Disruption prediction and its analysis using multimodal data in KSTAR via Deep Learning

JinSu KIM, Yong-Su Na

asdwlstn@snu.ac.kr, ysna@snu.ac.kr

Department of Nuclear Engineering, Seoul National University, Seoul, Korea

PLARE

Plasma Laboratory for Advanced REsearch

2022.10.12

SEOUL NATIONAL UNIVERSITY

Contents

- ➢ **Introduction**
- **Method**
- **Result**
- **Discussion**
- **Conclusion**

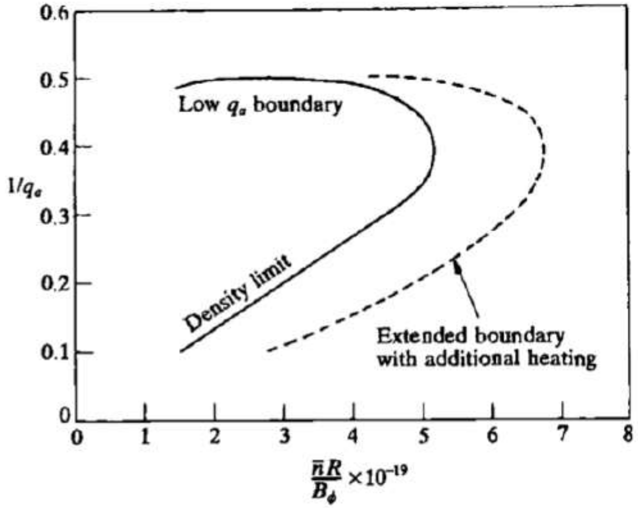
PLARE

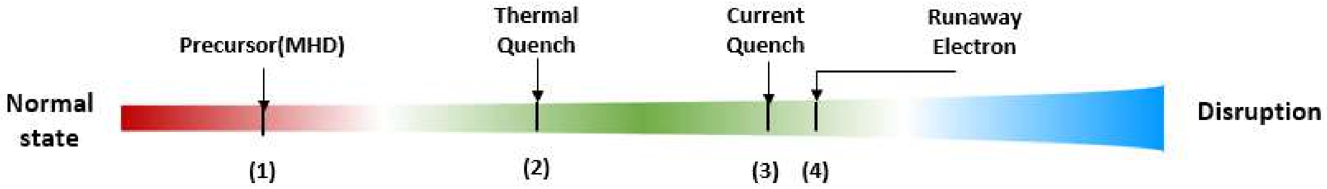
Introduction

- **Plasma disruption**

  - **Definition**

- Global and sudden losses of plasma which result in abrupt termination with massive magnetic and thermal energies and consequently transfers harmful damages to the device[1].
- There are 4 phases in disruption phenomenon : Pre-precursor phase, Precursor phase, Fast phase, Current phase
- Many different precursor events and quench events are possible and induced by different physical mechanisms[2].
- Two basic causes of disruption
  - Low q disruptions : $q_a \geq 2$
  - Density limit disruptions : $\frac{1}{q_a} = \alpha \frac{(n_e n_z)^{\frac{5}{8}} R}{B_\phi}$





Introduction

▪ Plasma disruption

❑ General process of disruptions

- The **evolution of an unstable current profile** leading to the growth of a tearing mode (m = 2 mode being particularly important)
- The **nonlinear growth** of the **tearing mode**
- A **sudden relaxation of the equilibrium** : current profile being flatten and dramatic loss of confinement with a collapse of plasma temperature → **Thermal quench (TQ)**
- The **total current decays** → **Current quench (CQ)**
- The **increased toroidal E-field** associated with increased plasma resistance generates **runaway electrons**, which carries a **large current** and sometimes **persists after CQ**
- Both the loss of plasma energy and the current decay **induce currents in the vessel** which can produce **large forces on the vessel**.

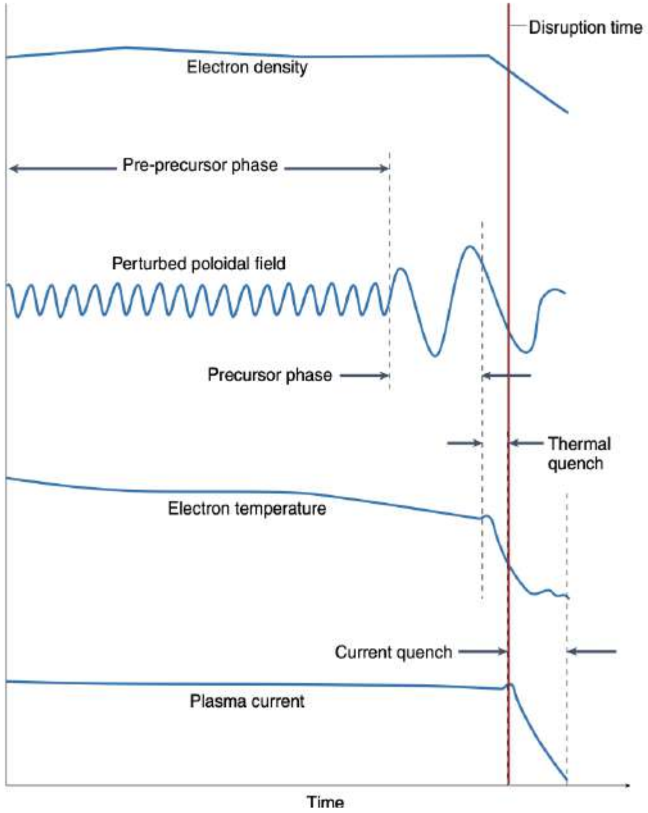
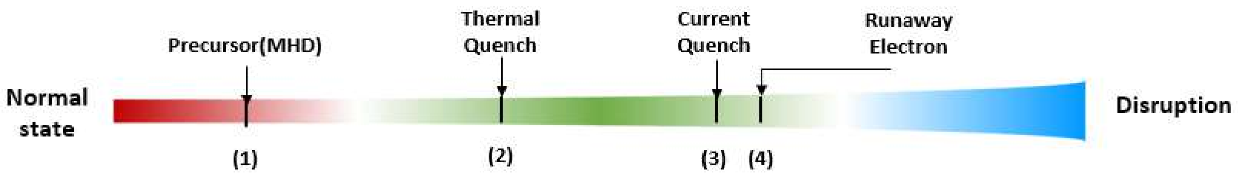


Fig 1, J.Vega et al, 2022, Nature Physics



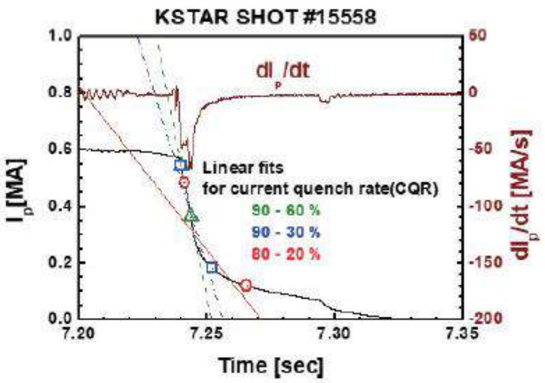
Introduction

▪ Plasma disruption

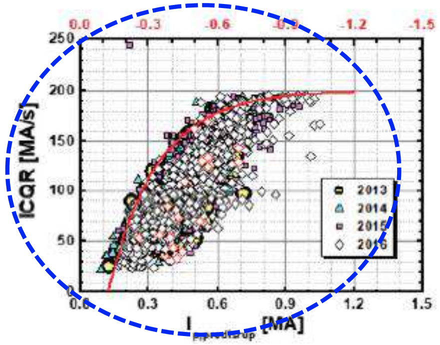
❑ Characteristics of plasma quench in the KSTAR tokamak

- There was the linear dependence between ICQR and plasma current less than 0.6MA, but ICQR was saturated as increasing plasma current more than 0.6MA → **exponential structure** estimated
- **Larger discrepancy** between the LCQR and ICQR : **due to the long tail** at the level of less than 30% of plasma current induced by the contribution of the Runaway Electrons as reported in JET
- The **current quench rate does not linearly depend** upon the **magnitude of plasma current**.

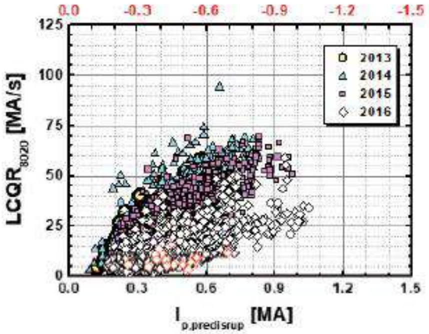
Upper bound of $(\frac{dI_p}{dt})_{max}$ from exponential fit



(a)



(b)



(c)

* ICQR : Instantaneous current quench rate, evaluated by exponential fit

* LCQR : Linear current quench rate, evaluated by the linear slope

* CQR : Current quench rate

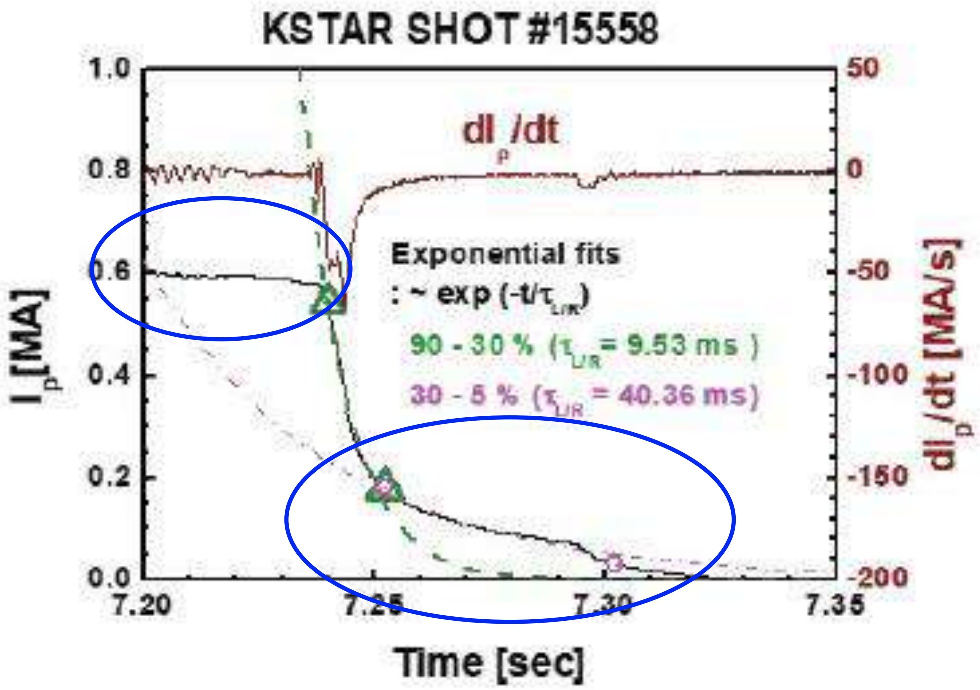
[BAK, J. G., et al. *Characteristics of plasma current quench during disruptions in the KSTAR tokamak*. 2018]

Introduction

- **Plasma disruption**

❑ <u>Characteristics of plasma quench in the KSTAR tokamak</u>

- **Due to the long tail**, the current quench curve had **a double exponential decay structure** with faster and slower R/L times.
- The slower slope might be due to the **formation of the RE plateau** at the lower plasma current under 0.1MA in the phase of the quench.
- The RE plateau had **small slope** which was different from the RE plateau with almost constant level, reported in the JET.



[BAK, J. G., et al. *Characteristics of plasma current quench during disruptions in the KSTAR tokamak*. 2018]

Introduction

- **Disruption prediction**

❑ **Importance and Limitation**

- Disruption carries large amount of magnetic and thermal energy loss and causes harmful damages on the device.

  → the prediction, avoidance and mitigation of disruption is **prerequisite** for tokamak fusion.

- Conventional approaches by MHD theory and simulation have remained **limitations**.

  → Disruption is **highly non-linear dynamics** with complex interaction of different physical processes.

- Data-driven approaches based on a posteriori observation can be alternative for disruption predictor

  - Machine Learning : Random Forest, Catboost, Xgboost, Light GBM, GBM, Decision Tree Classifier, SVM
  - Deep Learning : Neural Network (e.g. FRNN, https://github.com/PPPLDeepLearning/plasma-python)

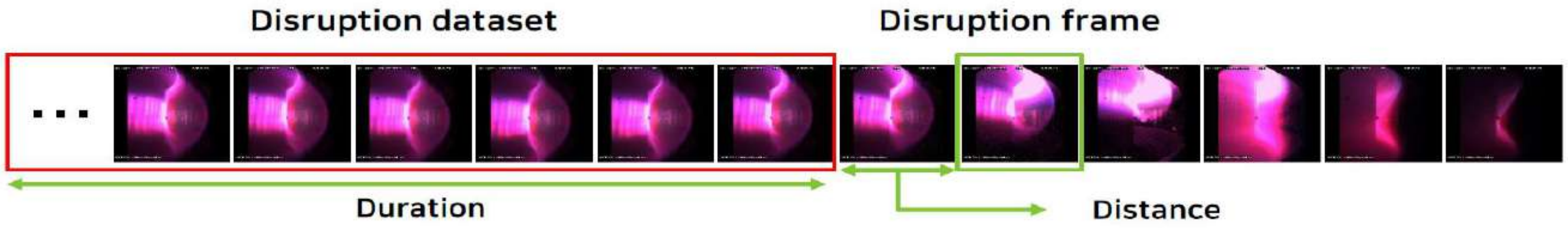
➡ **Disruption predictor model based on Deep Neural Network**

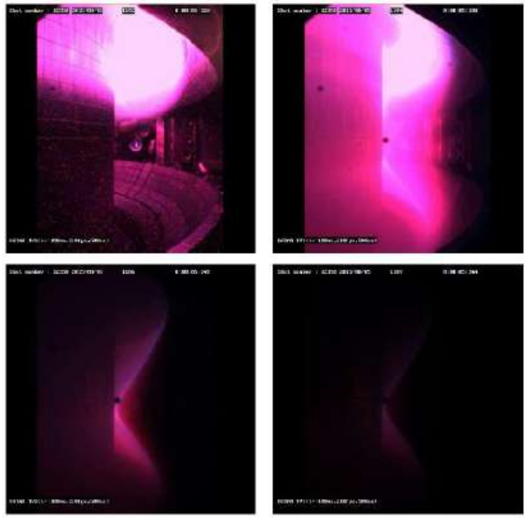
PLARE

Introduction

- **Motivation : Why use video data?**

❑ **KSTAR IVIS data : Real-time Image sequence (Video) data**





Tokamak visible image sequence recognition using nonlocal spatio-temporal CNN for attention needed area localization*

Giil Kwon *, Hanmin Wi, Jaesic Hong

Control Team, National Fusion Research Institute, Daejeon, South Korea

ARTICLE INFO

Keywords:
Tokamak visible image diagnostic system
Deep learning
Video classification

ABSTRACT

In this paper, we report a study that was conducted to explore the feasibility of developing a system that classifies the image sequence as 'disruptive' shot images or 'non-disruptive' shot images. The classifier identifies an image sequence as 'non-disruptive' shots or 'disruptive' shots using a non-local spatio-temporal 3D convolution neural network (CNN) from image sequences from the Tokamak visible imaging diagnostic system. To analyze the classification result, we localize an area that has contributed to the classification of an image sequence. We use class activation mapping (CAM) for CNN with global average pooling to localize the area. To train this classifier, we created a plasma disruption image sequence dataset using the data acquired from the KSTAR experiment. This classifier recognized disruption image sequences on the test dataset with 91.11% accuracy. Analysis of the CAMs of these image sequences revealed that this classifier recognizes the disruption of the plasma with a relative change in brightness over time in areas other than the plasma area of the image. Through this work, we will be able to develop a system that automatically classifies plasma disruption image sequence after experiments.

