Nuclear Engineering and Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Development of machine learning model for automatic ELM-burst detection without hyperparameter adjustment in KSTAR tokamak

Jiheon Song [a], Semin Joung [b], Young-Chul Ghim [b], Sang-hee Hahn [c], Juhyeok Jang [c], Jungpyo Lee [a, *]

[a] Hanyang University, Department of Nuclear Engineering, 222 Wangsimni-ro, Seongdong-gu, Seoul, 04763, Republic of Korea
[b] Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
[c] Korea Institute of Fusion Energy, Daejeon, 34133, Republic of Korea

ARTICLE INFO

Article history:
Received 26 May 2022
Received in revised form
28 July 2022
Accepted 24 August 2022
Available online xxx

ABSTRACT

In this study, a neural network model inspired by a one-dimensional convolution U-net is developed to automatically accelerate edge localized mode (ELM) detection from big diagnostic data of fusion devices and increase the detection accuracy regardless of the hyperparameter setting. This model recognizes the input signal patterns and overcomes the problems of existing detection algorithms, such as the prominence algorithm and those of differential methods with high sensitivity for the threshold and signal intensity. To train the model, 10 sets of discharge radiation data from the KSTAR are used and sliced into 11091 inputs of length 12 ms, of which 20% are used for validation. According to the receiver operating characteristic curves, our model shows a positive prediction rate and a true prediction rate of approximately 90% each, which is comparable to the best detection performance afforded by other algorithms using their optimized hyperparameters. The accurate and automatic ELM-burst detection methodology used in our model can be beneficial for determining plasma properties, such as the ELM frequency from big data measured in multiple experiments using machines from the KSTAR device and ITER. Additionally, it is applicable to feature detection in the time-series data of other engineering fields.

1. Introduction

To operate an advanced tokamak nuclear fusion facility, the undesirable filamentary eruption of plasma around the edge area, known as the edge localized mode (ELM), must be controlled [1–5]. This magnetohydrodynamic instability mode occurs when the plasma enhanced confinement mode (H-mode) is achieved with edge transport barrier. The accompanying ELM-burst event releases a significant amount of energy and particles in an extremely short duration (~1 ms), causing significant damage to the diverter tile in future device ITER owing to the high heat flux ($> 10\, MW/\, m^2$) [6], which must be addressed for the commercialization of fusion reactors. More specifically, not all ELM bursts involve a high heat flux. For example, a type-I ELM with a low frequency and a large burst peak and heat flux should be avoided in ITER-type devices, although a type-III ELM, which has a high frequency and a small burst peak known as grassy ELM, is desirable [6]. Therefore, the ELM must be detected and characterized from diagnostic signals to understand and control the ELM for achieving high-performance tokamaks. However, an analysis of the ELM manually using large amount of diagnostics data is challenging. To accelerate ELM analysis, we developed a novel technique for detecting ELM bursts automatically from a significant amount of diagnostic signal data in the Korea Superconducting Tokamak Advanced Research (KSTAR) tokamak [7].

ELM-burst events can be directly identified by a kind of radiation signal (D-alpha) D_α, which represents the visible light of the deuterium spectral lines of the Balmer series. When ELM-burst events happened, the peaks of D_α signals occur because the particles emitted from the plasma toward open magnetic fields at low temperatures are preferred for electron recombination, thereby resulting in the emission of visible light. However, because a significant amount of temporal data (more than thousands of time steps in each channel) is generated during the discharge of the KSTAR tokamak, the ELM-burst peak detection method must be automated. Several methods have been used to detect ELM-burst events. Kim et al. [8] used a prominence selection algorithm,

* Corresponding author.
E-mail address: jungpyo@hanyang.ac.kr (J. Lee).

https://doi.org/10.1016/j.net.2022.08.026

which identifies peaks via a simple comparison of neighboring values, to detect the ELM peak in the D_α signal. Eldon et al. [9] adopted a method of fine-tuning the difference by selecting points that exceed the smoothed differences of the D_α signal. However, the prominence algorithm relies on the threshold for height information; therefore, a small-intensity ELM peak cannot be detected. The difference method can detect small-intensity ELM peaks by adjusting the threshold for the difference value; however, it is unlikely to be automatic because a threshold optimization process is required for each discharge.

The automatic ELM-peak detection algorithm was developed by Berta et al. [10] based on a similarity comparison with a fitted single distribution function of the COMPASS tokamak [11], where a single function known as the Wald distribution was used. However, according to the authors of [12], ELM data in a wide range of discharges are unlikely to fit in a single distribution function, as indicated by the Joint European Torus (JET) tokamak data. To increase the reliability of the detection algorithm for versatile ELM signals, we introduced a machine learning technique based on convolutional neural networks (CNN) [13].

Recently, the field of machine learning involving CNNs for pattern recognition has developed extensively. Since results from the neural network model are typically more accurate than those of some conventional methods [14], the neural network model is widely adopted for data processing in the fusion devices [15,16]. We developed a new machine learning model based on the U-net [17], a type of CNN, and applied it to detect one-dimensional (1-D) ELM signals. U-net was proposed for medical image segmentation, in which encoding (downsampling) and decoding (upsampling) structures are connected using data of different dimensions (e.g., one [18], two [17], and three dimensions [19,20]). Our model had skip connections by simply connecting not cropping as U-net and based on a 1-D time signal. For out model, no assumption is required for the ELM-burst shape; therefore, the model is expected to be applicable to various types of ELM bursts. However, for simplicity, we do not differentiate between the ELM types (e.g., type-I or type-III ELMs). Because we labeled the ELM data without assigning the type information in training, the model result only shows the time range in which the ELM burst exists. The classification of ELM types is to be performed in future studies.

The remainder of this paper is organized as follows: In Section 2, the process by which the ELM data are obtained from the D_α measurement of the KSTAR experiments and the data labeling for supervised learning are explained. In Section 3, we introduce our CNN detection model, training method, and optimization process. In Section 4, the model results are presented in terms of accuracy based on a comparison with existing algorithms and their application to ELM physics. Finally, the conclusions of this study are presented in Section 5.

