Received: 25 September 2024 | Revised: 14 March 2025 | Accepted: 16 April 2025

DOI: 10.4218/etrij.2024-0478

ORIGINAL ARTICLE

ETRI Journal WILEY

Distance mean-square loss function for ordinal text classification of emergency service response codes in disaster management

Eungyeol Lee[1,2] | **Sungwon Byon**[1] | **Eui-Suk Jung**[1] | **Eunjung Kwon**[1] | **Hyunho Park**[1,3]

[1]Police Science & Public Safety ICT Research Center, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

[2]Department of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea

[3]Department of Artificial Intelligence, University of Science and Technology (UST), Daejeon, Republic of Korea

Correspondence
Hyunho Park, Police Science & Public Safety ICT Research Center, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea; and Department of Artificial Intelligence, University of Science and Technology (UST), Daejeon, Republic of Korea.
Email: hyunhopark@etri.re.kr; hyunhopark@ust.ac.kr

Funding information
This research was supported by CONNECT-AI platform (CONnected Network for EMS Comprehensive Technical-support using Artificial Intelligence) of the Korea Planning & Evaluation Institute of Industrial Technology (KEIT) funded by the Korea National Fire Agency (No. RS-2023-00237687).

Abstract

The National Fire Agency (NFA) and National Police Agency (NPA) have defined risk levels based on the severity of disasters. Risk-level data possess the characteristics of ordinal data such as NPA's Emergency Service Response Code (ESRC) data, which are classified based on their magnitudes (from C0 to C4). In this study, we propose a distance mean-square (DiMS) loss function to improve the accuracy of ordinal data classification. The DiMS loss function calculates loss values based on the distances between the predicted and true labels: value distances (commonly used in regression analysis for magnitude data) and probability distances (typically used in classification analysis). Therefore, the DiMS loss function contributes to improved accuracy when classifying ordinal data, such as ESRC. In addition, using the DiMS loss function, we achieved state-of-the-art performance in classifying the SST-5 data, which is a representative ordinal dataset. The DiMS loss function for ordinal classification enabled accurate risk recognition. Thus, accurate risk recognition using the DiMS loss function enhances disaster response.

KEYWORDS
disaster response, distance mean-square loss function, emergency service response codes, ordinal classification, text classification

1 | INTRODUCTION

To respond to disasters and minimize damage, it is important to recognize the severity of disaster situations. The National Fire Agency (NFA) and National Police Agency (NPA) of South Korea established risk levels for emergency situations, such as the Emergency Service Response Code (ESRCs), ranging from C0 to C4 [1]. C0

refers to a state of major crimes in progress, C1 indicates an imminent or ongoing threat to life, C2 represents a potential threat to life, C3 denotes a state in which an investigation or expert consultation is required, and C4 signifies a situation in which no dispatch is required. For more information, please refer to Section 4.1. These ESRCs are ordinal data classified according to severity. When emergency calls are received, such as the 112 emergency calls in South Korea, the emergency dispatch response system recommends appropriate response guidelines based on ESRCs. ESRCs are classified as ordinal data based on their severity.

Our study focused on ordinal classifications such as the ESRC classification. Ordinal classification is a classification task for multiclass problems in which labels have an ordered sequence, and classification respects this monotonic relationship. Ordinal classification refers to the classification of target variables that exhibit natural ordering [2]. Suppose that $U = x_1, ..., x_n$ is the set of objects, and x_i has attributes $a_i \in A$ or label $d_i \in D$. The ordinal features of the data denote ordered labels, such that $a_j \leq a_k$ or $d_j \leq d_k$, where $x_j \leq x_k$ [3]. For instance, evaluations can include "very bad," "bad," "neutral," "good," and "very good," and grades can include "A," "B," "C," and "D."

Recent studies on ordinal classification have addressed ordinal characteristics by integrating ranking methods or adjusting loss functions. Ranking-based approaches aim primarily to enhance ordinal regression performance for unseen continuous labels [4–6]. Meanwhile, modifying the loss function often involves assigning weights to each label to improve learning efficiency [7, 8]. In the case of ESRCs, the labels are predefined, and the dataset is sufficiently large to mitigate concerns regarding unseen values or data imbalance. Given these conditions, instead of approaching the task as a regression problem for continuous values, we framed it as a classification problem.

