Data-driven transient frequency stability assessment: A deep learning method with combined estimation-correction framework

Yunfeng Wen¹ | Rongzhen Zhao² | Mingzeng Huang¹ | Chuangxin Guo³

¹ College of Electrical and Information Engineering, Hunan University, Changsha, Hunan Province, China
² Nanjing NR Electric Co. Ltd., Nanjing, Jiangsu Province, China
³ College of Electrical Engineering, Zhejiang University, Hangzhou, Zhejiang Province, China

Correspondence
Yunfeng Wen, College of Electrical and Information Engineering, Hunan University, Changsha, Hunan Province, China.
Email: yfwen@hnu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 52077066; Hunan Provincial Natural Science Foundation for Excellent Young Scholars, Grant/Award Number: 2020JJ3011

Abstract
Transient frequency stability assessment (TFSA) is important for operators to understand dynamic power system frequency characteristics. However, because of the problem of combinatorial explosion, exacerbated by uncertain renewable generation and various contingencies, applying traditional time-domain simulation methods is very time-consuming for TFSA of a low-inertia power system. To address this, this study evaluates a fast and online TFSA using a data-driven tool, based on deep learning. A novel combined estimation–correction learning framework is proposed. In the estimation step, a deep neural network is used to obtain the network parameters based on actual input–output feature data, which effectively realize automatic dimension reduction and feature extraction. In the subsequent correction step, the multidimensional frequency metrics produced by the deep neural network-based estimator are input to a stacked extreme learning machine-based corrector to further reduce the prediction error. A case study on a modified IEEE RTS-79 system demonstrates that the proposed approach has extremely high computation speed, compared to the time-domain simulation method, and achieves higher prediction accuracy and superior generalisation performance in comparison with other state-of-the-art algorithms.

1 INTRODUCTION

With the increasing penetration of renewable generation, a considerable portion of conventional synchronous units (coal and hydro) has been replaced by renewable energy generation units. However, because of their power–electronic interfaces (non-synchronous) to power systems and tendency to fluctuation, randomness, and intermittency, renewable units typically provide low to zero inertial response. Consequently, the increasing penetration of renewable generation drastically reduces grid inertia, thus significantly challenging transient frequency stability [1,2].

Presently, the time domain simulation (TDS) method is commonly used for transient frequency stability assessment (TFSA) [3], which can accurately model the transient frequency characteristics of a power system following an active power disturbance. However, because of the inherent problems of model complexity and excessive computational time, the TDS method is only suitable for offline analysis. In the ‘combinatorial explosion’ case, in view of multiple uncertainties (stochastic wind/solar generation, load fluctuation, and contingencies), it would be challenging to apply TDS for online TFSA. Operators in a control centre cannot understand post-fault dynamic power system frequency characteristics timely and precisely in the absence of online TFSA tools. Therefore, preventive measures to deal with large disturbances may not be properly prepared in advance. This causes an urgent need to develop a fast method that is applicable to the online evaluation of transient frequency stability.

Recently, the application of machine learning technologies in power systems has received considerable attention [4]. A few preliminary works in the field of machine-learning-based TFSA have been carried out in [5–10]. Wu et al. [5] and Djukanovic et al. [6] adopted a back propagation (BP)-algorithm-based neural network to predict the post-fault frequency nadir metric of a power system under sudden disturbances. However, the BP algorithm has many shortcomings, such as the saw-tooth phenomenon, low efficiency, and poor generalisation ability [11]. Chang et al. [7] applied a regression tree to select the most significant system characteristics between generation/load
imbalance and frequency decline. The main drawbacks of the regression tree are its ignorance of the correlation between the data attributes and the overfitting problem [4]. Bo et al. [8] predicted the frequency nadir using a support vector machine (SVM) method. An SVM uses inner product kernel functions instead of mapping the non-linearities to high-dimensional space, to avoid a dimensionality disaster. Nevertheless, the SVM has slow convergence and relatively low prediction accuracy, making it difficult to apply in large-scale data training [11].

Hong et al. [9] employed a genetic algorithm (GA) to determine the minimum curtailment at each stage for the under-frequency protection relays. Dai et al. [10] proposed a transient frequency feature evaluation method based on a single-layer extreme learning machine (ELM). Although the single-layer ELM has a high offline training speed, it is prone to making the weights of some neurons unreasonable because of the random generation of parameters between the input and hidden layers, thus compromising the prediction accuracy of the trained neural network [11].

The network structures in the above works are generally simple, making the analytical ability of typical features of the input data inefficient. It is difficult for them to map the input-output relationship under complex conditions. Additionally, over-fitting and under-fitting problems are difficult to resolve [12]. Feature selection is a critical step in these procedures. However, previous studies used a hand-engineered feature extraction method whose process is excessively arbitrary. This may result in the loss of a large amount of valuable information and reduced accuracy of the TFSA based on shallow learning algorithms [13]. Therefore, a more efficient method is necessary to achieve automatic extraction and optimize the feature set without any human involvement.

