that date have no previous data. These time series data are not used immediately. We will change it to a state that is easy to learn through some preprocessing process.

A. Meaning values

First, get a moving average of 40 days to avoid confusion in learning with seemingly meaningless noise. The formula for this is shown in equation (1).

$$d_n = (x_n + x_{n-1} + x_{n-2} + \dots + x_{n-39})/40 \quad (1)$$

If it remove the NaN value, it will get 40 day moving average data except for the first 40 days.

B. Slice values

Now it will cut as much as it will be want to use to learn the data. The 40-day averaged time series data will be cut into blocks for 40 days. In order to learn the continuity of the block and to the movement unit, it will be limited to 10. This can be expressed by equation (2) and in Figure 1.

$$\begin{aligned} b_0 &= \{d_0, d_1, d_2, \dots, d_{39}\} \\ b_1 &= \{d_{10}, d_{11}, d_{12}, \dots, d_{49}\} \\ b_m &= \{d_n, d_{n+1}, d_{n+2}, \dots, d_{n+39}\} \end{aligned} \quad (2)$$

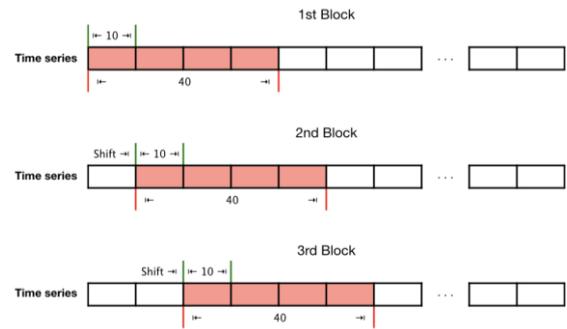


Fig. 1. Sliding window block

The data blocks generated by the above method will now be used for two purposes.

C. Standardize values

The data blocks generated above are different for each stock price of the company. So it is difficult to specify specifications for comparison or prediction. Therefore we will go through the standardization process. The standardization process divides each element in a block by the sum of the blocks to find the ratio of each number in a block. The sum of these ratios is 1.

$$SB(b_m) = \sum_{i=0}^{39} d_{10m+i} \quad (3)$$

$$B_m = N(b_m) = \left\{ \frac{d_m}{SB(b_m)}, \frac{d_{m+1}}{SB(b_m)}, \dots, \frac{d_{m+39}}{SB(b_m)} \right\} \quad (4)$$

D. Divide into two sets

Now we will split the data into two sets. It will make 80% of the data as training data and 20% as test data. The data used in this paper are shown in Table 1.

Table I: Training period, Test period

Company	Training period	Test period
Samsung C&T Corporation	2014/12/18 ~ 2017/06/15	2017/06/19 ~ 2018/01/31
Samsung Life Insurance Co., Ltd.	2010/05/13 ~ 2016/07/12	2016/07/14 ~ 2018/01/31
Samsung SDS Co., Ltd.	2014/11/17 ~ 2017/06/08	2017/06/12 ~ 2018/01/31
Samsung Electronics Co., Ltd.	2001/01/03 ~ 2014/08/21	2014/08/25 ~ 2018/01/31
Samsung Electronics Co., Ltd. (participating preferred)	2001/01/03 ~ 2014/08/21	2014/08/25 ~ 2018/01/31
Shinhan Financial Group Co., Ltd.	2001/09/11 ~ 2014/10/15	2014/10/17 ~ 2018/01/31
Amorepacific Corporation	2006/06/30 ~ 2015/09/30	2015/10/02 ~ 2018/01/31
Korea Electric Power Corporation	2001/01/03 ~ 014/08/21	2014/08/25 ~ 2018/01/31

Hyundai Mobis Co.,Ltd	2001/01/03 ~ 2014/08/21	2014/08/25 ~ 2018/01/31
Hyundai Motor Company	2001/01/03 ~ 2014/08/21	2014/08/25 ~ 2018/01/31
KB Financial Group Inc.	2008/10/13 ~ 2016/03/16	2016/03/18 ~ 2018/01/31
LG Household & Health Care Ltd.	2001/04/26 ~ 2014/09/16	2014/09/18 ~ 2018/01/31
LG Electronics Inc.	2002/06/19 ~ 2014/12/08	2014/12/09 ~ 2018/01/31
LG Chem, Ltd.	2001/04/26 ~ 2014/09/16	2014/09/18 ~ 2018/01/31
Naver Corporation	2002/10/30 ~ 2015/01/06	2015/01/08 ~ 2018/01/31
POSCO	2001/01/03 ~ 2014/08/21	2014/08/25 ~ 2018/01/31
SK Holdings Co., Ltd.	2009/11/12 ~ 2016/06/07	2016/06/09 ~ 2018/01/31
SK Innovation Co., Ltd.	2007/07/26 ~ 2015/12/15	2015/12/17 ~ 2018/01/31
SK Telecom Co., Ltd.	2001/01/03 ~ 2014/08/21	2014/08/25 ~ 2018/01/31
SK Hynix Inc.	2001/01/03 ~ 2014/08/21	2014/08/25 ~ 2018/01/31

IV. GENERATE PATTERNS

There are many ways to create patterns. Typically, there is a clustering algorithm such as the K-means algorithm. The K-means algorithm first needs to specify the number of centroids in k. In addition, there are many clustering algorithms and classification algorithms, but this paper is designed to set a threshold and add a new pattern for data that exceeds the threshold. Now let's create a pattern to be used for prediction.