1. Introduction

The in-vessel visible inspection system (IVIS) is used for monitoring the plasma and interior of the vessel in the Korean Superconducting Tokamak Advanced Research (KSTAR) device. Using the image acquired from the IVIS, the experiment participants analyze images to obtain information, such as the shape and location of the plasma. It is thought that the images generated by the IVIS contain more information that has not yet been discovered in addition to the shape and location of the plasma. Participants in KSTAR's experiments have only analyzed one image at a time to obtain the plasma information by manually inspecting each image [1]. Rather than analyzing only the static information (such as the plasma position and shape) of the plasma obtained from a single image at a specific point in time, we can better understand the state of the plasma if we also analyze the motion information (such as the plasma position or shape change) that the plasma moves over time. Motion information can be calculated by obtaining dynamic features (such as pixel displacement between two frames, change in brightness of pixels between two frames) that appear over time. However, these dynamic features were difficult to find manually, as was previously done by researchers, because the amount of image data is very large, and dynamic features appearing in image sequences are more difficult to capture than features obtained statically from a single image. The IVIS produces massive amounts of image data (480 × 640 resolution images at 210 frames/s, 15.75 MB/s) using a colored programmable progressive scan charge-coupled device CCD camera [1]. The IVIS produces many images, and these images must be analyzed between shot intervals so that the analysis results can be applied to the plasma control of the next shot. When the plasma pulse length is a 300 s shot, the operator must examine 63,000 images by hand to analyze the plasma status (KSTAR milestone is achieving a 20–300 s long-pulse H-mode operation). Therefore, it is necessary to develop an approach that can analyze the dynamic characteristics of images by recognizing a sequence of images over time. In this paper, we developed an IVIS image sequence classifier using deep learning for a feasibility test in developing a program that analyzes the dynamic characteristics of image sequences from the IVIS to test its performance. To classify an image sequence, we use non-local inflated 3D convolution networks (I3D) [2]. I3D is an inflated 2D convolution network. By inflating the kernel of the 2D convolution network (ConvNets), we can make I3D from 2D ConvNets. I3D is widely

Giil Kwon et al, Fusion Engineering and Design, 2021

- **Video data** contain **spatial-temporal information** including **time-varying position** and **shape** of plasma.
- **Vertical Displacement Event** may **be captured** from video directly.
- In the computer vision area, there are many implemented codes which show **high performance** with several tasks (e.g. Video Action Recognition, Video Prediction)
- Kwon et al used video data to classify the disruptive event and showed the neural network recognized disruption image using relative change in brightness from the plasma area.

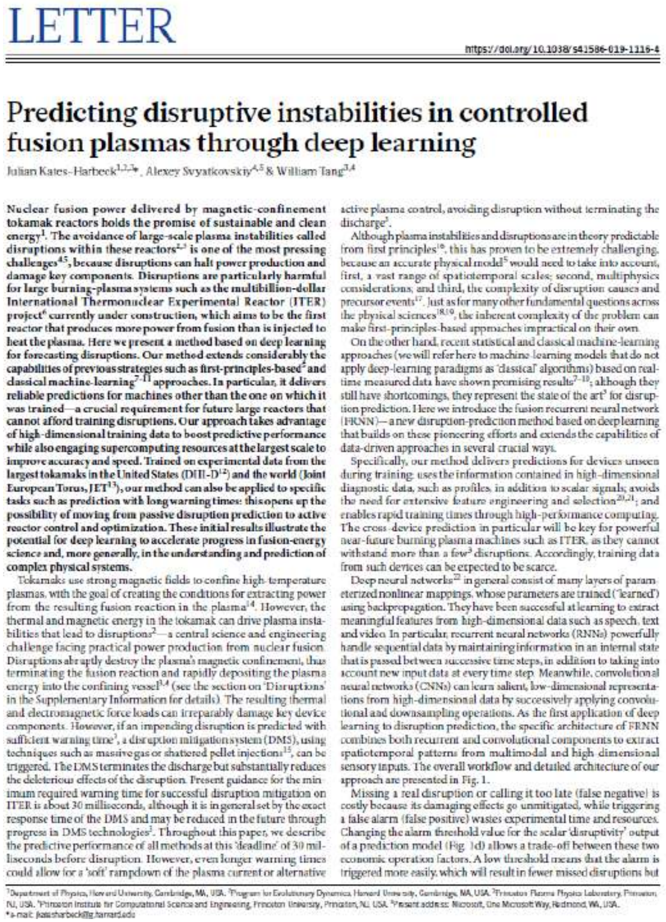
➡ **Disruption prediction using Video data**

PLARE

Introduction

- **Related work**

❑ **Research for predicting disruptions with deep learning**



LETTER

https://doi.org/10.1038/s41586-019-1116-4

Predicting disruptive instabilities in controlled fusion plasmas through deep learning

Julian Kates-Harbeck[1,2,3]*, Alexey Svyatkovskiy[4,5] & William Tang[3,4]

- Paper : Predicting disruptive instabilities in controlled fusion plasmas through deep learning
- RNN-based deep learning algorithm with 0-D signals and 1D profiles
- Predictive performance : successful prediction of disruption prior to 30ms
- Cross-machine experiment also proceeded : D3D, JET(without 1D profiles)

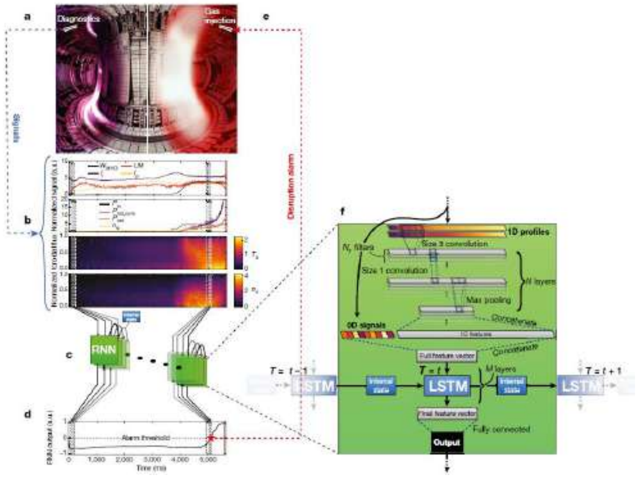


Fig 1, Julian Kates Harbeck et al, 2019, Nature

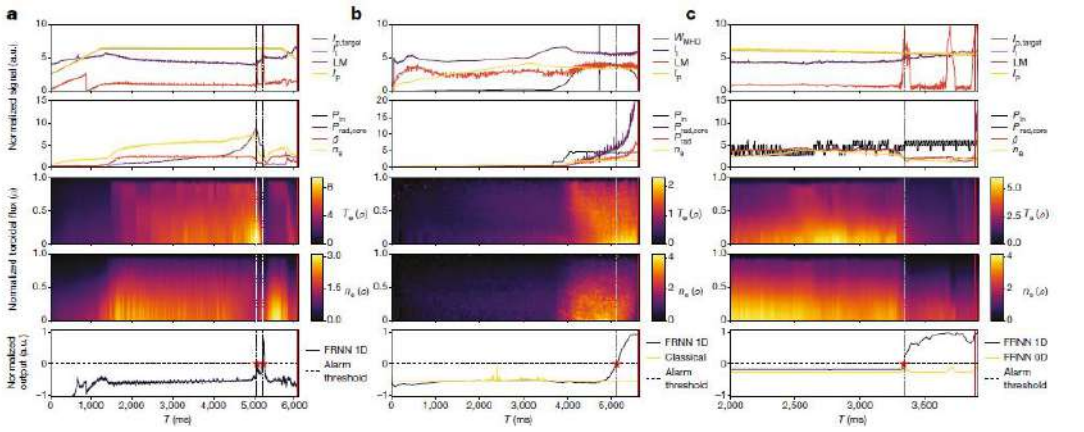


Fig 2, Julian Kates Harbeck et al, 2019, Nature

PLARE

Introduction

- **Related work**

  - ❑ **<u>Expansion and Variation : Application with different Neural Network architectures and data modality (type)</u>**

Disruption prediction using a full convolutional neural network on EAST

B H Guo[1,2], B Shen[1], D L Chen[1], C Rea[3], R S Granetz[3], Y Huang[1], L Zeng[1], H Zhang[4], J P Qian[1], Y W Sun[1] and B J Xiao[1,2]

[1] Institute of Plasma Physics, HFIPS, Chinese Academy of Sciences, Hefei 230031, People's Republic of China
[2] University of Science and Technology of China, Hefei 230026, People's Republic of China
[3] MIT Plasma Science and Fusion Center, Cambridge, MA 02139, United States of America
[4] Chongqing University of Posts and Telecommunications, Chongqing 400065, People's Republic of China

E-mail: cdalong@ipp.ac.cn and yaohuang@ipp.ac.cn

Received 10 October 2020, revised 9 November 2020
Accepted for publication 18 November 2020
Published 17 December 2020

CrossMark

Abstract

In this study, a full convolutional neural network is trained on a large database of experimental EAST data to classify disruptive discharges and distinguish them from non-disruptive discharges. The database contains 14 diagnostic parameters from the ~10^4 discharges (disruptive and non-disruptive). The test set contains 417 disruptive discharges and 999 non-disruptive discharges, which are used to evaluate the performance of the model. The results reveal that the true positive (TP) rate is ~0.827, while the false positive (FP) rate is ~0.067. This indicates that 72 disruptive discharges and 67 non-disruptive discharges are misclassified in the test set. The FPs are investigated in detail and are found to emerge due to some subtle disturbances in the signals, which lead to misjudgment of the model. Therefore, hundreds of non-disruptive discharges from training set, containing time slices of small disturbances, are artificially added into the training database for retraining the model. The same test set is used to assess the performance of the improved model. The TP rate of the improved model increases up to 0.875, while its FP rate decreases to 0.061. Overall, the proposed data-driven predicted model exhibits immense potential for application in long pulse fusion devices such as ITER.

Keywords: plasma, disruption prediction, deep learning, convolutional neural network

(Some figures may appear in colour only in the online journal)

1. Introduction

Disruption events are an inevitable and prominent issue that must be resolved to ensure the implementation of CFETR

[2]. However, for mitigation to be effective, a certain response time needs to be considered. On EAST, the response time of the MGI system, which is a main component of the disruption mitigation system, from triggering to entering the core is

Deep Learning for Plasma Tomography and Disruption Prediction from Bolometer Data

Diogo R. Ferreira, Pedro J. Carvalho, Horácio Fernandes, and JET Contributors

***Abstract*—The use of deep learning is facilitating a wide range of data processing tasks in many areas. The analysis of fusion data is no exception, since there is a need to process large amounts of data collected from the diagnostic systems attached to a fusion device. Fusion data involves images and time series, and are a natural candidate for the use of convolutional and recurrent neural networks. In this work, we describe how CNNs can be used to reconstruct the plasma radiation profile, and we discuss the potential of using RNNs for disruption prediction based on the same input data. Both approaches have been applied at JET using data from a multi-channel diagnostic system. Similar approaches can be applied to other fusion devices and diagnostics.**

***Index Terms*—Nuclear Fusion, Plasma Diagnostics, GPU Computing, Deep Learning**

I. Introduction

Deep learning [1] has become the state-of-the-art approach to many problems, especially those related to image processing and natural language processing. Convolutional neural networks (CNNs) have been extremely successful in image classification [2], [3], image segmentation [4], [5] and object detection [6], [7], to cite only a few examples. On the other hand, recurrent neural networks (RNNs) have been used for speech recognition [8], [9], language modeling [10], [11] and machine translation [12], [13], among other applications.

In general, it could be said that CNNs are appropriate for problems involving images and computer vision, whereas RNNs are especially useful for text and other sequential data, including time series [14]. However, this distinction is not clear-cut since, for example, it is possible to analyze images with RNNs [15], it is possible to perform sequence learning with CNNs [16], and there are hybrid models combining features from both CNNs and RNNs [17].

For the purpose of this work, we will focus on two broad categories: (1) image processing for plasma tomography

Here, we focus on JET (Joint European Torus), a D-shaped tokamak with a major radius of 2.96 m and a minor radius of 1.25–2.10 m. JET has a vast assortment of diagnostics, including magnetic coils to measure plasma current and instabilities, interferometers and reflectometers to measure plasma density, Thomson scattering to determine the electron temperature, spectroscopy to measure ion temperature, and X-ray cameras to measure electromagnetic radiation, among others.

In this work, we will be using on a specific diagnostic, the bolometer system [18], which measures the plasma radiation on a poloidal cross-section of the fusion device. The signals collected from the bolometer system can be used to monitor the plasma state across an entire pulse. Several phenomena, such as impurity transport and accumulation at the plasma core, can be detected from the bolometer signals. Since these impurity-related phenomena are one of the most frequent precursors of disruptions at JET, this diagnostic plays an important role in disruption studies as well.

The data coming from this diagnostic is the basis for tomographic reconstructions that provide a 2D image of the plasma radiation profile. The reconstruction process itself is time-consuming. However, with a CNN trained on a large collection of sample tomograms, it becomes possible to produce those results much faster and with high accuracy.

In addition, the bolometer signals can be used to study disruption precursors. With a RNN trained on these signals, it is shown that bolometer data can provide a useful input for disruption prediction, both in terms of probability of disruption and time remaining to an impending disruption.

The paper is structured as follows. Section II provides a brief overview of the bolometer system, where the data is coming from, and of the tomographic method used at JET to reconstruct the plasma radiation profile. Section III describes the CNN that has been developed for plasma tomography, and

Disruption predictor based on neural network and anomaly detection on J-TEXT

W Zheng[1], Q Q Wu[1], M Zhang[1,3], Z Y Chen[1], Y X Shang[2], J N Fan[2], Y Pan[2] and J-TEXT Team[1]

[1] International Joint Research Laboratory of Magnetic Confinement Fusion and Plasma Physics, State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, 430074, People's Republic of China
[2] School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China

E-mail: zhengwei@hust.edu.cn and zhangming@hust.edu.cn

Received 10 October 2019, revised 19 December 2019
Accepted for publication 13 January 2020
Published 25 February 2020

CrossMark

Abstract

Disruption prediction is essential for the safe operation of a large scale tokamak. Existing disruption predictors based on machine learning techniques have good prediction performance, but all these methods need large training datasets including many disruptions to develop their successful prediction capability. Future machines are unlikely to provide enough disruption samples since these cause excessive machine damage and the prediction models used are difficult to extrapolate to a machines that the predictor was not trained on. In this paper, a disruption predictor based on a deep learning and anomaly detection technique has been developed. It regards the disruption as an anomaly, and can learn on non-disruptive shots only. The model is trained to extract the hidden features of various non-disruptive shots with a convolutional neural network and a long-shot term memory (LSTM) recurrent neural network. It will predict the future trend of selected diagnostics, then using the predicted future trend and the measured signal to calculate an outlier factor to determine if a disruption is coming. It was tested with J-TEXT discharges in flat top phase and can demonstrate comparable performance to current machine learning disruption prediction techniques, without requiring a disruption data set. This could be applied to future tokamaks and reduce the dependency on disruptive experiments.