As one of the cutting-edge approaches, deep learning [14] can automatically extract spatial and temporal features via the multiple hidden layer structure, and thus perform complex classification and regression tasks. The effectiveness of the features extracted by deep learning is much better than that by artificial feature engineering, and the integrity of information is preserved to the greatest extent [15]. Therefore, the application of deep learning methods can not only improve the prediction accuracy and generalisation ability but also save manpower and simplify the processes. In recent years, deep learning methods have been applied to load forecasting [16], angle stability assessment [17], and early fault detection of gearbox bearings [18].

To achieve quick and accurate awareness of power system frequency response characteristics for transmission system operators and designers, this study proposes a data-driven method based on deep learning for online TFSA. Multiple post-fault frequency metrics can be predicted quickly and accurately using the developed intelligent tool, and effective preventive measures can be applied to keep the post-fault frequency metrics within tolerable limits, avoiding/reducing load curtailment and generation spillage. The main contributions of this work are summarised as follows:

(i) A novel deep-learning-based tool is proposed for transient frequency stability evaluation. This data-driven tool has a combined estimation-correction learning framework, which consists of two types of deep learning networks: a deep neural network (DNN) based estimator and a stacked ELM (SELM)-based corrector. In the estimation phase, the DNN is used to obtain the network parameters based on real input–output feature data, which can realize automatic feature extraction. In the correction phase, the frequency metrics produced by the estimator are input to the SELM-based corrector to further reduce the prediction error. We show that high accuracy and superior generalisation performance can be achieved by leveraging the combined estimation-correction deep learning framework.

(ii) Rather than focus on the prediction of a single post-fault frequency metric (e.g. frequency nadir) [5,6], multidimensional frequency metrics, including the maximum rate-of-change of frequency, frequency nadir, time to reach frequency nadir, and quasi-steady-state frequency deviation, are simultaneously predicted by the proposed approach. The obtained multidimensional frequency metrics are more useful for operators to comprehensively evaluate the power system transient frequency stability characteristics.

(iii) To verify the validity of the proposed method, a comprehensive comparative study is conducted on the training time, accuracy, and generalisation ability, in comparison with a time-domain simulation and a number of traditional and more recent machine-learning-based methods (BP, DNN, Stacked Denoising Autoencoders [SDAE], and ELM). The study results demonstrate that the performance of the proposed method enables real-time use, making it a practical tool for the industrial applications.

The remainder of the study is organised as follows: Section 2 presents the multidimensional metrics for assessing transient frequency stability. Section 3 proposes a deep learning method with a combined estimation-correction framework. Section 4 presents numerical results to illustrate the advantages of the proposed method. The conclusion is drawn in Section 5.

2 | MULTIDIMENSIONAL METRICS FOR ASSESSING TRANSIENT FREQUENCY STABILITY

System frequency dynamics \( f(t) \) following a sudden power disturbance incident can be expressed by the following differential equation [19]:

$$2H \frac{df(t)}{dt} = \Delta P_{G_i}(t) - \Delta P_L - D \Delta f(t)$$  \hspace{1cm} (1)

where \( \Delta f(t) \) represents the frequency deviation, \( H \) denotes the total inertia of the power system, which refers to the ability of the system to resist frequency change, \( D \) is the load damping rate, \( \Delta P_L \) denotes the amount of active power imbalance caused by the disturbance, and \( \Delta P_{G_i}(t) \) represents the active power variation of the \( i \)-th synchronous unit, which is determined by the unit's primary frequency control (PFC).

A typical waveform for the post-contingency transient frequency excursion is shown in Figure 1. During the initial several seconds (\( \Delta t \leq 5 \text{ s} \)), the frequency drop is only limited by
the inertial response of the synchronous units owing to the lag of the prime mover adjustment. After the governor dead-band, active spinning reserves are gradually deployed by the units’ PFC to reduce the power imbalance. The duration ($\Delta t$) of the PFC is typically 30 s. The transient frequency stability of a power system can be evaluated by the following multidimensional frequency metrics:

(i) **Frequency nadir (FN)**: FN (Hz) is the most significant frequency index, which represents the lowest point (when $\Delta P_L > 0$) or highest point (when $\Delta P_L < 0$) during the transient frequency excursion. To avoid triggering under-frequency load shedding (UFLS) and over-frequency generator tripping (OFGT) relays, the value of FN should be kept within its tolerable limits ($F_{N_{\text{min}}} \leq FN \leq F_{N_{\text{max}}}$).