A. Comparison patterns of stock value

When the function to create and compare the pattern list receive a pattern, it compares the received pattern with the previously saved pattern list. When there is no pattern data, the first block is saved as the first pattern. Subtract the pattern received from the saved patterns as shown in Equation (5) and convert each element to an absolute value. Finally we sum the absolute values of the elements as shown in Equation (6). Let's call this value the comparison value.

$$c_n = |P_n - B_m| \quad (5)$$

$$SC(c_n) = \sum_{n=0}^{39} c_n \quad (6)$$

B. Selection the pattern of stock value

Now we get the comparison value of all patterns by the above method. The pattern with the smallest value among these comparison values is the most similar pattern (if the threshold value is set and the smallest comparison value

exceeds the threshold value, it will be made into a new pattern). The pattern data and the incoming block are averaged to update the existing pattern. If it has created a pattern using all blocks, the pattern save and use it next. In this paper, 29 patterns were created and used. Figure 2 is a picture of 29 patterns. The vertical axis represents the percentage of each block element in the block, and the horizontal axis shown time series for 40-day.

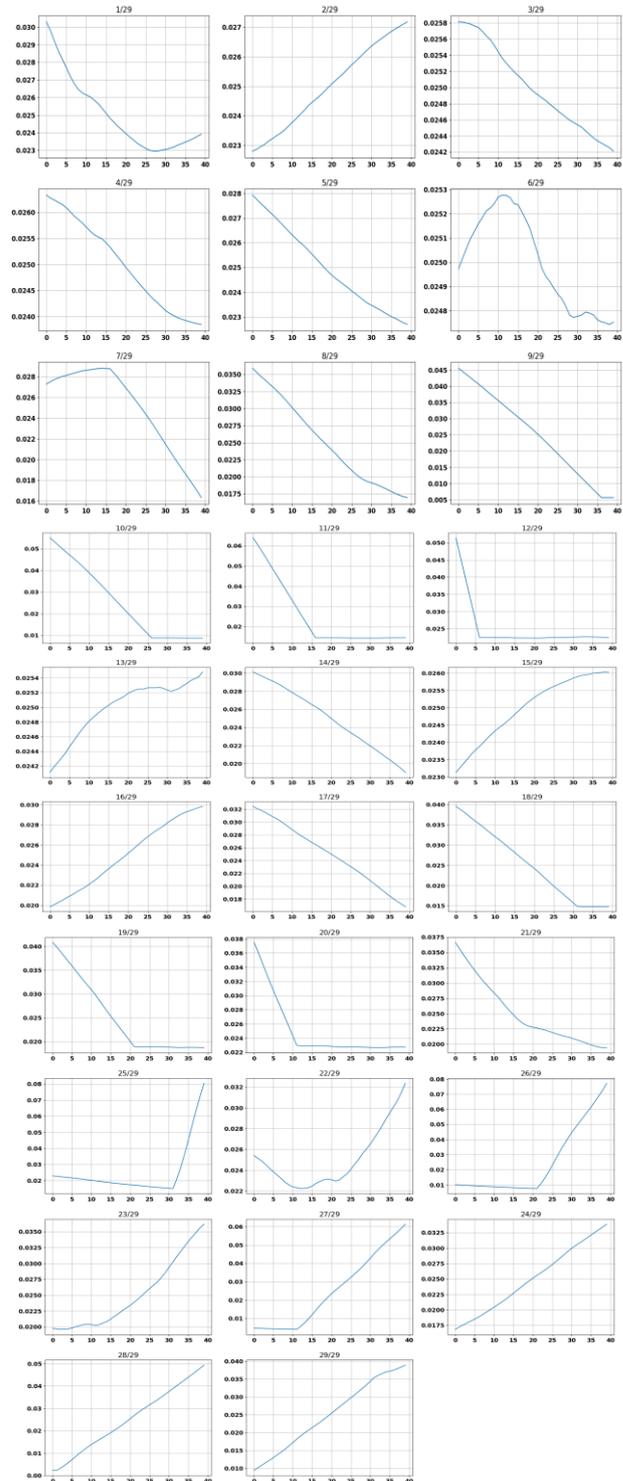


Fig. 2. 29 Patterns of stock value

V. LEARNING USING DNN CLASSIFIER

Now we will use the Deep Neural Network (DNN) to learn blocks and predict the next pattern. In this paper we used a DNN learning model of TensorFlow made by Google. DNN is an artificial neural network consisting of several hidden layers between the input and output layers as shown in Figure 3. Deep Neural Network can model complex nonlinear relationships like ordinary artificial neural networks.

