IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 10, 2021, accepted August 31, 2021, date of publication September 14, 2021, date of current version October 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3112682

Smart Policing Technique With Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation

MYUNG-SUN BAEK[1], (Member, IEEE), WONJOO PARK[1], JAEHONG PARK[2], KWANG-HO JANG[3], AND YONG-TAE LEE[1]

[1]Electronics and Telecommunication Research Institute (ETRI), Daejeon 34129, South Korea
[2]MyCQ Inc., Daejeon 34129, South Korea
[3]Smart Police Intelligence Center, Police Science Institute, Asan-si, Chungcheongnam-do 31539, South Korea

Corresponding author: Yong-Tae Lee (ytlee@etri.re.kr)

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government [Ministry of Science and ICT (MSIT)] under Grant 2018-0-00440 (ICT-Based Crime Risk Prediction and Response Platform Development for Early Awareness of Risk Situation).

ABSTRACT In order to quickly and effectively respond to a newly received criminal case, information regarding the type and severity of the case is crucial for authorities. This paper designs and develops a crime type and risk level prediction technique based on machine learning technology and verifies its performance. The designed technology can predict crime type and crime risk level using a text-based criminal case summary, which is criminal case receipt data. For the text-based criminal case summary data, the KICS data format is considered, which is actual policing data that contains information about criminal cases. For the crime type, 21 representative types of crimes are considered; therefore, the system can predict one of 21 types of crime for each criminal case. Furthermore, to predict the crime risk level, we developed a crime risk calculation formula. The developed formula calculates the crime risk level and outputs the risk score in numerical terms considering the severity and damage level of the criminal case. To predict the crime type and crime risk score, both DNN and CNN-based prediction models were designed and developed. The performance evaluation section shows that, in the case of crime type prediction, the proposed prediction models can achieve better performance than traditional classification algorithms such as naïve Bayes and SVM. The performance of the CNN-based crime type prediction model is about 7% and 8% better than those of the SVM algorithm and the naïve Bayes algorithm, respectively. The performance of the designed technology was comprehensively analyzed and verified through various performance measurement parameters. It is also developed in the form of a software platform with a GUI, allowing field personnel (e.g. police officers) to intuitively identify the type of criminal case and the level of risk from a text-based criminal case summary upon receipt of a new criminal case.

INDEX TERMS Crime response technology, smart policing, machine learning, natural language processing.

I. INTRODUCTION

Digital transformation based on artificial intelligence (AI) technology and big data has received considerable attention [1], [2] and has become a huge trend. In addition, AI technology is being utilized in various research areas such as smart communications, medical research, natural language process, and robots [3]–[9]. Recently, smart policing technologies using AI and big data have been actively investigated to secure the lives and properties of the public [10]–[17]. Yoo *et al.* [10] has investigated a fingerprint detection technique using computer vision technology based on deep learning. Deep reinforcement learning technology has been used in [11] to detect criminal networks. Sangkaran *et al.* [12] has developed community detection in criminal networks using the graph theory, and the method differs from the traditional method by allowing law enforcement agencies to be able to compare the detected communities and thereby be able to assume a different view-point of the criminal network. Shafi *et al.* [13] has proposed simplified yet adaptable

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

framework that uses a novel features extraction algorithm for extracting features from the textual part of social media contents. Various technologies for predicting and preventing crimes have been developed using accumulated security data and AI technology [14]–[17]. In addition, the need to develop smart policing technologies that can provide detailed information related to newly received criminal incidents to police officers and investigators dispatched to the crime scene is constantly being raised. The details provided in connection with the newly received criminal case allow field staff and investigators to respond quickly and efficiently to the crime.

In this paper, we propose and implement a smart policing technique that can predict crime types and crime risk levels based on machine learning so that dangerous situations can be recognized early. The proposed technique predicts crime types by analyzing text-based crime summaries. In addition, the proposed technique predicts the crime risk level as well. Before developing the crime type and crime risk prediction technique, a calculation formula that numerically calculates the crime risk score (CRS) has also been developed. The CRS calculation formula calculates the severity of a criminal event taking into account the type of crime and the level of damage to the victim. ETRI and the Korea Police Science Institute (PSI) collaborated together to establish a reasonable crime risk score calculation formula. Besides, to reflect practical crime risk levels, the opinions of the Korean police officers were gathered and included in the formula. The proposed technique predicts the type of criminal case newly received, infers the severity of the criminal case, and outputs the information to the police. In addition, the developed technology is implemented in the form of a GUI-based SW platform, so field staff can easily use the developed system. To develop the proposed technology, data from the Korea Information System of Criminal Justice Services (KICS) was considered. The KICS data is actual policing data that contains information about each criminal case. For security reasons, we have created and used virtual KICS data according to the actual KICS data format. The created virtual KICS data includes 5000 criminal cases and is utilized to train and verify the developed system. In the virtual KICS data, text data on criminal cases is extracted and used for system development. The developed system can predict crime type and CRS from text-based crime data. Table 1 describes the examples including text-based criminal case descriptions, crime types, and CRS. The second column of Table 1 is a text-based criminal case description that follows the KICS format. The simple descriptions of criminal cases are shown in this column. The first criminal case is about larceny. The second criminal case is about fraud, and the last criminal case is about arson. The third column shows the crime types for the criminal cases. And the fourth column shows the calculated CRS. Therefore, the crime type and CRS can be labels for the prediction system. The developed system can predict the crime type and CRS based on the text-based criminal description part, as shown in Table 1. The developed system takes into account 21 types of crimes that fall into the middle category. Therefore, the developed system can predict one of the 21 criminal types for each criminal case. This prediction system is developed for security data based on the Korean language. However, in this paper, the proposed technology and developed system are introduced using English examples in order to improve readability.

TABLE 1. Example of text part of KICS data, Crime type and CRS about criminal cases.

