



Fault Detection in Nuclear Power Plants using Deep Learning based Image Classification with Imaged Time-series Data

Y. Shi, X. Xue, J. Xue, Y. Qu

Yong Shi

1. School of Economics and Management, University of Chinese Academy of Sciences
Beijing 100190, China
2. Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences
Beijing 100190, China
3. Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences
Beijing 100190, China
4. College of Information Science and Technology, University of Nebraska at Omaha
Omaha, NE 68182, USA

Xiaodong Xue

1. School of Economics and Management, University of Chinese Academy of Sciences
Beijing 100190, China
2. Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences
Beijing 100190, China

Jiayu Xue*

1. School of Computer Science and Technology, University of Chinese Academy of Sciences
Beijing 101408, China
2. Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences
Beijing 100190, China
3. Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences
Beijing 100190, China

*Corresponding author: xuejiayu18@mailsucas.ac.cn

Yi Qu*

1. School of Economics and Management, University of Chinese Academy of Sciences
Beijing 100190, China
2. Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences
Beijing 100190, China
3. Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences
Beijing 100190, China

*Corresponding author: quyi17@mailsucas.ac.cn

Abstract

Fault detection is critical to ensure the safely routine operations in nuclear power plants (NPPs), requiring very high accuracy and efficiency. Meanwhile, the rapid development of modern information technologies have profoundly changed and promoted various sectors including nuclear industry. Inspired by the great progress and promising performance of deep learning based image classification recent years, a two-stage fault detection methodology in NPPs has been proposed in this paper. First the time-series data describing the operating status of NPPs have been transformed into two-dimensional images by four methods, preserving the time-series information in images and converting the fault detection problem into a supervised image classification task. Then four specific image classifying models based on three primary deep learning architectures have been separately experimented on the imaged time-series data, achieving excellent accuracies. Further the performances of different combinations of transforming means and classifying models have been compared and discussed with extensive experiments and detailed analysis of throughput for four transforming methods. This methodology proposed has obtained remarkable results by reshaping data format and structure, making image classifying models applicable, which not only efficiently detect and warn possible faults in NPPs but also enhances the capability for safety management in nuclear power systems.

Keywords: fault detection, nuclear power plants, deep learning, image classification, imaged time-series data.

1 Introduction

Nuclear energy has become a strategic source of global power supply and is essential to national power security. As a well-known green energy, it could provide stable and huge amount of electricity with many advantages like no pollution, low operating cost, low carbon emission, etc. According to the annual report of *World Nuclear Association*, the electricity generated by nuclear power has accounted for more than 10% in global electricity supply and it's still growing, which has already been a significant component. A common sense is that nuclear energy is probably the only option that could secure adequate energy consumption but doesn't trash this planet, by its many great strengths in dealing with problems from growing energy demand, global carbon reduction, climate change and environmental protection. However, nuclear power systems have one particular specialty as huge disaster might happen once there's a small accident occurred, and that's precisely why **safety management** has always been the top priority in daily operations of NPPs. Therefore, it's definitely necessary to detect the faults (or failures) in NPPs accurately and effectively, then make warnings in advance so that these troubles could be addressed and eliminated in a timely manner. And this subject **fault detection and diagnosis** has become a common focus for both academia and industry.

Original fault detection in NPPs was simple and easy to implement, normally using the "pre-set values" as standard criteria. Once the parameters on meters are beyond the pre-set values or range, then it's assumed the fault has occurred, with voting scheme sometimes been introduced to avoid false alarms [18]. However, with the rapid development of information technology especially artificial intelligence (AI), advanced and intelligent tools represented by machine learning have infiltrated many areas recent years. This has resulted in more intelligent transformations in real industries, also the data-driven fault detection and diagnosis has become a hot research topic of the intelligent cybernetics in the big data era [53]. Former studies of fault detection and diagnosis can be mainly divided as two parts according to the chronological development and they are "model-driven" methods like statistical models, "data-driven" methods like machine learning models. Reasons for the transition from "model-driven" to "data-driven" are the tremendous expansion in data and its complexity, with the great promotion in computational capacity [18] [42]. These two points above have also dramatically accelerated the progress of intelligent techniques especially machine learning and deep learning models, generating a large variety of intelligent approaches and enabling deep learning to make a significant influence in fields such as image classification, voice recognition, natural language processing, etc.

Motivated by the outstanding performances of deep learning in image processing, we propose a two-stage fault detection methodology in NPPs, which is a supervised classification task under this scenario. Due to the lack of real images from NPPs, we also consider to transform the numeric time-series data into images with four major methods to make deep learning based image classifying models

applicable. The fault detection methodology proposed in this paper has obtained remarkable results by reshaping data format and structure, with the results that it's capable of efficiently detecting and warning possible faults in NPPs and further enhancing the capability of safety management in many power systems. The major highlights of this research are described as follows:

- The time-series data describing the operating status of NPPs have been transformed into two-dimensional images, preserving the time-series information as pixel information in images and making deep learning based models applicable.
- Four specific image classifying models based on three primary deep learning architectures have been separately experimented on the imaged time-series data and it has achieved excellent accuracies.
- Performances of different combinations of transforming means and classifying models have been compared and discussed through extensive experiments, with detailed analysis of throughput for four methods.

The rest of this paper will be arranged as follows: Section 2 will briefly review the most related work previously, mainly focused on fault detection in NPPs, data transformation and deep learning based image classification. Section 3 will specifically illustrate our proposed two-stage fault detection methodology in NPPs. Data, experiments and results will be described in Section 4 with detailed analysis and finally, the conclusion will be presented in Section 5 as well as limitations and future directions.

2 Related work

In this section, related work will be briefly reviewed and they could be categorized into three major parts. They are **fault detection in NPPs** of the exact subject, **data transformation** and **deep learning based image classification** concerning with the two-stage fault detection methodology proposed later. Focusing on these three subjects, this section will outline and introduce the most related cases previously.

2.1 Fault detection in NPPs

Fault detection and diagnosis is a multidisciplinary task involving several fields like engineering, management science, information technology, etc, also it could be regarded as an information technology & decision making problem. Here the original and ancient research and practice of fault detection in NPPs will not be mainly discussed, but rather cases of intelligent "data-driven" techniques recent years, which can be categorized into two primary sections, the machine learning ones including the well-known artificial neural networks (ANN), support vector machines (SVM), ensemble methods and the deep learning applications like convolutional neural networks (CNN). These applications are mostly supervised learning tasks and have demonstrated that it's feasible and promising for the implementation of deep learning based image classifying techniques.

Machine learning. To support the strategic decision-makings in nuclear fuel management, regression model has been used to estimate the composition of spent nuclear fuel [16]. Logistic regression (LR) also was applied in fault diagnosis of NPPs [2], with Naive Bayes (NB) introduced into fault diagnosis of thermocouple [4] and artificial cognitive system for recognizing current status of nuclear reactors [34] respectively. The powerful ANN has shown great adaptability by being applied in control rod drive system and accident prevention system [1], also in conceptual accident prevention system for light water reactors [35], while the SVM used for designing more accurate identification and diagnosis of faults [52] and detecting loose components in the primary circuit cooling system of NPPs [30]. The classical model decision tree (DT) has been used to assess the probabilistic risk of precursor events, with the aim of reducing operational risk in NPPs [44] or accurately and quickly identifying possible root causes of faults [33]. As a widely-used ensemble learning approach, XGBoost with extracted features has performed effective fault detection of screen cleaners in NPPs [6] and another representative

ensemble learning algorithm, random forests (RF), has been integrated as a learner for fault diagnosis with SVM, kNN and multilayer perceptron (MLP) [23].

