Deep Learning-Based Adaptive Remedial Action Scheme with Security Margin for Renewable-Dominated Power Grids

Yinfeng Zhao [1], Shutang You [1], Mirka Mandich [1], Lin Zhu [1,*], Chengwen Zhang [1], Hongyu Li [1], Yu Su [1], Chujie Zeng [1], Yi Zhao [1], Yilu Liu [1,2], Huaiguang Jiang [3], Haoyu Yuan [3], Yingchen Zhang [3] and Jin Tan [3]

[1] Department of Electrical Engineering & Computer Science, Tickle College of Engineering, The University of Tennessee at Knoxville, Knoxville, TN 37996, USA; yzhao53@utk.edu (Y.Z.); syou3@utk.edu (S.Y.); mmandich@vols.utk.edu (M.M.); czhang70@vols.utk.edu (C.Z.); hli90@utk.edu (H.L.); ysu10@vols.utk.edu (Y.S.); czeng8@vols.utk.edu (C.Z.); yzhao77@utk.edu (Y.Z.); liu@utk.edu (Y.L.)
[2] Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA
[3] National Renewable Energy Laboratory, Golden, CO 80401, USA; Huaiguang.Jiang@nrel.gov (H.J.); Haoyu.Yuan@nrel.gov (H.Y.); Yingchen.Zhang@nrel.gov (Y.Z.); Jin.Tan@nrel.gov (J.T.)
* Correspondence: lzhu12@utk.edu

Abstract: The Remedial Action Scheme (RAS) is designed to take corrective actions after detecting predetermined conditions to maintain system transient stability in large interconnected power grids. However, since RAS is usually designed based on a few selected typical operating conditions, it is not optimal in operating conditions that are not considered in the offline design, especially under frequently and dramatically varying operating conditions due to the increasing integration of intermittent renewables. The deep learning-based RAS is proposed to enhance the adaptivity of RAS to varying operating conditions. During the training, a customized loss function is developed to penalize the negative loss and suggest corrective actions with a security margin to avoid triggering under-frequency and over-frequency relays. Simulation results of the reduced United States Western Interconnection system model demonstrate that the proposed deep learning–based RAS can provide optimal corrective actions for unseen operating conditions while maintaining a sufficient security margin.

Keywords: adaptive capability; deep learning; customized loss function; Remedial Action Scheme (RAS); security margin

1. Introduction

Transient stability is a crucial aspect of power system stability, which refers to the ability to keep all generators synchronized after a large disturbance. If a power system is not transient stable, necessary corrective actions, e.g., load shedding and generation trip, will take place by Remedial Action Scheme (RAS) or Special Protection System (SPS) [1–4]. Usually, these corrective actions are designed offline based on the time-domain simulation of a few selected severe contingencies under some representative operating conditions (e.g., spring light, summer peak, and winter peak).

Recently, due to the effort to reduce greenhouse gas emissions, conventional synchronous generators are being replaced by inverter-based renewables (IBRs), resulting in lower system inertia, unusual power flow paths, and larger angular differences spreading across the system [5]. Since IBRs have different dynamic behaviors than conventional synchronous generators, power grids with high renewable penetration exhibit more complex dynamics. More importantly, the intermittence features introduce more frequent and dramatic variations in operating conditions. Additionally, the increasing integration of dispatchable load [6] and distributed battery energy storage systems (e.g., electric vehicles) [7] can impact system dynamics significantly. These changes make the traditional RAS or SPS design method inadequate to guarantee its effectiveness under all possible

operating conditions. The RAS designed based on the traditional method may be too conservative or even useless in some extreme cases. Moreover, the traditional method based on time-domain simulations requires a tremendous amount of time to cover all possible operating conditions.

Artificial Intelligence (AI) technologies have been widely used in many areas, such as image recognition, speech recognition, self-driving cars, and fraud detection. Many researchers also propose AI-based methods for wind speed prediction [8], situational awareness [9,10], emergency control [11,12], and oscillation damping control [13]. In this paper, AI technologies provide a promising solution for adaptive RAS by mapping the operating conditions to the optimal corrective actions.

- The deep learning–based adaptive remedial action scheme (RAS) is proposed in this paper. This new method suggests optimal load shedding and generation trip amount to avoid excessive load shedding or generation trip compared to traditional RAS.
- Considering a safety margin, a customized loss function is proposed to make the prediction error in the desired direction as much as possible to guarantee the frequency is inside the secure zone.
- This new method can be generalized for extreme events that may not be covered by traditional RAS due to complex operating conditions with higher renewable penetration levels.

The deep learning–based adaptive RAS is developed using the reduced 240-bus Western Electricity Coordinating Council (WECC) system model [14] and WECC-1 RAS as case study [15].

The remainder of this paper is organized as follows. Section 2 introduces the study system and WECC-1 RAS, followed by the training and testing dataset generation in Section 3. The deep learning model and the customized loss function are proposed in Section 4. In Section 5, the model performance is validated, and the adaptive RAS is compared with traditional RAS. Finally, Section 6 concludes this paper. The flow chart of the whole process in the paper is shown in Figure 1.
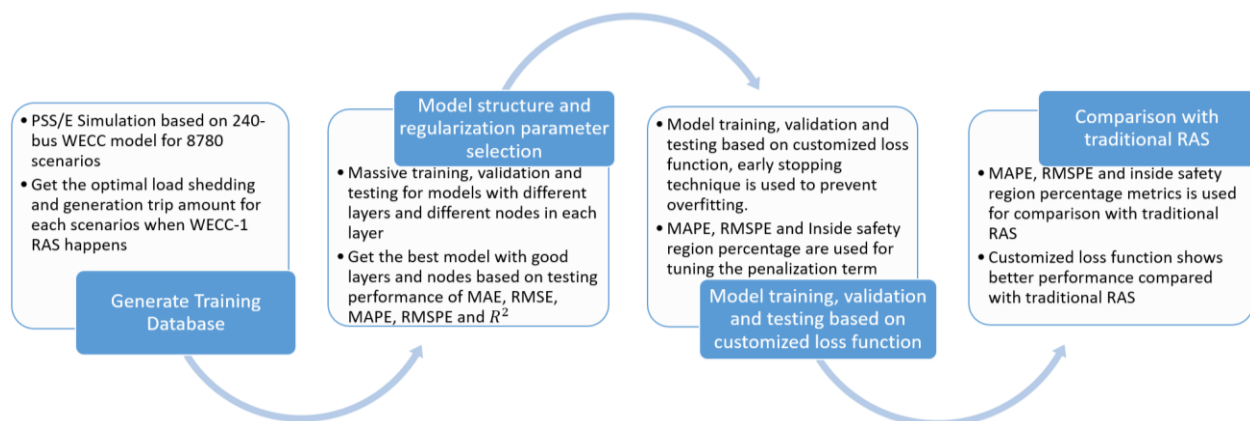

Figure 1. Flow chart of process.

2. Study System and WECC-1 RAS

2.1. Study System

This study is based on the reduced WECC 240-bus system model developed by the National Renewable Energy Laboratory (NREL), which has the generation of different fuel types, e.g., coal, gas, bio, nuclear, hydro, wind, and solar [16]. The wind and solar generation penetration level varies between 0.2% and 49.2% across 8786 hourly dispatches for an entire year of 366 days. Figure 2 shows the renewable penetration level across one year. In the figure, it shows that the renewable penetration is higher during the daytime than during the nighttime due to the availability of renewable energy.
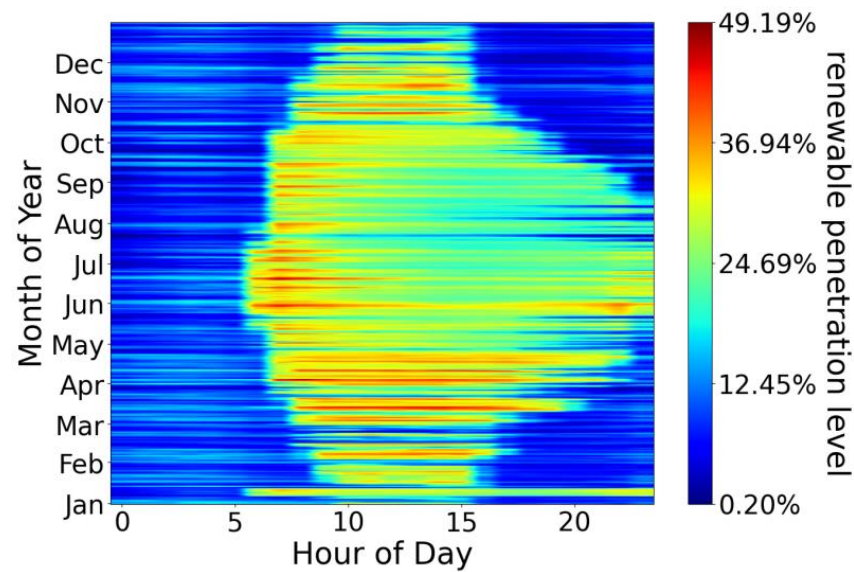
Figure 2. Renewable penetration across one year.

