Diagnosis-based prediction of abnormal situations using multi-task learning in nuclear power plants

*Jiyong Lee[a], Hyojin Kim[b], Jonghyun Kim[a]**
[a]Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology (KAIST) 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
[b]Department of Nuclear Engineering, Chosun University, 10 Chosundae 1-gil, Dong-gu, Gwangju, Republic of Korea, 61452
**Corresponding author: jonghyun.kim@kaist.ac.kr*

***Keywords : Transformer, Multitask learning, Prediction, Diagnosis, Nuclear power plant**

1. Introduction

Nuclear power plants (NPPs) are critical infrastructure for large-scale energy production, where ensuring safety and reliability is of utmost importance. Various systems, such as the reactor control system and plant protection systems, along with their components, degrade over time, increasing the likelihood of abnormal conditions. When a such condition occurs, operators must quickly identify the symptoms and take appropriate actions to restore normal conditions [1]. However, this process places significant psychological stress on operators, increasing the potential for human error. These errors can lead to severe consequences, such as unexpected reactor trips, highlighting the importance of predicting and preventing these situations in advance.

Current technology allows the development of predictive models for a single event scenario. However, traditional predictive models only consider a single abnormal scenario at a time, and they are unable to predict multiple scenarios simultaneously. This limitation requires the development of separate predictive models for each scenario, which significantly increases computational costs, as each model only operates effectively within its specific scenario. Creating a model capable of accurately predicting multiple abnormal scenarios in a single model is a highly challenging task. Each scenario involves different variables that are intricately interwoven, and a model optimized for a single scenario may perform poorly when applied to multiple scenarios. Therefore, there is a need for a new approach that can simultaneously consider different abnormal scenarios.

This paper proposes a diagnosis-based Multi-Task Learning (MTL) prediction algorithm utilizing the Transformer mechanism. The key advantage of the Transformer lies in its ability to effectively capture complex dependencies within the data. By leveraging this strength, the Transformer enables the simultaneous execution of diagnostic and predictive tasks within the MTL framework. Furthermore, the diagnostic results obtained during this process are then fed back into the prediction process, allowing for the accurate prediction of various abnormal scenarios.

The objective of this study is to develop a model that predicts abnormal conditions in NPPs by utilizing MTL. This model incorporates diagnostic outcomes into predictive tasks, enabling the prediction of multiple abnormal scenarios. Through this approach, the model aims to improve the accuracy of predictions in complex situations, thereby enhancing the safety of NPPs.

2. Method

To predict various abnormal scenarios in nuclear power plants (NPPs), this study employs a multitask learning (MTL) framework. MTL enables simultaneous learning of related tasks by sharing information and leveraging common features across them. This approach improves the overall model performance by allowing knowledge gained from one task to benefit others.

The MTL framework is used for two main tasks: (1) diagnosing the current system state and (2) using the diagnosis to predict future states. Each task shares a common network backbone, enabling the model to learn generalized features while maintaining task-specific outputs. We used a shared encoder network with task-specific output layers to ensure that the model could leverage shared knowledge across tasks while focusing on the unique aspects of each.

To achieve this, this section will describe the transformer mechanism used within the MTL framework, while section 3 will present the MTL-based abnormality prediction process utilizing the transformer mechanism.

2.1 Transformer Encoder

The transformer, based on the attention mechanism, consists of multiple identical layers, each with multi-head self-attention layer, feed-forward neural network (FFNN) and residual connections and layer normalization (Add & Norm). Residual connections and layer normalization are applied to improve performance and mitigate the vanishing gradient problem.