Deep learning. Based on different architectures, deep learning has developed many variants like CNN, recurrent neural networks (RNN), restricted boltzmann machines (RBM), deep belief networks (DBN), MLP, etc. These can be summarized as a general architecture called deep neural networks (DNN) while the variants can be designed or constructed using different quantities of hidden layers. By the convolutional operations and the specialty of parameters sharing, CNN is good at and capable of processing high-dimensional data like images or videos, constructing a fast and accurate detection system of cracks in underwater metal surfaces using inspection videos in NPPs [4]. Also it can be applied to detect scratches in nuclear fuel assembly appearances [11] and diagnose fault with a specific reactor dataset [55]. Different from CNN, RNN is suitable to deal with sequential data especially in natural language understanding and time-series signal processing, which has been integrated with PPCA, multi-resolution wavelet analysis to build a nuclear power machinery fault detection model [25]. To improve the reliability of engineering systems of NPPs, RBM was used to detect anomalies and conduct diagnosis & prognosis by analyzing massive data [38]. DBN has also been applied to indicate online faults for thermocouple in NPPs [29] while MLP has been introduced into transient detection [31] and identification for fault scenarios [41] in NPPs.

2.2 Data transformation: from numeric to imaged

Data transformation plays a vital role in solving some tough problems, which converts the nature of problems by reshaping data format or structure. The renowned CNN in image classification is a representative. Large quantities of temporal data generated in NPPs have been mapped into images and then anomaly detection has been performed using CNN [12] [22] by its great strength in image processing. In addition, this kind of data transformation from numeric to imaged has also been observed in other areas like bankruptcy prediction. By dividing the numerical data, the original numeric indicators from financial statements have been transformed into grey-scale images [17], then the issue of predicting bankruptcy for firms was interpreted as image classification task using CNN, with the results of better and more robust performances than conventional machine learning classifying models. The promotional effect of precision by data transformation has been experimentally verified in many cases, suggesting that this way could be considered and adopted in the field of fault detection in NPPs.

2.3 Deep learning based image classification

Image classification has been a longstanding research topic in the field of computer vision (CV). As a critical and fundamental problem, it aims to identify the label of classes for a given image. Traditional image classification using machine learning methods rely on two separate steps, first is applying hand-designed operations to extract features (e.g., SIFT, HOG, SURF, LBP) from input images, then is to train a classifier (e.g., SVM, kNN, AdaBoost) to recognize the class of input signals. These methods can deal with the tasks quite well with relatively small datasets, but the performance is not satisfactory in more complicated tasks. With the appearance and availability of huge datasets (e.g., ImageNet [7]) and the increasing computational capacity, advanced models with better performances are desperately needed and there comes up the deep learning based image classification techniques.

AlexNet [21], achieving record-breaking classification results in the *ImageNet Large-scale Visual Recognition Challenge*, has led to a spurt of various deep convolutional neural networks (CNNs) [14] [43] [45]. Instead of using hand-crafted features, CNNs apply flexible and trainable architectures to automatically extract and integrate low/mid/high-level features in an end-to-end multilayers fashion. The inductive biases inherent to CNNs, like translation invariance and local connectivity, are of great help in image feature extraction. In addition to classification, deep CNNs models have got state-of-the-art (SOTA) performances on a wide range of computer vision applications, such as object detection [24] [26] [37], image segmentation [5] [13] [40], super-resolution [8] [20], etc. With CNNs becoming the de-facto standards for computer vision, two major architectures, namely Transformer and MLP, have achieved competitive results compared with CNN-based models and attracted significant

research interest globally from the CV community.

Transformer [50] is a model architecture based solely on the self-attention mechanism to draw global dependencies between input and output, dispensing with recurrence and convolutions. Inspired by the great success of Transformer in natural language processing (NLP), numerous studies have started to migrate the Transformer to visual recognition tasks. Compared with CNN-based models that focus only on local features, Transformer is able to learn long-range dependencies, which means it can easily derive global information. Meanwhile, it also learns visual features with fewer priors and further reduces human intervention, while it's of larger computational costs and can't be generalized well when training with insufficient amount of data. To tackle this problem, a series of works [3] [27] [49] have been proposed in researches.

Multilayer perceptron (MLP) is not a new tool in the community of computer vision, which is actually one of the first classifiers tested on MNIST and has become the most widely-used model and benchmark in computer vision tasks at that time. Nevertheless, limited by a large amount of calculation and prone to over-fitting when the amount of data is insufficient, MLP has once been lying dormant for decades. Recent years, inspired by Transformer's great success in vision tasks, researchers have proposed deep MLP models such as MLP-Mixer [47] and ResMLP [48] for promotional performances. Compared with conventional MLP, these models have more hidden layers and have changed the input from full flattening to patch flattening, indicating that a purely MLP-based architecture which drops the artificially designed attention mechanism and learns from the raw data, can be fairly competitive with CNNs or Transformer. But it also suffers from the issue of data-hungry and furthermore, the complexity of MLP is quadratic to the size of images, making MLP-like models intractable on high-resolution images.

3 Methodology

In this section, detailed description of the two-stage fault detection methodology in NPPs will be presented. For better understanding and directly visual senses, the precise process of this methodology proposed is shown in Figure 1. In short, after the dimension reduction, the time-series numeric data of the operating status collected from sensors in NPPs will be firstly encoded into images by different transforming methods. Then these transformed images are concatenated and will be fed into different deep classification models to conduct the image classification and output the fault class as detection results. In addition, we will focus on the processes of exact methods for encoding time-series numeric data into images and deep learning based image classification models, which will be elaborated while the specifics of experiments like dimension reduction of data and image concatenation will be fully illustrated in Section 4.

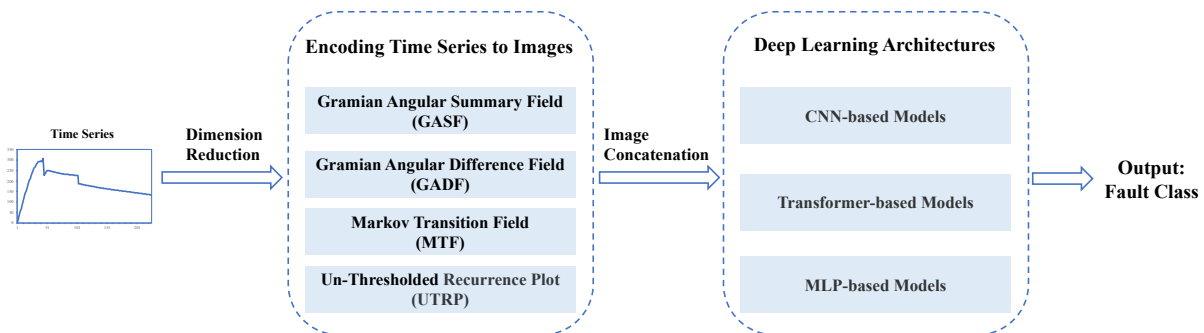


Figure 1: The precise process of the two-stage fault detection methodology in NPPs

3.1 Data processing & transformation

Obviously the transformed images as experimental data certainly will affect the results and performance of fault detection. A suitable and appropriate transforming method should be able to preserve temporal relations in images. Thus, four major transforming methods to convert 1-D time-series

data into different types of 2-D images have been adopted in our methodology which have been applied or tested in other studies. And they are Gramian Angular Summation Field (GASF) [46][54], Gramian Angular Difference Field (GADF) [46][54], Markov Transition Field (MTF) [12][46][54] and Un-Thresholded Recurrence Plot (UTRP) [46], of which the basic principles and detailed processes will be introduced in the following.