2.2. WECC-1 RAS

WECC-1 RAS is used for this study. WECC-1 RAS monitors the transmission system within California, Oregon, Arizona, Nevada, etc. When the predefined criterion is met, e.g., the trip of multiple tie lines, WECC-1 RAS will lead to a controlled separation of the WECC system into two islands to maintain system stability. In this study, as shown in Figure 3, once two tie lines between California and Oregon are tripped, the tie lines connecting the south area (California, Arizona, and New Mexico) and the north area (Oregon, Nevada, Utah, and Colorado) are tripped in WECC-1 RAS and followed by corrective actions to maintain the system stability. In reality, load shedding and generation trip at predetermined locations in two islands respectively are used to maintain the two islands' frequency between 59.5 hertz (Hz) and 60.5 Hz, according to the WECC grid code [17]. However, the optimal locations could be varying under different operating conditions, which makes it quite difficult to obtain the ground truth by considering both optimal generation trip/load shedding amount and optimal locations. For simplicity, proportional load decrease in one island and proportional load increase in the other island are used to generate the training database in this study. In practice, the load decrease and increase amounts can be used to further determine the load shedding and generation trip actions.
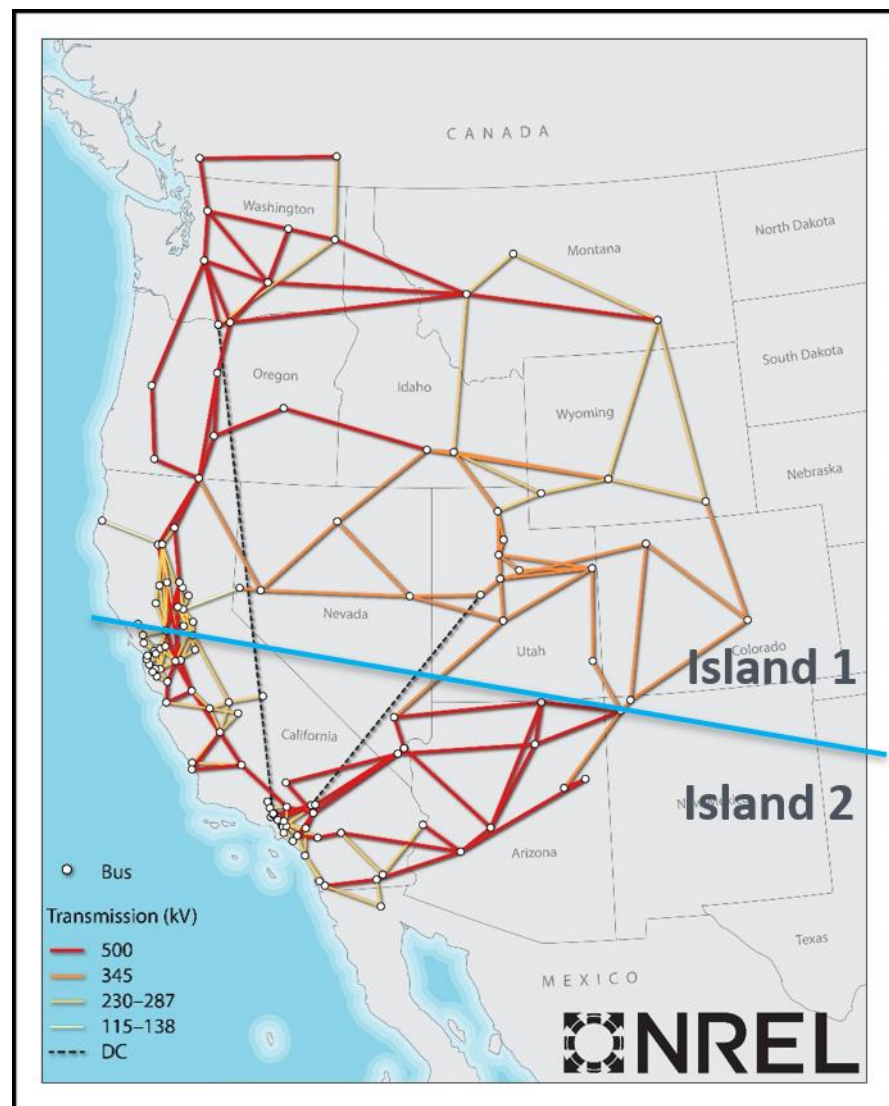
Figure 3. One-line diagram of 240-bus WECC test system model.

3. Training and Testing Dataset Generation

The hourly dispatch data of one year developed by NREL are used to simulate different operating conditions. Siemens PTI's PSS/E software is used to simulate the WECC-1 RAS under different operating conditions. At 1 s, the tie lines shown in Figure 3 are tripped, and the entire WECC system is separated into two islands. One hundred milliseconds (ms) after this islanding, load increase and load decrease in two islands are performed respectively to maintain the load-generation balance. The value 100 ms is the RAS action delay in communication and breakers. The minimum load increase and decrease in the two islands to maintain the two systems' frequency below 60.5 Hz and above 59.5 Hz, which are typical over-frequency and under-frequency monitoring thresholds, are considered optimal load increase and load decrease, respectively.

For each dispatch, multiple simulations are performed to find the optimal load increase or load decrease amount. Those values are always less than the tie line active power flows. Taking one dispatch (1st January, hour 22), for instance, Figure 4a shows the frequency of two islands when WECC-1 RAS performs the load increase and load decrease according to the total tie-line flows, while Figure 4b shows the frequency of two islands when WECC-1 RAS performs the optimal load increase and load decrease. Note that the average frequency of each bus is used to present the bus frequency of each island. Under this dispatch, the

total tie-line flows between the two islands are 6177.9 megawatt (MW) and -6073.3 MW, respectively. The optimal load increase and decrease are 2996.9 MW and -4928.9 MW, respectively. The optimal load increase and load decrease values are less than the total tie-line flows while maintaining the frequency above 59.5 Hz and below 60.5 Hz.
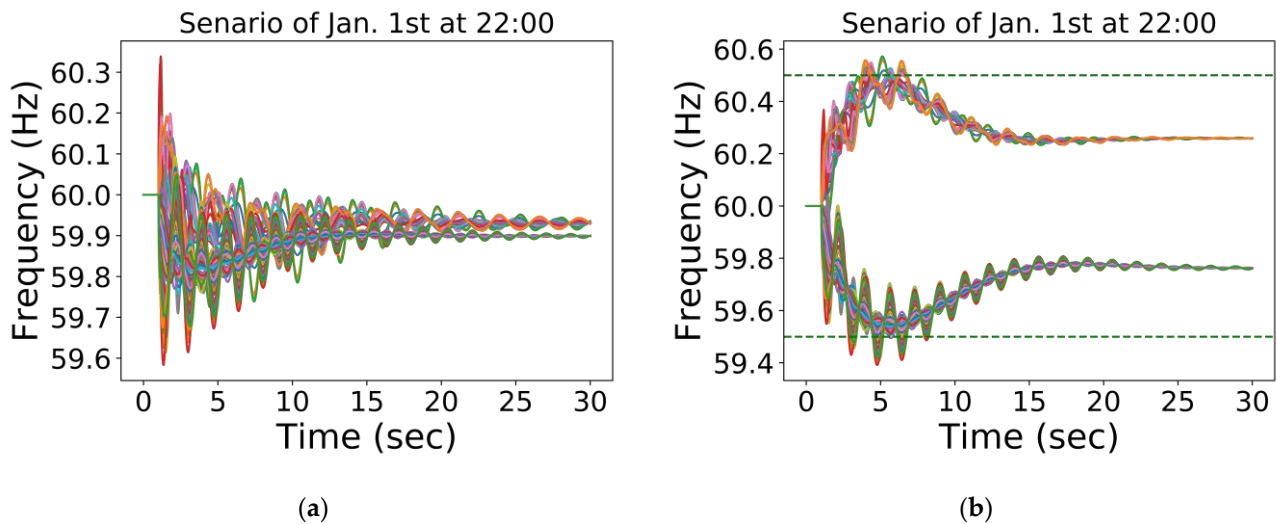

(**a**)   (**b**)

Figure 4. Frequency of two islanded grids. (**a**) Load decrease and load increase according to total tie line flows; (**b**) Optimal load decrease and load increase.