2. ELM data acquisition and labelling in KSTAR experiments

The plasma performance and heat flux in the tokamak were significantly affected by the presence of the ELM bursts. The H-mode with explosive ELMs is shown in Fig. 1. In particular, the figure shows the plasma current (I_p), plasma beta (β), heating power by electron cyclotron heating (ECH) and neutral beam injection (NBI), core and edge electron temperature via electron cyclotron emission (ECE), magnetic field fluctuation from the Mirnov signal, and D_α signal [21] of the KSTAR discharge #29487 ($I_p = 600\ kA$, toroidal magnetic fields $B_\phi = 1.8$ T). The D_α signal shows ELM-burst events as the plasma transforms into the H-mode with enhanced plasma beta (β), which signifies a high plasma pressure. The ELM and its characteristics (intensity, frequency, etc.) are correlated well with other plasma performance parameters. In this discharge, the increase of the NBI beam power likely results in the increase of the ELM intensity (see the time frames of 5–8 and 11–13 s in Fig. 1(a)). As shown in the enlarged plot in Fig. 1, the ELM is shown in the fluctuation of the magnetic field measurement by the Mirnov signal [22]. This causes an abrupt decrease in the edge electron temperature (purple line in Fig. 1(b)) measured in the ECE if the ELM intensity is large, as highlighted in green. A small ELM does not significantly affect the edge electron temperature, as highlighted in gray.

For supervised learning in our approach, a sufficiently high amount of high-quality labeled data is preferred. To ensure the versatility of our model, which can be applied to various plasma environments, we selected and labeled training data from 10 KSTAR discharges with different normalized plasma beta (β_N), poloidal current (I_p), and total neutral beam power (P_{NBI}) values, as shown in Fig. 2. Additionally, the fraction of the ECH power to the total heating power of each discharge is indicated by color. It is important to consider the shape of the plasma for finding ELM characteristics such as the ELM frequency. One of the important shape parameters is triangularity (δ), which is a factor in how close the plasma poloidal cross section is formed to the D-shape. The range of δ included in the training set is shown in Fig. 2. The data were sliced into 11091 inputs of length 12 ms to train the model. These selected discharge samples included large parameter spaces of similar weights. A well-distributed sample over the parameter space can be useful for learning the general ELM-burst shapes in our model. Subsequently, two additional discharges, #18396 and #29487, used for the test set data, were labeled to compare our model's accuracy with those of other algorithms, as explained in Section 4.

The labeling of the D_α signal to the ELM-burst area, was conducted by confirming the additional signals based on diagnostics using Mirnov coils [22]. The Mirnov coil measures magnetic field fluctuations [23] based on the peeling-ballooning instability with temporal collapse of edge density. Hence, by considering both the Mirnov signal with a low-pass filter[1] and the D_α radiation signal, we can ensure the labeling reliability for determining the existence of the ELM in the ELM-burst area. Fig. 3(a) shows the ELM-burst region labeled from the D_α signal, which is compared with the filtered Mirnov measurement of KSTAR #21228 in Fig. 3(b). The D_α peak was labeled when a large fluctuation in Mirnov occurred simultaneously. For example, the small peak of the D_α signal of about 5.62 s in Fig. 3 was not labeled.

Additionally, among the discharge data, some outliers were eliminated from the sampling. We exclude the signal peak appeared during the disruption and the dithering peaks during the L–H transition. Therefore, the initial time data prior to the L–H transition (e.g., at approximately 2 s in Fig. 1) were eliminated because the ELM is supposed to be absent. The labeled data were selected only in the period (e.g., from 2 to 14 s in Fig. 1) slightly before the first ELM burst occurred until the current flat-top control ended.

3. Development of neural network model

3.1. Architecture of the ELM-burst detection model

Our neural network model is suitable for automatic ELM-burst detection because its ELM-burst intensities and frequencies change significantly in many KSTAR discharges, and the peak shapes are non-uniform. The key feature of automation is that after training model with specific hyperparameter once, no adjustment

[1] An analog Butterworth filter cutoff at 8 kHz was used numerically in the post-process.

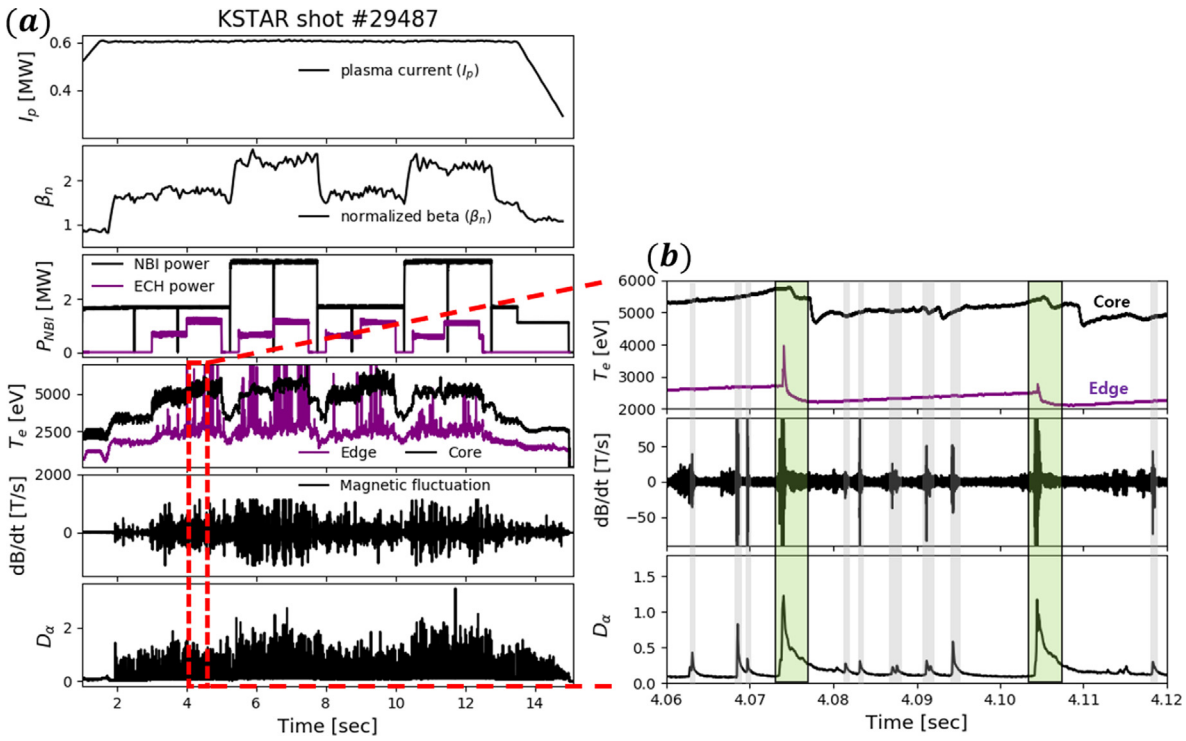

Fig. 1. Time history of several plasma parameters in KSTAR discharge #29487 (left) and expanded view around t = 4.06 s (right). Y-axes show plasma current (I_p), normalized plasma beta (β_N), total neutral beam power, electron cyclotron heating power (P_{NBI}, P_{ECH}), electron temperature (T_e) at plasma core and edge, magnetic field fluctuation, and D_α signal. H-mode is obtained at approximately 2.0 s with explosive edge localized modes (ELMs).

is required between several discharge samples. A model that accurately learns and classifies burst features is required to perform classification without changing hyperparameters. An autoencoder [24] based on a CNN model has been reported to be effective for feature extraction from data because of its bottleneck structure, which compresses the signal. Therefore, the autoencoder model is suitable for classifying ELM peaks. Furthermore, the input information before encoding should be preserved to determine whether each data point is an ELM burst based on its features. Thus, the U-net [17] model shown in Fig. 4, which adds a layer (yellow arrow in Fig. 4) that connects the encoding and decoding steps by adding a crop layer to the autoencoder, was selected for ELM-burst detection.