	Crime summary	Crime type	CRS
1	The suspect Yoo-Jeong Moon and the victim Lee-Hyun Ko are friends. The suspect slept at the victim's house at 46, Yeongchang-ro 163beon-gil (Changjeon-dong), Icheon-si, Gyeonggi-do around the evening of May 03, 2014. Two pieces, a hand mirror worth $10.00, a necklace worth $180.00, and one earring worth $50.00 were stolen. As a result, the suspect stole a total market value of $330.00 from the victim.	Larceny	132.5
2	Suspect Se-Yoon Choi wrote a post that he would sell a smartphone for $450 under the title "iPhone xs Gold 256g Full Box (Unused)" in a second-hand app using his ID "seun0190". And he lied that he would sell it to the victim Ga-Hee Do who had contacted him. The suspect did not have the ability or willingness to sell the smartphone to the victim. On this day, the suspect received $452, including bank transfer fees, to the IBK (94648289425175) account through the deceived victim. As a result, the suspect deceived the victim and obtained a total profit of $452 from the victim.	Fraud	44.5
3	The suspect, Kang-Min Ban, worked at the forest factory in Bian-myeon, Uiseong-gun, Gyeongsangbuk-do from January 2016 to November 2016. However, because the salary was not properly paid, he decided to set fire to the factory in order to revenge. On December 29, 2016, around 00:36, the suspect crumpled and stacked toilet paper in a wooden waste collection station at a forest factory, sprinkled thinner and lit it on fire. The fire immediately spread throughout the building. As a result, the suspect destroyed a building worth $200,000.	Arson	309

In the first step of the prediction system development process, a keyword dictionary is built. In the keyword

TABLE 2. Assigned WC values.

	Crime Type	Weight Value
1	Murder	100
2	Rape	75
3	Imitative rape	75
4	Arson	70
5	Indecent assault	60
6	Kidnapping	60
7	Robbery (mugging)	60
8	Injury	60
9	Battery (violence)	55
10	Drug	50
11	False arrest, illegal Confinement	40
12	Specific Economic Crimes	40
13	Intimidation	30
14	Larceny	30
15	Extortion	30
16	Breach of duty	30
17	Fraud	10
18	Gambling Crime	10
19	Embezzlement	10
20	Destruction of Property	10
21	Crimes of Sexual Morals	10

TABLE 3. Assigned values for *WG* and *WA*.

Variable	Item	Weight Value
Victim's gender (*WG*)	Man	2.0
	Woman	2.4
	Unknown	2.2
Victim's Age (*WA*)	Under 7 years old	2.5
	Over 6 years old and under 13 years old	2
	Over 12 years old and under 16 years old	1.5
	Over 15 years old and under 21 years old	1.2
	Over 20 years old and under 60 years old	1
	Over 59 years old	1.2
	Unknown	1

TABLE 4. Assigned values for *WP* and *WM*.

Variable	Item	Weight Value
Victim's Physical Damage (*WP*)	No damage	0.0
	2 weeks or less	0.8
	Over 2 weeks and 1 month or less	1
	Over 1 month and 2 months or less	1.2
	Over 2 months and 4 months or less	1.4
	Over 4 months and 6 months or less	1.6
	Over 6 months	2
	Dead	5
Victim's Material Damage (*WM*)	No damage	0.0
	Damage < \$10	0.01
	\$10 ≤ Damage < \$100	0.02
	\$100 ≤ Damage < \$1000	0.05
	\$1000 ≤ Damage < \$10000	0.08
	\$10000 ≤ Damage < \$100000	0.1
	\$100000 ≤ Damage < \$ 1 million	0.2
	\$ 1 million < Damage	0.3
	Unknown	0.05

dictionary build process, about 27 keywords are extracted for each crime type. So, given 21 types of crime, the keyword dictionary consists of 568 keywords. And then, the dataset is established. In this process, input data and output labels are determined. In addition, the training dataset, validation dataset, and test dataset are arranged, deep learning-based prediction models are developed, and the models are trained by the training dataset. The real-time GUI system is implemented and applied to the prediction models. The developed GUI-based system can predict crime type and CRS in real-time for the new input data.

The rest of this paper is organized as follows. Section II explains the formula for CRS calculation. Section III describes the prediction models. In Section IV, performance evaluation results and discussions are reported. The concluding remarks are in Section V.

II. FORMULA FOR CRIME RISK SCORE CALCULATION

In this Section, the development of a crime risk score calculation formula is explained which calculates the crime risk as a numerical value and then output. The crime risk score is calculated by taking into account the crime type and various damage information for the victim of the crime. The formula for calculating CRS can be written as:

$$CRS = WC \times (WG + WA) + WP \times (WG + WA) + WM \times 10 \tag{1}$$

where the meanings of the variables are as follows:

- *WC* (weight for crime type): weight according to crime type
- *WG* (weight for gender): weight according to victim's gender
- *WA* (weight for age): weight according to victim's age
- *WP* (weight for physical damage): weight according to victim's physical damage
- *WM* (weight for material damage): weight according to victim's material damage

