ISSN 2466-135X
Vol.11, No.1

The 11th International Conference on

BIG DATA APPLICATIONS AND SERVICES (BIGDAS2023)

Proceeding

MULTIMEDIA
LIFE
STORAGE
NETWORK
DATABASE
SYSTEM
BIG DATA
SOCIETY
SCIENCE
CLOUD
BUSINESS
ANALYSIS
TREND
CLUSTER
GRAPHICS
VISUALIZATION

August 16-18, 2023
Danang, Vietnam

Hosted by
Korea Big Data Service Society

THE KOREA BIG DATA SERVICE SOCIETY
한국빅데이터서비스학회
충북대학교 빅데이터연구소
ĐẠI HỌC DUY TÂN

Table of Contents

NLP-Based Predictive Model for Identifying and Presenting Construction Accident Scenarios

Seung-Hyeon Shin[1] and Jeong-Hun Won[2]

[1] Department of Big Data, Chungbuk National University, Cheongju, Republic of Korea
[2] Department of Safety Engineering & Department of Big Data, Chungbuk National University, Cheongju, Republic of Korea, shshin0317@chungbuk.ac.kr

Abstract. This study presents an NLP-based predictive model for identifying and presenting similar construction accident scenarios with the aim of enhancing safety and health management. The model used a BERT-based deep learning approach trained on 18,000 accident cases from the Korea Authority of Land and Infrastructure Safety. By inputting construction site information, the model provides users with relevant accident cases, similarity measures, and on-site information at the time of the accident, enabling stakeholders to develop effective risk reduction measures. The proposed model demonstrated the potential to raise awareness of possible accidents and hazards, ultimately contributing to a safer construction industry with reduced fatality rates.

Keywords: NLP-based predictive model, construction accident scenarios, construction S&H

1 Introduction

The construction industry is one of the most dangerous industries in the world [1]. Construction workers account for approximately 7% of the total number of workers in all industries, but the number of fatalities accounts for a much higher proportion, including in South Korea (hereafter Korea) [2]. In the last ten years, the work-related fatality rate in the construction industry has been higher than that in the Korean manufacturing industry. As a result of analyzing the work-related fatality rate by industry in Korea from 2013 to 2022, the fatality rate in the construction industry is 3.09 times higher on average than that in the manufacturing industry [3]. Compared to the change in the fatality rate, the construction industry has been gradually decreasing since 2021, after showing an increasing trend from 2016 to 2020.

Through many studies on the prevention of fatalities at construction sites, it has been found that one of the measures to prevent fundamental accidents is to include all construction stakeholders in the safety and health(S&H) management system [4,5]. The Korean government introduced various systems, such as 'Design for Safety' and 'Safety and Health ledgers, ' to include stakeholders, such as construction clients and designers, in the S&H management system based on previous studies. The Korean government implemented several policies to prevent industrial accidents at construction sites.

However, the effectiveness of preventing industrial accidents is insufficient. The fundamental reason for the low effectiveness of the system is that construction stakeholders lack knowledge of the S&H. Owing to the low S&H management capabilities of construction stakeholders, stakeholders have limitations in S&H management activities, such as identifying hazards that occur at their sites and preparing risk reduction measures. One of the most effective ways to improve stakeholders' S&H management capabilities is to make them aware of the accidents that may occur during construction work and the hazards that cause them. Therefore, this study proposes a model that presents accident scenarios at construction sites for stakeholders.

2 Method

In this study, we develop an accident prediction model for the construction industry using a two-step approach: data collection and model training.

2.1 Data collection

The model was developed using data from accident reports comprising 18,000 cases collected by the Korea Authority of Land and Infrastructure Safety (KALIS). The data included information on the type of accident, date and time of the accident, weather conditions, type of facility, type of work being performed, and the number of fatalities and injuries. Data collection was conducted through web crawling on the sites provided by KALIS. The collected data were cleaned, pre-processed, and stored in a separate database to ensure quality and consistency.

2.2 Model training

These data were used to train a BERT model, which is a type of deep learning model that is particularly effective for processing natural language data. During the training phase, the dataset was split into training and validation sets to evaluate the performance of the model and to prevent overfitting. The BERT model was trained to predict the cause of accidents based on the other variables in the dataset using techniques such as fine-tuning, cross-validation, and hyperparameter optimization to enhance the model's predictive capabilities.

2.3 Model evaluation

Once the model was trained, its performance was assessed through qualitative evaluation conducted by construction safety experts. These experts examined the output of the model to ensure its interpretability, relevance, and applicability in real-world construction scenarios. They provided valuable feedback and insights, which helped

refine the model and ensure its effectiveness in predicting the cause of accidents and presenting similar accident scenarios in the construction industry.

3 Results

As illustrated in Fig.1, the model presented in this study enables users to input construction site information, such as the process rate, construction period, construction cost, weather, number of workers, and work details, in the form of natural language. Upon input, the model retrieves accident information from similar sites based on the embedded accident data. To provide a comprehensive understanding of the similarity between the user-input site information and retrieved accident sites, the model calculates and presents a similarity score using cosine similarity. This score aids users in evaluating the relevance of the accident scenarios to their construction sites. Furthermore, the model provides site information at the time of an accident, offering users a valuable context for accident occurrence.

By examining the similarity score and on-site information at the time of the accident, users can assess the likelihood of a proposed accident occurring at a construction site. This information empowers them to devise and implement targeted measures to prevent potential accidents, ultimately enhancing S&H management in the construction industry.

Fig. 1. Accident case proposed model

4 Conclusion

This study aims to develop a model that presents similar field accident cases in the construction industry by leveraging Natural Language Processing (NLP) techniques. Our research contributes to the body of knowledge on safety and health (S&H) management in the construction industry and proposes a novel approach to improve

stakeholder awareness of potential accidents and hazards at construction sites. The developed model uses a BERT-based deep-learning approach trained on a comprehensive dataset of 18,000 accident cases collected from the Korea Authority of Land and Infrastructure Safety (KALIS).

Our findings indicate that the proposed model can effectively identify similar accident cases based on the input of construction site information, such as the process rate, construction period, construction cost, weather, number of workers, and work details. By presenting users with accident information from similar sites, along with similarity measures and on-site information at the time of the accident, the model enables stakeholders to better understand potential risks and hazards and develop more effective risk-reduction measures. While the current study demonstrated promising results, there are some limitations that should be addressed in future research. First, the model's performance can be further improved by incorporating additional variables and data sources, such as geographical factors and safety-related regulations. Additionally, the applicability of the model can be expanded to encompass a wider range of construction activities and industries. Finally, future studies should investigate the practical implications of the model in real-world construction projects to assess its effectiveness in reducing accident rates and in enhancing S&H management. This study presents a novel NLP-based model for predicting and presenting similar field accident cases in the construction industry. The model demonstrates the potential to enhance the safety and health management capabilities of construction stakeholders by increasing their awareness of possible accidents and hazards. By adopting this model, construction stakeholders can proactively identify and address risks, ultimately contributing to a safer construction industry with lower fatality rates.

Acknowledgments. This work was supported by the National Research Foundation(NRF), Korea, under project BK21 FOUR; and a Human Resources Development of the Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korean government. (No. 20224000000070).

References

1. Pinto, A., Nunes, I.L., Ribeiro, R.A.: Occupational risk assessment in construction industry – Overview and reflection. Saf. Sci. vol. 49, no. 5, pp. 616--624. (2011)
2. Sunindijo, R.Y., Zou, P.X.W.: Political skill for developing construction safety climate. J. Constr. Eng. Manag. vol. 138, no. 5, pp. 605--612. (2019)
3. Ministry of Employment and Labor: Analysis of industrial accident status. (2022)
4. Huang, X. and Hinze, J., "Owner's role in construction safety", Journal of Construction Engineering and Management, vol. 132, no. 2, pp. 164-173. (2006)
5. Choe, S. Y., Seo, W. K. and Kang, Y. C., "Inter– and intra-organizational safety management practice difference in the construction industry", Safety Science, vol.128, no. 2, pp. 164-173. (2020)

Cloud and Edge Computing Model for Smart IoT Systems

In Joo[1], Kwan-Hee Yoo[1*]

[1] Dept. of Computer Science, Chungbuk National University, South Korea
{jooin95, khyoo}@chungbuk.ac.kr
*Corresponding Author

Abstract. Smart IoT is a concept that utilizes advanced technologies such as cloud and artificial intelligence (AI) to enhance agricultural productivity. In a Smart IoT scenario, a large amount of data is collected from various sources such as wireless sensor networks, connected external data centers, monitoring cameras, and smartphones. However, one of the major challenges with such big data is the diversity in formats and meanings. Additionally, the lack of standardized practices for data and system integration in the Smart IoT ecosystem limits the interoperability of functionalities. These challenges pose significant issues in terms of collaborative service provisioning, data and technology integration, and data sharing. To address these challenges, this whitepaper proposes a platform-centric approach to effectively and reliably build Smart IoT systems. The proposed platform approach considers edge computing and cloud computing, integrating them with offloading programs. This enables the Smart IoT platform to improve efficiency, reduce costs, and enhance performance of connected devices. This paper aims to introduce the concept of a platform approach for Smart IoT and review the associated requirements.

Keywords: Smart IoT; Platform approach; Interoperability; Offloading; Cloud computing; Edge computing;

1 Introduction

The smart IoT system is being digitalized due to the rapid advancements in ICT, smart technology, IoT, AI, and other related fields. It is considered as part of the fourth industrial revolution [1,2]. For instance, in a smart farm utilizing smart IoT, farmers can manage inputs such as fertilizers, pesticides, and animal feed, leading to waste reduction, labor and cost savings, and achieving a more sustainable environmental impact [3,4]. In a smart IoT scenario, a large volume of real-time and high-resolution data is generated from remote and automated sensor systems. Analyzing this data helps filter out inaccurate or erroneous information and calculate user recommendations to enhance productivity [6]. Although data-driven solutions offer various benefits in agriculture, challenges remain in terms of data integration, processing, usage processes, and protocols. Moreover, the technology required to analyze and refine this data can be too burdensome for execution on clients' devices.

Additionally, transmitting heavy real-time data to a central cloud server can incur substantial network costs. Consequently, edge computing is necessary for managing and analyzing the generated data. This paper proposes a strategy for remote management of smart IoT facilities using edge computers and central cloud computers. Such a strategy fosters data sharing and collaboration in agricultural applications, thereby increasing productivity and reducing resource waste. Moreover, this approach strengthens the relationship between clients and system providers, enabling safe tracking of production cycles, livestock, dairy products, and facilitating decision-making and data management.

2 Key components of a smart IoT system

This paper proposes a design strategy for real-time data processing using the integration of various systems to enhance the efficiency of smart IoT data processing. By utilizing a smart IoT platform approach, farmers, researchers, technology providers, and other stakeholders can have standardized and reliable solutions to collect and share information, resources, and experiences to improve productivity and performance in various environments such as smart cities, factories, and farms.

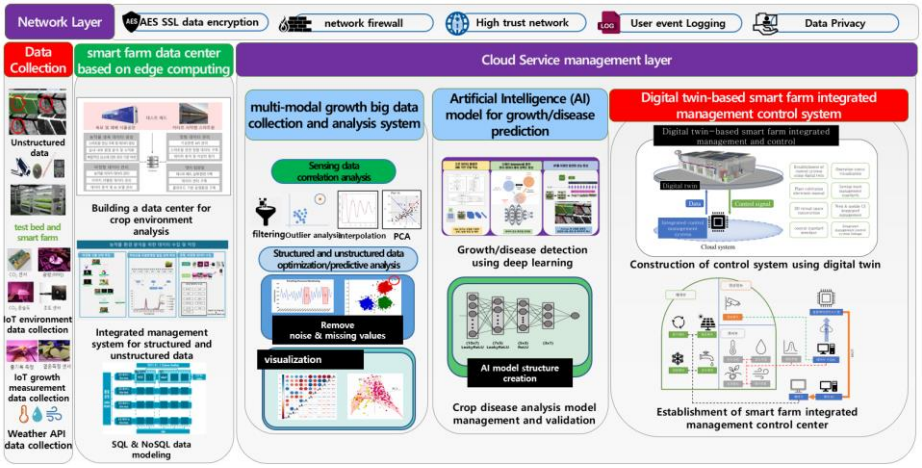


Fig. 1. Smart farming component that facilitates the integration, processing and use of farm data.

According to Figure 1, in the data collection phase of a Smart IoT system, structured and unstructured data is collected from various sources such as sensors, workers, real-time data, weather information, and external databases like API data. These data from different sources may have incompatible formats, so data preprocessing is necessary to store them in a unified database. Additionally, the collected data may contain incomplete records, missing values, outliers, and abnormal instances. Cloud-based data preprocessing can be employed to refine and manage the data, ensuring data consistency, completeness, and accuracy [28]. The cloud layer consists of edge computers and central cloud components, which are highly interconnected and capable of handling numerous Smart IoT data processing tasks.

The decision on whether data processing should take place at the edge computers or the central management layer depends on factors such as network status, data size, and the condition of cloud nodes.

Table 1. Data processes used in a smart IoT strategy and the resulting amount of computation.

	Processing Order	Calculation Amount
Structured data collection	0	0
Unstructured data collection	0	0
Structured Data Preprocessing	1	1
Unstructured Data Preprocessing	1	1
Data Integration	3	0
Data analysis through algorithms	2, 4	2
Data analysis through AI	2, 4	2
Data Monitoring	5	0
Automatic System Control	5	1
Integrated System Control	5	1

The above chart categorizes the functions used in Smart IoT based on their sequence and computational complexity. Data collection is positioned at the beginning of the program flow. Next, data preprocessing takes place, which involves verifying the validity of each data and processing it into a usable format. In the analysis phase, algorithms and AI algorithms generally have a size that is difficult to handle on the client side. Therefore, in the analysis phase, considering the data size, network status, distance, and required computational complexity, it is necessary to decide whether to execute the functions at the edge computer or the central cloud, taking into account the overall network latency cost and computing energy consumption. Various scoring techniques exist for this purpose, such as Markov-based offloading decisions, graph-based offloading decisions, optimization-based offloading decisions, and deep learning-based offloading decisions. These operations are performed in the data center, which also plays a role in data integration and management. Subsequently, the entire system is integrated and controlled automatically or manually at the central cloud. Furthermore, ensuring the trustworthiness of the entire network is an essential mechanism in the Smart IoT platform for data security and privacy protection. This allows service providers to evaluate the trustworthiness of participants in the system and restrict activities of participants with low trustworthiness. Security techniques such as AgriTrust, Attribute-Based Access Control (ABAC), Ametepe et al., Wen Xue et al., etc., form a higher level of reliability when using smart applications. Additionally, monitoring the interactions of the system and updating trust metrics such as reliability, robustness, and stability over time. By utilizing this approach, Smart IoT devices can broadcast feedback for each transaction to the network, which can be utilized in the trust management framework. This enables the improvement of trust management and evaluation for Smart IoT devices.

3 Concluding Remarks

Table 2. Existing offloading method and proposed offloading method.

Nodes	existing strategy	Suggested Strategy
Sensing Node	Sensor control, transmission of collected data Collection Node	Collected data processing and sensor control, collection node transmission
Collection Node	Data integration and transfer to database	Primary analysis and database integration, analyzed data collection and sensing node control
Control Node	Visualization of analysis and analysis results, sub-node control	Visualization of analysis and analysis results, sub-node control
Database	One data specification control and storage	Multiple data standard control and storage
Offloading setup	Offload generation considering Sensing Node and Control Node	Create offload considering Sening Node, Collection Node, and Control Node

Table 2 compares the existing strategy with the proposed strategy. To facilitate understanding, we refer to the Leaf Node as the Sensing Node, the Edge Node as the Collection Node, and the Central Cloud Node as the Control Node. In the existing approach, the Sensing Node and Collection Node were mainly responsible for data collection, but going forward, they need to perform more functional roles. These functional roles include managing sub-nodes and performing tasks such as data analysis and integration. As the communication distance increases, the actual communication cost and time also increase, making it inefficient to process large-sized data at the Central Node. Additionally, performing computationally intensive tasks at the Collection Node may not be suitable. Therefore, in the future strategy, offloading should be determined based on factors such as data size, network status, and cost considerations. This development strategy enables collaboration among various service providers, thereby improving the quality of services available in the Smart IoT industry. Additionally, this development strategy reduces network and computing resource costs and provides a trusted environment to IoT data holders, making data sharing practices easier. In future work, the goal is to develop and implement a sample of the proposed framework. To achieve this, compatible available software and hardware products will be utilized since they follow similar protocols.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2023-2020-0-01462) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

This work was partly supported by the Technology development Program of MSS S3290113 and the ICT development R&D program of MSIT S3290113

References

1. Knierim, A.; Kernecker, M.; Erdle, K.; Kraus, T.; Borges, F.; Wurbs, A. Smart farming technology innovations—Insights and reflections from the German Smart-AKIS hub. NJAS-Wagening. J. Life Sci. 2019, 90, 100314.
2. Yang, X.; Shu, L.; Chen, J.; Ferrag, M.A.; Wu, J.; Nurellari, E.; Huang, K. A survey on smart agriculture: Development modes, technologies, and security and privacy challenges. IEEE/CAA J. Autom. Sin. 2020, 8, 273–302.
3. Finger, R.; Swinton, S.M.; El Benni, N.; Walter, A. Precision Farming at the Nexus of Agricultural Production and the Environment. Annu. Rev. Resour. Econ. 2019, 11, 313–335.
4. Mushi, G.E.; Serugendo, G.D.M.; Burgi, P.Y. Digital Technology and Services for Sustainable Agriculture in Tanzania: A Literature Review. Sustainability 2022, 14, 2415.
5. Islam, N.; Rashid, M.M.; Pasandideh, F.; Ray, B.; Moore, S.; Kadel, R. A review of applications and communication technologies for internet of things (Iot) and unmanned aerial vehicle (uav) based sustainable smart farming. Sustainability 2021, 13, 1821.
6. Chukkapalli, S.S.L.; Mittal, S.; Gupta, M.; Abdelsalam, M.; Joshi, A.; Sandhu, R.; Joshi, K. Ontologies and artificial intelligence systems for the cooperative smart farming ecosystem. IEEE Access 2020, 8, 164045–164064.
7. Awan, K.A.; Din, I.U.; Almogren, A.; Almajed, H. Agritrust—A trust management approach for smart agriculture in cloud-based internet of agriculture things. Sensors 2020, 20, 6174.
8. Chukkapalli, S.S.L.; Piplai, A.; Mittal, S.; Gupta, M.; Joshi, A. A Smart-Farming Ontology for Attribute Based Access Control. In Proceedings of the 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Baltimore, MD, USA, 25–27 May 2020; pp. 29–34.
9. Ametepe, A.F.X.; Ahouandjinou, S.A.R.M.; Ezin, E.C. Secure encryption by combining asymmetric and symmetric cryptographic method for data collection WSN in smart agriculture. In Proceedings of the 2019 IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, 14–17 October 2019; pp. 93–99.
10. Tan, W.X.; Zhao, C.J.; Wu, H.R.; Wang, X.P. An innovative encryption method for agriculture intelligent information system based on cloud computing platform. J. Softw.2014, 9, 1–10.

Alarm Signal based Machine status estimation using Deep Learning

Dimang Chhol, Kwan-Hee Yoo

Department of Computer Science, Chungbuk National University, South Korea
{dimangchhol, khyoo}@chungbuk.ac.kr

Abstract. Machine Health Stability provides an insight into understanding the overall status of the manufacturing process, providing a report when the anomaly production occurred and more importantly, giving the manufacturing company an upcoming estimation of failure. Many researchers have proposed various techniques for solving issues regarding machine stability including using machine learning techniques, deep learning techniques, and statistical process control. This research study proposed deep learning techniques for the estimation of machine health stability which consists of collected and analyzed data from a manufacturing company from January 2020 to August 2022 in South Korea. Followed by a data extraction procedure that uses essential features that work a major role in identifying the machine's health in the experiment. To understand which methods perform better than others we used various evaluation metrics in our test. In conclusion, we found that the GRU method performs better than other methods.

Keywords: Machine Health Stability, Alarm Signals, Deep Learning, Machine Learning, LSTM

1 Introduction

The software and hardware part of the system is critical to understanding the overall machine health stability. This research motivates us to consider additional features to the existing work of Bakhit, S. [1] and Pheng, T. et al. [2], and Borith, T. et al. [3], which aim to improve the current manufacturing system that we are currently working on by providing the manufacturer with the output result essential to improve their production. For instance, the machine's product quantity can give us the necessary information regarding how well the machine performs daily. Furthermore, by comparing it with the target product, which is input by the company, we could see and identify the overall process performance of the machine. Moreover, in making a product daily, the critical alarm might appear before the stop state of the machine occurs. In smart factories, there are many types of alarms using this alarm information could give insightful information for the future estimation of machine failure [4].

When there is any problem in the production process of a machine, PLC transferred alarm data to the system database [1]. Various research has pointed out the importance of using the alarm for predicting and preventing machine faults in smart manufacturing.

Alarms Signals are random events that should usually process in real-time, which could give manufacturing companies various advantages for instant predictive maintenance of the equipment [5] [6].

In our system environment, alarm type is a critical feature since before any upcoming STOP state of the machine appears, it is alert to the system. In making a product daily, the critical alarm might appear before the stop state of the machine occurs. In smart factories, there are many types of alarms using this alarm information could give insightful information for the future estimation of machine failure [4].

2 Related Studies

In the recent year, various studies have been conducted to deal with Machine Health Assessment including the prediction of upcoming failure by using Machine Learning and Deep Learning Techniques with different types of features.

Abas, N. et al. 2020 proposed five steps and methods with statistical control to distinguish the category of wind turbine faults [7]. Bruneo, D. et al. 2019 used Long Short Term memory networks for the estimation of the engine's Remaining useful life with the tunning of long short-term memory[8]. Lee, G.Y. et al. 2018 Presented a review of machine health management the author review machine health management which found essential tools used for Prognostics and Health Management (PHM) [9]. The authors define Prognostics as checking product health and estimating the remaining useful life of the machine [9].

We applied some techniques as mentioned by Khan, S. 2018 regarding the use of RNN, LSTM, and GRU for machine health monitoring status [10] and integrated with another researcher on features extraction technique from Pheng, T. et al. 2022 [2], Bahit, S. 2021 [1] and Borith, T. et al. 2020[3] with additional features.

3 Proposed Methods

We used two types of deep learning including LSTM and GRU. In the test we config LSTM and GRU with loss function “mse”, optimization “adam” and metrics “acc”. In terms of machine learning, we used LinearSVR with configuration random_state equal to 10 and tol equal to 0.001.

We define the framework and formula to label our Machine Health Stability as follows [1, 4]:

$$\begin{aligned} MHS = {} & \omega_1 \times (1 - Rate_{alarm\_type}) + \omega_2 \times (1 - Rate_{non\_active}) \\ & + \omega_3 \times (1 - Rate_{product\_making\_condition}) \\ & + \omega_4 \times (1 - Rate_{NG\_product}) + \omega_5 \times (1 - Probability_{failure}) \\ & + \omega_6 \times (1 - Rate_{defective\_product}) + \omega_7 \times (1 - Rate_{cp}) \\ & + \omega_8 \times (1 - Rate_{cpk}) + \omega_9 \times (1 - Rate_{mr}) \end{aligned}$$

We define machine health stability constant parameter with $\omega_1 = 0.1, \omega_2 = 0.1, \omega_3 = 0.2, \omega_4 = 0.2, \omega_5 = 0.2, \omega_6 = 0.15, \omega_7 = 0.05, \omega_8 = 0.05\ and\ \omega_9 = 0.05$ as the probability value of failure.

Regarding the MHS output, we specify the output range from 0 to 1. When the output is from 0.0 to 0.5 the Machine Stability is terrible, 0.5 to 0.65 the Machine Stability poor, 0.65 to 0.75 is average, 0.75 to 0.85 is good, and 0.85 to 1.00 the Machine Stability is excellent.

4 Experiment Result

In this experiment, we used two types of deep learning and compare them to machine learning so-called Linear SVM. In this test, we used one machine dataset from the line of the manufacturing car part industry in South Korea with 8061 rows. In the process, we split data into training sets consisting of 70% and testing 30%. We use various evaluation metrics including MAE, MSE, and RMSE to see which model performs better.

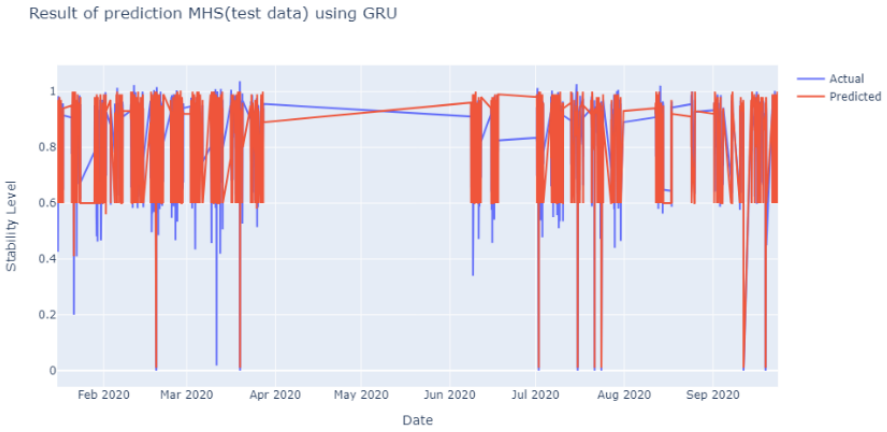


Fig 1. Result of Prediction of MHS with GRU

Table 1. Experiment Result in Comparison of the test dataset

Model	MAE	MSE	RMSE
GRU	0.07785	0.01098	0.10479
LSTM	0.08002	0.01138	0.10668
Linear SVM	0.09869	0.41131	0.64134

In Figure 1 and Table 1, we demonstrate the experiments result of our testing. We found that GRU performs better compared to other methods by using evaluation metrics MAE, MSE, and RMSE.

5 Conclusion and Future Work

In this experiment, We work on the implementation of a data analysis program for data preprocessing features by adding additional features to our study. We found that GRU performs better compared to other methods in the test such as LSTM and Machine Learning technique SVM Linear with MAE, MSE, and RMSE as evaluation metrics. In the future, we plan on extending additional features and Deep Learning methods to further improve the performance of the prediction accuracy and integration of the techniques to the current smart manufacturing industry that we are working with.

Acknowledgments. This research was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE)(No. P0022332, Digital data platform for material development)

References

1. Bakhit, S.: Prediction of Machine Health Stability in Smart Factory Systems using Machine Learning. Department Of Computer Science, vol. Master. Chungbuk National University (2021)
2. Pheng, T., Chuluunsaikhan, T., Ryu, G.-A., Kim, S.-H., Nasridinov, A., Yoo, K.-H.: Prediction of Process Quality Performance Using Statistical Analysis and Long Short-Term Memory. Applied Sciences 12, (2022)
3. Borith, T., Bakhit, S., Nasridinov, A., Yoo, K.-H.: Prediction of machine inactivation status using statistical feature extraction and machine learning. Applied Sciences 10, 7413 (2020)
4. Dimang, C.: Prediction of Machine Health Stability using Deep Learning. Computer Science, vol. Master of Engineering. Chungbuk National University, Cheongju, South Korea (2023)
5. Villalobos, K., Suykens, J., Illarramendi, A.: A flexible alarm prediction system for smart manufacturing scenarios following a forecaster–analyzer approach. Journal of Intelligent Manufacturing 32, 1323-1344 (2021)
6. Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., Vasilakos, A.V.: A Manufacturing Big Data Solution for Active Preventive Maintenance. IEEE Transactions on Industrial Informatics 13, 2039-2047 (2017)
7. Abas, N., Dilshad, S., Khalid, A., Saleem, M.S., Khan, N.: Power quality improvement using dynamic voltage restorer. IEEE Access 8, 164325-164339 (2020)
8. Bruneo, D., De Vita, F.: On the Use of LSTM Networks for Predictive Maintenance in Smart Industries. 2019 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 241-248 (2019)
9. Lee, G.-Y., Kim, M., Quan, Y.-J., Kim, M.-S., Kim, T.J.Y., Yoon, H.-S., Min, S., Kim, D.-H., Mun, J.-W., Oh, J.W., Choi, I.G., Kim, C.-S., Chu, W.-S., Yang, J., Bhandari, B., Lee, C.-M., Ihn, J.-B., Ahn, S.-H.: Machine health management in smart factory: A review. Journal of Mechanical Science and Technology 32, 987-1009 (2018)
10. Khan, S., Yairi, T.: A review on the application of deep learning in system health management. Mechanical Systems and Signal Processing 107, 241-265 (2018)

Application of Statistical Analysis for Recommending Ceramic Raw Material Blending

Ga-Ae Ryu[1*], Sung-hun Kim[2]

[1] Dept. of Materials Digitalization Center, Korea Institute of Ceramic Engineering & Technology, South Korea
[2] Dept. of Computer Science, Chungbuk National University, South Korea
garyu@kicet.re.kr, sidsid84@chungbuk.ac.kr
*Corresponding Author

Abstract. Proper raw material blending is crucial for achieving desired properties and performance in ceramic materials. This study focuses on utilizing statistical analysis techniques, specifically the FP-Growth method, to recommend optimal raw material blending in the ceramic manufacturing process. The FP-Growth algorithm is a widely used association rule mining technique that allows the identification of significant itemsets and association rules from large datasets. The research involved collecting data on various ceramic compositions and their corresponding properties. The Apriori method was applied to analyze the relationships between raw material compositions and the resulting ceramic properties. The application of the Apriori method in this study facilitated the identification of frequent itemsets and association rules, enabling the extraction of valuable knowledge regarding raw material blending for ceramic production. The findings contribute to the development of a data-driven approach for optimizing raw material combinations and enhancing the efficiency and effectiveness of ceramic manufacturing processes.

Keywords: Ceramic Materials, Recommendation, Statistical Analysis, Digital Transformation

1 Introduction

The composition derived from material blending is a crucial factor that significantly impacts the properties of ceramic materials. Various research studies are being conducted to discover new composition information using diverse simulation techniques.[1,2] However, as the number of variables to consider in simulations increases, the computational complexity rises, resulting in longer processing times.[3] To address this issue, research is also underway to utilize data-driven analysis in order to explore diverse composition information.[4-6]

In this study, association analysis is employed to recommend material blending for finding composition information based on property conditions. The recommendation approach utilizes the FP-Growth (Frequency Pattern-Growth) algorithm[7], which is a statistical analysis-based technique for discovering association rules through frequency patterns.

By recommending material blending using this approach, guidelines can be provided for obtaining material blending and composition information for specific property values. Moreover, this methodology can be applied not only to ceramic materials but also to various fields.

2 Proposed Method

The quality of ceramic materials is significantly influenced by the blending of various materials. In this study, we provide recommendations for ceramic material blending based on property characteristics through data-driven statistical analysis. The recommendation of material blending involves dividing the dataset based on property attributes and using the FP-Growth algorithm[7] to create and recommend sets of material blending based on the association rules for each property. The overall analytical framework is illustrated in Fig 1.

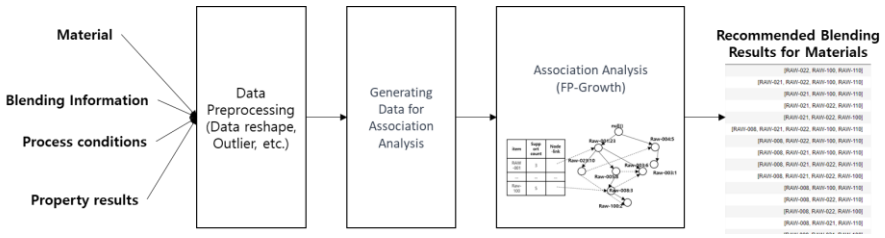


Fig. 1. Recommended Method for Material Blending through Association Analysis

To conduct data-driven statistical analysis for ceramic materials, we compiled data on composition information, process conditions, and property outcomes of ceramic material processes. Subsequently, we utilized these property outcomes, process conditions, and composition information to establish connections and perform analyses.

The property outcomes exhibited correlations among three different properties. Based on the data distribution, we defined target properties by dividing them into two intervals using the 50% value as a threshold. The defined data was then transformed into association analysis-ready data using one-hot encoding, considering the utilization status of each material.

The generated data groups were subjected to the FP-Growth algorithm[7] to calculate the minimum support. FP-Growth algorithm[7] is a method that discovers association rules by analyzing frequency patterns. It constructs an FP-tree for the entire dataset, where each item is incrementally added to the tree based on its frequency in the dataset. Using this FP-tree, patterns with support above the minimum threshold are extracted, providing recommendations for the data. In the context of ceramic composition, it is common to combine three or more materials based on their characteristics. Therefore, we extracted only those recommendations with itemsets consisting of three or more items. The results are presented in Fig 2.

	support	itemsets	length
72	0.780000	[RAW-022, RAW-100, RAW-110]	3
79	0.780000	[RAW-021, RAW-022, RAW-100, RAW-110]	4
78	0.780000	[RAW-021, RAW-100, RAW-110]	3
77	0.780000	[RAW-021, RAW-022, RAW-110]	3
76	0.780000	[RAW-021, RAW-022, RAW-100]	3
94	0.673333	[RAW-008, RAW-021, RAW-022, RAW-100, RAW-110]	5
93	0.673333	[RAW-008, RAW-022, RAW-100, RAW-110]	4
92	0.673333	[RAW-008, RAW-021, RAW-100, RAW-110]	4
91	0.673333	[RAW-008, RAW-021, RAW-022, RAW-110]	4
90	0.673333	[RAW-008, RAW-021, RAW-022, RAW-100]	4
89	0.673333	[RAW-008, RAW-100, RAW-110]	3
88	0.673333	[RAW-008, RAW-022, RAW-110]	3
87	0.673333	[RAW-008, RAW-022, RAW-100]	3
86	0.673333	[RAW-008, RAW-021, RAW-110]	3
85	0.673333	[RAW-008, RAW-021, RAW-100]	3
84	0.673333	[RAW-008, RAW-021, RAW-022]	3
13	0.433333	[RAW-017, RAW-100, RAW-110]	3
44	0.426667	[RAW-018, RAW-100, RAW-110]	3
60	0.413333	[RAW-018, RAW-021, RAW-022, RAW-110]	4
57	0.413333	[RAW-018, RAW-021, RAW-022, RAW-100]	4
55	0.413333	[RAW-018, RAW-022, RAW-100, RAW-110]	4
54	0.413333	[RAW-018, RAW-021, RAW-100, RAW-110]	4
51	0.413333	[RAW-018, RAW-021, RAW-022]	3

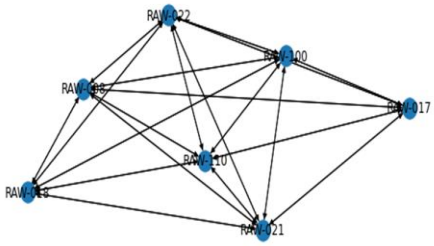
(a) Recommended Material Blending Results for Group 1

	support	itemsets	length
21	0.917647	[RAW-017, RAW-100, RAW-110]	3
25	0.823529	[RAW-003, RAW-017, RAW-110]	3
22	0.823529	[RAW-003, RAW-100, RAW-110]	3
41	0.823529	[RAW-003, RAW-017, RAW-100, RAW-110]	4
31	0.823529	[RAW-003, RAW-017, RAW-100]	3
34	0.682353	[RAW-003, RAW-020, RAW-100]	3
56	0.682353	[RAW-003, RAW-017, RAW-020, RAW-100, RAW-110]	5
51	0.682353	[RAW-003, RAW-017, RAW-020, RAW-100]	4
47	0.682353	[RAW-003, RAW-017, RAW-020, RAW-110]	4
44	0.682353	[RAW-003, RAW-020, RAW-100, RAW-110]	4
37	0.682353	[RAW-003, RAW-017, RAW-020]	3
42	0.682353	[RAW-017, RAW-020, RAW-100, RAW-110]	4
23	0.682353	[RAW-020, RAW-100, RAW-110]	3
28	0.682353	[RAW-003, RAW-020, RAW-110]	3
26	0.682353	[RAW-017, RAW-020, RAW-110]	3
32	0.682353	[RAW-017, RAW-020, RAW-100]	3
33	0.611765	[RAW-005, RAW-017, RAW-100]	3
50	0.611765	[RAW-003, RAW-005, RAW-020, RAW-110]	4
61	0.611765	[RAW-003, RAW-005, RAW-017, RAW-020, RAW-100]	5
60	0.611765	[RAW-003, RAW-005, RAW-017, RAW-020, RAW-110]	5
59	0.611765	[RAW-003, RAW-005, RAW-020, RAW-100, RAW-110]	5
58	0.611765	[RAW-005, RAW-017, RAW-020, RAW-100, RAW-110]	5

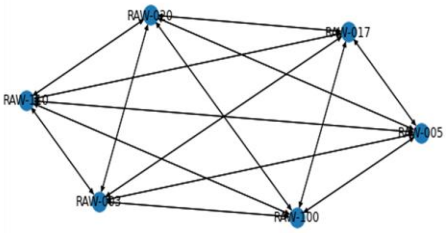
(b) Recommended Material Blending Results for Group 2

Fig. 2. Recommended Material Blending Results for Property Conditions

When visualizing the extracted results as a network graph, it is depicted as shown in Figure 3. In Group 1 data, the recommended material usage mainly includes raw-017, raw-021, raw-018, raw-008, raw-022, raw-100, raw-021, and raw-110. On the other hand, in Group 2 data, the recommended material usage includes raw-017, raw-005, raw-100, raw-003, raw-110, and raw-020. Additionally, the direction of the arrows indicates the association between the recommended materials.



(a) Recommended Material Blending Results for Group 1



(b) Recommended Material Blending Results for Group 2

Fig. 3. Visualization Results of Recommended Material Blending for Property Conditions

3 Conclusion and Future Works

In this study, we propose a method for recommending material blending for ceramic material properties through statistical analysis. The proposed method utilizes the FP-Growth algorithm for association analysis to recommend material blending for specific property conditions. This approach can provide guidelines for material blending based on property conditions in the context of ceramic material composition.

While this study focused only on considering the types of materials for recommending blending, future research will aim to consider factors such as material types, blending ratios, and elemental characteristics to provide recommendations for more comprehensive material blending.

Acknowledgment

This research was supported by the Ministry of Trade, Industry & Energy(MOTIE, Korea) under Virtual Engineering Platform Project, P0022336, 'Development of Virtual Engineering & Materials data Platform for Digital Transformation of Ceramics' and supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE)(No. P0022332, Digital data platform for material development)

References

1. Mishra, S., S. Maiti, and B. Rai, Computational property predictions of Ta–Nb–Hf–Zr high-entropy alloys. Scientific Reports, 2021. 11(1): p. 1-12.
2. Zhang, X., et al., Insight into the elastic and anisotropic properties of BiMg2MO6 (M= P, as and V) ceramics from the first-principles calculations. Ceramics International, 2019. 45(8): p. 11136-11140.
3. Troyer, M. and U.-J. Wiese, Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. Physical review letters, 2005. 94(17): p. 170201.
4. Himanen, L., et al., Data-driven materials science: status, challenges, and perspectives. Advanced Science, 2019. 6(21): p. 1900808.
5. Zhang, R.-Z., et al., Data-driven design of ecofriendly thermoelectric high-entropy sulfides. Inorganic chemistry, 2018. 57(20): p. 13027-13033.
6. Oses, C., C. Toher, and S. Curtarolo, Data-driven design of inorganic materials with the Automatic Flow Framework for Materials Discovery. MRS Bulletin, 2018. 43(9): p. 670-675.
7. Han, J., et al., Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data mining and knowledge discovery, 2004. 8: p. 53-87.