Our study proposes a distance mean-square loss function for ordinal classification, which is a modified version of ordinal log loss [7]. DiMS calculates loss by measuring the distance between the predicted and true labels using two metrics: value distances for magnitude data and probability distances for classification. Accordingly, our research demonstrated the high performance of the DiMS loss function in the classification of ESRCs. We showed that DiMS can be effectively applied to ESRC classification, achieving high accuracy. Furthermore, we demonstrated the effectiveness of the DiMS loss function in general ordinal classification tasks. Through the experiments, we obtained state-of-the-art results for the SST-5 sentiment classification problem, which is a well-known benchmark for text classification with ordinal properties. This result proves that DiMS is suitable not only for specific tasks such as ESRC classification but also for broader ordinal text classification settings such as sentiment analysis, in which the data have an inherent order. As disaster impact patterns become more diverse and complex, risk levels, such as ESRCs, can be further divided into more detailed ordinal datasets. Using DiMS, it is possible to accurately detect finer risk levels during disasters. Accurate detection of risk levels using DiMS will enable appropriate disaster responses, ultimately reducing the damage caused by disasters.

2 | RELATED WORKS

2.1 | Emergency response system

The artificial intelligence (AI) technical architecture for call centers, which forms the overall structure of the AI-based support system, was proposed previously [9]. A previous study [10] proposed a system for general emergency-response situations. Another study [1] proposed a crime response system that can be applied to real-world scenarios based on the degree of danger. In addition, [11] introduced a crime spot prediction model that can be used to prevent emergencies by estimating the number of crimes. [12] proposed a multimodal model that combines bidirectional long short-term memory and graph convolutional networks (GCN) for text and convolutional neural networks and transformers for images, which showed strong performance in crisis classification.

2.2 | Text classification

Traditionally, neural network-, convolutional neural network-, and recurrent neural network-based models have been used in classification tasks [13–16]. Feed-forward neural networks with Word2Vec, GloVe, and recurrent neural network-based long short-term memory have been used for text classification [15, 17, 18]. In addition, graph-based methods using GCN and BertGCNs have been proposed [19, 20]. The most frequently used method is the transformer-based approach [21–24]. Fine-tuning pretrained large language models is the most common method for achieving good performance in natural language processing tasks, including classification [25].

2.3 | Ordinal classification

Some research studies have been conducted on ordinal classification [26]. Ordinal entropy-based methods that apply weights at the activation-function stage before using cross-entropy in the loss function have been

proposed [27]. Weighted Kappa is a well-known method for ordinal classification [28]; however, this method exhibits moderate accuracy. Ordinal log loss and class--distance-based cross entropy are effective methods for ordinal sentence classification [7, 8].

2.4 | Ordinal regression

Ordinal regression is a technique that enhances regression performance by leveraging the ordinal nature of datasets. Tasks such as age estimation and average score prediction are typically framed as regression problems because of the continuous nature of their labels. To address this problem, ranking-based approaches were introduced for ordinal regression [4–6]. However, these methods differ slightly because they primarily focus on regression while effectively learning from infrequent labels in the training data.

3 | METHODS

3.1 | Distance mean-square loss function

We propose a distance mean-square (DiMS) loss function motivated by the weight determination process of the ordinal log loss and class-distance-based cross-entropy [7, 8]. The DiMS loss function replaces the weight with the power distance between the label and the target. The DiMS loss function is defined as follows and can be expressed mathematically as

$$L(T,Y^{\theta};\alpha)=\frac{1}{nl}\sum_{i=1}^{n}\sum_{j=1}^{l}(|A(T_i)-j|+1)^{\alpha}\left(T_{ij}-Y_{ij}^{\theta}\right)^2, \quad (1)$$

where $T\in\mathbb{R}^{l\times n}$ is the target matrix representing one-hot encoding, $Y^{\theta}\in\mathbb{R}^{l\times n}$ is the prediction from a classification model with parameters θ,α is the hyperparameter that needs to be tuned experimentally, l is the number of labels, and n is the batch size. T_i is an array consisting of zeros except when T_{ij} is 1, which indicates the target index j. $A(T_i)$ represents the argmax of T_{ij}, as shown below.

$$A(T_i)=\arg\max_{1\le j\le l}T_{ij}. \quad (2)$$

Figure 1 shows the DiMS calculation using six labels. DiMS assigns higher penalties to predictions that deviate further from the correct label, whereas smaller penalties are assigned to predictions that deviate less from the correct label.