Fig. 4 shows the overall structure of the proposed model. Our U-net-based model can accommodate a 1-D signal input of size 256× 1, and the interval time of each point is $5 \cdot 10^{-5}$ s. The model first reduces the length of the 256-value input to 32 through three convolution layers with a max pooling layer and then decodes the data with the original length. Therefore, a 12 ms signal is processed, and the output measures 256 × 2. To classify the ELM-burst area, the output was configured using a one-hot encoding method. If the column of each point is closely located the ELM burst (1,0) signal than the normal (0,1) signal, then it is considered an ELM-burst region. The general U-net [17,19] comprises a connection layer that crops some information before downsampling and concatenates it with the upsampling layer in the decoding section. Similarly, our model contains simplified connection layers that preserve the information prior to encoding. The model uses an activation function known as a rectified linear unit in the entire convolution layer [25], except for the final layer, where the sigmoid activation function is used. Thus, each component of the output is between 0 and 1. The model was implemented in TensorFlow2.0 framework based on Python.

3.2. Training and optimization

The model was successfully trained using 11091 input data points sliced from 10 labeled D_α discharges, as shown in Fig. 2. The labeled data were resampled to unify the sampling time to $5 \cdot 10^{-5}$ s. This was conducted by simply dropping the points between the intervals. Subsequently, each labeled data point was sliced to a length comprising 256 points (12 ms) with a moving window comprising 100 points to increase the amount of data. For example, discharge data #19306 provides 1665 sliced data measuring 256 × 1, and the total number of sliced data is 11091. The sliced data were segregated into training and validation sets at a ratio of 8:2. The data classified as validation data were not used for training because the number of sliced data was sufficient to train our model. The validation set was entirely different from the test set described in Section 2. In this model, the data were normalized by the largest value in each slice to prevent the incorrect overweighting of some input data. This is a key advantage of our model compared to other algorithms because our model can differentiate not the intensity but the shape of the peaks. Data normalization is important to ensure the insensitivity of the signal intensity. An ELM detection model can be implemented without changing hyperparameter setting between KSTAR discharges with different intensity values of D_α. We also found that our model can differentiate well the small peaks in ELM-free data from the small ELM peaks because the small peaks (or noise) in the ELM-free data do not have enough time duration (time width) on the x axis to form a shape to be detected.

Our model was trained using the cross-entropy function for a loss function; meanwhile, adaptive moment estimation (Adam) was used as the L2 optimizer. The model was trained up to 1500 epochs at a learning rate of 0.00001. The training was performed in an environment involving Intel Xeon Gold 6240R CPU with three numbers of NVIDIA 3090 RTX 24GB, and the total training time was

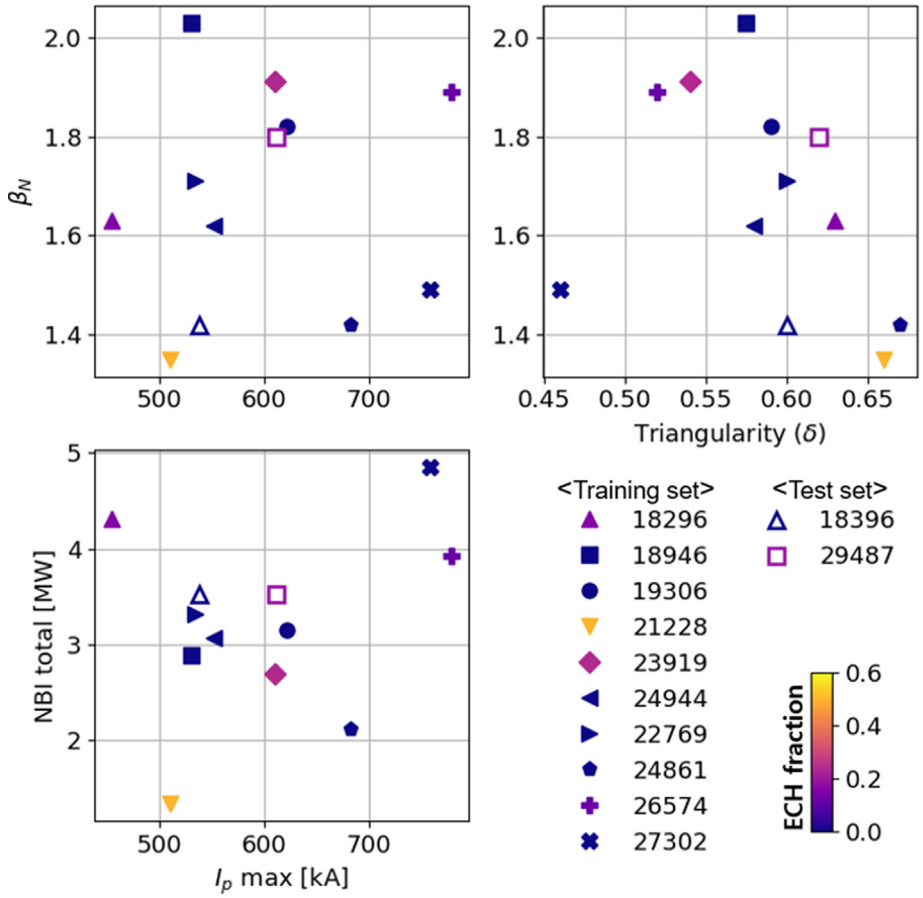

Fig. 2. Scattered plots of 10 KSTAR discharges parameters selected for our training in terms of NBI total power, colloidal current (I_p) and normalized plasma beta (β_N), and plasma shape triangularity parameter (δ). The ECH power ratio to the total heating power were indicated by color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

less than 1 h. Fig. 5 shows the training results for the loss and accuracy. Accuracy refers to the matching rate between the label and output. As the learning progressed, the accuracy increased for both the training and validation datasets until 1000 epochs. Subsequently, unlike the training data, the validation data indicated a decrease in accuracy. This clearly shows that the model overfitted after 1000 epochs.