Deep Learning Approach for Property Intrusion Detection Using CCTV Video

Vungsovanreach Kong[1], Saravit Soeng[1], Munirot Thon[1], Tae-Kyung Kim[2], Wan-Sup Cho[1]

[1] Department of Big Data, Chungbuk National University, Cheongju, South Korea
[2] Department of Computer Information Technology, Incheon Jaeneung University, Incheon, South Korea
{kv.sovanreach, soengsaravit, munirot.thon, misoh049, wscho63}@gmail.com

Abstract. Automated systems for detecting crimes play a significant role in reducing crime rates and facilitating crime monitoring. One type of crime that can be detected using these systems is property intrusion, which occurs when an individual enters another person's property without permission. With the advancements in technology, it has become easier to detect property intrusion and other common crimes using machine learning or deep learning techniques. This paper presents an approach that utilizes deep learning technology (specifically, YOLOv5) to detect property intrusions in CCTV videos. The system has demonstrated high performance in detecting intrusions within a predefined property area. This research is significant because it showcases the potential of deep learning for property intrusion detection, and its results can be utilized to assist property owners in preventing such intrusions more easily.

Keywords: Intrusion Detection, Invasion Detection, Deep Learning, YOLOv5

1 Introduction

Intrusion onto a property, whether intentional or inadvertent, can constitute a criminal act. Intrusion manifests in various forms, ranging from home burglary to simply trespassing on a property without the owner's consent. As per Forbes, the United States witnesses approximately one million home intrusions annually, leading to a staggering $737 billion worth of damage and theft of valuables in 2021 [1]. A home burglary transpires every 15 seconds in the country, with the prime hours being between 10am and 3pm, a period when homeowners are usually absent [2]. Intruders on the premises can engage in a spectrum of illicit activities, such as burglary or armed robbery. Property owners typically employ conventional preventive measures like erecting fences, displaying signage, and installing surveillance cameras to deter such intrusions. In addition, modern technology has enabled the development of various automated systems to enhance prevention of property intrusion.

To bolster security and safety measures, certain researchers have employed various types of sensors to monitor and regulate door opening-closing activities, thereby alerting the property owner [3]. Another research team has designed a smart intrusion system utilizing repurposed edge devices linked to home Wi-Fi, which sends alerts to

the property owner upon detection of unauthorized door opening [4]. In this study, we propose a system that leverages security camera footage to automatically identify instances when an individual enters a predefined area. The system allows property owners to specify their property boundaries at the initiation of the system. Consequently, they can remotely oversee their property and detect any intrusion in real-time via deep learning technology applied to live security camera feeds from the property site.

2 Proposed Methodology

Figure 1 illustrates the comprehensive workflow of the proposed detection system, starting from CCTV video streaming and culminating in the detection of intrusions. Users can define their property boundaries in two specific shapes: a regular rectangle or a polygon. Following this, the YOLOv5 model is employed to identify human figures. The system then verifies whether the detected figure is situated within the predetermined property boundaries.

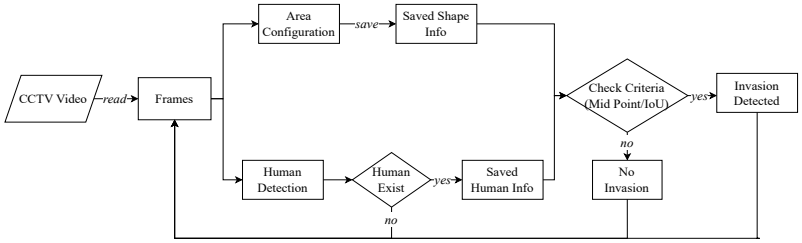


Fig. 1. The overall flowchart of the proposed intrusion detection system.

2.1 Area Configuration

The operation of the detection system commences with the user delineating their property area via the provided graphical user interface. The property boundaries can be outlined as a standard rectangle, or a polygon constituted by numerous points. In this stage, the system collects metadata pertaining to the area, including the coordinates of each point, which can be employed to replicate the area designated by the user. This metadata is instrumental in ascertaining potential property intrusions by cross-checking it during the post-processing step.

2.2 Human Detection with YOLOv5

The next phase is to detect human objects using the YOLOv5 algorithm. Since the person object is one of the 80 default classes supported by YOLOv5, we do not need to train the model again with the custom dataset [5]. We detect the human object using the default YOLOv5 model and return a collection of results comprising the detected human object's metadata. These metadata mainly represent the location of the detected human object with respect to the original image that streamed from the CCTV camera.

In the post-processing step, we may check the intrusion criterion using the results of area configuration and the YOLOv5 model.

2.3 Post-Processing Step

The system merges two results, one from the area-configuration step and the other from the detecting human objects step, to ascertain the occurrence of an intrusion. The system determines whether the human object detected by YOLOv5 is located within the predefined property area. If the object is located within the area, it is classified as an intrusion; otherwise, it is not considered as one. Presently, the system has two methods for identifying an intrusion: when the center point of the human object intersects the property area or when the Intersection over Union (IoU) of the two results is greater than 60% [6].

3 Results and Discussion

The video dataset collected via CCTV camera were used to evaluate the proposed system. The result shows that the system has an accurate detection using the proposed methodology. Fig. 2 shows the detection result produced by the system and the full demonstration video can be found at https://youtu.be/3Wd3Bhrna6Q.

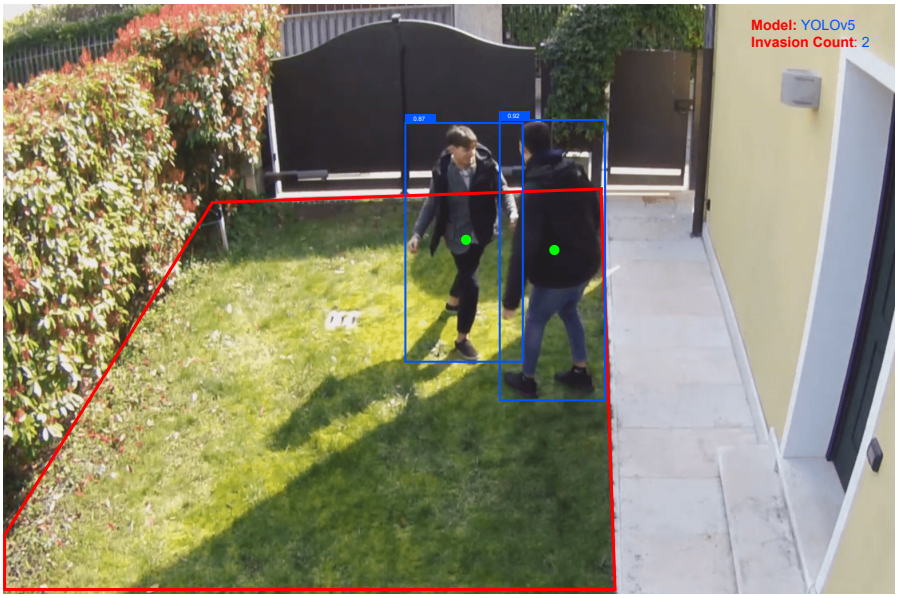


Fig. 2. The intrusion detection result of the proposed system applied to CCTV video.

The red line delineates the boundaries of the property, whereas the green dots symbolize the central point of each corresponding human figure. The blue rectangles

mark the instances of intrusion and also depict the probability calculated by YOLOv5. We posit that this system could be a highly effective tool for property owners in identifying any unauthorized entries onto their private premises. Moreover, Table 1 shows the evaluation of our proposed system for detecting area intrusion applied to sample CCTV camera videos.

Table 1. The evaluation result of proposed system on sample CCTV videos.

	# frames	Correct	Incorrect	Accuracy(%)	Error(%)
video #1	750	734	16	97.87	2.13
video #2	1,110	1069	41	96.31	3.69
video #3	450	422	28	93.78	6.22
Overall	**2,310**	**2,225**	**85**	**96.32**	**3.68**

4 Conclusion

This paper proposes a deep learning-based system for detecting the intrusion of private property using CCTV video. The system process starts by configuring the area as two shapes, whether a regular rectangle or a polygon made by multiple random points. Then, YOLOv5 is used as a backbone model to detect the human object present in the image frame. Then, we combine the two results from area configuration and human detection to determine the intrusion using the criteria described in the methodology section. The system showed good performance in detecting the intrusion, but there are still some limitations. It shows slightly less accuracy with nighttime detection due to the accuracy drop of YOLOv5 during nighttime. By the way, we believe that the system can be expanded in the future to give a real-time signal to the property owner whenever an intrusion is detected on their property.

References

1. Forbes, https://www.forbes.com/home-improvement/home-security/home-invasion-statistics
2. The Zebra, https://www.thezebra.com/resources/research/burglary-statistics/
3. Daramas, A., Pattarakitsophon, S., Eiumtrakul, K., Tantidham, T., Tamkittikhun, N.: HIVE: home automation system for intrusion detection. In: 2016 Fifth ICT International Student Project Conference (ICT-ISPC), pp. 101-104. IEEE, (2016)
4. Kwon, D., Song, J., Choi, C., Eun-Kyu, L.: A Home Intrusion Detection System using Recycled Edge Devices and Machine Learning Algorithm. International Journal of Advanced Computer Science and Applications 11, (2020)
5. Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Yifu, Z., Wong, C., Montes, D.: ultralytics/yolov5: v7. 0-YOLOv5 SOTA Realtime Instance Segmentation. Zenodo (2022)
6. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 658-666. (2019)

Performance Evaluation of Helmet and Traffic Light Detection Models: Time Consumption Analysis

Munirot Thon[1], Vungsovanreach Kong[1], Saravit Soeng[1], Tae-Kyung Kim[2], Wan-Sup Cho[1]

[1] Department of Big Data, Chungbuk National University, Cheongju, South Korea
[2] Department of Computer Information Technology, Incheon Jaeneung University, Incheon, South Korea

{munirot.thon, kvsovanreach, soengsaravit, misoh049, wscho63} @gmail.com

Abstract. With the increasing reliance on automated systems in areas such as traffic management and industrial safety, the development of efficient object detection models has become a crucial area of research. This study evaluated the time efficiency of four distinct models designed for the detection of helmets and traffic lights. A specified device tracked time in milliseconds for a range of processes, encompassing model loading and compiling, inference, comprehensive execution, and batch detection, using OpenVINO runtime. The findings indicated that each model consistently consumed approximately 3-4 seconds per image, irrespective of whether it was a complete frame or a cropped segment. Notably, the inference process was found to require less time compared to the loading and compiling stages, thereby indicating potential areas for optimization to enhance detection performance. Additional research is necessitated to investigate methods for optimizing the time consumption during the loading and compiling stages.

Keywords: Object Detection, Helmet Detection, Traffic Light Detection, Deep Learning, YOLO, Optimization

1 Introduction

Object detection is crucial for numerous practical applications, such as surveillance, autonomous driving, and robotics [1]. The effectiveness of these applications is largely determined by the accuracy and efficiency of the object detection models [2][3]. As such, it is essential to continuously improve and refine these models to ensure their effectiveness.

In this study, we evaluate the time efficiency of four different models for helmet and traffic light detection. These models were tested and evaluated on a device operating on a Central Processing Unit (CPU), without a high-performance Graphics Processing Unit (GPU), using OpenVINO runtime to assess their performance. The main goal of this study is to analyze the time consumption of these models in various execution tasks, including model loading and compiling, inference, full code execution, and batch detection.

This study outlines the methodology, presents the findings, and draws conclusions from the evaluation, offering insights into the utility and dependability of these models for production applications. The development of efficient models that can diminish time consumption and bolster performance is indispensable, particularly considering the current object detection architecture [4]. This architecture necessitates loading and compiling the model for every instance of image detection, which can potentially cause significant delays, as illustrated in Figure 1.

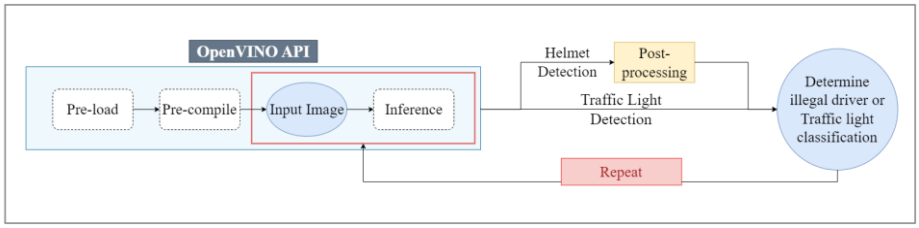


Fig. 1. Current Architecture for both Helmet Detection and Traffic Light Detection

2 Evaluation Methodology

In this study, we evaluated four different models using two categories of images: full-frame and cropped images. To ensure reliable and accurate outcomes, each model was subjected to ten evaluations. These evaluations were performed on a device with relatively modest computational power, reflecting more realistic conditions for many real-world applications [5]. Time consumption was gauged in milliseconds across four different categories: load and compile time, inference time, full code run time, and batch detection time. Load and compile time pertains to the duration spent on loading and compiling the model. Inference time refers to the duration needed for inference on a single image. Full code run time encompasses the complete end-to-end process of the code, including both pre-processing and post-processing. Batch detection time measures the duration required to detect a batch of ten images, following a single instance of loading and compiling the model.

These measurements were critical in accurately determining each model's performance and optimizing its results. By measuring and comparing load and compile time, inference time, full code run time, and batch detection time, the study offered a comprehensive and nuanced understanding of the time consumption associated with each model.

3 Evaluation Results

The evaluation performed on the four distinct models for helmet and traffic light detection offers a valuable understanding of their performance, specifically in terms of time efficiency. Table 1 presents an in-depth overview of the average time consumed by each model during various processes.

Table 1. The average time consumption of each model

Model	Load + Compile	Inference	Full Code Run	Batch Detection
Helmet Detection - Full Frame	3305.7	48.3	4339.7	4924.8
Helmet Detection - Crop Image	3413.1	41.6	4120.3	4448.3
Traffic Light Detection - Full Frame	3452.7	36.4	4022.6	4763.1
Traffic Light Detection - Crop Image	3343.5	41.2	3822.1	4327.3

The evaluation showed that image consumption takes 3-4 seconds per image for all models, whether it's a full-frame or a crop. Results were consistent across all models, implying that image type affects the detection process. A graph (Fig 2.) was plotted to compare the average time consumption of each model, indicating no significant difference in time consumption for various processes. Inference time was the quickest, averaging around 41 milliseconds.

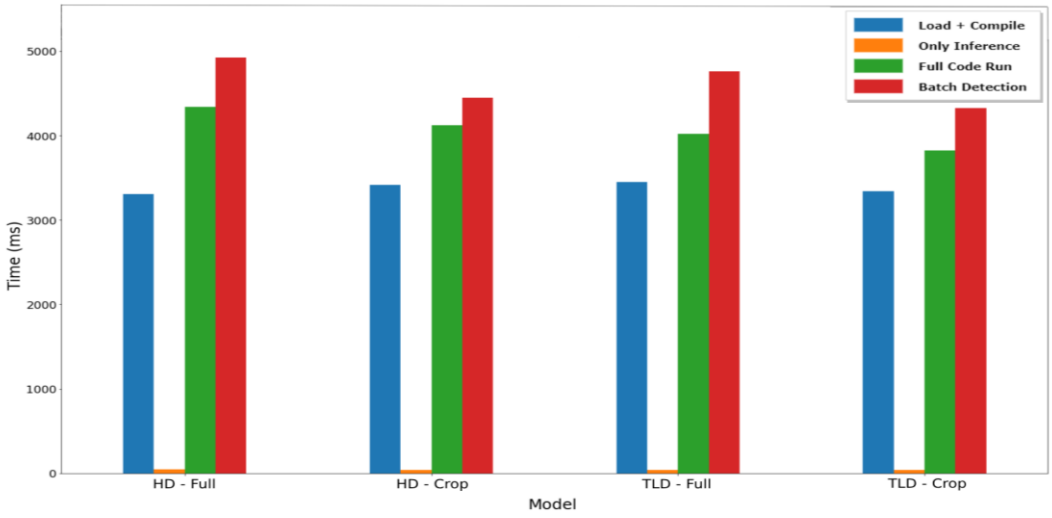


Fig. 2. Comparison of the Average Time Consumption for Different Model Execution Tasks

4 Discussion

One significant finding from the evaluation is that the inference time is relatively short compared to the loading and compiling time. This indicates that the current architecture of loading and compiling models for every image detection is not efficient and can cause significant delays in the object detection process.

To address this issue, several techniques were explored, such as optimizing the model from fp32 to int8 to reduce the model size [6]. Additionally, caching techniques were tested hoping to improve time performance, but it still takes an average of 3-4 seconds for single detection [7]. Redesigning the architecture to pre-

load and pre-compile models at program execution and storing them in memory could significantly reduce time consumption and improve efficiency (Fig 3).

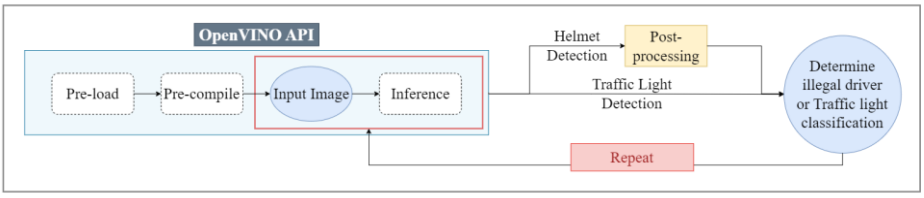


Fig. 3. Suggested Architecture Redesign

5 Conclusion

The performance analysis of four distinctive models for Helmet Detection and Traffic Light Detection was conducted. Key performance metrics indicate that these models require an average of 3-4 seconds for image processing. Interestingly, models operating on cropped images demonstrate superior efficiency compared to those analyzing full-frame images. It's important to note that the Helmet Detection model exhibits a longer processing time than the Traffic Light Detection model due to an additional post-processing phase. The evaluation further illustrates that the time dedicated to inference is considerably less than that required for the loading and compilation processes. Consequently, the current architecture, which mandates model loading and compilation for each image detection instance, lacks optimal efficiency. A reconsideration and modification of this process could result in substantial time savings and enhancement of the overall performance of the object detection mechanism.

References

1. Jun Deng, X.X., Weifeng Wang, Zhao Li, Hanwen Yao, Zhiqiang Wang: A review of research on object detection based on deep learning. In: Journal of Physics: Conference Series (2020)
2. Mingxing Tan, R.P., Quoc V. Le: EfficientDet: Scalable and Efficient Object Detection. In: IEEE (2020)
3. A. Tripathi, M. K. Gupta, C. Srivastava, P. Dixit, S. K. Pandey: Object Detection using YOLO: A Survey. In: 5th International Conference on Contemporary Computing and Informatics (2022)
4. What is OpenVINO? – The Ultimate Overview in 2023, https://viso.ai/computer-vision/intel-openvino-toolkit-overview/
5. A. Raza, M. H. Yousaf, S. A. Velastin: Human Fall Detection using YOLO: A Real-Time and AI-on-the-Edge Perspective. In: 12th International Conference on Pattern Recognition Systems (ICPRS) (2022)
6. What Is int8 Quantization and Why Is It Popular for Deep Neural Networks?, https://www.mathworks.com/company/newsletters/articles/what-is-int8-quantization-and-why-is-it-popular-for-deep-neural-networks.html
7. Optimize Inference — OpenVINO™ documentation, https://docs.openvino.ai/latest/openvino_docs_deployment_optimization_guide_dldt_optimization_guide.html

Unsupervise transfer learning using DANN in Color contact lenses

Ginam-Kim, Kwan Hee-Yoo

Dept. of Computer Science, Chungbuk National University
1, Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Republic of Korea
{kgn4192, khyoo}@chungbuk.ac.kr

Abstract. In this paper, we explored methods to enhance the generalization performance of models that detect defects occurring in the color contact lens molding process by applying Domain adaptation (DA), specifically Domain-adversarial neural networks (DANN). This process of generalizing a model trained in a source domain to another target domain is crucial when training and test data follow different distributions. Through this, we aimed to strengthen the ability to detect defects in various lens types. In the experiment, we set two different lens types, Lens A and Lens B, as D_S and D_T, respectively, and applied DANN for learning. However, we confirmed that domain adaptation from a specific lens type to another type remains a challenging issue. Therefore, in future research, we intend to develop new unsupervised learning-based DA methods applicable to the color contact lens dataset.

Keywords: Domain adaptation, DANN, Color contact lens, Unsupervise Learning

1 Introduction

Domain adaptation (DA) [1] is one of the major challenges in artificial intelligence, which refers to the process of generalizing a model trained in a source domain to a different target domain. This process is particularly crucial when the training and testing data follow different distributions.

Domain adversarial neural networks (DANN) [2] is one of the techniques designed to overcome such differences between domains. DANN is trained to minimize the distribution difference between the source domain dataset (D_S), which has labels for classification, and the target domain dataset (D_T), which does not have labels.

Recently, research on defect detection in the color contact lens molding process [3] has been conducted. A notable example includes defect detection of color contact lenses through a convolutional neural network (CNN) [4] proposed by Kim et al [5]. This research classified defects that could occur during the sandwiching process [6,7] of lens manufacturing into 'center defects', 'colorpoor', 'dotmissing', 'inkcut', 'line defects', and 'etc defects'. However, a limitation of this research is that training was only conducted for a specific lens type, meaning the data distributions of the trained lens type D_S and the untrained lens type D_T are different, leading to inaccurate predictions for D_T. Even

if labeling is carried out for numerous lens types D_T, it is challenging to secure data as defects do not occur frequently in the actual process. Therefore, this research applies DA using DANN to the color contact lens dataset. By doing so, we aim to effectively overcome the distribution differences between domains and improve the model's generalization performance.

2 Proposal method

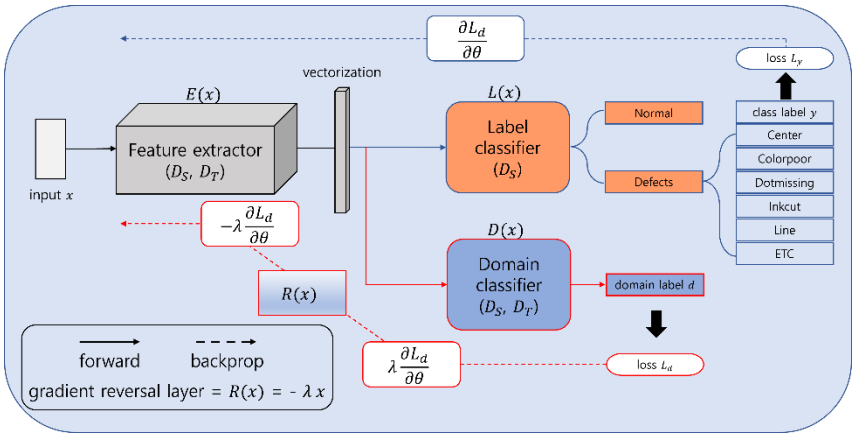


Fig. 1. DANN architecture

In this study, we utilized DANN, one of the DA techniques, for the transfer learning of color contact lenses. As shown in Figure 1, the pair (D_S, D_T) is processed through $E(x)$, composed of a CNN, to extract features. D_S, which has classification labels, proceeds to $L(x)$, a defect detection classifier, and undergoes learning via backpropagation. Conversely, after (D_S, D_T) passes through $D(x)$, it is classified according to its domain. During the backpropagation process, it is multiplied by a constant (λ) and converted in the negative direction. In other words, we aim to extract similar feature maps between the two domains in $E(x)$ so that D_T can also be effectively classified in $L(x)$.

The dataset used in this study, as depicted in Figure 2, consists of two different types of lenses, Lens A and Lens B, set as D_S and D_T, respectively. The quantity of data for each domain is as shown in Table 1.

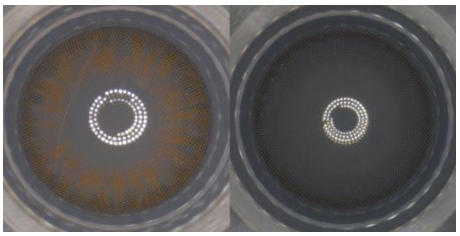


Fig. 2. (left) Lens Type A for $\boldsymbol{D_S}$ (right) Lens Type B for $\boldsymbol{D_T}$

Table 1. Number of datasets

Class	D_S (Lens A)	D_T (Lens B)
Normal	98	98
Center defects	97	97
Colorpoor	93	93
Dotmissing	95	95
Inkcut	99	99
Line defects	97	97
ETC	90	90

3 Result

The learning graph of DA through DANN proceeded as shown in Figure 3, and the evaluation metrics were as stated in Table 2, based on the highest accuracy of the class classifier in D_T during 500 epochs. The accuracy in $L(x)$, which classifies defects, was 0.9701 and 0.2422 for D_S and D_T, respectively, and the loss in $D(x)$, which classifies whether the two domains are the same, was 0.5308 and 0.5754, respectively.

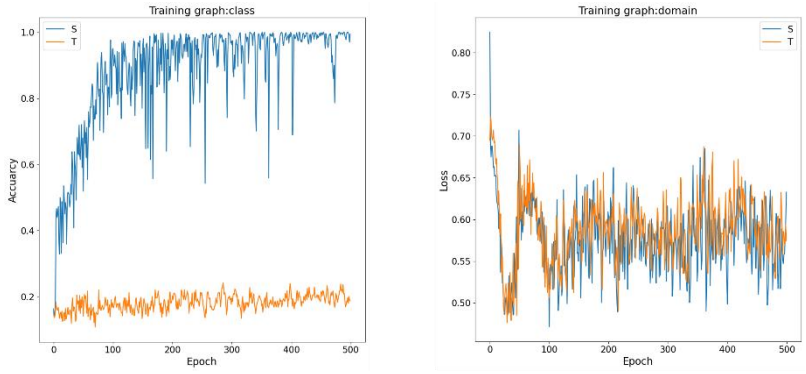


Fig. 3. Training graph by epochs. (left) class classifier accuracy (right) domain classifier loss

Table 2. Evaluation matrix

Classifier	Metric	D_S	D_T
Class classifier	**Accuracy**	0.9701	0.2422
	Loss	0.5049	17.1829
Domain classifier	**Loss**	0.5308	0.5754

4 Concluding Remarks and Future Work

In this study, we carried out DA learning for another type of color contact lens using DANN. However, we did not obtain significant results through this method. This

suggests that we failed to find a common distribution for classifying defects in D_S and D_T. Therefore, for future research, the authors aim to implement a new unsupervised-based DA method applicable to the color contact lens dataset.

Acknowledgments. This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2023-2020-0-01462) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

References

1. Csurka. G.: Domain adaptation for visual applications: A comprehensive survey. arXiv preprint arXiv:1702.05374 (2017)
2. Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M.: Domain-adversarial neural networks. arXiv preprint arXiv:1412.4446 (2014)
3. Fernandes, C., Pontes, A. J., Viana, J. C., Gaspar-Cunha, A.: Modeling and Optimization of the Injection-Molding Process: A Review. Advances in Polymer Technology. 37, 429-4494 (2018)
4. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 86, 2278-2324 (1998)
5. Kim, G.; Kim, S.; Joo, I. Yoo, K.H. Classification of Color Contact Lens Defects using Various CNN Models. The Journal of the Korea Contents Association, 22, 160-170 (2022)
6. Kim, M. A Process of Sandwich for Color Coating Contact Lenses. Publication No. KR100647133B1 (2005)
7. Kim, M. Coating method for cosmetic color contact lenses. Publication No. Publication No. WO2011019100A1 (2009)

Prediction of Alzheimer's Disease Progression using Multiview Dense Residual Attention and Stack Polynomial Attention

Ngoc-Huynh Ho[1] , Hyung-Jeong Yang[1], Jahae Kim[2]

[1] Department of AI Convergence, Chonnam National University, Gwangju, 61186, South Korea
[2] Department of AI Convergence and Nuclear Medicine, Chonnam National University and Hospital, Gwangju, 61469, South Korea

Abstract. Alzheimer's Disease (AD) is a chronic neurodegenerative disease that affects millions of people worldwide. In this study, a novel approach is proposed for the prediction of AD progression using Multiview Dense Residual Attention (MDRA) and Stack Polynomial Attention (SPA). MDRA is used to extract features from multiple views of brain MRI scans, while SPA is used to encode these features and predict the progression of AD. The proposed method was evaluated on a dataset of 676 MRI scans of Mild Cognitive Impairment (MCI) patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The results show that the proposed method reaches state-of-the-art performance in terms of identifying early MCI with late MCI stages. The proposed method has the potential to be used as a non-invasive tool for early prediction of AD progression, which can lead to better management and treatment of the disease.

Keywords: AD progression, Multiview Dense Residual Attention, Stack Polynomial Attention, MCI-to-AD conversion, eMCI vs lMCI.

1 Introduction

Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative disease that affects millions of people worldwide. It is the most common cause of dementia in the elderly, characterized by the accumulation of amyloid-beta plaques and tau protein tangles in the brain, leading to cognitive decline and memory loss. Early detection and prediction of AD progression are crucial for effective intervention and treatment. However, diagnosing AD at an early stage is challenging due to the absence of specific biomarkers and the overlapping symptoms with other neurodegenerative disorders.

Recent studies have shown promising results on the use of machine learning and deep learning in predicting the progression of Alzheimer's disease. In 2021, a study by Lee et al. [1] proposed a deep learning model that utilized multimodal neuroimaging data to improve diagnostic accuracy. A year later, an article by Sun et al. [2] introduced a machine learning model that incorporated cognitive and genetic markers to predict the disease progression. Another study by Wang et al. [3] used a convolutional neural network to extract features from MRI data, achieving high accuracy in predicting

Alzheimer's disease progression. In 2023, a research paper by Zhang et al. [4] compared the performance of several machine learning models and found that a hybrid approach combining deep learning and support vector machine achieved the best results. Lastly, a study by Kim et al. [5] leveraged the power of transfer learning to improve the prediction accuracy of Alzheimer's disease progression. These studies demonstrate the potential of machine learning and deep learning in predicting Alzheimer's disease progression and offer insights into the development of more accurate diagnostic tools and therapies. However, even though it is crucial to identify the early and late stages of MCI and predict the conversion rate to AD, there is a dearth of research on this topic.

In this paper, we propose a novel method for predicting AD progression, called Multiview Dense Residual Attention (MDRA) and Stack Polynomial Attention (SPA), on brain MRI data. This study seeks to identify patients in the early or late stages of MCI and predict their possibility of developing Alzheimer's disease. The proposed MDRA module integrates a Multiview Attention block with a Dense Residual block in order to extract features from multiview MRI scans, including axial, coronal, and sagittal dimensions. In addition, it is proposed that the SPA module encode low-level features for distinguishing early versus late MCI states and predicting the progression of MCI to AD. Our proposed method achieves state-of-the-art performance on an ADNI dataset of MCI patients that is publicly available.

2 Related Works

In recent years, various studies have proposed classification models for differentiating between early and late MCI stages in AD prediction. One approach involves using multimodal neuroimaging data and machine learning algorithms. Zhang et al. [6] proposed a joint classification model that utilizes multiple imaging modalities to identify both early and late MCI stages. Similarly, Zhang et al. [7] proposed a machine learning model that predicts the progression of MCI to AD using multimodal MRI data. Song et al. [8] proposed a deep learning model that combines positron emission tomography (PET) images and hybrid features to identify early-stage MCI. Another approach involves using MRI data and deep learning models. Kwon et al. [9] proposed a deep learning-based classification model that uses multi-modal MRI data to differentiate between early and late MCI stages. The model achieved high accuracy and outperformed traditional machine learning models.

In another field, there are few studies focusing on predicting the conversion from MCI to AD using survival analysis models. Suh et al. [10] proposed a model that combined cortical thickness and serum neurofilament light chain to predict conversion. Fang et al. [11] developed an interpretable machine learning algorithm based on multimodal neuroimaging and clinical data for conversion prediction. Chen et al. [12] proposed a survival model based on Random Survival Forest for the same purpose. Finally, Zuliani et al. [13] presented a machine learning-based 5-year prognostic model for conversion from MCI to AD using the ADNI cohort.

In summary, these models use different combinations of biomarkers, including multimodal neuroimaging data, CSF biomarkers, and genetic data, and employ different machine learning and deep learning algorithms. However, further studies are

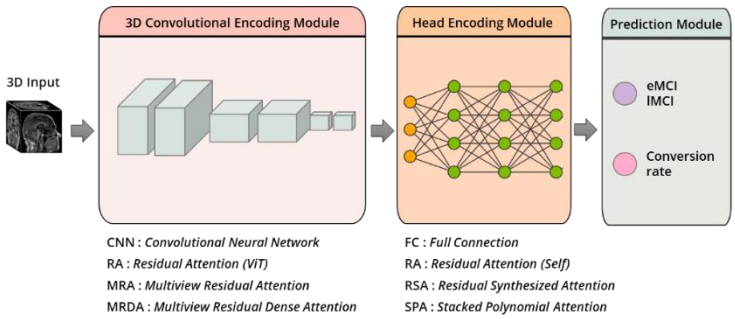


Figure 1 Overall architecture of prediction of AD progression

needed to validate the models, compare their performance across different populations and settings, and optimize their clinical utility.

3 Proposed Method

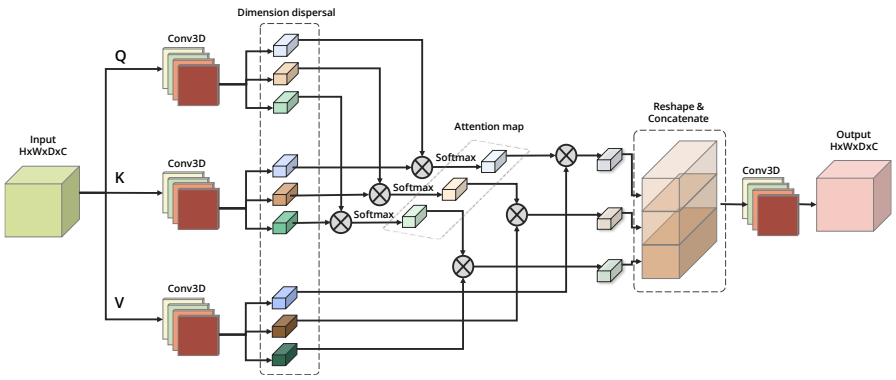


Figure 2 Architecture of Multiview Attention block.

The method proposed consists of three modules: 3D convolutional encoding, head encoding, and prediction modules. The 3D convolutional encoding module seeks to extract spatial features from 3D images that can be classified more readily. CNNs have the ability to acquire key characteristics on their own, eliminating the need for hyperparameters and manually designed filters. Next, the head encoding module uses the output of the 3D convolutional encoding module to generate the final representation for the prediction module, which yields the results of eMCI/lMCI classification and MCI-to-AD conversion prediction. The overall architecture of predicting progression of Alzheimer's disease is depicted in Figure 1.

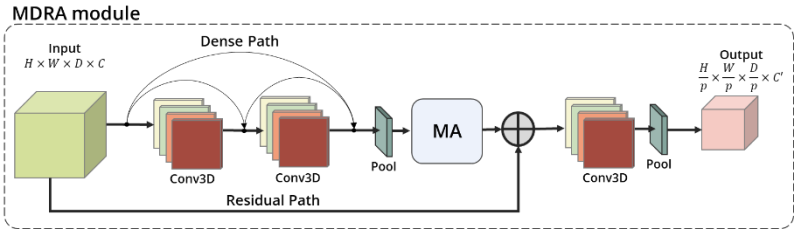


Figure 3 The architecture of Multiview Dense Residual Attention

3.1 3D Convolutional Encoding Module

In this paper, we propose a novel CNN, called Multiview Dense Residual Attention (MDRA), to explore the key features in three-view of brain MRI: axial, coronal, and sagittal. The MRDA module includes two blocks: Multiview Attention (MA) and Dense Residual (DR). Figure 2 displays the structure of the MA block. The MA block requires a 3D input with shape of $H \times W \times D \times C$, where H, W, D, C are the height, width, depth, and channel of the 3D input. The key technique we use in this block is self-attention [14]. Self-attention can capture global dependencies across different spatial locations within an image. By using self-attention, a model can selectively attend to important regions in an image and ignore the less relevant regions, which can improve the model's performance.

First, the 3D input is duplicated to the three features standing for Queries (Q), Keys (K), and Values (V). For each feature, we apply dimension dispersal on axial, coronal, and sagittal views to generate three 2D inputs. Then, we apply self-attention on each 2D input by assigning weights to different spatial positions of the input image based on their relevance to each other. This is done by comparing every position to every other position in the image to compute a similarity score. The similarity scores are then normalized using a Softmax function to produce a weight distribution for each position. These weight distributions are used to compute a weighted sum of the features at every position, with the weights serving as attention coefficients. This results in a new set of feature vectors that capture the most important information from the original image, taking into account the dependencies between different spatial positions. Finally, we reshape and concatenate all three feature vectors to the 3D shape. We apply convolutional operation to transform the features to the original shape of the 3D input.

Figure 3 presents the whole architecture of the Multiview Dense Residual Attention which includes MA block, Dense and Residual paths. Given input shape of $H \times W \times D \times C$, it is fed into a Dense path with includes multiple convolutional layers, where each layer is connected to every subsequent layer. It takes an input tensor and passes it through a series of convolutional layers, with each layer concatenating the feature maps from all preceding layers. This dense connectivity promotes feature reuse, enables the network to capture complex patterns by leveraging information from multiple layers, and improves the overall performance of the network. Then, a 3D max pooling layer is added to downsample the spatial dimensions of a three-dimensional

input tensor while retaining the most salient features. Next, the MA block is utilized to capture and integrate information from multiple perspectives of the 3D encoded features. We combine the output features with the original 3D input to complete the residual path, which is to address the problem of vanishing gradients and enable effective training of very deep neural networks. Finally, we add one convolutional layer and one pooling layer to capture more meaningful features and reduce the spatial information to the size of $\frac{H}{p} \times \frac{W}{p} \times \frac{D}{p} \times C'$, where p is the pooling size and C' is the filter size of the last convolutional layer.

3.2 Head Encoding Module

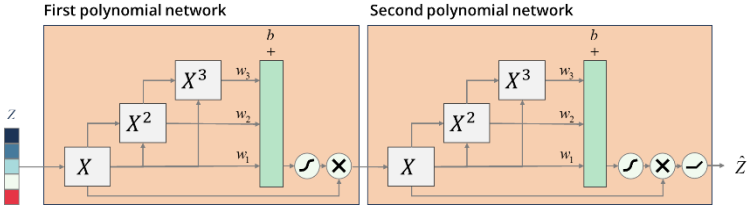


Figure 4 The architecture of the stacked polynomial attention

In recent years, there has been an increase in the utilization of attention mechanisms within the field of academia. These mechanisms serve the dual purpose of improving the performance and enhancing the explainability of deep learning techniques. In a previous study [15], researchers indicated that a stacked deep polynomial network (S-DPN) could enhance the performance of the extracted characteristics, showing potential for AD diagnosis through neuroimaging. Building upon these findings, we devised a new attention mechanism based on S-DPN, which we named the stacked polynomial attention (SPA) network. This novel approach leverages attended representation from constrained indeterminates.

Figure 4 illustrates the architecture of the SPA block. It stacks two polynomial attention networks, each containing polynomial units. A polynomial unit is a nonlinear transformation that involves combinations of polynomial terms, allowing the network to capture intricate patterns in the data. These polynomial layers help in capturing higher-order interactions and nonlinearities, enhancing the representation power of the network. In addition to the polynomial layers, the network also employs attention mechanisms, which are executed by the Sigmoid function. The attention mechanism enables the network to selectively focus on specific parts of the input data that are deemed more important or informative for the given task. By assigning attention weights to different elements of the input, the network can prioritize and attend to relevant features or patterns.

3.3 Prediction Module

The prediction module of the proposed network utilizes two activation functions for different tasks. A sigmoid function is employed for eMCI versus lMCI classification, producing a probability score between 0 and 1. A linear function is used for AD conversion risk prediction, generating a continuous value. This dual approach enables the network to handle binary classification and regression tasks effectively, providing accurate predictions for both the classification of MCI subtypes and the estimation of time-to-AD conversion.

4 Experiments

We used the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, which includes 249 eMCI and 427 lMCI patients at baseline diagnosis. To determine time-to-AD conversion, for uncensored patients, we assume that the conversion time is the time span between the baseline diagnosis and the first observation of AD. When considering the censored patients, the conversion time is calculated by adding the delaying time to their most recent visits.