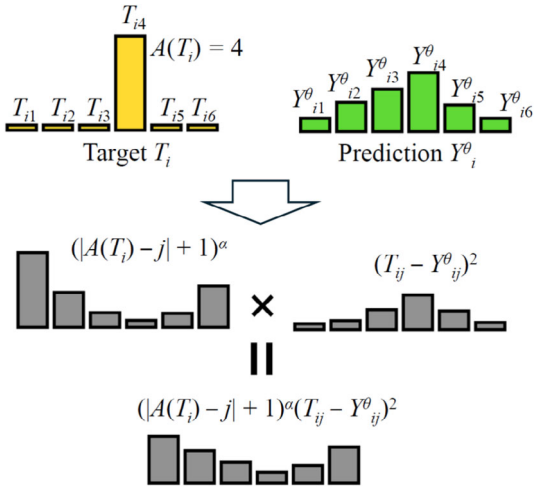


FIGURE 1 Example of the distance mean-square (DiMS) calculation with six labels.

3.2 | Training algorithm

The training algorithm followed the standard procedure used in known neural network model learning processes, with the only difference being the loss function, as shown in Algorithm 1. The model parameters were updated by minimizing the loss function using the values predicted by the model. The optimization direction was determined using an optimizer. Algorithm 1 describes the learning process of the proposed model.

Algorithm 1 Pseudocode for numerical ordering learning

```
Initialize model M_θ with random parameters θ
Dataset batch B with size n
repeat
    for b = b_1 ··· b_n, B do
        X, T ← b // Divide b into input X and label T
        Y ← M_θ(X)
        l ← ∇_θ L_D(T, Y; α) // Calculate DiMS loss
        θ ← Optimizer(θ, l) // Update parameter θ
    end for
until M_θ is trained enough
```

4 | CLASSIFICATION FOR EMERGENCY SERVICE RESPONSE CODES

In this section, we describe the methods used for ESRC classification. To respond effectively to crime scenes, it is crucial to classify and prioritize the urgency and

importance of police responses to reported situations. Therefore, the ESRC was proposed to enable a more efficient and appropriate police response. To further improve this system, an emergency response decision support system should be developed, and more accurate text classification techniques should be applied.

4.1 | ESRC dataset description

ESRCs is an unpublished dataset that matches emergency response codes to report sentences received in Korea. Emergency response codes categorize responses into five levels, from the most urgent (C0) to situations in which dispatch is not required (C4). A detailed breakdown of the emergency code classification criteria is provided below:

1. **C0**: Code 1 cases, including crimes in progress and violent crimes.
2. **C1**: Cases where there is an imminent or ongoing danger to life or body, immediately after the incident, or when the person is an active criminal.
3. **C2**: Situations where there is a potential risk to life or body, or for crime prevention purposes.
4. **C3**: Cases where immediate on-site action is unnecessary, but an investigation or professional consultation is required.
5. **C4**: Non-urgent complaints or counseling reports.

The data were organized by labeling the response codes based on 16376 police reports, including codes C0–C3. The data were divided into 11522, 2432, and 2422 entries (in a 6:2:2 ratio) and classified as training, validation, and test data, respectively. Examples of the emergency service response data are as follows:

1. **C0**: We are fighting, and my husband is holding a knife, threatening to kill me.
2. **C1**: A major accident occurred between a bus and a vehicle.
3. **C2**: A group tried to assault one person, but they stopped. They ran away, so please check.
4. **C3**: A male passenger is drunk in a taxi and will not wake up.

4.2 | Training strategy

ESRC is an ordinal classification problem. The target labels are based on urgency, and this inherent characteristic can be leveraged to improve performance.

The training data consisted of Korean input texts with ESRCs labels. These labels were categorized into five levels (from C0 to C4) depending on the urgency of police support required.

Improving performance by upgrading the model alone may be ineffective because of the curse of dimensionality. However, the ordinal classification approach presented in this study is efficient and demonstrates excellent performance.

The model-learning process follows the procedure outlined in Algorithm 1 in Section 3.2. We fine-tuned a pretrained encoder-based large language model—currently one of the best-performing models for classification problems—using supervised learning to achieve optimal performance [29].

4.3 | Experiments

Experiments were conducted using the proposed method and the DiMS loss function. Our methods were compared with cross-entropy (CE), mean-square error (MSE), and ordinal log loss (OLL) functions under the same conditions. In previous studies, OLL exhibited significantly better performance in ordinal classification compared to the others; therefore, we selected OLL as the primary comparison target [7] (Table 1).