To select the optimized epoch number, analysis was conducted using the true positive (TP) and positive prediction (PP) rates. The TP rate represents the proportion of peaks predicted by the model among the labeled peaks. The peak implies a point with the largest D_α signal in the ELM-burst area, and a point with a size of less than 0.1 ms is disregarded. The PP rate represents the proportion of peaks that match the labeled data among the peaks predicted by the model. Fig. 6 shows the change in the TP and PP rates for each discharge number as the epoch number increases for all the sliced data. As the learning progressed, the model learned more peaks, and the TP values increased. By contrast, the PP values decreased gradually, except in the early phases of training. The point at which both the TP and PP values are maximized would be the ideal optimized epoch number. Thus, we selected the model parameters based on an epoch number of 660, where the sum of the TP and PP is the highest for the training and validation data. After 660 epochs, the model determines the ELM-burst area for an excessively large number of peaks.

4. Detection results and analysis

4.1. Comparison of accuracy with other algorithms

In this section, we compare the detection results of our model with those of several existing ELM detection algorithms, such as the methods of prominence [8], simplified difference [26], and fine-tuned difference [9]. Fig. 7 shows the procedures of the different peak detection algorithms ((a) prominence, (b) fine-tuned difference method, and (c) our model) for the KSTAR discharge. For methods (a) and (b), the ELM peak was well-classified by selecting an appropriate threshold.

The prominence method estimates the ground level of each peak and is typically used to classify peaks in topography [27]. As shown in Fig. 7(a), the ground level indicated by the red line for each section is obtained from nearby local minimum points. If the height of the peak point from the ground level is greater than a predetermined threshold value, it is regarded as a peak. To achieve the best detection, if a smaller peak point exists within 0.5 ms, then the local peak is disregarded. The threshold of prominence is set as $1.1 \cdot 10^{20}$, as shown in Fig. 7(a). The effects of different thresholds are shown in Figs. 8–10. Notably, the prominence method failed to detect a small peak at 2.782 s.

The simplified difference method [26] uses the distribution of signal differences without any filter. Among the distribution of

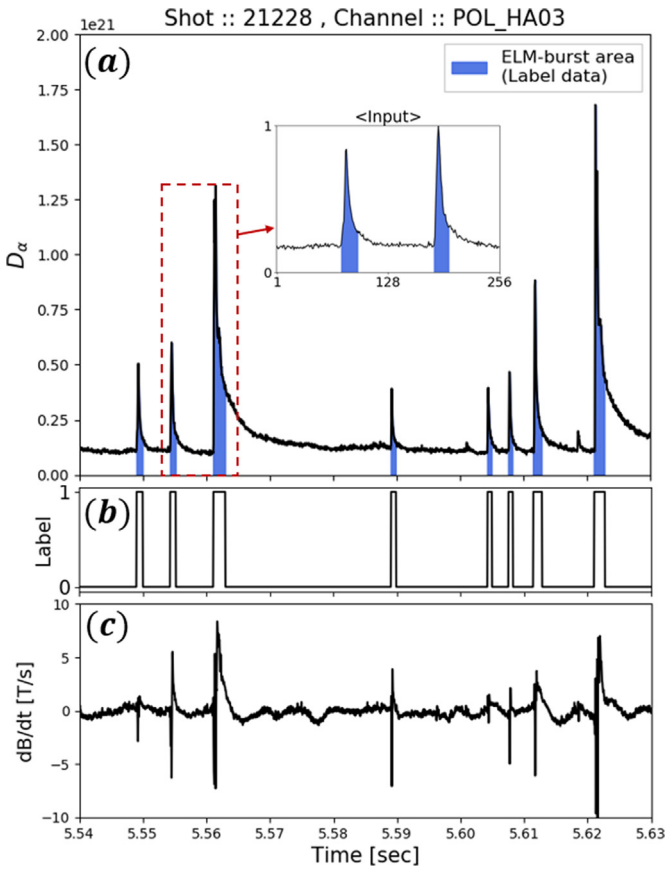

Fig. 3. An example of labeled data for supervised learning. (a) D_α signal with labeled ELM-burst area, (b) label data (0:none, 1:ELM-burst area) (c) low-pass filtered (8 kHz) Mirnov signal of KSTAR experiment #21228. The ELM-burst showing the clear magnetic perturbations in (c) was labeled as (b) and highlighted as blue in (a). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

differences, the points farther than the threshold from the mean are regarded as peaks when a threshold (= h*σ) of value proportional to the standard deviation is used ($\sigma = \int (D-\overline{D})^2 dt$, $D \simeq \frac{dy}{dt}$ is the time derivative of data), where h is a predetermined hyperparameter. For this method, additional points assessed as peaks within 1 ms of the forward peak are disregarded.

Eldon et al. [9] used a method using fine-tuned differences of smoothed D_α signals for ELM detection, which is similar to the Gaussian edge detection method [28]. Specifically, D_α signals were smoothed using a low-pass Butterworth filter at two different cut-off frequencies of 10^3 and 200 $rad\,s^{-1}$. By applying the height threshold of each smoothed difference ($SD_1 \simeq \frac{dS_1(y)}{dt}$, $SD_2 \simeq \frac{dS_2(y)}{dt}$), the points exceeding the threshold are classified as ELM areas. As shown in Fig. 7(b), the highest D_α signal among the areas without considering the small gap (> 0.5 *ms*) between the areas is the peak point. Fig. 7(b) shows two smoothed differences (SD_1 *and* SD_2) and their respective thresholds, i.e., $SD_{1,th} = 4.3 \cdot 10^{18}$ and $SD_{2,th} = 1.6 \cdot 10^{18}$.

The performance of each algorithm can be compared by adjusting its threshold in the PP and TP domains to draw a line known as the ROC curve, as shown in Figs. 8–10. In the TP and PP domains, as the results approach the upper right corner, the accuracy of the detection model increases. A comparison was performed using the discharge #18296 data, which were used for training. Fig. 8 shows the ROC curves of the other algorithms with different hyperparameters and our CNN model results. To draw the ROC curves, the threshold of the prominence algorithm was set between $4.3 \cdot 10^{20}$ and 10^{19}. For the algorithm using fine-tuned differences, $SD_{1,th}$ was set between $2 \cdot 10^{18}$ and $5 \cdot 10^{19}$, and $SD_{2,th}$ was assumed to be $SD_{1,th}$. The threshold for the simplified difference method is shown in Fig. 8. In the ROC curves, the optimized threshold value for the prominence, simplified difference, and fine-tuned differences were $1.6 \cdot 10^{20}$, 14.5, and $SD_{1,th} = 1.7 \cdot 10^{19}$, respectively. Meanwhile, our model found PP and TP values of 0.924 and 0.935, respectively, which indicates similar levels of high accuracy without hyperparameter adjustment.