For comparison, we implemented 3DCNN, 3D Residual Attention (3DRA), and Multiview Residual Attention (MRA) for the 3D convolutional encoding module. For the head encoding module, we implemented the fully connected (FC) layer, residual attention (RA) layer, and residual synthesized attention (RSA) layer to compare to the proposed SPA. The evaluation of predictive models for time-to-AD conversion involves various metrics, including Concordance Index (CI), Brier Score (BS), and Mean Absolute Error (MAE). These metrics assess the performance of the model in accurately predicting the time until an individual's conversion to Alzheimer's Disease (AD). For the classification task of distinguishing between eMCI and lMCI, several evaluation metrics are commonly employed. These include accuracy (Acc), precision (Pre), recall (Rec), average precision (AP), F_1 score, and Area Under the Curve (AUC).

Table 1 presents performance on prediction time-to-AD conversion. The results indicate that [MDRA + SPA] achieved a CI of 0.669, outperforming all other approaches in the table. This suggests that the proposed approach has a higher discriminative ability and better concordance between predicted and observed time-to-AD conversion. Moreover, [MDRA + SPA] demonstrated a lower BS of 0.192, indicating better calibration and overall accuracy in probability estimation. Additionally, it achieved the lowest MAE value of 413, indicating better accuracy in predicting the time-to-AD conversion.

Table 2 presents the performance on classification of eMCI and lMCI. The results show that the proposed method [MDRA + SPA] achieved an accuracy of 82.35%, which is consistent with other approaches in the table. In terms of Average Precision, [MDRA + SPA] obtained a high value of 95.07%, indicating its ability to rank positive samples higher in the classification. Moreover, [MDRA + SPA] demonstrated competitive Precision, Recall, and F1 scores, with values of 81.99%, 84.37%, and 81.96% respectively, which implies a good balance between correctly identifying positive cases (eMCI and lMCI) and minimizing false positives. Additionally, [MDRA

+ SPA] achieved an impressive AUC score of 90.12%, indicating its strong discriminative ability and effectiveness in distinguishing between eMCI and lMCI classes.

Table 1 performance on prediction time-to-AD conversion

Approach	CI	BS	MAE
3DCNN + FC	0.585	0.216	868
3DCNN + RA	0.646	0.203	451
3DRA + RA	0.613	0.208	767
MRD + RA	0.622	0.213	818
MDRA + RA	0.632	0.212	962
MRDA + RSA	0.53	0.194	911
MDRA + SPA	**0.669**	**0.192**	**413**

Table 2 Performance on classification of eMCI vs lMCI.

Approach	Acc	AP	Pre	Rec	F_1	AUC
3DCNN + FC	75	92.23	73.35	71.44	72.08	84.44
3DCNN + RA	69.85	88.69	69.74	61.93	61.61	81.12
3DRA + RA	81.62	94.87	**82.26**	**84.63**	81.39	88.93
MRD + RA	78.68	93.94	78.49	80.63	78.25	87.05
MDRA + RA	**82.35**	94.31	81.68	83.95	81.88	87.74
MRDA + RSA	**82.35**	95.03	81.43	83.54	81.79	89.98
MDRA + SPA	**82.35**	**95.07**	81.99	84.37	**81.96**	**90.12**

Overall, these superior performance metrics highlight the effectiveness and potential of the [MDRA + SPA] approach for accurately predicting AD conversion and classifying eMCI and lMCI, suggesting its potential as a valuable tool in the diagnosis and management of cognitive impairments.

5 Conclusion

In this article, we proposed a novel model for the prediction of Alzheimer's disease progression, which involved the classification of early and late MCI stages and the prediction of the conversion from MCI to AD. Our proposed model utilized a Multiview Dense Residual Attention module for extracting features from three dimensions of the MRI brain, and a Stack Polynomial Attention module for encoding low-level representation for final prediction. The experimental results demonstrated that our model achieved good performance in terms of accuracy for classification and C-index for conversion prediction. Despite the promising results, we acknowledge that there is still room for improvement in the proposed model, particularly in terms of predicting AD progression. In future works, we aim to incorporate more advanced techniques, such as graph-based learning, to capture the complex relationships between brain regions and to further enhance the performance of our model. Additionally, we plan to explore the use of more diverse datasets and imaging modalities to validate the effectiveness of our proposed model and to improve its generalization capabilities.

Ultimately, we hope that our proposed model can contribute to the early diagnosis and prevention of Alzheimer's disease, leading to better patient outcomes and quality of life.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT). (RS-2023-00208397)
This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629) grant funded by the Korea government(MSIT)

References

1. Lee, H., Lee, J. H., Kim, H. J., Lee, K., & Kim, Y. K.: Deep learning-based multimodal neuroimaging analysis for prediction of Alzheimer's disease. Journal of Alzheimer's Disease, 83(1), 1-11. (2021)
2. Sun, J., Zhou, Y., Wang, P., & Zhang, D.: A machine learning approach for predicting Alzheimer's disease progression by integrating cognitive and genetic markers. Frontiers in Neuroscience, 16, 801. (2022)
3. Wang, S., Li, X., Li, W., & Xu, J.: A convolutional neural network for predicting Alzheimer's disease progression using MRI. Frontiers in Neuroscience, 16, 823. (2022)
4. Zhang, L., Li, Y., Zhang, X., & Xu, X.: A comparison of machine learning models for predicting Alzheimer's disease progression. Journal of Alzheimer's Disease, 86(1), 75-85. (2023)
5. Kim, Y., Lee, J. H., & Kim, Y. K.: Transfer learning-based prediction of Alzheimer's disease progression using multimodal neuroimaging data. Journal of Alzheimer's Disease, 86(2), 569-579. (2023)
6. Zhang, Y., Li, X., Wang, Y., Li, H., Guo, L., & Wang, Y.: Joint classification of early and late mild cognitive impairment using multimodal neuroimaging and machine learning. Frontiers in Neuroscience, 15, 658135. (2021)
7. Song, Y., Lu, H., Sun, X., Shen, D., & Li, J.: Deep learning model for early stage mild cognitive impairment identification using PET images and hybrid features. Frontiers in Neuroscience, 15, 690789. (2021)
8. Zhang, Y., Li, X., Wang, Y., Li, H., Guo, L., & Wang, Y.: Predicting the progression of mild cognitive impairment to Alzheimer's disease using multimodal MRI data and machine learning. Frontiers in Neuroscience, 16, 834280. (2022)
9. Kwon, H., Lee, J., & Lee, S.: Deep learning-based classification of early and late mild cognitive impairment stages using multi-modal MRI data. Pattern Recognition, 128, 109476. (2023)
10. Suh J, Shin DY, Kim J, et al.: Combining cortical thickness and serum neurofilament light chain predicts conversion from mild cognitive impairment to Alzheimer's disease. Alzheimers Res Ther., 13(1), 56. (2021)
11. Fang X, Yang Z, Chen K, et al.: Predicting conversion from MCI to Alzheimer's disease using an interpretable machine learning algorithm based on multimodal neuroimaging and clinical data. Front Aging Neurosci., 13, 741719. (2021)
12. Chen J, Xie F, Wu Y, Zhang W, Zhu Y, Fang Y. Predicting conversion from mild cognitive impairment to Alzheimer's disease using a survival model based on Random Survival Forest. Front Aging Neurosci., 13, 739417. (2021)

13. Zuliani G, Tardelli M, Marini M, et al. A 5-year prognostic model for conversion from mild cognitive impairment to Alzheimer's disease using the Alzheimer's Disease Neuroimaging Initiative cohort: a machine learning approach. Neurobiol Aging., 112, 183-191. (2022)
14. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I.: Attention is all you need. In Advances in neural information processing systems, 5998-6008. (2017)
15. Zheng, X., Shi, J., Li, Y., Liu, X. & Zhang, Q. Multi-modality stacked deep polynomial network based feature learning for alzheimer's disease diagnosis. In 2016 IEEE 13th international symposium on biomedical imaging (ISBI), 851–854. (2016)

Classification of tomato leaf diseases using various CNN models

JuHan-Song, Sunghoon Kim, Kwan Hee-Yoo

Dept. of Computer Science, Chungbuk National University
1, Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Republic of Korea
{2017038048, sidsid84, khyoo}@chungbuk.ac.kr

Abstract. Plant diseases have a profound impact on global food safety, farmers' livelihoods and plant yields. In order to solve these problems, research on automated plant disease recognition systems through deep learning is being actively conducted. In this paper, we compared the performance of each model using various CNN models to select which model is the most suitable for classifying tomato leaf diseases. To this end, disease classification accuracy was compared and analyzed by using CNN models of ResNet152, DenseNet169, U-net, and GoogleNet V4 on PlantVilage preprocessed tomato data images. The accuracy of the above model was 93.85%, 96.49%, 61.48%, and 88.76%, respectively, and DenseNet169 had the highest accuracy.

Keywords: Plant Disease, DeepLearning, ResNet, DenseNet, GoogLeNet, U-Net

1 Introduction

Every year, an average of 26% of global crop production is lost due to pre-harvest plant diseases. [1] Since it is difficult for most farmers to immediately identify diseases due to limited resources, a method for farmers to rapidly identify plant diseases is essential. [2] One such method is computer vision technology, which has seen significant advancements recently with the utilization of deep learning techniques. Among these techniques, Convolutional Neural Network (CNN) [3] has demonstrated superior performance in plant disease classification compared to humans [4-10].

However, most of the currently presented plant disease classification studies using CNNs are cases in which accuracy is improved by improving specific models, and various CNN models are not selected for comparative analysis. Therefore, there is a lack of research on which CNN model is best for a particular plant leaf. Therefore, in this paper, we compared ResNet-152, DenseNet-169, U-Net, and GoogLeNet V4 models to the same dataset and hyperparameters to study the most suitable CNN model for tomato leaf disease classification.

2 Related Works

In this study, we want to find a model suitable for tomato leaf disease classification, and related studies include a model comparison study suitable for tomato leaf disease classification and related research to improve the accuracy of plant disease detection. As a study on improving the accuracy of plant disease detection, Penghui et al. There is a study of [3], and this study applied Data Augmentation, Channel Orthogonal Constraint, Species Classification Task, and three loss functions using the ResNet-50 model to improve the accuracy of plant disease detection in the real environment, and the accuracy was 41.81%. improved to 71.03%. In addition, Lee et al [11] obtained 98% accuracy using the VGG-13 model after using various preprocessing techniques. However, both studies lack experiments in various models. As a comparative study of suitable models for tomato leaf disease, Suryawati et al [12] There is a study. This study obtained 91.52%, 95.24%, and 89.68% accuracy, respectively, using Alexnet, VGGNet, and GoogleNet to find a model suitable for tomato leaf disease classification. However, in this paper, we want to compare models using more diverse techniques.

3 Proposed Method

To find a model suitable for tomato leaf disease classification, we used segmented tomato images from PlantVillage[13], an open dataset, and selected ResNet-152, DenseNet-169, U-Net, and GoogLeNet V4 models as comparative model groups.

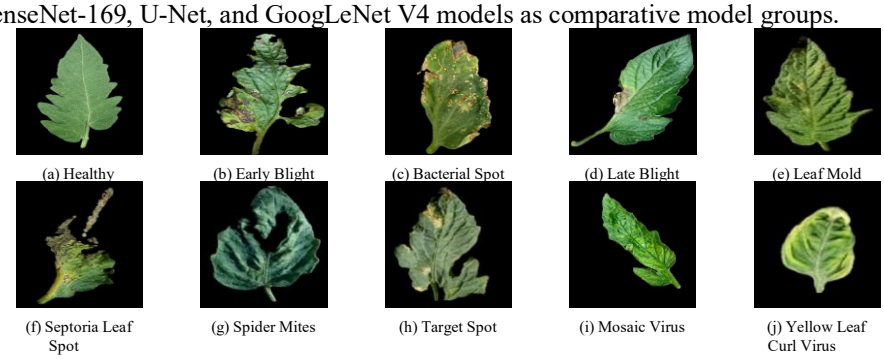


Fig 1. Segmented images of PlantVillage Dataset

Fig 1. is an example of each class of Segmented images of PlantVillage Dataset, and the background is removed by separating the background and the object (Leaf). Fig 1. - (a) is a healthy tomato leaf. (b) is characterized by the appearance of small black or brown spots, and (c) starts from the lower leaves and turns brown, with black dots appearing around the leaf edges. (d) shows brown blisters, and (e) is characterized by a thick layer of green or yellowish-brown fungus on the top of the leaf. (f) is characterized by the appearance of small brown dots followed by white spores in brown centers. In (g), many holes appear, and in (h), round black dots appear, followed by white tan dots. (i) is characterized by a bright green rose pattern. (j) is characterized by turning yellow and having ring-shaped leaf veins. The number of each class is shown in Table 1.

Table 1. Number of each class

Class	Number of data
Bacterial Spot	2127
Early Blight	1000
Healthy	1591
Late Blight	1908
Leaf Mold	953
Septoria Leaf Spot	1771
Spider Mites	1676
Target Spot	1404
Mosaic Virus	373
Yellow Leaf Curl Virus	5357

The compared models and major modules are shown in Table 2 below.

Table 2. Target models and core concept

Model	***core*** **concept**
ResNet-152	Skip Connections+ Add
DenseNet 169	Dense Conected +Concatenate
U-Net	Upsampling +Concatenate
GoogleNet-V4	Inception Module, Reduction Module

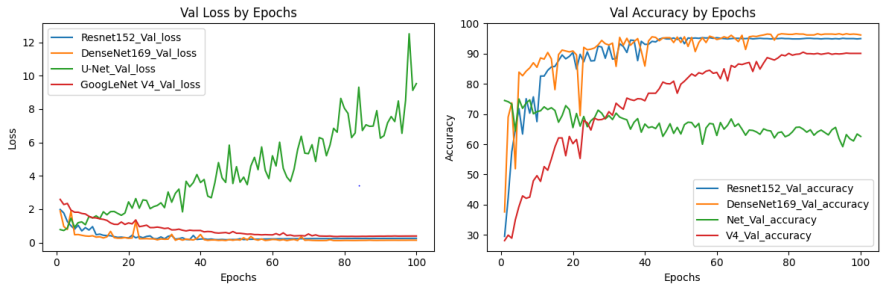
ResNet [14] introduced Skip Connections as a solution to address the issue of vanishing gradients, where the gradients for layers closer to the input tend to approach zero as the network becomes deeper [15]. By skipping the convolutional layers and directly adding the input to the output, ResNet ensures that the gradients are preserved with a minimum value of 1, thereby improving the problem of gradient vanishing. DenseNet [15] connects the feature maps of all layers within a Dense Block. It has a structure that connects the feature maps of the previous layer to the feature maps of all subsequent layers. Unlike ResNet, concatenation is performed instead of addition to improve the gradient loss problem. This structure has a problem that the amount of computation increases as the feature map continues to increase. This is reduced by reducing the feature map through 1x1 convolution. In U-Net [16], the encoder and decoder form a U-shaped structure to extract low-resolution abstract features of the image, and concatenate the features extracted from the encoder with features upsampled by the decoder to obtain not only abstract features but also spatial locations. can be extracted. U-Net is mainly used for object segmentation tasks, but we added an MLP layer for classification to see if it is suitable for classification models as well. The GoogLeNet V4[17] architecture consists of three major modules: InceptionA, InceptionB, and InceptionC. The InceptionA module employs a parallel structure, utilizing both 1x1 and 3x3 convolutions simultaneously. The InceptionB module also adopts a parallel structure, combining 1x1, 7x1, and 1x7 convolutions. The InceptionC module employs a parallel structure as well, using 3x1 and 1x3 convolutions alongside a 1x1 convolution module. Due to the various kernel sizes of each module, it is possible to learn various feature maps by concatenating the extracted feature maps.

4 Experiment Result

Table 3. Hyperparameter of Models

Hyperparameter	Value
Epoch	100
Optimizer	SGD
Batch Size	5
Image Channel	RGB
Image Size	256*256
Train/Val/Test ratio	0.8/0.1/0.1

Table 3 shows the hyperparameter settings of the model used in the experiment. The accuracy and loss of each model for each epoch is shown in fig 6, and unlike other models, U-Net's loss value increases and accuracy decreases as epochs progress.



(a) Val Loss for each Model (b) Val Accuracy for each Model

Fig 2. Val Loss and Accuracy by Epochs for each Model

Table 4. Comparison of Final Test Results of Each Model

Model	Precision	Recall	F1 Score
ResNet-152	0.9385	0.9386	0.9384
DenseNet-169	**0.9649**	**0.9649**	**0.9646**
U-Net	0.6148	0.6416	0.6213
GoogLeNet V4	0.8876	0.8898	0.8879

Table 4. shows good results for Testdata. Since each of the experimental models has a small gap in Precision and Recall, the F1 Score, which is the harmonic average, also shows a similar value. Among the experimental models, DenseNet-169 showed the best performance and showed an F1 score of 0.9646. Unlike the rest of the models, the performance of U-Net shows a difference of more than 0.25 based on the F1 score.

5 Concluding Remarks and Future Work

In this paper, ResNet152, DenseNet169, GoogleNet V4, and U-Net were compared and analyzed to find a CNN model suitable for tomato leaf disease classification. In the case of U-Net, it is a model used for image segmentation, and it can be inferred that the performance of the model deteriorated because it did not use the spatial information of the image well while performing classification by adding an MLP layer at the end. In the case of DenseNet169, it can be inferred that the characteristics of the Dense block structure were more effective in classifying tomato leaf diseases than other models. Also, although not mentioned, it was seen that the accuracy of Early blight and Leaf Mold was much lower than that of other diseases. For this reason, we plan to study by applying a method to better classify these two features to DensNet.

Acknowledgments. This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2023-2020-0-01462) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

References

1. Oerke, E.-C., 2006. Crop losses to pests. J. Agric. Sci. 144 (1), 31−43.
2. Gui, Penghui, et al. "Towards automatic field plant disease recognition." Computers and Electronics in Agriculture 191 (2021): 106523, 2-3.
3. Yann LeCun Leon Bottou Yoshua Bengio and Patrick Haffner ″Gradient-Based Learning Applied to Document Recognition″ Proceedings of the IEEE, Vol.86, pp.2278-2324, 1998
4. da Silva Abade, Andre, Ana Paula GS de Almeida, and Flavio de Barros Vidal. "Plant Diseases Recognition from Digital Images using Multichannel Convolutional Neural Networks." VISIGRAPP (5: VISAPP). 2019.
5. Brahimi, Mohammed, et al. "Deep learning for plant diseases: detection and saliency map visualisation." Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent (2018): 93-117.
6. Chen, Junde, et al. "Using deep transfer learning for image-based plant disease identification." Computers and Electronics in Agriculture 173 (2020): 105393.
7. Li, Yang, Jing Nie, and Xuewei Chao. "Do we really need deep CNN for plant diseases identification?" Computers and Electronics in Agriculture 178 (2020): 105803.
8. Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." Frontiers in plant science 7 (2016): 1419.
9. Nazki, Haseeb, et al. "Unsupervised image translation using adversarial networks for improved plant disease recognition." Computers and Electronics in Agriculture 168 (2020): 105117.
10. Too, Edna Chebet, et al. "A comparative study of fine-tuning deep learning models for plant disease identification." Computers and Electronics in Agriculture 161 (2019): 272-279.
11. Cheolwon Lee, and Hyungtae Ahn. "A Study on Image-based Greenhouse Tomato Leaf Disease Detection Technique Using Deep Learning." Proceedings of the Korean Society of Communications and Communications Conference (2020): 117-118.

12. Suryawati, Endang, et al. "Deep structured convolutional neural network for tomato diseases detection." 2018 international conference on advanced computer science and information systems (ICACSIS). IEEE, 2018.
13. Hughes, David, and Marcel Salathé. "An open access repository of images on plant health to enable the development of mobile disease diagnostics." arXiv preprint arXiv:1511.08060 (2015).
14. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, ″Deep Residual Learning for Image Recognition″CVPR, pp.770-778, 2016
15. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, ″Densely Connected Convolutional Networks ″CVPR, pp.4700-4708 ,2017
16. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.
17. Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi, ″Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning″ AAAI'17 pp.4278-4284

Cross Attention-based Multimodal Fusion for Depression Prediction

Duy-Phuong Dao[1], Hyung-Jeong Yang[1,1], Eun-Chae Lim[1], Soo-Hyung Kim[1]

[1] Department of Artificial Intelligence Convergence, Chonnam National University, 77 Yongbong-dong (61186), Bukgu, Gwangju, South Korea
phuongdd.1997@gmail.com, {hjyang, 218354, shkim}@jnu.ac.kr

Abstract. Depression is a frequent and severe medical condition that has an impact on how you feel, think, and behave. Early detection and intervention can potentially reduce the escalation of the disorder. In this study, we proposed multimodal deep learning model to predict depression status using audio and visual characteristics from the self-recorded video. Our proposed model employs Temporal Convolutional Network to capture temporal features from sequential input data in each modality. Afterwards, the cross-attention module is adopted to learn relationships across multiple extracted features from the corresponding modalities. The experimental results show that our method achieves the highest score in terms of weighted average F1-score, weighted average precision, and weighted average recall metrics with 0.6860, 0.6990, and 0.6840, respectively.

Keywords: Depression Detection, Deep Learning, Multimodal Fusion, Attention Mechanism, Temporal Feature.

1 Introduction

Millions of people worldwide are impacted by depression, which is a prevalent mental health disorder [1]. Recent advances in deep learning techniques have opened up new avenues for accurate and reliable depression detection. Deep learning algorithms can process large amounts of data, including speech, facial expression, and physiological signal, to identify patterns and markers associated with depression.

One of the major advantages of deep learning for depression detection is its ability to learn from complex and heterogeneous data sources. For example, deep learning models can analyze speech patterns to identify changes in tone, pitch, and speed that may be indicative of depression. Similarly, facial recognition algorithms can detect subtle changes in facial expressions that are associated with depression. Therefore, multiple modality learning can capture complementary information from different modalities, which can improve learning accuracy and robustness.

[1] Corresponding author: Hyung-Jeong Yang (email: hjyang@jnu.ac.kr)

On another hand, the type of speech and facial expression are commonly sequential data. Recent deep learning methods employ recurrent neural networks (RNNs), such as gated recurrent unit (GRU) [2] or long-short term memory (LSTM) [3], to capture

temporal features. However, these methods have limitations in handling long variable length sequences of input data. The problem of vanishing and exploding gradients hampers the learning of long data sequences when using the RNNs. Temporal convolutional network (TCN) [4] is used to overcome this issue. The TCN uses dilated causal convolutions to incorporate information from past and future steps. Dilated convolutions involve skipping some time steps during the convolutional operation, allowing the network to capture long-term dependencies while avoiding the vanishing gradient problem that can occur in RNNs.

Besides, multimodal fusion is used to combine information from multiple modalities. Traditional fusion techniques are based on concatenation or summarization. Recently, inspired by advance of deep learning, the attention-based multimodal fusion methods have a great concert from many researchers. The attention mechanism [5-7] can give more importance to the most informative modalities and less importance to the less informative ones, resulting in a more accurate and robust representation.

In this study, we propose a multimodal deep learning model to detect depression status based on vlog data that data contains multiple modalities such as audio and visual characteristic. Our proposed method consists of three main modules: (1) temporal feature learning, (2) multimodal fusion, and (3) depression prediction. The temporal feature learning module is used to extract modality-specific temporal features from each input modality. The purpose of multimodal fusion module is to capture complementary information across different modalities. The depression prediction module uses several fully connected layers to predict the depression status.

The structure of this paper is organized as follows. Section 2 briefly depicts related works in depression detection using multimodality. The proposed network is presented in detail in Section 3. In section 4, the dataset is described, and the comprehensive results are shown with comparison to the other competing methods. Finally, the conclusion is presented in Section 5.

2 Related Work

The creation of objective AI tools for automatic depression diagnosis has been actively researched in recent years using physiological markers such self-reported symptoms, facial expressions, audio characteristics, and text data from conversation. Conventional methods [8-9] select one of these physiological data as unimodal input to analyze the depression status. However, depression is a complex disorder that can manifest in many different ways. Using only one type of input may not provide enough information to make an accurate diagnosis. For example, relying solely on audio characteristics may not capture the full extent of a patient's condition. Therefore, the use of multimodal data can help capture the complexity of depression by integrating information from multiple sources.

A.H. Jo et al [10] employed audio and text information from the interview to diagnose depression. The authors extracted the audio features using bidirectional long short-term memory (Bi-LSTM) and CNN-based transfer learning models. Meanwhile, word encoding was adopted to tokenize the text. Finally, the outputs of audio and text models were integrated via the SoftMax function to produce results.

L. Zhou et al [11] extracted contextual information of audio and visual sequential data using CAM-BiLSTM module. Next, the authors proposed the local information fusion module and the global information interaction module to capture a fine-grained relationship between multimodal features and global relationship from local and global views. These features were then concatenated and passed through the fully connected layer (FCL) to classify the depressed or normal person.

C. Lin et al [12] adopted CNN and pre-trained BERT models to capture image and text features posted by users on social media. Then, the extracted features were concatenated and fed into deep visual-textual multimodal learning module to classify the normal or depressed user.

3 Proposed Method

In this section, we describe the details of our proposed method, as shown in Figure 1. Our proposed method comprises three main components: (1) temporal feature learning, (2) multimodal representation fusion, and (3) depression prediction. The audio and visual extracted features are two input modalities of the proposed method. First, the TCN-based temporal feature learning module process sequential data using one-dimensional causal and dilated convolutional layers. Thanks to the use of causal convolution, the output at time t is only convolved with the input elements that occurred before t. Meanwhile, the purpose of dilation is to increase the receptive field without increasing the number of parameters. The causal and dilated convolution are visually shown in Figure 2.

After extracting the temporal feature representation from each modality, we fed them into the cross-attention module. Different to previous cross-attention modules, which commonly calculate attention using query, key and value vectors for each modality, our proposed module does not require query weight for each modality. We set key of audio modality as query of visual modality and vice versa (Equation (3)). Afterward, we apply the row-wise and column-wise SoftMax function multiplied with the corresponding value to generate complemented features representations Z_{av} and Z_{va}, respectively. Lastly, two complemented feature representations Z_{av} and Z_{va} are fused via channel-wise concatenation to produce the fused feature representation. Given the temporal feature representations Z_{audio} and Z_{visual}, the cross attention can be expressed as follows:

$$V_a = Z_{audio}W_{V\_a}\,; \qquad K_a = Z_{audio}W_{K\_a} \tag{1}$$

$$V_v = Z_{visual}W_{V\_v}\,; \qquad K_v = Z_{visual}W_{K\_v} \tag{2}$$

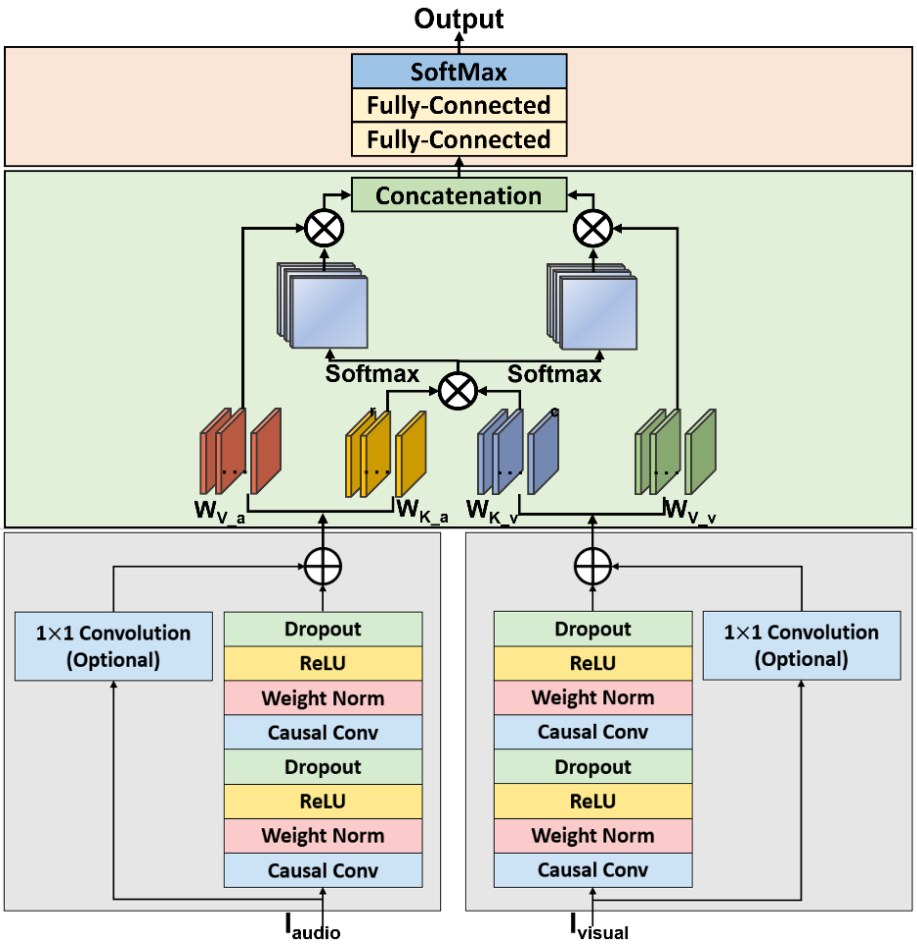


Fig. 1. Overall architecture of the proposed model for depression detection using multimodality.

$$S = \frac{K_a K_v^T}{\sqrt{d_k}} \tag{3}$$

$$Z_{av} = \text{Softmax}_r(S)V_a\,; \qquad Z_{va} = \text{SoftMax}(S)V_v \tag{4}$$

$$Z = \text{Concat}\,([Z_{av}; Z_{va}]) \tag{5}$$

Finally, the fused feature representation Z is passed through the FCLs followed by the SoftMax function to return probability of depression status. It is formed as follows:

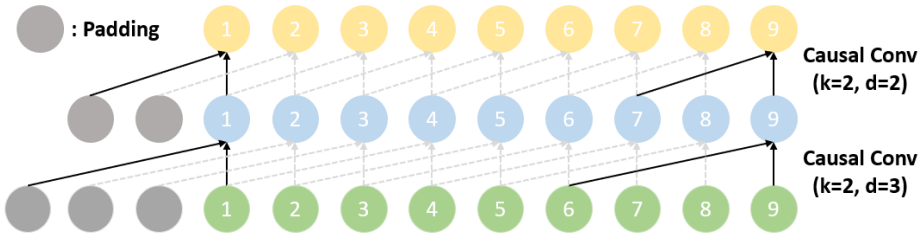


Fig. 2. Causal and dilated convolution in temporal feature learning module.

4 Materials and Experimental Results

4.1 Materials

In this study, we use D-Vlog dataset [13] to conduct comprehensive experiments. The D-Vlog dataset, which was collected from YouTube, consists of 961 vlog videos with

Table 1. The number of samples in the training, validation, and testing sets.

	#samples (depression/non-depression)
Training	647 (375/272)
Validation	102 (57/45)
Testing	212 (123/89)

555 depressed and 406 non-depressed samples from 816 different people. Due to personal privacy, the D-Vlog dataset only provides the extracted audio and visual extracted features for each sample. The visual extracted features are 68 coordinates of facial landmarks for each frame. Therefore, there are 136 elements (x and y coordinates) in each timepoint. Meanwhile, the audio features are extracted from the OpenSmile library that extracts 25 low-level audio features such as Mel-frequency cepstral coefficients (MFCCs), loudness, spectral flux, etc. The length of timepoint is from 24 to 3969.

4.2 Experimental Setting

The experiments are comprehensively conducted in Pytorch framework. We employ Adam optimizer [14] with a learning rate of 0.0001 to optimize parameters of the proposed model. The number of epoch and batch size is set to 30 and 32, respectively. Following the previous literature [11, 13, 15], we fix the length of input data and dimension of hidden feature representation at 596 and 256, respectively. The number of samples in the training, validation and testing sets are shown in Table 1.

Cross-entropy loss is adopted to measure difference between the predicted and actual labels. To compare the proposed model with previous methods, we use weighted average F1 score, weighted average precision, and weighted average recall metrics. The loss function and evaluation metrics can be expressed as follows:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{7}$$

$$F1_i = \frac{2TP_i}{2TP_i + FP_i + FN_i}; \quad wa\_F1 = \frac{|y_1|F1_1}{|y|} + \frac{|y_2|F1_2}{|y|} \tag{8}$$

$$Pre_i = \frac{TP_i}{TP_i + FP_i}; \quad wa\_Precision = \frac{|y_1|Pre_1}{|y|} + \frac{|y_2|Pre_2}{|y|} \tag{9}$$

$$Rec_i = \frac{TP_i}{TP_i + FP_i}; \quad wa\_Recall = \frac{|y_1|Rec_1}{|y|} + \frac{|y_2|Rec_2}{|y|} \tag{10}$$

Table 2. Performance comparison between the proposed method and the competing methods.

Method	wa_F1 (x10²%)	wa_Precision (x10²%)	wa_Recall (x10²%)
LR	0.5478	0.5486	0.5472
SVM	0.5297	0.5310	0.5519
RF	0.5784	0.5769	0.5949
KNN-Fusion	0.5425	0.5786	0.5943
BLSTM	0.5970	0.6081	0.6179
TFN	0.6100	0.6139	0.6226
Fusion_Concat	0.6110	0.6251	0.6321
Fusion_Add	0.5811	0.5911	0.6038
Fusion_Multiply	0.6309	0.6348	0.6415
Depression Detector [13]	0.6350	0.6540	0.6557
TAMFN [15]	0.6582	0.6602	0.6650
CAIINET [11]	0.6656	0.6657	0.6698
Ours	**0.6860**	**0.6990**	**0.6840**

Table 3. Ablation study on different combinations of proposed modules. (CA: our proposed cross-attention module; TCN: temporal convolutional network)

Method	wa_F1 (x10²%)	wa_Precision (x10²%)	wa_Recall (x10²%)
TCN w/o CA	0.6710	0.6755	0.6792
TCN w CA	0.6860	0.6990	0.6840

4.3 Experimental Results

In this section, we present the results of our proposed method and the previous method using the D-Vlog dataset. As can be seen in Table 2, the proposed method outperforms the other competing methods in terms of all the three metrics. The proposed method enhances 2.04-5.10%, 3.33-4.5%, and 1.42-2.83% compared to the other methods in terms of weighted average F1, weighted average precision, and weighted average recall, respectively. These improvements demonstrate that the combination of temporal feature learning and multimodal fusion modules produces the fused feature representation that can identify the patterns and relationships between the features and the classes better than the other methods.

In addition, we conducted an experiment to analyze the impact of the proposed multimodal fusion learning module that based on the cross-attention mechanism. As seen in Table 3, the combination of CA and TCN modules has higher results in terms of all three evaluated metrics.

4 Conclusion

Depression has a great attention from many researchers worldwide due to its mental health impacts. In particular, the use of vlog data, for example the audio and visual features that are extracted from the vlog, from social media network provide a fast way to predict. In this paper, we proposed the multimodal fusion technique that uses the attention-based mechanism to predict depression status. Our model consists of three components, namely, temporal feature learning component, multimodal representation fusion component, and prediction component. Experimental results show that our proposed model outperforms all the competing models in terms of all the evaluation metrics.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00208397)

This study was supported by a grant (HCRI 23026) Chonnam National University Hwasun Hospital Institute for Biomedical Science

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629) grant funded by the Korea government(MSIT)

References

1. World Health Organization. (2017). Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2). World Health Organization.
2. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
3. Hochreiter, S., & Schmidhuber, J. Long short-term memory. Neural computation, 9(8), 1735-1780 (1997)
4. Bai, S., Kolter, J. Z., & Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)
5. Wei, X., Zhang, T., Li, Y., Zhang, Y., & Wu, F. Multi-modality cross attention network for image and sentence matching. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10941-10950) (2020)
6. Hori, C., Hori, T., Lee, T. Y., Zhang, Z., Harsham, B., Hershey, J. R., ... & Sumi, K. Attention-based multimodal fusion for video description. In Proceedings of the IEEE international conference on computer vision (pp. 4193-4202) (2017).
7. Nagrani, A., Yang, S., Arnab, A., Jansen, A., Schmid, C., & Sun, C. Attention bottlenecks for multimodal fusion. Advances in Neural Information Processing Systems, 34, 14200-14213 (2021)
8. Choi, D., Zhang, G., Shin, S., & Jung, J. Decision Tree Algorithm for Depression Diagnosis from Facial Images. In 2023 IEEE 2nd International Conference on AI in Cybersecurity (ICAIC) (pp. 1-4). IEEE.
9. Tadesse, M. M., Lin, H., Xu, B., & Yang, L. Detection of depression-related posts in reddit social media forum. IEEE Access, 7, 44883-44893. (2019)
10. Jo, A. H., & Kwak, K. C. Diagnosis of Depression Based on Four-Stream Model of Bi-LSTM and CNN from Audio and Text Information. IEEE Access. (2022)
11. Zhou, L., Liu, Z., Yuan, X., Shangguan, Z., Li, Y., & Hu, B. CAIINET: Neural network based on contextual attention and information interaction mechanism for depression detection. Digital Signal Processing, 137, 103986 (2023)
12. Lin, C., Hu, P., Su, H., Li, S., Mei, J., Zhou, J., & Leung, H. Sensemood: depression detection on social media. In Proceedings of the 2020 international conference on multimedia retrieval (pp. 407-411).
13. Yoon, J., Kang, C., Kim, S., & Han, J. D-vlog: Multimodal Vlog Dataset for Depression Detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 11, pp. 12226-12234) (2022)
14. Kingma, D. P., & Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
15. Zhou, L., Liu, Z., Shangguan, Z., Yuan, X., Li, Y., & Hu, B. TAMFN: Time-aware Attention Multimodal Fusion Network for Depression Detection. IEEE Transactions on Neural Systems and Rehabilitation Engineering. (2022)

A Transformer-based Approach to Video Frame-level Prediction in Affective Behavior Analysis In-the-wild

Dang-Khanh Nguyen[1], Sudarshan Pant[1], Aera Kim[1],
Soo-Hyung Kim[1] and Hyung-Jeong Yang[1*]

[1] Department of Artificial Intelligence, Chonnam National University, Gwangju, Republic of Korea
* Corresponding author: Hyung-Jeong Yang (email: hjyang@jnu.ac.kr)

Abstract. The analysis of facial behavior has been extensively studied at the intersection of computer vision, psychology, and physiology. Its applications are widespread across diverse fields, ranging from healthcare, medicine to education and entertainment. Only in recent times, with the expansion of datasets and the utilization of robust machine learning techniques like deep neural networks, the field of automatic facial behavior analysis has experienced significant growth and progress. In the deep learning discipline, transformer architecture has been a dominating paradigm in many applications, including affective computing. In this paper, we propose our transformer-based model to handle frame-level prediction in a video to solve three well-known affective analysis tasks such as emotion classification, valence-arousal estimation, and action unit detection. Our proposed network is a straightforward model using only one deep-learned feature. It can respond fast to a real-time system requiring a tight timing constraint while maintaining a moderate performance. By using the attentive model and the synthetic dataset, our model outperforms the baseline in all tasks on the Aff-Wild2 dataset.

Keywords: Affective Behavior Analysis, Transformer, Video Frame-level Prediction.

1 Introduction

Recently, extensive research has focused on the analysis of facial expressions as a means of recognizing and understanding the emotions, behavior, and reactions of individuals, intending to develop machines capable of such comprehension. Effectively addressing the problem of analyzing and recognizing feelings plays a critical role in behavioral modeling, human-machine interaction, and the field of affective computing. Numerous applications encompass diverse domains, including medicine, healthcare, e-learning, marketing, entertainment, and law.

Representing human emotion is a fundamental topic in affective computing [3,6]. A naïve approach is using a discrete classification of 7 basic emotions: neutral, happiness, sadness, surprise, anger, disgust, and fear. Besides, The Facial Action Coding System (FACS) [17] proposes representing facial expressions as the presentation of muscle

movements on the human face, called the action units. Apart from these above categorial approaches, emotion could also be represented in a continuous two-dimensional space that consists of valence and arousal axis. Analyzing human interaction from various perspectives can help researchers deeply understand their feelings and behavior.