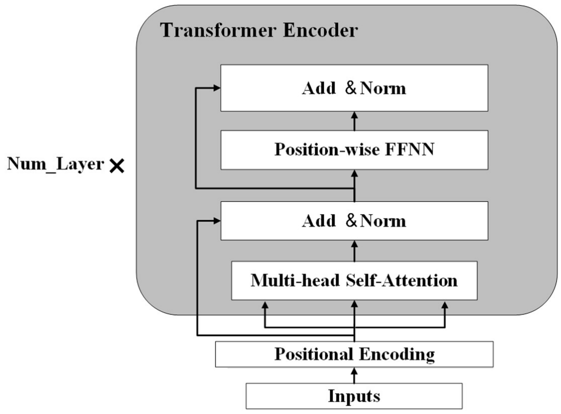


Fig. 1. The architecture of the transformer encoder

2.1.1 Positional Encoding

Since the transformer model processes input data in parallel rather than sequentially, PE is used to introduce sequence order. To encode positional information, the model uses two functions: the sine and cosine functions. These functions generate oscillating values that are added to the embedding vectors, thereby embedding the sequence order into the model. According to Eq. (1), the sine function is applied to even indices of the embedding vector dimensions, while the cosine function is applied to odd indices.

$$(1)\quad PE_{(pos,2i)} = \begin{cases} \sin\left(\frac{pos}{10000^{i/d_{model}}}\right) & (i = 2k) \\ \cos\left(\frac{pos}{10000^{(i-1)/d_{model}}}\right) & (i = 2k+1) \end{cases}$$

where pos and i represent the rows and columns of the embedding vector, d_{model} is the output dimension of all layers in the transformer.

2.1.2 Multi-head Self-Attention

Multi-head self-attention is a core mechanism of the transformer model, enabling the effective learning of complex relationships within an input sequence. Each attention mechanism transforms the input data into three vectors: Query, Key, and Value, using learnable weight matrices (W) [2]. The attention scores, computed as the dot product of Query and Key vectors, represent the relevance of each word in the sequence, as shown in Eq. (2). These scores are then converted into probabilities via softmax, determining the focus each word should receive within the sequence. When this process is extended to multiple heads, as shown in Eq. (3), several attention heads independently perform attention, and their results are concatenated to produce the final output. Each head is defined as in Eq. (4), where each head independently transforms the Query, Key, and Value matrices using their respective weight matrices, followed by performing attention. After concatenating the outputs of all heads, a final linear transformation is applied using the W^O matrix to obtain the final output.

$$(2)\quad Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

In this equation, Q is the query matrix, K is the key matrix, V is the value matrix, and d_k is the dimension of the key vectors used for scaling.

$$(3)\quad Multihead(Q,K,V) = Concat\ (head_1, head_2, \dots, head_h)W^O$$

where:

$$(4)\quad head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

Here, QW_i^Q, KW_i^K and VW_i^V are the weight matrices associated with the query, key, and value matrices for each head, and W^O is the weight matrix applied after concatenating the outputs of all heads. The number of attention heads is denoted by h.

3. Development of Diagnosis-based Prediction Algorithm

The overall structure of the diagnosis-based prediction algorithm is depicted in Fig. 2. This framework is composed of three main stages: 1) Preprocessing, 2) Transformer encoder, and 3) Prediction. Each stage plays a critical role in ensuring the accuracy and reliability of the final prediction results.

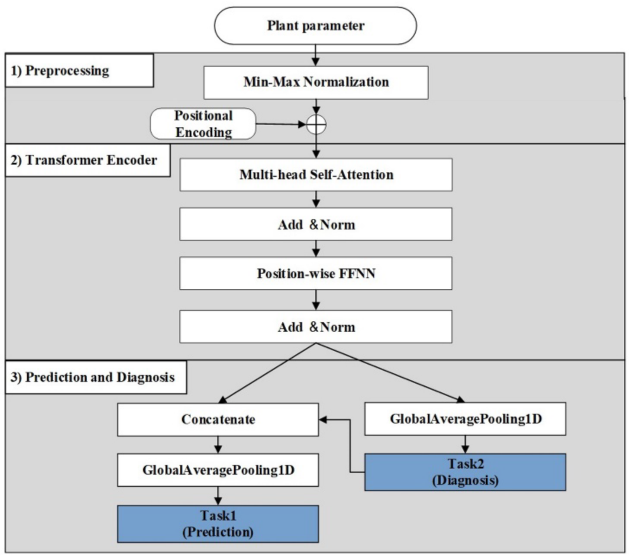


Fig. 2. Process of diagnosis-based prediction function using MTL.