A. DNN Learning model

We will now design the DNN Learning Model in Figure 4. The size of a block is 40, so the input layer of DNN is also 40. The output layer has 29 patterns, so it has 29 output layers. The hidden layer discretionally created three hidden layers with 200, 400, and 200 units respectively.

B. Training step

When the DNN model learns each block, you teach it the number of patterns that will appear next. Equation (7) is as follows.

$$P(B_{m+1}) = fit(B_m) \quad (7)$$

Training was learned 10,000 times for each company.

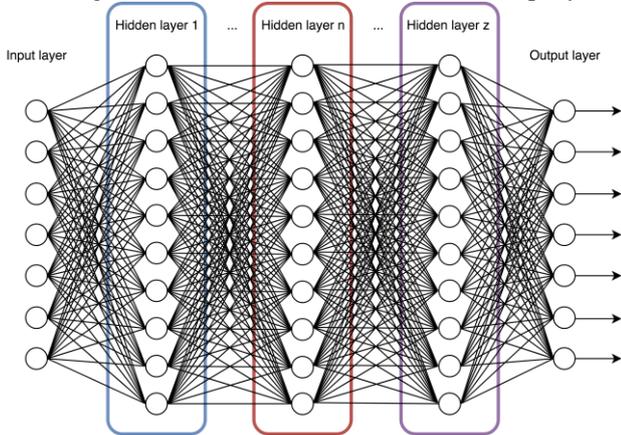


Fig. 3. DNN Model

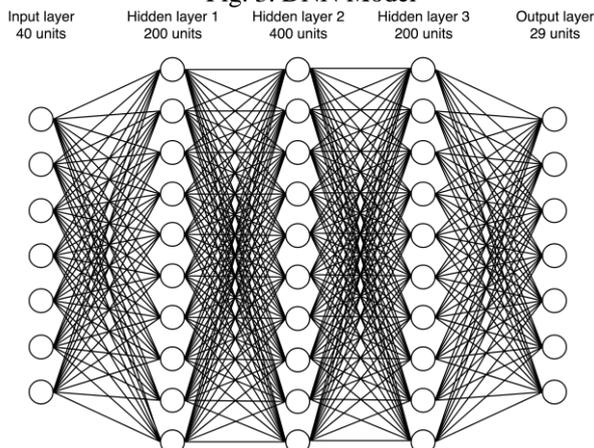


Fig. 4 DNN Learning Model

C. Test step

To see how accurate DNN learning is, we use the rest of the test set that is not used in the study for evaluation. The data used in the test step is not used for learning in the DNN model. The hit ratio compares the pattern obtained by predicting all test sets with the correct answer patterns, and obtains true if they match. Obtains the number of true values predicted in the test set. And we compute the percentage. For instance if 70

pattern match, when it there are 100 blocks in test set, and it shows a hit rate of 70%.

VI. EXPERIMENTAL RESULTS

Table 2 shows the results of evaluating the test set of 20 companies.

Table 2: Evaluation percentage of Test sets

Company	Evaluation percentage
Samsung C&T Corporation	89.99%
Samsung Life Insurance Co., Ltd.	96.96%
Samsung SDS Co., Ltd.	81.81%
Samsung Electronics Co., Ltd.	84.81%
Samsung Electronics Co., Ltd. (participating preferred)	88.60%
Shinhan Financial Group Co., Ltd.	92.10%
Amorepacific Corporation	84.61%
Korea Electric Power Corporation	93.67%
Hyundai Mobis Co., Ltd	77.21%
Hyundai Motor Company	78.48%
KB Financial Group Inc.	92.68%
LG Household & Health Care Ltd.	88.46%
LG Electronics Inc.	83.56%
LG Chem, Ltd.	85.89%
Naver Corporation	88.57%
POSCO	83.54%
SK Holdings Co., Ltd.	91.42%
SK Innovation Co., Ltd.	80.85%
SK Telecom Co., Ltd.	75.94%
SK Hynix Inc.	87.34%

This is a very high accuracy compared to 50-70% of the average accuracy, that obtained by chart analysis using deep learning.

VII. CONCLUSION

In this paper we designed the pattern and the AI model predicted the following pattern through the 40-day time series price. As a result the DNN deep-learning algorithm yielded an accuracy of about 86%. This is a high hit rate obtained from chart analysis alone. But there is still room for improvement. Better deep-learning algorithms such as CNN and RNN are still being designed. Also as there create patterns with more abstract and clearer points, when patterns creating, the expected hit rate will go up.