For an accurate comparison, the same analysis was performed using the test dataset of KSTAR discharges #18396 and #29487, mentioned in Section 2. The ROC curves along with the model performance of the test data, and the detection results of the model are shown in Figs. 9 and 10, respectively. To draw the ROC curve for the method using fine-tuned differences, the relation of thresholds was set as $SD_{1,th}/3 = SD_{2,th}$ at #18396 and $SD_{1,th} = SD_{2,th}$ at #29487, which were adjusted for the optimization of ELM-burst detection of each signal. As shown in Figs. 9(b) and 10(b), the intensity and frequency of the ELM burst were different in the two

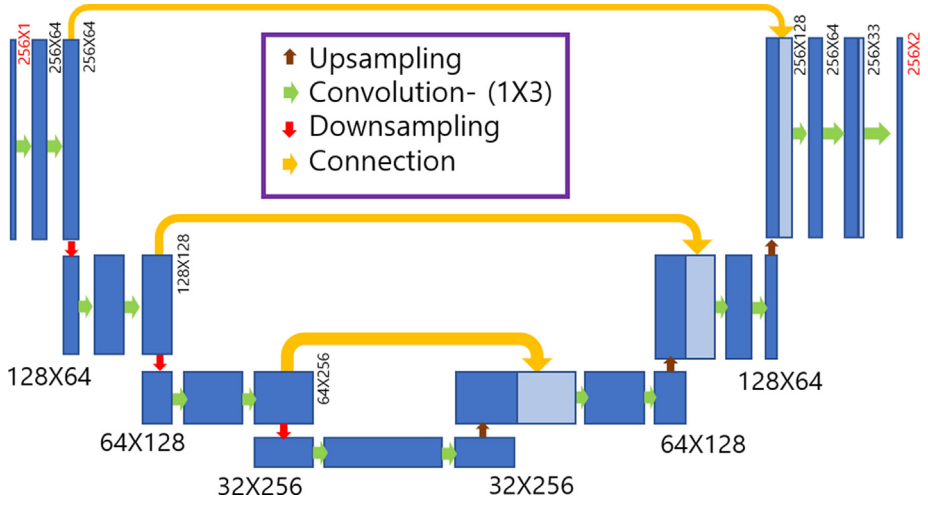

Fig. 4. Architecture of U-net convolutional model in this study. The number of length and channel of data is denoted on each box. The 1D input data is D_α signal of 12 ms. The encoding layers and decoding layers are connected by simply copying the entire layer, not cropping like the conventional U-net [17,19].

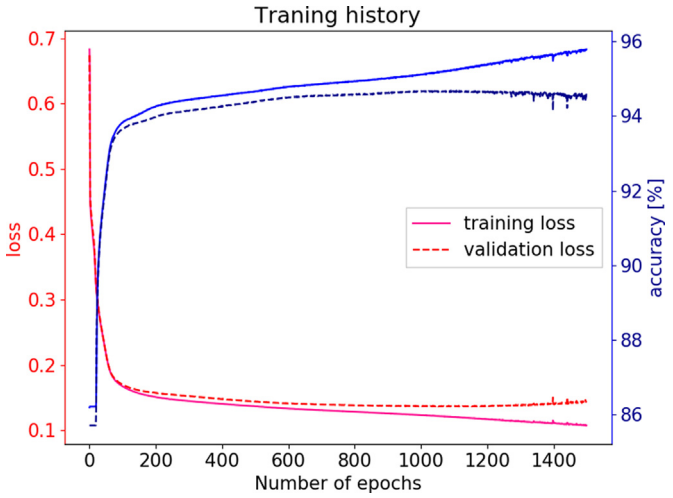

Fig. 5. The training result (solid lines) of the loss function (left y) and the accuracy (right y) in terms of epoch number for our model. The dashed lines are the result using the validation data.

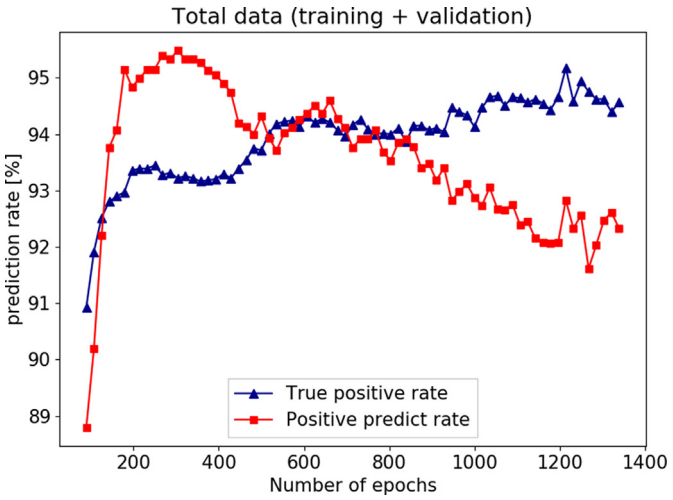

Fig. 6. True positive (TP) and positive prediction (PP) rates in terms of epoch number for the same case of Fig. 5.

experimental D_α signals. Accordingly, the optimal threshold value for each existing method was changed. In #18396, the optimized threshold value for the prominence difference was $1.9\cdot 10^{20}$, simplified difference 12.5, and fine-tuned difference $SD_{1,th} = 1.4\cdot 10^{19}$. In #29487, the optimized threshold value for the prominence difference was $1.6\cdot 10^{19}$, simplified difference 10.5, and fine-tuned difference $SD_{1,th} = 1.3\cdot 10^{19}$. Because the optimized thresholds of the other algorithms were different for each discharge, a good performance position in the upper right corner was difficult to achieve for multiple discharge data. However, our model was located on the upper right, which yielded good accuracy without setting any parameters, whereas it was located slightly above the ROC curve of other methods.