3.1. preprocessing

The first step in this algorithm involves processing the plant parameters to make them suitable for use as inputs to the network. The input data for the network is typically adjusted to have a value range between 0 and 1, which is achieved through Min-Max normalization. This normalization process helps prevent larger values from disproportionately influencing the network's output and enhances the stability and speed of the training process. According to Eq. (7), the input data x_i is normalized using the minimum x_{min} and maximum x_{max} values of the dataset, resulting in normalized values within the range of 0 to 1. After normalization, the data is embedded into a high-dimensional space, and PE is applied to incorporate positional information, enabling the Transformer model to recognize the order of the sequence.

$$(5) \quad x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

3.2. Transformer Encoder

The core of the model is the transformer encoder, which includes several important components such as multi-head self-attention, Add & Norm, and FFNN. The multi-head self-attention mechanism enables the model to simultaneously attend to different parts of the input sequence, capturing complex dependencies within the data. The Add & Norm layers, incorporating residual connections and layer normalization, ensure that critical information is retained and that the model remains stable during training. The position-wise FFNN further processes the data, transforming it into representations that are more suitable for prediction tasks.

3.3. Prediction and Diagnosis

The final stage involves making predictions based on the processed data. The output from the transformer Encoder is passed through two branches: Task 1 (Prediction) and Task 2 (Diagnosis). Task 1 focuses on predicting future trends, while Task 2 utilizes diagnostic results to improve the accuracy of these predictions by incorporating the diagnostic outcomes into the prediction process. The outputs from both tasks are aggregated through GlobalAveragePooling1D layers, ensuring that the model provides comprehensive and reliable predictions.

4. Experiments

4.1. Data collection

The proposed algorithm was implemented using the 3KEYMASTER simulator, which is based on a 1400 MWe pressurized water reactor. For diagnostic-based anomaly prediction, four abnormal scenarios were selected: Steam generator tube leak, Containment spray system malfunction, and Condenser tube leak, Letdown leak. All simulations were initiated from a 100% full-power operation, and data were collected at one-second intervals for a duration of one hour. Each scenario introduced the fault 10 minutes after the simulation commenced. Table I illustrates the scenarios used for the development of the diagnosis-based MTL model.

Table I: Acquired abnormal scenarios

No.	Scenario	Number of Data
1	Steam generator tube leak	20
2	Containment spray system malfunction	7
3	Condenser tube leak	42
4	Letdown leak	52
Total	-	121

4.2. Results

In this study, 10 trip parameters from the reference plant were selected as the outputs of the algorithm, and the developed MTL model utilized 60 input sequences and 561 input parameters as inputs to the network. The model was designed to predict 40 minutes of operation by setting the prediction time length to 240 steps (i.e., 2,400 seconds) with the aim of achieving accurate predictions. Notably, the model was configured to incorporate the results of the diagnostic task into the predictive task.

The model's performance was evaluated using accuracy, mean squared error (MSE), mean absolute percentage error (MAPE), and the coefficient of determination (R-square, R^2), as shown in Table II. Accuracy was used as the evaluation metric for the anomaly diagnosis task (Task 1), where higher accuracy indicates better model performance. Meanwhile, MAPE, MSE, and R^2 were employed to assess the variable state prediction task (Task 2). A lower MSE and an R^2 value closer to 1 indicate superior model performance. These MSE and R^2 values were calculated using Eq. (6) through (9).

$$Accuracy = \frac{100}{n}\sum_{i=1}^{n}(Y_i^{real} = Y_i^{pred}) \tag{6}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\left|Y_i^{real} - Y_i^{pred}\right|\right)^2 \tag{7}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i^{real} - Y_i^{pred}}{Y_i^{real}}\right| \tag{8}$$