Historically, research primarily concentrated on controlled environments because of the lack of large-scale real-life datasets. However, the landscape has shifted with the widespread utilization of social media and platforms, leading to a significant availability of substantial data. Furthermore, deep learning has emerged as a viable solution for addressing visual analysis and recognition challenges. As a result, significant research efforts have been dedicated to the advancement and application of deep learning techniques and deep neural networks across various domains. This includes the exploration of affect recognition in uncontrolled environments.

In various modern multimedia applications such as social media platforms or customer service, it is crucial to estimate the users' feelings, interests, or reactions in a real-time manner. It motivates us to develop an artificial intelligence framework capturing facial clues and generating human behavior analysis for each frame in a video. This paper introduces our deep neural network for in-the-wild video frame-wise prediction focusing on affective behavior analysis. Concisely, we utilize Transformer to encode the temporal information from the sequence of deep-learned facial features. The model is trained for three individual tasks including expression classification (EX), valence-arousal estimation (VA), and action unit detection (AU). Noticeably, in expression classification task, we exploit the generated facial images [1] to resolve the imbalance limitation in the dataset and create more data samples for the training process.

2 Related Works

Recently, various studies have been conducted on applying deep learning in facial behavior analysis. Savchenko [13] proposed a lightweight model using EfficientNet backbone to recognize emotion, estimate valence-arousal, and detect action units. The backbone was pre-trained with face recognition task and was fine-tuned on emotion recognition. Afterward, the trained network was used to effectively extract the facial features for many face-related tasks. Peng Zou et. al. [14] used the multi-head attention in multimodal learning and temporal learning. In their solution, they combined VGGFace2 Resnet-50 and AffectNet Resnet-50 features as the input of the deep learning network.

Furthermore, several works have attempted to infer various tasks simultaneously, which is also known as multi-task learning approach. Zhang et. al. [15] combined various feature representations comprising masked autoencoder-based features, IResNet-based features, and DenseNet-based features. They also observed the importance of the relationship among temporal neighbor frames and employed a temporal encoder to obtain that useful information. Nguyen et. al. [11] employed cross-attention to learn the association between the presence of AUs and human emotion. Moreover, they built a graph neural network to explore the relation among facial movements.

Even though Savchen [13], Nguyen [11], and Peng Zou [14] leveraged the networks pre-trained on various facial behavior tasks, they did not utilize the long-range interaction between frames in a sequence. This useful information can be exploited to capture the temporal relationship between features in different time steps. In this work, we employ the transformer as a sequence decoder to extract this information. Meanwhile, Zhang [15] used a cumbersome set of deep-learned features. It may be unsuitable for frame-wise prediction tasks because they require fast inference in real-world applications. To counter this issue, we suggest our simplified model that relies on a single deep-learned feature. This model exhibits swift responsiveness to real-time systems with strict timing requirements while maintaining a satisfactory level of performance.

3 Proposed Method

Our proposed model includes two parts, Facial-feature Learning (FFL) and Task-specific Learning (TSL), with cascaded connection as in Figure 1. Three paradigms, corresponding to three tasks, share the same FFL configuration but are different in TSL settings. Given a sequence of frames from a video clip, we fragment it into multiple segments with a common length L. Consequently, the dataset could be considered as a set of samples, each sample is a segment with a fixed number of consecutive frames. FFL takes a segment as an input and generates a sequence of facial features of length L, and feeds them to the TSL block. This block creates a series of output vectors and updates the model's parameters via a loss function. The output size and the target criterion of TSL depend on the task that the model is working on.

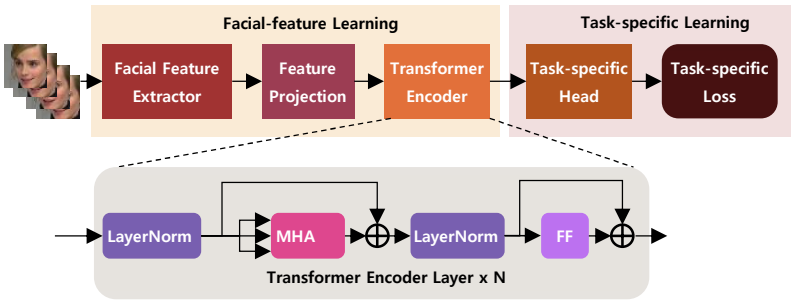


Fig. 1. Block diagram of our proposed method.

3.1 Facial Feature Learning

Given an i^{th} sample X_i, which is a series of facial frames $\{x_j \mid j = 1, \ldots, L\}$, we create a sequence of feature vectors by exploiting the pre-trained EfficientNet [13] well trained with multiple facial analysis tasks. Afterward, we project these vectors to a

latent feature space by a linear layer to acquire the sequence embeddings E_i. The procedure is formulated as the function below.

$$E_i = Linear\big(Effnet(X_i)\big) \tag{1}$$

The projected embeddings are now ready to be put into a transformer encoder. The encoder resembles the conventional transformer encoder [12] with multi-head attention and feed-forward networks except that we use pre-normalization instead of post-normalization. The encoder captures the long-range temporal dependency among the embeddings and generates the facial features $F_i = \{f_j \mid j = 1, \dots, L\}$.

$$F_i = TransformerEcoder(E_i) \tag{2}$$

3.2 Task-specific Head and Loss

The output from the encoder is sent to the task-specific head, which is simply a multilayer perceptron (MLP) with one hidden layer. The output dimension of the network depends on the target emotional space. Particularly, it could be either 2, 8, or 12 corresponding to VA, EX, and AU task, respectively. The output $\hat{Y}_i = \{\hat{y}_j \mid j = 1, \dots, L\}$ is denoted as below.

$$\hat{Y}_i = MLP_t(F_i) \tag{3}$$

Concerning the task-specific loss, following the idea of Nguyen [11], we train each model with a particular function corresponding to the task that it handles. For instance, the emotion recognition model is trained with weighted cross-entropy loss. An emotion with less occurrence rate in the training set will associate with a higher coefficient to re-balance their contribution to the loss.

$$l_{EX} = -P_y \, log\left(\frac{e^{\hat{y}}}{\sum e^{\hat{y}_j}}\right) \tag{4}$$

where P_k is the re-weighting factor for each class, which is computed from the training set data distribution, and $y \in \{0, \dots, 7\}$ is the ground truth emotion of the frame. The total loss of a batch is the average of losses of all frames in a batch:

$$\mathcal{L}_{EX} = -\frac{1}{B \times L}\sum_{i=1}^{B}\sum_{j=1}^{L} l_{EX}(i,j) \tag{5}$$

In the VA-estimation task, we use the summation of the Concordance Correlation Coefficient (CCC) loss of valance and arousal to explicitly maximize the metric of this task, which is also the average CCC of the two emotional dimensions. The loss and score are computed as shown in equations (6) and (7):

$$\mathcal{L}_{VA} = 1 - CCC^V + 1 - CCC^A \tag{6}$$

$$CCC = \frac{2s_{y\hat{y}}}{s_y^2 \; + \; s_{\hat{y}}^2 + \left(\bar{y} - \bar{\hat{y}}\right)^2} \tag{7}$$

In equation (7), $s_{\hat{y}}$ and s_y are variances of predicted values and ground truth, respectively. Their mean values are $\bar{\hat{y}}$ and $\bar{y}$, $s_{y\hat{y}}$ denotes their covariance. For AU-detection model, we choose the weighted average of binary cross-entropy (BCE) loss of 12 AUs. We also balance the influence of each AU a to the function by assigning a weight w_a for each BCE element. The loss is formulated as follows:

$$l_{AU} = -\frac{1}{12}\sum_{a=1}^{12} w_a[y_a\, log(\hat{y}_a) + (1 - y_a)\, \hat{y}_a log(1 - \hat{y}_a)] \quad (8)$$

4 Dataset and Experiments

4.1 Dataset

We used Aff-Wild2 [7] and the synthetic facial dataset [1] for our training process. The Aff-Wild2 for expression classification is a collection of 546 videos. More than 2 million frames are extracted from the videos and annotated as one among 8 emotional classes. The classification comprises 6 basic emotions: “anger”, “disgust”, “fear”, “happiness”, “sadness” and “surprise”. In the dataset, two additional classes, namely "neutral" and "other," are included, and they also exhibit the highest sample counts. -

To enlarge the scale of the training dataset, we leveraged the synthetic facial dataset from the 4th ABAW competition, Learning from Synthetic Data challenge. The corpus includes 277,251 synthesized images classified into 6 basic classes. We merge this dataset with the training split of the Aff-Wild2 dataset. It does not only increase the size of training samples but also decreases the influence of the imbalance issue.

The Aff-Wild2 is provided in three versions: raw videos, cropped images, and cropped-aligned images. For our experiments, we utilized cropped-aligned images, which maintain a consistent size of 112x112 pixels. As for the synthetic data, the image size is set to 128x128 pixels.

4.2 Experiment Setting

The sequence of frames in each video is split into segments comprising 64 frames. All images are normalized and resized to 224x224 pixels. Regarding the transformer, we use a hidden size and feed-forward size of 512, dropout ratio is 0.1. The emotion detection head is a neural network with one hidden layer with a size of 256. We train our models in 20 epochs with a batch size of 64 and a learning rate of 0.001. The Adam optimizer is used with a weight decay of $\frac{1}{64}$.

To boost the performance, we apply the ensemble of 3 models with different configurations of the transformer encoder. The number of heads and layers in 3 settings are (4, 4), (8, 4), and (4, 6), respectively. Soft average voting is used to acquire the final prediction from 3 logit vectors of the models.

4.3 Results

We compare our models in various settings. Particularly, the performance on the validation set of each approach for the expression recognition task is shown in Table 1. Using a pre-trained EfficientNet (Effnet) followed by fully connected layers, we obtained a score of 0.3327. Based on that model, we added a transformer encoder to the model and improved the F1-score. Moreover, when exploiting the synthetic data to train the transformer encoder, we boosted the results significantly to more than 0.44. Next, we tried to increase the scale of the transformer, particularly the number of heads and encoder layers. The performance was slightly enhanced but not noticeable. Afterward, we used an ensemble to combine the logit output of each model setting. Consequently, we attained a better output compared to a single model. The combinations of two transformer-based models got the F1-score from 0.4663 to 0.4729. The ensemble of the 3 best configurations accomplished a score of 0.4775. This is also our best result on the validation set of Aff-Wild2.

Table 1 The performance on the validation set of various settings. Effnet and FCs stand for Pre-trained EfficientNet and Fully Connected Layers. N and h are the number of layers and heads in transformer encoder. "Syn" implies that the experiments use synthetic data in the training process.

Model configurations	F1-score
Effnet+FCs	0.3327
Effnet+Encoder (N=4, h=4) +FCs	0.3615
(1) Effnet+ Encoder (N=4, h=4)+FCs+Syn	0.4400
(2) Effnet+Encoder (N=4, h=8)+FCs+Syn	0.4424
(3) Effnet+Encoder (N=6, h=4)+FCs+Syn	0.4555
Soft average voting (1) (2)	0.4663
Soft average voting (1) (3)	0.4672
Soft average voting (3) (2)	0.4729
Soft average voting (1), (2), and (3)	**0.4775**

Regarding the two remaining challenges, we applied the same methodology and configuration except that the synthetic dataset was not used. Similarly, the metric scores were enhanced after we fused the prediction of multiple models by soft average voting. Tables 2 and 3 provide the detailed results of AU detection and VA estimation tasks, respectively.

Table 2 The results of our proposed method on the validation set of Aff-Wild2 in AU detection task.

Model configurations	F1-score
(1) Effnet+Encoder (N=4, h=4)+FCs	0.51696
(2) Effnet+Encoder (N=4, h=8)+FCs	0.51146
(3) Effnet+Encoder (N=6, h=4)+FCs	0.51192
Soft average voting (1) (2)	0.51960
Soft average voting (1) (3)	0.52021
Soft average voting (3) (2)	0.51709
Soft average voting (1), (2), and (3)	**0.52085**

Table 3 The results of our proposed method on the validation set of Aff-Wild2 in VA estimation task.

Model configurations	F1-score
(1) Effnet+Encoder (N=4, h=4)+FCs	0.48296
(2) Effnet+Encoder (N=4, h=8)+FCs	0.48819
(3) Effnet+Encoder (N=6, h=4)+FCs	0.47389
Soft average voting (1) (2)	0.49684
Soft average voting (1) (3)	0.49679
Soft average voting (3) (2)	0.49874
Soft average voting (1), (2), and (3)	**0.50290**

Eventually, we chose the model performing the best on the validation set in each task to evaluate our method on the test set. As a result, the average F1 score in AU detection was 0.4563, and the average CCC of valance and arousal in VA estimation was 0.4640. On the other hand, the macro-F1 score of 8 emotions in EX recognition task attained a score of 0.2949. As shown in Table 4, our proposed model outperforms the baseline models of the datasets and attains higher scores in EX and AU tasks compared to the method of Peng Zou [14].

Table 4 The results of various methods on the Aff-Wild2 test set.

Model	EX task F1-score	AU task F1 score	VA task Avg-CCC
Baseline [16]	0.2050	0.3650	0.2010
Peng Zou [14]	02846	0.3776	**0.4842**
Ours	**0.2949**	**0.4563**	0.4640

5 Conclusion

In this study, we provided a straightforward and effective method for frame-wise classification in videos. The transformer is employed to learn the correlation among the frames in a sequence. The proposed network is feasible to be embedded in real-world behavior applications because it is lightweight and well-trained on a large in-the-wild dataset. However, our future research into affective computing is focusing on constructing a multi-level transformer architecture for multimodal fusion to exploit the acoustic and linguistic features in the video.

Acknowledgments.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00219107).

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629) grant funded by the Korea government (MSIT).

References

1. Kollias, Dimitrios. ABAW: Learning from Synthetic Data & Multi-Task Learning Challenges. arXiv preprint arXiv:2207.01138, 2022.
2. Kollias, Dimitrios. Abaw: Valence-arousal estimation, expression recognition, action unit detection & multi-task learning challenges. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2328–2336), 2022.
3. Kollias, Dimitrios, Viktoriia, Sharmanska, and Stefanos, Zafeiriou. Distribution Matching for Heterogeneous Multi-Task Learning: a Large-scale Face Study. arXiv preprint arXiv:2105.03790, 2021.
4. Kollias, Dimitrios, and Stefanos, Zafeiriou. Analysing affective behavior in the second abaw2 competition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3652–3660), 2021.
5. Kollias, Dimitrios, and Stefanos, Zafeiriou. Affect Analysis in-the-wild: Valence-Arousal, Expressions, Action Units and a Unified Framework. arXiv preprint arXiv:2103.15792 (2021)
6. Kollias, D, A, Schulc, E, Hajiyev, and S, Zafeiriou. Analysing Affective Behavior in the First ABAW 2020 Competition. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG) (pp. 794–800), 2020.
7. Kollias, Dimitrios, and Stefanos, Zafeiriou. Expression, Affect, Action Unit Recognition: Aff-Wild2, Multi-Task Learning and ArcFace. arXiv preprint arXiv:1910.04855, 2019.
8. Kollias, Dimitrios, Viktoriia, Sharmanska, and Stefanos, Zafeiriou. Face Behavior a la carte: Expressions, Affect and Action Units in a Single Network. arXiv preprint arXiv:1910.11111, 2019.
9. Kollias, Dimitrios, Panagiotis, Tzirakis, Mihalis A, Nicolaou, Athanasios, Papaioannou, Guoying, Zhao, Bjorn, Schuller, Irene, Kotsia, and Stefanos, Zafeiriou. Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond. International Journal of Computer Vision, 2019.
10. Zafeiriou, Stefanos, Dimitrios, Kollias, Mihalis A, Nicolaou, Athanasios, Papaioannou, Guoying, Zhao, and Irene, Kotsia. Aff-wild: Valence and arousal 'in-the-wild' challenge. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on (pp. 1980–1987), 2017.
11. Nguyen, Dang-Khanh, Sudarshan Pant, Ngoc-Huynh Ho, Guee-Sang Lee, Soo-Hyung Kim, and Hyung-Jeong Yang. Affective Behavior Analysis Using Action Unit Relation Graph and Multi-task Cross Attention. In European Conference on Computer Vision, pp. 132-142. Springer, Cham, 2023.
12. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in neural information processing systems 30, 2017.
13. Savchenko, Andrey V. Frame-level prediction of facial expressions, valence, arousal and action units for mobile devices. arXiv preprint arXiv:2203.13436, 2022.
14. Zou, Peng, Rui Wang, Kehua Wen, Yasi Peng, and Xiao Sun. Spatial-temporal Transformer for Affective Behavior Analysis. arXiv preprint arXiv:2303.10561, 2023.
15. Zhang, Tenggan, Chuanhe Liu, Xiaolong Liu, Yuchen Liu, Liyu Meng, Lei Sun, Wenqiang Jiang, Fengyuan Zhang, Jinming Zhao, and Qin Jin. Multi-Task Learning Framework for Emotion Recognition In-the-Wild. In European Conference on Computer Vision, pp. 143-156. Springer, Cham, 2023.
16. Kollias, Dimitrios, Panagiotis, Tzirakis, Alice, Baird, Alan, Cowen, and Stefanos, Zafeiriou. ABAW: Valence-Arousal Estimation, Expression Recognition, Action Unit Detection & Emotional Reaction Intensity Estimation Challenges. arXiv preprint arXiv:2303.01498, 2023.
17. Charles Darwin and Phillip Prodger. The expression of the emotions in man and animals. Oxford University Press, USA, 1998.

Cooperation Research Network Analysis: AI-Based BioHealth

Seongsu Jang[1], Junghwan Lee[2]

1 Graduate School of Big data Collaboration Courser, Chungbuk National University, Cheongju, Korea, jangss0525@gmail.com
2 Department of Management Information Systems, Chungbuk National University, Cheongju, Korea, junghwan@cbnu.ac.kr

Abstract. In order to fortify the future ecosystem of science and technology innovation, this study conducted a global cooperation analysis using research papers, published on the Web of Science between 2016 and 2021 in the field of artificial intelligence-based biohealth. As a result, a total of 29,237 papers were retrieved during this period, and it was found that 62.9% (18,405 papers) were conducted as cooperation research. Regarding collaborative activities, it was discovered that institutions located in the United States, such as Harvard University and Stanford University, engaged in more cooperation research than institutions in other countries. The network analysis further revealed that a significant cooperation network has been formed around Harvard University. However, since most of these cooperation networks consist of three or fewer institutions (nodes), it was confirmed that the cooperation network ecosystem in this field is still in its early stages. Based on these findings, the study concludes by emphasizing the need for and suggesting ways to strengthen the collaborative research ecosystem in the field of future artificial intelligence-based biohealth.

Keywords: Artificail Intelligence(AI), Biohealth Innovation, Network Analysis, Research Cooperation, Innovation Ecosystem

1 Introduction

In recent years, computer performance has significantly improved and the amount of data available has increased exponentially, causing the advancement of artificial intelligence to accelerate at a rate seven times faster than Moore's Law suggests (OpenAI, 2018). As we step into the era of the fourth industrial revolution, artificial intelligence has become a crucial base technology for intelligent systems. It is being employed in diverse domains such as autonomous vehicles, smart robots, and

healthcare, acting as a catalytic force that propels future societal evolution and economic growth (Kim Byung-Woon , 2016).

In light of these developments, several social and institutional changes are being initiated to establish new value chains and growth opportunities by integrating artificial intelligence technology into the biohealth field. The Korean government, for instance, has announced its intent to harness artificial intelligence technology to advance the digital technology required in the medical field and continue to support national new drug development projects (Ministry of Health and Welfare, 2023).
In accordance with these national efforts, the R&D scope in the domestic biohealth field has been progressively expanding since 2016, and there has been an exponential increase in the volume of literature, including theses and patents (Kim Jong-ran et al., 2021). However, despite these efforts, South Korea's biochip and sensor technology levels and capabilities in major bio fields such as infectious disease response are assessed to be inferior to those of the leading countries worldwide (Kim Hong-yeol, 2022).

In response to this, the Korean government is establishing a cooperative ecosystem between technology-leading companies and start-ups, as well as venture firms. By forming a biohealth consultative group, they aim to build a mutually beneficial cooperation network with local businesses, universities, research institutes, and hospitals that covers everything from research and development to clinical trials. Efforts are underway to strengthen the support system at all stages, including research and commercialization.

To tackle the uncertainties of the future and societal challenges, a cooperative R&D strategy for biotechnology innovation is required. For a long time, technological cooperation has been leveraged to mitigate uncertainty in technology development and to gain a competitive advantage by enabling organizations to acquire lacking resources from external sources and create new outcomes that couldn't be achieved with internal capabilities alone (Powell et al., 1996; Stuart, 1998). With this in mind, this study aims to suggest methods to invigorate the future ecosystem of science and technology innovation. It does so by analyzing the characteristics of global R&D cooperation as reflected in recent AI-based biohealth research from 2016 to 2021.

2 Method

In this study, our goal was to understand the characteristics of technological collaboration in the field of AI-based biohealth. Our approach involved collecting related research papers, purifying data, constructing networks, and visualizing processes. Firstly, to gather the necessary data for this analysis, we collected research papers from 2016 to 2021 from the international academic paper database, Web of Science (WoS). To ensure a high quality of the collected papers, we included only those listed in SCIE and SSCI. Next, we undertook a data cleansing process for the collected data. For this step, we extracted the names of the institutions that published the papers and the countries to which those institutions belonged from the 'Addresses' field in the analysis data. Following that, as depicted

in <Table 1>, we standardized the names of the institutions through a four-stage process: (1) standardization, (2) simplification, (3) removal of sub-institutions, and (4) clarification. Furthermore, we simplified the names of the countries. Finally, we examined the characteristics of the cooperative network using the refined analysis data. Network analysis is a text mining technique employed to discern the relationships between nodes that appear in a document. In this study, we conducted a collaborative network analysis to identify the cooperative relationships between research institutions. Specifically, we utilized 'Degree centrality,' a structural attribute of the network. Centrality is an index that numerically conveys the relative importance of research institutions within the entire network. In other words, it signifies the degree to which a particular research institution is connected to others. For our analysis, we employed NetMiner, a network analysis tool developed by the company Cyram.

Table 1. Data Preprocessing

Types of Activities		Existing	Change
institutions name	standardization	1) BOSCH CO 2) BOSCH AG …	BOSCH …
	simplification,	1) LG ELECT INC 2) KOREA ADVANCED INSTITUTE OF SCIENCE & TECHNOLOGY …	1) LG ELECT 2) KAIST …
	elimination of subsidiary institutions, and	1) HARVARD LAW SCH 2) BOSCH CORP RES …	1) HAVARD UNIV 2) BOSCH …
	clarification	1) USTC 2) UNSW …	1) UNIV SCI & TECHNOL CHINA 2) UNIV NEW SOUTH WALES …
Country name	simplification	1) PEOPLES R CHINA 2) SWITZERLAND	1) CHINA 2) SWISS

3 Result

The number of research papers published in the field of AI-based biohealth from 2016 to 2021 totals 29,237. The rate of collaborative research during this period was 62.9%, with little variation in the rate of collaborative research across each period.

Table 2. The number of Paper

계	The number of article	The number of Cooperated article	Cooperated Ratio
Total	29,237	18,405	62.9
2016~2018	5,039	3,187	63.2
2019~2021	24,198	15,218	62.8

As depicted in <Table 3>, the United States led in paper publication frequency across both periods, closely followed by China, England, and Germany. Notably,

over the past three years, there's been a significant increase in research collaboration efforts from Asian countries such as South Korea and India.

Table 3. The number of Country

2016~2018			2019~2021		
Rank	Country	Count	Rank	Country	Count
1	USA	1,437	1	USA	5,356
2	China	557	2	China	3,556
3	England	487	3	England	2,097
4	Germany	250	4	Germany	1,167
5	Australia	193	5	South Korea	994
	Italy	193	6	Canada	915
7	Canada	183	7	India	898
8	Spain	178	8	Australia	877
9	France	164	9	Italy	859
10	Netherlands	149	10	France	745

Institutions like Harvard University, Stanford University, and UCL have consistently led in terms of the number of collaborative research publications over the past six years. However, some institutions experienced a decline in their rankings over time, such as Johns Hopkins University (from 4th to 6th) and the University of Washington (from 5th to 7th). On the other hand, the University of Oxford made a notable ascent to the 5th rank from 2019 to 2021, with 244 research collaborations recorded over the last three years.

Table 4. The number of Institution

2016~2018			2019~2021		
Rank	Institution	Count	Rank	Institution	Count
1	HARVARD UNIV	214	1	HARVARD UNIV	779
2	STANFORD UNIV	93	2	STANFORD UNIV	350
3	UCL	75	3	UCL	298
4	JOHNS HOPKINS UNIV	69	4	UNIV OF TORONTO	266
5	UNIV OF WASHINGTON	68	5	UNIV OF OXFORD	244
6	UNIV OF MICHIGAN	66	6	JOHNS HOPKINS UNIV	237
7	MIT	65	7	UNIV OF WASHINGTON	226
8	US DEPARTMENT OF VETERANS AFFAIRS	63	8	MIT	223
9	UNIV OF PENNSYLVANIA	61	9	IMPERIAL COLLEGE LONDON	221
10	IMPERIAL COLLEGE LONDON	58	10	UNIV OF PENNSYLVANIA	217

An examination of centrality to analyze the characteristics of the collaborative research network suggests that a fully established technological collaboration ecosystem is yet to be realized. This conclusion is based on the fact that, as presented in <Table 4>, the centrality value of most collaborative research networks formed over the past six years is less than 0.1, indicating a relatively low score. To further scrutinize the characteristics of the current technology collaboration ecosystem, we analyzed the relationships between nodes (organizations) within the collaborative network. However, it's worth noting that the AI-based biohealth field tends to produce many research papers from diverse perspectives, resulting in relatively large scale and volume. This characteristic poses a challenge when trying to determine which institutions are at the core of the overall collaboration network

and understanding the composition of these institutions. Thus, a detailed analysis might be limited in this regard.

Table 5. The number of Degree centrality in each periods

2016~2018			2019~2021		
Rank	Institution	Index	Rank	Institution	Index
1	HARVARD UNIV	0.148	1	HARVARD UNIV	0.188
2	UCL	0.075	2	UNIV OF OXFORD	0.120
3	JOHNS HOPKINS UNIV	0.072	3	UCL	0.119
4	STANFORD UNIV	0.070	4	UNIV OF TORONTO	0.114
5	MIT	0.064	5	JOHNS HOPKINS UNIV	0.113
6	CTR DE INVEST BIOMEDICA EN RED	0.061	6	STANFORD UNIV	0.109
7	UNIV OF ZURICH	0.059	7	UNIV PARIS CITE	0.107
8	UNIV OF OXFORD	0.058	8	IMPERIAL COLLEGE LONDON	0.01
9	UNIV OF MICHIGAN	0.053		UNIV OF CAMBRIDGE	0.01
	KINGS COLLEGE LONDON	0.053	10	COLUMBIA UNIV	0.094

Therefore, In this study, we set a minimum frequency for connections between organizations (nodes) in the network graph. This approach allowed us to reduce the number of organizations included in the visualization and identify the key collaborating organizations. Consequently, a collaborative network was established for organizations that had more than 6 collaborations from 2016 to 2018, and over 19 collaborations from 2019 to 2021.

As shown in <Fig. 1> and <Fig. 2>, the size and scale of the collaborative network, primarily revolving around Harvard University, have grown over time. Between 2016 and 2018, UCL constructed a distinct network with itself at the core, but after 2019, Harvard University seemingly assumed a mediating role within the central network. Hence, while the dimensions and scale of the network, along with the roles of the nodes within it, have been dynamically evolving over time, a significant number of smaller networks, consisting of three or fewer nodes, continue to persist.

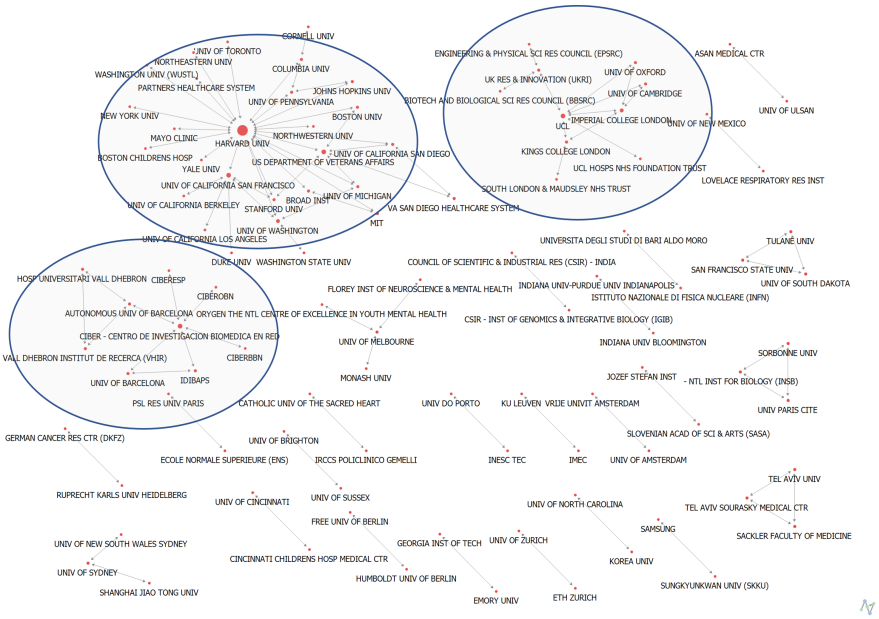


Fig. 1. Cooperation Network(2016~2018)

Fig. 2. Cooperation Network(2019~2021)

4 Conclusion And Implications

This study collected and analyzed 29,237 papers from Web of Science, over a span of six years, from 2016 to 2021. The papers were all centered on the field of artificial intelligence-based biohealth, a rapidly expanding research area. This study aim to observe the dynamic changes occurring within the cooperation research network.

From the analysis, we drew two main conclusions. Firstly, we discovered that the ratio of collaborative research to the total number of studies published over the last six years was 62.9%. Secondly, when examining the cooperation network, it became clear that while Harvard University held a central position and contributed to strengthening the cooperation ecosystem, most other networks formed by different organizations consisted of three or fewer entities. This finding confirms that the cooperation ecosystem in the corresponding field is not yet fully developed.

The results of this study have several important implications. Firstly, given that local governments are currently promoting investment in the biohealth industry and biohealth clusters to stimulate the local economy, it's crucial to select institutions that have shared interests, particularly in terms of time and cost. Secondly, there is a need for institutional efforts to preserve and enhance the collaborative networks of various institutions over an extended period. In the case of smaller institutions, despite having high technological potential, they often encounter difficulties in transforming excellent ideas into tangible technology due to the lack of analytical equipment and development funds. Therefore, it is imperative to devise technology innovation strategies that maintains the research capabilities of traditionally excellent institutions in the region and enhances the capacities of smaller institutions. This can be achieved by strengthening regional collaboration networks, which is essential for creating new opportunities and reinforcing the domestic AI-based biohealth ecosystem.

References

1. Kim, Byung-Woon: Artificial Intelligence Trend Analysis and National Policy Suggestions, Informatization Policy, vol.23.1, pp.74-93. (2016)
2. Kim Heung-yeol, Lee Ji-hyun, Kim Hyun-soo, Lee Ji-yeon, Seol-min, Moon Seong-hoon: A study on the establishment of mid- to long-term R&D investment strategies in the field of bio and health (2021)
3. Kim Jong-ran, Kang Yu-jin, Hong Mi-young: Biohealth policy and investment trends
4. Ministry of Health and Welfare: Biohealth New Market Creation Strategy (2023)
5. Open AI: AI and Compute – OpenAI (2018), https://openai.com/research/ai-and-compute
6. Powell, Walter W., Kenneth W. Koput, and Laurel Smith-Doerr: Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology, Administrative science quarterly, pp.116-145.(1996)

7. Stuart, Toby E.: Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry, Administrative science quarterly, pp. 668-698. (1998)

Pose Attention-based Knowledge Distillation for Human Action Recognition

Jeong-Hun Kim[1], Yoo-Sung Kim[2], Aziz Nasridinov[1,3,*]

[1] Bigdata Research Institute, Chungbuk National University, Cheongju, South Korea
[2] Department of Artificial Intelligence, Inha University, Incheon, South Korea
[3] Department of Computer Science, Chungbuk National University, Cheongju, South Korea
etyanue@chungbuk.ac.kr, yskim@inha.ac.kr, aziz@chungbuk.ac.kr

Abstract. Human action recognition is a representative task in video understanding, which involves modeling the motion information of humans within a video. This motion information is closely related to the human pose estimation. In this work, we propose a multi-task framework for human action recognition by distilling knowledge related to motion information from human 2D pose estimation. To establish the combination of two distinct tasks, we propose a new network architecture that distills knowledge through the activation differences between a main network, which utilizes RGB frame sequences as input, and auxiliary network, which utilizes heatmap sequences generated by human 2D pose estimation as input. We show that the knowledge distilled through human 2D pose estimation leads to the attention of action-salient regions in videos, thereby improving the action recognition performance.

Keywords: Human Action Recognition, Pose Estimation, Multi-Task Learning, Knowledge Distillation, Video Understanding

1 Introduction

Human action recognition is an important problem which benefits high-level video understanding tasks, such as video question answering [1-3], video retrieval [4-6], and video object detection [7-9]. However, recognizing human actions commonly requires temporal modeling to learn the motion information of humans [10]. To accomplish effective temporal modeling for the human action recognition, many researchers have proposed novel network architectures based on 2D and 3D convolutional neural networks (CNNs) [11-15]. These architectures significantly improve the action recognition performance, but they have some drawbacks: 1) as the layers deepen, motion information tends to be lost; 2) instead of focusing on human-centric motion information, they recognize human actions based on visual compositions such as background and objects [10, 16]. In particular, recognizing human actions based on visual compositions increases the intra-class variance for each action, leading to the degradation in action recognition performance.

* Corresponding author.

In this paper, we propose a novel multi-task framework to enhance human-centric motion information by distilling human pose information. It is based on the observation that the multi-modal distillation is sufficient to learn good feature representations for perception tasks [17]. More precisely, our framework consists of a main network that utilizes RGB frame sequences as input for human action recognition and an auxiliary network that extracts human pose information through the human 2D pose estimation. The auxiliary network then infuses this human pose information into the main network. Based on this knowledge distillation, our framework pays attention to human-centric motion information. To evaluate our framework, we conduct experiments on three video benchmark datasets: Kinetics-400 [18], UCF101 [19], and Something-Something V2 (SS-V2) [20], comparing the results with existing action recognition methods. Experimental results demonstrate that our framework highlights human-centric motion information and outperforms the existing action recognition methods.

2 Related work

Attempts for human action recognition could be categorized into two approaches: two-stream networks and 3D convolutional neural networks (CNNs). Specifically, the two-stream networks approach [12] often utilizes two separate CNNs for RGB frame sequences and optical flows, respectively, along with an average pooling operation to fuse visual and motion information. However, it often mismatches both visual and motion information since it fuses them in the later layers only. To address this problem, some researchers have proposed variant methods of two-stream networks approach: R(2+1)D [21] and TSM [22]. R(2+1)D utilizes 2D CNNs to extract features from RGB frames and 1D CNNs to aggregate these features. Based on these tightly coupled 2D CNNs and 1D CNNs, R(2+1)D addresses the mismatch problem between visual and motion information. TSM replaces the average pooling operation in the two-stream networks approach with a shift module to enhance the correlation between visual and motion information.

On the other hand, 3D CNN-based approach jointly learns spatial and temporal features, which correspond to visual and motion information, by stacking 3D convolutions. C3D [13], I3D [14], and SlowOnly [15] are representative methods in the 3D CNN-based approach. C3D employs 3D convolutions with the input of a short clip to learn motion information. Inflated 3D convolutional neural network (I3D) extends the GoogleNet [23] to extract strong motion information by utilizing inception layers. SlowOnly is one of the two streams in SlowFast [15], and it preserves the global and local consistencies of human actions by learning motion information at different frame rates.

However, despite the importance of human-centric motion information and interactions between humans, objects, and the environment for recognizing human actions, existing human action recognition methods do not focus on human-centric motion information. For this reason, existing methods often recognize human actions with bias towards visual compositions or motion information of irrelevant objects, rather than human-centric motion information.

3 Proposed method

This section describes the proposed framework, which enhances human-centric motion information based on the distillation of human pose information, to recognize human actions. An overview of the proposed framework is illustrated in Fig. 1. To implement the proposed framework, two separate 3D CNNs, the main network f and the auxiliary network g, need to be designed properly. The detail specification of each network architecture is demonstrated in Fig. 2. The input shape of the main network is [3x32x128x128], where the dimensions represent the number of channels, sequence length, height, and width, respectively. In the case of the auxiliary network, the input shape is [15x32x128x128]. Here, in the auxiliary network, the number of channels corresponds to the number of joints (i.e., the number of heatmaps) identified by the human 2D pose estimation. Both main and auxiliary networks adopt the ResNet layers to prevent the information loss when the layers are deepened. Specifically, the activations of each ResNet layer in the main and auxiliary networks have the same shape for knowledge distillation. As shown in Fig. 1, we first detect humans in the entire frames of an input video. Then, we estimate heatmaps, which represents human 2D joints, by utilizing the pre-trained human 2D pose estimation model. Afterward, we feed the RGB frame and heatmap sequences into the main network f and the auxiliary network g, respectively. During extracting the visual and motion information in the main and auxiliary networks, we infuse the human pose information of the auxiliary network to the main network by the knowledge distillation. Based on this knowledge distillation, the main network learns powerful human-centric motion information; and thus, it recognizes human actions by considering human-centric motion information

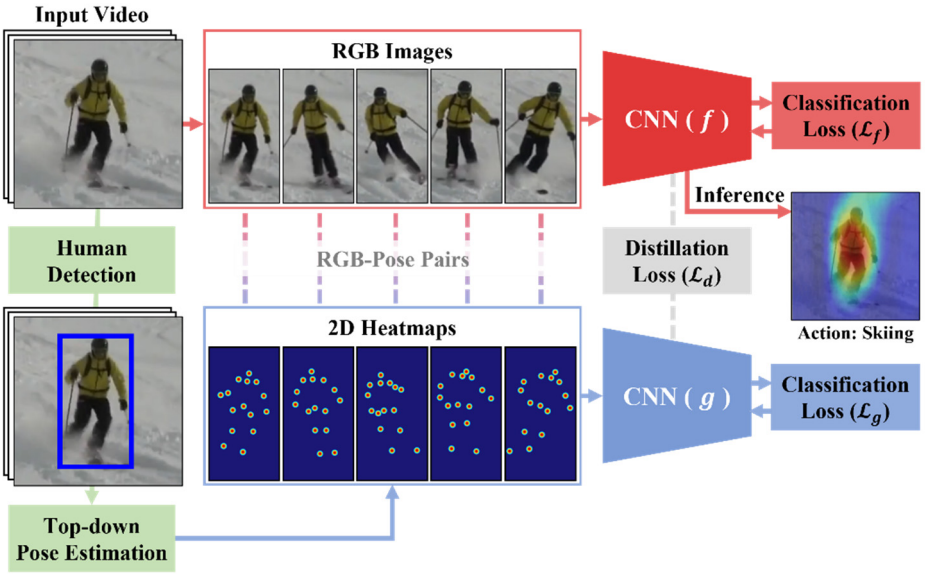


Fig. 1. Overview of the proposed framework.