4.2. ELM frequency calculation

Our model's ELM automatic detection results can be applied to determine the related physics, such as the ELM frequency. Using our model, the ELM frequency can be calculated easily as the number of peaks captured automatically. By comparing the

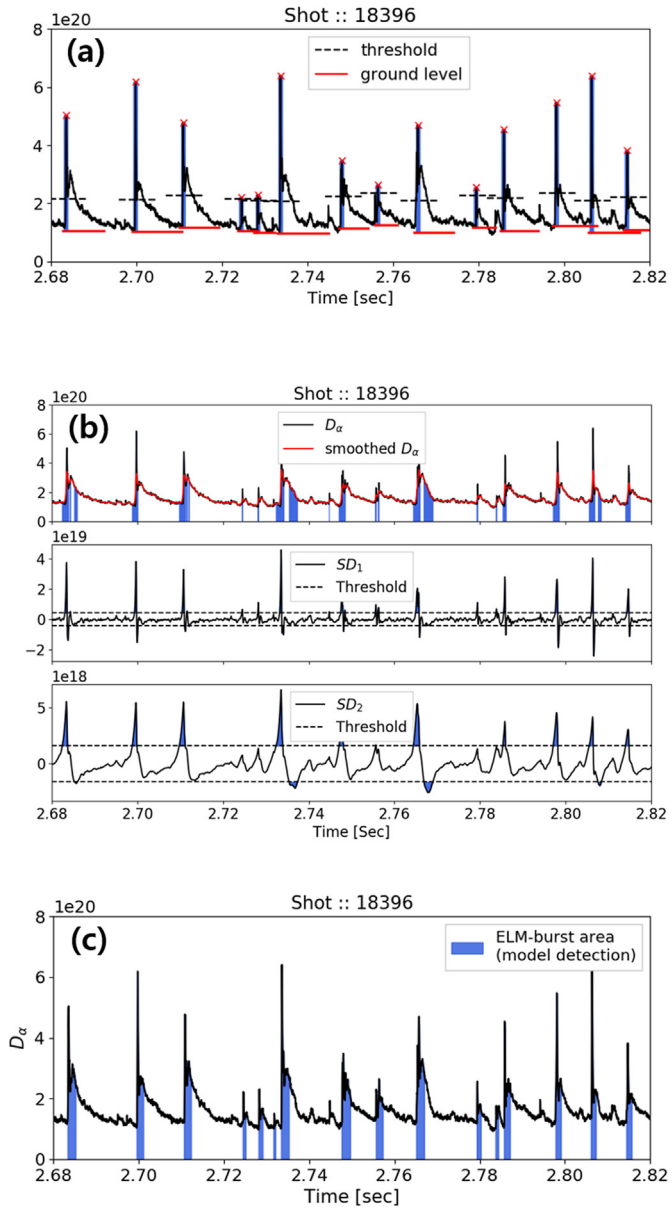

Fig. 7. ELM peak detections from D_α signal based on different algorithms: (a) prominence (threshold = $1.1\cdot 10^{20}$ case), (b) fine-tuned by smoothed differences, and (c) our model for time slice of KSTAR discharge #18396. Thresholds of methods (a) and (b) are indicated by dotted lines, respectively.

frequency between multiple discharges in KSTAR, we can obtain the characteristics of the ELM in different plasma environments. The shape of the plasma poloidal cross section is known to be important when determining the ELM frequency. One of the important shape parameters is the triangularity, as indicated in the previous study [29]. In our training data set, the triangularity varies from 0.45 to 0.7, and the peaks are successfully detected in the range of the triangularity. Fig. 11 shows the correlation of the ELM frequency obtained from our model output for discharge #29487 with the heating power and ion toroidal rotation velocity (V_{tor}). The frequency was calculated for 0.4 s of the moving window at 0.2 s intervals of D_α. V_{tor} was obtained from the charge exchange spectroscopy channel, which was located at the top of the edge pedestal region. As ECH power was additionally applied to the plasma at 3 s, the ELM frequency from our calculation changed significantly, which is consistent with the measurements of V_{tor}.

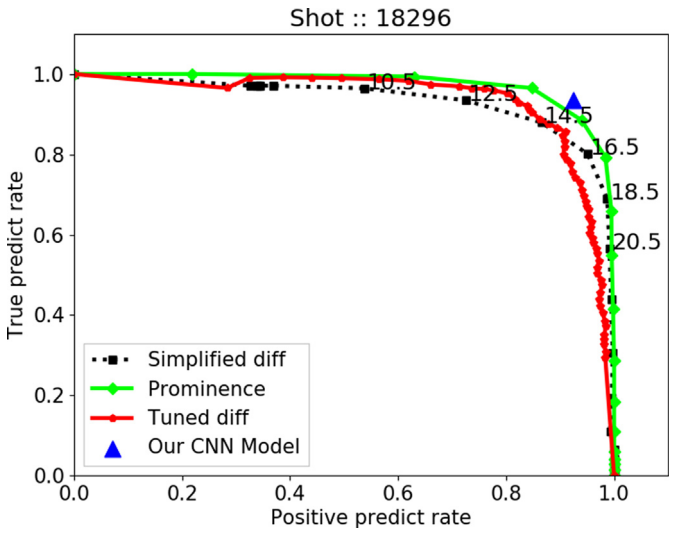

Fig. 8. Receiver operating characteristic (ROC) curves of several ELM detection algorithms from D_α signal of KSTAR experiment #18296, which was used as a training set. The simplified difference deviation method [26] (black square dashed), prominence method [8] (green dot) threshold, and tuned difference method [9] (red pentagon) have optimized thresholds of 14.5, 1.6×10^{20}, and $SD_{1,th} = SD_{2,th} = 1.7 \times 10^{19}$, respectively. The numbers on the graphs represent the thresholds of the simplified difference method. Our trained model's result is shown in a blue triangle point. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

For type-I or II ELM, the frequency increases as the heating power [30] or edge collisionality [31] increases, and the frequency decreases when co-current direction V_{tor} increases [32]. A reduction in V_{tor} in the co-current direction strengthens the radial electric field (E_r) toward the negative direction at the edge, which affects the plasma confinement and stabilization of the ELM trigger [33,34]. However, ELM physics is still not clearly understood. Contrary to the aforementioned trend, the frequency of type-III ELM decreases as the heating power increases [35] and the ELM energy loss of type-III slightly decreases with decreasing collisionality [36]. Therefore, further analysis of ELM characteristics is required. Our model can automatically classify ELM-peak detections between several discharges with a relatively constant performance, which would facilitate ELM investigations pertaining to the big data of KSTAR and other tokamaks.

5. Conclusions

In this study, an automatic ELM-burst detection model using the D_α signal in a fusion reactor was developed using a 1-D U-net model. We directly labeled 10 KSTAR experiments involving D_α data for training and two additional experiments for testing based on a comparison with the Mirnov signal. The epoch number of training was selected where the detection performance for ELM-peak is maximize and the trained model showed high performance with a more than 90 percent of positive prediction rate (PP). We compared the learning results yielded by the proposed model with those of other existing algorithms. As shown by the training and test dataset results in Figs. 8–10, our trained model demonstrated a similar level of accuracy as the optimized results of other algorithms without requiring any hyperparameter adjustment in the experimental data. To illustrate a simple example of the usage of our model, the ELM frequency was calculated using the model, and its correlation with the plasma parameters was presented.