(ii) **Time to reach the frequency nadir (TFN)**: TFN (s) is one of the main features of the dynamic system characteristics after a power disturbance, which represents the time duration before reaching the extreme frequency point.

(iii) **Rate-of-change of frequency (RoCoF)**: RoCoF (Hz/s) represents the slope of the frequency, which falls/increases immediately after a power imbalance. RoCoF is related to system inertia and the value of active disturbance. To avoid triggering the RoCoF protection relays (with measurement windows of several hundred ms, i.e. 500 ms), the value of RoCoF at the time of measurement should satisfy $|\text{RoCoF}| \leq \text{RoCoF}_{\text{max}}$.

(iv) **Quasi-steady-state frequency (QssF)**: QssF (Hz) is an indicator for secondary frequency control (SFC). If the value does not meet the steady-state frequency quality requirement, the SFC is activated to return the frequency within the tolerable band.

If the frequency variation satisfies $F_{N_{\text{min}}} \leq FN \leq F_{N_{\text{max}}}$ and $|\text{RoCoF}| \leq \text{RoCoF}_{\text{max}}$ following a sudden active power disturbance, the transient frequency stability of the power system is guaranteed; otherwise, the transient frequency is considered unstable [2,20, 21].

### 2.1 Deep learning based TFSA with a combined estimation-correction learning framework

Rather than simulating the power system post-fault frequency dynamics using the inefficient offline time-domain method, the main aim of this study is to propose a novel data-driven approach based on deep learning, which is surprisingly fast and accurate for online use. The core of this method is the introduced combined estimation-correction learning framework (illustrated in Figure 2). In the estimation phase, the DNN is utilised to approximate the power system frequency dynamics and obtain the preliminary values of the multidimensional frequency metrics ($F_{N}'$, $TFN'$, $\text{RoCoF}'$, and $QssF'$) for each disturbance under consideration. The multidimensional frequency metrics produced by the DNN-based estimator are then fed into the correction phase, in which the SELM is used to further reduce the prediction errors and obtain the corrected multidimensional frequency metrics ($F_{N}'', TFN''$, $\text{RoCoF}''$, and $QssF''$).

The DNN-based estimator transforms the sample features from the original space into a new feature space that can effectively implement complex and highly non-linear mapping [14]. Representation learning is the most important component of DNN, which allows it to be directly fed with raw data and realise automatic extraction and selection [22] of TFSA features. The reasons for the necessity of integrating an SELM-based corrector with the DNN-based estimator to build a combined estimation-correction learning framework are illustrated as follows:

- Conventional neural networks adopt an error back propagation algorithm to optimize the network parameters, which may suffer from gradient dispersion and readily fall into a local optimum. This drawback is more apparent when the neural network architecture goes deeper because there are more parameters to be optimised [22]. Although some optimisation methods, such as the stochastic gradient descent (SGD), can be used to obtain the network parameters [17], there may be certain non-ignorable errors between the estimated frequency metrics and true values because of the inherent defects of the back-propagation algorithm. Consequently, to further improve the prediction accuracy and generalisation performance, it is necessary to correct the multidimensional frequency metrics produced by the DNN-based estimator.

- Because the dimension of the corrector’s input data (the multiple frequency metrics output by the estimator) is relatively small, the relationship between the estimated frequency metrics and true values can be directly established without reducing the dimension of the corrector’s input data. SELM, an advanced deep learning computing paradigm with an extremely high training speed and high learning accuracy [26,27], can serve as a good candidate corrector for the combined estimation-correction learning framework. Because the SELM algorithm does not need any iterative solution during offline training, integrating the SELM-based corrector into the learning framework only slightly increases the overall solution time, while considerably improving the prediction accuracy.

The proposed data-driven TFSA via deep learning mainly consists of three parts: database formation, combined
estimation-correction deep learning framework, and online evaluation.

2.2 Database formation

The amount of active power imbalance, speed regulation gains, inertia constants, unit commitment statuses, primary reserves, governor response time, turbine parameters, and load damping coefficient are all considered input data. These data can be easily accessed in a practical power system. The historical frequency metrics of disturbance events (load fluctuation and unit outage) monitored by the power system control centre are used as test data to comprehensively evaluate the transient frequency stability.

When sufficient frequency monitoring data and historical disturbance event data are available in the control centre of a power system, the historical dataset can be used directly as training data. In the absence of a historical dataset, offline simulation can be conducted to obtain a certain amount of input–output samples. Further, the historical and simulation-based dataset can be simultaneously considered to create more samples. The input and output data are normalised and then divided into two parts: training and test samples. The training samples are used to obtain the parameters of the combined estimation-correction deep learning network, whereas the test samples are used to assess the performance of the trained network.

2.3 DNN-based estimation

The neural network layer of the DNN-based estimator is divided into three parts: input layer (first layer), hidden layers (middle layers), and output layer (last layer). The offline training of the DNN includes forward and backward propagation phases [22,23].