The experimental setup is presented in Table C2 in Appendix C.

An accuracy of 87.8% was achieved using the DiMS and the Korean RoBERTa model. The DiMS loss function outperformed the existing CE loss and showed a slight improvement over OLL in ordinal classification. We confirmed that the DiMS loss function outperformed the strong baseline OLL in terms of accuracy, mean absolute error, and quadratic-weighted kappa. This suggests that applying ordinal weights to mean-square error may be more effective than applying them to cross-entropy. Interestingly, MSE, which was not originally designed for classification, performed better than CE. This suggests that DiMS, which applies distance-based weighting to MSE, may be more effective than OLL, which applies distance-based weighting to CE.

TABLE 1 Benchmark results for ESRC classification.

Model	ACC	QWK	MAE
KLUE RoBERTa large + CE	0.783	0.570	0.245
KLUE RoBERTa large + MSE	0.852	0.761	0.151
KLUE RoBERTa large + $OLL_{\alpha=2}$	0.863	0.771	0.147
KLUE RoBERTa large + $DiMS_{\alpha=2}$	**0.878**	**0.788**	**0.141**

Note: Our method "DiMS" shows the best performance; therefore, we emphasize it in bold.

5 | CONCLUSION

In this study, we proposed a distance loss function called DiMS, which can efficiently train neural network models for ordinal classification tasks. The contributions of this study are as follows.

First, DiMS demonstrated slightly better performance than other loss functions across various ordinal classification tasks. Based on DiMS, we present a new approach for classifying emergency service response codes. Instead of a model-based method, we adopted a method that uses the characteristics of classified data.

In future studies, we plan to conduct a comprehensive quantitative analysis to rigorously validate the effectiveness of the DiMS loss function in ordinal classification. Although we have demonstrated its performance and provided qualitative insights, establishing its advantages through quantitative metrics remains challenging. Furthermore, we aim to apply DiMS to broader decision-making systems beyond emergency response classification, potentially enhancing support for police operations. These directions will be the focus of our future studies.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

ORCID

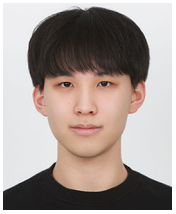
Hyunho Park https://orcid.org/0000-0002-9943-1497