Keywords: major disruption, disruption prediction, deep learning, anomaly detection

(Some figures may appear in colour only in the online journal)

Introduction : Basic concept

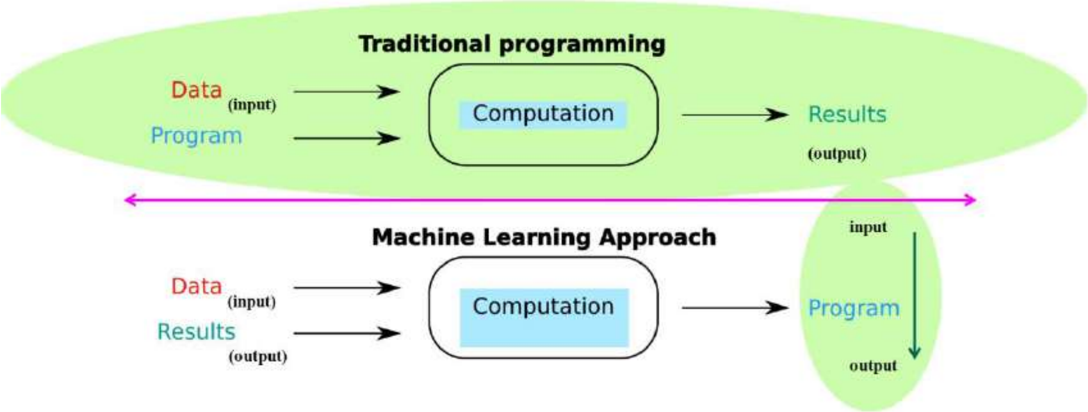
- **Prediction and Forecast**
  - Prediction : Data-driven approach
    - An estimate of future events from subjective considerations
    - Probabilistic statement
    - Based on intuition
    - The results of predictions are dependent upon unique representations

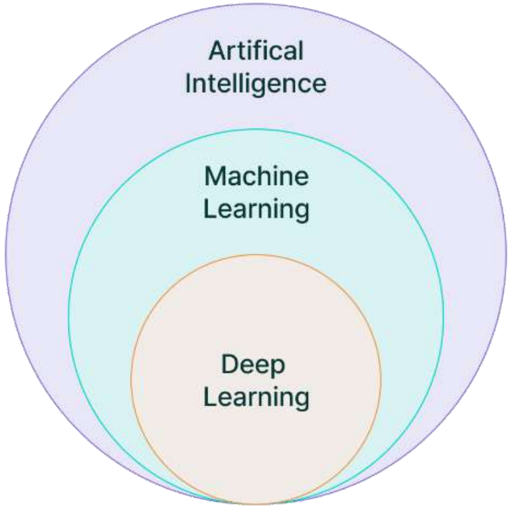
  - Forecasting : Model-driven approach
    - An estimation of future events by incorporating and casting forward data and systematic manner(physics)
    - Definitive and specific statement (Deterministic)
    - Based on statistical model
    - The results of forecasting are replicable

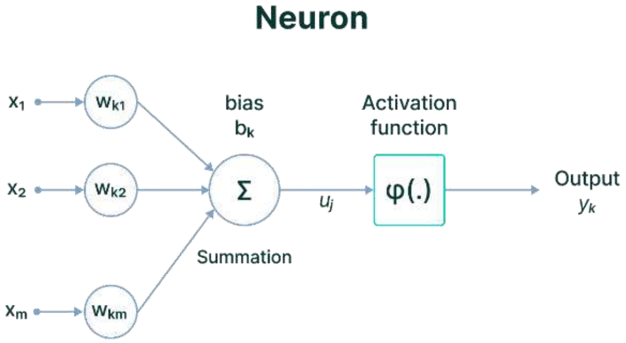
Introduction : Basic concept

▪ Machine Learning and Deep Learning

- **Machine Learning :** A set of methods that can automatically **detect patterns** in data → **Data representation**
- **Components :** Dataset (Training data + Test data), Probabilistic model, Objective function, Algorithm (How to learn**)**
- **Goal of machine learning :** to **learn a mapping** from **input data distribution** to **output data distribution** by optimizing the objective function with given dataset and learning algorithms → **to Learn probability distribution $\boldsymbol{p(y|x, D)}$**
- **Deep Learning :** A subset of machine learning algorithm which use **Artificial Neural Network** as a function approximator



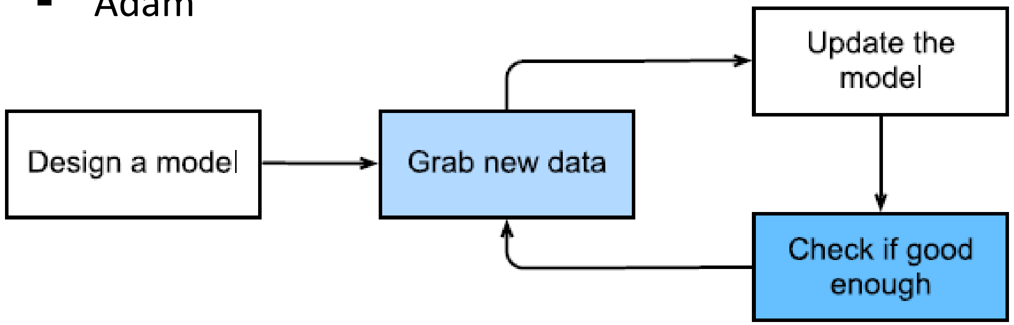




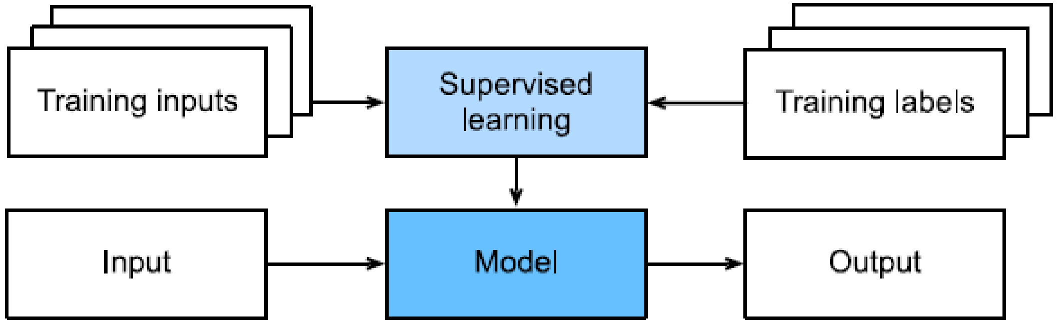
Introduction : Basic concept

▪ Machine Learning and Deep Learning

- **Dataset** : A set of data point collected from given prior data distribution
- **Model** : A set of mapping functions which generate the mapping from input data to output data or probability distribution.
- **Objective function** : A set of **targets** as formal measures of **how good the models are**. **Training** or **Learning** is equal to **optimization process of objective function**. (We also call it as '**Loss function**')
- **Algorithm** : A set of methods for optimizing a well-defined objective function given from the task
    - Stochastic Gradient Descent (SGD)
    - Momentum method
    - Adagrad
    - Adam



Training process per batch data

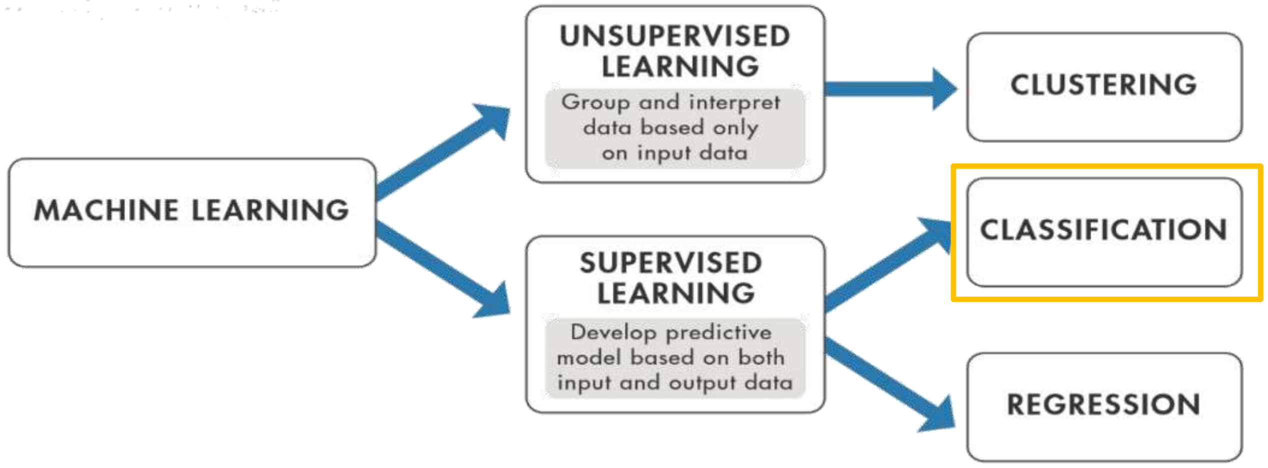


Simple process for supervised learning

Introduction : Basic concept

▪ Disruption prediction and Deep Learning

- **To learn the disruption prediction** is **to learn whether the state of the plasma** estimated from the given data **is disruptive or non-disruptive** → Binary classification
- We can train the neural network with supervised learning
  - Input : Plasma 0D, 1D data, IVIS data, Tomography …
  - Output : Disruption probability, Binary label (0 : disruptive, 1 : non-disruptive)



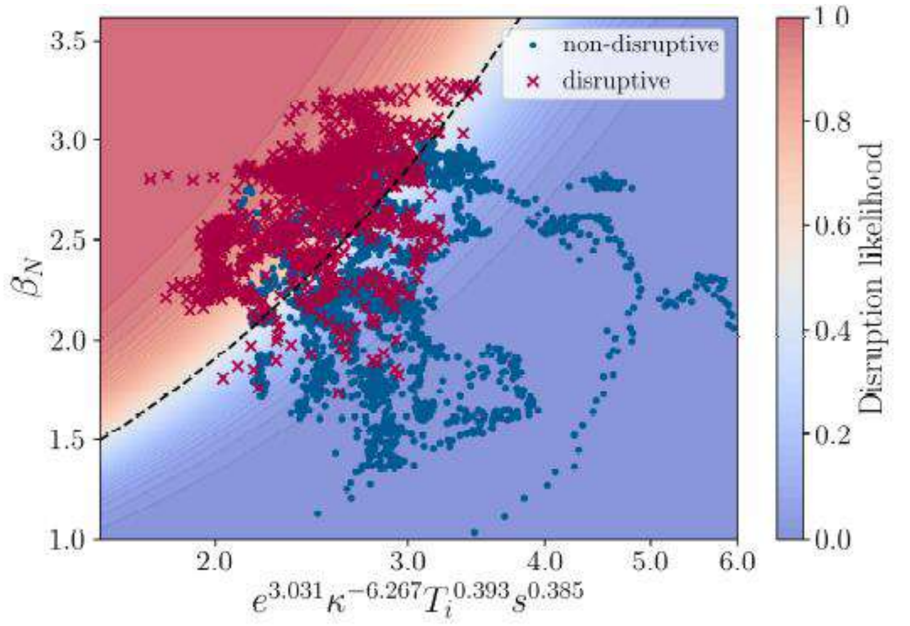


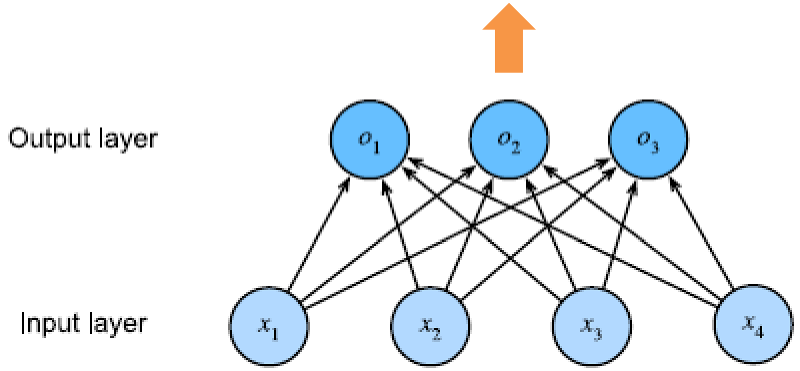
Fig 15. from 2022 Review of Data-Driven Plasma Science, Rushil Anirudh et al

Introduction : Basic concept

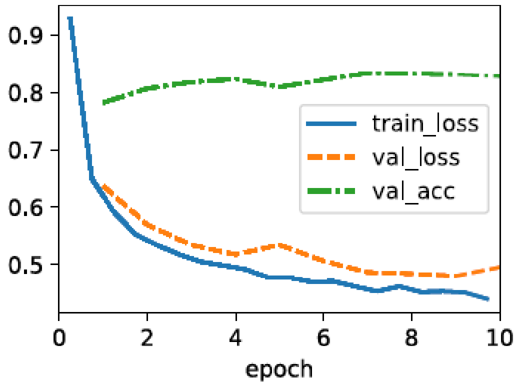
▪ Disruption prediction and Deep Learning

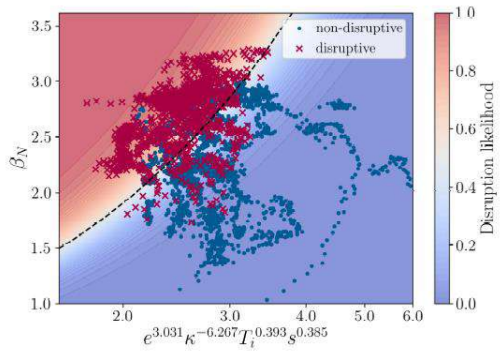
- The **objective** for binary classification : **to maximize likelihood** which means conditional probability of output y with given input x
- The objective function (loss function) : Cross Entropy Loss
- ***Maximization of likelihood = Minimization of cross entropy loss***

$$y_i = \frac{\exp(o_j)}{\sum \exp(o_j)} : P(y_i|x_i, \theta, D)$$



$$\max_{\theta} \sum logP(y_i|x_i, \theta, D) \longrightarrow \min_{\theta} \sum -y_i \log P(y_i|x_i, \theta, D)$$





PLARE

Introduction : Basic concept

- ## Issues on Deep Learning

  - **The goal of machine learning :** to learn general patterns / data distribution with given data
  - If we increase model complexity or use few data compared to the model complexity, generalization error increases
  - **Generalization :** Fundamental issue of machine learning about how to discover general patterns from given data
  - **Overfitting :** the phenomenon of **fitting closer** to the **training data than** to the **underlying distribution** (prior distribution)
  - **Underfitting** : the phenomenon of **limiting** the **reduction of training error** due to the **low complexity** of model or **small size** of data

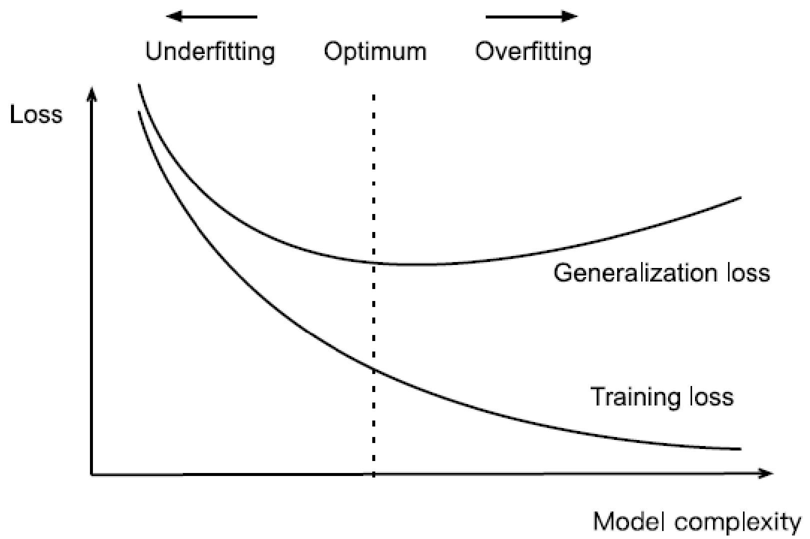
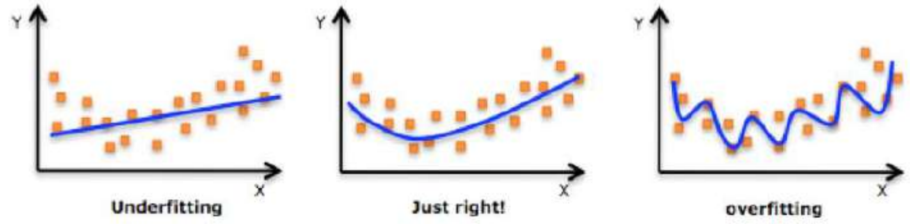


Fig 3.6.1. from Dive into Deep Learning, Aston Zhang et al



Complexity: low
Sample bias: low

Complexity: High
Sample bias: High

Introduction : Basic concept

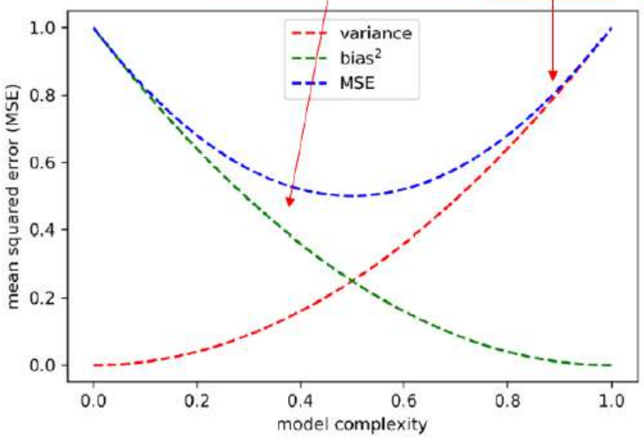
- **Issues on Deep Learning**

  - A good statistical model minimizes the loss by **finding the optimal balance between bias and variance**
  - Factors for underfitting

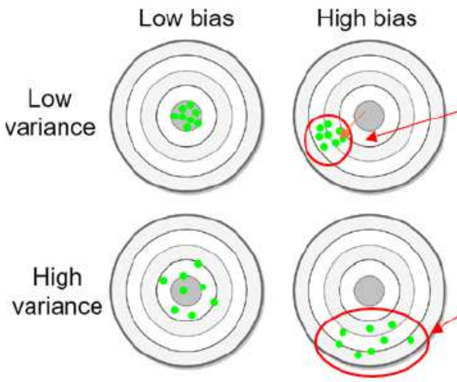
    (1) **High bias and low variance** (2) **Not enough features** of training dataset (3) The model is **too simple** (4) Noise on data
  - Factors for overfitting