The weight matrix $W$ and bias matrix $b$ of the hidden layers are randomly generated in the forward propagation process. The feature vector $a_j$ of the $l$-th hidden layer is determined by:

$$a_j = s_j(W_ja_{j-1} + b_j)$$  \hspace{1cm} (2)

where $W_j$ and $b_j$ are the weight and bias matrices of the $l$-th hidden layer, respectively, and $s_j$ is the activation function, which
is represented by:

\[ s_j(W, a_{j-1} + b) = \frac{1}{1 + \exp(-W_j a_{j-1} + b)} \]  

(3)

After obtaining the feature vector \( a_n \) of the output layer, the errors between the predicted and true values can be calculated by the loss function, which is given as follows:

\[ J(W, b, x, y) = \frac{1}{2} \| \hat{y} - y \| \]  

(4)

where \( \hat{y} \) is the vector of the predicted value of the multidimensional frequency metrics \( y \) is the vector of the true values, and \( \| \hat{y} - y \| \) is the \( L_2 \) norm of \( \hat{y} - y \). The parameters, weight and bias \( (W, b) \) are then obtained using the SGD method. The traditional SGD method with a fixed learning rate usually causes slowness or non-convergence. To reduce the local optimum of DNN and improve the network prediction accuracy, this study introduces some improvements into the SGD; the involved techniques include the mini-batch technique \[24\] and the adaptive learning rate gradient descent method \[25\]. Instead of choosing a fixed learning rate hyper parameter, leveraging the adaptive learning rate gradient descent method, the learning rate can be adjusted in response to the performance of the model on the training dataset, that is, the improved SGD can dynamically adjust the updating speed of the weights and bias by \( (9) - (12) \). Therefore, the risk of converging to a local optimum can be reduced, and the algorithm converges much faster than simple back-propagation with a poorly chosen fixed learning rate. The steps of the DNN-based estimator are described in Algorithm 1.

**Algorithm 1** Steps of the DNN-based estimator

(a) Input the training samples \( [(x_1, y_1), (x_2, y_2), \ldots, (x_K, y_K)] \). Set the number of mini-batches \( M \), layers \( s \), neurons in each hidden layer, learning rate \( \alpha \), maximum number of iterations \( e_{\text{max}} \), decay rate \( \rho \), and momentum \( \beta \). Initialize the weight matrix \( W \) and bias matrix \( b \). Set the initial speed \( V_x = 0, V_b = 0 \), and cumulative variables \( r_1 = 0, r_2 = 0 \).

(b) Randomly divide the training sample dataset \( K \) into \( M \) mini-batches, then calculate the feature vector of each layer using the forward propagation method.

(c) Update the network parameters based on the mini-batch technique and adaptive learning rate gradient descent method:

\[ \text{for} \ e = 1 \text{ to } e_{\text{max}} \]

\[ \text{for} \ m = 1 \text{ to } M \]

\[ \text{for} \ i = 1 \text{ to } K/M \]

Calculate the partial gradient \( \delta_{i,a} \) of the output layer:

\[ \delta_{i,a} = \frac{\partial J(W, b, x, y)}{\partial \sigma_i} = (a_i - y) \odot f'_i(z_i) \]  

(5)

where \( \odot \) represents the Hadamard product.

\[ \text{for} \ l = a - 1 \text{ to } 2 \]

Calculate the partial guidance \( \delta_{i,j} \) layer-wise:

\[ \delta_{i,j} = \delta_{i,j} \frac{\partial \sigma_{i,j}}{\partial a_{i,j}} = (W_{i+1})^T \delta_{i+1} \odot f'_i(z_i) \]  

(6)

Calculate the cumulative gradient:

\[ r_1 = \rho r_1 + (1 - \rho) \sum_{i=1}^{N/M} \delta_{i,j} (a_{i,j-1})^T \odot \sum_{i=1}^{N/M} \delta_{i,j} (a_{i,j-1})^T \]  

(7)

\[ r_2 = \rho r_2 + (1 - \rho) \sum_{i=1}^{N/M} \delta_{i,j} \odot \sum_{i=1}^{N/M} \delta_{i,j} \]  

(8)

End

\[ \text{for} \ i = 1 \text{ to } a - 1 \]

Update weight and bias matrices based on the gradient descent with adaptive learning rate:

\[ W_x' = \beta W_x - \frac{r_1}{\sqrt{r_2}} \odot \sum_{i=1}^{N/M} \delta_{i,j} (a_{i,j-1})^T \]  

(9)

\[ b_x' = \beta b_x - \frac{r_2}{\sqrt{r_2}} \odot \sum_{i=1}^{N/M} \delta_{i,j} \]  

(10)

\[ W' = W + W_x' \]  

\[ b' = b + b_x' \]  

(11)

End

End

End

(d) Output the weight matrix \( W \) and bias matrix \( b \) of the hidden layers and formulate the well trained DNN-based estimator.

