Exploring Methodologies to Increase the Reliability of Data for Verifying Nuclear Power Plant Artificial Intelligence Software.

Da-Young Lee*, Han-Gil Lee, Jeong-Hun Lee
FNC TECHNOLOGY CO.,LTD / 13 Heungdeok 1-ro, 32F, Giheung-gu, Yongin-si, Gyeonggi-do, 16954, Korea
**Corresponding author: ekdud93@fnctech.com*

1. Introduction

The impact of artificial intelligence technology on the industry is gradually increasing, and as a result, the technology readiness level is getting higher.

Software used in the nuclear field must be verified and validated to ensure reliability. Therefore, artificial intelligence should also have sufficient reliability through the same manner.

In this paper, we conducted a study on software verification to apply artificial intelligence-based software to MMIS, and explored a methodology that can improve the explanability of data through related cases.

2. Background

So far, it is difficult to find cases where AI technology has been applied to nuclear power plants, satisfying regulatory requirements for incorporating AI into the operational support functions of nuclear reactors. On the other hand, in order to apply artificial intelligence to the MMIS of the APR1400, it is necessary to have a minimum grade of ITA (SIL level2 according to IEEE 1012 standards).[1]

In this paper, we aim to utilize the verification methodology of artificial intelligence software presented in the LIME(Local Interpretable Model-agnostic Explanation) model and TTA (Telecommunications Technology Association) Journal No. 201[2] for software testing and verification. The followings are examples of the methodology.

2.1 Explainable artificial intelligence methodology

The first methodology is an approach that explains the area of the black box.

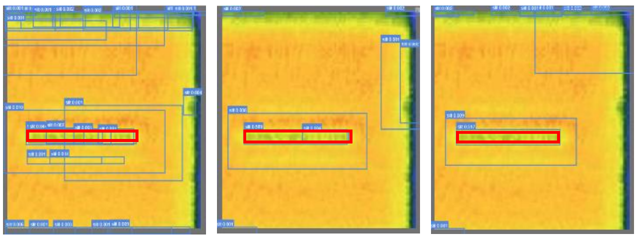

Fig 1. Detection Data-1

Fig1. shows the process of artificial intelligence detecting defects in self-produced nuclear reactor components, using the Yolov7 model with CNN (Convolutional Neural Networks) algorithm applied for data training.[7] The training process progresses from the left image to the right image in Fig1, and in the leftmost image, multiple bounding boxes are detected by AI detection, indicating low performance. As the process moves to the right, the number of bounding boxes decreases, and it can be confirmed that they become closer to the correct area in the center. By visualizing the black box area in this way, it is possible to enhance the explanatory power for quality verification of the artificial intelligence module.

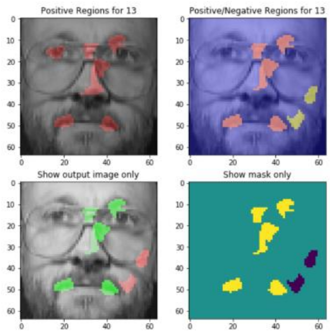

Fig 2. Detection Data-2

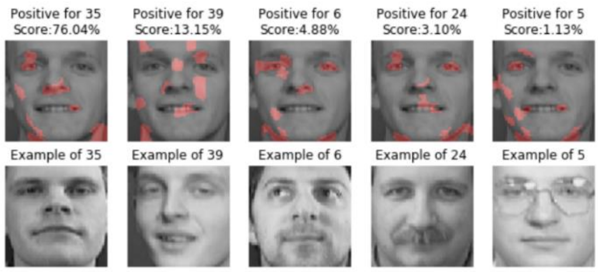

Fig 3. Detection Data-3

Fig 2. and Fig 3. are cases of verifying the reliability of artificial intelligence through models designed with a focus on eXplainable Artificial Intelligence (XAI).[6] The Local Interpretable Model-agnostic Explanations (LIME) of XAI analyzes the process interpreted by the black box by focusing on one data point. This analysis technique finds the superpixels that the model is most influenced by, by transforming the superpixels in the training data, and Fig 2. shows the results of the superpixel output after image data learning of the model. Fig 3. is a case of misclassification where the highlighted part of the person image in the first row is judged to be similar. By visualizing and diagnosing misclassified data in the black box area, we can confirm the characteristics of data that are important from an artificial intelligence perspective.

2.2 Appropriate Data Balance Methodology

The second approach is to pursue the reliability of the data used for AI learning itself. When designing AI-based software, the total amount of training data is important, but from the perspective of handling "various scenarios" in the real world, it is necessary to evaluate the reliability level using a dataset with a balanced distribution.[2] If the balance of the training data is sufficient, we can confirm the AI's stable performance with the following example.