4. Deep Learning-Based Adaptive RAS

4.1. Deep Learning Neural Network Structure for Adaptive RAS

An artificial neural network (ANN) is an extremely simplified representation of the human brain and incorporates two fundamental components of biological neural nets: neurons and synapses. Neurons are computation nodes summing signals from previous neurons and decide whether to fire this neuron or not by applying an activation function, which is usually nonlinear. Synapses are the connection between neurons and are represented by weights. Figure 5 shows those two fundamental elements of an artificial neural network. Synapses connect neurons from the previous layer with the neuron in the next layer through weights. The $u_1, \ldots, u_n$ are outputs of n neurons from the previous layer, and are connected to a neuron in the next layer by synapses. The neuron sums up all weighted output from neurons from the previous layer connected by synapses and goes through an activation function to decide whether to fire this neuron or not. Finally, the output of this neuron z_1 are connected to other neurons from the next layers by synapses again.
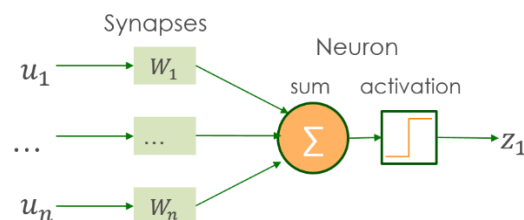

Figure 5. Fundamental components of artificial neural network.

ANN claims that one hidden layer is able to represent any nonlinear function; however, deep learning neural networks can use multiple layers to abstract the information based on the previous layer and progressively abstract the input features, resulting in better generalization [18]. As shown in Figure 6, this paper uses a fully connected feedforward neural network with more than one hidden layer to map the operating conditions to the optimal WECC-1 RAS corrective actions. The operating conditions include total generation,

total load, total inertia, power output of each generator, inertia of each generator, and each load. The outputs are the corrective actions, such as load increase MW amount and load decrease MW amount in two islands, respectively.
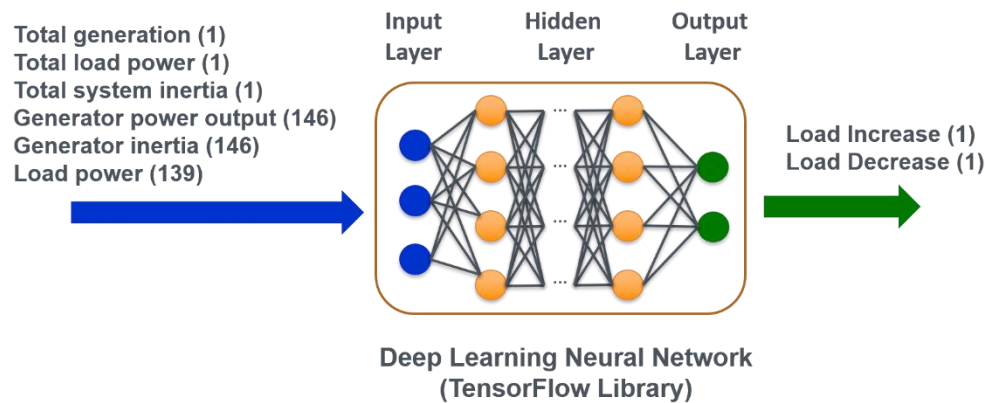

Figure 6. Fundamental components of deep learning neural network.

4.2. ANN Model Training, Validation, and Testing

Take a simplest neural network with one neuron for example, as in Figure 5. For a regression problem, the mapping between n input feature u and target output $\hat{z}$ can be represented per Equation (1). In the equation, φ represents any nonlinear or linear activation function, and b is the bias for that neuron. By substituting the m samples into Equation (1), it can get m equations about weights w. Solving those sets of nonlinear equations for weights w and bias b directly is very hard, especially when there are thousands of samples.

The purpose of training is to find the correct weights in order to make the model output as close to the ground truth as possible. In other words, to make the error as small as possible. Taking mean squared error loss function for example, it can be expressed as Equation (2) for m training samples. Gradient descent is an optimization algorithm for finding minimum value of an objective function which ties to minimize the error between the model output and the ground truth. By using gradient descent, each weight w_i can be updated towards a smaller error direction per the function (3), α is a hyperparameter learning rate.

$$\hat{z} = \varphi\left(b + \sum_{i=1}^{n} u_i * w_i\right) \tag{1}$$

$$E = \sum_{j=1}^{m} \left(\hat{z}_j - z_j\right)^2 \tag{2}$$

$$w_i = w_i - \alpha \frac{dE}{dw_i} \tag{3}$$

In Equation (3), loss function E is a complicated nonlinear function related to weights w, so it is not easy to get the derivative of the loss function to weights directly. Backpropagation provides a way to backpropagate the loss of the output to previous nodes' weight. By using the chain rule of the derivative and backpropagation, $\frac{dE}{dw_i}$ is easy to be calculated. In the Figure 5 example, we can set $out = b + \sum_{i=1}^{n} u_i * w_i$. For example, for calculating $\frac{dE}{dw_1}$, by using backpropagation and chain rule, we can get the following Equation (4)

$$\frac{dE}{dw_1} = \frac{dE}{d\hat{z}} * \frac{d\hat{z}}{d\left(b + \sum_{i=1}^{n} u_i * w_i\right)} * \frac{d\left(b + \sum_{i=1}^{n} u_i * w_i\right)}{dw_i} = \frac{dE}{d\hat{z}} * \frac{d\hat{z}}{d(out)} * \frac{d(out)}{dw_i} \tag{4}$$

During the training process, weights are updated to minimize the loss function. Without validation, model training will continue until the error between the ground truth

and the output of the model becomes zero. Finally, the mode would be overfitting to the training dataset and would not be able to generalize to the testing dataset.

The validation process is to prevent overfitting. Normally after each epoch during training, the validation data will be fed into the model, and the validation error will be calculated. When the validation error does not decrease anymore, but the training error continues to decrease, the training process should be stopped; otherwise, the model is overfitting.

Figure 7 shows a model training process. In the figure, after 1200 training epochs, the validation performance does not improve, but the training error continues to decrease. The model should stop training around 1200 epochs to avoid overfitting.
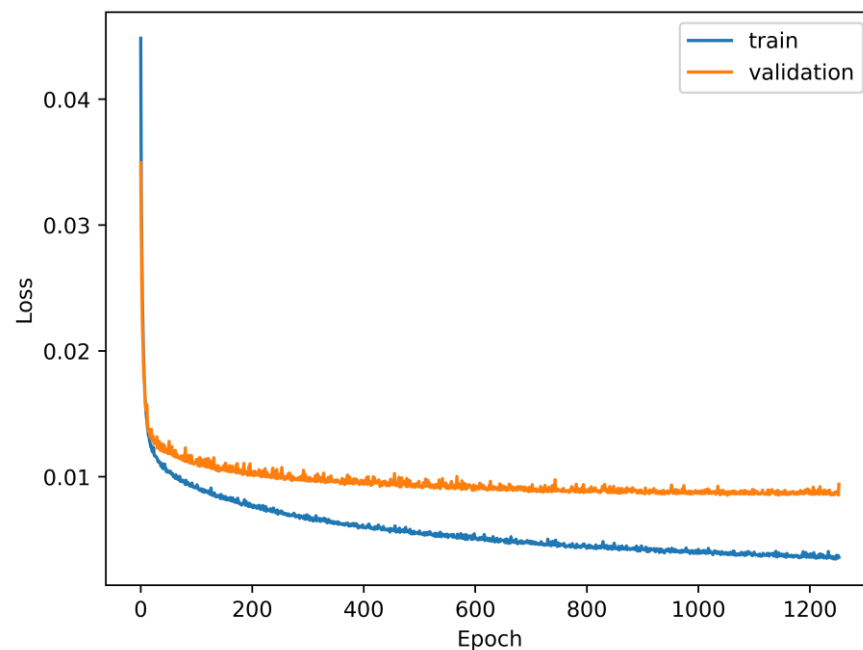

Figure 7. A model raining and validation process.