REFERENCES

1. H. Park, J. Yoon, I. Han, S. Byon, E. Kwon, and E.-S. Jung, *Crime response system based on danger degree calculations of crime scenes*, (13th International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea), 2022, pp. 2221–2224. DOI 10.1109/ICTC55196.2022.9952396
2. J. R. Cano, P. A. Gutiérrez, B. Krawczyk, M. Woźniak, and S. García, *Monotonic classification: an overview on algorithms, performance measures and data sets*, Neurocomput. **341** (2019), 168–182. DOI 10.1016/j.neucom.2019.02.024
3. H. Zhu, E. C. C. Tsang, X.-Z. Wang, and R. A. Raza, *Monotonic classification extreme learning machine*, Neurocomput. **225** (2016), 205–213. DOI 10.1016/j.neucom.2016.11.021
4. Y. Gong, G. Mori, and F. Tung, *RankSim: Ranking similarity regularization for deep imbalanced regression*, (Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA), 2022, pp. 7634–7649.
5. N.-H. Shin, S.-H. Lee, and C.-S. Kim, *Moving window regression: a novel approach to ordinal regression*, (Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA), 2022. DOI 10.1109/CVPR52688.2022.01820
6. R. Wang, P. Li, H. Huang, C. Cao, R. He, and Z. He, *Learning-to-rank meets language: boosting language-driven ordering alignment for ordinal classification*, (Proceedings of the 37th International Conference on Neural Information Processing Systems, New Orleans, LA, USA), 2023, pp. 76908–76922.
7. F. Castagnos, M. Mihelich, and C. Dognin, *A simple log-based loss function for ordinal text classification*, (Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea), 2022, pp. 4604–4609. https://aclanthology.org/2022.coling-1.407
8. G. Polat, I. Ergenc, H. T. Kani, Y. O. Alahdab, O. Atug, and A. Temizel, *Class distance weighted cross-entropy loss for ulcerative colitis severity estimation*, arXiv preprint, (2022). DOI 10.48550/arXiv.2202.05167
9. T. Zhang, *Solving large scale linear prediction problems using stochastic gradient descent algorithms*, (Proceedings of the Twenty-First International Conference on Machine Learning, Banff, Canada), 2004. DOI 10.1145/1015330.1015332
10. H. Park, S. Byon, E. Kwon, D. M. Jang, and E.-S. Jung, *Emergency dispatch support system for police officers in incident scenes*, (International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea), 2021, pp. 1542–1544. DOI 10.1109/ICTC52510.2021.9620958
11. G. Hajela, M. Chawla, and A. Rasool, *Crime hotspot prediction based on dynamic spatial analysis*, ETRI J. **43** (2021), no. 6, 1058–1080. DOI 10.4218/etrij.2020-0220
12. J. Wang, S. Yang, H. Zhao, and Y. Chen, *A crisis event classification method based on a multimodal multilayer graph model*, Neurocomput. **621** (2025), 129271. DOI 10.1016/j.neucom.2024.129271
13. A. A. Elngar, M. Arafa, A. Fathy, B. Moustafa, O. Mahmoud, M. Shaban, and N. Fawzy, *Image classification based on cnn: a survey*, J. Cybersecur. Inf. Manag. **6** (2021), no. 1, 18–50.
14. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, *Large-scale video classification with convolutional neural networks*, (Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Columbus, OH, USA), 2014, pp. 1725–1732.
15. T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient estimation of word representations in vector space*, (1st International Conference on Learning Representations, ICLR 2013, Scottsdale, AZ, USA), 2013. DOI 10.48550/arXiv.1301.3781
16. H. Phan, P. Koch, F. Katzberg, M. Maass, R. Mazur, and A. Mertins, *Audio scene classification with deep recurrent neural networks*, (Interspeech, Stockholm, Sweden), 2017. DOI 10.21437/Interspeech.2017-101
17. J. Pennington, R. Socher, and C. Manning, *GloVe: global vectors for word representation*, (Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar), 2014, pp. 1532–1543. DOI 10.3115/v1/D14-1162
18. M. Shi, K. Wang, and C. Li, *A C-LSTM with word embedding model for news text classification*, (IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), Beijing, China), 2019, pp. 253–257. DOI 10.1109/ICIS46139.2019.8940289
19. T. N. Kipf and M. Welling, *Semi-supervised classification with graph convolutional networks*, International Conference on Learning Representations, 2017. https://openreview.net/forum?id=SJU4ayYgl

20. Y. Lin, Y. Meng, X. Sun, Q. Han, K. Kuang, J. Li, and F. Wu, *BertGCN: transductive text classification by combining GNN and BERT*, (Findings of the Association for Computational Linguistics: ACL-IJCNLP), 2021, pp. 1456–1462. DOI 10.18653/v1/2021.findings-acl.126
21. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *Bert: Pre-training of deep bidirectional transformers for language understanding*, (Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and short papers)), 2019, pp. 4171–4186.
22. J. Jung, *Cr-m-spanbert: Multiple embedding-based dnn coreference resolution using self-attention spanbert*, ETRI J. **46** (2024), no. 1, 35–47.
23. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, *RoBERTa: a robustly optimized Bert pretraining approach*, arXiv preprint, (2019). DOI 10.48550/arXiv.1907.11692
24. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention is all you need*, 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
25. J. Howard and S. Ruder, *Universal language model fine-tuning for text classification*, (Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia), 2018, pp. 328–339. DOI 10.18653/v1/P18-1031
26. S. R. Kasa, A. Goel, K. Gupta, S. Roychowdhury, P. Priyatam, A. Bhanushali, and P. Srinivasa Murthy, *Exploring ordinality in text classification: a comparative study of explicit and implicit techniques*, (Findings of the Association for Computational Linguistics, Bangkok, Thailand), 2024, pp. 5390–5404. DOI 10.18653/v1/2024.findings-acl.320
27. S. Zhang, L. Yang, M. B. Mi, X. Zheng, and A. Yao, *Improving deep regression with ordinal entropy*, (The Eleventh International Conference on Learning Representations, Kigali, Rwanda), 2023. https://openreview.net/forum?id=raU07GpP0P
28. J. de la Torre, D. Puig, and A. Valls, *Weighted kappa loss function for multi-class classification of ordinal data in deep learning*, Pattern Recognit. Lett. **105** (2018), 144–154. DOI 10.1016/j.patrec.2017.05.018
29. S. Chi, X. Qiu, Y. Xu, and X. Huang, *How to fine-tune bert for text classification?*, arXiv preprint, (2019). DOI 10.48550/arXiv.1905.05583
30. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: machine learning in Python*, J. Mach. Learn. Res. **12** (2011), 2825–2830.
31. R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, *Recursive deep models for semantic compositionality over a sentiment treebank*, (Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA), 2013, pp. 1631–1642.