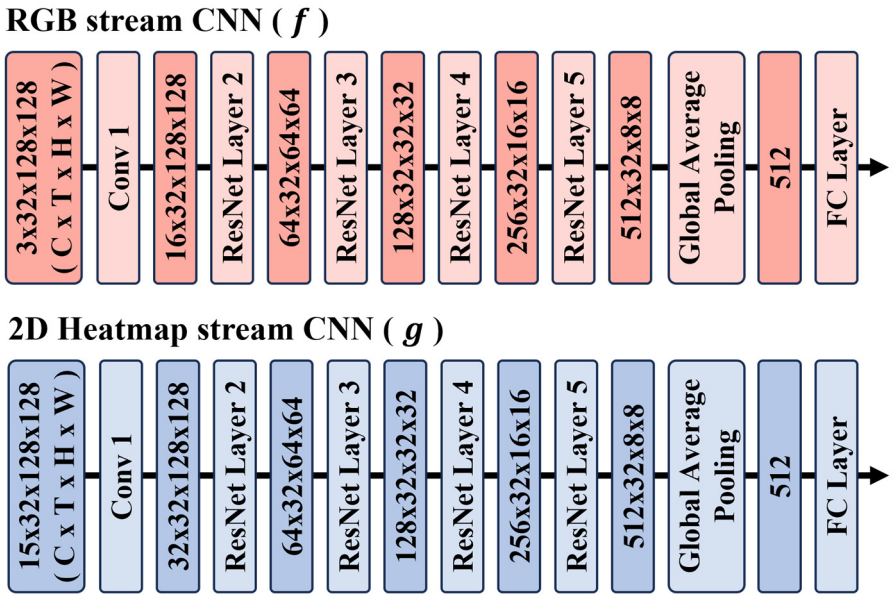


Fig. 2. The detail specification of main and auxiliary network architectures.

primarily. Lastly, both main and auxiliary networks are simultaneously optimized by multiple losses.

3.1 Knowledge distillation

Knowledge distillation was typically introduced for two purposes: training smaller networks by matching the representations of deeper ones and transferring knowledge from pre-trained networks to other networks. In this study, we use the knowledge distillation to infuse human pose information into the main network. To accomplish the successful knowledge distillation, we define the distillation loss $\mathcal{L}_d$ for transferring human pose information to the main network as following Equation 1.

$$\mathcal{L}_d(f, g) = \frac{1}{\ell}\sum_{i=2}^{\ell} \|f_i - g_i\|_2 \quad \textbf{(1)}$$

where ℓ is number of layers, f_i and g_i are activations of i-th layer in the main and auxiliary networks, respectively. The distillation loss $\mathcal{L}_d$ is the average L_2 difference between the activations between layers of both networks. Such difference equalizes the activations of main and auxiliary networks, infusing human pose information into the main network.

We adopt multi-task learning to simultaneously optimize the main and auxiliary networks based on the distillation loss $\mathcal{L}_d$. While it is possible to optimize the main and auxiliary networks based on the single task of the main network in end-to-end manner, during training, there is a risk of losing human pose information relevant to

human actions in the auxiliary network. To establish the separate optimization, we infer the human action recognition results from each network and calculate classification losses $\mathcal{L}_f$ and $\mathcal{L}_g$, respectively. Afterwards, we optimize each network based on the integrated loss $\mathcal{L}_i$, which combines its classification loss and the distillation loss. The integrated loss $\mathcal{L}_i$ is defined as following Equations 2 and 3.

$$\mathcal{L}_i(f) = \lambda_1 \mathcal{L}_f + (1 - \lambda_1)\mathcal{L}_d(f, g) \quad \textbf{(2)}$$

$$\mathcal{L}_i(g) = \lambda_2 \mathcal{L}_g + (1 - \lambda_2)\mathcal{L}_d(f, g) \quad \textbf{(3)}$$

where λ_1 and λ_2 denote weights of main and auxiliary networks, respectively.

3.2 Implementation

In this subsection, we describe the implementation of the proposed framework for human action recognition. We utilizes HRNet [24] pre-trained on COCO-keypoint [25] as the human 2D pose estimator. The output of HRNet is set to [15x128x128], where the dimensions represent the number of heatmaps, height, and width, respectively. The network architectures are specified in Fig. 2. However, the architecture can be changed according to the distillation strategy, as shown in Fig. 3. There are two distillation strategies, full and late distillations. The full distillation strategy calculates the average distillation loss between all layers of the main and auxiliary networks, except for their first convolutional layer. On the other hand, the late distillation strategy only calculates the distillation loss between the last ResNet layers of the main and auxiliary networks. While the full distillation strategy tightly connects the main and auxiliary networks, it also increases the complexity of the framework. Conversely, the late distillation strategy reduces the complexity of the framework, but it may not appropriately infuse human pose information into the main network.

To train and evaluate the proposed framework, the sequence length T is set to 32, which is mainly employed in existing human action recognition methods. The batch size is set to 32. We use Adam optimizer to optimize our framework. The initial learning rate is set to 0.1 and decreases by 0.1 times when the action recognition accuracy saturates. The training epoch is set to 100 in the UCF101 and Kinetics-400 datasets and 50 in the Something-Something V2 (SS-V2) dataset. The weights of the main and auxiliary networks, λ_1 and λ_2, are set to 0.5 and 0.6, respectively.

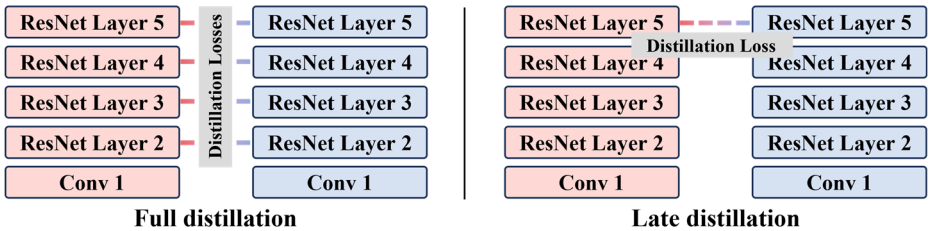


Fig. 3. The network architectures of two different distillation strategies.

4 Experiments

In this section, we present the experimental results of the proposed framework. We first describe the experimental settings and then conduct an ablation study on the effectiveness of knowledge distillation. Finally, we demonstrate the superiority of the proposed framework compared to the existing human action recognition methods and visualize human action recognition results to further analyze our framework.

4.1 Experimental settings

We evaluate the proposed framework using the three video benchmark datasets. For Kinetics-400 dataset, we utilizes 240,618 and 19,404 videos as the training and validation sets, respectively. We randomly partition the UCF101 dataset into training and validation sets. The training and validation sets consist of 9,324 and 3,996 videos, respectively. For the Something-Something V2 (SS-V2) dataset, we utilizes 168,913 and 24,777 videos as training and validation sets, respectively. Following common practice, the shorter side of video resolution is resized to 256. In addition, we randomly crop the videos to 224x224 and apply left-right flipping for data augmentation. For testing, the shorter side of video resolution is also resized to 256 and center-cropped to 224x224.

To measure the human action recognition accuracy, we adopt the top-1 accuracy, which corresponds to the classification accuracy. The top-1 accuracy measures how accurately human actions are recognized. Based on the top-1 accuracy, we evaluate the proposed framework compared with existing action recognition methods, C3D, I3D, SlowOnly, R(2+1)D, and TSM.

4.2 Experimental results

We first conduct an ablation study on the proposed framework to analyze the effectiveness of knowledge distillation. We compare the top-1 accuracy, floating point operations (FLOPs), and the number of parameters for the RGB-only and the two distillation strategies, late and full distillation, architectures under the same hyperparameter setting on the Kinetics-400 dataset. Table 1 presents the result of the ablation study. The architecture adopting the full distillation strategy achieved a top-1 accuracy that is approximately 2.2% to 10.6% better than the other architectures. The result indicates that distilling human pose information can significantly improve the performance of human action recognition. It is worth noting that the full distillation strategy achieved significant improvement in human action recognition performance with only a small difference in complexity compared to the late distillation strategy.

Table 1. The result of an ablation study on knowledge distillation.

Architecture	Top-1 accuracy (%)	FLOPs	Parameters
RGB-only	65.2	13.6 G	3.4 M
Late distillation	73.6	26.8 G	8.0 M
Full distillation	75.8	26.3 G	8.0 M

Table 2. The result of comparative experiments on three video benchmark datasets.

Method	Dataset		
	Kinetics-400	**UCF101**	**SS-V2**
C3D [13]	58.1	85.2	40.8
I3D [14]	72.8	98.0	58.6
SlowOnly [15]	74.9	98.1	61.3
R(2+1)D [21]	73.9	96.8	42.4
TSM [22]	74.7	96.0	63.4
Ours (full)	75.8	98.6	64.6

Next, we conduct comparative experiments on three video benchmark datasets. Table 2 demonstrates the results of comparative experiments. Our framework achieved better top-1 accuracy on all video benchmark datasets compared to existing human action recognition methods. Specifically, our framework achieved an average of 3.2% better top-1 accuracy compared to I3D, and an average of 1.6% better top-1 accuracy compared to TSM. These results imply that the proposed framework effectively learns motion information for human action recognition.

Fig. 4 visualizes the class activation maps of baseline (I3D) and our framework adopting each distillation strategy with GradCAM [26]. As shown in Fig. 4, I3D often struggles to find action-salient regions. Conversely, our framework successfully finds out the action-salient regions. It is worth noting that the full distillation strategy identified action-salient regions more clearly than the late distillation strategy. These results suggest that distilling human pose information enables the learning of powerful motion information for human action recognition and modeling the human-centric motion information.

Fig. 5 shows the human action recognition results for ‘skiing’, ‘vault’, ‘snowboarding’, and ‘frisbee’ videos. As shown in Fig. 5, our framework accurately

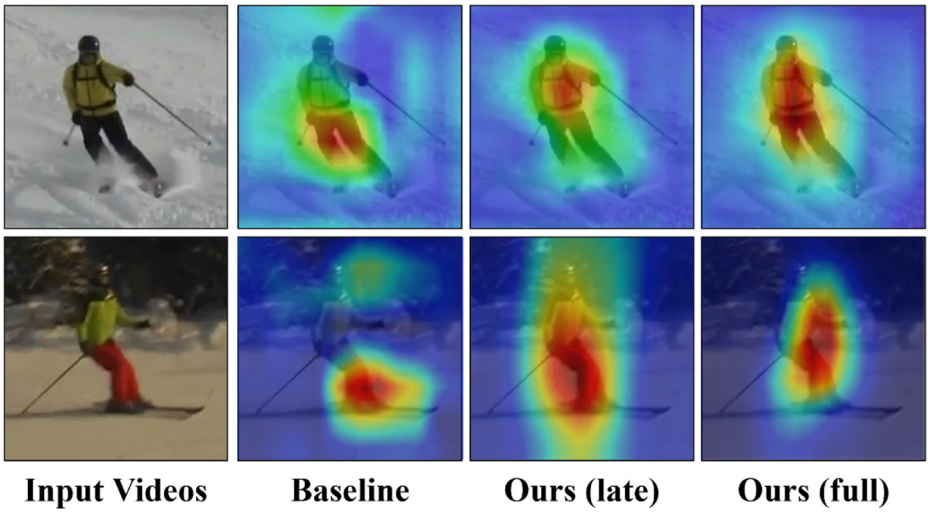


Fig. 4. Visualization of activation maps with GradCAM.

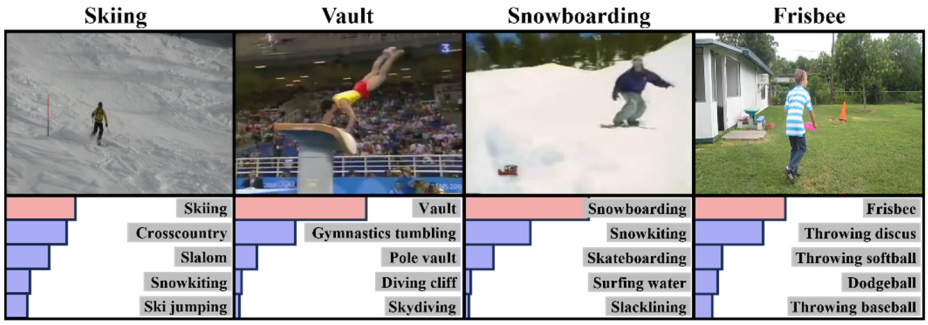


Fig. 5. The classification results for four videos that include different human actions.

recognizes human actions on the four videos. Interestingly, our framework can recognize human actions even when the movements were similar. These results imply that the proposed framework can capture not only human-centric motion information but also interactions between human and objects as well as the environment.

5 Conclusion

In this paper, we proposed a novel multi-task framework that emphasizes human-centric motion information by distilling human pose information. The proposed framework consists of a main network that recognizes human actions from RGB frame sequences and an auxiliary network that distills human pose information into the main network through human 2D pose estimation, enabling the learning of powerful human-centric motion information. The results of comparative experiments under various video benchmark datasets show that the proposed framework outperforms existing human action recognition methods.

Acknowledgments

This work was supported by institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (00167198, AI-PRISM).

References

1. Zeng, K. H., Chen, T. H., Chuang, C. Y., Liao, Y. H., Niebles, J. C., Sun, M.: Leveraging video descriptions to learn video question answering. In: AAAI Conference on Artificial Intelligence. AAAI (2017)
2. Le, T. M., Le, V., Venkatesh, S., Tran, T.: Hierarchical conditional relation networks for video question answering. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9972--9981. IEEE (2020)

3. Xiao, J., Zhou, P., Chua, T. S., Yan, S.: Video graph transformer for video question answering. In: European Conference on Computer Vision, pp. 39--58. Springer Nature (2022)
4. Gabeur, V., Sun, C., Alahari, K., Schmid, C.: Multi-modal transformer for video retrieval. In: 16th European Conference on Computer Vision, pp. 214--229. Springer Nature (2020)
5. Dong, J., Li, X., Xu, C., Yang, X., Yang, G., Wang, X., Wang, M.: Dual encoding for video retrieval by text. IEEE Trans. Pattern Anal. Mach. Intell. 44(8), 4065--4080 (2021)
6. Wang, Z., Wu, Y., Narasimhan, K., Russakovsky, O.: Multi-query video retrieval. In: European Conference on Computer Vision, pp. 233--249. Springer Nature (2022)
7. Zhu, X., Dai, J., Yuan, L., Wei, Y.: Towards high performance video object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 7210--7218. IEEE (2018)
8. Wu, H., Chen, Y., Wang, N., Zhang, Z.: Sequence level semantics aggregation for video object detection. In: IEEE/CVF International Conference on Computer Vision, pp. 9217--9225. IEEE (2019)
9. Jiao, L., Zhang, R., Liu, F., Yang, S., Hou, B., Li, L., Tang, X.: New generation deep learning for video object detection: A survey. IEEE Trans. Neural Netw. Learn. Syst. 33(8), 3195--3215 (2021)
10. Li, Y., Ji, B., Shi, X., Zhang, J., Kang, B., Wang, L.: Tea: Temporal excitation and aggregation for action recognition. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 909--918. IEEE (2020)
11. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1725--1732. IEEE (2014)
12. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition. In: advances in neural information processing systems, pp. 568--576. (2014)
13. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: IEEE International Conference on Computer Vision, pp. 4489--4497. IEEE (2015)
14. Carreira, J., Zisserman, A.: Quo vadis, action recognition? A new model and the kinetics dataset. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6299--6308. IEEE (2017)
15. Feichtenhofer, C., Fan, H., Malik, J., He, K.: SlowFast networks for video recognition. In: IEEE International Conference on Computer Vision, pp. 6202--6211. IEEE (2019)
16. Kim, J. H., Hao, F., Leung, C. K. S., Nasridinov, A.: Cluster-guided temporal modeling for action recognition. Int. J. Multimed. Inf. Retr. 12(2), 15 (2023)
17. Piergiovanni, A. J., Angelova, A., Ryoo, M. S.: Evolving losses for unsupervised video representation learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 133--142. IEEE (2020)
18. Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., Suleyman, M., Zisserman, A.: The kinetics human action video dataset. arXiv preprint arXiv:1705.06950 (2017)
19. Soomro, K., Zamir, A. R., Shah, M.: UCF101: a dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402 (2012)
20. Mahdisoltani, F., Berger, G., Gharbieh, W., Fleet, D., Memisevic, R.: Fine-grained video classification and captioning. arXiv preprint arXiv:1804.09235 (2018)
21. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A close look at spatiotemporal convolutions for action recognition. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6450--6459. IEEE (2018)
22. Lin, J., Gan, C., Han, S.: TSM: temporal shift module for efficient video understanding. In: IEEE/CVF International Conference on Computer Vision, pp. 7083--7093. IEEE (2019)
23. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1--9. IEEE (2015)

24. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5693--5703. IEEE (2019)
25. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Lawrence Zitnick, C.: Microsoft coco: Common objects in context. In: European Conference on Computer Vision, pp. 740--755. Springer Nature (2014)
26. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2921--2929. IEEE (2016)

Enhancing IoT Network Intrusion Detection Model through Over-sampling Techniques

Eun-Beom Sung[1], Sung-Jin Im[2], Jin-Soo Kim[1], and Kwan-Hee Yoo[3]

[1,2] GNSOFT, 56, Munji-ro 299beon-gil, Yuseong-gu, Daejeon, Republic of Korea
[2,3] Chungbuk National University Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Republic of Korea
[1 2]{takoptak, isj}@gn-soft.co.kr
[3] khyoo@chungbuk.ac.kr

Abstract. There has been a growing trend in the adoption and integration of IoT devices into various industries and everyday life. However, along with the benefits, the rapid expansion of IoT has also raised concerns regarding security. Therefore, machine learning based intrusion detection system has been increasingly used. To improve the performance of detection model, this paper provides a comprehensive review of the problem of imbalanced data and investigates various approaches to address it by experimental results.

Keywords: IoT, Machine Learning, SMOTE, ADASYN, Borderline-SMOTE

1 Introduction

IoT stands for Internet of Things, which refers to the network of physical devices, vehicles, appliances, and other objects embedded with sensors, software, and connectivity, enabling them to connect and exchange data. The proliferation of IoT devices has brought about significant changes in sectors such as smart factory or smart home. However, along with the benefits, the rapid expansion of IoT has also raised concerns regarding security. With a large number of interconnected devices, there is an increased risk of cyberattacks and data breaches. For Example, in a smart factory system, numerous IoT devices are connected to the network, and these devices can become sources of DDoS attacks.

Ensuring the security of IoT systems has become a critical focus. Therefore, the combination of IoT security and machine learning-based intrusion detection models has been essential in safeguarding to IoT ecosystems and ensuring the reliability of connected devices.

To improve the performance of intrusion detection model, our research focuses on resolving imbalance data problem by employing an appropriate method named Over-sampling. Specifically, this paper provides a comprehensive review of the problem of imbalanced data and investigates various approaches to address it by experimental results.

2 Related Research

In this section of Related Research, we will present the Bot-IoT dataset and the research related to the oversampling method used in this paper right below.

2.1 Bot-IoT Dataset

The Bot-IoT dataset was created by designing a realistic network environment in the Cyber Range Lab of UNSW Canberra. The captured pcap files are 69.3GB in size with more than 72,000,000 records.

In this paper, to ease the handling of the dataset, we extracted 5% of the original dataset via the use of select MySQL queries. The extracted 5%, is comprised of 4 files of approximately 1.07 GB total size, and about 3 million records. Although the Bot-IoT dataset was designed and created a realistic network environment, the important point is the Bot-IoT consists of over 99% of Botnet traffic and under 1% of normal traffic samples.[1] This can cause data imbalance problem.

Table 1. The Number of Samples for each Class

Label	Support
Normal	477
DoS	1,650,260
DDoS	1,926,624
Reconnaissance	91,082
Theft	79
Total	3,668,522

2.2 Over-sampling

Over-sampling is a technique used to address the data imbalance problem in machine learning, particularly when dealing with imbalanced datasets where one class has significantly fewer instances than the others. The goal of Over-sampling is to increase the representation of the minority class in the dataset to achieve a more balanced distribution.

There are different approaches to Over-sampling, but the common idea is to artificially increase the number of instances belonging to the minority class. This can be done through duplication or by generating synthetic samples that resemble the characteristics of the existing minority class instances. SMOTE(Synthetic Minority Over-sampling Technique) [2] is a generates synthetic samples by interpolating between neighboring minority class instances. This helps to introduce new and diverse samples into the dataset, reducing the risk of overfitting. ADASYN(Adaptive Synthetic) [3] is an extension of the SMOTE. The ADASYN algorithm works by calculating a density distribution ratio for each minority class sample. Samples with lower densities, indicating higher difficulty in classification, are assigned higher

importance and are more likely to be selected for the generation of synthetic samples. This adaptive process helps to address the issue of overfitting that can occur with traditional Over-sampling techniques. Borderline-SMOTE [4] is an enhanced version of the SMOTE (Synthetic Minority Over-sampling Technique) algorithm that specifically targets the samples at the borderline between different classes. While traditional SMOTE randomly oversamples the minority class, Borderline-SMOTE focuses on increasing the number of samples in the challenging regions where the minority class is near the majority class.

3 Experimental Results

3.1 Performance Evaluation

In this paper, performance evaluation is conducted by using Accuracy, Precision, Recall, F1-Score based on the Confusion.

Table 2. Confusion Matrix

Confusion Matrix		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

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

$$\text{Recall} = TP/(TP+FN) \quad (3)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

(1) Accuracy is a measure of how many correct predictions the model makes out of all the predictions it has mad. It is calculated as the ration of the number of correct predictions to the total number of predictions made.
(2) Precision is crucial when the cost of false positives is high. For example, in medical testing, a high precision means fewer health individuals are wrongly classified as having a disease(false positive).
(3) Recall is a metric that assesses how many of the actual positive instances were correctly identified by the model. It is the ratio of true positive predictions to the total number of actual positive instances(both true positives and false negatives).

(4) F1-score is the harmonic mean of precision and recall. It is a single metric that provides a balance between precision and recall.

3.2 Evaluation Results

Overall, this experiment aimed to demonstrate the impact of Over-sampling techniques, namely SMOTE, ADASYN, and Borderline SMOTE, on the performance of the Random Forest model in an IoT network environment using the Bot-IoT dataset. The results provide insights into the effectiveness of these techniques in addressing data imbalance and improving the classification performance in IoT-related applications.

Table 3. Accuracy Result

Category	Original	SMOTE	ADASYN	Borderline-SMOTE
DDoS	99%	99%	99%	99%
DoS	99%	99%	99%	99%
Normal	96%	99%	99%	99%
Reconnaissance	99%	99%	99%	99%
Theft	92%	99%	99%	99%

Table 4. Precision Result

Category	Original	SMOTE	ADASYN	Borderline-SMOTE
DDoS	99%	99%	99%	99%
DoS	99%	98%	99%	99%
Normal	100%	100%	99%	100%
Reconnaissance	100%	100%	99%	100%
Theft	100%	100%	100%	100%

Table 5. Recall Result

Category	Original	SMOTE	ADASYN	Borderline-SMOTE
DDoS	99%	98%	99%	99%
DoS	99%	99%	99%	99%
Normal	99%	100%	100%	100%
Reconnaissance	100%	100%	100%	100%
Theft	25%	100%	100%	100%

Table 6. F1-score Result

Category	Original	SMOTE	ADASYN	Borderline-SMOTE
DDoS	99%	99%	99%	99%
DoS	99%	99%	99%	99%
Normal	99%	100%	100%	100%
Reconnaissance	100%	100%	100%	100%
Theft	40%	100%	100%	100%

4 Conclusion

In this paper, in order to address the data imbalance issue in Bot-IoT dataset, we performed Over-sampling and created intrusion detection models for various attacks. The performance of these models was evaluated. As a result, we showed that the accuracy of the 'normal' class increased by approximately 3%. Additionally, the accuracy of the 'Theft' attack class increased by around 7%. The ‘Theft’ attack is a significant problem such as in smart factory system. For example, the problem with a ‘Theft’ attack in a smart factory lies in the potential for attackers to gain valuable insights and knowledge that can be exploited for further malicious activities.

As an experimental result, the recall for the 'Theft' attack exhibited a significant increase of about 75%. Recall is a significant evaluation metric in cases where false negatives can have a substantial impact on business operations, such as in diagnostic models for diseases or fraud detection models in finance. Web attack detection falls into this category as well. Although the increase in the recall measure may appear modest, its effect becomes more significant as the volume of malicious traffic increases.

To enhance the performance of intrusion detection models in IoT network environments, addressing the data imbalance issue is crucial. This research aimed to explore methods to alleviate this problem and employed various Over-sampling techniques to achieve meaningful results.

By utilizing different Over-sampling methods, the study effectively tackled the data imbalance challenge. The results demonstrated notable improvements in the performance of the intrusion detection models.

Given that this research was conducted using the Bot-IoT dataset and within a limited environment, further studies exploring performance enhancement strategies are needed. It would be beneficial to apply different datasets and methods to investigate improvements in performance. This would provide a more comprehensive understanding of the effectiveness and generalizability of the proposed approaches.

Acknowledgments. This work was partly supported by the Technology development Program of MSS S3290113 and the ICT development R&D program of MSIT S3290113

References

1. K. Nickolaos, N. Moustafa, E. Sitnikova, & B.Turnbull. (2019). Towards the development of realistic Botnet dataset in the internet of things for network forensic analytics Bot-Iot dataset. Future Generation
2. Fernández, Alberto, et al. "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." Journal of artificial intelligence research 61 (2018): 863-905.
3. He, Haibo, et al. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning." 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). Ieee, 2008.
4. Hui Han, Wen-Yuan Wang, and Bing-Huan Mao. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In International conference on intelligent computing, 878–887. Springer, 2005.
5. Hyun Mi Jin (2021) A comparative study of the performance of machine learning algorithms to detect malicious traffic in IoT networks. Research on Digital Convergence,19(9),463-468.
6. Peterson, Jared M., Joffrey L. Leevy, and Taghi M. Khoshgoftaar. "A review and analysis of the bot-iot dataset." 2021 IEEE International Conference on Service-Oriented System Engineering (SOSE). IEEE, 2021.
7. Leevy, Joffrey L., et al. "An easy-to-classify approach for the bot-iot dataset." 2021 IEEE third international conference on cognitive machine intelligence (CogMI). IEEE, 2021.
8. Zorić, Petra, Mario Musa, and Tibor Mijo Kuljanić. "Smart factory environment: Review of security threats and risks." International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures. Cham: Springer International Publishing, 2021.

The Smart Factory with Variable System Design

Chae-Hyun Lee[1], Sung-Jin Im[1,2], Ja-Yeon Heo[1], Jin-Soo Kim[1] and Kwan-Hee Yoo[2]

[1] GNSOFT, 56, Munji-ro 299beon-gil, Yuseong-gu, Daejeon, Republic of Korea
[2] Chungbuk National University Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, Republic of Korea
[1]{dlcogus0425, isj, jayheo23}@gn-soft.co.kr
[2] Khyoo@chungbuk.ac.kr

Abstract. By proposing a program that can flexibly reflect data such as process, equipment configuration, and environmental variables in real time in a rapidly changing manufacturing factory environment, this paper intend to overcome the maintenance limitations of the program and provide improvement in factory efficiency and productivity.

Keywords: Smart Factory, MongoDB, Dynamic, Real Time

1 Introduction

Modern manufacturing plants face a rapidly changing environment. Advances in technology and business requirements, such as manufacturing processes, equipment configurations, and environmental parameters, are constantly changing. This dynamic environment has required new changes to existing plant management and control systems. Existing manufacturing plant management systems are designed in a static structure, and thus have limitations that make it difficult to respond to a variably changing environment. This causes problems of process and cost according to maintenance and expansion of the system.

Due to the limitations of such a static environment, there is a need to improve the existing system so that it can be flexibly modified for setting and adjusting suitable for the actual factory environment. Therefore, this paper proposes a program that allows the factory administrator to directly set and adjust the process, manufacturing process, and environmental variables through smart factory design tailored to the variable factory environment.

Through this, it is possible to flexibly respond to and manage the changing factory environment in real time, overcome the limitations of program maintenance, and provide opportunities to improve factory efficiency and productivity.

2 Related Research

Smart Factory. It is a factory that aims to produce products at the minimum cost and time by integrating the entire process from product planning to design, distribution, and sales by combining ICT with the manufacturing industry. The production resources of a smart factory are Man, Machine, Material, and Method, and it is called 4M1E including Environment. The definition of 4M1E is shown in Table 1 [1].

Table 1. Smart Factory Production Resource 4M1E Definition of Terms

4M1E	Glossary	definition
4M	Man	Workers performing work or running equipment inside and outside the workplace
	Machine	Various facilities installed to produce products
	Material	Various raw materials (raw materials, parts) required to produce products
	Method	Working standards and working conditions necessary to produce products using production resources
1E	Environment	Environmental information related to production

MongoDB. It is used in variable smart factories, is called NoSQL, which means that it does not use SQL. It operates without schema and does not require definition or change of structure. It has the advantage of facilitating distributed expansion, and there is no fixed form of schema as documents and key values are not defined in advance. Since there is no fixed schema, fields can be added or removed as needed [2]. It is easy to implement a variable smart factory because documents and key values are not defined in advance.

3 Variable smart factory technology and design

The variable smart factory is composed of a universally usable menu that is not tailored to a specific factory in order to allow users to directly build and modify data including production resource (4M1E) information without separate specialized maintenance. Fig 1 is the menu of the variable smart factory for the administrator, and it is possible to directly set the management system, design, input system, and environment setting. Since the administrator has database access and modification rights, it is possible to respond to variable changes in the factory with the administrator account.

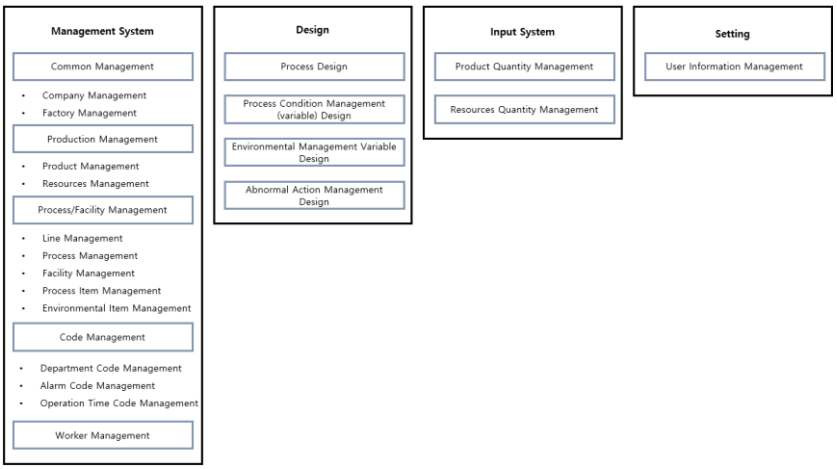


Fig. 1. Structure Diagram of Variable Smart Factory Menu for the Administrator

In Fig 2, it is possible to monitor factory facilities, materials, products, etc. with a user account, and the authority to modify the DB is not separately granted.

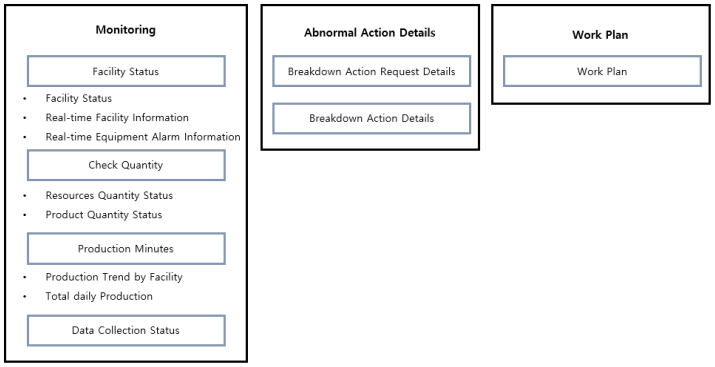


Fig. 2. Structure Diagram of Variable Smart Factory Menu for the User

The core of the variable smart factory is that the administrator can access the database and reflect the actual factory situation to the smart factory. In order to put data into a smart factory, data must be designed first. The administrator accesses the data design page of the DB to be built, adds the type and name of the data to be put in, and saves it. When saved in the data design page, it is reflected in the DB, and the design is completed so that the designed items can be viewed in the management page. Items can be modified or deleted on the design page, and are reflected in the database according to the saving of the design page, which can be checked on the management page. If this is shown in a flowchart, it is shown in Fig 3.

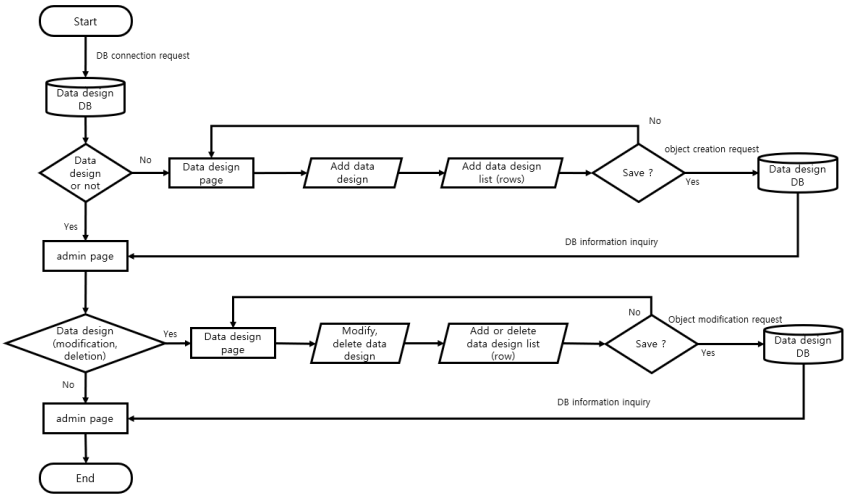


Fig. 3. Variable Smart Factory Data Design and Management Flow Chart

A variable smart factory using MongoDB has several Collections in the database as shown in Fig 4. It can be subdivided according to function, and several Documents can be additionally created in one Collection. When designing factory data in a variable smart factory, designed items are created in Company Documents in Structures Collection, and when data is entered, they are created in the format and variable name set in Company Documents in Documents Collection.

In other words, by connecting the Structures Collection and the Documents Collection internally, the administrator can create and manage the database for the smart factory on the web page without directly accessing the database.

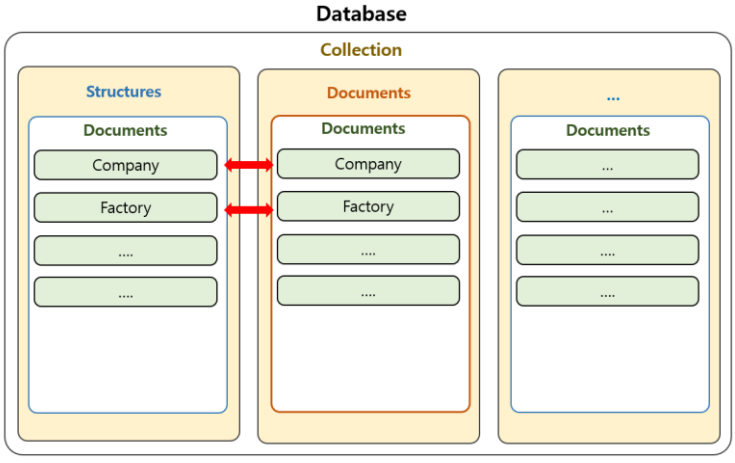


Fig. 4. MongoDB Structure

4 Experiments and Considerations

Fig 5 is a page for setting the data and form required in a smart factory. It is possible to check that DB variables that will be items on the management page can be added with the add button. Additional database management is possible by deleting or modifying on the current page.

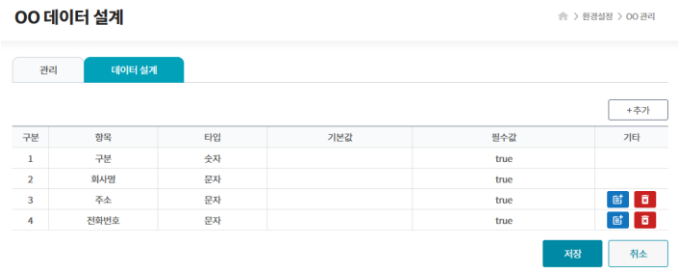


Fig. 5. Data Design Web Page

Fig 6 is a page created by adding sample data. It can be seen that the variables created in the data design appear in the items of the management page, and the data can be added according to the items.

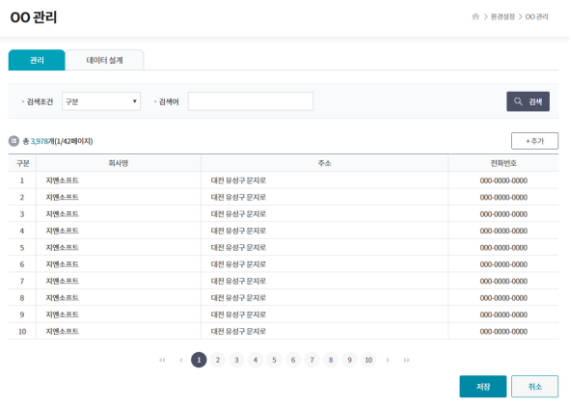


Fig. 6. Management Web Page

When data is entered on the management page after data design, it is reflected as shown in MongoDB Fig 7. Seq, Name, test, test1, and boo fields of the Structures Company Documents are created as Objects in Documents Company Documents, and data is entered in the form of JSON.

Through this, it was confirmed that two Documents are connected to enable design the data and input it on a web page.

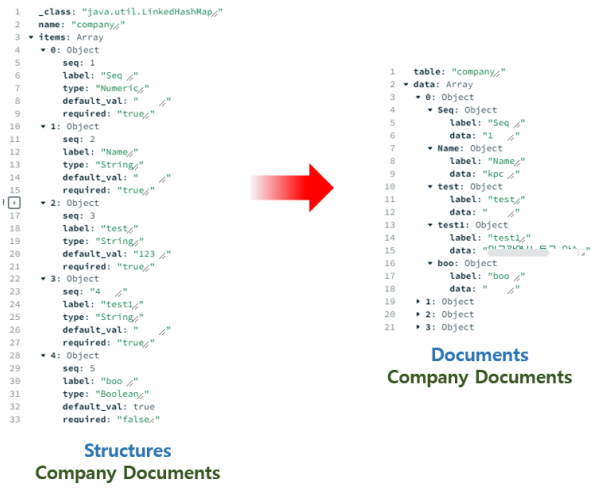


Fig. 7. MongoDB Data Design, Documents Change After Input

5 Conclusion

In this paper, it was confirmed whether the situation of the factory can be reflected in the smart factory in real time to the variable factory without expert maintenance according to the variable factory situation. It is possible for managers to directly access databases through web pages to build and manage factories in real time. In addition, it is expected that there will be no maintenance difficulties by providing a universally usable manual so that it can be distributed regardless of the type of factory, so that no separate maintenance is required when distributing to multiple factories. In addition, it is expected that it will be easily accessible by implementing it as a web page through a variable smart factory design, and in the future, the cloud environment and ERP/MES data will be linked to provide services so that each company can use a variable smart factory.

Acknowledgments. This work was partly supported by the Technology development Program of MSS S3290113 and the ICT development R&D program of MSIT S3290113

References

1. TTA, Information Management of Smart Factory based on Production Resources (4M1E) – Part 1 : Reference Model, Telecommunications Technology Association, (2017)
2. Hyun-Joo Kim, “Web Service Data Processing Using Database Mongo”, The Journal of the Korea Academia-Industrial Cooperation Society, pp. 233-236, (2014)

3. Yong-Hyun Kim, Eui-Nam Huh, “Dynamic Big Data Analytics System Architecture on Hadoop and MongoDB”, Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp. 128-129, (2014)
4. Wan, Jiafu, et al. “Artificial intelligence for cloud-assisted smart factory.” IEEE Access 6 (2018): 55419-55430.
5. Nguyen, Huy Toan, et al. “Deep learning-based defective product classification system for smart factory.” The 9th International Conference on Smart Media and Applications. (2020)

Automated Quality Control of Dried Peppers : Image Preprocessing and Classification using Deep Learning

Ki-Tae Park[1], Woo-Seok Choi[1], Sang-Hyun Choi[2*]

[1]Dept. Bigdata, Chungbuk National University, Cheongju, South Korea
[2]Dept. Management Information System, Chungbuk National University, Cheongju, South Korea
{chois}@cbnu.ac.kr, {rlxogustn, cdt3017}@naver.com

Abstract. This study aimed to automate the quality control of dried peppers for imported frozen peppers. Using an image classification model based on deep learning and the EfficientNetB7 architecture, a dataset of 2,362 dried pepper images was classified into "Normals" and "Abnormals" categories. Preprocessing techniques such as the contour algorithm and image crop were employed to enhance the model's performance. The results showed that the crop image dataset achieved an accuracy of 97.8%, outperforming the original dataset's accuracy of 76.9%. The study concluded that higher area representation of dried peppers in images led to better image classification performance.