Subsequently, compute the estimated values of the multidimensional frequency metrics (\( F_{n'} \), \( T_{n'} \), \( RoCoF_{n'} \), and \( QoS_{n'} \)).
2.4 | SELM-based corrector

To maximally ensure the correctness of the TFSA results, an SELM-based corrector is formulated to further improve the accuracy of the estimated multidimensional frequency metrics produced by the DNN-based predictor.

Motivated by the idea of deep learning networks, the SELM is formulated as a stacked ELM with multilayer neural networks.

Unlike other algorithms, the SELM algorithm does not require an iterative solution; therefore, it can be trained much faster and requires less memory space [26]. The structure of the SELM is shown in Figure 3. To prevent the learning efficiency reduction of the representative feature data by the randomly generated weights and bias [26] and improve the model generalisation, the ELM Auto-Encoder (ELM-AE) and regularisation coefficient techniques [27] are applied to obtain the network

---

**ALGORITHM 2** Steps of the SELM-based corrector

(a) The output data (FN', TFN', RoCoF', and QssF') produced by the DNN-based estimator are fed into the corrector as input data.

(b) Set the number of hidden layers and number of neurons in each hidden layer of the SELM.

(c) Solve the weight matrix of the first hidden layer according to the ELM-AE and regularisation coefficient techniques:
   i) Generate the randomly orthogonal weight matrix $W$ and bias matrix $b$ of the encoder.

   $W^T W = I, b^T b = 1 \quad (13)$

   ii) Calculate the feature vector $h^1$ of the first hidden layer.

   iii) Solve the reconstruction matrix $W^1$. The feature vector $h^1$ is reconstructed into the original input data using the decoder.

   $X = h^1 W^1 \quad (14)$

To avoid local optimum, a regularisation term is added to the objective function of SLEM:

$$\min_W \left\{ \frac{1}{2} \|W^1 X - X\|^2 + \frac{1}{2} \|W^1\|^2 \right\} \quad (15)$$

Aided by the regularisation coefficient $\lambda$, the weights will decay to significantly mitigate the local optimum problem.

The output weight matrix $W^1$ is then calculated by:

$$W^1 = \left( h^1 h^1 + \frac{1}{2} \right)^{-1} h^1 \quad (16)$$

Finally, the transition rank matrix $W^1$ of the reconstruction matrix is used as the weight matrix of the input layer and first hidden layer of the original network structure.

(d) Solve the weight matrix of the remaining hidden layers. Leveraging the ELM-AE, the feature vector $h^p$ of the $p$-th hidden layer is taken as input data to obtain the input weight matrix $W^{p+1}$ and feature vector $h^{p+1}$ of the $p+1$-th hidden layer.

(e) Establish the trained SELM-based corrector using the network parameters obtained in the previous steps and obtain the corrected multidimensional frequency metrics (FN'', TFN'', RoCoF'', and QssF'').
204

WEN ET AL.

FIGURE 4 Construction of data-driven model

2.5 Construction of data-driven model

The problem of predicting the post-fault multi-dimensional frequency metrics based on the combined estimation-correction learning framework can be described as a non-linear mapping relationship between the input feature set and output frequency metrics. The constructed input feature set and frequency metrics are treated as the input \(X\) and output \(Y\) of the proposed data-driven approach, respectively, as shown in (17) and (18):

\[
X = \begin{bmatrix}
    x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,m} \\
    x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,m} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    x_{n,1} & x_{n,2} & x_{n,3} & \cdots & x_{n,m}
\end{bmatrix}
\]

\[
y = \begin{bmatrix}
    FN_{1,1} & QssF_{1,2} & RoCoF_{1,3} & TFN_{1,4} \\
    FN_{2,1} & QssF_{2,2} & RoCoF_{2,3} & TFN_{2,4} \\
    \vdots & \vdots & \vdots & \vdots \\
    FN_{n,1} & QssF_{n,2} & RoCoF_{n,3} & TFN_{n,4}
\end{bmatrix}
\]

where \(n\) is the number of samples and \(m\) is the number of input features.

The DNN-based estimator and SELM-based corrector can be trained through the input feature set and frequency metrics. After the offline training process, the combined estimation-correction learning framework establishes a non-linear mapping relationship between the input feature variables and frequency metrics. Therefore, the well-trained learning framework can be used directly to predict the frequency stability of the system after a new disturbance. The concept is shown in Figure 4.