Gramian Angular Field. Gramian Angular Field [51] is a method that encodes time-series, which are represented in a polar coordinate system rather than the Cartesian coordinates, into images. According to different calculation methods, it can be divided into two types: Gramian Angular Summation Field (GASF) and Gramian Angular Difference Field (GADF).

Given a time series $X = \{x_1, x_2, \dots, x_n\}$ composed of n real-valued monitoring values obtained from a certain sensor, it is first normalized to the interval $[-1,1]$ according to:

$$\tilde{x}_i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)}$$

Then, encoding time stamp t_i as the radius and the value x_i as the angular cosine to represent the rescaled X in polar coordinates, which is defined as:

$$\begin{cases} \phi = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, & \tilde{x}_i \in \tilde{X} \\ r = \frac{t_i}{N}, & t_i \in \mathbb{N} \end{cases}$$

where t_i is the time stamp corresponding to x_i , N is the total length of the time stamp to normalize the span of polar coordinates. Encoding time series into polar coordinates system has two important properties. First, it can produce one and only one result with a unique inverse map, because it is bijective as $\cos(\phi)$ is monotonic when $\phi \in [0, \pi]$. Second, compared with Cartesian coordinates, it is able to preserve absolute temporal relations.

The final step is to use the transformed polar coordinate system to generate 2-D images which are represented as a Gramian-like matrix. Each element in the Gramian-like matrix is actually the trigonometric sum/difference between each point, so it can be used to identify the temporal correlation within different time intervals. The GASF and GADF are defined as follows:

$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} = \tilde{X}' \cdot \tilde{X} - \sqrt{I - \tilde{X}^2}' \cdot \sqrt{I - \tilde{X}^2}$$

$$GADF = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \cdots & \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) & \cdots & \sin(\phi_2 - \phi_n) \\ \vdots & \ddots & \vdots \\ \sin(\phi_n - \phi_1) & \cdots & \sin(\phi_n - \phi_n) \end{bmatrix} = \sqrt{I - \tilde{X}^2}' \cdot \tilde{X} - \tilde{X}' \cdot \sqrt{I - \tilde{X}^2}$$

where I is the unit row vector $[1, 1, \dots, 1]$. In GAF, time increases as the position moves from top-left to bottom-right, making it possible to retain the temporal dependency. The main diagonal of GAF contains the original value information.

Markov Transition Field. Markov Transition Field [51] is a method which expands the dynamic transition statistical information by sequentially representing the Markov transition probability.

Given a time series $X = \{x_1, x_2, \dots, x_n\}$, first determine its quantile bins Q and allocate each x_i to the corresponding bin $q_j (j \in [1, Q])$. Then construct a $Q \times Q$ weighted adjacency matrix W by computing the Markov transition probabilities among quantile bins. For example, w_{ij} means the probability of $x_t \in q_i$ and $x_{t+1} \in q_j$. After normalization by $\sum_j w_{ij} = 1$, the Markov transition matrix W is defined as follows:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,Q} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ w_{Q,1} & w_{Q,2} & \cdots & w_{Q,Q} \end{bmatrix}$$

W is not sensitive to the distribution of X as well as the temporal dependency on time stamp t_i , thus it can not reflect the time characteristic. The final step is to construct MTF based on W .

$$MTF = \begin{bmatrix} w_{i,j}|x_1 \in q_i, x_1 \in q_j & \cdots & w_{i,j}|x_1 \in q_i, x_n \in q_j \\ w_{i,j}|x_2 \in q_i, x_1 \in q_j & \cdots & w_{i,j}|x_2 \in q_i, x_n \in q_j \\ \vdots & \ddots & \vdots \\ w_{i,j}|x_n \in q_i, x_1 \in q_j & \cdots & w_{i,j}|x_n \in q_i, x_n \in q_j \end{bmatrix}$$

Each element in MTF denotes the transition probability of point in q_i at time stamp i move to q_j at time stamp j . The information and time characteristics contained in the MTF vary with the value of quantile bins Q .

Un-Thresholded Recurrence Plot. Recurrence Plot (RP) [10] is a powerful tool to analyze the periodicity, chaos and non-stationarity of time-series data and identify hidden regularities in scalar time-series. Whether using the threshold, it can be divided into two methods: Thresholded Recurrence Plot (TRP) and Un-Thresholded Recurrence Plot (UTRP). Since the distance parameter between recursive points in UTRP is preserved, for the same time series, UTRP often contains more information than TRP. Thus, we choose to apply UTRP to encode the time-series data.

Given a time series $X = \{x_1, x_2, \dots, x_n\}$, the first step is to form a reconstructed phase space by extending the one-dimensional time-series into a higher-dimensional space, which is represented as:

$$\tilde{X} = \begin{bmatrix} \vec{x}_0 \\ \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix} = \begin{bmatrix} x_0 & x_\tau & \cdots & x_{(m-1)\tau} \\ x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n+\tau} & \cdots & x_{n+(m-1)\tau} \end{bmatrix}$$

where $m \geq 2$ is the embedding dimension and $\tau \geq 1$ is the time delay.

The recurrence plot is constructed by calculating:

$$R_{i,j} = \|\vec{x}_i - \vec{x}_j\|, \quad i, j = 1, 2, \dots, n$$

where $\|\cdot\|$ is the usual Euclidean norm.

The pixel value of each position (i, j) in the un-thresholded recurrence plot reflects the correlation between vector \vec{x}_i and vector \vec{x}_j , therefore it could reveal the dynamic characteristics of the signal in time.

3.2 Deep learning based architectures for image classification

In the research field of image classification, CNNs have dominated this area over the last decade since 2012’s powerful AlexNet. With the introduction of Vision Transformer and MLP-Mixer, it seems that a new round of paradigm shift has ushered in. Therefore, a series of image classification models including CNN-based models, Transformer-based models and MLP-based models have been incorporated into this proposed methodology, with extensive experiments carried out in a unified framework to compare these three major deep learning architectures.

CNN-based models. Convolutional neural network is a special case of neural network, which makes use of the 2-D structure of images and the advantage that pixels within a neighborhood are usually highly correlated. A standard CNN-based model consists of an input layer, alternating convolutional and pooling layer, fully connected layer as well as output layer. Figure 2 shows the architecture of a standard CNN-based model. The convolutional layer performs convolutional operations on the input image matrix to extract features, then apply the pooling layer to compress feature map obtained by the convolutional layer. Pooling layers not only reduce the number of parameters to prevent over-fitting, but also are able to maintain model invariance (e.g., rotation, translation, scale). If the depth of the network increases, the receptive field continues to expand, enabling the network to incorporate and learn more abstract characteristics. Finally the fully connected layer acts as a classifier to obtain classification results based on the feature maps gain by multiple convolutional and pooling

layers. In this past decade, a series of networks have been designed to improve the performances of CNNs, for instance, the ResNet [14] proposed residual structures, which significantly alleviate the gradient disappearance problem and make it possible to train deeper models. And the EfficientNet [45] has achieved better accuracy with less parameters after carefully acquiring a balance between network width, depth and resolution. Therefore, the ResNet and EfficientNet have been chosen as representatives of CNN-based models.