Keywords: Dried Peppers, Image Classification, Image Preprocessing, Quality Control, DL, Deep Learning

1 Introduction

The global imports of frozen peppers have been witnessing a notable upsurge. The cost advantage of imported frozen peppers over domestically grown ones and the attractive low tariff rates in many countries have made them a preferred choice [1]. In Korea, for instance, the import volume reached 105,622 tons in 2015, and the demand for imported frozen peppers continues to grow [1]. China, Mexico, Turkiye, Indonesia, and the United States are among the major pepper producers worldwide, with China alone accounting for 51.3% of global pepper production in 2012 [2]. However, during the drying process of these imported frozen peppers, mold formation and decomposition are common issues [3]. As dried peppers are extensively used as raw materials in various food products, ensuring quality control throughout the drying process is crucial.

In order to analyze the decay of imported dried peppers, Based on ICA(Independent Component Analysis), there is a study that suggests a method of selecting peppers that have changed color due to decomposition during the pepper drying process or have no major ingredients such as capsaicin[4], In addition, there is research on analyzing the size of dried peppers and recognizing colors based on artificial neural networks [5]. Furthermore, there is a study that analyzes the process of changing the freshness of food

* Corresponding author

using image classification model based on the Transfer learning for deep learning for automation of quality control [6].

In light of these findings, this study aims to manage the quality of frozen peppers imported from China by analyzing the state of imported dried peppers based on "Normals(edible)" or "Abnormals"(decayed) using an image classification model based on deep learning. Our goal is to propose methods for automating the quality management of dried peppers. By doing so, we hope to contribute to the enhancement of overall food safety and quality standards in the importation of frozen peppers.

2 Data preprocessing

In this study, dried peppers were stored in an experimental bag, and a random selection of 30 to 50 dried peppers was taken out and photographed once a week. To ensure the randomness of the study, the location and direction of the dried peppers were adjusted for each photography session, and rotation was applied to capture various angles. he collected images were then categorized into two labels: "Normals(edible)" and "Abnormals" (decayed). Table 1 shows that during the 32nd week, a total of 2,362 image data was collected, consisting of 1,407 "Normals" and 955 "Abnormals" image data.

Table 1. About Dried Red Pepper Image Data Collection.

No	Date	Normals	Abnormals
1	2022.10.06	48	46
2	2022.10.14	48	41
3	2022.10.21	46	42
		⋮	
32	2023.03.30	40	48

To focus on the color (RGB) of the dried peppers as a key feature, this study utilized the original image data without image preprocessing. However, it was observed that the area ratio of dried peppers in the original images was only 21%. To better represent the color of the dried peppers, all images were cropped to increase the area ratio of the dried peppers to 31%. This was achieved by cropping the images to emphasize the area occupied by the dried peppers. The OpenCV's Contour Algorithm was employed to detect the outline of the dried pepper objects and calculate their areas, as shown in Fig. 2.

Fig. 1 provides a sample of the cropped image, while Fig. 2 illustrates the process of obtaining the area of each dried pepper on the image through contour detection.

Fig 1. Crop image Sample.

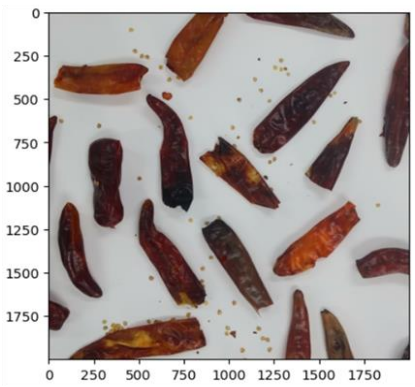
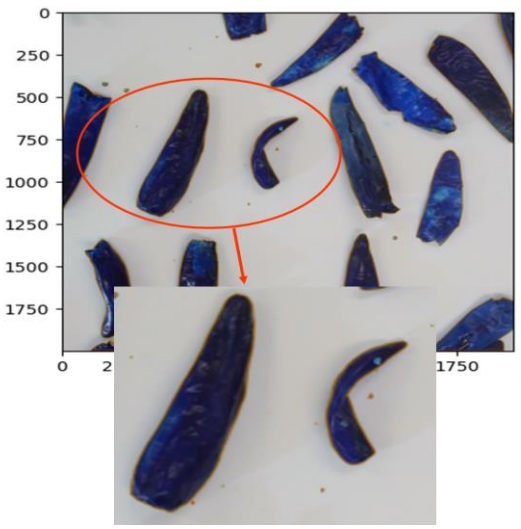


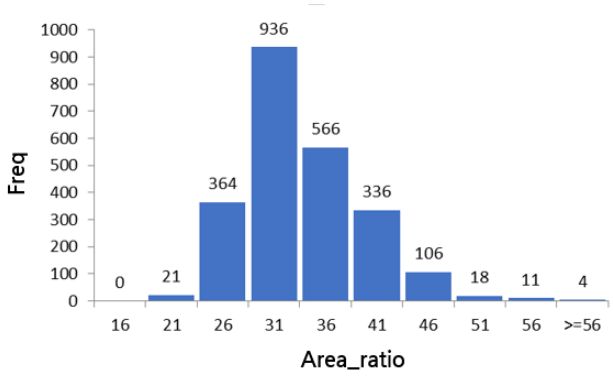
Fig 2. Area of dried peppers by contour detection.



The distribution of the dried pepper area for each cropped image is depicted in Fig. 3. "Area_ratio" represents the area of all dried peppers in each image, and "Freq" indicates the frequency of the respective area ratios. The most frequent area ratio was in the third class, with a maximum area ratio of 59%, a minimum area ratio of 18%, and an average of 31%. These statistics provide insights into the distribution of dried pepper areas, which will be valuable for the subsequent image classification and analysis tasks.

The rigorous data preprocessing steps, including random sampling, rotation, and contour-based cropping, ensure the dataset's consistency and improve the accuracy and reliability of the image classification model used in this stud.

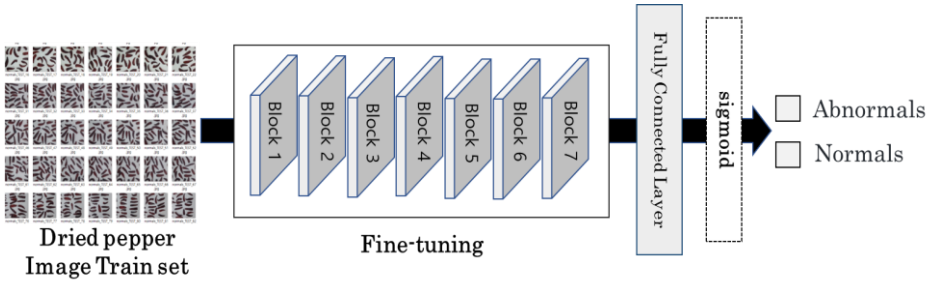
Fig 3. Histogram for dried pepper area distribution.



3 Analysis

In this study, a dataset of 2,362 crop images of dried peppers was used for binary classification to predict whether each dried pepper belonged to the "Normals" or "Abnormals" class. we fine-tuned the EfficientNet model, which was developed by Google Brain in 2019. EfficientNet is an image classification model that employs compound scaling, resulting in higher performance with fewer parameters compared to previous models used in Image Classification Tasks. Among the EfficientNet variants (B0 to B7), B7 is known for its superior performance. Hence, for this study, we utilized the pre-trained weights of the EfficientNetB7 model and fine-tuned the fully connected layer to suit the specific objectives of this research.

Fig 4. Classification model Fine-tuning Process.



The dataset of 2,362 crop images was divided into training and validation sets, with 80% (1,890 crop images) used for model training and 20% (472 crop images) used for model verification. For binary classification, the Fully Connected Layer employed the

Sigmoid activation function. The learning rate was set to 0.001, and the momentum was set to 0.9. The training and validation batch_size consisted of 16, and the training process was carried out for 30 epochs. However, the training was set to terminate early if the model's performance no longer improved, to optimize the learning process. The performance of model verification was evaluated by “Accuracy”. Python (3.7 Ver), Keras for Framework, and GPU of GeForce RTX2060 were used as development environments for learning and verification.

The model's verification involved comparing the results obtained using the original dried pepper image dataset with those from the preprocessed crop image dataset. The accuracy achieved with the original dataset was **76.9%**, while the crop image dataset resulted in significantly higher accuracy of **97.8%**.

Table 2. Comparison of validation datasets.

Index	Original dataset	Crop image dataset
Accuracy	76.9%	97.8%

The original dataset had an average area ratio of dried peppers on the image of 21%, about 10% lower than the Crop image dataset, and had more noise than the Crop image. The Crop image data set used the contour algorithm to crop the dry pepper area to be expressed more in the image, and the area ratio of dry peppers was 31% on average. As a result, the more dried peppers are reflected in the image, the higher the performance of the image classification model and the higher the stability of the model.

4 Conclusion and Future work

In this study, 2362 dried pepper images were collected and the “Normals” and “Abnormals” of dried peppers were classified to automate the quality control of imported frozen peppers that have been dried. As a result of model training and verification, the accuracy of the original data set without preprocessing was 76.9%, and the accuracy of the Crop image data set with preprocessing was 97.8%. The 97.8% accuracy was a significant enough result to automatically classify the quality of dried peppers. The contour algorithm and Image crop pretreatment method were able to reflect the feature of dried peppers and had a positive effect on improving model performance. In conclusion, it was confirmed that the higher the area of the object expressed on the image, the more RGB feature of the object are included, and the more the area of the object is reflected, the better the image classification performance.

In future studies, genetic testing of dried peppers will be conducted to advance the quality control of dried peppers, and research will be conducted on how to more accurately classify the quality of dried peppers into grades A, B, and C by matching the genetic test results with dried pepper images.

References

1. KOREA RURAL ECONOMIC INSTITUTE, Causes and Implications of Dried Red Pepper Industry in Korea (2017)
2. KOREA RURAL ECONOMIC INSTITUTE, World Agriculture (2014)
3. So-soo Kim, Seul Gi Baek, Injun Hwang, Se-Ri Kim, Gyusuck Jung, Eunjung Roh, Ja Yeong Jang, Jeomsoon Kim, Theresa Lee, Fungi of red pepper occurs during the production stage of dried red pepper, Journal of Food Hygiene and Safety Vol. 34 No. 6 pp. 571--575 (2019)
4. Kihyeon Kwo, Jung-Dae Li, Dried pepper sorting using independent component analysis on RGB image, Journal of The Korea Society of Computer and Information, Vol. 17 No. 4, April (2012)
5. O. Cruz-Domínguez, J.L. Carrera-Escobedo, C.H. Guzman-Valdivia, A. Ortiz-Rivera, M. García-Ruiz, H.A. Duran-Munoz, C.A. Vidales-Basurto, V.M. Castano, A novel method for dried chili pepper classification using artificial intelligence, Journal of Agriculture and Food Research, Vol. 3, March (2021)
6. Jiangong Ni, Jiyue Gao, Limiao Deng, Zhongzhi Han, Monitoring the Change Process of Banana Freshness by GoogLeNet, Vol. 8 pp. 228369—228376 IEEE Access, December (2020)

Electric Vehicle Power Load Prediction Using Machine Learning Ensemble Techniques and Charge-Based Derived Variables

Seong-ju Joe[1], Dong-kyu Yun[1], Sang-hyun Choi[2]

[1] Dept. Bigdata, Chungbuk National University, Cheongju, South Korea
[2] Dept. Management Information System, Chungbuk National University, Cheongju, South Korea
{fulans2}@naver.com, {dongkyu.yun, chois}@cbnu.ac.kr

Abstract. Recently, electric vehicles have gained worldwide popularity due to their ability to reduce fossil fuel consumption and promote carbon neutrality. With the rapid growth of the electric vehicle market in Korea, there is a need for research to predict the demand for electric vehicle charging power and develop an effective management system for charging stations to accommodate the expanding business size. To achieve more accurate predictions, this study aims to utilize ensemble-based machine learning algorithms instead of relying on a single existing model. Additionally, a data-driven model of the charging amount at EV charging stations, derived from public data portals, is developed. The study yielded a 90.42% accuracy rate in the MAPE performance evaluation index, confirming its effectiveness.

Keywords: EV, Electricity Demand, Prediction, Machine Learning.

1 Introduction

Recently, electric vehicles have gained worldwide popularity as a means to reduce fossil fuel consumption and achieve carbon neutrality. In Korea, industrial power consumption has consistently accounted for over 50% of the total energy usage in the past decade. With continued government support, the adoption of electric and eco-friendly vehicles is expected to expand further. As the electric vehicle market grows rapidly, research efforts are actively underway to predict the demand for electric vehicle charging power. This is essential for developing an effective management system for charging stations capable of accommodating the increasing business size.

2 Data & Analytics

In this study, data on EV charging usage in Buk-gu, Gwangju, was obtained from the public data portal, while weather forecast data was sourced from the Korea Meteorological Administration. The data collection period spanned from December

31, 2019, to March 3, 2022, with each data point representing a unit of charge. The researcher generated two derived variables based on the charging date. Firstly, to account for date-related variations, variables were set to indicate holidays and weekends/weekdays. On holidays or alternative holidays, the variable was set to 1 or 0, respectively, and a variable of 1 for weekends and 0 for weekdays was created. Additionally, a derivative variable was generated based on the dependent variable, the charging amount. It represented the average charging amount for the corresponding month-hour, and the average charging time for the month-hour. In addition, in order to remove outliers and missing values of the data, it was removed when the charging amount was less than 1 kwh and the charging time was less than 1 minute. Additionally, to enable hourly predictions, cases that overlapped at the same time were aggregated. In the case of electric vehicles, the driving distance tends to decrease slightly in winter compared to summer due to increased battery resistance caused by low temperatures. As a result, the amount of electric vehicle charging is also influenced by temperature. Therefore, temperature was included as an independent variable. The final dataset consisted of 10,334 observations, and the independent variables used in the model were month, hour, week, holiday, temperature, precipitation, kWh_y, and C_time_sum_y. The dependent variable, kWh, exhibited differences in charge amounts depending on the 'year', so data scaling was performed through standardization. Finally, for prediction purposes, three algorithms—Random Forest, XGBoost, and LightGBM—were selected as ensemble algorithms based on decision trees. These algorithms were employed for AI model learning, and their performances were compared.

3 Conclusion

Table 1. Electric charge prediction accuracy comparison by dataset and model.

Modelt / dataset	RandomForest	XGBoost	LightGBM	Average (%)
Set1	88.31 %	86.40 %	79.51 %	84.74 %
Set2	92.53 %	89.35 %	85.49 %	89.12 %
Average (%)	90.42 %	87.87 %	82.50 %	86.93 %

Table 1 presents the performance test results for three algorithms using the dataset. Set1 refers to the dataset that excludes kWh_y and C_time_sum_y, which are derived variables based on the charge amount. Set2, on the other hand, includes kWh_y and C_time_sum_y. The study revealed that the models performed better when the derivatives generated from the charge amount were excluded. Notably, Random Forest demonstrated the highest accuracy at 90.42%.

The study involved generating and utilizing several derivatives related to electric vehicle charging for prediction purposes, and it confirmed high accuracy. These results enable charging stations to operate stably by proactively preparing for demand. Additionally, accurate predictions of charging volume help evaluate the feasibility of businesses and prevent the indiscriminate installation of charging stations.

Analysis of Distribution of Fishing Vessels Using AIS Data

Eun A Song, Eun Ju Jeong, Kwang Il Kim

College of Ocean Sciences, Jeju National University, Jeju 63243, Korea
c0209@jejunu.ac.kr, zkrkdi01@gmail.com, kki@jejunu.ac.kr

Abstract. It is important to check the distribution of resources for sustainable fisheries resource management. Therefore, in this study, as a basic study to identify fishery resources, the fishing boat's fishing grounds are analyzed using the ship's AIS data. In future studies, we would like to analyze the movement of fishery resources by combining the distribution of fishing vessel's fishing grounds and the analysis of fisheries resource catch.

Keywords: Fishing ground, AIS data, Fishing vessel

1 Introduction

Given the changes occurring in the marine environment, such as increasing pollution and rising temperatures, impacts on habitats, abundance and availability of fisheries resource continue to pose challenges to the sustainability. It is therefore crucial to identify and manage resources to avoid depletion and total collapse. However, determining the precise location of fishing grounds solely through fishermen's reports poses challenges due to concerns related to personal information and security. Consequently, this study sought to address this issue by utilizing the Automatic Identification System (AIS) data from fishing vessels to analyze the real-time fishing locations and timing of fishing activities.

2 Distribution Paper Using AIS Data

The AIS data comprises essential information about the ship, including its name, location (latitude and longitude), speed, and course. For this study, we focused on the research period from January 1, 2018, to December 31, 2019, with the target waters confined to the geographical region ranging from 29°00' to 35°00' N and 124°00' to 129°00' E. To facilitate the analysis of fishing activities in this region, we divided the area into grids based on 30' intervals in both the upper and lower longitude. Furthermore, for a more detailed examination of fishing operations within the same sea area, we additionally subdivided the latitude and longitude into a grid with 15' intervals. This study utilized the ship's speed as a key indicator to determine its operational status. If the ship's speed fell within the specified standard, it was

classified as actively fishing in the designated fishing ground. The duration of time when fishing vessels maintained the reference speed for their operations was recorded. By analyzing the latitude and longitude data collected through AIS, we were able to pinpoint the exact fishing locations. Subsequently, the operating time of each fishing vessel was assigned to the corresponding grid, allowing for a comprehensive mapping of fishing activities within the study area.

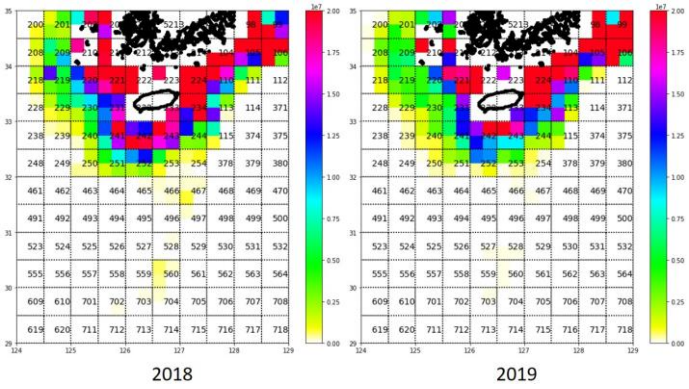


Fig. 1. Result of Fishing Distribution Analysis of Fishing Vessels.

3 Conclusions

During the entire periods of 2018 and 2019, a noticeable decline in the frequency of fishing operations was observed in the western waters of Jeju Island in 2019 as compared to 2018. Specifically, when comparing the period from January to March 2018 to the same period in 2019, a decrease in fishing activities was evident in the 220 and 221 fishing grids located northwest of Jeju Island. Conversely, during the same period in 2019, an increase in fishing activities was observed in the 243 and 252 fishing grids situated south of Jeju Island. For future studies, we intend to delve into the analysis of fishery resources' movement by examining the distribution of fishing grounds of vessels and the catch of fishery resources.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1C1C1013773). and was a part of the project titled, “Development of AI Based Smart Fisheries Management System” which is funded by the Ministry of Oceans and Fisheries, Korea.

References

1. Owiredu, S. A., & Kim, K. I. (2021). Spatio-Temporal Fish Catch Assessments Using Fishing Vessel Trajectories and Coastal Fish Landing Data from around Jeju Island. *Sustainability*, *13*(24), 13841.

Estimation of Spatial and Temporal Impacts of Commercial Fishing Using Catch and Vessel Mobility Data

Solomon A. Owiredu, Shem Otoi Onyango and Kwang Il Kim

College of Ocean Sciences, Jeju National University, Jeju 63243, Korea
saowiredu@jejunu.ac.kr, shemotoi@jkuat.ac.ke, kki@jejunu.ac.kr

Abstract. Applications of mobility data in marine research has provided deeper insights into vessel activity and their impacts on the marine environment. The use of tracking devices to monitor spatial and temporal dynamics of fishing vessels has become critical in marine fisheries research in recent times. Declines in global fishery productivity has been attributed to vessel overcapacity resulting in excessive overfishing and management measures that have yet to address these impacts in a changing environment. Investigating the linkages between spatial and temporal distribution of fisheries resources and vessel activity is necessary in estimating the extent of fishing impact on marine ecosystems. In our study, using a data fusion approach, we combined AIS and fish catch datasets of commercial fishing vessels that operate in the waters around Jeju Island. We proposed a method of allocating catch amounts to fishing segments of trajectories by reconstructing trajectories into fishing and non-fishing activities using vessel speed profiles. We produced spatio-temporal distributions of catch, vessel activity and reliance on fishing grounds and discussed opportunities of combining larger datasets collected over longer periods to provide estimates and reference points that can inform sustainable resource management decisions on a local and regional scale.

Keywords: commercial fish stocks, fish population declines, automatic identification system (AIS), overfishing.

1 Introduction

The assessment of fishing effort intensities on marine ecosystems through the spatial and temporal monitoring of fishing vessel activity is very important in understanding the dynamics of the fish populations and distributions, migratory patterns, and fishing fleet behaviour. Fisheries investigations have previously used a variety of different tracking devices to monitor the spatio-temporal dynamics of fishing vessels and to quantify fishing efforts and intensities in the fishing grounds and fisheries resources therein. In recent times, these assessments have been enhanced using vessel monitoring system (VMS) and automatic identification system (AIS) data which provide fishing vessel trip data with high patio-temporal resolution using the VHF radio frequency band [1]. However, AIS data is characterized by high persistence, hence a better a better choice [2]. In addition, the availability of satellite AIS (S-AIS) data from a growing

number of satellite-based data providers [3] has made it readily available for fisheries resource assessments. AIS was originally intended to help improve ship safety and to transmit at high frequencies to avoid ship collisions. However, its expansion in recent years has increased its popularity in academic research and has now been practically applied in various disciplines, including its application in addressing resource management challenges as it provides data on a scale that is critical to ensure effective assessment of marine fishery resources. Most studies that have used fishing vessel trajectory data have focused on determining the fishing gear type or fishing activity status. It is very difficult, however, to identify the quantity of fish caught. Therefore, we used a data fusion approach by combining fish landing and fishing vessel trajectory data of coastal and offshore fishing fisheries and proposed a method to discriminate vessel activity into port entry and exit, fishing, and navigation.

2 System Model and Methods

2.1. Study Area and Fishing Vessel Trajectory and Fish Landing Data Aggregation

In the waters surrounding the island, various water masses, such as Tsushima Middle Water, Tsushima Surface water, Yellow Sea Bottom Cold water, and the China Coastal water, merge, depending on the season. Due to the characteristics of these water masses, many fish species have altered when they appear to access the more suitable fishing grounds. Daily landing reports were obtained from the Korean Fishery Union, which receives and manages fish landing data from fishing ports located on Jeju Island. Using the proposed method, we collected AIS-based fishing vessel trajectory data and fish landing data from the Korean Fishery Union over one year (January to December 2018) in the southern part of the Korean waters around Jeju Island. The AIS static data included the ‘fishing ship’ category of code ‘30’; however, the fishing vessel type information was not present. Therefore, we extracted the fishing vessel type information of the corresponding fishing vessel based on the MMSI (maritime mobile service identity) ID of the AIS from the Korean Fishing Ship Register Database, which is managed by the National Federation of Fisheries Cooperatives of Korea.

2.2. Vessel Activity Identification and Spatio-Temporal Distribution of Fish Catch

Classification of fishing activity based on trajectory data is essential for identifying fishing grounds. Since catch data are rarely recorded by vessel owners by location and time, we identified the vessel status using trajectory data and fishing operation characteristics. Fishing vessels have several fishing trajectory patterns, depending on the target species and gear. Usually, fishing vessels travel at low speeds when they are engaged in fishing activities. During non-fishing activities, they sail at high speeds or are stationary when they are in the harbor. Combining fishing vessel trajectory data and

fish catch data to understand the interrelations between vessel activities and impact is plausible using developing technologies and methods that establish the interconnection between these datasets and provide information that informs management of fisheries resources and marine planning [4]. We proposed a method that allocates catch data to fishing grounds by extracting fishing trajectories and redistributing the catch uniformly to these trajectories, and creating a fine scale spatiotemporal map of catch distribution in the study area. Figure 1 represents the process as detailed above.

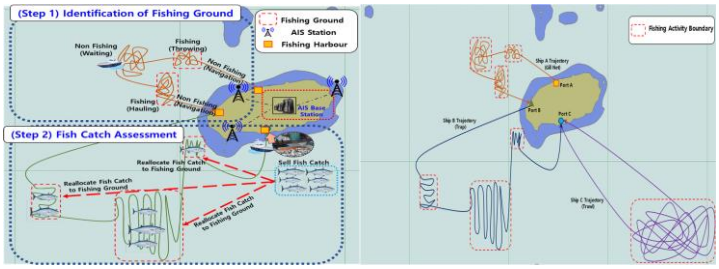


Fig. 1. The process of extracting fish catches from ship trajectory and reallocating to fishing ground

To reallocate catch data to fishing trajectories, we established fishing vessel status, i.e., fishing, non-fishing, port entry, and exit. Here, fishing refers to the throwing, hauling, and towing (deployment and retrieval) of fishing gear (Figure 2). Non-fishing refers to the movement between the fishing harbor and fishing ground and waiting periods after deploying fishing gear. Port entry refers to vessel entering a port to land and sell their catch and port exit refers to vessels steaming out of port to fishing grounds

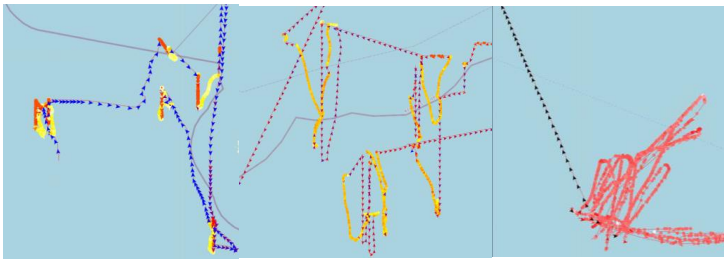


Fig. 2. Some examples of fishing trip trajectories and vessel activity ("Fishing" or "Non-fishing") for different gear types

3 Results

Estimation of Spatio-Temporal Distribution of Fish Catch Data To estimate the spatio-temporal distribution of catch, we identified the locations of fishing activity, the number of fishing hauls (NFH), and the corresponding fish catch information per fishing trip. The total catch is divided by the number of fishing hauls, each of which is allocated to the fishing segment of the trajectory. For a fishing vessel which has a k fishing gear

type selling fish species fs of m kg, we propose the equation of catch, FC in each fishing area, in a given trajectory, at a given time, t and for a given species, which is as follows

$$FC[area, datetime, k, fs] = \frac{m[k, fs]}{NFH}$$

3.1. Mapping spatio-temporal distribution of catch

Fishing vessel trajectories and a set of standards were applied to identify individual fishing trips as described in section 2. The standards were adapted for the five types of fishing gear considered during this study to enable the assessment of spatio-temporal variations in their activities. To assign catch values to fishing trajectories, the area under study was divided into square grid cells. The study area was divided into square grid cells and the spatial resolution was set at 0.1° (5° by 5° pixel for each 0.5° grid cell) and 0.2° (2.5° by 2.5° pixel for each 0.5° grid cell) for fishing vessel trajectory distribution (Fig. 3) and catch distribution (Fig. 4). The calculated fish catch values are assigned to each grid cell to generate maps of reallocated fish catch distribution for each gear type. The resulting maps show distributions of catch values to fishing trajectories and highlights spatio-temporal variations in fish catch amounts and intensities of fishing vessel activity. We produced vessel trajectory and catch distributions mapped at fine spatial scales and show variations in fishing activities and reliance on fishing grounds and variations in locations fished by fishing gears. Figure 3 a–e shows the fishing vessel trajectory distribution, while Figure 4 a–e shows the fish catch distribution as computed by the proposed method. The trajectory maps show that the areas fished varied for each gear type. Vessels operating purse seine and squid jigging gears were dependent on nearshore areas, and occasionally moving to areas further offshore (Figure 3 a,d). Fishing vessels that operate longlines, gill nets, and trawl nets operated within fishing grounds in locations further offshore, with longline vessels and gill net vessels operating from near coastal areas and moving offshore to the northern, western, and southern areas of the island. Trawlers, on the other hand, showed restricted operations in the south and northeastern parts of the island.

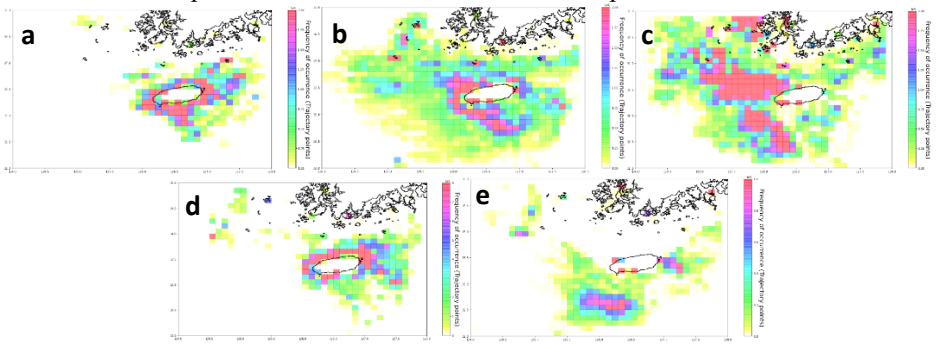


Fig. 3. Map of fishing vessel trajectory distribution in the study area for a) purse seine, b) longline, c) gill net, d) squid jigging and e) trawl fishing gears throughout January to December 2018.

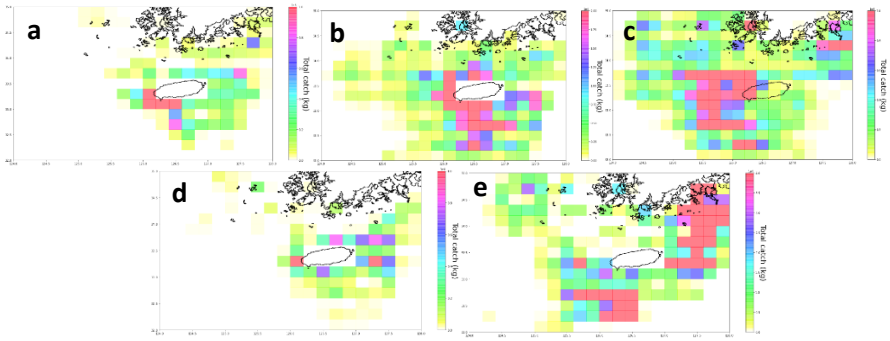


Fig. 4. Map of catch distribution in the study area for a) purse seine, b) longline, c) gill net, d) squid jigging and e) trawl fishing gears throughout January to December 2018.

4 Discussion and Conclusions

This study presents the initial results for the combined use of vessel trajectory data and fish landing data to assess the spatial and temporal patterns of distribution for gear and fishing vessel activity and their level of dependence on fishing grounds. The method proposed in this study was useful in identifying reference points for major gears and improves our ability to investigate seasonal variations and make predictions of potential changes in fishing activities of vessels in the coastal and offshore fishing industry. We used a short temporal extent of one year for both AIS and fish landing dataset in our research to analyze the spatial and temporal extent of fishing activities. The current work did not incorporate some variables like environmental datasets such as SST, salinity and wind which can help assess multifactor impact on the marine environment. With access to datasets collected over a longer period and by applying available machine learning methods [2, 3, 5], we can analyze impacts of climate change on fishing activities, provide a more extensive assessment of seasonal changes in catch and effort and make predictions that can support sustainable resource management and marine spatial planning [5].

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1C1C1013773).

References

1. International Telecommunication Union. Recommendation itu-r m.1371-3. Technical Characteristics for an Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Band. 2014. Available online: http://www.itu.int/ (accessed on 23 March 2021).

2. Adibi, P.; Pranovi, F.; Raffaetà, A.; Russo, E.; Silvestri, C.; Simeoni, M.; Soares, A.; Matwin, S. Predicting fishing effort and catch using semantic trajectories and machine learning. In International Workshop on Multiple-Aspect Analysis of Semantic Trajectories; Springer: Cham, Switzerland, 2019; pp. 83–99.
3. Skauen, A.N. Quantifying the tracking capability of space-based AIS systems. Adv. Space Res. **2016**, 57, 527-42.
4. de Souza, E.N.; Boerder, K.; Matwin, S.;Worm, B. Improving fishing pattern detection from satellite AIS using data mining and machine learning. PLoS ONE **2016**, 11, e0158248.
5. James, M.; Mendo, T.; Jones, E.L.; Orr, K.; McKnight, A.; Thompson, J. AIS data to inform small scale fisheries management and marine spatial planning. Mar. Policy **2018**, 91, 113–121.

Item-Oriented Mining of Rare Patterns from Big Data Applications and Services

Elieser Capillar, Chowdhury Abdul Mumin Ishmam, Carson K. Leung ✉ [0000-0002-7541-9127], Hoang Hai Nguyen, Adam G.M. Pazdor, Prabhanshu Shrivastava, Ngoc Bao Chau Truong

Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada
Carson.Leung@UManitoba.ca

Abstract. In numerous real-world big data applications and services (e.g., interpreting biological data, identifying rare associations between diseases and their causes, detecting anomalies), rare patterns play a crucial role. Nonetheless, uncovering these rare patterns presents challenges. This paper presents an effective algorithm for mining minimal rare patterns from sparsely correlated data. The algorithm creatively combines and customizes the vertical frequent pattern algorithm VIPER to efficiently discover minimal rare patterns. Evaluation results of our algorithm, RP-VIPER, demonstrate its superiority over current horizontal rare pattern mining algorithms. Additionally, the results underscore the performance enhancements achieved by our optimized strategies.

Keywords: Big data, big data applications, big data services, data mining, data analytics, knowledge discovery, pattern mining, rare patterns, vertical mining, bitwise representation, sparse data.

1 Introduction

Nowadays, big data are everywhere. Embedded in these big data are implicit, previously unknown, and potentially useful information (e.g., in the form of patterns), which can be discovered by data science. It makes good use of data mining [1-3], machine learning [4], visualization, mathematics, and statistics. Among data mining tasks (e.g., clustering [5, 6], classification [7, 8]), association rule mining and frequent pattern mining [9-12] are popular in various real-life big data applications and services [13, 14]. Examples include medical informatics [15-17], transportation analytics [18, 19], and social analysis [20-22].

In general, frequent pattern mining aims to discover frequently co-occurring items. These frequent patterns can be served as building blocks for antecedent A and consequence C of association rule of the form:

$$A \rightarrow C \tag{1}$$

which reveal associative relationships or correlations between items within A and C. It has played an important role in mining other patterns such as emerging patterns [23], sequential patterns [24], periodic patterns [25], and quantitative patterns [26].

Besides frequent pattern mining, *rare pattern mining* [27] can also result in interesting results because they represent infrequent patterns in data. These infrequent patterns are particularly useful in various fields (e.g., biology, medicine, security). Consider the following two examples.

Example 1. In the medical field of pharmacovigilance (which detects, assesses, and studies adverse drug effects), mining for *rare patterns* (*RPs*) can result in the discovery of associating drugs with adverse effects [28]. Provided with a database of drug effects, if "{drug} ∪ {effect A}" is found to be a frequent pattern, and "{drug} ∪ {effect B}" is found to be a RP, this information can be used to determine whether these are desired and expected effects.

Example 2. In computer network security, given a database of connections to a computer network, mining for RPs can easily isolate uncommon and unusual connections. The resulting list of rare connections can then be further analyzed to determine if any were malicious.

In terms of related works, there were few algorithms for mining rare patterns or minimal rare patterns (mRP). Examples include AprioriRare [28], MRG-Exp [28], and Walky-G [29]. As MRG-Exp and Walky-G were built on the foundation of AprioriRare, they performed best on dense and highly correlated data. When it came to sparse and weakly correlated data, their runtime and search space were similar to those of AprioriRare. As a remedy for this, we aim to improve the runtime on finding mRPs in sparse and weakly correlated data. Hence, in this paper, we present an efficient algorithm called RP-VIPER for mining mRPs from sparse and weakly correlated data. The algorithm non-trivially integrates and adapts vertical frequent pattern algorithm VIPER to discover mRPs in an efficient manner. Our *key contributions* of this paper include our design and development of this RP-VIPER algorithm.

The remainder of this paper is organized as follows. The next section provides background and related works. Then, Section 3 describes our RP-VIPER algorithm. Section 4 shows evaluation results. Finally, Section 5 draws conclusions.

2 Background and Related Works

Definition 1. A pattern X is a *rare pattern* (*RP*) if its support (i.e., frequency) is less than the user-specified minimum support threshold *minsup*:

$$\sup(X) < minsup \tag{2}$$

Definition 2. A pattern X is a *minimal rare itemset* (*mRP*) if it is rare but all of its proper subsets are frequent:

$$\sup(X) < minsup \tag{3}$$

$$\forall W \subset X, \sup(W) \geq minsup \tag{4}$$

A *vertical database* is an item-centric representation of data. It can be considered as a collection of bit vectors (BVs), each of which represent a domain item. The *i*-th bit of a bit vector representing a domain item *x* indicates the presence or absence of *x* in the *i*-th transaction. Specifically:

- a "0"-bit reveals that *x* is absent from the *i*-th transaction, whereas
- a "1"-bit reveals that *x* is present in the *i*-th transaction.

To mine frequent patterns, VIPER [30] represents data vertically by using bit vectors. It mines frequent patterns in a bottom-up, breadth-first fashion. Candidate (*k*+1)-itemsets (i.e., patterns of cardinality *k*+1) are generated by applying bitwise AND operations on bit vectors of two frequent *k*-itemsets. These bitwise operations are cheap.

A *horizontal database* is a transaction-centric representation of data. It can be considered as a collection of transactions, each of which captures co-occurrence of items—represented by their transaction IDs—present in a transaction. The following are algorithms for mining mRPs:

- AprioriRare [28] mines mRPs from a horizontal database in a bottom-up, breadth-first fashion. It performs Apriori [10] as normal, finding all mRPs as a by-product.
- MRG-Exp [28] also mines mRPs from a horizontal database in a bottom-up, breadth-first fashion. However, it reduces the search space of AprioriRare by looking for generators instead of itemsets. By the nature of generators, if an item is not a frequent generator, then it is an mRP.
- Walky-G [29] is similar to MRG-Exp in the sense that it also looks for generators when mining mRPs. However, Walky-G mines mRPs from a vertical database in a depth-first fashion.

Among these three algorithms, MRG-Exp and Walky-G run faster than AprioriRare partially due to their use of generators. They can efficiently mine mRPs from dense and strongly correlated data. However, their runtimes are similar to that of AprioriRare when mining sparse and weakly correlated data.

3 Our RP-VIPER Algorithm

Due to the nature of sparse and weakly correlated data, frequent patterns usually do not tend to get too large. See Fig. 1, in which Fig. 1(a) captures a dense and highly correlated dataset, whereas Fig. 1(b) captures a sparse and weakly correlated dataset. We observe from their superset lattices that mRPs discovered from these datasets usually lie just past the border between frequent and rare patterns. This observation motivate our current work.

Our RP-VIPER algorithm is a level-wise, bottom-up, breadth-first vertical algorithm for mining mRPs. Specifically, it first converts a given database to a vertical format if it is not already in a vertical format. It then stores all singleton bit vectors in a hash structure. Afterwards, it performs breadth-first candidate generation and mines mRPs. Pseudo code is shown in Algorithm 1.