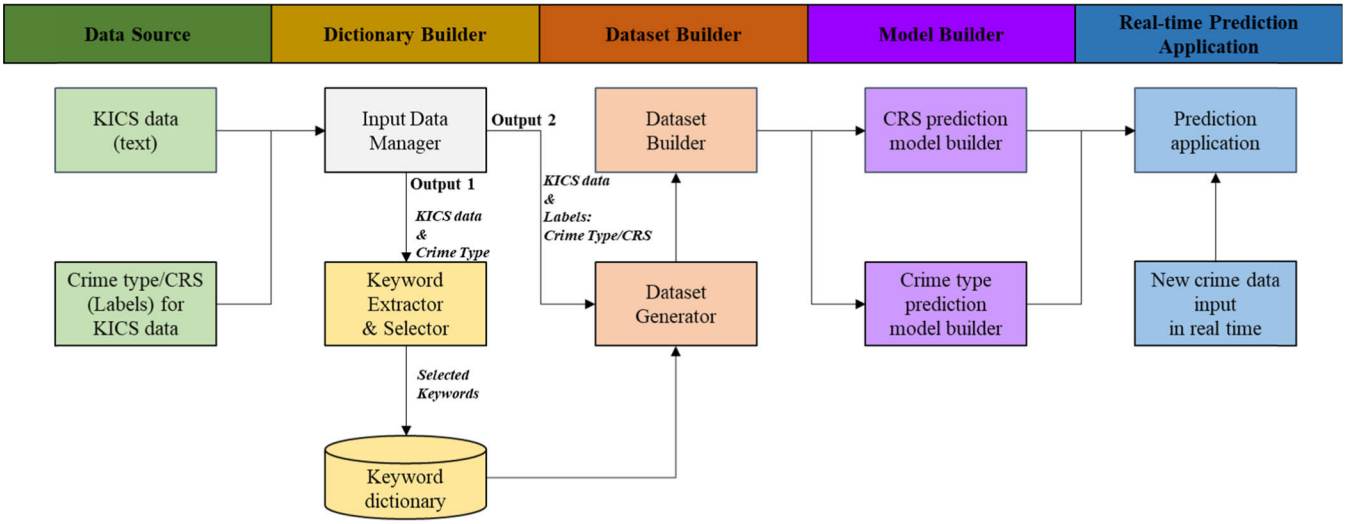

FIGURE 1. Structure diagram of crime type and CRS prediction system.

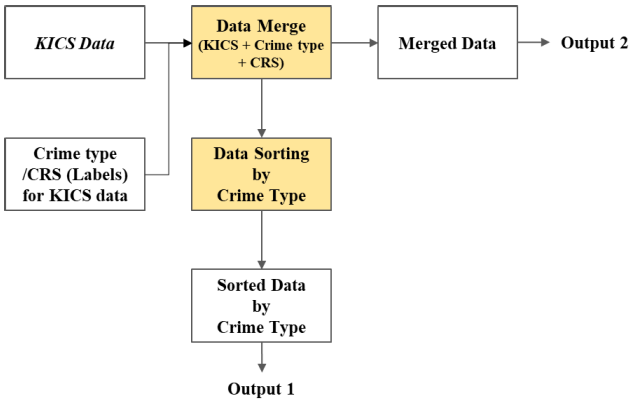

FIGURE 2. Input data manager process.

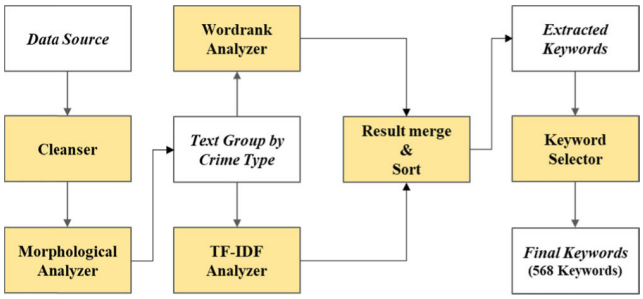

FIGURE 3. Keyword extraction process.

For the *WC*, this paper considers 21 crime types that are representative and practical in the policing environment, and the considered crime types and weight values are listed in Table 2. The weight values of *WC* are assigned considering severity as well as sentencing guidelines according to crime types. *WG* and *WA* are variables that are set taking into account the victim's physical information. Therefore, WG and WA can be regarded as information that determines how vulnerable a victim is to a crime. The weight values of WG and WA are described in Table 3. As shown in Table 3, a higher weight value is assigned to women who are relatively vulnerable to criminal damage, and from this perspective, higher weight values are assigned to young children and the elderly who are vulnerable to crime. *WP* and *WM* are variables that present the damage level from crime. *WP* is the victim's physical damage. The weight value of *WP* is set according to the time required to fully heal the damage.

WM indicates the degree of property damage: the higher the amount, the higher the assigned weight value. Table 4 shows the weight values for *WP* and *WM*. As mentioned in Section I, ETRI and PSI worked jointly to establish a reasonable crime risk score calculation formula. To reflect practical crime risk levels, the opinions of Korean police officers were gathered and included in the formula.

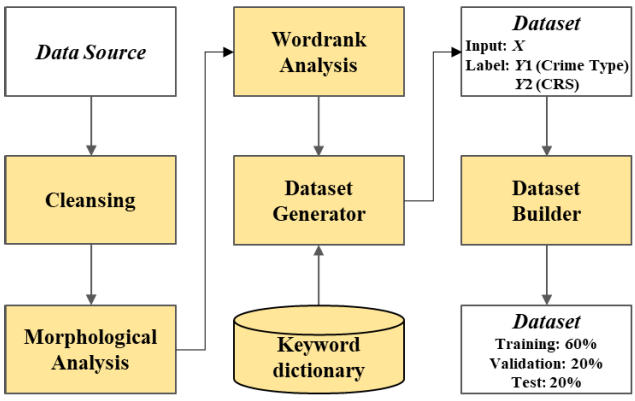

FIGURE 4. Process of dataset builder.

III. PREDICTION SYSTEM FOR CRIME TYPE AND CRIME RISK SCORE

Fig. 1 shows the proposed crime type and CRS prediction system. The proposed system consists of five functional parts which are as follows:

- **Data source:** Text-based crime data is used as a data source. The crime data contains criminal information (criminal case summary), and the form of the crime data source is the same as actual Korean police security data from KICS. In addition, crime type and CRS according to the text-based crime data are also inserted for labeling of the training process.
- **Dictionary builder**: In this process, keywords are extracted from the text-based crime data source, and a keyword dictionary is built.
- **Dataset builder**: Datasets for training, validation, and test are built using the text-based crime data source and the keyword dictionary.
- **Model builder**: In this process, both the CRS prediction model and crime type prediction model are designed, built, and trained by using the datasets.
- **Real-time prediction application platform**: This is a GUI application system that can predict crime type and CRS from text-based crime data input in real-time.

A. INPUT DATA MANAGER

Fig. 2 describes the process of the input data manager. At the first step, the input data manager merges the KICS data with the crime type/CRS that can be labels for the prediction model. The input data manager produces two kinds of output. The first is Output 1 which is used for the keyword dictionary builder. To extract keywords according to the crime types, the data manager sorts the entered KICS data by crime type and outputs the sorted data as Output 1. The second output of the input data manager is Output 2 which is used for the dataset builder. In the dataset builder, the training dataset, validation dataset, and test dataset are built. Therefore, the data in which KICS data, crime type, and CRS are merged can be Output 2.