    (1) **High variance and low bias** (2) The model is **too complex** (3) **Not enough size** of training dataset

$$E[(y-\hat{f}(x)^2)] = \underbrace{\boxed{E[(f(x)-\hat{f}(x))^2]}}_{\text{Bias Squared}} + \underbrace{\boxed{V[\hat{f}(x)\ ]}}_{\text{Variance}} + \underbrace{\boxed{V[\epsilon]}}_{\text{Noise}}$$



$$E[(y-\hat{f}(x)^2)] = \underbrace{E[(f(x)-\hat{f}(x))^2]}_{\text{Bias Squared}} + \underbrace{V[\hat{f}(x)\ ]}_{\text{Variance}} + \underbrace{V[\epsilon]}_{\text{Noise}}$$



★Noise: caused by datasets

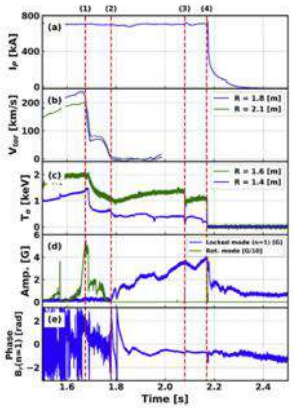
Introduction

- **Aims of this research**
  - **Step 1. Development of plasma disruption prediction model using Deep Learning**
    - KSTAR IVIS Video data : Image sequence data / 210 fps and 480 X 640 resolution
    - KSTAR 0D data : Time-series data / 20kHz sampling rate
    - Multimodal data : IVIS data + 0D data
  - **Step 2. Analysis of the experimental results**
    - Experiment for model comparison
    - Experiment for different prediction time
    - Experiment for different learning algorithms (due to imbalance data distribution)
    - Image localization with attention rollout
    - Visualization of the latent vectors for single-modal(Video, 0D) / multi-modal (Video + 0D) data
  - **Step 3. Real-time prediction with test shot**
    - Continuous prediction with video / 0D / Multimodal data

Introduction

- ## Short summary

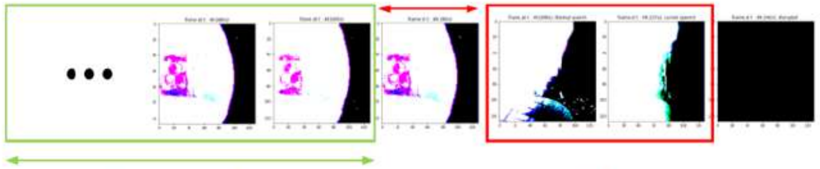
Single-modal data (Video, 0D) / Multi-modal data(Video + 0D)



Jeongwon Lee et al., Fusion Eng. and Des., 2019

0D data
(batch size, sequence length, number of variables)

Distance or Prediction time

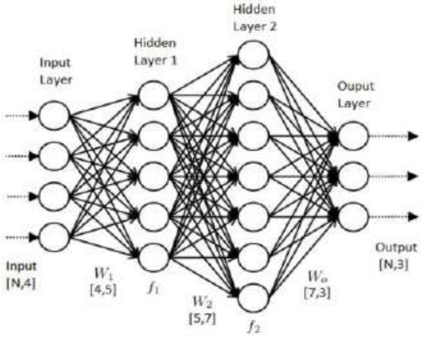


Duration or Sequence length

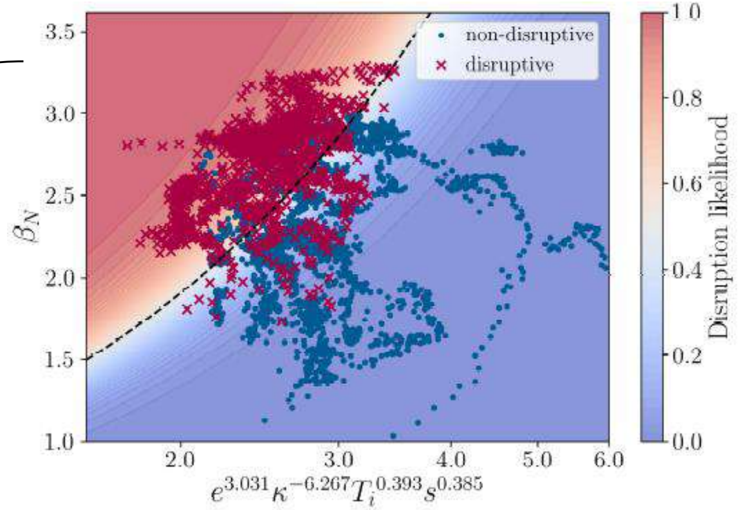
Disruption frame
(Thermal Quench + Current Quench)

Image sequence data
(batch size, sequence length, channel, height, width)

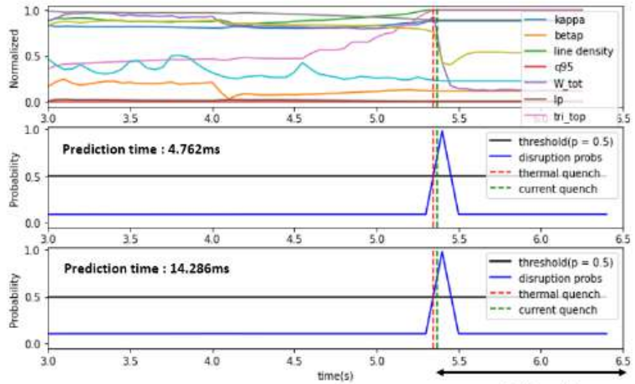
Neural Network



Disruptive classification



Shot 21310, 0D data and disruption probability curve



Continuous prediction

PLARE

Contents

- Introduction
- ➢ **Method**
- Result
- Discussion
- Conclusion

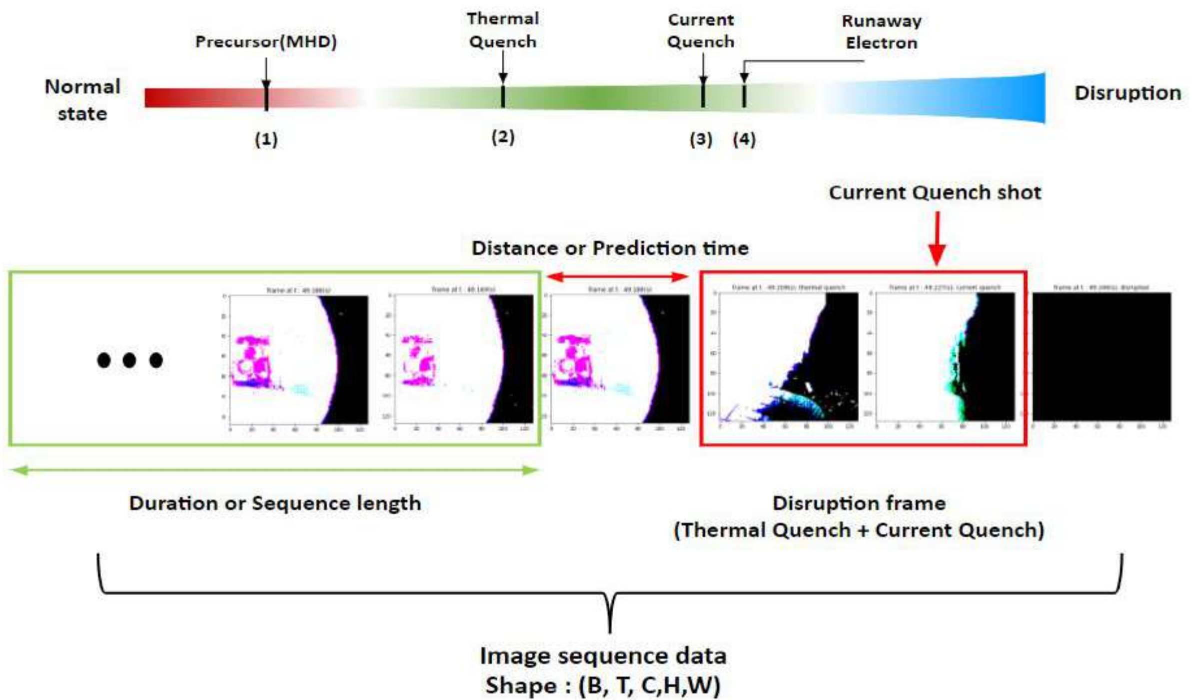
Method

- **Method**
  - Dataset setup
    - KSTAR IVIS Video data : Image sequence data / 210 fps and 480 X 640 resolution
    - KSTAR 0D data : Time-series data / 20kHz sampling rate
    - Multimodal data : Video data + 0D data
  - Model setup
    - Model for video data: SlowFast / R2Plus1D / ViViT
    - Model for 0D data: Self-attention 1D CNN – LSTM
    - Model for multimodal data (Video + 0D) : Tensor Fusion Network
  - Learning Algorithm
    - Objective / Loss function : Cross Entropy Loss, Focal Loss, LDAM Loss
    - Learning algorithm for **imbalance data distribution** : Re-Sampling, Re-Weighting, Deferred Re-Weighting
    - Learning algorithm for **multimodal data** : Gradient Blending, Deep CCA (not complete)

Method : Dataset setup

- **KSTAR IVIS data for Video Model (SlowFast, R(2+1)D, ViViT)**

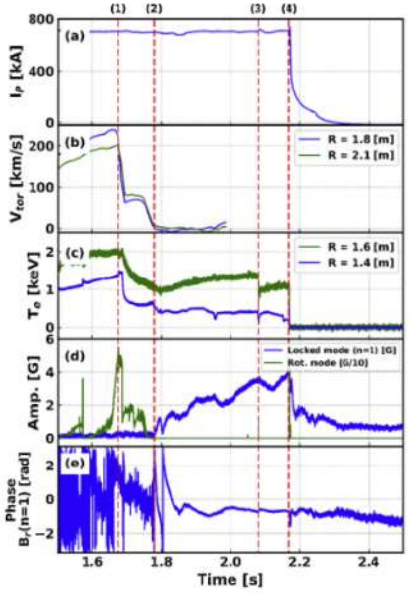
  - KSTAR In-vessel Visible Inspection System(IVIS) : Video data used for monitoring the plasma in a vessel
  - **82 video data** with **210 frame** per seconds collected from KTSAR IVIS were used for training and evaluation
  - We set the **last frame** as a **disruptive event** and considered the **last second frame** as a **current quench state**.
  - **21 frames** as sequence length, **resize as 128 x 128 due to GPU memory limit**

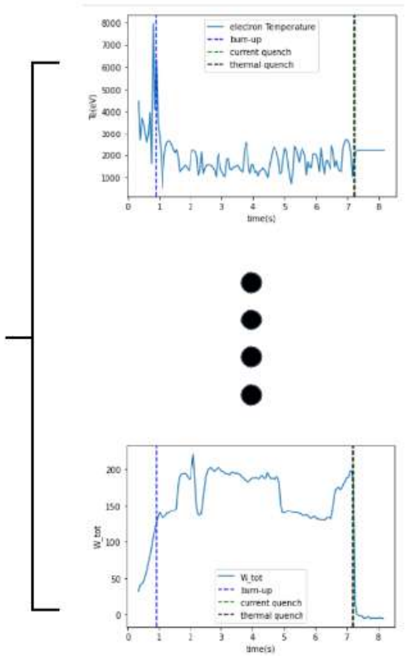


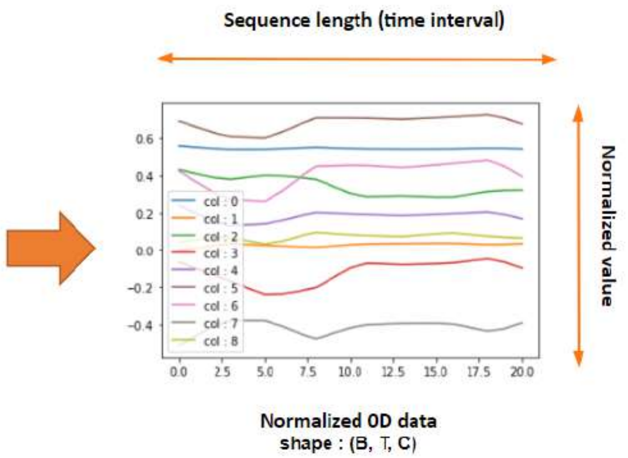
T : Duration or Sequence length
H : Height of the image
W : Width of the image
C : Channel of the image

Method : Dataset setup

- **KSTAR 0D data for 1D CNN-LSTM model**
  - **Plasma 0D data** : $I_p, \beta_n, \beta_p, \kappa, W_{tot}, n_e, li, q_{95}, \delta_{tri}$
  - **Different sampling rates** from experimental data: 20kHz, 50kHz
  - **Linear interpolation** with **constant time interval** (=19.04ms, 4 times of video fps)
  - Scaling : **Robust scaler** was used for ignoring anomality







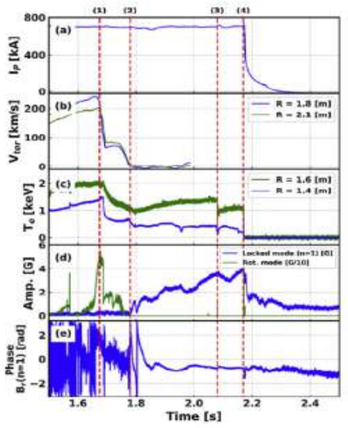
T : Duration or Sequence length
C : Number of variables

PLARE

Method : Dataset setup

▪ Muti-modal Dataset for disruption prediction

- KSTAR IVIS video data : **84 frames** as sequence length, **4 frames** as **frame interval**
- Plasma 0D data : same as input data of single-modal model (1D CNN –LSTM)
- Time-synchronization : Due to **delay issue** from **video capture process** induced by the **limit of IVIS Cam device**, We **re-matched video frames and 0D data** from **backward** with respect to time.



Jeongwon Lee et al., Fusion Eng. and Des., 2019

0D data
(batch size, sequence length, number of variables)

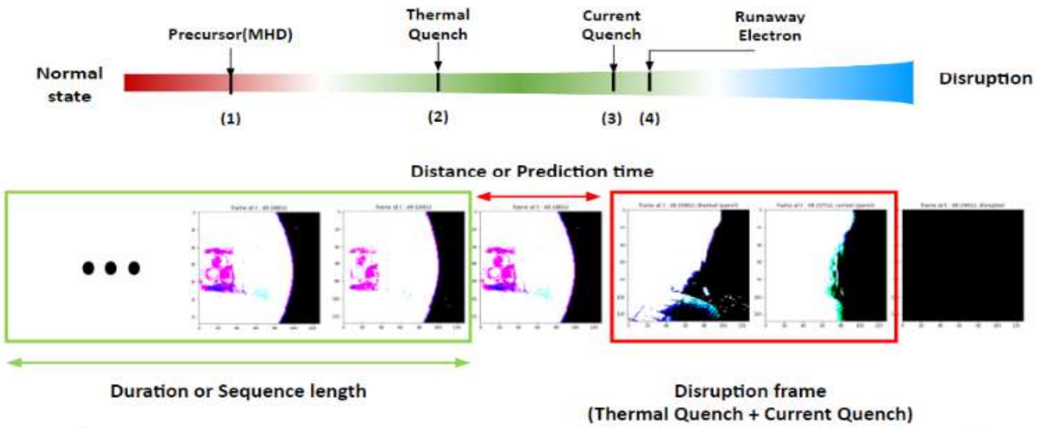
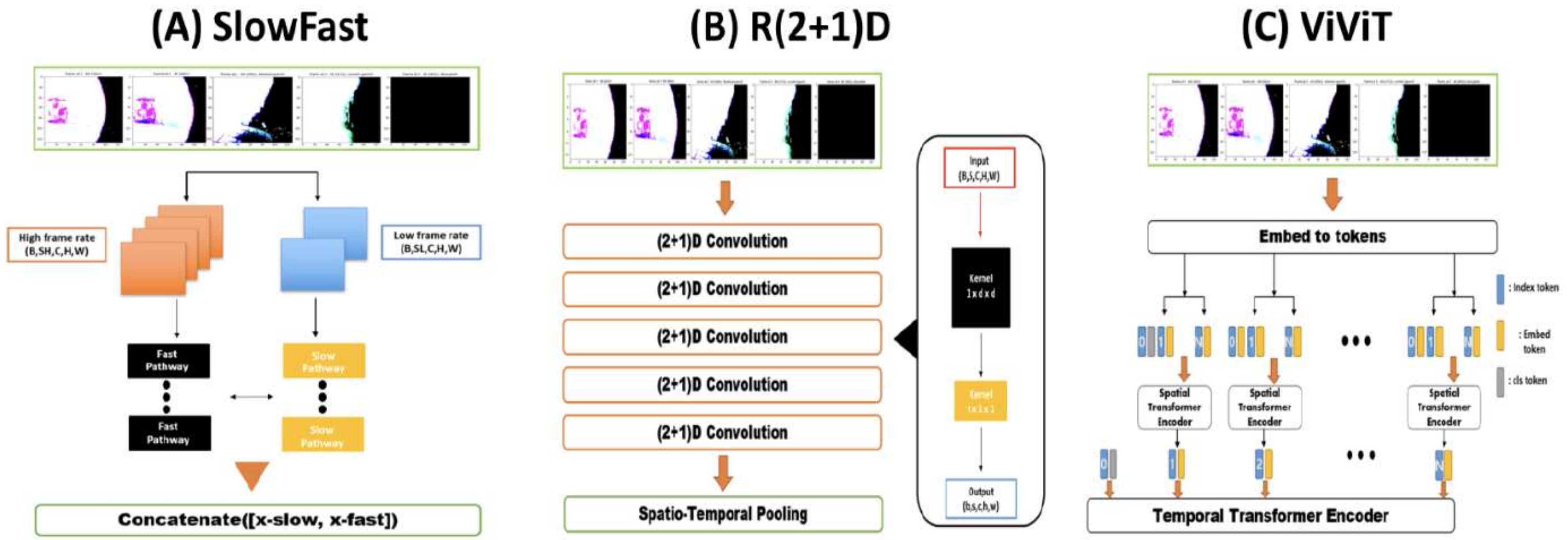