Our model offers the following two advantages in terms of hyperparameter-independent characteristics. First, our model is independent of the signal intensity. Because CNN methods depend primarily on the signal pattern and not the signal magnitude, they are more suitable for signal normalization. The D_α input in our model can be less dependent on the sample intensity by normalizing it with the highest value within each input slice, whereas the other algorithms are not suitable for normalization because they depend on the global threshold of the magnitude in the entire input.

Second, we can avoid the sensitivity problem arising from manually setting thresholds among different experimental data, which is important because KSTAR has quite different wall condition each year which comes up with a different D_α base level. The other methods have different threshold values optimized for capturing the peaks in each discharge, as shown in Section 4.1. This implies that multiple discharges are difficult to compare using a uniform standard. By contrast, our model does not involve such thresholds, and it can classify peaks via the shape and pattern with consistently good accuracy in different environments. Therefore, accurate and automatic ELM-burst detection using our model can

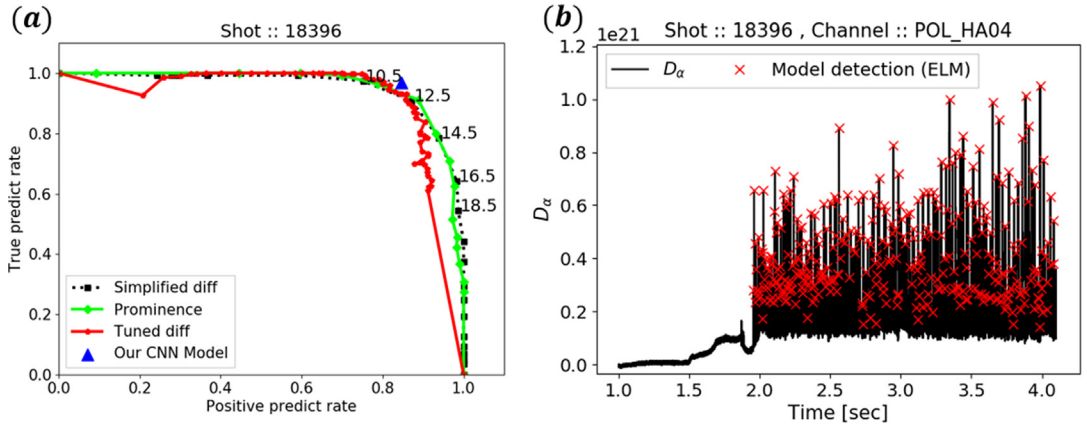

Fig. 9. (a) ROC curve of several ELM detection algorithms from D_α signal of KSTAR experiment #18396, which was used as a test set, and (b) peak detections by our model. The simplified difference deviation method [26] (black square dashed), prominence method [8] (green dot) threshold, and tuned difference method [9] (red pentagon) have optimized thresholds of 12.5, 1.9×10^{20}, and $SD_{1,th} = SD_{2,th} = 1.4 \times 10^{19}$, respectively. Our trained model's result is shown in blue triangle point. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

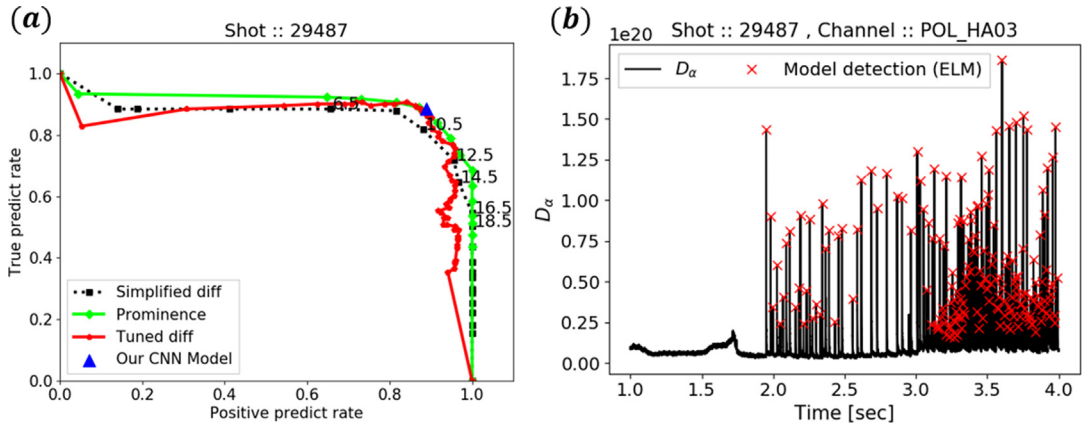

Fig. 10. (a) ROC curve of several ELM detection algorithms from D_α signal of KSTAR experiment #29487, which was used as a test set, and (b) peak detections by our model. The simplified difference deviation method [26] (black square dashed), prominence method [8] (green dot) threshold, and tuned difference method [9] (red pentagon) have optimized thresholds of 10.5, 1.6×10^{19}, and $SD_{1,th} = SD_{2,th}/3 = 4.3 \times 10^{18}$, respectively. Our trained model's result is shown in blue triangle point. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

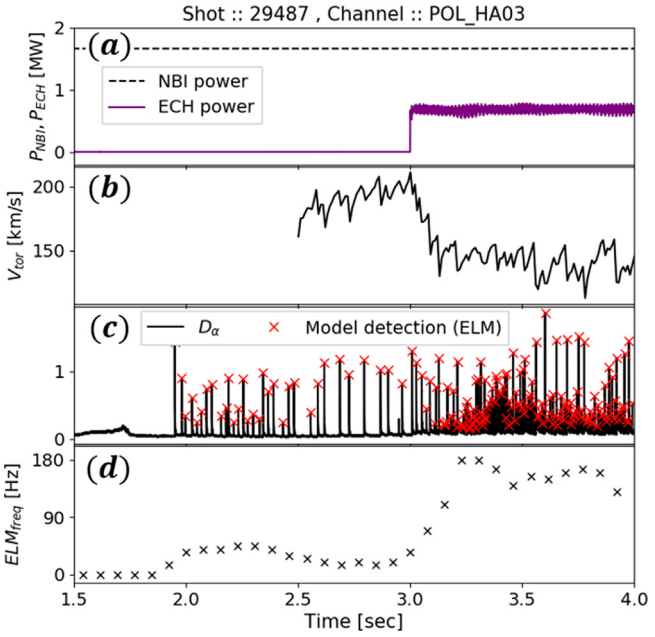

Fig. 11. Time history of (a) heating power, (b) measured toroidal rotation, (c) our ELM detection from the measured D_α signal, and (d) ELM frequency calculated from ELM detection results of (c) KSTAR experiment #29487, which is the same case as that shown in Fig. 1.

be useful for ELM analysis between multiple experiments in the KSTAR and other tokamaks, such as the ITER. The machine learning techniques can be used to detect the precursor signal of the disruption, which is one of the critical issues for the success of ITER. We expect this methodology to be applicable to other engineering fields involving a significant amount of time-series data and the identification of some features.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

J.H.S. and J.P.L. were supported by the National R&D Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science and ICT (NRF-2021M1A7A4091137, NRF-2021M3F7A1084421). Y.C.G was supported by NRF Grant No. RS-2022-00155917. This work was also supported by Ministry of Science and ICT under KFE R&D Program of "KSTAR Experimental Collaboration and Fusion Plasma Research (KFE-EN2201-13)".