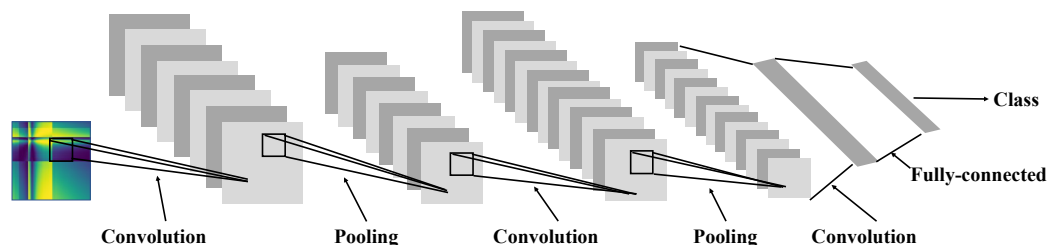


Figure 2: Architecture of a standard CNN-based model

Transformer-based models. ViT [9] is the first full-transformer model that can achieve state-of-the-art performance on image classification task. The overview of the ViT architecture is depicted in Figure 3.

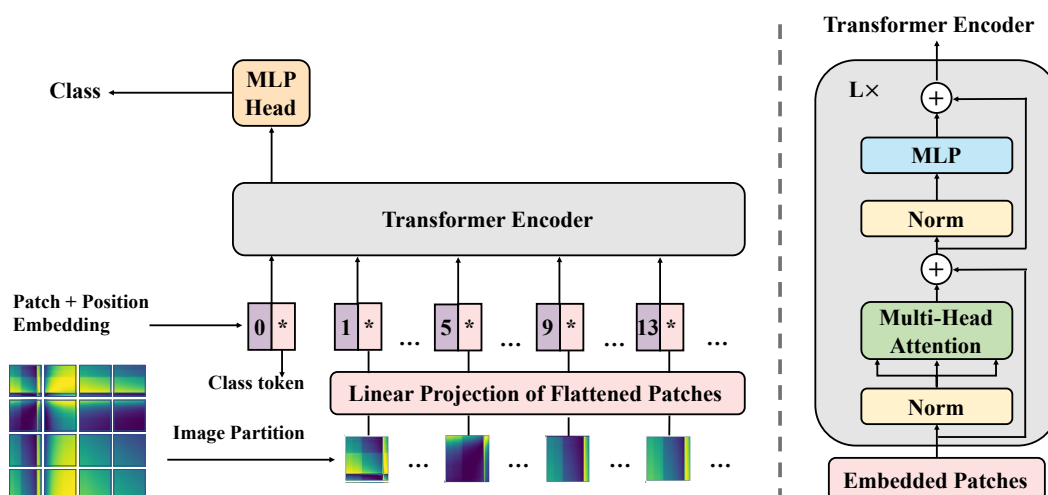


Figure 3: Architecture of ViT

Since the original Transformer receives a 1-D sequence of token embeddings as input, the input 2-D images have to be firstly split into a sequence of non-overlapped patches, similar to words in NLP application. The patches are then projected into patch embeddings by a trainable linear projection. The patch embeddings with an additional class token are fed together into the Transformer encoder after adding position embeddings to retain positional information. In classification tasks, only the class token can be predicted by a MLP head to output image class. The Transformer encoder in ViT is composed of a stack of L identical blocks. Each block consists of alternating layers of multi-head self-attention (MSA) and multilayer perceptron (MLP). Residual connection and layer normalization are employed to enhance the scalability.

As the core component of Transformer, MSA is an extension of standard self-attention. The self-attention mechanism estimates the relevance of each position to all positions and aggregates global information from the complete input sequence to update the output vector. Specifically, the input sequences X are mapped into three different sequential vectors and they are query vectors Q , key vectors K and value vectors V , according to:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V$$

where $W^Q \in R^{d \times dk}$, $W^K \in R^{d \times dk}$, $W^V \in R^{d \times dv}$ are linear matrices; d and dk denote the length and dimension of queries, keys and values respectively.

Compute the dot products of the query with all keys, then scale it and use the softmax function to obtain the weights on the values. This process can be formulated as:

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

In order to capture the multiple complex relationships and extract richer information, the MSA mechanism has been introduced [50], comprising h parallel self-attention layers, and the outputs of each layer are concatenated and projected to the final output, which is defined as:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V), \quad i = 1, 2, \dots, h$$

where $W^O \in R^{hd_v \times d}$. Furthermore, the Data-efficient image Transformer (DeiT) [49] has been proposed to improve vision transformer applicability when training on ImageNet-1k. Finally, the DeiT has been chosen as a representative of Transformer-based models.

MLP-based models. The pioneering work of deep MLP models, MLP-Mixer [47], which consists of three modules as shown in Figure 4. The per-patch fully-connected layer and classifier layer play the same role as in ViT, mapping image patches into embedding vector and generating the classification results respectively. In the second module, MLP-Mixer stacks N Mixer layers instead of self-attention layers. Each Mixer layer contains one token-mixing MLP and one channel-mixing MLP. Apply token-mixing MLP to communicate between different patches, and apply channel-mixing MLP to communicate between different channels. Every MLP includes two fully-connected layers and a GELU [15] activation function. The layers of Mixer can be written as follows:

$$U_{*,i} = X_{*,i} + W_2\sigma(W_1 LayerNorm(X)_{*,i}), \quad for \ i = 1 \dots C,$$

$$Y_{j,*} = U_{j,*} + W_4\sigma(W_3 LayerNorm(U)_{j,*}), \quad for \ j = 1 \dots S.$$

where $X \in R^{S \times C}$ is the input of each mixer layer, S and C are the numbers of patches and channels, σ is the GELU function, $W_1 - W_4$ represents weights of different fully-connected layers. In addition, the ResMLP [48] replaces the layer normalization in MLP-Mixer with the affine transform layer, aiming to stabilize the training process, so the ResMLP has been chosen as a representative of MLP-based models.

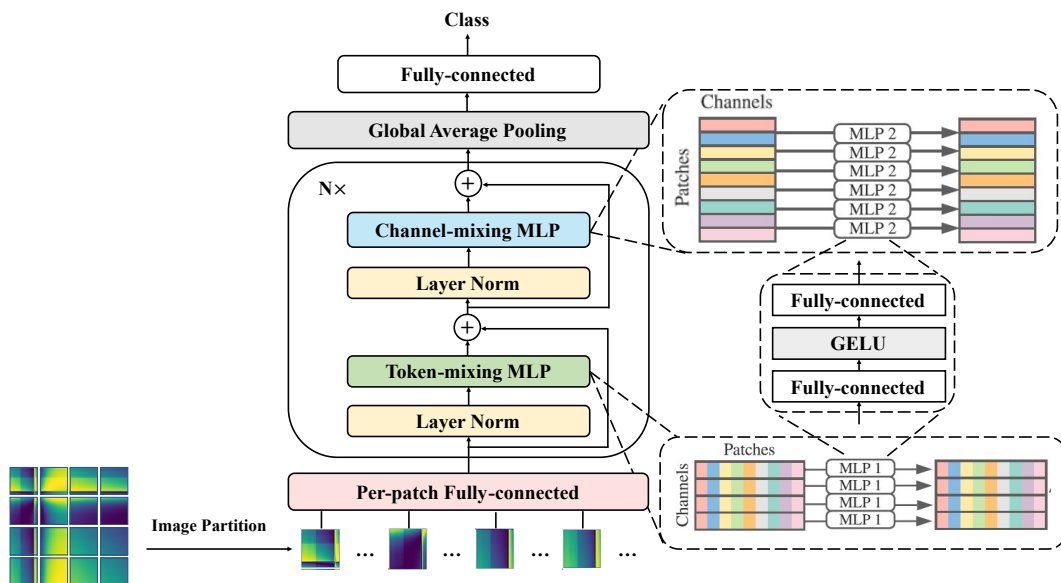


Figure 4: Architecture of MLP-Mixer

4 Experiment & results

This section will primarily illustrate the experiment & results. First the data acquisition & processing will be summarized with details and secondly are the experimental settings. Finally, results and analysis will be introduced with a discussion of experiments.