(a) Density=70% (i.e., sparsity=30%)	(b) Density=42% (i.e., sparsity=58%)
BV(a) = [10111 01001]	BV(a) = [11000 11110]
BV(b) = [11111 11100]	BV(b) = [11010 10101]
BV(c) = [01011 11010]	BV(c) = [00011 01111]
BV(d) = [10101 11111]	BV(d) = [00100 00000]
BV(e) = [11111 00011]	BV(e) = [01001 00000]

Fig. 1. (a) A dense dataset with a density of 70% (i.e., sparsity = 30%), and (b) a sparse dataset with a density of 42% (i.e., sparsity = 58%).

Algorithm 1 RP-VIPER
Input: Transactional database (*T*), max transaction length (*maxXactLen*), *minsup*
Output: List of all minimal rare patterns (*minRare*)
1. *level* ← 1
2. *minRare* ← {infrequent singletons} // All minimal rare itemsets
3. *Frequents* $_{level}$ ← {frequent singletons} // All frequent 1-itemsets
4. construct vertical bit vectors of each frequent item
5. **while** (|*Frequents* $_{level}$| ≥ *level*+1 **and** *level*+1 ≤ *maxXactLen*)
6. *Frequents* $_{level+1}$ ← generateCandidates(*Frequents* $_{level}$); *level* ← *level* +1
7. **return** *minRare*

In terms of implementation, RP-VIPER takes advantage of a hash structure (hashMap) to capture distinct domain item in a key-value store, where the item is the key and its corresponding bit vector is the value. With this data structure, support of singletons can be easily computed. To generate candidate 2-itemsets, bitwise AND operations are applied to frequent singletons. This step is repeated to generate candidate k-itemsets (for all $k \geq 2$) until:

- no more frequent itemsets can be generated, or
- the level of the largest transaction length is reached.

During the candidate generation process, RP-VIPER first calls a subroutine generateCandidate(list) to generate candidates. It, in turn, calls checkSubset(itemset, list) to generate all proper subsets and eliminate one item at a time from the itemset. It then checks its presence in the Frequents list and returns whether or not it is found. True returns are added to the Frequents list if they meet *minsup*. True returns that do not meet *minsup* are saved as an mRP.

4 Evaluation

To evaluate our RP-VIPER, we compared with AprioriRare. Both algorithms were implemented in Java. Experiments were performed on an Intel Core i5-8265U 1.60GHz CPU on Windows 10 with 8 GB of RAM. We applied both algorithms to four benchmark datasets (which are available at SPMF library[1]):

[1] https://www.philippe-fournier-viger.com/spmf/index.php?link=datasets.php

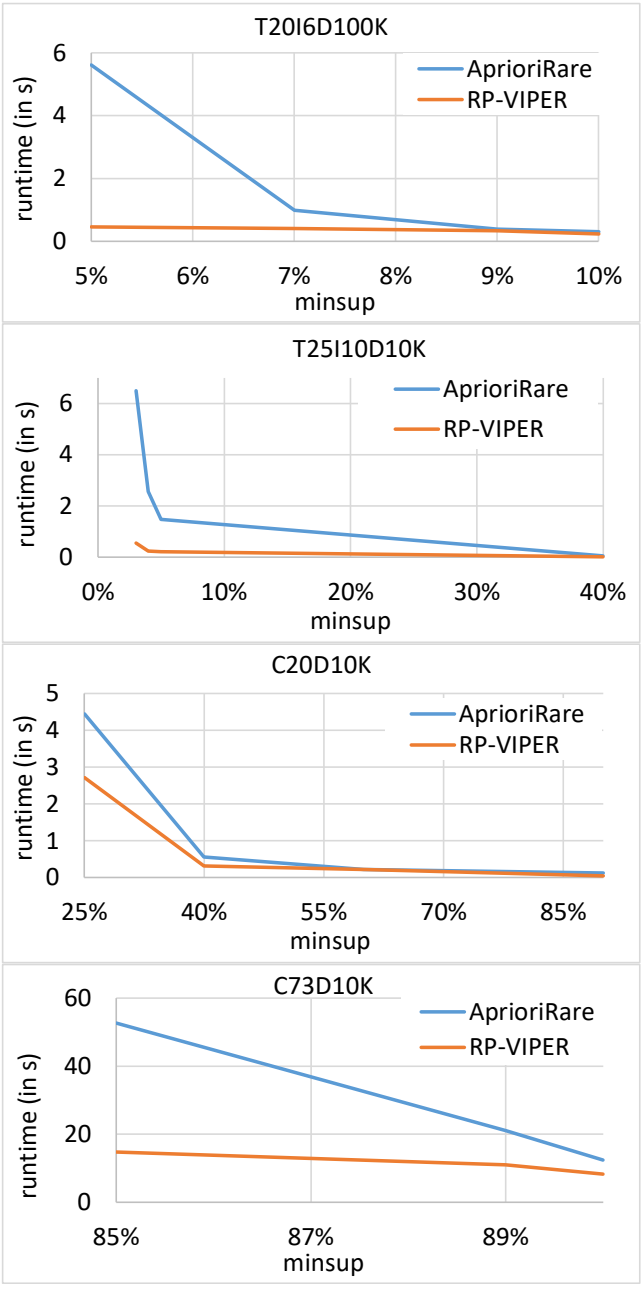


Fig. 2. Runtimes of our RP-VIPER and existing AprioriRare on four datasets.

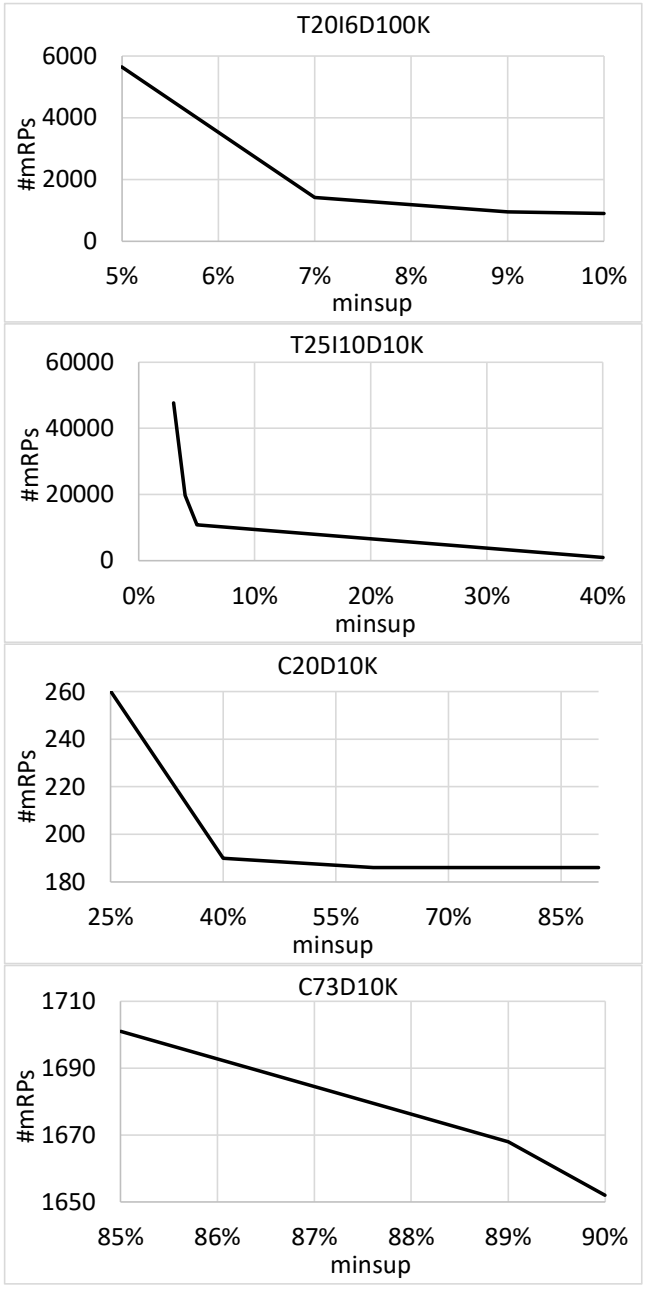


Fig. 3. The number of mRPs mined by our RP-VIPER (and AprioriRare) from four datasets.

- T20I6D100K, which is a sparse and weakly correlated dataset with a sparsity of 97.77% with 893 distinct items;
- T25I10D10K, which is a sparse and weakly correlated dataset with a sparsity of 97.33% with 929 distinct items;
- C20D10K, which is a slightly denser and highly correlated dataset with a sparsity of 89.58% with 192 distinct items; and
- C73D10K, which is also a slightly denser and highly correlated dataset with a sparsity of 95.41% with 1592 distinct items.

We measured both runtime and the number of mRPs. The reported runtimes are average of multiple runs. Results show that, when *minsup* increased, the number of mRPs dropped and thus reducing the runtime. The gap between the two algorithms increased when *minsup* decreased. Our RP-VIPER took shorter to mine mRPs than the existing AprioriRare. See Figs. 2 and 3.

5 Conclusions

In this paper, we presented RP-VIPER as a foundation for mining minimal rare patterns in sparse and weakly correlated data. The evaluation results show that, when using sparse datasets regardless of correlation strength, our RP-VIPER outperforms the existing AprioriRare when given low minimum support thresholds. The algorithm is enhanced with an optimization of using a hash structure for all levels of candidate generation (instead of just at the second level). Moreover, the algorithm takes advantage of vertical item-centric representation of data—namely, bitwise representation. By doing so, candidates can be generated by performing bitwise AND operations. As *ongoing and future work*, we explore additional ways to further optimize our RP-VIPER algorithm with an aim to further reduce its runtime in mining minimal rare patterns. We would also like to extend RP-VIPER by incorporate Q-VIPER [32] for mining minimal rare *quantitative* patterns.

Acknowledgement. This work is partially supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and University of Manitoba.

References

1. Aggarwal, C.C.: Data Mining: The Textbook. Springer (2015)
2. Han, J., et al.: Data Mining: Concepts and Techniques, 4th edn., MK (2022)
3. Leung, C.K., et al.: Constrained big data mining in an edge computing environment. In: Big Data Applications and Services 2017. AISC, vol. 770, pp. 61-68.
4. Sarumi, O., Leung, C.K.: Scalable data science and machine learning algorithm for gene prediction. In: BigDAS 2019, pp. 118-126.
5. Brown, P.O., et al.: Mahalanobis distance based k-means clustering. In: DaWaK 2022. LNCS, vol. 13428, pp. 256-262.
6. Dierckens, K.E., et al.: A data science and engineering solution for fast k-means clustering of big data. In: IEEE TrustCom-BigDataSE-ICESS 2017, pp. 925-932.

7. Choudhery, D., Leung, C.K.: Social media mining: prediction of box office revenue. In: IDEAS 2017, pp. 20-29.
8. Min, B., et al.: Image classification for agricultural products using transfer learning. In: BigDAS 2020, pp. 48-52.
9. Agrawal, R., et al.: Mining association rules between sets of items in large databases. In: ACM SIGMOD 1993, pp. 207–216.
10. Agrawal, R., Srikanth, R.: Fast algorithms for mining association rules. In: VLDB 1994, pp. 487-499.
11. Han, Z., Leung, C.K.: FIMaaS: scalable frequent itemset mining-as-a-service on cloud for non-expert miners. In: BigDAS 2015, pp. 84-91.
12. Hoi, C.S.H., et al.: Constrained frequent pattern mining from big data via crowdsourcing. In: Big Data Applications and Services 2017. AISC, vol. 770, pp. 69-79.
13. Leung, C.K.: Big data mining applications and services. In: BigDAS 2015, pp. 1-8.
14. Leung, C.K., Nasridinov, A. (eds.): Proceedings of the 2015 International Conference on Big Data Applications and Services (BigDAS 2015). ACM Press.
15. de Guia, J., et al.: DeepGx: deep learning using gene expression for cancer classification. In: IEEE/ACM ASONAM 2019, pp. 913-920.
16. Fung, D.L.X., et al.: Self-supervised deep learning model for COVID-19 lung CT image segmentation highlighting putative causal relationship among age, underlying disease and COVID-19. BMC Journal of Translational Medicine 19, 318:1-318:18 (2021)
17. Leung, C.K., et al.: Health analytics on COVID-19 data with few-shot learning. In: DaWaK 2021. LNCS, vol. 12925, pp. 67-80.
18. Balbin, P.P.F., et al.: Predictive analytics on open big data for supporting smart transportation services. Procedia Computer Science 176, 3009-3018 (2020)
19. Leung, C.K., et al.: Urban analytics of big transportation data for supporting smart cities. In: DaWaK 2019. LNCS, vol. 11708, pp. 24-33.
20. Braun, P., et al.: MapReduce-based complex big data analytics over uncertain and imprecise social networks. In: DaWaK 2017. LNCS, vol. 10440, pp. 130-145.
21. Leung, C.K., et al., Big data analytics of social network data: Who cares most about you on Facebook? In: Highlighting the Importance of Big Data Management and Analysis for Various Applications, pp. 1-15 (2018)
22. Leung, C.K., et al., Personalized DeepInf: enhanced social influence prediction with deep learning and transfer learning. In: IEEE BigData 2019, pp. 2871-2880.
23. Dong, G., Bailey, J.: Contrast Data Mining: Concepts, Algorithms, and Applications. Chapman & Hall/CRC (2012)
24. Agrawal, R., Srikant, R.: Mining sequential patterns. In: IEEE ICDE 1995, pp. 3-14.
25. Madill, E.W., et al.: Enhanced sliding window-based periodic pattern mining from dynamic streams. DaWaK 2022, LNCS, vol. 13428, pp. 234-240.
26. Srikant, R., Agrawal, R.: Mining quantitative association rules in large relational tables. In: ACM SIGMOD 1996, pp. 1-12.
27. Weiss, G.M.: Mining with rarity: a unifying framework. ACM SIGKDD Explorations 6(1), 7-19 (2004)
28. Szathmary, L., et al.: Towards rare itemset mining. In: IEEE ICTAI 2007, pp. 305–312.
29. Szathmary, L., et al.: Efficient vertical mining of minimal rare itemsets. In: CLA 2012, pp. 269–280.
30. Shenoy, P., et al.: Turbo-charging vertical mining of large databases. In: ACM SIGMOD 2000, pp. 22–33.
31. Czubryt, T.J., et al.: Q-VIPER: quantitative vertical bitwise algorithm to mine frequent patterns. In: DaWaK 2022. LNCS, vol. 13428, pp. 219-233.

Deep learning-based method for real-time safety helmet detection in construction/manufacturing sites

Woochan Park[1], Joonghun Cho, Sang-hyun Choi[2*]

[1]Dept. Bigdata, Chungbuk National University, Cheongju, South Korea
[2]Dept. Management Information System, Chungbuk National University, Cheongju, South Korea
{2022278010, chois}@cbnu.ac.kr

Abstract. In hazardous construction and manufacturing environments, worker safety is imperative. This paper introduces an AI-based solution to detect non-compliant workers, particularly related to protective gear adherence. The YOLOv5 model's implementation achieved robust real-time safety helmet detection, demonstrating its effectiveness with an mAP@0.5 score of around 65%. This enhances workplace safety by identifying rule non-compliance. Future efforts include expanding the approach to detect critical incidents like fire, smoke, and falls, aiming for a comprehensive safety solution.

Keywords : Object detection, safety, Deep Learning

1 Introduction

Construction and manufacturing sites are dangerous environments where workers are exposed to various potential safety accidents. According to the Ministry of Employment and Labor, about 16% of industrial accidents are caused by not wearing protective equipment. It, simply, follows that if employees comply with the safety regulations by wearing protective equipment, then industrial accidents will naturally decrease. However, ensuring that all workers comply with safety regulations can be difficult for site supervisors. The development of Artificial Intelligence (AI), especially, computer vision has potential solution to overcome that problems. In this study, we propose an automated system that detects and identifies non-compliant workers based on deep learning.

2 Methodology

In this study, we focus on detecting the protective equipment. For this YOLO models were used to detect wearing helmet and vest. The Kaggle HardHat-Vest dataset was utilized for this study, containing a diverse range of images from construction and

* Corresponding author

manufacturing sites. The dataset was split into three subsets: 78% for training, 11% for testing, and 11% for validation. The collected data is summarized in table1.

Table 1. datasets were preprocessed in three ways.

Instances	Train	Test	Validation	Total
Helmet	44,240	6,895	6,725	57,860
Vest	7,794	1,156	1,074	10,024
Head	98,652	12,565	124,826	124,826

3 Conclusion

In conclusion, the implementation of the YOLOv5 model has proven to be a robust and effective approach in achieving real-time safety helmet detection within the complex environments of construction and manufacturing sites. With an impressive mAP@0.5 score of approximately 65%, the model demonstrates its proficiency in enhancing workplace safety by identifying individuals not adhering to safety regulations.

Our future trajectory involves further enriching this method's capabilities by introducing additional classes for detecting critical industrial accidents such as fire, smoke, and falls. By extending the model's scope, we aim to provide a more holistic safety solution that addresses a wider range of potential hazards, reflecting our dedication to safeguarding workers and preventing accidents.

The successful implementation of this deep learning-based approach underscores its potential to revolutionize safety practices in industrial settings, enhancing proactive monitoring and risk mitigation. As we move forward, the integration of advanced features promises to make workplaces even safer, contributing to a more secure and productive environment for all.

Prediction of New Solar Power Generation Using Machine Learning Ensemble Techniques

Dong-kyu Yun[1], Wooseok Choi[1], Sang-hyun Choi[2*]

[1]Dept. Bigdata, Chungbuk National University, Cheongju, South Korea
[2]Dept. Management Information System, Chungbuk National University, Cheongju, South Korea
{dongkyu.yun, chois}@cbnu.ac.kr, {cdt3017}@naver.com

Abstract. In order to reduce fossil fuels, which are the cause of environmental pollution, there is a lot of interest in the use of renewable energy around the world. Among them, solar power generation is growing in size worldwide because it is inexpensive and easy to install. Accurately predicting the amount of solar power generation is very important for stable power plant operation by checking the feasibility in advance or checking the efficiency of facility operation before constructing the power plant. However, predictions are difficult in areas where power plants are scheduled to be built without data or in small power plants that lack data. Therefore, this study predicted new power plants by establishing a 'solar power generation integrated prediction model', and as a result of the prediction, the model proposed in this study recorded an error rate of 11.02%, 17.38% lower than the single model of the new power plant model.

Keywords : Photovoltaic, Prediction, Machine-Learning, Data Scaling

1 Introduction

In addition to interest in renewable energy generation due to environmental pollution, solar power plants are growing in size due to their relatively low construction costs and easy construction compared to other power plants. However, indiscriminate construction of solar power plants is another problem because discarded solar panels cannot be recycled or materials from processing destroy the environment. In addition, accurately predicting power generation is very important because it can reduce operating costs due to losses and indiscriminate installation in the process of starting and stopping power plant facilities. However, in the case of new or small power plants, there is a problem that is difficult to predict accurately because there is insufficient data to build predictive models.

Therefore, this study aims to establish a ‘solar power integrated prediction’ model that learns prediction models by scaling solar power in various regions so that it can be used to predict and operate new or small power plants in advance.

* Corresponding author

2 Data & Analytics

A total of four power generation data were used in the analysis, and 34,920 solar power generation data were collected from solar power plants located in Guro-gu, Seoul (30.75kWp), Haenam-gun, Jeollanam-do (998.4kWp), and 903.1kWp, Gyeongsangbuk-do, respectively. The other is a research testbed solar power plant (1.48 kWp) located on the roof of N13-dong, Chungbuk National University, with 2,016 solar power generation data by hour (2022-09-08-00 to 2022-11-30-23). Since solar power generation is heavily influenced by meteorological factors, weather data (temperature, humidity, solar radiation, and cloud) over time in the area where the power plant is located were also collected. In addition, time variables such as month and time were used.

In this study, two models were created, and the prediction accuracy was compared to confirm the performance of the integrated solar power generation prediction model. Model 1 is a model learned using 1,814 data sets, which is 90% of the data sets of small testbed site, and Model 2 is a model learned using a dataset with 106,574 cases each added by scaling the data used in Model 1 and data from the remaining three regions. The test data used to compare the accuracy of the two predictive models are 202 cases, or 10% of the remaining testbed data set.

Model 1 and Model 2 were respectively learned with Random Forest, XGBoost, and LightGBM, which are machine learning-based ensemble models, and the prediction error rate was checked and compared using NMAE (Normalized Mean Absolute Error). The predicted error rate of each model is shown in Table 1.

Table 1. Comparison table of prediction results by Methodology & Algorithm.

Methodology \ Algorithms	Random Forest	XGBoost	LightGBM	Average
Model1	29.86 %	29.33 %	28.12 %	29.1 %
Model2	12.46 %	**11.02 %**	11.68 %	11.72 %
Average	21.16 %	20.17 %	19.9%	

As a result of the prediction, Model 2, a method of scaling power generation in other regions and using it as one data, was 17.38% lower than Model 1, which uses only data from small plants, and it was confirmed that XGBoost was the best with an error rate of 11.02%.

3 Conclusion

This study confirmed that the use of the ‘solar power integrated prediction model’ can more effectively predict future power generation of new or small plants with insufficient data for predictive model learning than the existing method of predicting only by learning data from one plant. In addition, this prediction method can be used to check profitability in advance before installing a power plant by predicting the expected amount of power in the region before building a solar power plant, and to prevent indiscriminate construction of a solar power plant.

Disaster Resilience analysis of Urban Planning Facilities upon Urban Flood Risk

Kiyong Park[1],

[1]Dept. of Big data, Chungbuk National University, 1 Chungdae-ro, Seowon-gu, Cheongju, Chungbuk, 28644, Republic of Korea, pky3489@chungbuk.ac.kr

Abstract. Influenced by recent climate change, frequencies of urban disasters have increased in scales and diversified in types. Such trends have highlighted the importance of urban prevention schemes. This study analyzed the effect of urban planning facilities on the resilience from disaster with the focus on the urban space as a non-structural measure against urban flood damage. The effect of urban planning facilities was analyzed by evaluating relative values using a decision making system with network structure based on four functional features of the recovery: robustnes, redundancy, resourcefulness and rapidity. As a result, the disaster-prevention facilities showed the highest recovery efficiency followed by space, traffic, public-cultural-athletic, environmental, health and sanitary, and distribution and supply facilities, in respective order. Therefore, the facilities with high recovery efficiency should be primarily planned and installed in the areas vulnerable to urban flood damages. And the facilities with less recovery efficiency should be planned in the areas relative safe from flooding to increase overall recovery efficiency to minimize damage from flooding.

Keywords: Climate Change, Urban Flood Risk, Urban Planning Facilities, Disaster Resilience

1 Introduction

Global warming is one of the main reasons for the increase in frequency and magnitude of extreme rainfalls (therefore floods), since the warmer atmosphere with enhanced humidity leads to a more active hydrological cycle (Mailhot et al., 2007). Katz and Brown (1992) stated that even a small change in the mean rainfall due to global warming can cause significant changes in extreme rainfalls. Also, urbanization is another driver which increases intensity and frequency of floods. The effects of urbanization and global warming on floods will increase in future due to more urbanization to accommodate increased population in cities and rises in greenhouse gas emissions (IPCC 2013). Cities are complex and interdependent systems, extremely vulnerable to threats from both natural hazards. The very features that make cities feasible and desirable—their architectural structures, population concentrations, places of assembly, and interconnected infrastructure systems—also put them at high risk to floods attacks (Godschalk, 2003).

This paper issues a call for advanced planning and action to reduce those risks through the development of resilient cities.

2 Methods

The flood risk was analyzed based on the Flood Risk = Hazard (Exposure) × Vulnerability equation (Kron, 2002; Brooks et al., 2005). The hazard (exposure) analysis map and vulnerability analysis map were used to analyze the flood risk. Flood risk reduction approaches have proposed that can resilience of urban planning facilities. Bruneau et al. (2003) demonstrated four functional goals of resilient infrastructure: robustness, redundancy, resourcefulness, and rapidity. 4R framework of resilience is proposed. ANP(Analytic Network Process) analysis of urban planning facilities based on 4R framework was analyzed.

3 Results and Discussion

Flood risk of analysis area is shown in Figure 1. Resilience of urban planning facilities are shown in Figure 2.

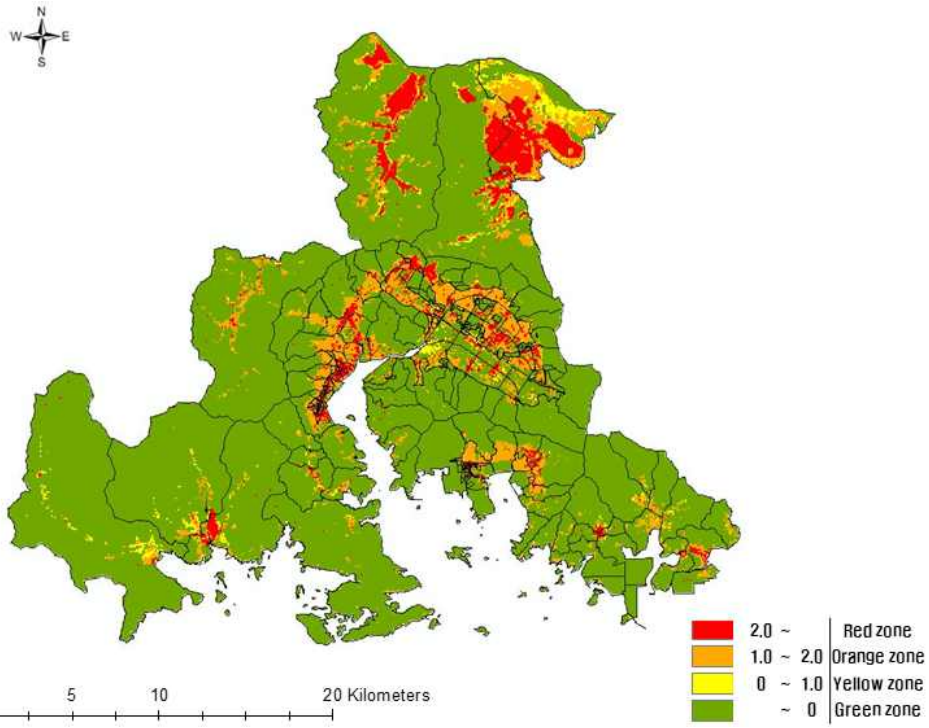
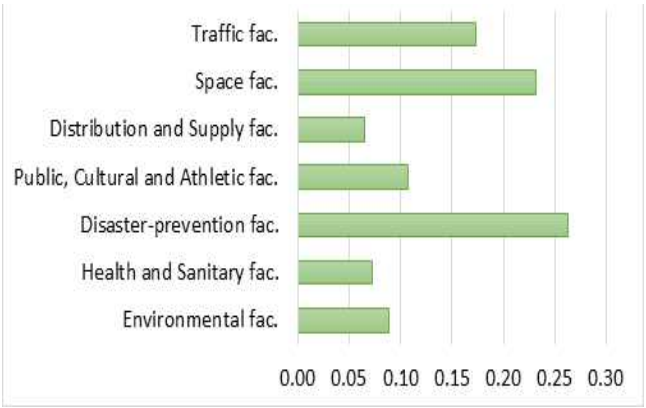
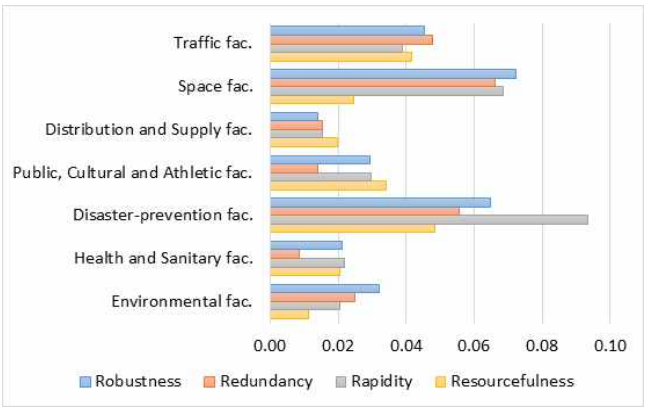


Fig. 1. Flood risk



(a) Weight of Resilience



(b) Weight of Resilience Element

Fig. 2. Resilience of Urban Planning Facilities

4 Conclusion

The results of this urban flood risk analysis demonstrated that the areas have high levels of flood risk that downtown and lowland of analysis area. As a result of rresilience analysis of urban Planning Facilities, the disaster-prevention facilities (0.26268) for damage prevention showed the highest recovery efficiency followed by Space(0.23158), traffic(0.17345), public, cultural and athletic(0.10694), environmental(0.08835), health and sanitary(0.07241), and distribution and supply facilities(0.06460), in respective order. Therefore, the facilities with high recovery efficiency shall be primarily planned and installed in the areas vulnerable to urban flood damages and the facilities with less recovery efficiency, in the areas relative safe from flooding to increase overall recovery

efficiency to minimize damage from flooding. I argue that urban flood risk mitigation is a branch of disaster risk mitigation practice, and that its overriding goal should be to develop resilient cities.
A nonstructural measure is difficult to quantify and analyze due to its qualitative aspects compared to a structural measure. However, it is necessary to investigate a long-term perspective of this issue as it can address the limitations of current structural measures. The results of this study can help further research on preparing nonstructural measures from various perspectives.

References

1. Mailhot, A, Duchesne, S, Caya, D and Talbot, G, Assessment of future change in intensity–duration–frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM), *Journal of Hydrology*, 347 (2007), 197–210.
2. Katz, RW & Brown, BG, Extreme events in a changing climate: variability is more important than averages, *Climatic Change*, 21 (1992), 289-302.
3. IPCC, Contribution of working group I to the fifth assessment report of the Intergovernmental panel on climate change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2013).
4. Godschalk, D. R., Urban Hazard Mitigation: Creating Resilient Cities," *Natural Hazards Review*, 4 (2003), 136-143.
5. W. Kron, Flood risk = hazard x exposure x vulnerability, Flood Defence, Science Press, New York. (2002) 82-97.
6. N. Brooks, W.N. Adger, P.M. Kelly, The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation, Global Environ. Chang. 15 (2005) 151-163.
7. Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D. and Winterfeldt, D., A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities, Earthquake Spectra, 19 (2003), 733–752.

An exclusive digital map system for alumni connection: University of "U' Case

Zeyuan Yu[1], Guowei Ou[1], Seon-Phil. JEONG[1],

[1] Computer Science & Technology, Department of Computer Science, Faculty of Science and Technology, BNU-HKBU United International College, Zhuhai, P.R. China
{p930026150, p930026101}@mail.uic.edu.cn, spjeong@uic.edu.cn

메모 포함[S1]: Chinese authors should write their first names in front of their surnames. This ensures that the names appear correctly in the running heads and the author index.

Abstract. University of 'U'('UU') has been expanding for years as an internationalized college with whole-person education characteristics, consequently, a large number of graduates from various departments and majors, many of whom have spread all over the world. Therefore, the alumni network between the graduates is also an advantage for 'UU' graduates. This alumni map project is to meet the potential demand of 'UU' alumni management and the need for alumni to establish contact with each other. This alumni map project aims to enhance the connection between 'UU' graduates who are scattered over the world. Of course, there are already well-established commercial SNSs available, but exclusive services for the graduates are needed. With this exclusive map service, new functions can be added quickly, and internal changes can be reflected promptly. We explored various possible web architectures and detailed cutting-edge technologies for the implementation of this project. In this project, we adopted several open APIs and implemented them on a cloud server with recommendation service functions based on a clustering algorithm.

Keywords: Web-based Map Service, Hierarchical Clustering, Recommendation System, Open source map API.

1 Introduction

Several universities are now offering alumni map systems. However, despite the demand for a variety of features from both alumni and administrators, many of these systems lack functionality. For instance, many do not have automated functions to recommend engaging alumni to each other. As such, there is a need for the development of a secure and algorithm-backed recommendation system to enhance the user experience. [1][2]

In the system, alumni could register and login to the alumni map system, where they could upload their contact info, location and other information such as major and year of admission. After being verified, they can view other alumnus on the map. Alumni separated around the world can view each other's location and other information on this alumni map platform, and then they may find each other and establish contact.

Alumni around the world could easily find others with some shared characteristics such as the same major, same year of admission or same region. After viewing and having contact with each other, the alumni could have meetings with others nearby (the geographical distance will be shown on map). 'UU' office will gain a visualized tool to easily grab location information, and contact one of them if needed. 'UU' may also use such maps as visual material to promote itself as an international college with great prospects to study.

2 Materials and Methods

This Alumni map project surely requires a visualized interface to put alumni data into graphical maps. Such a process is GIS mapping. [1],[2],[3],[4]

GeoPandas allows users to easily do implementations in Python which do require a spatial database.Folium builds on the data processing functions of the Python ecosystem and the mapping ability of the Leaflet.js. to control target data in Python, and then visualizing it in a Leaflet map through folium. [5],[6],[7]

Hierarchical clustering

1. Start with n clusters each consisting of a single object.
2. Calculate the matrix of distances.
3. Merge the 2 nearest clusters into a single (new) cluster.
4. Calculate the distance between the new cluster & each of other clusters. Repeat step 3 and 4 (n-1) times.
5. Draw the dendrogram or a tree.

Above is how you hand-write a hierarchical clustering, since in each round of iteration, distances between each node were calculated, that explains why the time complexity of this algorithm is O(kn^2) where n is the number of clusters, and k here is number of alumni

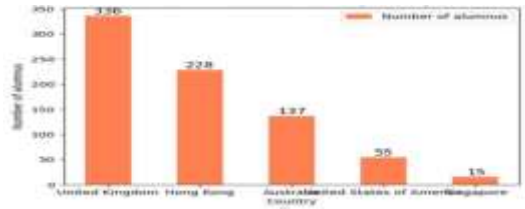


Fig. 1. Geographic distribution of 2019 graduates of 'UU' [8]

3 Implementation Results

Below is the pseudo code of backend supporting this recommendation

```
//clustering part
login_user = get_user_by_session
alumni_list = get_all_alumni //except the login user itself
clusternum = int(alumnus.count() / cluster) //cluster is adjective by user on GUI
hierarchical_cluster = AgglomerativeClustering(n_clusters=clusternum,
affinity='euclidean', linkage='ward')  //n_clusters is the number of cluster
initially
labels = hierarchical_cluster.fit_predict(data) //data has 2 dimensions, latitude
and longitude
```

Above is the hierarchical clustering part of the weighted sort, the recommendation is based on the finalized weight including more factors than geographical cluster result.

```
Compare cluster //if each alumni is in the same cluster with user
  Compare year and major
  Weight = w1(if_same_cluster) + w2(major) + w3(year) //w1>w2>w3
  Sort and output first 5 alumni
```

Here if an alumni has the same cluster attribute with login user, w1 will be added to the weight of this alumni, and w2 for major, w3 for year. The logic is at first if you want to actually meet someone the location between you two should not be too far away, that is why w1 takes the most value in weight. And people tend to be interested in other people that share the same job/major, that is why w2 is larger than w3, but of course it would be better if they are students of the same admission year.

Fig. 2. A visual feature and a query result of the suggested system.

4 Conclusion

In In this project, we first examined the context of 'UU' alumni and the current situation and explained why there is a particular need for an alumni map system. The purpose was elaborated, such as guiding 'UU' admin to manage all graduated students participating in this system and assisting alumni in locating and contacting each other using visualized data and queries. We explained how they were implemented in our project then, the architecture and other diagrams were illustrated to show the structure of this suggested system clearly.

In addition, we introduced a recommendation system based on the weighted sort involving hierarchical clustering. As our future plan, blockchain technology would be introduced to enhance the security and efficiency of this system.[9]

References

1. ArcGIS Pro Helps You Get Work Done Faster. (n.d.). Retrieved from esri: https://www.esri.com/about/newsroom/arcuser/arcgis-pro-helps-you-get-work-done-faster/?rmedium=arcuser&rsource=https://www.esri.com/esri-news/arcuser/summer-2014/arcgis-pro-helps-you-get-work-done-faster
2. Computer graphics using OpenGL. (n.d.). Retrieved from http://hi.baidu.com/xun1573/blog/item/2295fa2580c41f6735a80f0c.html
3. Different Types of Distance Measures in Machine Learning. (2020). Retrieved from Data Analytics:https://vitalflux.com/different-types-of-distance-measures-in-machine-learning/
4. Ma, J,. Computer World. Mcgraw Hill (2009)..
5. Sevtsuk, A., & Mekonnen, M., Urban network analysis: A new toolbox for ArcGIS. Revue internationale de géomatique (2012).
6. What is Geographic Information Systems (GIS)?, Retrieved from GISGeography: https://gisgeography.com/what-gis-geographic-information-systems/ (2022)
7. Liu J. G., Zhou T., Wang B. H., research progress of personalized recommendation system "J1 progress in Natural Science 19 (1): 1-3.(2009).
8. Lyu C. H., Xavi, UIC Alumni Map, Zhuhai, UIC, (2019).
9. Cai, Ting & Yang, Zetao & Chen, Wuhui & Zheng, Zibin & Yu, Yang., A Blockchain-Assisted Trust Access Authentication System for Solid. IEEE Access. PP. 1-1. 10.1109, (2020).

A Comparative Study of Public Frameworks for Facial Landmark Detection

Tserenpurev Chuluunsaikhan[1], Jeong-Hun Kim[2], Aziz Nasridinov[1,*]

[1]Department of Computer Science, Chungbuk National University,
28644, Cheongju, South Korea
[2]Bigdata Research Institute, Chungbuk National University,
28644, Cheongju, South Korea

{teo, etyanue, aziz}@chungbuk.ac.kr

Abstract. Facial landmark detection is one of the computer vision techniques in face processing on images and videos. It identifies specific facial features (i.e., eyes, nose, and others) and locates them in 2D or 3D space. Facial landmark detection can be used in various applications, such as face alignment, emotion recognition, and augmented reality. There are many public frameworks (or libraries) for facial landmark detection. Selecting the optimal framework depends on the operating environment, device capacity, and others. This paper compares the two popular libraries called MediaPipe and DLIB. We consider several simple and essential factors like distance, angle, light, and speed. The experiment results help to determine the frameworks' performance and select the optimal one for a specific task.

Keywords: Facial landmarks, Face processing, MediaPipe, DLIB

1 Introduction

A facial landmark is the location of a specific feature, for example, the eye's corner, the nose's tip, or the face's edge. In computer vision, Facial landmarks become one of the essential features for emotion detection [1], face alignment [2], 3D construction, and augmented reality [3]. Utilizing a public framework for facial landmark detection tasks offers rich functionality, compatible integration, scalability, and community support. There are many public frameworks or libraries for facial landmark detection. Each framework is designed for its purpose. Therefore, a framework can have advantages and disadvantages depending on the task. So, the optimal framework depends on the task purpose, operating environment, and device capacity.

MediaPipe [4] and DLIB [5] are popular public frameworks used for computer vision tasks, including face detection, object detection, facial landmark detection, and others. MediaPipe is an open-source framework developed by Google. Its features include pose estimation, face mesh, hair segmentation, and hand landmark detection that analyze the human body. MediaPipe supports cross-platform web, desktop, mobile, and embedded systems. DLIB is a C++ open-source framework for various

computer science tasks used for machine learning, image processing, data mining, and many others. DLIB is used for robotics, embedded devices, and mobile phones. DLIB can be implemented by Python through it's bindings.

In this paper, we organized experiments to compare the two public frameworks. Our experiments include essential factors like speed, distance, light, and angle. Speed, distance, and light are tested in the office environment, and angle is tested in the factory environment. The experiment results help to determine the frameworks' performance and select the optimal one for a specific task.

2 Facial Landmark Detection Frameworks

2.1 MediaPipe

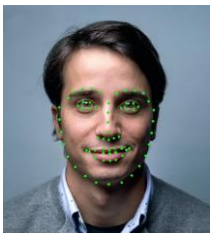
MediaPipe face mesh detector outputs 468 3D facial landmarks in images and videos. It consists of face detection and face landmark models. The face detection reveals a face on an input image and crops the face to an area of 256x256. The face landmark predicts 3D facial landmarks from the area. Fig 1 shows an example of the MediaPipe facial landmarks result as an image and return data.

```
[
   landmark {x: 0.65, y: 0.73, z: -0.02},
   landmark {x: 0.64, y: 0.68, z: -0.04},
   …
]
```

Fig. 1. MediaPipe result. The image from Unsplash.