AUTHOR BIOGRAPHIES



Eungyeol Lee has been pursuing undergraduate studies in electrical engineering and computer science at the Gwangju Institute of Science and Technology since 2021. In 2024, he began working as a student researcher at the Police Science and Public Safety ICT Research Center of the Electronics and Telecommunications Research Institute (ETRI). His main research interests include deep learning, reinforcement learning, and their applications in optimization and scientific computing.

Sungwon Byon received his BS degree in computer science from Soongsil University, Republic of Korea in 1986 and his MS degree in computer science from KAIST, Republic of Korea in 1989. From 1993 to 2014. Since 2014, he has been a research fellow at the Police Science and Public Safety ICT Research Center, ETRI, Daejeon, Korea. His research interests include deep learning-based field-assistant technologies for enhancing the efficiency of police operations.

Eui-Suk Jung received his BS and MS degrees in electronics from Korea Aerospace University, Goyang, Republic of Korea in 1992 and 1994. He earned his PhD from Yonsei University, Seoul, Republic of Korea, in 2015. From 1994 to 2002, he contributed as a senior researcher at ETRI, where he worked on the development of a CDMA mobile exchange system and ATM switching system. From 2002 to 2019, he worked on the FTTH WDM system platform, low-power set-top box technology, and multi-log platforms as a principal researcher. From 2019 to 2025, he served as the director of the Police Science and Public Safety ICT Research Center at ETRI. Since 2025, he has been conducting research on artificial intelligence systems for police and firefighting services as a principal researcher. His current research interests include WDM hybrid networks, access network protection, and the convergence of wireless and public safety technologies.

Eunjung Kwon received the ME degree in computer engineering from Ewha Womans University, Korea, in 2000. In 2000, she joined ETRI, Daejeon, Korea, where she worked as a principal researcher. She has contributed to the development of broadcasting systems for digital cable TV, cloud set-top boxes, multi-log technologies, and more. In 2024, she received her Ph.D. in data and software engineering from Chungnam National University, Korea. Since 2025, she has served as the Director of the Police Science & Public Safety ICT Research Center at ETRI. Her research interests include intelligent software platforms, computer vision technologies, deep learning-based disaster awareness, and response support modeling.

Hyunho Park received his BS degree in electrical engineering and computer science from Kyungpook National University, Daegu, Republic of Korea, in 2005, his MS degree in information and communications from the Gwangju Institute of Science and Technology, Republic of Korea, in 2007, and his PhD in broadband network technology from the University of Science and Technology (UST), Daejeon, Republic of Korea, in 2014. Since 2014, he has been a senior researcher at the Police Science and Public Safety ICT Research Center of ETRI, Daejeon, Republic of Korea. From 2011 to 2014, he served as the secretary of the IEEE 802.21c Task Group on Single Radio Handover Optimization. From 2015 to 2018, he was an editor of the ITU-T Q14/16 standard H.DS-CASF: Digital Signage: Common Alerting Service framework. Since 2025, he has been a professor at the Department of Artificial Intelligence at UST. His research interests include public safety technologies based on deep learning mechanisms.

How to cite this article: E. Lee, S. Byon, E.-S. Jung, E. Kwon, and H. Park, *Distance mean-square loss function for ordinal text classification of emergency service response codes in disaster management*, ETRI Journal (2025), 1–8, DOI 10.4218/etrij.2024-0478.

APPENDIX A: Effect of α

α is a hyperparameter used to adjust the weighting in the DiMS loss function. Because the degree of ordinal characteristics in a dataset is often unknown, optimizing α to obtain the best results is challenging. This parameter controls how much the contribution to the loss is weighted as the distance between predicted and true classes increases. To explore the impact of α, we conducted experiments using a custom dataset generated with Scikit-learn [30].

We conducted the experiments using a custom dataset generated with the *make_regression* function in the Scikit-learn package [30]. We created 10,000 data points with 99 features, of which only ten were related to the target values. Subsequently, we separated the range of values and labeled them in increasing order. Thus, the dataset exhibited ordinal characteristics.