Image sequence data
(batch size, sequence length, channel, height, width)

PLARE

Method : Model setup

- **Development of plasma disruption prediction model using Deep Learning**
  - **Vision model for KSTAR IVIS video data**
    - Convolutional Neural Network (CNN) - based model : SlowFast, R(2+1)D
    - Transformer – based model : Video Vision Transformer (ViViT) → Efficient for small dataset with effective regularization

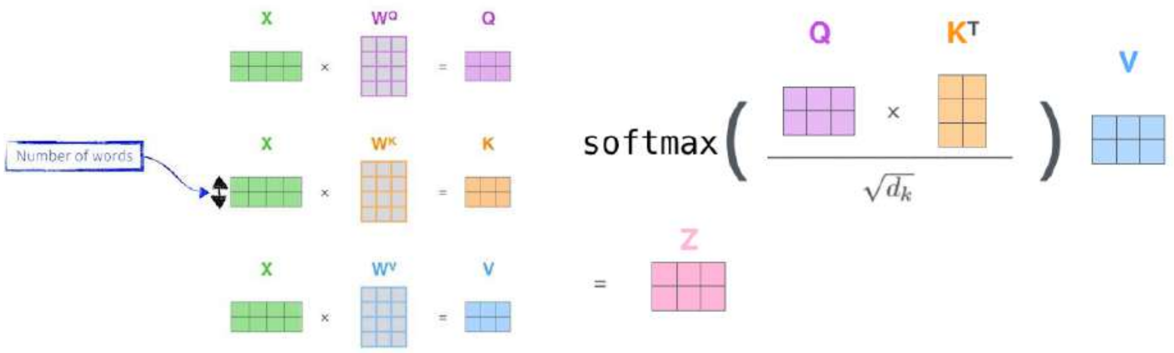


Different objective functions : Cross Entropy Loss, Focal Loss, LDAM Loss

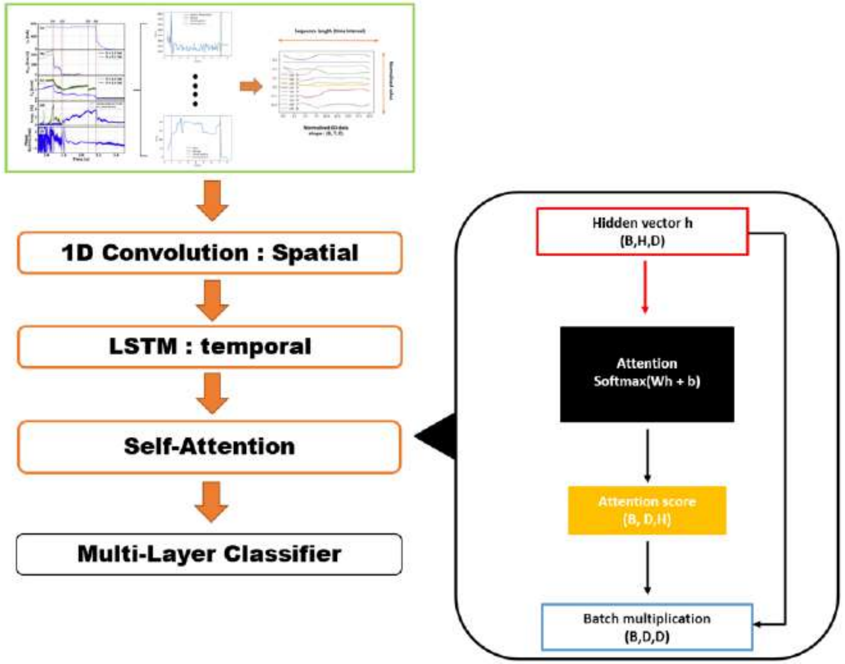
Different learning algorithms : Re-Sampling, Re-Weighting, Deferred Re-Weighting

Method : Model setup

- **Development of plasma disruption prediction model using Deep Learning**
  - **Self-attention 1D CNN – LSTM model for KSTAR 0D data**
    - 1D CNN – LSTM model : 1D Convolutional Neural Network (1D – CNN) + Long-short term memory (LSTM)
    - 1D CNN : To extract spatial components (=correlation between variables)
    - LSTM : To extract temporal components
    - Self-attention : Attention mechanism applied to LSTM output



1D CNN – LSTM

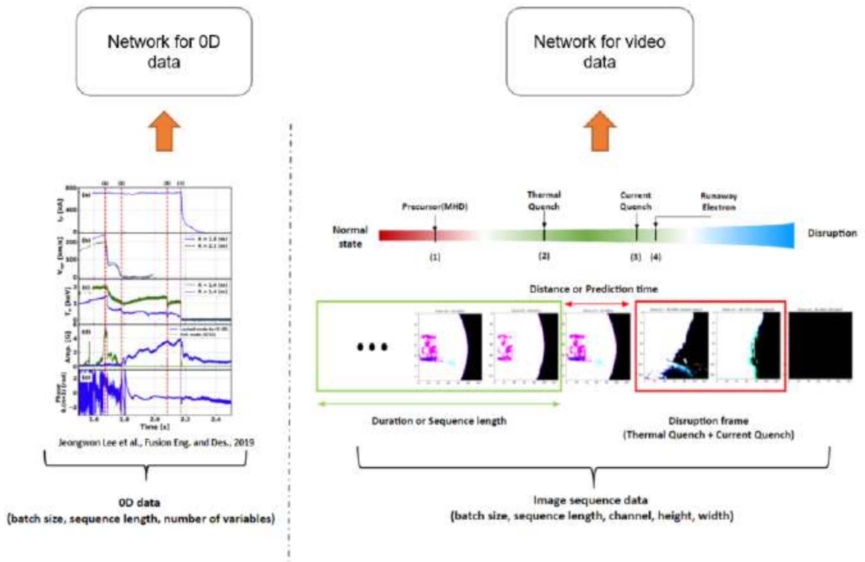


Different objective functions : Cross Entropy Loss, Focal Loss, LDAM Loss

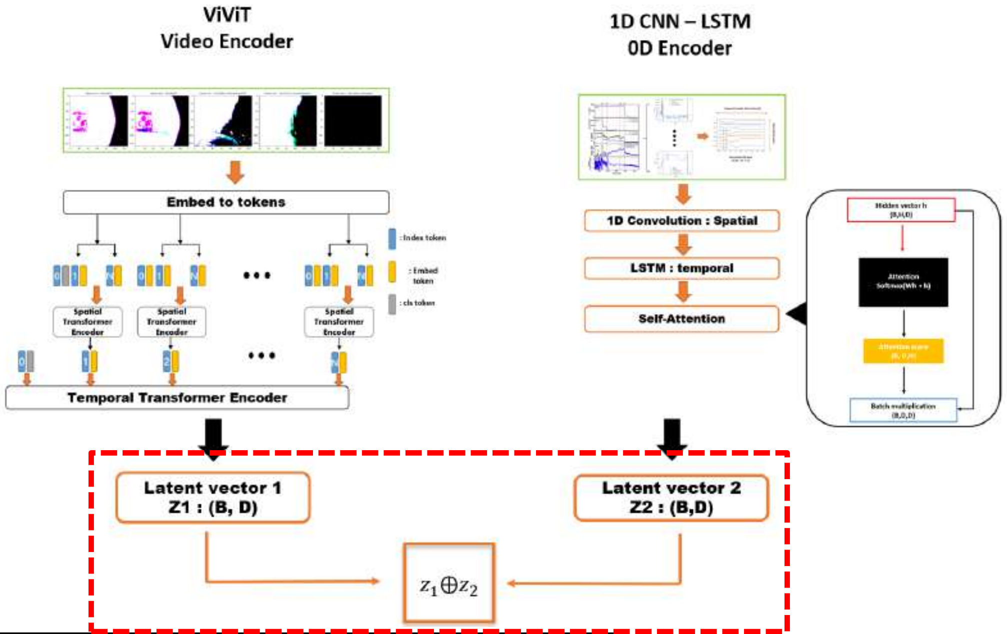
Different learning algorithms : Re-Sampling, Re-Weighting, Deferred Re-Weighting

Method : Model setup

- **Development of plasma disruption prediction model using Deep Learning**
  - **Tensor Fusion Network for Multimodal data (Video + 0D data)**
    - KSTAR Multimodal data : Video + 0D data



Tensor Fusion Network
Element-wise matrix multiplication with different modalities



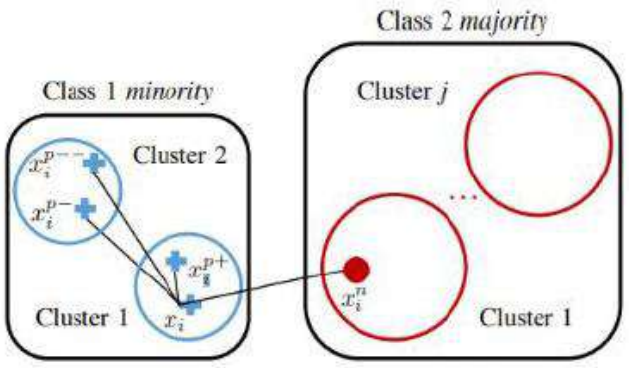
Different objective functions : Cross Entropy Loss, Focal Loss, LDAM Loss

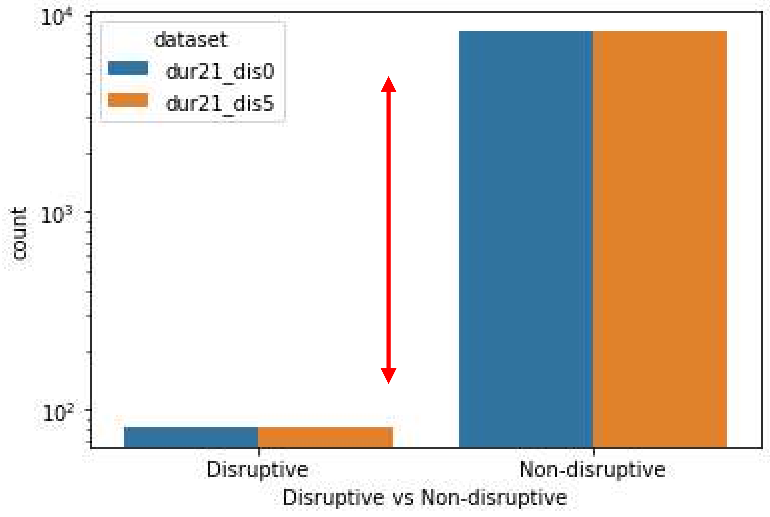
Different learning algorithms : Re-Sampling, Re-Weighting, Deferred Re-Weighting

Additional : Tensor Fusion Network Architecture, **Gradient Blending**, Deep Canonical Correlation Analysis (Deep CCA)

Method : Learning algorithms

- **Learning algorithms for imbalance data distribution**
  - Since the time scale of disruption is relatively short compared to the operation time, there exist severe data – imbalance (disruptive vs non-disruptive) problem.
  - Uniform class distribution : Cross Entropy Loss is enough → **Not effective for long-tailed distribution**





Method : Learning algorithms

- **Learning algorithms for imbalance data distribution**
  - Method for learning imbalanced datasets : **Boosting** vs **Alternative objective functions**
  - **Boosting :** Meta learning technique designed to improve classification performance
    - **Re-Sampling** : **Over-sampling** the **minority classes** or **Under-sampling** the **frequent classes**
    - **Re-Weighting** : **Assigning weights** for **different classes** or different samples to compensate the importance
    - **Deferred Re-Weighting** : 2-stage method (1-stage : training without re-weighting , 2-stage : training with re-weighting)
  - **Alternative objective functions**
    - **Focal Loss** : reshape the cross entropy with modulating factor $(1-p)^{\gamma}$ to compensate the importance for hard samples

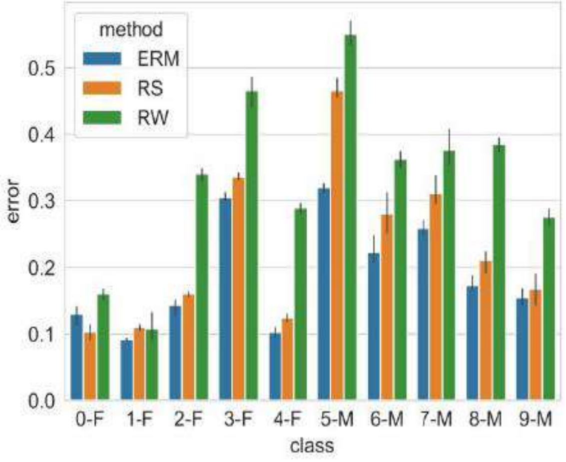
$$\textbf{Focal Loss} = -\sum (1-\mathrm{p_i})^{\gamma} \log \mathrm{p_i}$$

    - **LDAM Loss** : reshape the cross entropy for maximizing class margin according to label-distribution

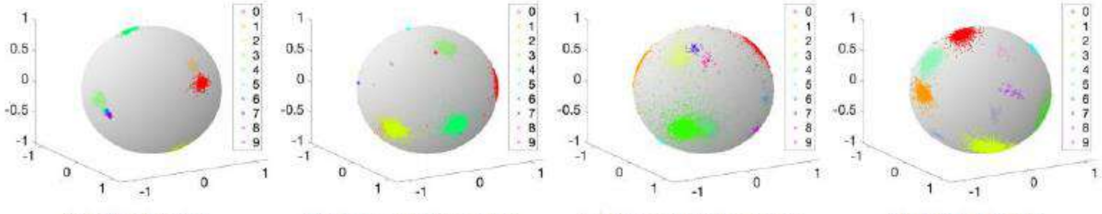
$$\textbf{LDAM Loss} = -\sum \log \frac{\exp(z_y - \Delta_y)}{\exp(z_y - \Delta_y) + \sum_{y \neq j} \exp(z_y)} \quad where\ \Delta_j = \frac{C}{n_j^{0.25}}$$

Method : Learning algorithms

- **Learning algorithms for imbalance data distribution**
  - Learning algorithms can generate significant different results with general imbalanced dataset (CIFAR, open-source dataset)
  - Learning algorithms also affect the data representation of the neural network : more significant for higher feature dimensions + severe imbalance distribution



Dataset	Imbalanced CIFAR-10				Imbalanced CIFAR-100			
Imbalance Type	long-tailed		step		long-tailed		step	
Imbalance Ratio	100	10	100	10	100	10	100	10
ERM	29.64	13.61	36.70	17.50	61.68	44.30	61.45	45.37
Focal [35]	29.62	13.34	36.09	16.36	61.59	44.22	61.43	46.54
LDAM	26.65	13.04	33.42	15.00	60.40	43.09	60.42	43.73
CB RS	29.45	13.21	38.14	15.41	66.56	44.94	66.23	46.92
CB RW [10]	27.63	13.46	38.06	16.20	66.01	42.88	78.69	47.52
CB Focal [10]	25.43	12.90	39.73	16.54	63.98	42.01	80.24	49.98
HG-DRS	27.16	14.03	29.93	14.85	-	-	-	-
LDAM-HG-DRS	24.42	12.72	24.53	12.82	-	-	-	-
M-DRW	24.94	13.57	27.67	13.17	59.49	43.78	58.91	44.72
LDAM-DRW	**22.97**	**11.84**	**23.08**	**12.19**	**57.96**	**41.29**	**54.64**	**40.54**