References

[1] H. Zohm, The physics of edge localized modes (ELMs) and their role in power and particle exhaust, Plasma Phys. Control Fusion 38 (1996) 1213–1223.
[2] A.W. Leonard, Edge-localized-modes in tokamaks, Phys. Plasmas 21 (2014), 090501.
[3] A. Loarte, et al., Characteristics of type I ELM energy and particle losses in existing devices and their extrapolation to ITER, Plasma Phys. Control Fusion 45 (2003) 1549.
[4] W. Suttrop, The physics of large and small edge localized modes, Plasma Phys. Control Fusion 42 (2000) A1.
[5] P.B. Snyder, et al., ELMs and constraints on the H-mode pedestal: peeling–ballooning stability calculation and comparison with experiment, Nucl. Fusion 44 (2004) 320.
[6] C. Ham, A. Kirk, S. Pamela, et al., Filamentary plasma eruptions and their control on the route to fusion energy, Nat. Rev. Phys. 2 (2020) 159–167.
[7] S.W. Yoon, et al., Characteristics of the first H-mode discharges in KSTAR, Nucl. Fusion 51 (2011), 113009.
[8] J. Kim, et al., Nonlinear energy transfer from low frequency electromagnetic fluctuations to broadband turbulence during edge localized mode crashes, Nucl. Fusion 60 (2020), 124002.
[9] D. Eldon, et al., Controlling marginally detached divertor plasmas, Nucl. Fusion 57 (2017), 066039.
[10] M. Berta, et al., Automatic ELM detection using gSPRT on the COMPASS tokamak, Fusion Eng. Des. 123 (2017) 950–954.
[11] R. Panet, et al., Status of the COMPASS tokamak and characterization of the first H-mode, Plasma Phys. Control Fusion 58 (2015), 014015.
[12] A. Murari, et al., Extensive statistical analysis of ELMs on JET with a carbon wall, Plasma Phys. Control. Fusion 56 (2014), 114007.
[13] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105.
[14] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[15] G. Shin, J.-W. Juhn, G.I. Kwon, S.-H. Hahn, Real-time classification of L–H transition and ELM in KSTAR, Fusion Eng. Des. 157 (2020), 111634.
[16] Semin Joung, et al., Deep neural network Grad-Shafranov solver constrained with measured magnetic signals, Nucl. Fusion 60 (2020), 016034.
[17] P. Ronneberger, T. Brox Fischer, U-net: convolutional networks for biomedical image segmentation, in: medical image computing and computer-assisted intervention (MICCAI), Lecture Notes Comput. Sci. Springer Cham. 9351 (2015) 234–241.
[18] A.D. Melnikov, Y.P. Tsentalovich, V.V. Yanshole, Deep learning for the precise

peak detection in high-resolution LC-MS data, Anal. Chem. 92 (2020) 588–592.
[19] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: medical image computing and computer-assisted intervention (MICCAI), Lecture Notes Comput. Sci. Springer Cham. 9901 (2016) 424–432.
[20] Q. Huang, J. Sun, H. Ding, X. Wang, G. Wang, Robust liver vessel extraction using 3D U-Net with variant dice loss function, Comput. Biol. Med. 101 (2018) 153–162.
[21] J.H. Lee, et al., Diagnostics for first plasma and development plan on KSTAR, Rev. Sci. Instrum. 81 (2010), 063502.
[22] J.G. Bak, et al., Initial measurements by using Mirnov coils in the KSTAR machine, in: Proceeding of the 37th EPS Conference on Plasma Physics p. 5 102, 2010.
[23] T. Kass, et al., Characteristics of type I and type III ELM precursors in ASDEX upgrade, Nucl. Fusion 38 (1998) 111.
[24] V. Pascal, H. Larochelle, Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (2010) 3371–3408.
[25] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning, 2010, pp. 807–814.
[26] P.K. Sahoo, S. Soltani, A.K.C. Wang, Y.C. Chen, A survey of thresholding techniques, Comp. Vis. Gr., Image Process. 41 (1988) 233–260.
[27] M. Llobera, Building past landscape perception with GIS: understanding topographic prominence, J. Archaeolog. Sci. 28 (2001) 1005.
[28] M. Basu, Gaussian-based edge-detection methods-a survey, IEEE Trans. Syst. Man. Cybern., Part C: App. Rev. 32 (2002) 252–260.
[29] Y. Kamada, et al., Disappearance of giant ELMs and appearance of minute grassy ELMs in JT-60U high-triangularity discharges, Plasma Phys. Control Fusion 42 (2000) A247.
[30] J.-W. Ahn, et al., Confinement and ELM characteristics of H-mode plasmas in KSTAR, Nucl. Fusion 52 (2012), 114001.
[31] N. Oyama, et al., Effects of edge collisionality on ELM characteristics in the grassy ELM regime, Nucl. Fusion 50 (2010), 064014.
[32] N. Oyama, et al., ELM frequency dependence on toroidal rotation in the grassy ELM regime in JT-60U, Plasma Phys. Control Fusion 49 (2007) 249.
[33] J. Schirmer, et al., The radial electric field and its associated shear in the ASDEX Upgrade tokamak, Nucl. Fusion 46 (2006) S780.
[34] A. Cathey, et al., Non-linear extended MHD simulations of type-I edge localised mode cycles in ASDEX Upgrade and their underlying triggering mechanism, Nucl. Fusion 60 (2020), 124007.
[35] J. Seol, et al., Study of type III ELMs in the KSTAR tokamak, 25th IAEA fusion energy conference, IAEA (2014).
[36] J. Rapp, et al., Integrated scenario with type-III ELMy H-mode edge: extrapolation to ITER, Nucl. Fusion 49 (2009), 095012. Q4