B. KEYWORDS DICTIONARY BUILDER

Fig. 3 shows the keyword extraction process. In this process, keywords are extracted by both the wordrank analysis algorithm [18], [19] and the term frequency – inverse document frequency analysis algorithm (TF-IDF) [20], [21].

The text-based input KICS data is processed by each criminal case. The cleansing process is applied to the input data. In the cleansing process, noise that may be unnecessary or interfere with keyword extraction from the data source is removed. In this process, date and time information is removed, and the morphological analysis is then applied to the output of the cleansing process [22]. In this morphological analyzer, general nouns (NNG) and verbs (VV) are extracted. The outputs of the morphological analyzer are sorted and grouped according to 21 crime types in order to extract keywords for each crime type. The text group by crime type is inserted into both the wordrank analyzer and the TF-IDF analyzer. The results of both analyzers are merged and sorted in the order of the highest score.

The extracted keywords are inserted into the keyword selector. Among the keywords extracted by crime type, keywords with over 100 points are finally selected and stored in the keyword dictionary. There are 568 keywords stored in the keyword dictionary, and the number

TABLE 5. The number of keywords of each crime type.

	Crime Type	The Number of Keywords
1	Murder	22
2	Rape	25
3	Imitative rape	29
4	Arson	21
5	Indecent assault	21
6	Kidnapping	29
7	Robbery (mugging)	30
8	Injury	37
9	Battery (violence)	33
10	Drug	25
11	False arrest, illegal Confinement	20
12	Specific Economic Crimes	33
13	Intimidation	26
14	Larceny	21
15	Extortion	22
16	Breach of duty	28
17	Fraud	35
18	Gambling Crime	24
19	Embezzlement	32
20	Destruction of Property	23
21	Crimes of Sexual Morals	32

TABLE 6. One-hot vector for 21 crime types.

One-Hot Encoded Label	Crime Type
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Larceny
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	False arrest, Illegal Confinement
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Destruction of Property
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Extortion
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Kidnapping
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Injury
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Battery (violence)
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	Intimidation
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0	Embezzlement
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0	Breach of duty
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	Fraud
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0	Specific Economic Crimes
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0	Rape
0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0	Robbery
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0	Indecent assault
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0	Murder
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0	Arson
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0	Imitative rape
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0	Drug
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0	Gambling Crime
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	Crimes of Sexual Morals

of stored keywords for each crime type is described in Table 5.

C. DATASET BUILDER

In this dataset builder, datasets for training, validation, and test are generated, as shown in Fig. 4. In the first step of creating the dataset builder, data sourced from the input data manager is processed by a cleanser and morphological analyzer. These cleansing and morphological analysis processes are the same as those of the keywords dictionary builder. The processed data is inserted into the wordrank analyzer block, and NNG and VV are extracted by each criminal case. The result of the wordrank analyzer is inserted into the dataset generator. In the dataset generator, the inserted the wordrank analyzer result is analyzed by each criminal case and compared with the keyword dictionary which consists of 568 keywords. According to the above process, input vector X with 568 $\times$ 1 is generated. The input vector generation process is as follows:

- An all zero vector with 568 $\times$ 1 length is generated.
- The extracted words from the wordrank analyzer result are compared with 568 keywords in the keyword dictionary.
- If one of the words in the result of the wordrank analyzer result is the same as one in the keyword dictionary, the value of the zero vector at the same position as the corresponding keyword position is changed to 1.
- This comparison process is applied to all words extracted from the result of the wordrank analyzer.
- The generated vector with 568 $\times$ 1 length can be an input vector X of the prediction model.

At the same time, the dataset generator produces two kinds of output ($Y1$ for crime type and $Y2$ for CRS) of the prediction model. The output of the crime type is the one-hot encoded label which presents one of the 21 crime types. Table 6 shows the one-hot encoded output vector for crime types, and numeric CRS values are used for CRS.

The generated datasets are transmitted to the dataset builder. The dataset builder separates the received dataset for training, validation, and test as shown in Fig. 4. As mentioned in Section I, virtual KICS data with 5000 crime cases is created and utilized in our research. Therefore, as shown in fig. 4, the dataset composition is as follows:

- Training: 60% (3000 crime cases)
- Validation: 20% (1000 crime cases)
- Test: 20% (1000 crime cases).

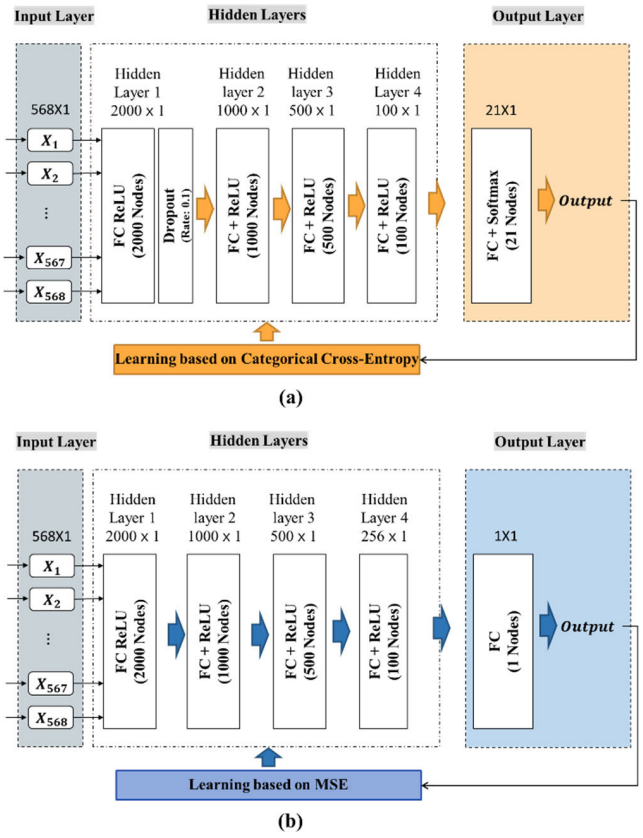

FIGURE 5. DNN-based prediction model architectures: (a) crime type prediction model, (b) CRS prediction model.