2.6 Performance evaluation

Three metrics are used to evaluate the accuracy of the multi-dimensional frequency predictions obtained by the proposed combined estimation-correction deep learning framework: the mean absolute error (MAE), and assessment accuracy (AC) [8,16].

\[
MAE = \frac{1}{S} \sum_{s=1}^{S} | y' - y_t |
\]

\[
MAPE = \frac{1}{S} \sum_{s=1}^{S} \left| \frac{y' - y_t}{y_t - y_{base}} \right|
\]

where \(y'\) and \(y_t\) denote the actual and predicted values of the \(s\)-th sample, respectively. \(y_{base}\) is the reference value; the reference values of FN, TFN, RoCoF, and QssF are 50 Hz, 0 s, 0 Hz/s, and 50 Hz, respectively. AC is used as an overall evaluation index for TFSA, which is expressed as:

\[
AC = \frac{TP + TN}{TP + FP + TN + FN}
\]

where TP and FP are the number of stable samples that are correctly and wrongly evaluated, respectively; TN and FN are the number of unstable samples that are correctly and wrongly evaluated, respectively.

2.7 Online transient frequency stability assessment

The online TFSA can be utilised after the completion of the combined estimation-correction learning framework offline training. The steps of the online TFSA are described as follows:

Step 1: Input the operating and contingency lists, using the dataset accessed from real-time monitoring of the power system.

Step 2: Data normalisation.

Step 3: Input data (i.e., operating scenarios, contingency lists) into the trained combined estimation-correction learning framework, and obtain output data (\(FN''\), \(TFN''\), \(RoCoF''\), and \(QssF''\)). Subsequently, use the resulting multidimensional frequency metrics of each disturbance incident to comprehensively evaluate the transient frequency stability.
Step 4: To further improve the robustness and generalisation ability of the combined estimation-correction learning framework, the input and output data obtained from the online evaluation can be fed back to the offline training samples to enrich the historical dataset.

3 | CASE STUDY

The proposed data-driven TFSA is tested on a modified IEEE RTS-79 system (as shown in Figure 5). This system has 32 generators, 38 branches, and 23 buses. The base case total system load is set at 2850 MW. Four wind farms are added to buses 107, 114, 121, and 123. The penetration of the wind generation is gradually increased from 0 to 50%, reducing the inertia of the system from 8250 to 4200 MWs. The thresholds of the UFLS, OFGT, and RoCoF protection relays are set at 49 Hz, 51 Hz, and 1 Hz/s, respectively. All the experiments are performed on Matlab R2015a running on an Intel Core i5, 4 GB RAM personal computer.

3.1 | Dataset preparation

The sample datasets are obtained by offline simulation performed on the MATLAB-SIMULINK platform. To evaluate the effect of the amount of active power disturbance on the post-fault multidimensional frequency metrics, the power imbalance of the base case total load is gradually varied from 0 to ±40% in steps of 10 MW. Further, the impacts of the system inertia and spinning reserve levels are considered when conducting the simulation. Because the stochastic wind generation load fluctuation and contingencies with different inertia and spinning reserve levels are considered exhaustively, the dataset is enormous. Around 37,392 samples are generated from the offline simulation, among which 21,473 are frequency stable.
and 15,919 are frequency unstable. We randomly selected 30,000 samples for the offline training of the proposed combined estimation-correction learning framework and the remaining 7392 samples were used for performance testing. Based on the input–output data selection scheme presented in Section 2.5, the dimension of the input data is 261 and the dimension of the output data is 4.

### 3.2 Parameter selection

The number of hidden layers and neurons in each hidden layer both have impacts on the assessment accuracy and offline training time. The number of hidden layer neurons are set as applied in [22–24]. First, the optimal number of neurons in the first hidden layer is fixed. The second hidden layer is then added, and the optimal number of neurons in this new layer is determined. The above steps are iterated until the accuracy shows no further improvement. For the DNN-based estimator, the optimal number of hidden layers is 5, and the optimal number of neurons in each layer is 150, 120, 90, 30, and 10, respectively. For the SELM-based corrector, the optimal number of hidden layers is 3, and the optimal number of neurons for each layer is 700.

### 3.3 Analysis and comparison of test results

The performance of the proposed data-driven TFSA with the combined estimation-correction deep learning framework is compared with that of four learning algorithms, that is, BP, ELM, DNN, and SDAE [13]. To ensure the validity of the comparison, the results from each machine-learning-based algorithm should be the best. This requires optimal parameter tuning of the machine learning models. To address this, we first determine the appropriate range of parameters based on multiple experiments and human experience. Subsequently, the parameter combinations of the determined range are traversed using the enumeration method. Consequently, we finally obtain the approximate optimal parameters of BP, ELM, DNN, and SDAE.