After training and validation, the model's weights are fixed. During the testing stage, the model makes predictions for the testing dataset, and the model's prediction error can be calculated. Model performance based on the testing dataset is used to measure the model's performance.

4.3. Customized Loss Function

Mean squared error and mean absolute error are commonly used loss functions. However, the error between ground truth and the output of the model is a normal distribution around a mean of zero when using these two standard loss functions.

If we use the standard loss function, the predicted load decrease value could be less than the actual value, which is a critical value to maintain one island's frequency above 59.5 Hz. This could trigger under-frequency load shedding unnecessarily. Similarly, the predicted load increase amount could be less than the ground truth that can maintain the other island's frequency below 60.5 Hz, resulting in an unwanted over-frequency generation trip. Therefore, in the training process, the model should not only minimize the loss but also make the loss in a conservative direction.

In this study, a customized loss function is proposed to make the model favor the positive error more than the negative error. The customized loss function is shown in Equation (5).

$$E = \begin{cases} (\hat{Z} - Z)^2 \; if \; |\hat{Z}| > |Z| \\ (N * (\hat{Z} - Z))^2 \; if \; |\hat{Z}| < |Z|, \; N > 1 \end{cases} \tag{5}$$

where Z is the ground truth of load increase or load decrease amount, $\hat{Z}$ is predicted load increase and load decrease amount, and N is the penalizing factor.

When $|\hat{Z}|$ is greater than $|Z|$, the two islands' frequency will not trigger under-frequency load shedding or over-frequency generation trip relays. When $|\hat{Z}|$ is less than $|Z|$, the loss function will be penalized by enlarging N times. With this customized loss function, the model will try to predict the value greater than the ground truth more frequently. In other words, N decides the degree of conservativeness: the greater the penalizing factor N, the more conservative the prediction results.

4.4. Feature Normalization

When the input features have different scales, the normalization of those features can help speed up the training process. The normalization could be the min–max normalization which scales all the features within 0 to 1, or the standard normalization which scales all the features with a mean of zero and a standard deviation of one. In this study, min-max normalization is used. For instance, min–max normalization for feature p can be expressed as follows:

$$p' = \frac{p - \min(p)}{\max(p) - \min(p)} \tag{6}$$

4.5. Evaluation Metrics

Five metrics—root mean squared error (RMSE), root mean squared percentage error (RMSPE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) [19–21]—are used to evaluate the prediction error of the model.

RMSE is the square root of mean squared error (MSE). MSE is the average of all the squared errors in the prediction which shows how much the predicted result deviates from the ground truth. RMSE is an extension of MSE, but has the same unit as the prediction value. The calculation of RMSE is given in Equation (7).

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2} \tag{7}$$

In Equation (7), m is the number of testing data, $\hat{y}_i$ is the prediction value from the model, and y_i is the ground truth.

RMSPE is very similar to RMSE, but the percentage error is used for calculation instead of the error itself. Percentage error is defined as the error divided by the ground truth. The load increase and load decrease amount are a scale of a thousand MW to several thousand MW in WECC RAS-1, and the error could be several MW to even hundreds of MW depending on different dispatches. Therefore, using the percentage error can normalize the error scale difference in different operating conditions. The calculation of RMSPE is given in Equation (8).

$$RMSPE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2} \tag{8}$$

The calculation of MAE is given in Equation (9). The MAE uses the same scale as the data being measured, while mean absolute percentage error (MAPE) can be used to make comparisons between data with different scales. The MAE and MAPE calculation are given in Equations (9) and (10).

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{9}$$

$$MAPE = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{10}$$

R^2 measures the degree of the linear correlation between the predicted value and the ground truth. R^2 is a value between 0 and 1. The bigger value, the better the model fits the data. R^2 can be expressed in (11).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{11}$$

5. Numerical Studies

5.1. Layer Number and Node Number Selection

The deep learning neural network model defined in Figure 6 is trained and validated with 80% of the total dataset. This work uses TensorFlow, which is an open-source software library for machine learning [22]. During the training for the model, the early stopping technique is used to prevent overfitting [23]. The remaining 20% dataset is used for performance evaluation after the model is trained. The model can have different layer numbers and different node numbers in each layer. A model with too many weights tends to be overfitting, but with insufficient weights tends to be underfitting. In this study, models with two, five, and seven hidden layers and with 50, 100, 300, 500, 800, 1000, 1500, 2000, and 3000 nodes in the first layer, and half the number of the previous layer nodes in the latter layer, are compared.

Figure 8 shows different models' performances with metrics MAE, RMSE, and R^2. The model with two hidden layers and the model with five hidden layers have similar performances. Figure 9 shows different models' performances with metrics MAPE and RMSPE. The model with two hidden layers and 1000 nodes in each layer has the best performance overall. However, the model with two hidden layers and 300 nodes in each layer's performance improved a lot from 50 nodes in each layer to 300 nodes in each layer. The model's performance with 1000 nodes in each layer did not improve dramatically when nodes were increased from 300 to 1000 in each layer. In this paper, considering that the complex model is more likely to be overfitted and hard to train, the model with two hidden layers with 300 nodes was selected as the optimal model structure instead.
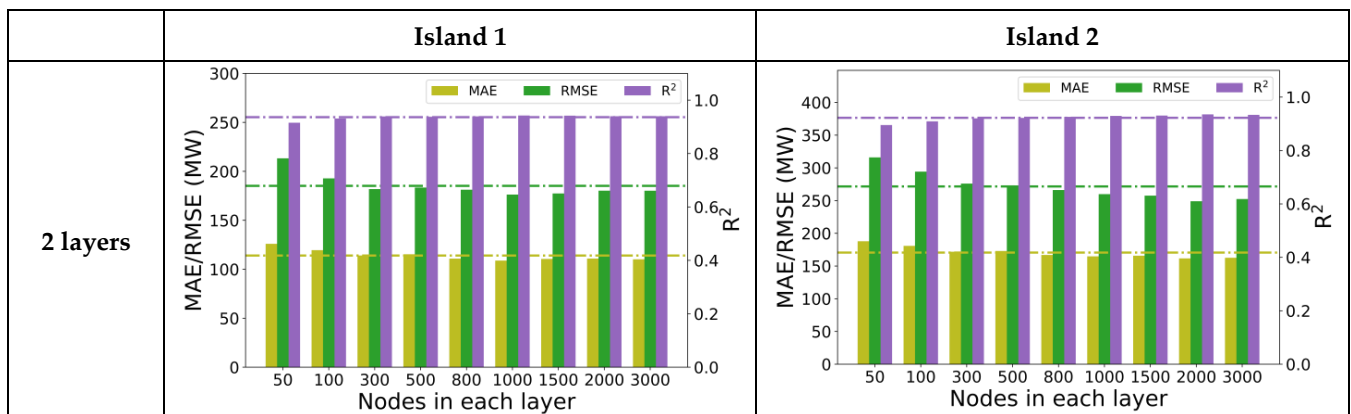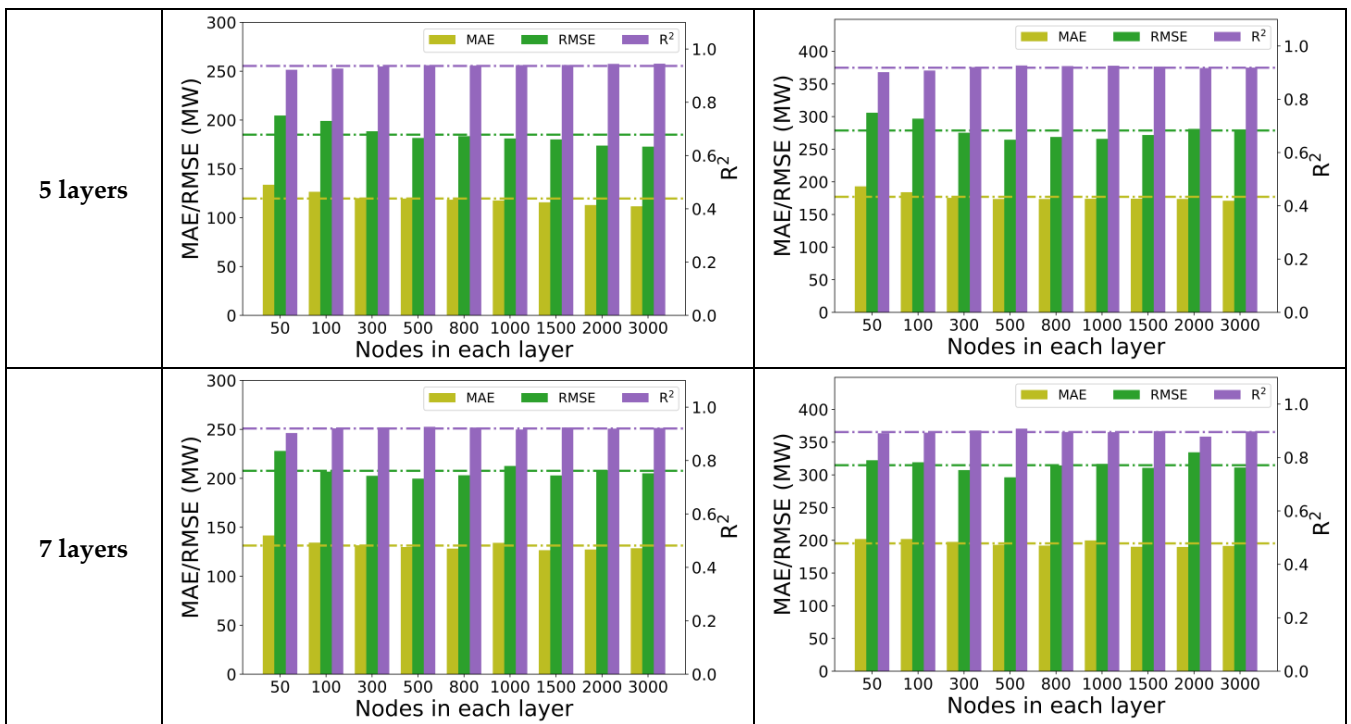