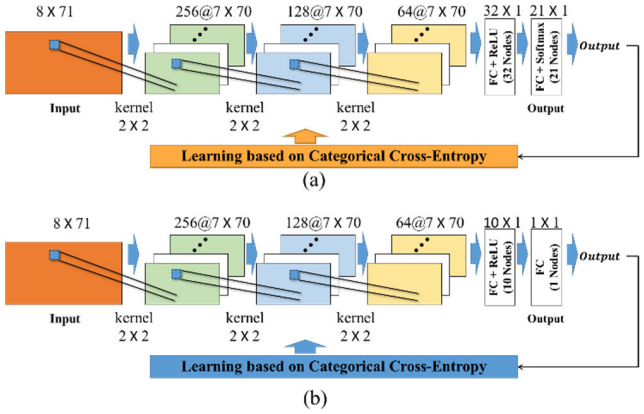

FIGURE 6. CNNN-based prediction model architectures: (a) crime type prediction model, (b) CRS prediction model.

D. MODEL BUILDER

In this model builder, the crime type prediction model and CRS prediction model are developed. For the model design, deep neural network (DNN) architecture and convolutional neural network (CNN) architecture are considered. For both prediction models, input vector X with 568 is applied. The dense layers are updated by the learning process.

Fig. 5-(a) depicts the detailed DNN architecture for the crime type prediction model. It consists of an input layer, hidden layers, and an output layer. The hidden layers include four hidden layers which consist of fully connected (FC) layers and activation functions. For the FC layers, feed forward NN architecture is considered, and for the activation function, the rectified linear unit (ReLU) activation function is considered [23]. The ReLU is the most general activation function that can relax the vanishing gradient problem. The number of nodes of each FC layer is also shown in Fig. 5-(a).

TABLE 7. Hyper-parameters of proposed prediction models.

	Crime type prediction model		CRS prediction model	
	DNN	CNN	DNN	CNN
Learning rate	0.001	0.001	0.001	0.001
Maximum number of epochs	1500	1500	1500	1500
Batch	16	16	16	16

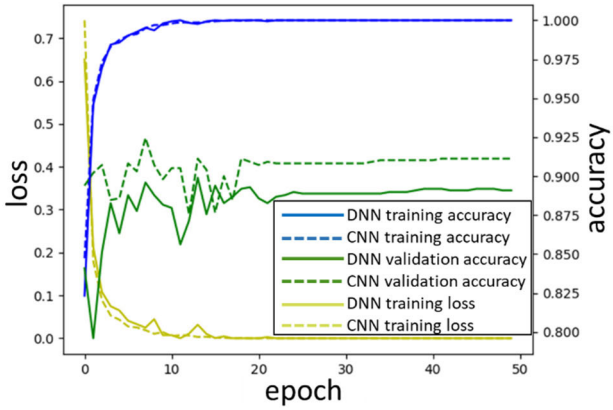

FIGURE 7. Training and validation performance of the crime type prediction model.

In the input layer, the input vector X from the dataset builder is inserted. In the hidden layers, the input data is processed by four FC layers and the ReLU activation function. The final output of the hidden layer is inserted into the output layer with the softmax activation function which is a representative classification function. The softmax function outputs a vector of length 21 which is the same as the length of one-hot vector of Table 5, and can be regarded as a probability distribution over all possible crime types. In the training procedure, all the layers are trained to reduce the cost between the real crime type and classified outputs. For the cost value, categorical cross-entropy is considered. The cost value is minimized by the RMSProp algorithm [24] which is one of the standard methods to train neural networks beyond stochastic gradient descent.

Fig. 5-(b) shows the DNN architecture for the CRS prediction model. In this model, the input dataset is X which is the same as the crime type prediction model. The final output is one numerical value which is the predicted CRS value. The DNN architecture of this model is similar to that of crime type prediction model. Therefore, for the hidden layers, FC layers with feed forward NN architecture is also considered. For the cost value, the mean square error (MSE) between the final output of the CRS prediction model and the real CRS value is considered, and the RMSProp algorithm is also used for the optimization of this model.

Fig. 6-(a) shows the conceptual diagram of the CNN-based crime type prediction model. It consists of input layer, three convolutional layers, one FC layer with ReLU, and output layer with softmax. The specific CNN architectures are described in Fig. 6-(a) including kernel size and CNN structure. Fig. 6-(b) describes the CNN-based CRS prediction model architecture in which the final FC layer does not consider the softmax function, just as in Fig. 6-(b). Although the CNN architecture requires much higher computational complexity, it can give better performance than DNN architecture, generally. Therefore, in our research, the CNN-based prediction models can give better performance than that of DNN-based prediction models. Table 7 shows the hyper-parameters according to the prediction models.

IV. PERFORMANCE EVALUATION AND IMPLEMENTED RESULT

In this section, the performance of the developed system is evaluated. For the performance evaluation, virtual KICS data with 21 crime types are considered. Since the system includes two prediction functions, which are the crime type prediction and CRS prediction, each prediction function performance is evaluated. Furthermore, the developed system also includes a real-time prediction application with GUI. In this section, the functional performance of the real-time prediction application is also checked.

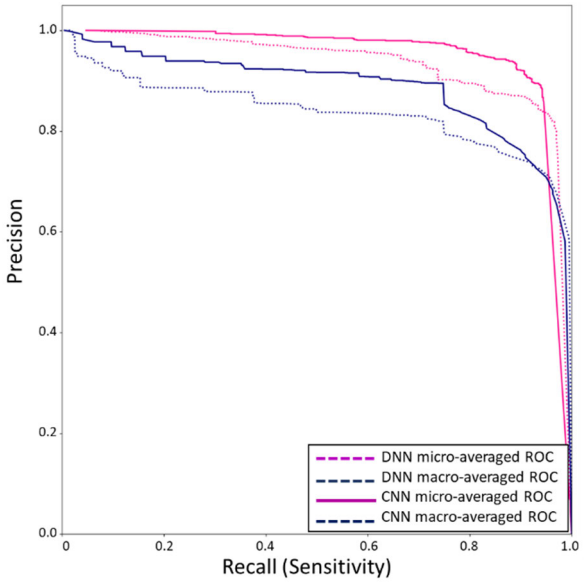

FIGURE 8. ROC curve and AUC for each prediction model.