(a) ERM train (b) Re-sampling train (c) Re-weighting train (d) LDAM train

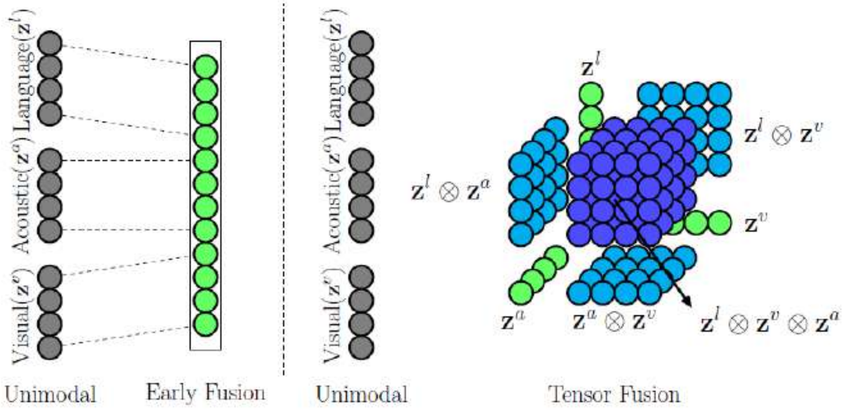
(e) ERM val (f) Re-sampling val (g) Re-weighting val (h) LDAM val

[Learning Imbalanced Dataset with Label-Distribution-Aware Margin Loss, Kaidi Cao et al, 2019]

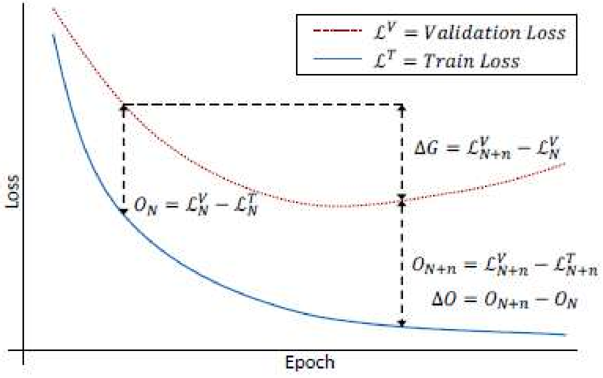
Method : Learning algorithms

▪ Learning algorithms for multi-modal data

- Muti-modal networks are often prone to overfitting due to increased capacity : Multi-modalities often cause negative effect
- Different modalities overfit and generalize at different rates : training multi-modal data can cause sub-optimal problems
- **Gradient-Blending** : Training multi-head model (training single-modal model + multi-modal model) with additional optimization process of overfitting-to-generalization-ratio for blending weights
- **Tensor Fusion Networks** : improved architecture to learn intra-modality (uni-modal interaction) and inter-modality (interaction between different modality) dynamics



Tensor Fusion Networks for Multimodal Sentiment Analysis, Amir Zadeh et al, 2017



$$OGR^2 = \left(\frac{\langle \nabla \mathcal{L}^T - \nabla \mathcal{L}^*, \hat{g} \rangle}{\langle \nabla \mathcal{L}^*, \hat{g} \rangle}\right)^2$$

$$w^* = \arg\min_w \mathbb{E}\left[\left(\frac{\langle \nabla \mathcal{L}^T - \nabla \mathcal{L}^*, \sum_k w_k v_k \rangle}{\langle \nabla \mathcal{L}^*, \sum_k w_k v_k \rangle}\right)^2\right]$$

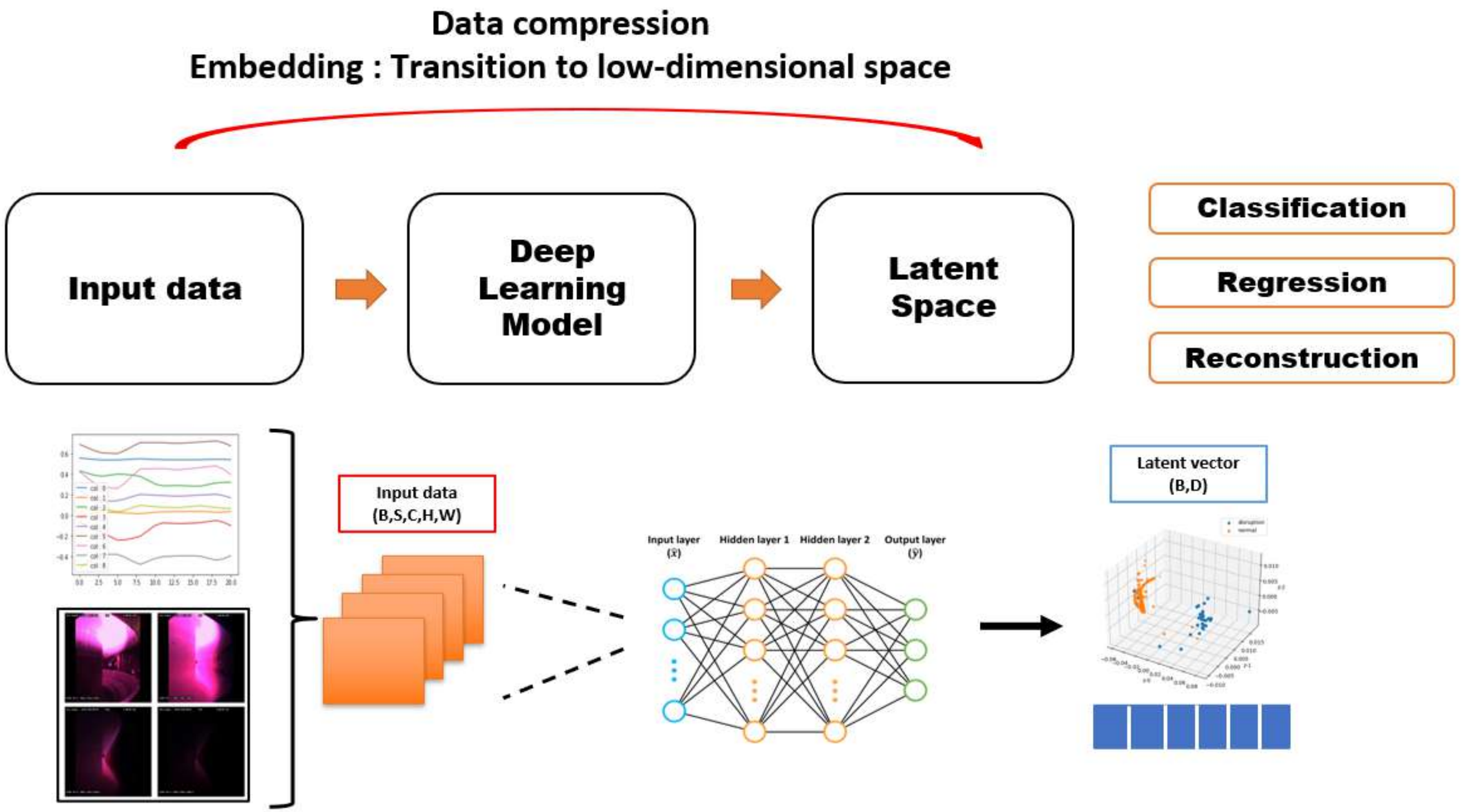
$$\mathcal{L}_{blend} = \sum_{i=1}^{k+1} w_i \mathcal{L}_i$$

What Makes Training Multi-modal Classification Networks Hard?, Weiyao Wang et al, 2020,

Method : Analysis

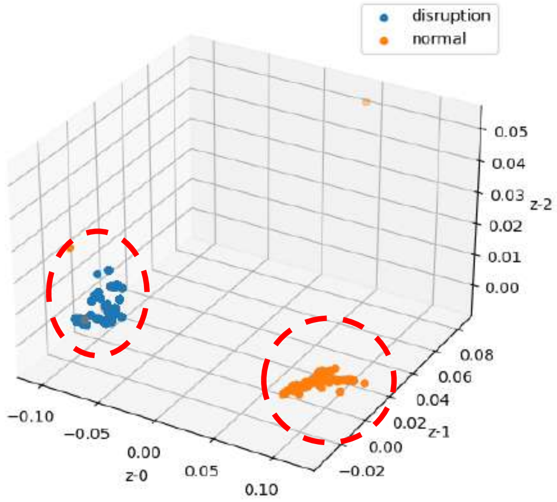
▪ Data Embedding and Latent vector

- Neural Network **transforms raw data** into a suitable internal representation **or hidden vector with dimension compression.**
- The output data computed from NN contains compressed feature information : We call this data as ‘**Latent vector**’.

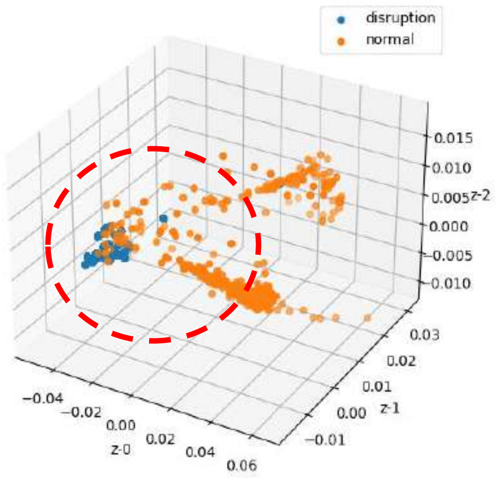


Method : Analysis

- **Visualization of latent vectors in 3-dimension space**
  - We use **Principal Components Analysis** (PCA) **to reduce the dimension** of the **latent vector** to 3-dimension for visualization
  - **The distance between latent vectors = similarity**
  - The **distance** between **disruptive** data and **non-disruptive** data **will increase** if the **model predict and classify** the disruptive and non-disruptive data **successfully**.



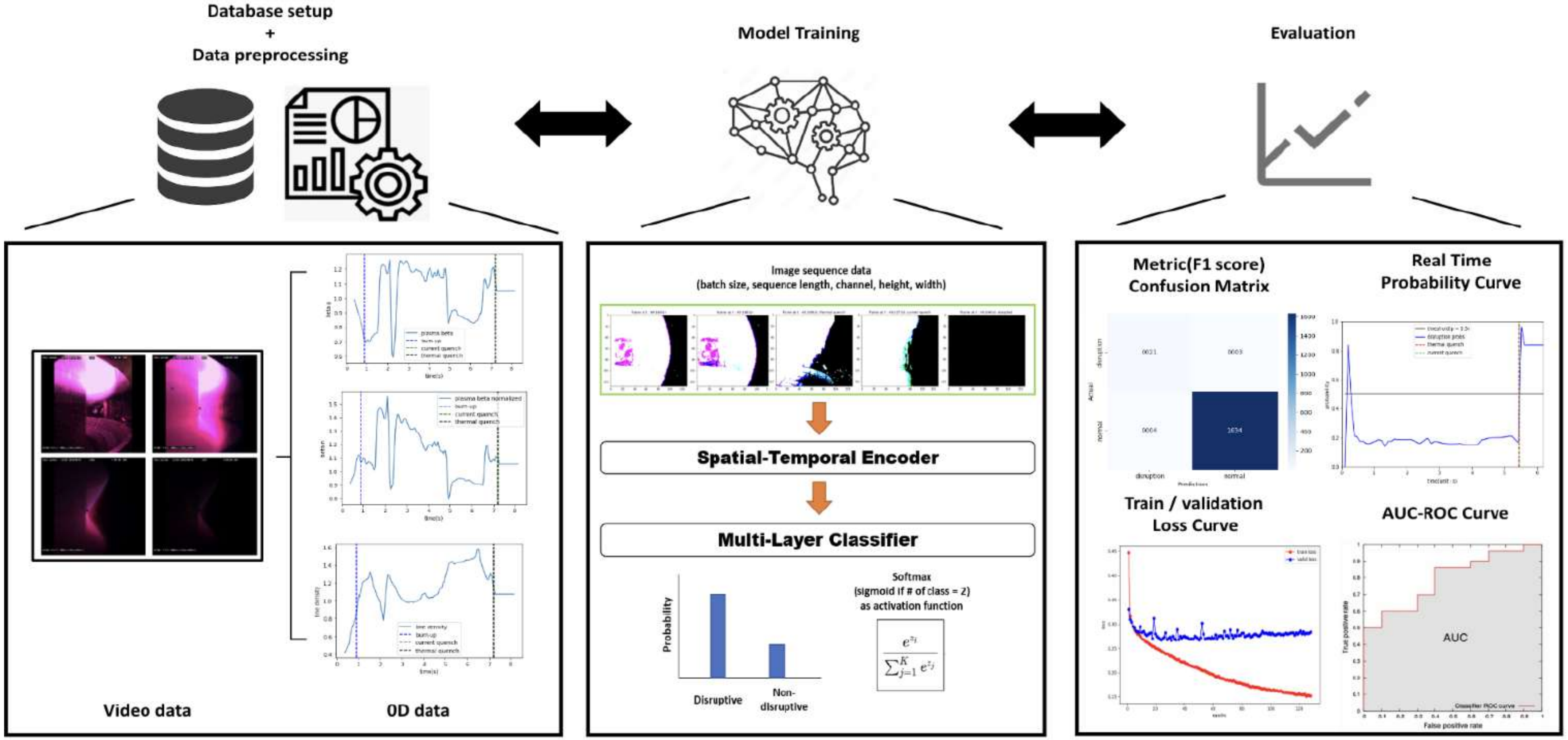
Prediction time : 9.52ms
F1 score : 0.99



Prediction time : 19.04ms
F1 score : 0.569

Method : ML pipeline for experiments

- **Process for training and evaluation**



PLARE

Method : ML pipeline for experiments

- **Metrics**
  - It is important to predict disruptive phase to alert prior to the disruption event **without false /missing alarms**.
  - So, We should monitor **both precision and recall** score for training and evaluation.
  - **Precision** : the ratio of true positives over the predicted positives
  - **Recall** : the ratio of true positives over the real disruption
  - **Macro-F1** : Mean of the F1 scores for each class → Main metric for this research
  - **Confusion Matrix** : Error matrix which reports the number of True-Positive(TP), False-Positive(FP), True-Negative(TN), False-Negative(FN)

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{tp}}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})}$$

Contents

- Introduction
- Method
- ➢ **Result**
- Discussion
- Conclusion

Result

- **Experimental setup**

  - Cross Entropy Loss : Baseline for comparison
  - Focal Loss : r = 2, Re-Weighting with inverse class frequency
  - LDAM Loss : Rescaling class-dependent margin loss with 0.5, Re-weighting with inverse class frequency
  - Training Epochs : 128 for all experiments
  - Learning algorithms for use
    - **Re-Sampling** : **Over-sampling** the **disruptive data**
    - **Re-Weighting** : Assigning **inverse class frequency** to weight loss functions
    - **Deferred Re-Weighting** : 4-stages Re-Weighting (update β and weight $w_j = \frac{1-\beta}{1-\beta^{n_j}}$ with respect to epochs)
  - Data Properties
    - Image size : 256
    - Crop size : 128
    - Sequence length : 21 frames for video data, 21 points for 0D data, 84 frames / 21 points for multi-modal data
    - Augmentation : Flip, Shift, Brightness, Contrast, Blur for video data

Result

- **Experiment list**
  - **Video data experiment**
    - Model comparison
    - Different prediction time
    - Different learning algorithm
    - Continuous prediction for test shot
  - **0D data experiment**
    - Different prediction time
    - Different learning algorithm
    - Continuous prediction for test shot
  - **Multi-modal data experiment**
    - Different prediction time

Result

- **Experiments for disruption prediction only with video data**
  - **Model comparison**
    - Video Vision Transformer is **effective** with **relatively small model size** and **high performance**.
    - We use ViViT for next experiments.