Figure 8. *Cont.*

Figure 8. Model performance MAE/RMSE/ R^2 of models with different layers and nodes.
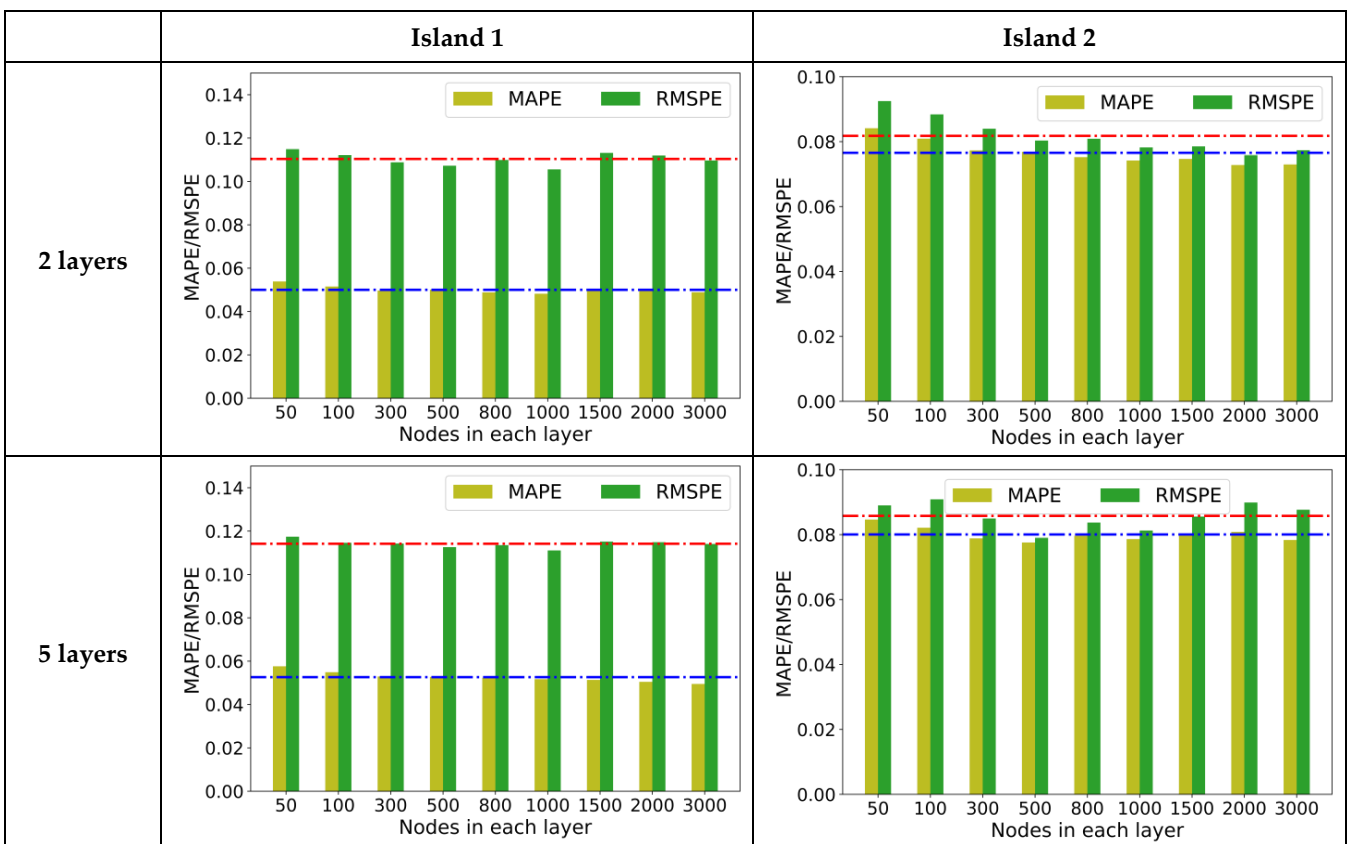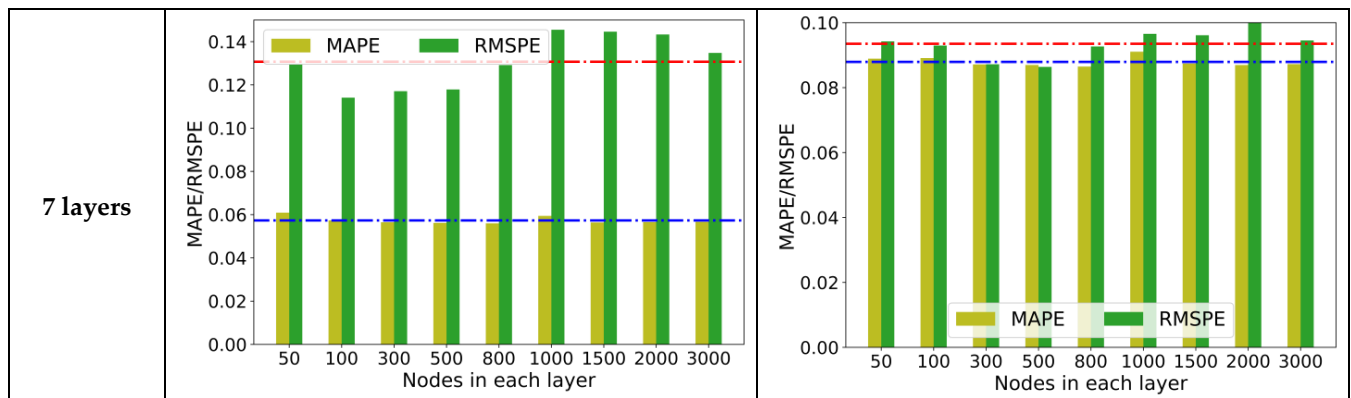

Figure 9. *Cont.*

Figure 9. Model performance MAPE/RMSPE of models with different layers and nodes.

5.2. Regularization Hyper Parameter Tuning

Table 1 shows the performance of the trained model with different L1, L2 regularization techniques [24]. When both L1 and L2 are equal to 1, the model can achieve the best performance. Tables 2 and 3 shows the performance of the trained model with different dropout regularization techniques and Gaussian noise levels with different standard deviation (std), respectively [25,26]. Neither dropout regularization nor Gaussian noise can help significantly improve the model performance, and thus they are not used during the model training process.

Table 1. Performance of trained model with L1 and L2 regularization techniques.

Hidden Layers /Nodes	L1, L2	RMSE (MW) Island 1/Island 2	MAE (MW) Island 1/Island 2
	0	182.31/275.62	112.05/169.51
	0.0001	185.24/276.63	116.20/173.58
	0.001	187.58/280.95	116.65/173.68
	0.01	185.62/279.33	116.19/175.02
2/300	0.1	175.12/259.69	107.88/163.45
	1	132.23/206.36	87.64/134.07
	5	133.17/209.35	89.67/136.25
	10	134.49/206.92	91.50/135.64
	20	142.41/224.85	101.61/152.53

Table 2. Performance of the trained model with Dropout regularization techniques.