Table 1 compares the offline training/online evaluation times and TFSA AC from the time-domain simulation and five neural network algorithms. For the time-domain simulation method, 7392 dynamic simulation calculations must be performed to assess the frequency stability of the 7392 test samples, which requires a total of 806.49 s. Considering the multiple uncertainties in a low-inertia power system, the time domain simulation cannot meet the requirements of online TFSA for operators. The total online evaluation time of the proposed data-driven TFSA approach is 0.36 s, which is only 0.045% of that required by the time domain simulation. In terms of the AC, the proposed approach has the highest accuracy (99.8%) in comparison with the other four neural network algorithms, which is very close to the time domain simulation results. Therefore, the proposed approach can realize fast and accurate TFSA and is suitable for on-line application in a power system control centre. If the SELM-based corrector is not deployed, that is, only the DNN-based estimator is used for TFSA, the AC produced by the DNN decreases to 97.87%, which demonstrates the significance of the SELM-based corrector in further improving the accuracy of the predictions. The accuracy of the TFSA obtained by the SDAE is slightly higher than that of the DNN. However, because the SDAE initializes the network parameters using pre-training, its offline training time (1175 s) is the longest among all the methods. The ACs of the BP and ELM are 58.64% and 64.05%, respectively. This low level of precision makes it difficult for them to meet the accuracy requirements for online application.

Figures 6 and 7 compare the MAE and MAPE of the four frequency prediction metrics (i.e. FN, QssF, RoCoF, and TFN), respectively, used to evaluate the proposed approach in comparison to the four other neural network methods. As can be observed in the figures, because deep learning can effectively characterize complex functions, the MAE and MAPE based on the deep learning methods (the proposed approach, DNN, and SDAE) are significantly lower than those based on the neural networks (BP and ELM). The MAE and MAPE, based on the proposed approach, are almost negligible compared to the other methods.

Figure 8 shows the detailed error distributions of the predicted multidimensional frequency metrics obtained by the proposed data-driven approach with a combined estimation-correction deep learning framework. The percentages of samples with FN, QssF, RoCoF, and TFN within a small error range of 0.01 Hz, 0.03 Hz, 0.01 Hz/s, 0.03 s are 99.36%, 98.43%, 99.92%, and 95.98%, respectively. The MAE values of the FN, QssF, RoCoF, and TFN are only 0.0045 Hz, 0.0060 Hz, 0.00148 Hz/s, and 0.0082 s. Meanwhile, the MAPE values of the FN, QssF, RoCoF, and TFN are only 0.61%, 1.12%, 0.364%, and 1.63%, respectively. Compared with the time domain simulation, the frequency metrics obtained by the proposed approach have extremely high precision.

To verify the generalisation capability of the proposed approach, 1000 sets of new test data are regenerated by changing the system operational mode (unit statuses, penetration of wind generation, reserves, and power imbalance amount). The test results compared with the other four neural-network-based methods are shown in Table 2. The MAE of the predicted
The frequency and MAPE of the proposed approach are the lowest, whereas the AC is the highest (98.1%). Therefore, the proposed approach has better generalisation capability and adaptability under different power system operation scenarios.

### CONCLUSION

To realize the online TFSA of power systems, this study proposed a novel data-driven approach based on deep learning. A combined estimation-correction learning framework with
a DNN-based estimator and SELM-based corrector was formulated. The DNN-based estimator effectively achieved dimensionality reduction, while automatically obtaining representative and significant feature information. The SELM-based corrector adjusted the estimated multidimensional frequency outputs produced by the DNN-based estimator to further improve the accuracy of the TFSA. The advantages of the proposed approach are summarised as follows:

(i) Compared with the offline time domain simulation, the proposed approach meets the requirements of the rapid evaluation of multidimensional frequency indicators (i.e. FN, QssF, RoCoF, and TFN) under multiple uncertainties, which significantly saves on the computational time. It is beneficial for operators to use the proposed approach to predict power system dynamic frequency behaviour.

(ii) Compared with the traditional and more recent algorithms, that is, BP, ELM, DNN, SDAE, the proposed method achieves higher forecasting accuracy and better generalisation.

ACKNOWLEDGEMENTS
This work was supported in part by the National Natural Science Foundation of China (52077066) and in part by the Hunan Provincial Natural Science Foundation for Excellent Young Scholars (2020JJ3011).
REFERENCES