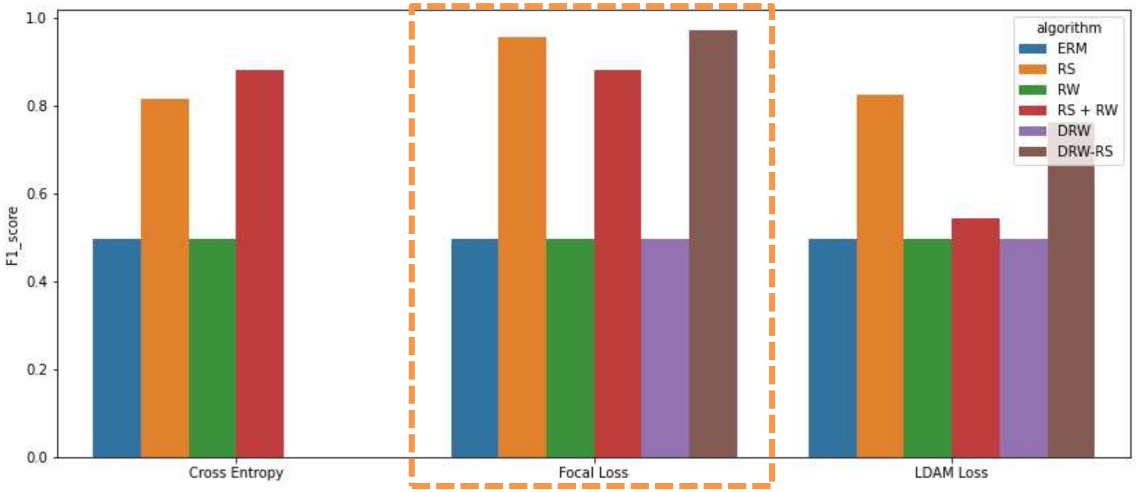
Model	Accuracy	F1 score	# of parameters	Precision (Disruption)	Recall (Disruption)
R(2+1)D	0.99	0.99	18,847,195	1.0	0.96
SlowFast	0.99	0.97	13,910,842	0.96	0.92
ViViT	0.99	0.98	1,513,026	0.96	0.96

  - **Different prediction time**
    - **Severe decrease of performance** is observed **after 19.04ms**.
    - Due to **abrupt decrease of precision**, it is **hard** to **identify disruption precursor** before **thermal quench** occurs.

Distance / Prediction time	Accuracy	F1 score	AOC (Disruption)	Precision (Disruption)	Recall (Disruption)
1(frame) / 4.762(ms)	0.99	0.99	0.98	1.0	0.96
2(frame) / 9.524(ms)	0.99	0.99	1.0	0.96	1.0
3(frame) / 14.286(ms)	0.99	0.93	0.94	0.84	0.88
4(frame) / 19.04(ms)	0.92	0.569	0.755	0.11	0.58
5(frame) / 23.810(ms)	0.92	0.59	0.84	0.12	0.75

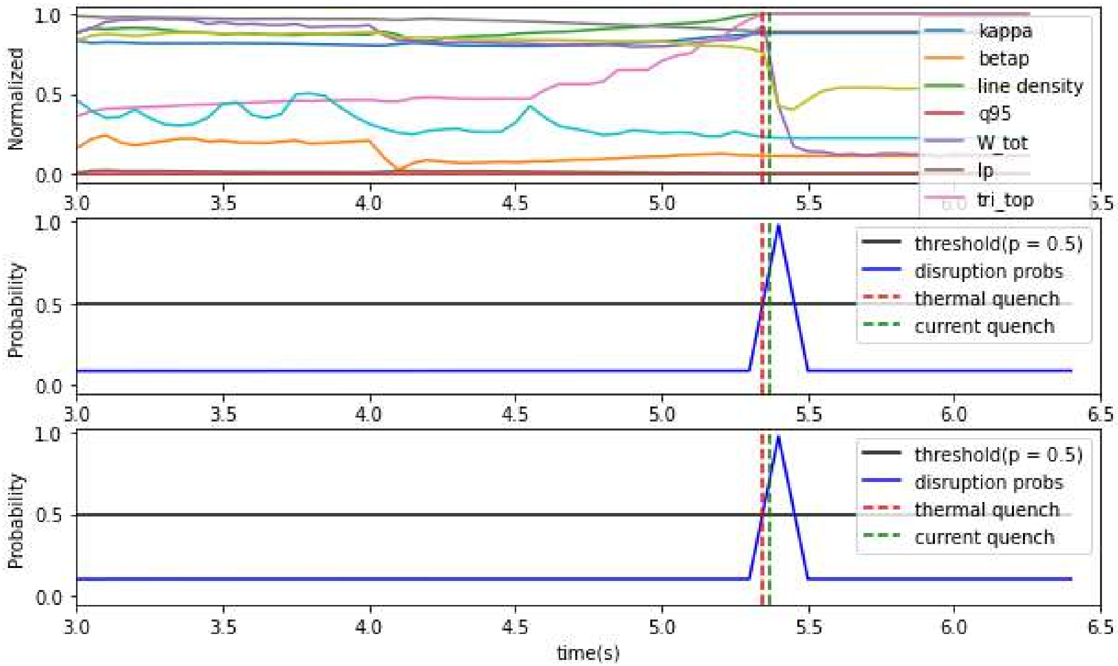
Result

- **Experiments for disruption prediction only with video data**
  - **Different learning algorithms**
    - Since there is limit to predict disruption prior to over 23.810ms with video data, we compare model performance prior to prediction time as 14.286ms with different learning algorithms.
    - **Re-Sampling** enhances **both precision and recall** for predicting disruption in **every case** while Re-Weighting should be used with Re-Sampling or used as Deferred Re-Weighting.
    - **Focal Loss** with **Deferred Re-Weighting** and **Re-Sampling** reached the best scores.



Result

- **Experiments for disruption prediction only with video data**
  - **Continuous prediction for test data - shot 21310**
    - We have shown that the ViViT – based model can predict disruption without false / missing alarms as a continuous disruption prediction for shot 21310 from test dataset.
    - The result below is for prediction time 4.762ms and 14.288ms



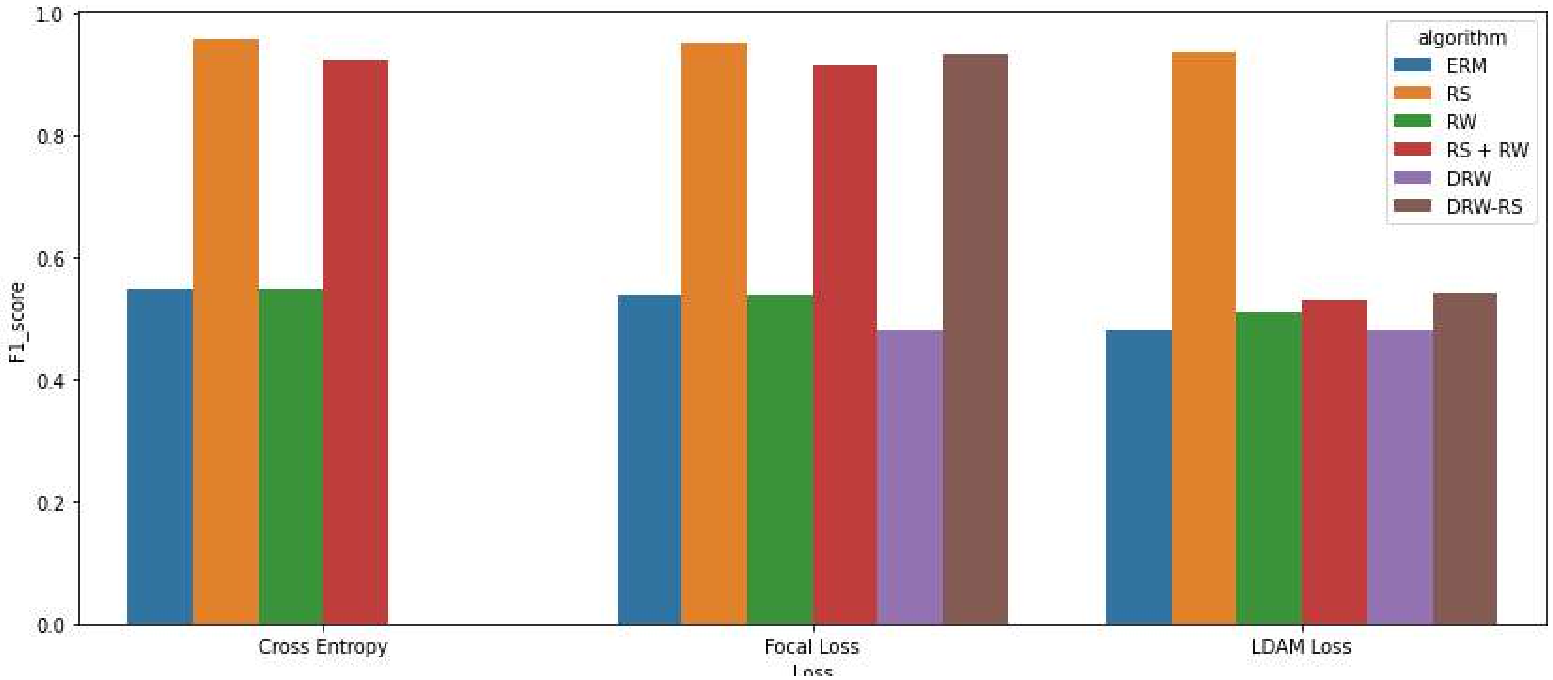
Result

- **Experiments for disruption prediction only with 0D data**
  - **Different prediction time**
    - **With 0D data**, it is obvious that the model can predict the disruption with **higher precision and recall compared to** those for only **video data → 0D data have more critical features for predicting disruptions**
    - It is shown that predicting disruptions **prior to over 57.14ms** with 0D data is **even possible** if ensemble method or cross-validation training for larger dataset is applied → **Disruption prediction** before **Thermal Quench**

Distance / Prediction time	Accuracy	F1 score	AOC (Disruption)	Precision (Disruption)	Recall (Disruption)
4(frame) / 19.04(ms)	0.99	0.967	0.941	1.0	0.88
8(frame) / 38.08(ms)	0.99	0.99	0.99	0.99	1.0
12(frame) / 57.14(ms)	0.99	0.957	0.969	0.89	0.94
16(frame) / 76.16(ms)	0.95	0.959	0.970	0.90	0.95
20(frame) / 95.2(ms)	0.97	0.877	0.852	0.83	0.71

Result

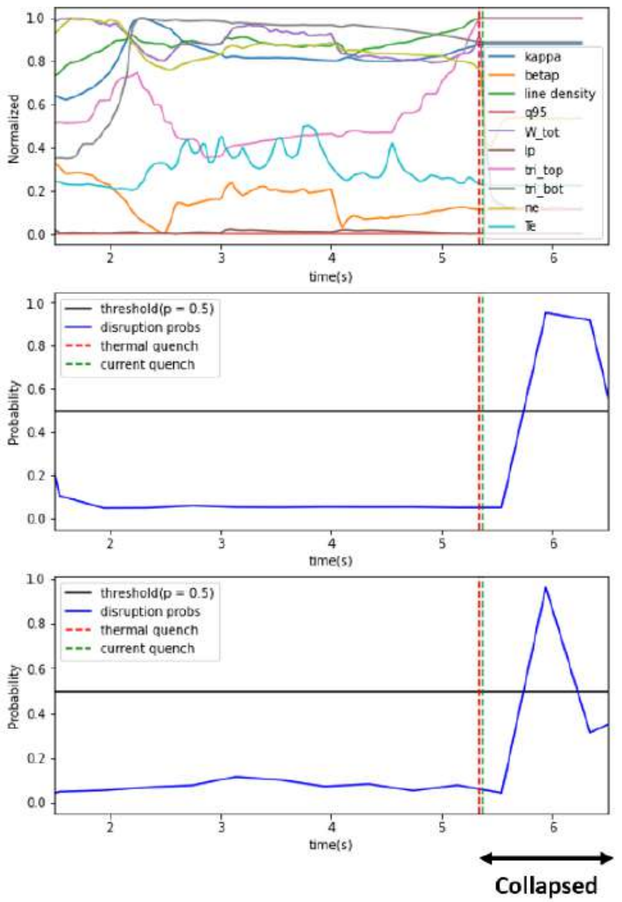
- **Experiments for disruption prediction only with 0D data**
  - **Different learning algorithms**
    - We also have compared model performance prior to 57.14ms with different learning algorithms since the realistic minimum prediction time for avoidance and mitigation is about 40ms.
    - **Focal Loss with Deferred Re-Weighting and Re-Sampling** or **Focal Loss with Re-Sampling** reached the high scores.



Result

- **Experiments for disruption prediction only with 0D data**
  - **Continuous prediction for test data - shot 21310**
    - We have also shown that the 1D CNN – LSTM can predict disruption without false / missing alarms as a continuous disruption prediction for shot 21310 from test dataset.
    - Since 1D CNN – LSTM can predict the disruption as early as 57.14ms with high performance and relatively small model size, the real-time disruption prediction with 0D data might be more efficient than video data only.
    - However, 0D data has relatively large time interval with compared to video data (about 4 times) which makes practical issue for real-time disruption prediction.
    - The result below is for prediction time 19.04ms, 57.14ms and 95.2ms.

Shot 21310, 0D data and disruption probability curve



Result

- **Experiments for disruption prediction with multi-modal data**
  - **Different prediction time**
    - Condition : Tensor Fusion Network (ViViT + 1D CNN – LSTM) + Gradient Blending
    - **Disruption prediction performance enhancement** : Using multimodal model for disruption prediction can help model precision and recall.

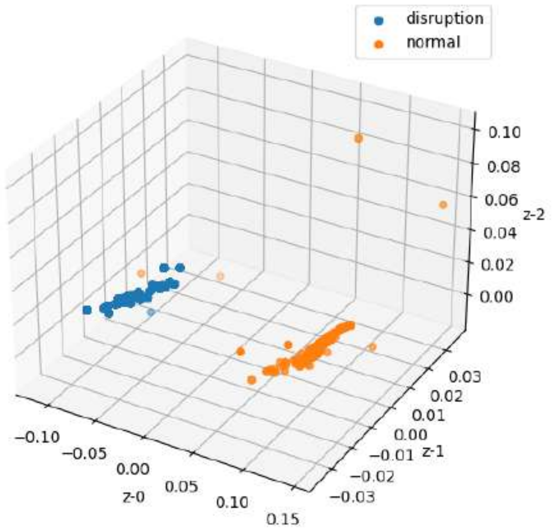
Distance / Prediction time **( ) : Video Data Only**	Accuracy	F1 score	AOC (Disruption)	Precision (Disruption)	Recall (Disruption)
4(frame) / 19.04(ms)	0.94(0.92)	0.785(0.569)	0.87(0.755)	0.49(0.11)	0.79(0.58)
8(frame) / 38.08(ms)	0.97	0.851	0.865	0.69	0.75

Contents

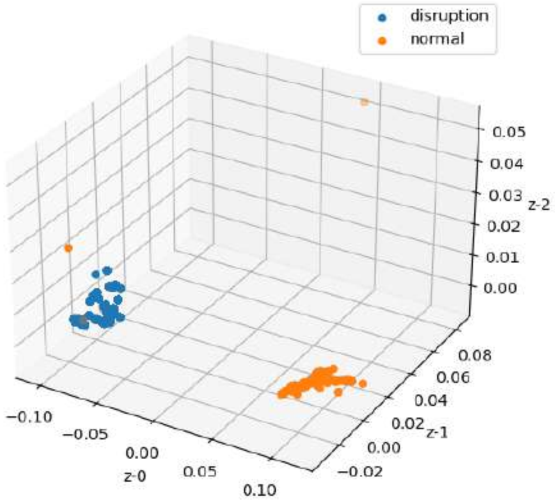
- Introduction
- Method
- Result
- ➢ **Discussion**
- Conclusion

Discussion

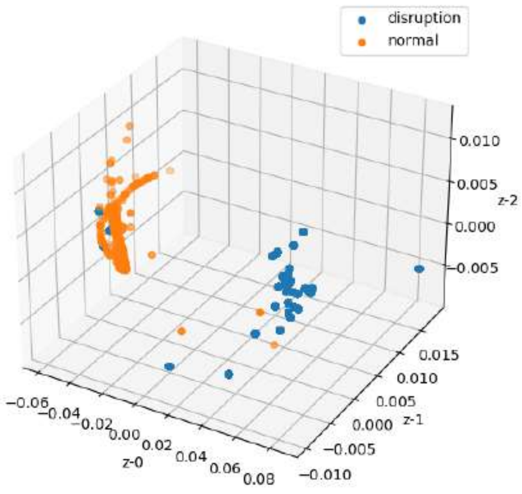
- **Experiments for disruption prediction only with video data**
  - **Visualization for 3D latent space**
    - Using Principal Component Analysis(PCA), We have visualized hidden vectors from the last layer of the Vision Video Transformers
    - As the accuracy higher, the model separated the disruptive and non-disruptive data more clearly.
    - With increasing prediction time, the difference between disruptive and non-disruptive data has decreased.
    - **Disruptive precursors** or **features might not be detected** when the prediction time is **over 15ms**.
    - Condition : **ViViT** with **DRW-RS and Focal Loss**, prediction time as 4.76ms, 9.52ms, 14.28ms, 19.04ms