Hidden Layers /Nodes	Dropout Rate (%)	RMSE (MW) Island 1/Island 2	MAE (MW) Island 1/Island 2
	0	182.31/275.62	112.05/169.51
	0.001	182.02/275.30	113.62/172.90
	0.01	182.51/276.03	114.89/171.89
2/300	0.05	191.58/281.00	124.78/181.59
	0.1	193.39/277.66	127.18/179.49
	0.3	208.59/298.81	141.53/197.86

5.3. Performance with Customized Loss Function

Although the trained model above can reach small RMSE and MAE, its performance is not good enough to maintain the frequency above 59.5 Hz in one island and below 60.5 Hz in the other island. By adding the customized loss function, the predicted results are more conservative to maintain the frequency. Table 4 gives the performance of the model with different penalizing factor *N*. The frequency between 59.5 Hz and 60.5 Hz is considered the secure region. As shown in Table 4, with a larger N, RMSE and

MAE get larger, but the frequency inside the security region is higher. This means the customized loss function can make the predicted value more conservative by penalizing the values out of the secure region. For instance, when N = 20, the frequency of 97.77% cases are within the secure region.

Table 3. Performance of trained model with Gaussian Noise regularization techniques.

Hidden Layers /Nodes	Gaussian Noise (std)	RMSE (MW) Island 1/Island 2	MAE (MW) Island 1/Island 2
	0	182.31/275.62	112.05/169.51
	0.0001	185.90/279.24	116.81/175.06
2/300	0.001	185.53/277.87	116.65/173.68
	0.01	179.49/262.96	115.21/172.00
	0.1	212.98/315.83	149.92/217.31
	1	454.10/656.37	345.28/494.78

Table 4. Prediction errors with different penalizing factors.

N	RMSPE Island 1/Island 2	MAPE Island 1/Island 2	Inside Security Region (%)
1	7.15%/4.94%	3.50%/3.04%	58.09
2	8.37%/5.43%	4.73%/3.49%	76.70
3	6.90%/5.09%	3.84%/3.14%	73.14
10	14.10%/7.89%	8.86%/5.93%	91.78
15	14.23%/9.50%	9.40%/6.83%	92.36
20	28.23%/14.54%	18.36%/11.44%	97.77

Figure 10 shows one-day testing cases' load shedding (load decreasing) and generation trip (load increasing) amount prediction after RAS in two islanded systems, respectively, based on the model without penalization (N = 1) and with penalization (N = 20). It shows that the prediction result based on the model without penalization has better accuracy and is closer to the ground truth compared to the result based on the model with penalization (N = 20), but it does not have the safety margin and will trigger the over-frequency generation trip and under-frequency load shedding. The prediction result based on the model with penalization shows a good safety margin in Figure 10, with very few cases having frequency violations.

Figure 11a shows the maximum and minimum frequency in two islanded systems, respectively, based on the prediction results without penalization (N = 1). The shadow area is the security region where UFLS and OVGT are not triggered. With no penalization, only 58.09% of the cases are in the security region. Figure 11b,c show the histogram of maximum frequency in Island 1 and the histogram of minimum frequency in Island 2.

Similarly, Figures 12 and 13 show the maximum and minimum frequency of the two islanded systems, respectively, and also show the histogram of maximum frequency in Island 1 and the histogram of minimum frequency in Island 2 when N =10 and N =20. In 91.78% (or 97.77%) of the total cases, the maximum and minimum frequency are inside the secure region when N = 10 (or N = 20).

5.4. Comparison with Traditional RAS

Traditional RAS is designed based on the rate of active power change to frequency change to estimate the generation trip and load shedding for each operating condition.

Since the WECC system is separated into two islands after the execution of WECC-1 RAS, the rate of active power change to frequency change is calculated separately for two islanded systems. Two scenario simulations are performed in order to calculate the rate r_{11} and r_{21} for one dispatch.
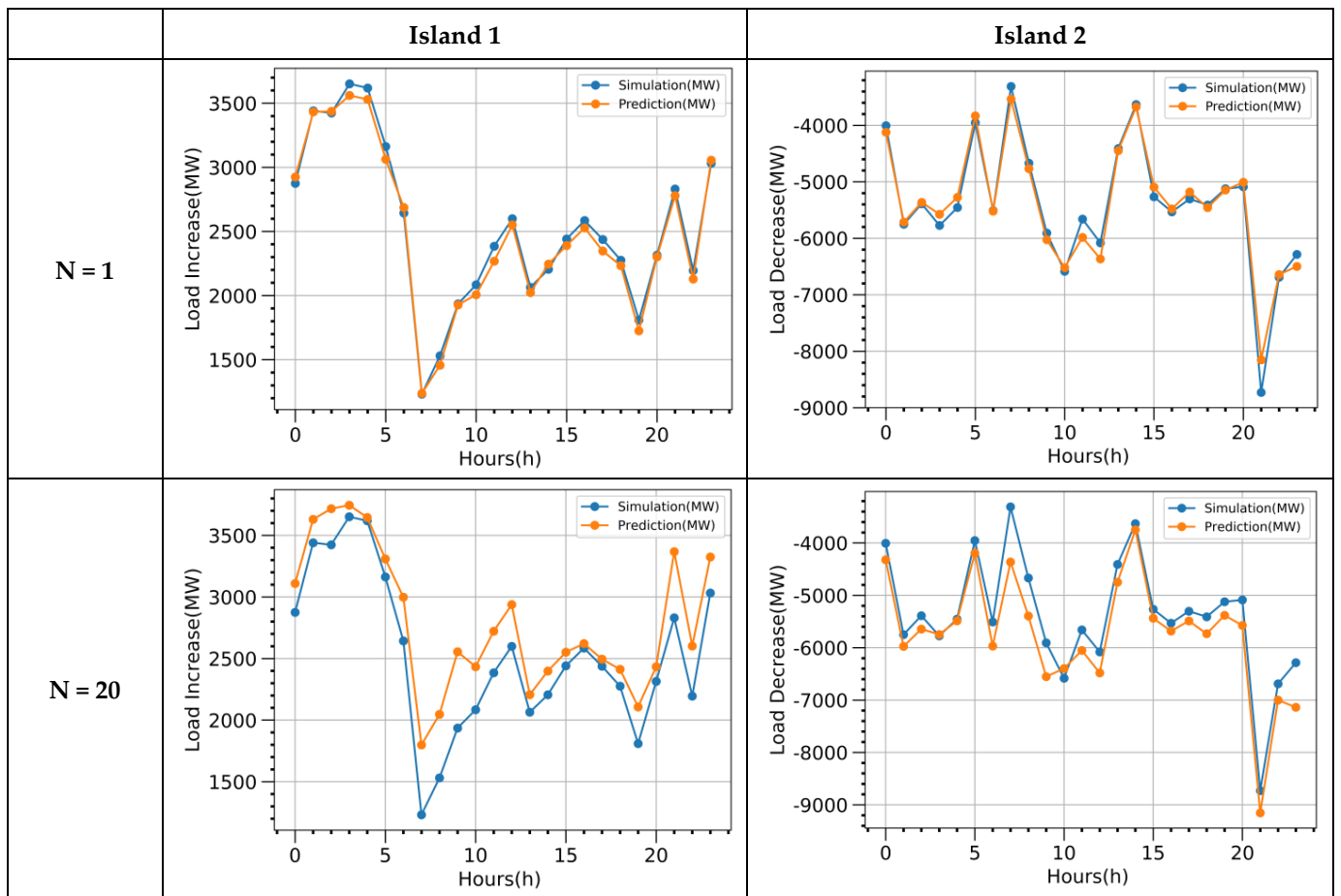
Figure 10. One-day prediction result based on model with penalization and without penalization.