4.1 Data acquisition & processing

Data acquisition. The data needed in our study is the time-series data of instances of operating status in NPPs, with labels of each instance indicating the current status like normal condition or abnormal conditions. Due to the specialty of NPPs and the difficulty of collecting real data from NPPs, a practical and realistic way for researchers is to obtain suitable data from simulation softwares [12] [22]. For example, the most widely-used nuclear reactor transient analyzer called Personal Computer Transient Analyzer (PCTTRAN) [36] is a simulation software for a variety of accidents and transient conditions occurred in NPPs, which could present and output the status of key parameters, and it allows simulation of operators' actions by interactive control. Thus, we choose the PCTTRAN software for experimental data acquisition.

After the simulation by the software, large quantities of numeric instances describing operating status in NPPs have been collected, containing 70 dimensions as 70 sensors have been monitoring the pressure, temperature, flow rate and many other indicators from different places in NPPs like reactor cores, steam generators, pipelines, heat exchangers, etc. In addition, 5 labels of current status have also been generated as 5 classes of each instance including one normal condition (Label 0) and four most common types of fault conditions, loss of coolant accident (Label 1), loss of feedwater accident (Label 2), steam line break (Label 3) and steam generator tube rupture (Label 4). Until here, the data acquisition process has been basically completed.

Data processing. The simulation process has generated tens of thousands of numeric instances indicating the operating status of a NPP, which are basically presented in seconds. Then the data transformation has been conducted using the four methods introduced in 3.1. The 1-D time-series data collected from each sensor has been transformed into one channel in an image, meaning a single image is consisted of 70 channels and each channel represents the temporal trend of each dimension or sensor. In order to better characterize the time-series information, the data have been sampled in the sliding window of 224 seconds with an interval of 5 seconds. Once an instance with fault condition labels (label 1-4) is included in this sliding window, the corresponding transformed image should be noted as fault labels. In addition, for better efficiency of fault detection and quick diagnosis, it's necessary to remove noisy and redundant information within the dataset by reducing the dimensions (70) of the data, before the transformation process. To do this, the most prevalently used dimension reduction algorithm, the Principal Component Analysis (PCA) [19], has been used to process the simulated data with the principal components set to 32 based on actual conditions. In this way, we have obtained 6300 images while each image contains 32 channels, as qualified experimental dataset. The distribution of classes of transformed images is shown in Table 1 and clearly it's an imbalanced supervised classification task.

Table 1: Distribution of classes of transformed images

Classes (Labels)	Number of images
Normal	4200
Loss of coolant accident	600
Loss of feedwater accident	300
Steam line break	600
Steam generator tube rupture	600

The four types of transformed images have been acquired and cases of these images are shown in Figure 5. Specifically, the key parameters in transformation are set as follows: MTF quantile bins = 8, UTRP delay time = 1 and embedding dimension = 1. In Figure 5, each row indicates four converting

means respectively, GASF, GADF, MTF, UTRP, and columns are different classes of images, as label 0 is normal condition images while label 1-4 represent different types of faults. Different classes of images are distinctly different from each other, which can be observed simply by visual inspection, and it has established strong foundations for image classification models and effective detection of faults using the methodology proposed in this research.

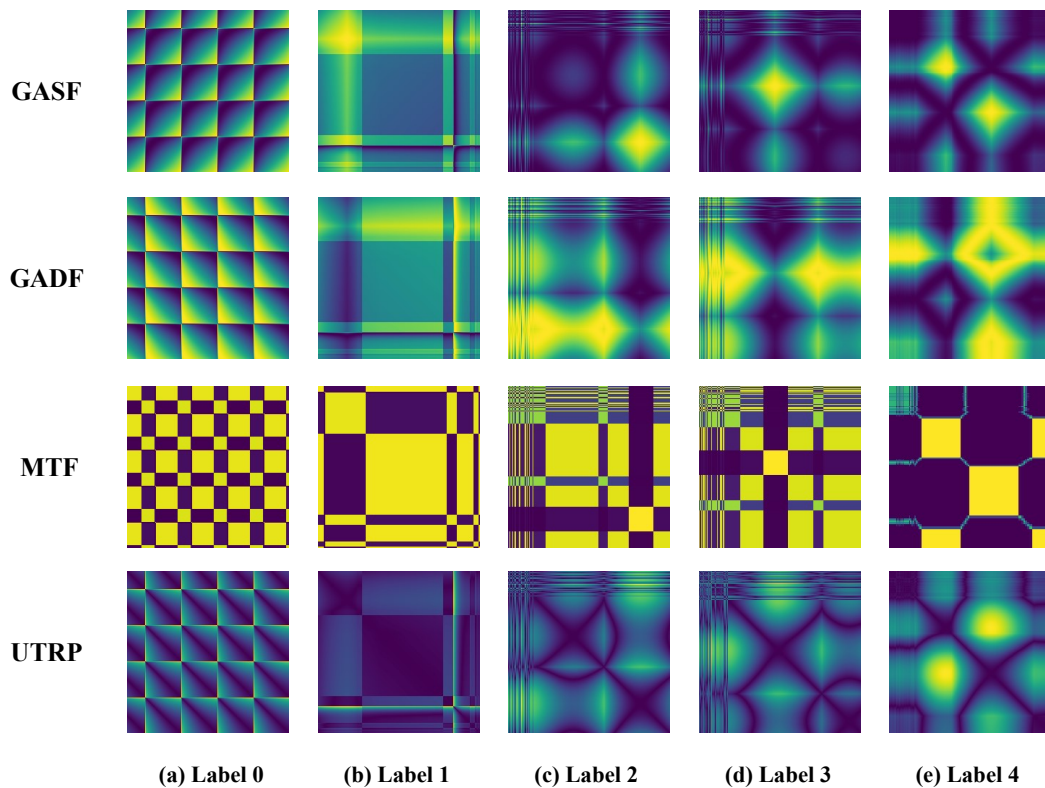


Figure 5: The transformed images of normal data and fault data

4.2 Experimental settings

According to the three primary deep learning architectures mentioned in Section 3.2, we have chosen the most currently popular models for experiments including ResNet (CNNs) [14], EfficientNet (CNNs) [45], DeiT (Transformer) [49] and ResMLP (MLP) [48]. Furthermore, to better achieve real-time fault diagnosis and lower the computational costs, we have chosen the exact models with minimal parameters among the variants, and they are ResNet18, EfficientNet-B0, DeiT-Tiny as well as ResMLP-S12. These models with pre-trained weights on ImageNet will be used for experiments subsequently. The details of network designs are summarized in Table 3.