We conducted the following experiments under certain conditions: The data were divided into training and test sets at a ratio of 8:2, and each experiment was conducted with 5, 7, 10, 15, 20, and 50 labels. We experimented with how DiMS works depending on the number of labels. The MLP models and experimental setup were almost identical, except for the following: The detailed experimental setup is presented in Appendices C and Table C2.

Table A1 lists the performances of each hyperparameter α for various numbers of labels. The experiments showed that α should be carefully adjusted for optimal performance. The best results were obtained when $\alpha = 0.3, 0.5, 1, 1.5, 2.5$, and 3. However, when α is excessively large (e.g., $\alpha = 5$), the performance decreases. Based on these results, we recommend tuning the α value within the range of 0.3 to 3 to achieve high performance. Additionally, we observed an interesting trend: As the number of labels increased, a smaller α value tended to be more suitable for learning.

TABLE A1 Experimental results of loss functions according to α.

α	**0.3**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**5**
5	96.5	96.8	97.1	96.9	97.1	97.0	**97.2**	97.1
7	95.7	96.2	96.6	**96.7**	96.5	**96.7**	96.3	96.0
10	93.3	94.3	**94.7**	**94.7**	94.6	94.6	94.4	93.1
15	91.0	92.3	**92.5**	92.3	92.3	92.0	91.8	87.3
20	89.4	**90.4**	89.8	89.9	89.9	89.6	88.6	77.5
50	**78.8**	78.6	73.3	76.7	73.2	67.8	61.7	32.8

Note: The bold text indicates the highest accuracy outside the 99% confidence interval error range.

APPENDIX B: DiMS on SST-5

In this section, we demonstrate the effectiveness of DiMS for other text classification problems. We trained the language model on the full SST-5 [31] dataset, which is widely used in text classification tasks. It is one of the most commonly used benchmarks for ordinal classification problems.

The experiments were conducted by fine-tuning the RoBERTa-large [23] pretrained models without altering the model architecture. Similar to the previous ESRCs classification task, it was trained using ordinal learning with DiMS (Algorithm 1 in Section 3.2). The detailed experimental setup can be found in Appendix C and Table C1.

As shown in Table B1, our RoBERTa-DiMS$_{\alpha=2}$ model outperformed all other methods, including OLL, across all metrics, such as accuracy, mean absolute error, and quadratic weighted Kappa. It achieved a classification accuracy of 0.618 on the SST-5 dataset, setting a new state-of-the-art score. This demonstrates that DiMS is more effective than OLL in various tasks.

TABLE B1 Benchmarks on the SST-5 dataset.

Model	ACC	QWK	MAE
RoBERTa large + CE	0.583	0.682	0.774
RoBERTa large + MSE	0.602	0.714	0.751
RoBERTa large + OLL$_{\alpha=1.5}$	0.610	0.718	0.739
RoBERTa large + DiMS$_{\alpha=2}$	**0.618**	**0.728**	**0.731**

Note: Our method "DiMS" shows the best performance; therefore, we emphasize it in bold.

APPENDIX C: Experimental Details

We experimented with ordinal classification using the ESRCs, custom Scikit-learn dataset [30], and SST-5 [31]. The experiments followed the procedure outlined in Algorithm 1. The detailed experimental setup is presented in Tables C1 and C2. The code was executed on an RTX 4090 GPU.

C.1 | Text classification

For text classification, the model was built by adding a linear layer for classification and dropout layers to pretrained models, such as BERT and RoBERTa. The text data were preprocessed into special tokens to enable learning with the BERT-based model. Specifically, tokens [**cls**] and [**sep**] were added before and after the data, respectively.

TABLE C1 Experimental setup for SST-5 text classification.

Parameter	Value
Maximum (max) text length	512
Batch size	16
Learning rate	1e-6
Dropouts	0.3
Optimizer	AdamW
Scheduler	Linear schedule with warmup

Note: All the experiments followed the setup detailed here.

C.2 | Numerical ordinal classification

The ordinal classification model is composed of two linear layers. To enhance training stability, batch normalization, dropout, and ReLU activation functions were applied between the two linear layers.

TABLE C2 Experimental setup for the ordinal classification.

Parameter	Value
Learning rate	1e−6
Hidden layer	250
Dropouts	0.1
Optimizer	AdamW
Scheduler	Step learning rate

Note: All the experiments followed the setup described here.