2.2 DLIB

DLIB library estimates 68 2D coordinates on the face using a pre-trained facial landmark detector. The coordinates include the facial landmarks of eyebrows, eyes, nose, mouth, and face edge. It first locates a human face by a rectangle (i.e., x, y, h, w). After that, it estimates the facial landmarks in the rectangle inside. Fig 2 shows an example of the DLIB facial landmarks result as an image and return data.



```
[
  [276, 192], [279, 209], [282, 225],
  [286, 241], [292, 255], [301, 268],
  [314, 278], [329, 286], [345, 287],
  [362, 284], [375, 276], [386, 265],
  …
]
```

Fig. 2. DLIB result.

3 Experiments

3.1 Experiment Design

In this paper, we compared the facial landmark detection of two MediaPipe and DLIB frameworks based on simple factors like speed, distance, light, and angle. We developed a real-time system to conduct the experiments using Python and its libraries, such as MediaPipe, DLIB, and OpenCV. The system can detect facial landmarks from various data types like folders with images, videos, or CCTV URLs. Speed, distance, and light experiments are conducted in the office, and angle experiments are conducted in factory environments. These experiments aim to find an optimal framework for emotion recognition in a factory environment.

3.2 Experiment Configuration

Table 1 shows the device configuration of our experiments. First, we use a computer with an Intel Core i7 CPU and 16 GB RAM. Second, we utilize a Galaxy Tab S5e as the camera. The app IP Webcam allows us to use the tab as a CCTV camera. It returns real-time 640x360 images with 30 FPS.

Table 1. Device configurations.

Device	Spec	Value
Server	CPU	Intel Core i7-1165G7
	RAM	DDR4 16GB
	OS	Windows 11 Home
	Python	3.8.10
Camera	Name	Galaxy Tab S5e
	Resolution	640x360
	FPS limit	30/30

3.3 Experiment Results

Table 2 shows the rate of frame per second (FPS). We set a maximum of 30 FPS for real-time video without facial landmark detection. Both frameworks show similar FPS performance between 28 - 30.

Table 2. FPS comparison of frameworks.

Initial	MediaPipe	DLIB
		
30	28 ~ 30	28 ~ 30

Table 3 compares the face landmark detection performance of the MediaPipe in an office environment. Here, the columns represent light between bright to dark. The rows represent distances from 90 cm to 180 cm. MediaPipe can detect facial landmarks sufficiently at a distance of 120 cm in all light environments. However, detecting at a distance of more than 150 cm is difficult.

Table 3. MediaPipe performance based on distance and light.

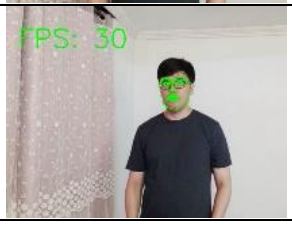
	Bright	Normal	Dark
90 cm			
120 cm			

Table 4 compares the face landmark detection performance of the DLIB in an office environment. Here, the columns represent light between bright to dark. The rows represent distances from 90 cm to 180 cm. Like MediaPipe, DLIB can detect facial landmarks at a distance of 120 cm. However, it can detect only at a distance of 90 cm in a dark environment.

Table 4. DLIB performance based on distance and light.

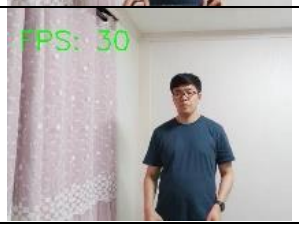
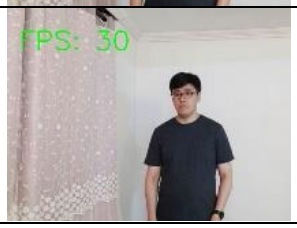
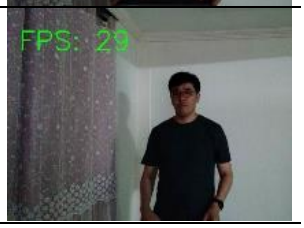
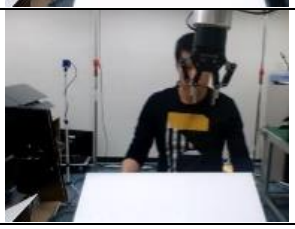
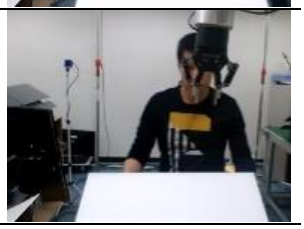
	Bright	Normal	Dark
90 cm			
120 cm			
150 cm			

Table 5 shows the results in a factory environment. We check both frameworks in normal, angle, and occlusion situations. The participant was recorded working about 100 cm away from the camera. In normal situations, both frameworks detect facial landmarks. Of course, facial landmark detection is difficult due to an occlusion blocking of the front face. MediaPipe could detect facial landmarks when the face is on some angles, but DLIB could not.

Table 5. Angle comparison in factory environment.

Type	MediaPipe	DLIB
Normal		
Angle		
Angle		
Occlusion		

4 Conclusions

In this paper, we compared the public frameworks for facial landmark detection. MediaPipe and DLIB are popular frameworks that contain many computer vision parts, such as face detection, pose estimation, or facial landmark detection. We tested each framework in environments of different distances, lights, and angles. The experimental results are summarized as follows:

- It is ideal if the distance to place the camera is a maximum of 120 cm.
- Distance and angle are more important than light.
- DLIB's FPS may drop more depending on the device's capability.
- MediaPipe has an advantage for tracking a face. After recognizing a face, it detects facial landmarks even from a distance and difficult angle.

Our comparative tests determine the capabilities of the frameworks and help choose the proper framework for a specific problem. For example, utilizing MediaPipe at a distance of 120 cm can detect facial landmark in a factory environment.

Acknowledgments. This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (00167198, AI-PRISM).

References

1. A. M. Pascual et al., "Light-FER: a lightweight facial emotion recognition system on edge devices," Sensors, vol. 22, no. 23, p. 9524, Dec. 2022, doi: 10.3390/s22239524.
2. C. Á. Casado and M. López, "Real-time face alignment: evaluation methods, training strategies and implementation optimization," Journal of Real-time Image Processing, vol. 18, no. 6, pp. 2239–2267, Apr. 2021, doi: 10.1007/s11554-021-01107-w.
3. W. Wei, E. S. L. Ho, K. D. McCay, R. Damaševičius, R. Maskeliunas, and A. Esposito, "Assessing Facial Symmetry and Attractiveness using Augmented Reality," Pattern Analysis and Applications, vol. 25, no. 3, pp. 635–651, Mar. 2021, doi: 10.1007/s10044-021-00975-z.
4. MediaPipe. https://developers.google.com/mediapipe
5. DLIB. http://dlib.net

Optimization of secondary users for Energy Efficient CSS over fading channels

Anand Nayyar[1], Nhu Gia Nguyen[2]

[1,2]School of Computer Science, Duy Tan University, Viet Nam.

[1]anandnayyar@duytan.edu.vn, [2]nguyengianhu@duytan.edu.vn

Abstract. This work proposes a unique technique for maximizing the energy efficiency of SUs across fading channels in CRNs. An energy-efficient framework that intelligently selects a subset of SUs from the available pool, considering the fading channel conditions and minimizing the overall energy consumption has been proposed. A fusion center (FC) is responsible for collecting sensing input from the designated SUs in the cooperative sensing system. The efficiency of the suggested optimization framework is shown by simulation results. Compared to conventional schemes, the proposed approach achieves significant improvements in energy efficiency (EE) while maintaining satisfactory detection performance. The results also show the impact of fading channel conditions on the selection of SUs and power allocation strategies.

Keywords: Energy Efficiency (EE), Cognitive Radio (CR), Cooperative Spectrum Sensing (CSS).

1 Introduction

Cooperative spectrum sensing (CSS) involves a group of SUs collaboratively sensing the wireless environment to identify spectrum opportunities. By exploiting the spatial diversity of SUs, cooperative sensing can improve the detection performance compared to individual sensing. However, energy consumption is a critical concern in CRNs, particularly in fading channel conditions where the SUs experience fluctuations in signal quality.

In fading channels, SUs may need to transmit at higher power levels to overcome the adverse effects of fading, leading to increased energy consumption. Therefore, optimizing the EE of SUs during CSS in fading channels becomes crucial to prolong the network's lifetime and enable sustainable operation. Energy-efficient spectrum sensing not only conserves energy resources but also reduces the environmental impact and enhances the reliability of battery-powered SUs.

2 Related Work

The CSS in CRNs has gained significant attention in recent years due to its potential to improve spectrum utilization and mitigate interference to primary users (PUs)

[1-3]. Several studies have focused on optimizing energy efficiency during cooperative spectrum sensing, particularly in fading channel conditions [4-6]. In this literature survey, we review the existing research and identify the key contributions and advancements in the field.

Cooperative spectrum sensing enables SUs to share sensing information and collectively detect the absence of PUs [5-7]. A comprehensive survey by in [8] provides an overview of various CSS techniques, including centralized and distributed schemes. The survey examines the advantages and disadvantages of cooperative sensing, as well as the significance of energy efficiency in CRNs. The CSS approaches use less energy strive to reduce SU energy usage while retaining accurate detection performance [9-10]. In [11], author proposed a centralized energy-efficient cooperative sensing scheme that dynamically adjusts the sensing time and power levels of SUs based on channel conditions. The results showed significant energy savings without compromising detection performance. Fading channel conditions have a substantial impact on the performance of CSS. Various fading channel models, such as Rician, Rayleigh and Nakagami-m, have been utilized in CRN studies. Authors in [12] investigated the effects of fading channels on cooperative sensing and proposed a CSS scheme based on Nakagami-m fading channels. The study highlighted the importance of considering channel characteristics in optimizing energy efficiency. Optimization techniques play a crucial role in achieving energy-efficient CSS [13-14]. Authors in [15] presented a joint optimization framework for power allocation and sensing time in fading channels. They devised an optimization problem in order to improve energy efficiency while meeting detection criteria. The proposed algorithm showed improvements in energy efficiency compared to traditional approaches. Machine learning techniques have also been explored to optimize energy efficiency in cooperative spectrum sensing. In [16], authors proposed a deep reinforcement learning-based framework for energy-efficient cooperative sensing. The framework dynamically adjusts the transmission power and sensing time of SUs based on the feedback from the environment, resulting in improved energy efficiency. Cross-layer design approaches that integrate physical layer characteristics with higher-layer optimization have shown promising results in energy-efficient spectrum sensing. In [17], authors proposed a cross-layer optimization framework that jointly considers power allocation, sensing duration, and data fusion in cooperative spectrum sensing. The study demonstrated significant energy savings while ensuring reliable detection performance. Optimizing resource allocation and SU selection strategies is essential for energy-efficient cooperative spectrum sensing. In [18], authors proposed a joint optimization framework for SU selection and power allocation based on a stochastic geometry approach. The framework considered fading channels and achieved energy-efficient spectrum sensing by selecting SUs strategically.

In this paper, we propose an optimization framework to address the EE challenges in CSS over fading channels in CRNs. The primary objective is to intelligently select a subset of SUs from the available pool and optimize their power allocation to minimize energy consumption while maintaining satisfactory detection performance.

The contributions of this research are significant, as they address the critical issue of energy efficiency in CRNs, specifically in the context of CSS over fading channels.

The proposed optimization framework can extend the lifetime of battery-powered SUs, leading to more sustainable and environmentally friendly wireless networks. Furthermore, the findings can facilitate the practical deployment of cognitive radio systems by enabling more efficient utilization of available spectrum resources.

The rest of the paper is structured as follows: The system model and problem formulation are presented in Section 3. The suggested EE computation for the system model is described in Section 4. Section 5 describes the optimization framework in full, including the SU selection procedure. Section 6 discusses the simulation setup and performance outcomes. Finally, Section 7 closes the work by discussing possible future research possibilities.

3 System Model

In the system model, "N" CR users and a licensed user are assumed. Each SU detects the presence of PU and communicates its findings to the FC. The Total frame time is considered as $\mathrm{T} = \tau_s + \mathrm{N}\tau_r + \tau_d$. Here τ_s is sensing time, τ_r is reporting time and τ_d is the data transmission time. During the data transmission, the licensed user can reoccupy the channel with the probability $P_1(\tau_d) = 1 - \exp(-\tau_d/a_0)$. Here a_0 is the mean value of the busy state of licensed user.

The energy detection mechanism is used by each CR user to find licensed users when they are present. The signal x(t) is sent into an energy detector (ED). With regard to the presence of a licensed user, the output of ED takes the form of a binary choice. The two hypotheses are as follows:

$$s_i(t) \triangleq \begin{cases} n_i(t), & H_0 \\ h_i x(t) + n_i(t) & H_1 \end{cases} \quad (1)$$

where, $s_i(t)$ denotes the received signal at the i^{th} SU $n_i(t)$ denotes the noise signal, and h_i denotes the channel gain. The existence of PU is determined by comparing the received energy with the detector threshold. As a result, the detection and false alarm probabilities for the AWGN channel are as follows: [5]

$$P_f{}^{(i)} = Pr\,[\,E_i > \lambda_i | H_0] = \frac{\Gamma\left(u, \frac{\lambda_i}{2}\right)}{\Gamma(u)} \quad (2)$$

$$P_d{}^{(i)} = Pr\left[\,[\,E_i > \lambda_i | H_1]\right] = Q_{\bar{u}}\left(\sqrt{2\Upsilon}, \sqrt{\lambda}\right) \quad (3)$$

where Υ and u are the channel SNR and the time-bandwidth product, respectively. By averaging Equation (3), the average detection probability under the fading scenario can be determined i.e.

$$\tilde{P}_d = \int_0^\infty Q_u\left(\sqrt{2\mathrm{x}}, \sqrt{\lambda}\right) f_\Upsilon(x)\,dx \quad (4)$$

where $f_\Upsilon(x)$ is the probability density function (PDF) of Υ under fading and $Q(a, b)$ is the Q function.

3.1 Cooperative Spectrum Sensing

The channel between FC and CR users is supposed to be noisy with an error rate P_e and a constant detector threshold λ is used for all the CRs then ${P_f}^i = P_f$ and ${P_d}^i = P_d$. The false alarm and detection probability for this noisy reporting may be calculated as follows:

$$P_{f1} = P_f(1 - P_e) + (1 - P_f)P_e \quad (5)$$

$$P_{d1} = P_d(1 - P_e) + (1 - P_d)P_e \quad (6)$$

According to the presumption, each SU makes a binary choice and communicates one bit of the decision D_i to the FC over the reporting channel. In D_i, bit 0 indicates the lack of PU, whereas bit 1 indicates the presence of PU. The FC uses the following logic rule to fuse all 1-bit choices:

$$Y = \sum_{i=1}^{N} D_i \begin{cases} \geq k, & \mathcal{H}_1 \\ < k & \mathcal{H}_0 \end{cases} \quad (7)$$

where $\mathscr{H}_0$ and $\mathscr{H}_1$ stand for the PU absence and PU presence hypotheses, respectively. The "*k*-out-of-*N*" voting rule is represented by the integer threshold *k*. In this case, $k = 1$ corresponds to the OR fusion rule, $k = N$ to the AND fusion rule, and $k>N/2$. to the Majority fusion rule. This logic rule gives the overall detection and false alert probability as

$$Q_f = \sum_{i=k}^{N} \binom{N}{i} {P_{f1}}^i \left(1 - P_{f1}\right)^{N-i} \quad (8)$$

$$Q_d = \sum_{i=k}^{N} \binom{N}{i} {P_{d1}}^i (1 - P_{d1})^{N-i} \quad (9)$$

4 Energy Efficiency of the proposed model

If probabilities of idle and busy state of PU are P_0 and P_1 respectively, θ_s is sensing energy, θ_t is data transmission energy and *C* is the channel's capacity over the data transmission period. Table 1 shows four different situations based on the findings of spectrum sensing and the state of the PU.

Table 1: Sensing Result Scenario

PU State	SU state	Probability	Energy Consumption (Joule)	Throughput (bits/Hz)
Busy	Detects the busy state of PU correctly	P_1P_d	$N(\tau_s\theta_s + \tau_r\,\theta_t)$	0
Idle	Detects PU state as busy	P_0P_f	$N(\tau_s\theta_s + \tau_r\,\theta_t)$	0
Busy	Detect PU state as idle	$P_1(1 - P_d)$	$N\tau_s\ \theta_s + \tau_d\,\theta_t$	0
Idle	Detect the PU idle state correctly	$P_0(1 - P_f)$	$N\tau_s\ \theta_s + \tau_d\,\theta_t$	$\tau_d\,C$

If reporting duration $\tau_r \ll T$, the normalized sensing time given as $(\tau) = \frac{\tau_s}{T}$ and data transmission duration given as $(\tau_d) = (1-\tau)T$.

Based on the 4 states, the energy consumption $\hat{\mathbb{E}}$ can be determined as

$$\hat{\mathbb{E}} = N\,\tau\,T\,\theta_s + \left(P_0\left(1-Q_f\right) + P_1\left(1-Q_d\right)\right)(1-\tau)T\theta_t \qquad (10)$$

The system's average throughput, $\hat{\mathbb{R}}$, which is the number of valid data bits transferred every frame, may be stated as

$$\hat{\mathbb{R}} = P_0\,\tau_d\,C\left(1-Q_f\right)\left(1-P_1(\tau_d)\right) \qquad (11)$$

The channel capacity can be determined as $C = \beta \log_2\left(1+\frac{\theta_t}{\Gamma}\right)$, where Γ is the channel's noise power over bandwidth β. Hence

$$\hat{\mathbb{R}} = P_0\,(1-\tau)T \times \left(1-Q_f\right) \times exp\left(\frac{-\left((1-\tau)T\right)}{a_0}\right) \times \beta \log_2\left(1+\frac{\theta_t}{\Gamma}\right) \qquad (12)$$

The Energy Efficiency (ξ), the amount of information that may be successfully conveyed per unit of energy cost is calculated as

$$\hat{\xi} = \frac{\hat{\mathbb{R}}}{\hat{\mathbb{E}}} \qquad (13)$$

5 Problem formulation and solution to the design problem

The detector threshold directly affects the EE for a fixed sensing time. By altering the detector threshold, the design problem seeks to maximize the EE. The optimization problem can be described mathematically as

$$\begin{aligned} &\text{To find: } (n_0) \\ &\text{Max.: } \xi \\ &\text{S.t.: } \lambda_{max} \geq \lambda \geq \lambda_{min}, P_d \geq \bar{P}_d \end{aligned} \qquad (14)$$

where n_0 is the optimal CR users, $\bar{P}_d$ is the target detection probability λ_{max} and λ_{min} are maximum and minimum values of detector threshold respectively.

6 Numerical results and discussion

The simulation parameters are taken as $T = 50$ ms, $N = 20$, $f_s = 6$MHz, $\theta_s = 0.2W, \theta_t = 3W$, $\tau_s = 1.5ms$, $\tau_r = 10\mu s$, $\bar{\Upsilon} = 10\ dB$, $u = 5$, $\bar{P}_d = 0.9$, $P_0 = P_1 = 0.5$ Fig. 1 shows the total error probability plotted against the threshold under various fading situations. The SNR remained unchanged at 10dB. According to the results, the overall error probability is lowest for the Nakgami channel and highest for the Rayleigh fading environment. The result also shows that error probability is minimum at threshold value 12 for Nakagami fading channel.

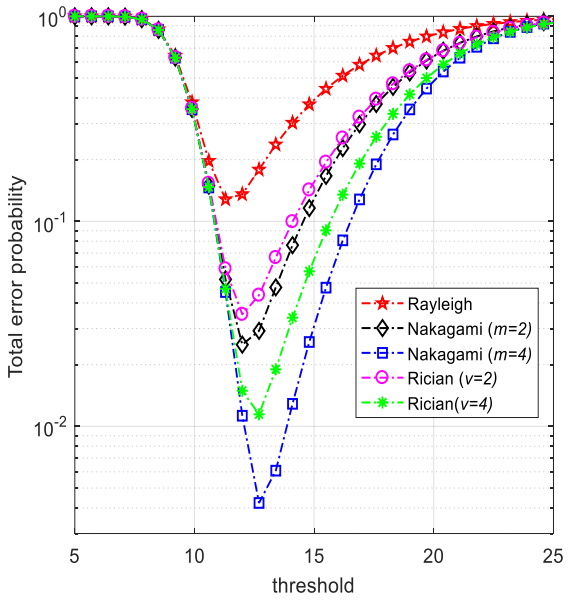


Fig.2: Total error probability vs. threshold over different fading

Fig. 2 depicts the variation of total error probability with respect to SNR across fading channels. Total error probability lowers as SNR grows because radio conditions improve as SNR increases. The Nakagami channel outperforms the other fading channels.

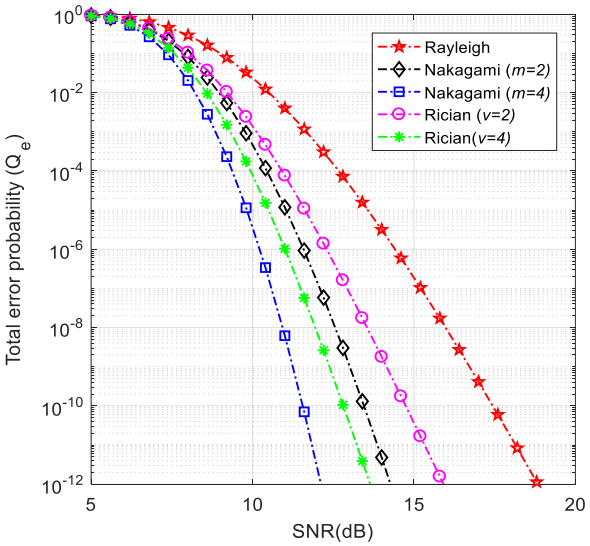


Fig.2: Total error probability vs. SNR over different fading

Total energy consumption is displayed versus the number of secondary users in Fig. 3. Because of the substantial overhead, energy usage rises as the number of secondary users grows. The results also reveal that Nakagami fading channels use less energy than other channels. The discussion makes it obvious that the number of secondary users is an important parameter in order to enhance the overall EE of the system.

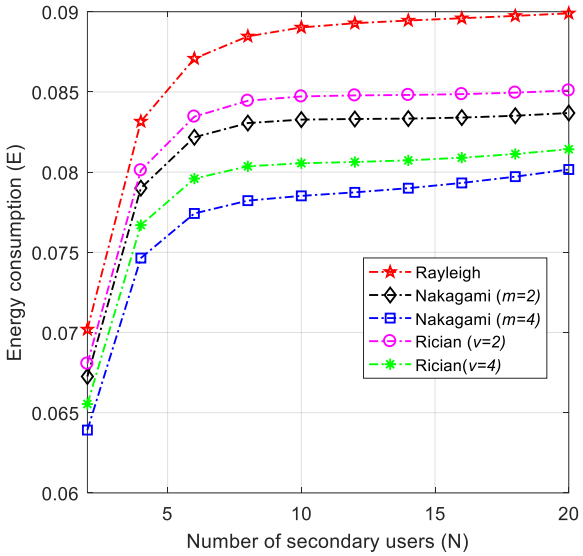


Fig.3: Energy consumption vs. number of secondary users over different fading

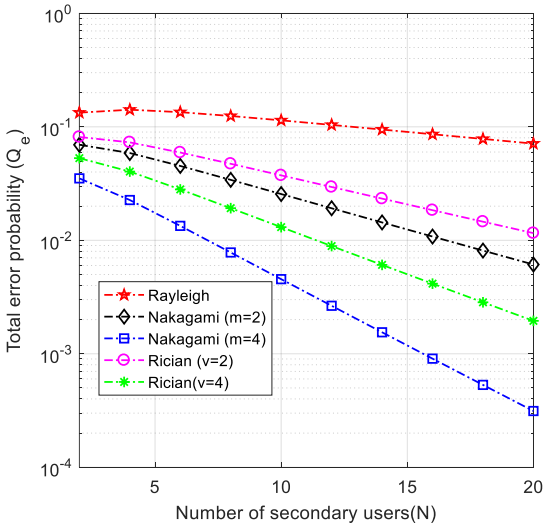


Fig.4: Energy consumption vs. number of secondary users over different fading

Fig. 4 depicts the variation of total error probability as a function of secondary user count across fading channels. The detector threshold is set at 15. It has been discovered that when the number of secondary users rises, the error probability falls due to secondary users' mutual cooperation, and the overall detection improves. It is clear that the error probability is minimum over the Nakagami fading.

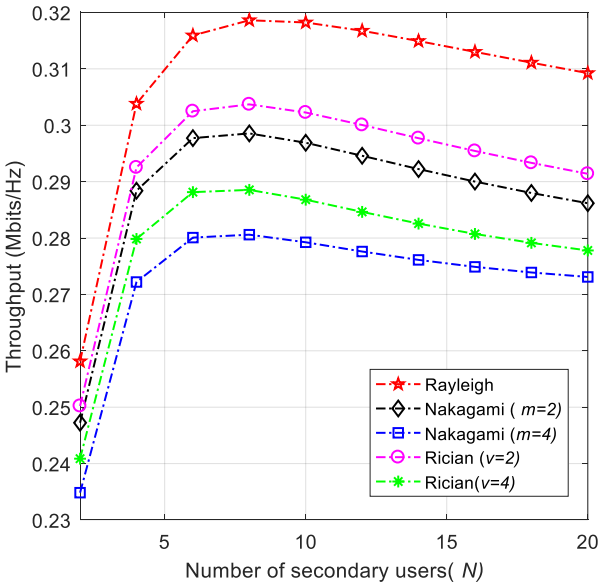


Fig.5: Throughput vs. number of secondary users over different fading

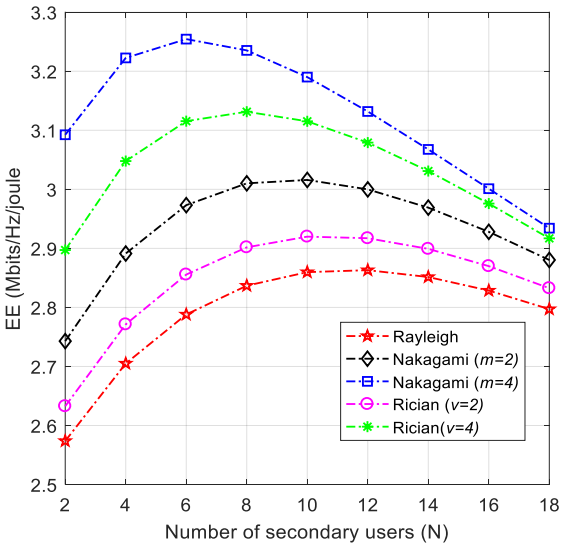


Fig.6: Energy Efficiency vs. the number of secondary users over different fading channels.

In Fig. 5, throughput is shown versus the number of secondary users under various fading circumstances. The study shows that when the number of users rises, the throughput initially increases because user collaboration raises the detection probability. As the number of users grows, the detection threshold begins to fall owing to the high overhead. The finding shows that throughput is a concave function of the number of secondary users, and that there must be an ideal value of secondary users at which throughput is maximal.

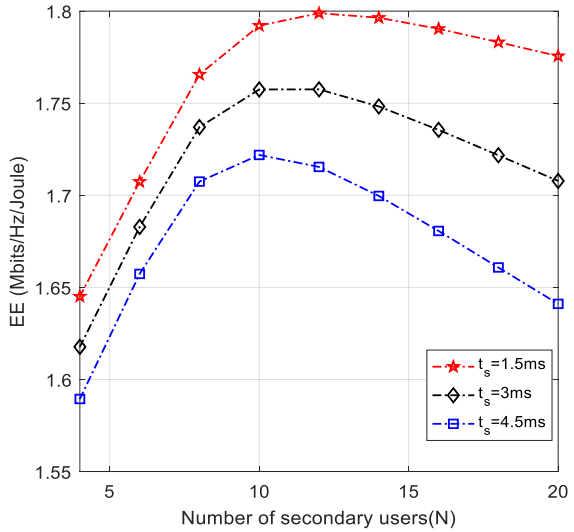


Fig.7: Energy Efficiency vs. number of sensing users at different sensing times

Energy efficiency is displayed versus the number of secondary users across various fading channels in Figure 6. EE is likewise shown to be a concave function of the number of SUs. If the number of secondary users rises, the EE initially increases because detection probability increases owing to user cooperation. When the number of users grows, the EE drops owing to increased overhead energy use.

Figure 7 depicts the change in energy efficiency as a function of the number of sensing users at various sensing periods. The graph is seen via the Rayleigh fading channel. The EE drops as the sensing time rises because the throughput decreases.

7 Conclusion

The primary objective of this paper is to enhance the energy efficiency of the CSS process while maintaining reliable sensing performance. Through a comprehensive analysis of the CSS system and the energy consumption model, several key findings have been derived. Firstly, it was observed that the energy efficiency of the CSS system is highly dependent on the fading channel conditions and the number of participating SUs. By incorporating these factors into the optimization framework, the pro-

posed approach effectively balanced the trade-off between energy consumption and sensing performance. The testing findings showed that the suggested optimization technique considerably increased the CSS system's energy efficiency without reducing sensing capability. Overall, the research presented in this paper contributes to the field of energy-efficient CSS over fading channels by providing a comprehensive optimization framework. The findings highlight the importance of considering fading channel conditions, cooperation among SUs, and adaptive sensing time allocation for achieving energy-efficient CSS. Future work could explore additional optimization techniques and further investigate the practical implementation of the proposed scheme in real-world scenarios

References

1. Z. Qin, X. Zhou, L. Zhang, Y. Gao, Y. Liang and G. Y. Li, "20 Years of Evolution from Cognitive to Intelligent Communications," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 6-20, March 2020
2. J. Lunden, V. Koivunen, A. Huttunen and H. V. Poor, "Censoring for Collaborative Spectrum Sensing in Cognitive Radios," *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2007, pp. 772-776.
3. C. Sun, W. Zhang and K. Ben Letaief, "Cooperative Spectrum Sensing for Cognitive Radios under Bandwidth Constraints," *2007 IEEE Wireless Communications and Networking Conference*, Kowloon, 2007, pp. 1-5.
4. G. Sharma and R. Sharma, "Energy efficient collaborative spectrum sensing with clustering of secondary users in cognitive radio networks," in IET Communications, vol. 13, no. 8, pp. 1101-1109, 2019.
5. Sharma, Girraj, and Ritu Sharma. "Optimised fusion rule in cluster-based energy-efficient CSS for cognitive radio networks." International Journal of Electronics, vol. 106, no. 5, pp 741-755, 2019. DOI: 10.1080/00207217.2018.1553248
6. Sharma, Girraj, and Ritu Sharma. "Performance evaluation of distributed CSS with clustering of secondary users over fading channels." International Journal of Electronics Letters, vol. 6, no. 3, pp. 288-301, 2018. DOI: 10.1080/21681724.2017.1357762
7. Ying Chang L, Yonghong Z, Edward CY, Peh A. Sensing-throughput tradeoff for cognitive radio networks. IEEE Transactions on Wireless Communications, 7(4):1326–1336, 2008
8. Yucek, T., & Arslan, H. (2009). A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE communications surveys & tutorials*, *11*(1), 116-130.
9. Sharma, Girraj, and Ritu Sharma. "Cluster-based distributed cooperative spectrum sensing over Nakagami fading using diversity reception." IET Networks, vol. 8, no. 3. Pp. 211-217, 2019. DOI: 10.1049/iet-net.2018.5002
10. Sharma, G., Upadhyaya, V., Kumar, A., Vyas, S., & Sharma, R. (2022). Fusion Rule Optimisation for Energy Efficient Cluster-Based Cooperative Spectrum Sensing. In *Emerging Electronics and Automation: Select Proceedings of E2A 2021* (pp. 275-284). Singapore: Springer Nature Singapore.

11. Zhou, Z., Zhou, S., Cui, J. H., & Cui, S. (2008). Energy-efficient cooperative communication based on power control and selective single-relay in wireless sensor networks. IEEE transactions on wireless communications, 7(8), 3066-3078.
12. Sun, H., Nallanathan, A., Jiang, J., & Wang, C. X. (2011, August). Cooperative spectrum sensing with diversity reception in cognitive radios. In *2011 6th International ICST Conference on Communications and Networking in China (CHINACOM)* (pp. 216-220). IEEE.
13. Sharma, G., & Sharma, R. (2022). Joint optimization of fusion rule threshold and transmission power for energy efficient css in cognitive wireless sensor networks. *Wireless Personal Communications*, *123*(3), 2107-2125.
14. Sharma, G., Sharma, Y., Upadhyaya, V., Kumar, A., & Sharma, R. (2021). Inter and intra fusion schemes for energy efficient CB-CSS in cognitive wireless networks. *International Journal of Electronics*, *108*(11), 1940-1956.
15. Zhao, N. (2016). Joint optimization of cooperative spectrum sensing and resource allocation in multi-channel cognitive radio sensor networks. *Circuits, Systems, and Signal Processing*, *35*, 2563-2583.
16. Guo, Z., Chen, H., & Li, S. (2023). Deep Reinforcement Learning-Based UAV Path Planning for Energy-Efficient Multitier Cooperative Computing in Wireless Sensor Networks. *Journal of Sensors*, *2023*.
17. Li, H., & Liu, C. (2019). Cross-layer optimization for full-duplex cognitive radio network with cooperative spectrum sensing. *International Journal of Communication Systems*, *32*(5), e3895.
18. Giri, M. K., & Majumder, S. (2022). Deep Q-learning based optimal resource allocation method for energy harvested cognitive radio networks. *Physical Communication*, *53*, 101766.

Enhancing computed tomography image with limited number of shooting angles

Dang Viet Hung, Vo Nhan Van
School of Computer Science, Duy Tan University, Viet Nam.

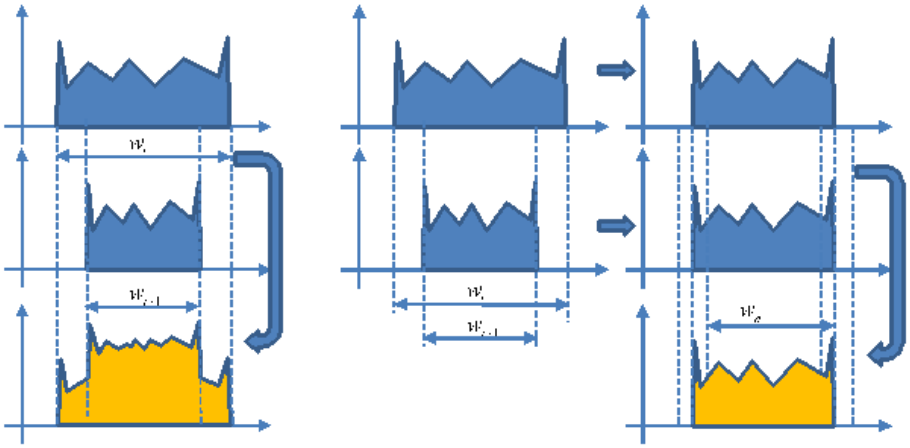
Abstract:

Computed tomography (CT) is a technique of scanning around a biological body part, and the results of the scan are combined to create a tomographic image of that part. This technique allows the doctor to have a view that is perpendicular to the scanning directions, and to visualize all the details as well as positions of internal tissue, muscle, and bone structures without surgery. The image rendering algorithm reconstructed by the back projection algorithm can achieve high accuracy if the shooting angle resolution is high, but it affects the health of the patients because the number of shots is too large. This paper presents a solution to improve the quality of reconstructed images with a lower number of shots. The results show that the solution reduces the number of shots (low resolution) but the tradeoff for quality is not high compared to commonly used methods.
Keyword: Computer Tomology, Radon, Back Projection, Image Enhancing, Inverse Radon

Introduction

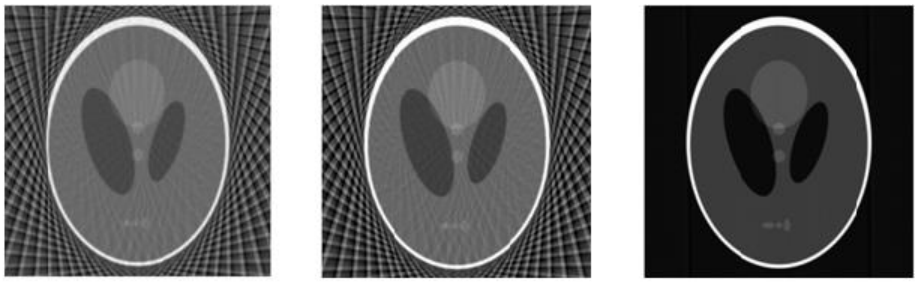
This paper will present a method to enhance the quality of CT images with low input resolution. In which, the main idea is to generate interpolation of intermediate sections, then combine the original images and the interpolated ones to perform the inverse Radon transform. Conventional interpolation techniques often do not achieve good results due to the simple interpolation between the results of two consecutive shooting angles without regard to large displacements in the width of the object. Therefore, our proposed idea is to consider these parameters during interpolation phase, as described in the following Figure 1.

2



(a)Classical interpolation (b) Proposed interpolation.

Figure 1. Proposed idea of interpolation before Radon inversion.adjusts



Classical interpolation with Linear filter (a)

Classical interpolation with Spline filter (b)

Proposed interpolation with linear filter (c)

Figure 2. Image results of applying (a), (b) classical interpolation and (c) proposed interpolation.

Conclusion

This paper briefly presents the Radon transform and the inverse Radon transform, which play an important role in constructing CT images from different X-ray angles. However, in order to obtain high-quality CT images, the number of angles must be large, resulting in the operator having to endure a large number of x-rays passing through the body. We have proposed a solution to improve the quality of CT images in

the interpolation step from two images taken with two consecutive shooting angles. This interpolation step deals with of the sudden displacement of the object width and position, thus giving better interpolated image quality. This results in higher CT image quality, as evidenced both analytically and experimentally. In the future, we will develop interpolation techniques using Machine Learning to further improve the quality of CT images with lower number of images.

References:

[1] F. Natterer, The mathematics of computerized tomography, John Wiley & Sons Ltd., Chichester, England 1986.

[2] F. Natterer, F. Wuebbeling, Mathematical Methods in Image Reconstruction, SlAM Monographs on Mathematical Modeling and Computation, ISBN:0- 89871472-9, Philadelphia, PA, USA, 2001.

[3] X. Lu, Y. Sun, G. Bai, "Adaptive wavelet-Galerkin methods for limited angle tomography", Image Vis. Comput. vol. 28, no. 4, pp.696–703, 2010.

[4] Nobumichi Yasunami2, Tatsuya Yatagawa, "Degree of local symmetry for geometry-aware selective part visualization on CT volume data", In the proccedings of the 11th Conference on Industrial Computed Tomography (iCT), Wels, Austria (iCT 2022), 2022.

[5] A. Slyamov1, J. Kim, M. Pedersen, K. Nielsen, A. Pedersen, T. Ramos1, M. Kagias, E. Lauridsen, "Towards lab-based X-ray scattering tensor tomography with circular gratings", In procceding(s) of the 11th Conference on Industrial Computed Tomography (iCT), Wels, Austria (iCT 2022), 2022.

MULTIMEDIA
LIFE
STORAGE
NETWORK
DATABASE
SYSTEM
BIG DATA
SCIENCE
CLOUD
TREND
CLUSTER
ANALYSIS
BUSINESS
SOCIETY
GRAPHICS
VISUALIZATION

THE KOREA
BIG DATA SERVICE SOCIETY
한국빅데이터서비스학회

N13 404-2, Chungbuk University, Chungdae-ro 1
Seowon-Gu, Cheongju, Chungbuk 28644, Korea

Email: kbigdataservice@gmail.com

Homepage: www.kbigdata.or.kr