A. CRIME TYPE PREDICTION

Fig. 7 shows the training and validation performance of crime type prediction models. In the graph, training loss performance, training accuracy performance, and validation accuracy performance are shown according to the prediction model architecture. The validation accuracy performance of the CNN-based prediction model is better than that of the DNN-based prediction model.

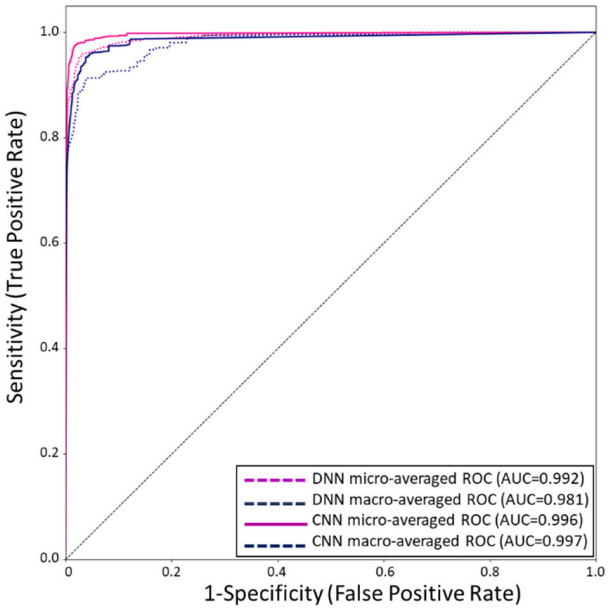

FIGURE 9. Precision-recall curve for each prediction model.

TABLE 8. Performance evaluation according to model architecture.

Classifier	Accuracy	Precision	Sensitivity (Recall)	F1-Score
CNN-based Prediction model	91	92	82	84
DNN-based prediction model	87	77	76	73
SVM	84	80	72	74
Naïve Bayes	83	68	80	72

Fig. 8 and Fig. 9 describe the receiver operating characteristic (ROC) curve and precision-recall curve, respectively [27], [28]. For the average method, both micro-average and macro-average are considered and depicted in Figs. 8 and 9. In Fig. 8, the area under the ROC curve (AUC) is also shown for each prediction model. In both average methods, the CNN-based prediction model shows the better performance than that of DNN-based prediction model. However, since both CNN and DNN-based models achieve more than AUC = 0.98, two prediction model can give superior prediction performance. The precision-recall curve is another tool to visualize the performance of each model. In precision-recall curve, precision is on y-axis and recall is on x-axis. The goal of precision-recall curve is the upper right-hand corner where precision and recall both are 1, which is optimal position for precision-recall curve. As shown in Fig. 8, the CNN-based prediction model can achieve better performance than that of DNN-based prediction model.

Table 8 describes the prediction performance of the proposed crime type prediction models. In addition, for performance comparison, SVM with polynomial kernel function of degree 2, and Bernoulli Naïve Bayes algorithms are also considered [29], [30]. The table describes the computed values of performance evaluation parameters including accuracy, precision, recall and F1-Score. The accuracy is used as a measures of how close the result is to the standard one. For all parameters, the CNN-based prediction model can achieve the best performances among other prediction models. In term of accuracy performance, the performance of the CNN-based prediction model is about 7% and 8% better than those of the SVM algorithm and the naïve Bayes algorithm, respectively.

B. CRS PREDICTION

Fig. 10 describes the training and validation performance of the CSR prediction model. In the validation cases, the performance of the CNN-based model is slightly better than that of the DNN-based model.

To evaluate the CRS prediction performance, the mean absolute percentage error (MAPE) is considered [25], [26]. The MAPE equation is written as

$$MAPE = \frac{100\%}{N} \sum_{n=1}^{N} \left| \frac{X_n - \hat{X}_n}{X_n} \right| \tag{2}$$

where X_n is the actual value and $\hat{X}_n$ is the estimated value of the neural network. Since the MAPE presents the error value as a percentage, a lower value means better prediction performance. In Table 9, the CNN-based CRS prediction performance is about 5% better than that of DNN-based performance. In conclusion, the CNN-based CRS prediction model can achieve an accuracy of more than 80%.

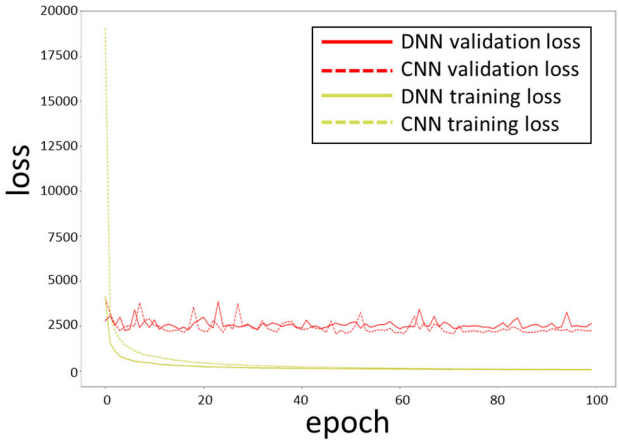

FIGURE 10. Training and validation performance of the CRS prediction model.