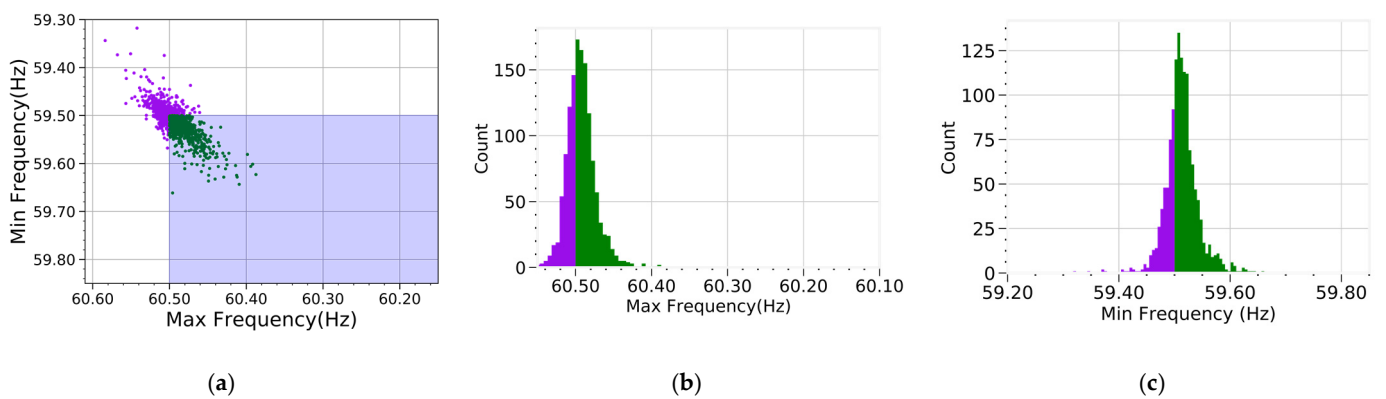
(**a**)　　　　　　　　　　　　　　(**b**)　　　　　　　　　　　　　　(**c**)

Figure 11. Maximum and minimum frequency and histograms of Island 1 and Island 2: N = 1 (purple: outside secure region, green: inside secure region, shaded area: secure region). (**a**) Max. and Min. frequency; (**b**) Max. frequency histogram in Island 1; (**c**) Min. frequency histogram in Island 2.
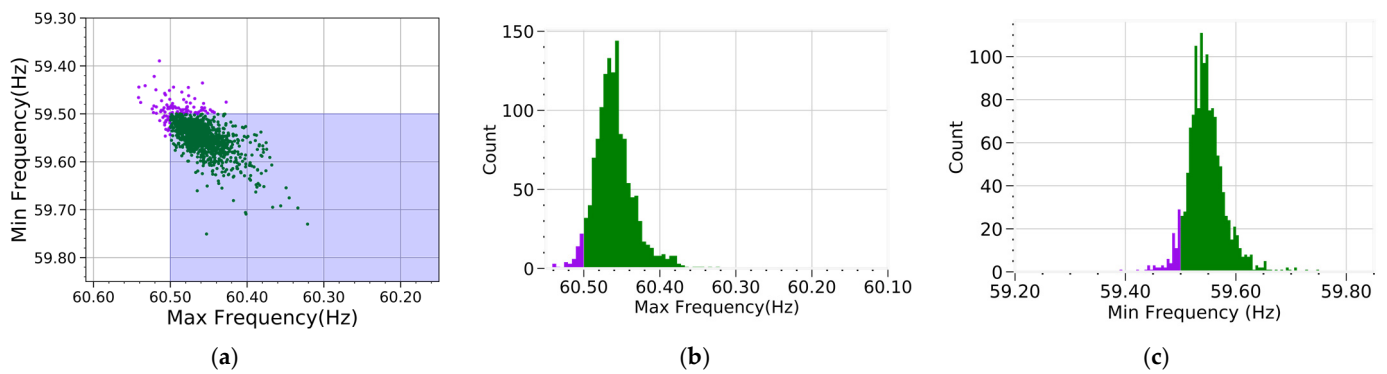
Figure 12. Max. & Min. frequency and histograms of Island 1 and Island 2: N = 10. (Purple: outside secure region, green: inside secure region, shaded area: secure region). (**a**) Max. and Min. frequency; (**b**) Max. frequency histogram in Island 1; (**c**) Min. frequency histogram in Island 2.
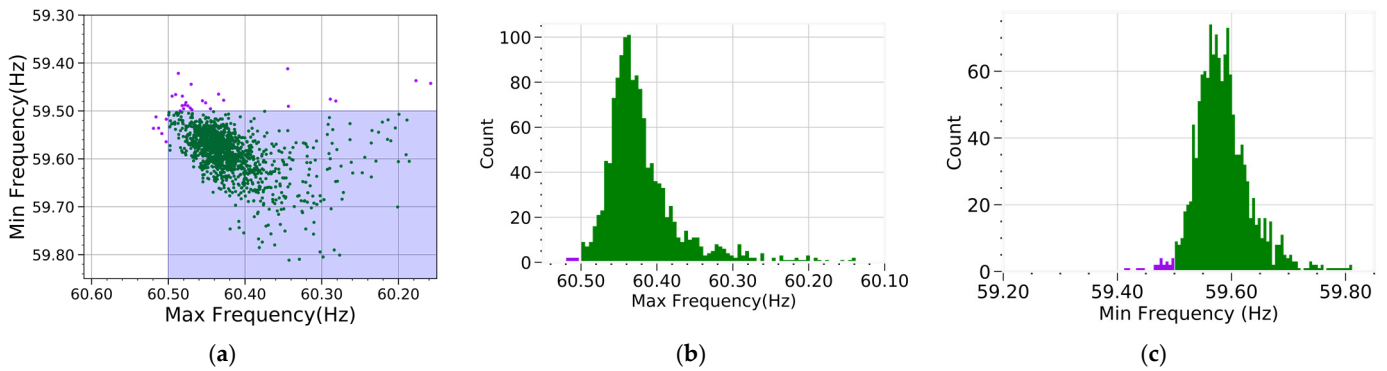

Figure 13. Maximum and minimum frequency and histograms of Island 1 and Island 2: N = 20 (purple: outside secure region, green: inside secure region, shaded area: secure region). (**a**) Max. and Min. frequency; (**b**) Max. frequency histogram in Island 1; (**c**) Min. frequency histogram in Island 2.

In scenario 1, if load decrease P_1 and load increase P_2 performed in two islanded systems equal to the power flow of the tie lines connecting these two systems before RAS, the two islanded systems' frequency nadir and frequency maximum will be f_{11} and f_{21}.

In scenario 2, if optimal load decrease P_{o1} and optimal load increase P_{o2} are performed in two islanded systems, the two islanded systems' frequency nadir and frequency maximum will be 59.5 Hz and 60.5 Hz. The rate r_{11} and r_{21} for two islanded systems can be calculated in Equation (12).

$$r_{11} = (P_1 - P_{o1})/(f_{11} - 59.5) \quad r_{21} = (P_2 - P_{o2})/(f_{21} - 60.5) \tag{12}$$

The average rates of four typical scenarios: heavy load case, light load case, high renewable penetration case, and low renewable penetration case are calculated and used as the final rates of active power change to frequency change for two islanded systems in WECC system in Equation (13). The average frequency nadir and frequency maximum of four typical scenarios in two islanded systems when performing load decrease and load increase using tie-line flows, are calculated in Equation (14).

$$\begin{cases} r_1 = \sum_{i=1}^{4} r_{1i} \\ r_2 = \sum_{i=1}^{4} r_{2i} \end{cases} \tag{13}$$

$$\begin{cases} f_1 = \sum\limits_{i=1}^{4} f_{1i} \\ f_2 = \sum\limits_{i=1}^{4} f_{2i} \end{cases} \tag{14}$$

In actual system operation, tie-line flows between two subsystems P_{1r} and P_{2r} can be monitored and obtained once RAS is detected. Then, the optimal load decrease amount P_{o1r} and load increase amount P_{o2r} in two islanded systems for that operation condition can be calculated in Equation (15).

$$P_{o1r} = P_{1r} - r_1 * (f_1 - 59.5) \tag{15}$$

where r_1, r_2, f_1 and f_2 can be calculated in Equations (13) and (14) using offline simulations on four typical scenarios. P_{1r} and P_{2r} are known for each operating condition.

Based on traditional RAS calculation in Equation (15), Table 5 shows the performance of the traditional RAS. Only in 33.78% of the cases, the frequency is inside the security region.

Table 5. Performance on testing data based on traditional RAS calculations.