As AI technology progresses and moves closer to human thinking, there will become more skilled at analyzing charts, analyzing corporate outlook, and investing as though people analyze them. AlphaGo did a move that people could not think of in the confrontation between AlphaGo and Se-Dol Lee in 2016(AlphaGo has been developed by DeepMind of Google, and Se-Dol Lee is professional Go player). In this way artificial intelligence may find a "signal" that a person does not find about it on a stock chart.

Because DNN has a problem of overfitting and high time complexity, it's hard to trust the current hit is the best hit rate. We can use CNN to find patterns such as image processing, and use RNN to predict your previous learning. If we use the volume of transactions, foreign investors, and technical indices appropriately based on pattern forecasting, we expect that it will be an appropriate model for grasping future share price trends.

ACKNOWLEDGMENT

This paper is sponsored of project funding of Industrial Academia Cooperation in Baekseok University.

REFERENCES

1. J. F. Chen, W. L. Chen, C. P. Huang, S. H. Huang and A. P. Chen, "Financial Time-Series Data Analysis Using Deep Convolutional Neural Networks," in 7th International Conference on Cloud Computing and Big Data (CCBD)2016, <https://doi.org/10.1109/CCBD.2016.027>
2. R. Akita, A. Yoshihara, T. Matsubara and K. Uehara, "Deep learning for stock prediction using numerical and textual information," in IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)2016, <https://doi.org/10.1109/ICIS.2016.7550882>
3. In-Jung Kim, "Deep Learning: New trend of Machine learning," Information and Communication Magazine of Korea Institute of Information and Communication Sciences (Information and Communication), 31(11), pp. 52-57, 2016, <https://www.dbpia.co.kr/Journal/ArticleDetail/NODE02502238>
4. Dong-Ha Shin, Kwang-Ho Choi, Chang-Boy Kim, "Deep Learning Model for Prediction Rate Improvement of Stock Price Using RNN and LSTM," Journal of Korean Institute of Information, 15(10), pp. 9-16, 2017, <http://doi.org/10.14801/jkiit.2017.15.10.9>
5. Yoojeong Song, Jongwoo Lee, "A Design and Implementation of Deep Learning Model for Stock Prediction using TensorFlow," in Proceeding of Korean Institute of Information Scientist and Engineers2017, <https://www.dbpia.co.kr/Journal/ArticleDetail/NODE07207386>
6. https://ko.wikipedia.org/wiki/deep_learning
7. <https://www.tensorflow.org>
8. <https://tensorflowkorea.gitbooks.io/tensorflow-kr/content>
9. K. Khare, O. Darekar, P. Gupta and V. Z. Attar, "Short term stock price prediction using deep learning," in 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)2017, <https://doi.org/10.1109/RTEICT.2017.8256643>
10. M. R. Vargas, B. S. L. P. de Lima and A. G. Evsukoff, "Deep learning for stock market prediction from financial news articles," in IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)2017. <https://doi.org/10.1109/CIVEMSA.2017.7995302D>
11. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in International Conference on Advances in Computing, Communications and Informatics (ICACCI)2017, <https://doi.org/10.1109/ICACCI.2017.8126078>
12. Ji-Hoon Lee, "Stock price prediction model using deep learning," MA Thesis, Soongsil University, 2017, https://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=80c3f1a991ea239bffe0bdc3ef48d419
13. Dae-Sup Song, "Predicting KOSPI index fluctuations with ensemble and deep learning," MA Thesis, Dankook University, 2016, Republic of Korea. http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=8d69978961c2f30affe0bdc3ef48d419
14. Dongyoung Kim, et al., "A Comparative Study between Stock Price Prediction Models Using Sentiment Analysis and Machine Learning Based on SNS and News Articles," Journal of Korea Information Technology Services, 13(3), pp. 221-233, 2014. https://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=fd5085ba8853d3aaffe0bdc3ef48d419
15. Ding, X., Zhang, Y., Liu, T., & Duan, J., "Deep learning for event-driven stock prediction," In IJCAI 2015.

AUTHORS PROFILE



Yoon-Ho Go student in division of ICT in Baekseok university of South Korea. His interest area is development of client and server platform programming and Back-end Server engineering and security, PaaS Cloud infrastructure for security, microservice, scale-out, serverless. Mainly using languages are Typescript, Python, Java. Contributed projects are GraphQL API deployed production, OAuth2 auth server, Tensorflow stock prediction, etc.



Jin-Keun Hong who serves professor in division of ICT in Baekseok University of South Korea. He was registered an excellent researcher at Marquis Who's Who in the world, IBC, and ABI human dictionary according to research results of ICT & Security. His research topic is convergence information and AI security technology. Especially, he is focusing on C-ITS and future security technology prediction.