C. REAL-TIME PREDICTION APPLICATION

Fig. 11 shows a real-time prediction application platform implemented as a GUI. The top of the platform is the input

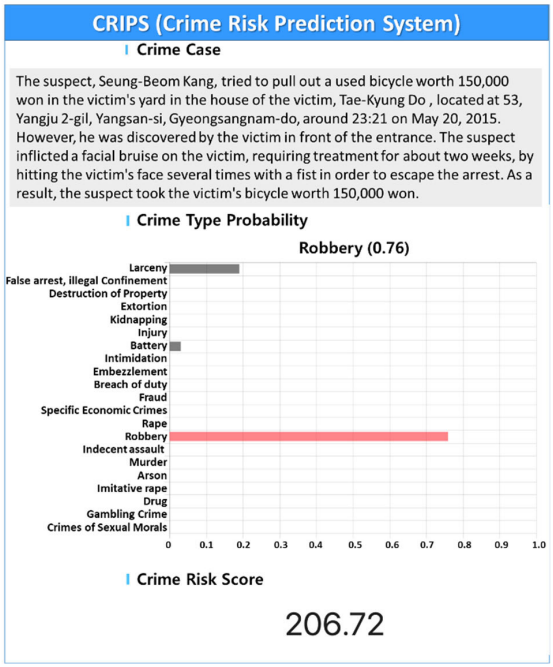

FIGURE 11. GUI-based prediction application result.

TABLE 9. Overall performance according to model architecture.

Prediction Function	Model Architecture	Performance
CRS	CNN-based prediction model	MAPE: 19.4 %
	DNN-based prediction model	MAPE: 24.2 %

part that receives text-based case contents from field personnel. The field personnel enter the crime summary, and the application uses the entered crime summary to predict the crime type and CRS in real-time.

The middle part of the application displays the results of the crime type prediction based on the input crime summary. The middle part of the application displays the results of the crime type prediction based on the input crime summary. The prediction results show the probability values for 21 crime types as a bar graph. At the top of this section, the types of crimes with the highest probability values are displayed. Based on the predicted results, field personnel can quickly identify the type of crime. The predicted CRS is displayed at the bottom of the application. This prediction result is also displayed in real-time.

The platform can predict crime type and CRS and display the prediction results in real-time; therefore, field staff, such as police officers, can easily check predictive information about crimes received through the platform.

V. CONCLUSION

To quickly respond to newly received crimes, this paper designs smart policing technology based on machine learning. The developed system includes two prediction models: crime type prediction model and a CRS prediction model. For both prediction model, DNN architecture and CNN architecture are designed and developed. The performance of each predictive model is also evaluated. In the case of crime type prediction model, the proposed architectures have better performance than those of existing techniques such as SVM and naïve Bayes algorithms. Especially, the accuracy performance of CNN-based crime type prediction model is 7% and 8% higher than SVM algorithm and naïve Bayes algorithm, respectively. In the case of CRS prediction, the CNN-based CRS prediction model can accomplish 80% accuracy. In addition, the real-time operation of the GUI-based smart polishing system developed from a functional point of view is also verified. Designed smart policing technology can predict crime type and CRS using text-based crime event data. A real-time GUI-based application platform is implemented and applied to the predictive models. The developed GUI-based system can predict crime type and CRS in real-time for new input data. In addition, the developed system provides an intuitive GUI, allowing field personnel to use the system efficiently.

REFERENCES

[1] C. G. Demartini, L. Benussi, V. Gatteschi, and F. Renga, "Education and digital transformation: The 'Riconnessioni' project," *IEEE Access*, vol. 8, pp. 186233–186256, 2020.
[2] M. Fanea-Ivanovici and M.-C. Pana, "From culture to smart culture. How digital transformations enhance Citizens' well-being through better cultural accessibility and inclusion," *IEEE Access*, vol. 8, pp. 37988–38000, 2020.
[3] W. L. Thompson and M. F. Talley, "Deep learning for IoT communications : Invited presentation," in *Proc. 53rd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2019, pp. 1–4.
[4] M. Baek, S. Kwak, J. Jung, H. M. Kim, and D. Choi, "Implementation methodologies of deep learning-based signal detection for conventional MIMO transmitters," *IEEE Trans. Broadcast.*, vol. 65, no. 3, pp. 636–642, Sep. 2019.
[5] D. C. Ertugrul and A. H. Ulusoy, "Development of a knowledge-based medical expert system to infer supportive treatment suggestions for pediatric patients," *ETRI J.*, vol. 41, no. 4, pp. 515–527, Jul. 2019.
[6] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, "Lung and pancreatic tumor characterization in the deep learning era: Novel supervised and unsupervised learning approaches," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1777–1787, Aug. 2019.
[7] Y. Kim, D. Ra, and S. Lim, "Zero-anaphora resolution in Korean based on deep language representation model: BERT," *ETRI J.*, vol. 43, no. 2, pp. 299–312, Oct. 2020.
[8] Y. J. Heo, D. Kim, W. Lee, H. Kim, J. Park, and W. K. Chung, "Collision detection for industrial collaborative robots: A deep learning approach," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 740–746, Apr. 2019.
[9] C. Nam, S. Lee, J. Lee, S. H. Cheong, D. H. Kim, C. Kim, I. Kim, and S.-K. Park, "A software architecture for service robots manipulating objects in human environments," *IEEE Access*, vol. 8, pp. 117900–117920, 2020.
[10] D. Yoo, J. Cho, J. Lee, M. Chae, B. Lee, and B. Lee, "FinSNet: End-to-end separation of overlapped fingerprints using deep learning," *IEEE Access*, vol. 8, pp. 209020–209029, 2020.
[11] M. Lim, A. Abdullah, N. Jhanjhi, and M. Khurram Khan, "Situation-aware deep reinforcement learning link prediction model for evolving criminal networks," *IEEE Access*, vol. 8, pp. 16550–16559, 2020.
[12] T. Sangkaran, A. Abdullah, and N. Jhanjhi, "Criminal community detection based on isomorphic subgraph analytics," *Open Comput. Sci.*, vol. 10, no. 1, pp. 164–174, Jul. 2020.
[13] I. Shafi, S. Din, Z. Hussain, I. Ashraf, and G. S. Choi, "Adaptable reduced-complexity approach based on state vector machine for identification of criminal activists on social media," *IEEE Access*, vol. 9, pp. 95456–95468, 2021.