Fig 4. Detection Data-4 [2]

The left image of Fig 4. shows the result of detecting "Load Truck" when trained with 50,000 general data, and the right image shows the result trained with 2,000 data with appropriate data balance. In the left image of Fig 4, there are difficulties in discrimination because multiple bounding boxes are detected, but in the right image, only one clear bounding box is detected, confirming the improved performance of the AI. Through this, it can be seen that a dataset with balanced data is significantly better in learning results than countless datasets without balance in AI data training.

The first methodology is widely researched in other fields, so in this paper, we will focus on examining the methodology of verifying the reliability of artificial intelligence software from the perspective of data balance.

3. Designing AI data balance for NPP

To ensure data balance, appropriate dataset design is necessary. This can be considered as one of the design phase tasks in IEEE 1012. Therefore, the design phase of IEEE 1012 is applied to the dataset design process.

At this point, our goal is to apply software design methodology to the design of necessary data balance. Therefore, we consider the software design phase tasks proposed in IEEE 1012 (*refer to Table 1) along with the design requirements for data balance to achieve our goal of improving the reliability of artificial intelligence.

SDLC(System Development Life Cycle)	Deliverables
Software Design	SDD, CT/IT Plan, CT/IT/ST Procedure, RTM, V&V Report

Table 1. IEEE 1012 Software Design Phase Output

We used the method presented in reference [3] to consider the design requirements of data balance in the software design phase. In reference [3], the following design procedure was applied as a minimum evaluation item to ensure balance in the dataset. This design procedure and method applies the techniques for each step of the specification-based test design method for software functional safety verification, which is a group standard of the Telecommunications Technology Association (TTA) in Korea.[4]

a) Contraction of Cause and Effect
b) Classification of Effect Group
c) Description of the Cause-Effect Graph
d) Cause-Effect Graph Inspection
e) Making an Assessment Scenario
f) Selecting the Base Assessment Scenario
g) Inducing Assessment Data
h) Drawing up Assessment Case

The number of evaluation items using the applied techniques above for each step corresponds to the EDC parameter defined for the analysis of artificial intelligence reliability as follows.

No	Testing Items	means
1	Evaluation Data Meta Attribute Count, (MC)	Meaning property complexity of data
2	Evaluation Data Meta Attribute Value Count, (MVC)	Indicates the complexity of the data attribute value
3	Balanced Evaluation Dataset Count, (EDC)	Minimum number of assessments required to balance data
4	Evaluation Applied Dataset Count, (AC)	Number of datasets used to measure the accuracy of actual AI-based software
5	Coverage of Evaluation Applied Dataset, (CAD)	This refers to the ratio of the actual data set to the total data set required for the evaluation

Table 2. Data Quality Metrics

The data reliability using the metrics in Table 2 can be calculated as follows.

A: The reliability of artificial intelligence software using a balance-based dataset

B: Accuracy from the perspective of an artificial intelligence model (mAP: mean Average Precision)

C: The number of datasets used in the actual evaluation. (AC)
D: The minimum number of datasets required for a balanced evaluation (EDC)

$$B = \text{mAP(Mean Average Precsion)} = \frac{1}{n}\sum_{i=1}^{N} AP_i \qquad (1)$$

$$A = B*(C/D)*100 \qquad (2)$$

The paper[5] was referred to as the basis for the utilization of mAP, and as the ratio of the number of datasets used for evaluation of the minimum evaluation dataset(EDC) to the number of datasets applied for evaluation(AC), increases in equation(2) the reliability is expected to improve. As a result, come to a conclusion to test case problem coverage.

4. Conclusion

In order to apply artificial intelligence technology to support the operation of nuclear power plants, we have introduced methodologies for software reliability verification, considering the black box characteristics of AI and the data quality aspect. We have also explored the design criteria for reliable data balance, which meets the regulatory requirements of the MMIS that AI software aims to achieve.
In the future, we aim to continuously research ways to improve the reliability of artificial intelligence through the process of expanding test data coverage.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2022-00144521).

REFERENCES

[1] IEEE, “IEEE Standard for Software Verification and Validation”, IEEE Std. 1012, 2004
[2] 박지환, “인공지능 소프트웨어 신뢰성 평가 방법”, TTA Journal 201st edition, 2022. 5
[3] TTA, A Method for Evaluating the Reliability of Artificial Intelligence Software Based on the Balance of the Validation Dataset - Part 1: Methodology and System, 2020.12
[4] TTA, “Test Design Method based on Specification for Software Functional Safety Verification”, 2018
[5] 정혜정, “Artificial Intelligence software evaluation plan ”, Advanced Industrial SCIence, 2022.1
[6] Marco Tulio Ribeiro, “Why Should I Trust You? Explaining the Predictions of Any Classifier”, University of Washington Seattle, 2016.8
[7] Chien-Yao Wang1, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”, 2022