1. Tuffner, F.K. et al.: Modeling load dynamics to support resiliency-based operations in low-inertia microgrids. IEEE Trans. Smart Grid 10(3), 2726–2737 (2019)
2. Wen, Y., Chung, C., Ye, X.: Enhancing frequency stability of asynchronous grids interconnected with HVDC links. IEEE Trans. Power Syst. 33(2), 1800–1810 (2018)
3. Wood, A., Wollenberg, B., Sheble, G.: Power Generation, Operation and Control, 3rd edn. John Wiley & Sons Press, Hoboken, NJ (2013)
4. Baipai, P., Dash, V.: Hybrid renewable energy systems for power generation in stand-alone applications: A review. Renew. Sustain. Energy Rev. 16(5), 2926–2939 (2012)
5. Wu, Y.K.: Frequency stability for an island power system: Developing an intelligent preventive-corrective control mechanism for an offshore location. IEEE Industry Appl. Mag. 23(2), 74–87 (2017)
6. Djukanovic, M., Popovic, D., Sobajic, D.: Prediction of power system frequency response after generator outages using neural nets. IEEE Proceedings C-Generation, Transmission and Distribution 140(5), 389–398 (1993)
7. Chang, R., Lu, C., Hsiao, T.: Prediction of frequency response after generator outage using regression tree. IEEE Trans. Power Syst. 20(4), 2146–2147 (2005)
8. Bo, Q., Wang, X., Liu, K.: Minimum frequency prediction of power system after disturbance based on the v-support vector regression. International Conference on Power System Technology, pp. 614–619. IEEE, Chengdu, China (2014)
9. Hong, Y., Chen, P.: Genetic-based underfrequency load shedding in a stand-alone power system considering fuzzy loads. IEEE Trans. Power Deliv. 27(1), 87–95 (2012)
10. Dai, Y., Xu, Y., Dong, Z.: Real-time prediction of event-driven load shedding for frequency stability enhancement of power systems. IET Gener. Transm. Distrib. 6(9), 914–921 (2012)
11. Zhang, C., Lin, P., Qin, A.: Multi objective deep belief networks ensemble for remaining useful life estimation in prognostics. IEEE Trans. Neural Netw. Learn. Syst. 28(10), 2306–2318 (2016)
12. Wang, S., Chen, H.: A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. Appl. Energy 235(1), 1126–1140 (2018)
13. Vincent, P., Larochelle, H., Lajoie, I.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. 11(12), 3371–3408 (2010)
14. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural Netw. 61(8), 85–117 (2015)
15. Wang, H., Li, G., Wang, G.: Deep learning based ensemble approach for probabilistic wind power forecasting. Appl. Energy 188(15), 56–70 (2017)
16. Shi, H., Xu, M., Li, R.: Deep learning for household load forecasting—a novel pooling deep RNN. IEEE Trans. Smart Grid 9(5), 5271–5280 (2018)
17. Al-Masri, A., Adr, M., Hizam, H.: A novel implementation for generator rotor angle stability prediction using an adaptive artificial neural network application for dynamic security assessment. IEEE Trans. Power Syst. 28(3), 2516–2525 (2013)
18. Bangalore, P., Tijm, B.: A neural network approach for early fault detection of gearbox bearings. IEEE Trans. Smart Grid 6(2), 980–987 (2017)
19. Wen, Y. et al.: Frequency dynamics constrained unit commitment with battery energy storage. IEEE Trans. Power Syst. 31(6), 5115–5125 (2016)
20. Ketabi, A., Fini, H.: An underfrequency load shedding scheme for hybrid and multiarea power systems. IEEE Trans. Smart Grid 6(1), 82–91 (2015)
21. Golpira, H., Seifi, H., Roman, A.: Maximum penetration level of microgrids in large-scale power system: Frequency stability viewpoint. IEEE Trans. Power Syst. 31(6), 5163–5171 (2016)
22. Gong, M., Liu, J., Li, H.: A multiobjective sparse feature learning model for deep neural networks. IEEE Trans. Neural Netw. Learn. Syst. 26(12), 3263–3277 (2015)
23. Shao, L., Wu, D., Li, X.: Learning deep and wide: A spectral method for learning deep networks. IEEE Trans. Neural Netw. Learn. Syst. 25(12), 2303–2308 (2014)
24. Konečný, J., Liu, J., Richtárik, P.: Mini-batch semi-stochastic gradient descent in the proximal setting. IEEE J. Sel. Top. Signal Process. 10(2), 242–255 (2016)
25. Zhang, R., Gong, W., Grzeda, V.: An adaptive learning rate method for improving adaptability of background models. IEEE Signal Process. Lett. 20(12), 1266–1269 (2013)
26. Huang, G., Wang, D., Lan, Y.: Extreme learning machines: a survey. Int. J. Mach Learn. Cybern. 2(2), 107–122 (2011)
27. Kasun, L., Yang, Y., Huang, G.: Dimension reduction with extreme learning machine. IEEE Trans. Image Process. 25(8), 3906–3918 (2016)

How to cite this article: Wen Y, Zhao R, Huang M, Guo C. Data-driven transient frequency stability assessment: A deep learning method with combined estimation-correction framework. Energy Convers Econ. 2020;1:198–209. https://doi.org/10.1049/enc2.12015