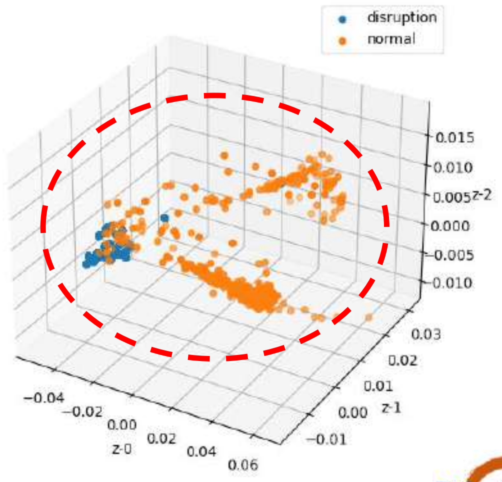
Prediction time : 4.76ms



Prediction time : 9.52ms



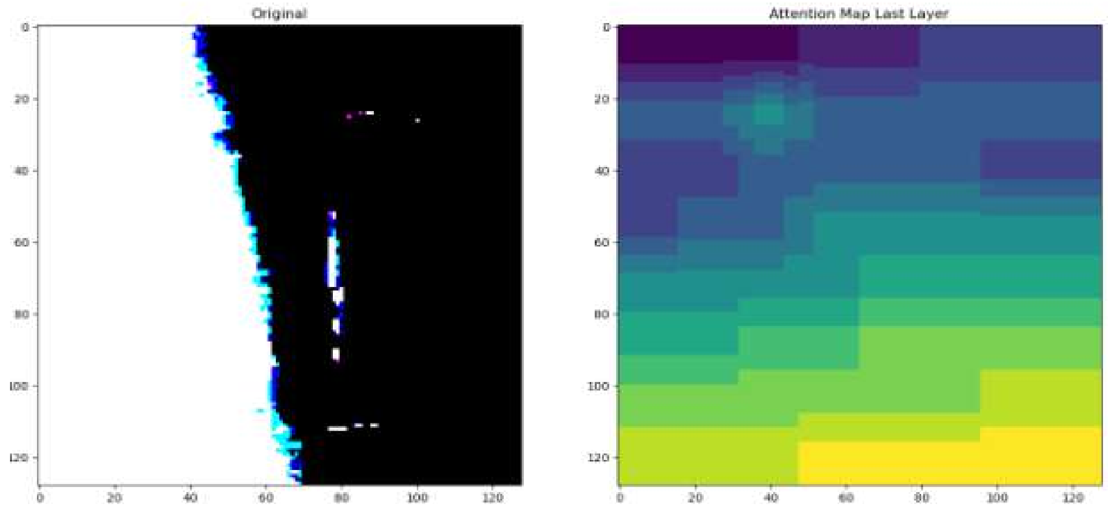
Prediction time : 14.28ms



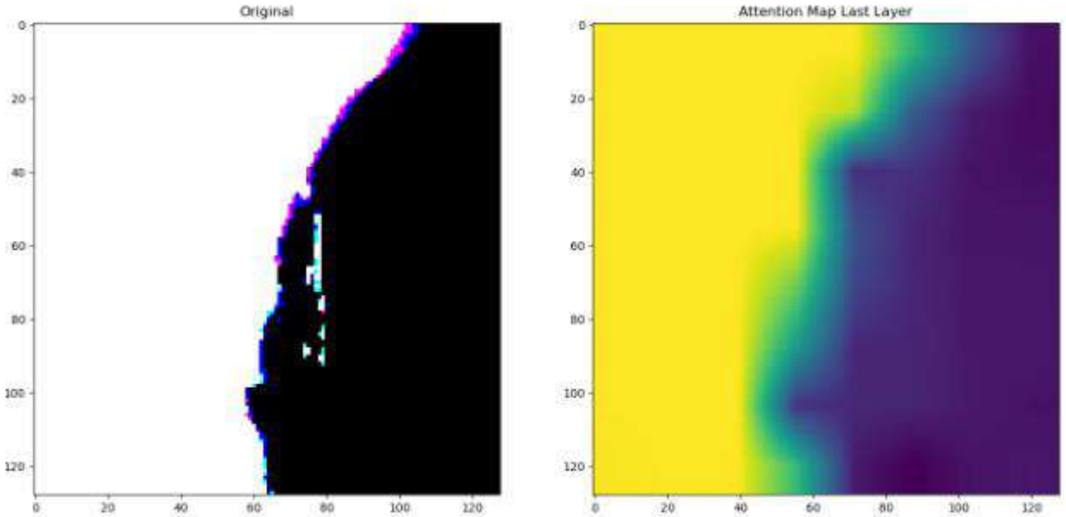
Prediction time : 19.04ms

Discussion

- **Experiments for disruption prediction only with video data**
  - **Visualization for attention mapping**
    - We have visualized the attention mapping which implies the 2D image of model weights used for feed-forward process.
    - The model weights focus on the specific part of the image if the prediction time is short → Some spatial criteria of the plasma might be learned from training model by video data.
    - The model weights just focus on the shape of the plasma more than the other side of the image as prediction time increases



ViViT trained by RS-DRW, prediction time : 4.76ms, F1 score : 0.966
Non-disruptive



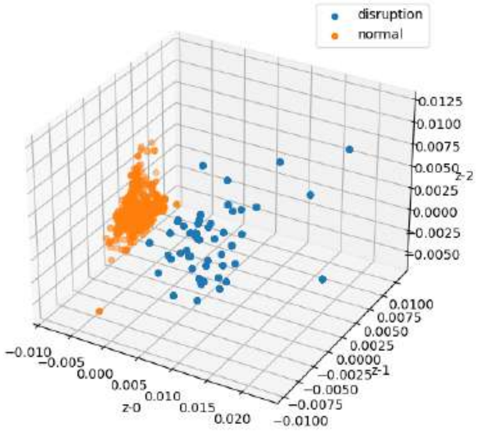
ViViT trained by RS-DRW, prediction time : 14.28ms, F1 score : 0.899
Disruptive

Discussion

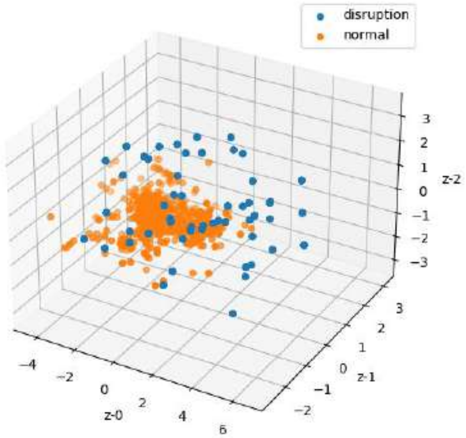
▪ **Experiments for disruption prediction only with 0D data**

- **Visualization for 3D latent space**

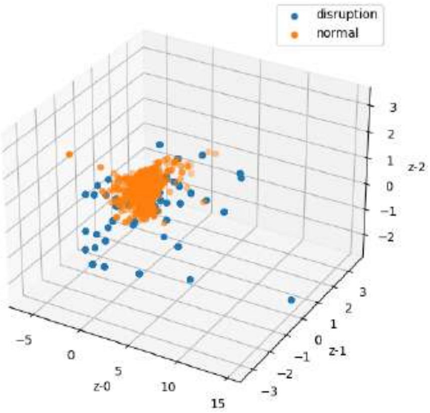
  - We have also visualized hidden vectors from the last layer of the 1D CNN – LSTM using PCA.
  - As the accuracy higher, the model separated the disruptive and non-disruptive data more clearly.
  - With increasing prediction time, the difference between disruptive and non-disruptive data has decreased.
  - Conditions : 1D CNN - LSTM with DRW-RS and Focal Loss, prediction time as 19.02ms, 38.04ms, 57.14ms, 76.16ms, 95.8ms



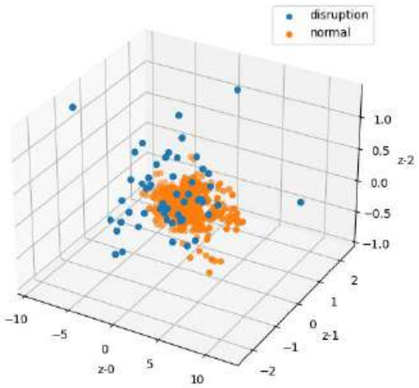
Prediction time : 19.02ms



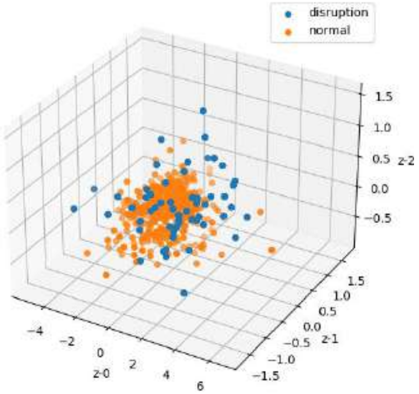
Prediction time : 38.04ms



Prediction time : 57.14ms



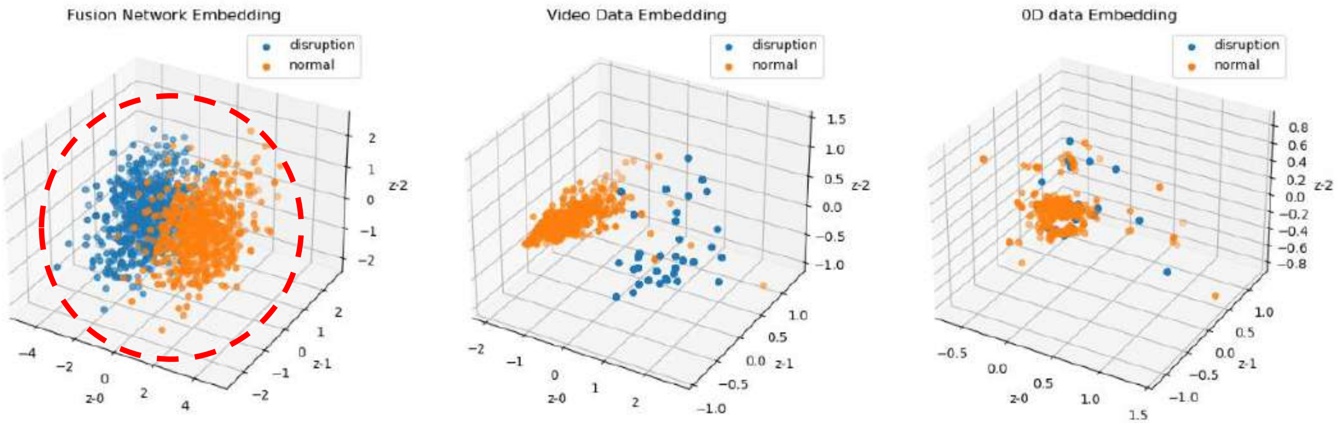
Prediction time : 76.16ms



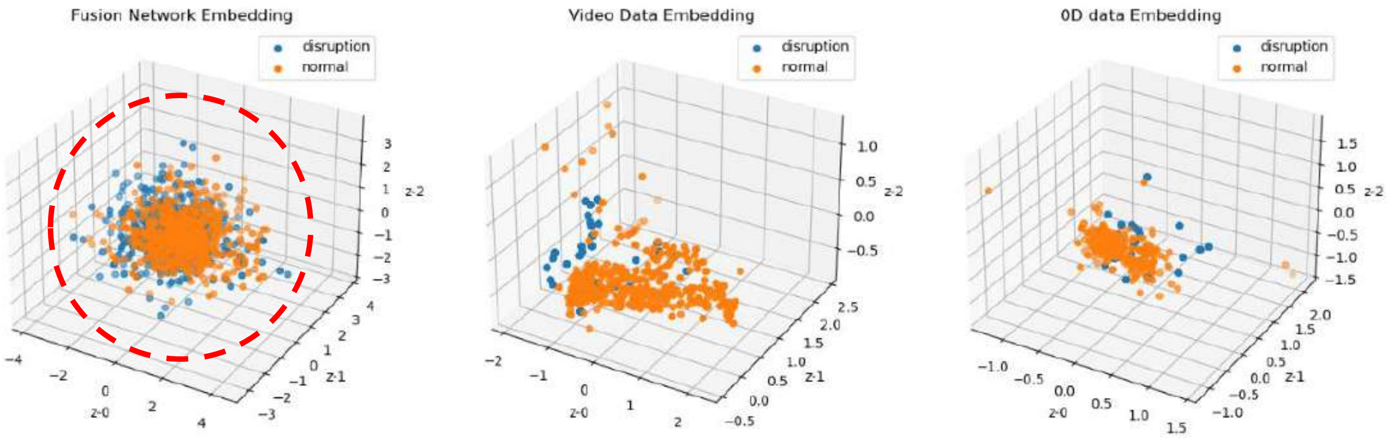
Prediction time : 95.8ms

Discussion

- **Experiments for disruption prediction with multi-modal data**
  - **Visualization for 3D latent space**
    - Gradient Blending



    - Without Gradient Blending



Contents

- Introduction
- Method
- Result
- Discussion

➢ **Conclusion**

Conclusion

- In this research, we propose the disruption prediction with video data in KSTAR using Deep Learning.
- We have developed video-based model and shown that predicting disruption from video data is limitedly possible as early as up to 15ms.
- Severe increase of false positive alarms occurs from video-based model when prediction time is over 20ms due to the lack of physical information.
- Meanwhile, we have found that the 0D data is important to predict the disruption from the previous experiments based on 0D data which implies that it is possible to predict the disruption as early as up to 95ms only with 0D data.
- Since video and 0D data contains different physical information, we propose multimodal learning using video and 0D data which learns features of disruptions from video and 0D data synchronously.
- We have improved the prediction performance with Tensor Fusion Network and Gradient Blending algorithm.
- In the future, we will proceed additional experiments for large dataset obtained from different devices and proceed the model compression for real-time disruption prediction from the real experiments.

Reference

[1] Zakharov, Leonid E., et al. “Understanding Disruptions in Tokamaks.” Physics of Plasmas, vol. 19, no. 5 (2012)

[2] Kates-Harbeck, Julian, Alexey Svyatkovskiy, and William Tang. “Predicting Disruptive Instabilities in Controlled Fusion Plasmas through Deep Learning.” Nature 568, no. 7753 (2019): 526–31.

[3] Sias, G., B. Cannas, S. Carcangiu, A. Fanni, A. Murari, A. Pau, and Jet Contributors. “Disruption Prediction Approaches Using Machine Learning Tools in Tokamaks.” 2019 PhotonIcs & Electromagnetics Research Symposium - Spring (PIERS-Spring) (2019)

[4] Chandrasekaran, Jayakumar, and Sangeetha Jayaraman. “Data-Driven Technique for Disruption Prediction in Golem Tokamak Using Stacked Ensembles with Active Learning.” Review of Scientific Instruments 93, no. 3 (2022)

[5] Diogo R. Ferreira, Pedro J. Carvalho, Carlo Sozzi, Peter J. Lomas & JET Contributors “Deep Learning for the Analysis of Disruption Precursors Based on Plasma Tomography.” Fusion Science and Technology, 76:8, 901-911 (2020)

[6] Kwon G, Wi H, Hong J, "Tokamak visible image sequence recognition using nonlocal spatio-temporal CNN for attention needed area localization". Fusion Engineering and Design (2021)

[7] F C Schuller, "Disruption in tokamak", Plasma Phys. Control. Fusion 37 A135 (1995)

[8] Christoph Feichtenhofer, Haoqi Fan, Jitendra Malik, Kaiming He, "SlowFast Network for Video Recognition", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), (2019)

[9] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, Manohar Paluri, 'A Closer Look at Spatiotemporal Convolutions for Action Recognition', Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), (2018)

[10] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, Cordelia Schmid, 'ViViT, A Video Vision Transformer', Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), (2021)

[11] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar, 'Focal Loss for Dense Object Detection', 'Proceedings of the IEEE International Conference on Computer Vision (ICCV), (2017)

[12] Kang, Haeyong, Thang Vu, and Chang D. Yoo. “Learning Imbalanced Datasets with Maximum Margin Loss.” 2021 IEEE International Conference on Image Processing (ICIP) (2021)

Thank You