For the supervised classification task, the quantities of images in training set and testing set have been arranged as 8:2 separately. We train the models illustrated above for 50 epochs using batch size 32 on one GPU while all models use 224×224 images as input and apply the cross-entropy loss. Most of our training settings are inherited from the original models as the ResNet18 and EfficientNet-B0 use a similar setting: SGD [39] as the optimizer with weight decay 0.0001 and momentum 0.9, learning rate starting from 0.1 and divided by 10 after 30 epochs. In DeiT-Tiny and ResMLP, we employ AdamW [28] optimizer with weight decay 0.05 and the initial learning rate is $\frac{0.0005}{512} \times batch\ size$ with cosine learning rate decay.

4.3 Analysis of results

Comparison of different transforming methods. We have compared the throughput and image size of each transformation method, details shown in Table 2. The throughput is measured as the number of images that can be processed per second on CPU. The results have demonstrated that

the transforming speed of MTF is almost four times faster than the other three means, and at the same time, the size of the image converted by MTF is also the smallest.

Table 2: Comparison of throughput and image size of different transformation methods

Transformation Methods	GASF	GADF	MTF	UTRP
Throughput (im/s)	30.48	29.12	109.57	30.14
Size (kb/per)	78.52	92.70	2.40	89.76

Before the transformation, the numeric 1-D time-series data has been processed by PCA for dimension reduction as instructed in 4.1, which has produced 32 key dimensions (channels). Then we have concated 32 channels into a single 32-channel image, similar with a 3-channel RGB image, as the feeding data into the deep learning models.

Performances of different combinations: accuracy vs. efficiency. We have trained the selected four models, ResNet18, EfficientNet-B0, DeiT-Tiny and ResMLP-S12 on the transformed image dataset, where the input size is $32 \times 224 \times 224$. Figure 6 shows the training process of different transformation methods with different deep learning models. It can be seen that, on the whole, all models and methods converge within 50 epochs and have achieved very high accuracy eventually.

Specifically, from the perspective of the transformation methods, the training process of MTF is the most stable one while other methods fluctuates greatly. MTF and GADF can converge in 20 epochs, and GASF and UTRP need 30 epochs to finish this process. However, UTRP achieves the highest accuracy in all four transforming methods while the classifying performance of MTF is inferior to other methods (see Figure 7). From the perspective of the classification models, Efficientnet-B0 achieves the best performance on all converting methods. DeiT-Tiny is slightly worse than Efficientnet-B0, but the training process is more stable. Regardless of convergence or accuracy, ResMLP-S12 is inferior to other models with all transforming means especially in GASF.

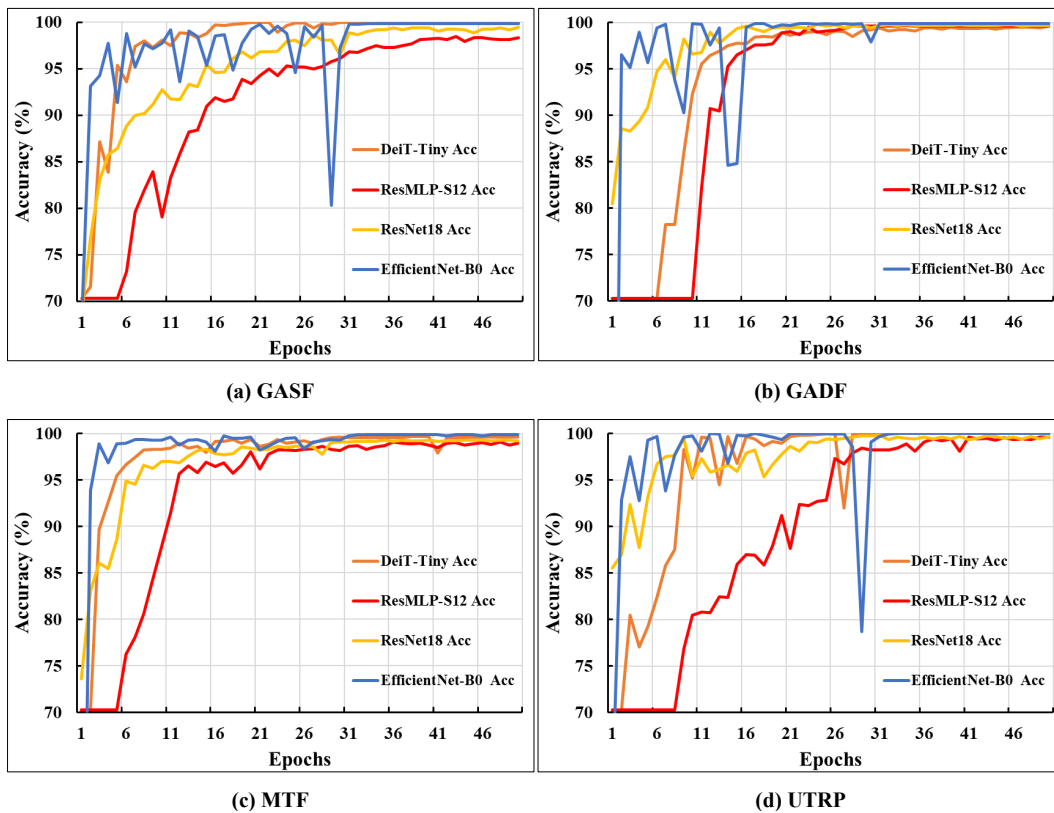


Figure 6: Training process of different transforming methods with different deep learning models

In order to evaluate different classifying means more accurately, the classification models are usually compared based on the trade-off between accuracy, number of parameters and throughput.

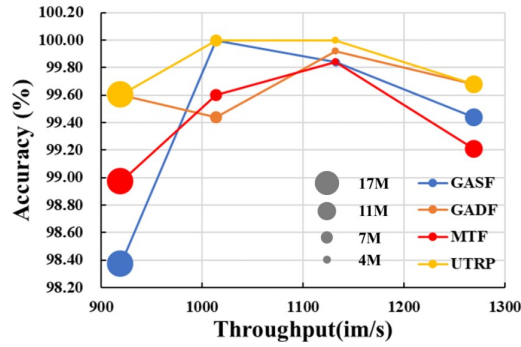


Figure 7: Comparison of accuracy, throughput and params of transforming methods

Table 3: Details of network designs of different models and classification accuracy of combinations in experiments

Architectures	Exact Model	Params (M)	Layers	Throughput (im/s)	Transforming Methods	Accuracy (%)
CNN-based models	ResNet18	11.27	18	1268.75	GASF	99.44
					GADF	99.68
					MTF	99.21
					UTRP	99.68
	EfficientNet-B0	4.02	18	1132.28	GASF	99.84
					GADF	99.92
					MTF	99.84
					UTRP	100.00
Transformer-based models	DeiT-Tiny	6.90	12	1013.83	GASF	100.00
					GADF	99.44
					MTF	99.60
					UTRP	100.00
MLP-based models	ResMLP-S12	17.79	12	919.81	GASF	98.37
					GADF	99.60
					MTF	98.97
					UTRP	99.60

Table 3 reports the classification accuracies of different architectures based on CNN, Transformer and MLP with throughput, and number of parameters. Because fault diagnosis in NPPs requires very quick response, we have chosen to conduct assessment based on the trade-off between accuracy and throughput. In here, the throughput is measured as the number of images that can be processed per second on one NVIDIA RTX 3060 GPU with largest possible batch size. Larger throughput means that the processing speed of the model is faster under the same hardware conditions. Among the four models, the throughput of ResNet18 is the fastest, followed by EfficientNet-B0 while the throughput of ResMLP-S12 is the worst.