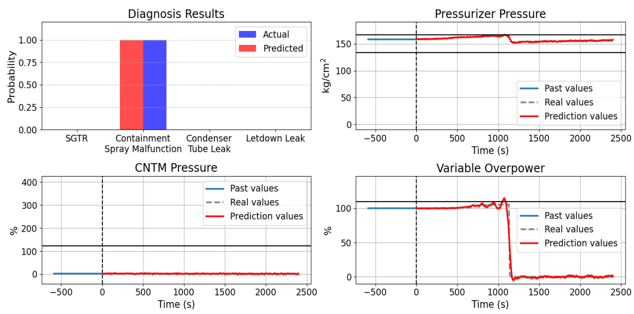
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i^{real} - Y_i^{pred})^2}{\sum_{i=1}^{n}(Y_i^{real} - \overline{Y})^2} \tag{9}$$

where Y_i^{real} and Y_i^{pred} are real and predicted values, respectively. $\overline{Y}$ represents the mean values of variable, and n is the number of samples.

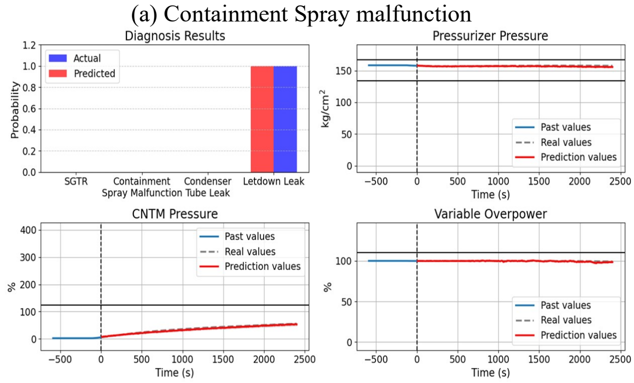
Table II: Diagnosis and Prediction performance of MTL

	Accuracy	MSE	MAPE	R^2
Diagnosis	0.992%	-	-	-
Prediction	-	0.0014	3.018	0.999

The trained model was evaluated using 10 test scenarios. The diagnostic task achieved an accuracy of 0.991%, while the predictive tasks yielded a MAPE of 3.216%, a MSE of 0.0016, and an $\boldsymbol{R^2}$ value of 0.984. Typically, a MAPE below 10% is considered to indicate strong predictive accuracy [3]. These results confirm that the proposed algorithm successfully predicted the trip parameters (i.e., 10 parameters and 240 steps) over a 40-minute period. Fig. 3 presents the outcomes of the diagnostic test, including uncertainty estimation, and the 40-minute prediction results for 10 trip parameters. The black and blue lines denote the trip setpoint and historical trends of the trip parameters, respectively. The red and orange lines show the algorithm's predicted results and the actual trends in the test scenarios, respectively. The light gray shaded regions represent the 95% confidence intervals for the predicted values



(a) Containment Spray malfunction



(b) Letdown Leak

Fig. 3. Diagnose and Predicted results

5. Conclusions

In this study, we developed a model to predict abnormal situations in NPPs using MTL. By integrating diagnostic results into the prediction task, this model enables the prediction of various abnormal scenarios within a single model. This demonstrates that the model can perform reliable and accurate predictions even in abnormal situations in nuclear power plants. This research has the potential to make a significant contribution to enhancing safety by predicting abnormal situations in advance.

The key contributions of this study are as follows. First, we developed a model capable of predicting multiple scenarios through MTL, overcoming the limitations of existing models that focus on predicting a single scenario. Second, we demonstrated that integrating diagnostic results into the prediction task improves prediction accuracy.

In future research, we plan to expand and validate the model's performance by incorporating additional abnormal scenarios into the learning process.

ACKNOWLEDGMENT

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (grant number: RS-2022-00144042).

REFERENCES

[1] Corcoran, W. R., & Zimmerman, B. F. (2006). The role of safety culture in preventing accidents in nuclear power plants. Nuclear Safety, 47(1), 1-11.

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *In Advances in neural information processing systems* (pp. 5998-6008).

[3] H. Kim, J. kim "Long-term prediction of safety parameters with uncertainty estimation in emergency situations at nuclear power plants." Nuclear Engineering and Technology. (2023).