RMSPE Island 1/2	MAPE Island 1/2	Inside Security Region (%)
9.65%/19.70%	5.65%/14.09%	33.78

Figure 14 shows the frequency distributions of two islanded systems based on traditional RAS. The frequency of one system is scattered in a larger range compared to the other system.
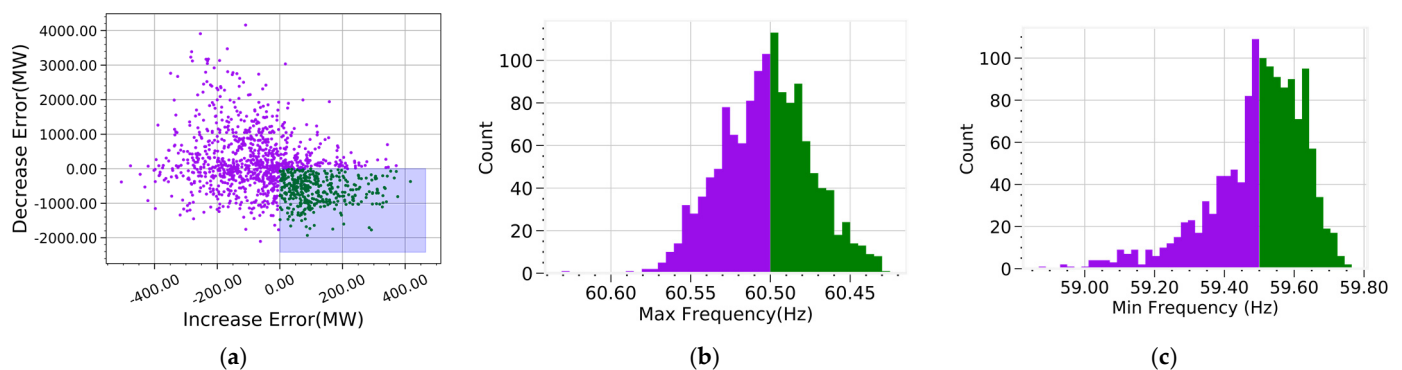

Figure 14. Maximum and minimum frequency and histograms of Island 1 and Island 2: Traditional RAS (purple: outside secure region, green: inside secure region, shaded area: secure region). (**a**) Max. and Min. frequency; (**b**) Max. frequency histogram in Island 1; (**c**) Min. frequency histogram in Island 2.

6. Conclusions

In this paper, the deep learning–based adaptive RAS is developed and validated on the reduced 240-bus WECC system model. The optimal load decrease and load increase values can be directly predicted based on the operating conditions. The optimal model structure is selected as two layers. When using the standard loss function, the prediction error complies with the normal distribution, and the frequency is below 59.5 Hz or above 60.5 Hz in about 50% of the cases. Therefore, a customized loss function is proposed to make the prediction error in the conservative direction. The simulation results demonstrate that the frequency is inside the secure region in 97.77% of the cases when using the customized loss function with N = 20, while only 58.09% when using the standard loss function. Moreover, the proposed adaptive RAS has better performance than the traditional RAS.

In reality, fast load shedding and generation trip are commonly used as corrective actions to maintain system transient stability. This paper uses proportional load increase in one island and proportional load decrease in the other island as the first step to determine the optimal load shedding and generation trip schemes for a specific and practical case. Future work will focus on using fast load shedding and generation trip to generate the training/testing dataset and further validation of the proposed adaptive RAS.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Disclaimer: This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This manuscript was also authored in part by Oak Ridge National Laboratory, operated by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

1. O'Brien, J.; Barrett, E.; Fan, X.; Diao, R.; Huang, R.; Huang, Q. Adaptive RAS/SPS System Settings for Improving Grid Reliability and Asset Utilization through Predictive Simulation and Controls, Task 1 Report—Survey on RAS/SPS Modeling Practice. October 2017. Available online: https://www.osti.gov/servlets/purl/1580707 (accessed on 4 June 2021).
2. Mahmoudi, M.M.; Kincic, S.; Zhang, H.; Tomsovic, K. Implementation and testing of remedial action schemes for real-time transient stability studies. In Proceedings of the 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 16–20 July 2017; pp. 1–5. [CrossRef]
3. Varghese, M.; Jin, L.; Ghosh, S.; Lin, G.; Pek, B. The CAISO experience of implementing automated remedial action schemes in energy management systems. In Proceedings of the 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Canada, 26–30 July 2009; pp. 1–5. [CrossRef]
4. Ajao, B.; Khaledian, P.; Johnson, B.K.; Chakhchoukh, Y. Implementation of Remedial Action Scheme for Transient Stability Index Improvement of Power System Island. In Proceedings of the 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 17–20 February 2020; pp. 1–5. [CrossRef]
5. Ackermann, T.; Prevost, T.; Vittal, V.; Roscoe, A.J.; Matevosyan, J.; Miller, N. Paving the Way: A Future without Inertia Is Closer Than You Think. *IEEE Power Energy Mag.* **2017**, *15*, 61–69. [CrossRef]
6. Brooks, A.; Lu, E.; Reicher, D.; Spirakis, C.; Weihl, B. Demand Dispatch. *IEEE Power Energy Mag.* **2010**, *8*, 20–29. [CrossRef]
7. Quiros-Tortos, J.; Ochoa, L.; Butler, T. How Electric Vehicles and the Grid Work Together: Lessons Learned from One of the Largest Electric Vehicle Trials in the World. *IEEE Power Energy Mag.* **2018**, *16*, 64–76. [CrossRef]
8. Chen, J.; Zhu, Q.; Li, H.; Zhu, L.; Shi, D.; Li, Y.; Duan, X.; Liu, Y. Learning Heterogeneous Features Jointly: A Deep End-to-End Framework for Multi-Step Short-Term Wind Power Prediction. *IEEE Trans. Sustain. Energy* **2020**, *11*, 1761–1772. [CrossRef]
9. Alimi, O.A.; Ouahada, K.; Abu-Mahfouz, A.M. A Review of Machine Learning Approaches to Power System Security and Stability. *IEEE Access* **2020**, *8*, 113512–113531. [CrossRef]

10. Hou, Q.; Du, E.; Zhang, N.; Kang, C. Impact of High Renewable Penetration on the Power System Operation Mode: A Data-Driven Approach. *IEEE Trans. Power Syst.* **2020**, *35*, 731–741. [CrossRef]

11. Yan, Z.; Xu, Y. Data-Driven Load Frequency Control for Stochastic Power Systems: A Deep Reinforcement Learning Method With Continuous Action Search. *IEEE Trans. Power Syst.* **2019**, *34*, 1653–1656. [CrossRef]

12. Huang, Q.; Huang, R.; Hao, W.; Tan, J.; Fan, R.; Huang, Z. Adaptive Power System Emergency Control Using Deep Reinforcement Learning. *IEEE Trans. Smart Grid* **2020**, *11*, 1171–1182. [CrossRef]

13. Hashmy, Y.; Yu, Z.; Shi, D.; Weng, Y. Wide-Area Measurement System-Based Low Frequency Oscillation Damping Control Through Reinforcement Learning. *IEEE Trans. Smart Grid* **2020**, *11*, 5072–5083. [CrossRef]

14. Yuan, H.; Biswas, R.S.; Tan, J.; Zhang, Y. Developing a Reduced 240-Bus WECC Dynamic Model for Frequency Response Study of High Renewable Integration. In Proceedings of the 2020 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Chicago, IL, USA, 12–15 October 2020; pp. 1–5.

15. Western Interconnection RAS Review. Available online: https://www.wecc.org/Reliability/Western%20Interconnection%20 RAS%20Review%20Guideline.pdf (accessed on 4 June 2021).

16. Test Case Repository for High Renewable Study. Available online: https://www.nrel.gov/grid/test-case-repository.html (accessed on 4 June 2021).

17. Available online: https://www.nrel.gov/docs/fy20osti/73856.pdf (accessed on 15 September 2021).

18. Deep Learning. Available online: https://en.wikipedia.org/wiki/Deep_learning (accessed on 4 June 2021).

19. Available online: https://www.mdpi.com/1996-1073/11/9/2328/htm (accessed on 15 September 2021).

20. Available online: https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.2419 (accessed on 15 September 2021).

21. Available online: https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/ep.12558 (accessed on 15 September 2021).

22. TensorFlow. Available online: https://www.tensorflow.org/ (accessed on 4 June 2021).

23. Available online: https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping/ (accessed on 2 April 2021).

24. Available online: https://www.machinecurve.com/index.php/2020/01/23/how-to-use-l1-l2-and-elastic-net-regularization-with-keras/ (accessed on 2 April 2021).

25. Available online: https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/ (accessed on 2 April 2021).

26. Available online: https://keras.io/api/layers/regularization_layers/gaussian_noise/ (accessed on 2 April 2021).