4.4 Summary

This section has conducted extensive experiments to compare the performances of different deep learning based models (CNN-based models, Transformer-based models, MLP-based models) in fault detection with images generated by different transforming methods (GASF, GADF, MTF, UTRP). It can be observed that MTF has the fastest converting speed and the converted image occupies the smallest memory, yet it is not the optimal choice because its classification accuracy is inferior to almost every methods on each model. In contrast, UTRP achieves excellent performances, no matter which classifying model has been used. In terms of the trade-off between throughput and accuracy, EfficientNet-B0 outperforms other deep learning classification approaches because it has achieved high accuracy with a faster throughput, which can meet the requirements of quick response in fault diagnosis of NPPs, according to the experimental results.

5 Conclusion & future directions

This research has proposed a two-stage fault detection method for NPPs based on imaged data and deep learning image classification. Experimental results have demonstrated that this methodology proposed is capable of detecting and warning possible faults of NPPs in advance and further it could be adapted to promote the safety management in many different power systems with high accuracy, fast response and great efficiency. The primary contributions of our research are as follows:

- The time-series data of the operating status of NPPs have been transformed into imaged data by four major methods, preserving the time-series information in images, which enables deep learning models to convert the fault detection problem into supervised image classification task.
- Extensive experiments have been carried out to compare and analyze the performances of different combinations of data transforming methods and image classifying models with the results of effective accuracies, which has also identified that UTRP and EfficientNet-B0 (CNN-based model) outperform others in the trade-off between precision and throughput.

This proposed methodology has improved the precision of fault detection and diagnosis in NPPs, providing effective and scientific support to decision-makings. However, there still leaves many limitations and shortcomings in our research. One is the data used in experiments. Though data transformation is considered to be a promising direction in solving many tough tasks, it has also raised doubts about the interpretability of artificial changes of data structure or format. Also, there's only one kind of data (numeric data) applied in our research and it's collected by simulation of software, not real operating status of a NPP, while it has been confirmed that more data involved in experiments will help achieve better performances, like the fusion of multiple-source heterogeneous data or integration of different forms of data. Another issue we need to pay attention to is those new models that have evolved rapidly recent years represented by deep learning based approaches. These emerging methodologies deal with data from different perspectives and could outperform than conventional ones to certain extent as long as appropriate data with compatible structures is available. Diversification of data and the significant development of computing capacity not only require more data-driven intelligent analyzing tools, but also make it possible for the implementations of these advanced techniques in both academic research and practice. Hence, to better promote the theoretical and practical development of fault detection in NPPs, it's necessary to further explore possible innovations from the two points illustrated above, more different types of data and more intelligent analyzing tools.

Acknowledgements

All authors hereby express sincere condolences for our loss of **Ioan Dzitac** (Founder & former Editor-in-Chief of IJCCC), and great appreciation to editors and reviewers for their careful work and valuable comments, especially **Florin G. Filip** and **Simona Dzitac** for their unwavering patience, understanding and support. The APC was funded by R&D center "Cercetare Dezvoltare Agora" of Agora University and this article was funded by the Key Project (grant number 71932008) of National Natural Science Foundation of China.

Author contributions

Y. Shi: Conceptualization; Formal Analysis; Methodology; Funding acquisition; Project administration; Supervision; Writing - review & editing. **X. Xue:** Conceptualization; Data curation; Formal analysis; Methodology; Software; Writing - original draft; Writing - review & editing. **J. Xue:** Conceptualization; Data curation; Formal analysis; Methodology; Software; Writing - review & editing. **Y. Qu:** Conceptualization; Data curation; Formal analysis; Methodology; Writing - review & editing.

Conflict of interest

All authors declare no conflict of interest.

References

- [1] Aizpurua, J. I.; McArthur, S.; Stewart, B. G.; Lambert, B.; Cross, J. G.; Catterson, V. M. (2019). Adaptive Power Transformer Lifetime Predictions Through Machine Learning and Uncertainty Modeling in Nuclear Power Plants, *IEEE Transactions on Industrial Electronics*, 66, 4726–4737, 2019.
- [2] Ayodeji, A.; Liu, Y. (2019). PWR heat exchanger tube defects: Trends, signatures and diagnostic techniques, *Progress in Nuclear Energy*, 112, 171–184, 2019.
- [3] Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. (2021). Emerging Properties in Self-Supervised Vision Transformers, *ArXiv, abs/2104.14294*, Available: <https://arxiv.org/abs/2104.14294>, 2021.
- [4] Chen, F.; Jahanshahi, M. R. (2018). NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion, *IEEE Transactions on Industrial Electronics*, 65, 4392–4400, 2018.
- [5] Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K. P.; Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40, 834–848, 2018.
- [6] Deleplace, A.; Atamuradov, V.; Allali, A.; Pelle, J. T.; Plana, R.; Alleaume, G. (2020). Ensemble Learning-based Fault Detection in Nuclear Power Plant Screen Cleaners, *IFAC-PapersOnLine*, 53, 10354–10359, 2020.
- [7] Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database, *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 248–255, 2009.
- [8] Dong, C.; Loy, C. C.; He, K.; Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 295–307, 2016.
- [9] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Deghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, *9th International Conference on Learning Representations (ICLR)*, 2021.
- [10] Eckmann, J.; Kamphorst, S. O.; Ruelle, D. (1987). Recurrence Plots of Dynamical Systems, *Europhysics Letters*, 4, 973–977, 1987.
- [11] Guo, Z.; Wu, Z.; Liu, S.; Ma, X.; Wang, C.; Yan, D.; Niu, F. (2020). Defect detection of nuclear fuel assembly based on deep neural network, *Annals of Nuclear Energy*, 137, 107078, 2020.
- [12] He, C.; Ge, D.; Yang, M.; Yong, N.; Wang, J.; Yu, J. (2021). A data-driven adaptive fault diagnosis methodology for nuclear power systems based on NSGAIL-CNN, *Annals of Nuclear Energy*, 159, 108326, 2021.
- [13] He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. B. (2020). Mask R-CNN, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 386–397, 2020.
- [14] He, K.; Zhang, X.; Ren, S.; Sun, J. (2016). Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, 2016.