[14] L. Elluri, V. Mandalapu, and N. Roy, "Developing machine learning based predictive models for smart policing," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, Jun. 2019, pp. 198–204.
[15] U. V. Navalgund and K. Priyadharshini, "Crime intention detection system using deep learning," in *Proc. Int. Conf. Circuits Syst. Digit. Enterprise Technol. (ICCSDET)*, Dec. 2018, pp. 1–6.
[16] E. S. Khan, H. Azmi, F. Ansari, and S. Dhalvelkar, "Simple implementation of criminal investigation using call data records (CDRs) through big data technology," in *Proc. Int. Conf. Smart City Emerg. Technol. (ICSCET)*, Jan. 2018, pp. 1–5.
[17] N. Esquivel, O. Nicolis, B. Peralta, and J. Mateu, "Spatio-temporal prediction of Baltimore crime events using CLSTM neural networks," *IEEE Access*, vol. 8, pp. 209101–209112, 2020.
[18] Z. Mason, "WordRank: A method for finding search-ad keywords for internet merchants," in *Proc. 2nd Int. Conf. Internet Web Appl. Services (ICIW)*, May 2007, p. 12.
[19] A. Kritikopoulos, M. Sideri, and I. Varlamis, "Wordrank: A method for ranking web pages based on content similarity," in *Proc. 24th Brit. Nat. Conf. Databases (BNCOD)*, Jul. 2007, pp. 92–100.
[20] S. Amin, M. I. Uddin, S. Hassan, A. Khan, N. Nasser, A. Alharbi, and H. Alyami, "Recurrent neural networks with TF-IDF embedding technique for detection and classification in tweets of dengue disease," *IEEE Access*, vol. 8, pp. 131522–131533, 2020.
[21] I. Yahav, O. Shehory, and D. Schwartz, "Comments mining with TF-IDF: The inherent bias and its removal," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 3, pp. 437–450, Mar. 2019.
[22] H. Kim and H. Kim, "Effective integration of automatic word spacing and morphological analysis in Korean," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2020, pp. 275–278.
[23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.
[24] M. C. Mukkamala and M. Hein, "Variants of RMSProp and Adagrad with logarithmic regret bounds," in *Proc. ICML*, Aug. 2017, pp. 2545–2553.
[25] L. Mendo, "Estimation of a probability with guaranteed normalized mean absolute error," *IEEE Commun. Lett.*, vol. 13, no. 11, pp. 817–819, Nov. 2009, doi: 10.1109/LCOMM.2009.091128.
[26] S. G. N and G. S. Sheshadri, "Electrical load forecasting using time series analysis," in *Proc. IEEE Bengaluru Humanitarian Technol. Conf. (B-HTC)*, Oct. 2020, pp. 1–6.
[27] G. De Carvalho Bertoli, L. A. Pereira Junior, O. Saotome, A. L. Dos Santos, F. A. N. Verri, C. A. C. Marcondes, S. Barbieri, M. S. Rodrigues, and J. M. P. De Oliveira, "An end-to-end framework for machine learning-based network intrusion detection system," *IEEE Access*, vol. 9, pp. 106790–106805, 2021.
[28] S. A. Khan and Z. Ali Rana, "Evaluating performance of software defect prediction models using area under precision-recall curve (AUC-PR)," in *Proc. 2nd Int. Conf. Advancements Comput. Sci. (ICACS)*, Feb. 2019, pp. 1–6.
[29] N. Kalcheva, M. Karova, and I. Penev, "Comparison of the accuracy of SVM kemel functions in text classification," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA)*, Sep. 2020, pp. 141–145, doi: 10.1109/BIA50171.2020.9244278.
[30] Noviantho, S. M. Isa, and L. Ashianti, "Cyberbullying classification using text mining," in *Proc. 1st Int. Conf. Informat. Comput. Sci. (ICICoS)*, Nov. 2017, pp. 241–246, doi: 10.1109/ICICOS.2017.8276369.

MYUNG-SUN BAEK (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in information and communication engineering from Sejong University, Seoul, South Korea, in 2003, 2005, and 2009, respectively. Since 2009, he has been with the Electronics and Telecommunication Research Institute (ETRI), where he is currently a Senior Researcher. His research interests include digital broadcasting systems, MIMO signal processing, full duplex, deep learning-based physical layer design, and smart policing technology.

WONJOO PARK received the B.S. and M.S. degrees in information and communications engineering from Chungnam National University, Daejeon, Republic of Korea, in 1998 and 2000, respectively. She joined the Electronics and Telecommunications Research Institute (ETRI), Daejeon, in 2000, where she is currently a Principal Researcher. She focuses on deep learning, NLP, and text mining.

JAEHONG PARK received the B.S. and M.S. degrees in computer engineering from Chungnam National University, Daejeon, South Korea, in 2000 and 2003, respectively. He is currently a Software Engineer, the Founder, and the CEO of MyCQ Inc. Prior to founding MyCQ Inc., in 2009, he was with the Electronics and Telecommunications Research Institute (ETRI), where he developed and conducted research on event-driven database systems. His research interests include invention and implementation of real-time stream data processing systems, such as continuous query language, high-performance messaging queue, and intelligent event detection engines.

KWANG-HO JANG received the B.S. and M.S. degrees from the Korean National Police University, Asan, South Korea, in 1997 and 2006, respectively, and the Ph.D. degree in public administration from Myongji University, Seoul, South Korea, in 2018. Since 2018, he has been with the Korean National Police University, where he is currently the Director of the SMART Policing Intelligence Center. His research interests include crime analysis planning, SMART policing, and data driven policing.

YONG-TAE LEE received the B.S. and M.S. degrees from Korea Aerospace University, Goyang, South Korea, in 1993 and 1995, respectively, and the Ph.D. degree from Yonsei University, Seoul, South Korea, in 2007.

Since 1995, he has been with the Defense & Safety ICT Research Department, ETRI, Daejeon, South Korea, where he is currently a Principal Researcher and the Assistant Vice President. Since 2014, he has been a Professor with the University of Science and Technology, Daejeon. His current research interests include digital transformation technology with AI, big-data, the IoT, and AR/VR/MR. He is an Associate Editor of the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS Publications Editorial Board.