- [15] Hendrycks, D.; Gimpel, K. (2016). Gaussian Error Linear Units (GELUs), *ArXiv, abs/1606.08415*, Available: <http://arxiv.org/abs/1606.08415>, 2016.
- [16] Holdsworth, A. F.; George, K. E.; Adams, S. J.; Sharrad, C. A. (2021). An accessible statistical regression approach for the estimation of spent nuclear fuel compositions and decay heats to support the development of nuclear fuel management strategies, *Progress in Nuclear Energy*, 141, 103935, 2021.
- [17] Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks, *Expert Systems with Applications*, 117, 287–299, 2019.
- [18] Hu, G.; Zhou, T.; Liu, Q. (2021). Data-Driven Machine Learning for Fault Detection and Diagnosis in Nuclear Power Plants: A Review, *Frontiers in Energy Research*, 9, 1–12, 2021.
- [19] Jolliffe, I. T. (1986). *Principal Component Analysis*, Springer, 1986.
- [20] Kim, J.; Lee, J. K.; Lee, K. M. (2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1646–1654, 2016.
- [21] Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems (NIPS)*, 25, 2012.
- [22] Lee, G.; Lee, S. J.; Lee, C. (2021). A convolutional neural network model for abnormality diagnosis in a nuclear power plant, *Applied Soft Computing*, 99, 106874, 2021.
- [23] Li, J.; Lin, M. (2021). Ensemble learning with diversified base models for fault diagnosis in nuclear power plants, *Annals of Nuclear Energy*, 158, 108265, 2021.
- [24] Lin, T.; Dollár, P.; Girshick, R. B.; He, K.; Hariharan, B.; Belongie, S. J. (2017). Feature Pyramid Networks for Object Detection, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936–944, 2017.
- [25] Ling, J.; Liu, G.; Li, J.; Shen, X.; You, D. (2020). Fault prediction method for nuclear power machinery based on Bayesian PPCA recurrent neural network model, *Nuclear Science and Techniques*, 31, 1–11, 2020.
- [26] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. E.; Fu, C.; Berg, A. C. (2016). SSD: Single Shot MultiBox Detector, *14th European Conference on Computer Vision (ECCV)*, 21–37, 2016.
- [27] Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, *ArXiv, abs/2103.14030*, Available: <http://arxiv.org/abs/2103.14030>, 2021.
- [28] Loshchilov, I.; Hutter, F. (2019). Decoupled Weight Decay Regularization, *7th International Conference on Learning Representations (ICLR)*, 2019.
- [29] Mandal, S.; Santhi, B.; Sridhar, S.; Vinolia, K.; Swaminathan, P. (2017). Nuclear Power Plant Thermocouple Sensor-Fault Detection and Classification Using Deep Learning and Generalized Likelihood Ratio Test, *IEEE Transactions on Nuclear Science*, 64, 1526–1534, 2017.
- [30] Meng, J.; Su, Y.; Xie, S. (2020). Loose parts detection method combining blind deconvolution with support vector machine, *Annals of Nuclear Energy*, 149, 107782, 2020.
- [31] Mo, K.; Lee, S. J.; Seong, P.H. (2007). A dynamic neural network aggregation model for transient diagnosis in nuclear power plants, *Progress in Nuclear Energy*, 49, 262–272, 2007.
- [32] Mohapatra, D.; Subudhi, B.; Daniel, R. (2020). Real-time sensor fault detection in Tokamak using different machine learning algorithms, *Fusion Engineering and Design*, 151, 111401, 2020.

- [33] Nicolau, A. D.; Augusto, J. P.; Schirru, R. (2017). Accident diagnosis system based on real-time decision tree expert system, *AIP Conference Proceeding*, 1836, 020017, 2017.
- [34] Oh, C.; Lee, J. I. (2020). Real time nuclear power plant operating state cognitive algorithm development using dynamic Bayesian network, *Reliability Engineering and System Safety*, 198, 106879, 2020.
- [35] Po, L. (2019). Conceptual Design of an Accident Prevention System for Light Water Reactors Using Artificial Neural Network and High-Speed Simulator, *Nuclear Technology*, 206, 505–513, 2019.
- [36] Racheal, S.; Liu, Y.; Ernest, M.; Ayodeji, A. (2021). A Systematic Review of PCTRAN-Based Pressurized Water Reactor Transient Analysis, *Proceedings of the 2021 28th International Conference on Nuclear Engineering*, 2021.
- [37] Redmon, J.; Divvala, S. K.; Girshick, R. B.; Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788, 2016.
- [38] Rezaeianjouybari, B.; Shang, Y. (2020). Deep learning for prognostics and health management: State of the art, challenges, and opportunities, *Measurement*, 163, 107929, 2020.
- [39] Robbins, H. E. (2007). A Stochastic Approximation Method, *Annals of Mathematical Statistics*, 22, 400–407, 1951.
- [40] Ronneberger, O.; Fischer, P.; Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation, *18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 234–241, 2015.
- [41] Saeed, H. A.; Peng, M.; Wang, H.; Zhang, B. (2020). Novel fault diagnosis scheme utilizing deep learning networks, *Progress in Nuclear Energy*, 118, 103066, 2020.
- [42] Shi, Y.; Xue, X.; Qu, Y.; Xue, J.; Zhang, L. (2021). Machine Learning and Deep Learning Methods used in Safety Management of Nuclear Power Plants : A Survey, *2021 International Conference on Data Mining Workshops (ICDMW)*, 917–924, 2021.
- [43] Simonyan, K.; Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition, *3rd International Conference on Learning Representations (ICLR)*, 2015.
- [44] Smith, C. L.; Borgonovo, E. (2007). Decision making during nuclear power plant incidents: a new approach to the evaluation of precursor events, *Risk analysis*, 27, 1027–1042, 2007.
- [45] Tan, M.; Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 6105–6114, 2019.
- [46] Tian, W.; Wu, J.; Cui, H.; Hu, T. (2021). Drought Prediction Based on Feature-Based Transfer Learning and Time Series Imaging, *IEEE Access*, 9, 101454–101468, 2021.
- [47] Tolstikhin, I. O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Keysers, D.; Uszkoreit, J.; Lucic, M.; Dosovitskiy, A. (2021). MLP-Mixer: An all-MLP Architecture for Vision, *ArXiv, abs/2105.01601*, Available: <http://arxiv.org/abs/2105.01601>, 2021.
- [48] Touvron, H.; Bojanowski, P.; Caron, M.; Cord, M.; El-Nouby, A.; Grave, E.; Izacard, G.; Joulin, A.; Synnaeve, G.; Verbeek, J.; J'egou, H. (2021). ResMLP: Feedforward networks for image classification with data-efficient training, *ArXiv, abs/2105.03404*, Available: <http://arxiv.org/abs/2105.03404>, 2021.

- [49] Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; J'egou, H. (2021). Training data-efficient image transformers & distillation through attention, *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 10347–10357, 2021.
- [50] Vaswani, A.; Shazeer, N. M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. (2017). Attention is All you Need, *Advances in Neural Information Processing Systems (NIPS)*, 30, 2017.
- [51] Wang, Z.; Oates, T. (2015). Imaging Time-Series to Improve Classification and Imputation, *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 3939–3945, 2015.
- [52] Wang, H.; Peng, M.; Wesley Hines, J.; Zheng, G.; Liu, Y.; Upadhyaya, B. R. (2019). A hybrid fault diagnosis methodology with support vector machine and improved particle swarm optimization for nuclear power plants, *ISA transactions*, 95, 358–371, 2019.
- [53] Wang, X.; Xu, Z. S.; Dzitac, I. (2019). Bibliometric Analysis on Research Trends of International Journal of Computers Communications & Control, *International Journal of Computers Communications & Control*, 14, 711–732, 2019.
- [54] Yang, C.; Chen, Z.; Yang, C. (2020). Sensor Classification Using Convolutional Neural Network by Encoding Multivariate Time Series as Two-Dimensional Colored Images, *Sensors*, 20, 1, 168, 2020.
- [55] Yao, Y.; Wang, J.; Long, P.; Xie, M.; Wang, J. (2020). Small-batch-size convolutional neural network based fault diagnosis system for nuclear energy production safety with big-data environment, *International Journal of Energy Research*, 44, 5841–5855, 2020.



Copyright ©2022 by the authors. Licensee Agora University, Oradea, Romania.

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

Journal's webpage: <http://univagora.ro/jour/index.php/ijccc/>



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).

<https://publicationethics.org/members/international-journal-computers-communications-and-control>

Cite this paper as:

Shi, Y.; Xue, X.; Xue, J.; Qu, Y. (2022). Fault Detection in Nuclear Power Plants using Deep Learning based Image Classification with Imaged Time-series Data, *International Journal of Computers Communications & Control*, 17 (1), 4714, 2022. <https://doi.org/10.15837/ijccc.2022.1.4714>