A Fault Aware Broad Learning System for Concurrent Network Failure Situations

ADEGOKE, Muideen; WONG, Hiu Tung; LEUNG, Chi Sing

Published in:
IEEE Access

Published: 01/01/2021

Document Version:
Final Published version, also known as Publisher’s PDF, Publisher’s Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1109/ACCESS.2021.3066217

Publication details:
https://doi.org/10.1109/ACCESS.2021.3066217

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

Download date: 24/09/2023
A Fault Aware Broad Learning System for Concurrent Network Failure Situations

MUIDEEN ADEGOKE, HIU TUNG WONG\textsuperscript{\rm ©}, AND CHI SING LEUNG\textsuperscript{\rm ©}, (Senior Member, IEEE)
Department of Electrical Engineering, City University of Hong Kong, Hong Kong
Corresponding author: Chi Sing Leung (eeleungc@cityu.edu.hk)

This work was supported research grant from City University of Hong Kong under Grant 7005223 and Grant 7610431.

ABSTRACT The broad learning system (BLS) framework gives an efficient solution for training flat-structured feedforward networks and flat structured deep neural networks. However, the classical BLS model and other variants focus on the faultless situation only, where enhancement nodes, feature mapped nodes, and output weights of a BLS network are assumed to be realized in a perfect condition. When a trained BLS network suffers from coexistence of weight/node failures, the trained network has a greatly degradation in its performance if a countermeasure is not taken. In order to reduce the effect of weight/node failures on the BLS network’s performance, this paper proposes an objective function for enhancing the fault aware performance of BLS networks. The objective function contains a fault aware regularizer term which handles the weight/node failures. A learning algorithm is then derived based on the objective function. The simulation results show that the performance of the proposed fault aware BLS (FABLS) algorithm is superior to the classical BLS and two state-of-the-arts BLS algorithms, namely correntropy criterion BLS (CBLS) and weighted BLS (WBLS).

INDEX TERMS Fault tolerance, broad learning system, regression, multiplicative noise, open fault, incremental learning.

I. INTRODUCTION

Without doubt, human brain is capable to handle fault and noise situations [1]. For instance, human being is able to recognize a partially occluded object without much difficulty. Moreover, human brain can still operate with a certain ability when a meager amount of synapses or neurons are malfunctioned. As the idea of artificial neural networks (ANNs) is based on human brain, it was presumed that a trained ANN should have an inherent ability against weight/node failures. In the inception of ANN researches, a common misconception was that ANNs have intrinsic tolerance against any form of failures, such as weight noise, node noise, and open fault. Meanwhile, if ANNs suffer from weight failures or node failures, without any fault aware measures in place, then the performance of such ANNs is greatly degraded [2], [3].

In the realization of a trained ANN in electronic systems, weight/node failures are not avoidable [4], [5]. For instance, when digital technology is employed to represent the weights of a trained ANN, there are precision errors and the errors can be modelled as multiplicative noise. Also, when analog technology is used, hardware components do have precision errors and thermal noise do exist. Furthermore, damages such as open weight faults or open node faults may be introduced when the communication link between two nodes are broken [6]. Besides, when realizing a ANN using the very large scale integration (VLSI) at nano-scale level, transient noise or failures are unavoidable during the operation [7], [8].

One of basic requirements of many electronic systems is dependability. Many safety critical systems would be using complex ANNs, such as deep neural network (DNN). It is expected that such systems must have high dependability. Hence, it is vital or crucial to embed good fault aware techniques in such systems, such as build-in self test and error detection/correction coding [9].

To enhance the behaviour of a well trained ANN during the failure situation, it is paramount to figure out how fault affects the behavior or performance of a trained ANN in an analytical way. There are related studies on fault resistant ability of ANNs. Recently, in [10], empirical results of a fault tolerant ability of a small convolutional neural network (CNN) are presented. Furthermore, a survey paper reported the effects of those imperfect conditions for various neural network models, including feedforward networks and...
radial basis function (RBF) networks [11], [12]. For RBF networks, fault aware ability and some fault aware training algorithms were extensively investigated in [3], [4].

The broad learning system (BLS) [13], [14] concept is very efficient for constructing flat structured ANN, and BLS networks are universal approximators [15]. In addition, due to its good properties, the BLS concept has caught many attentions [16]–[20]. Although many BLS results were reported, commonly available BLS algorithms focus on the faultless situation. To the best of our knowledge, there are not many literatures related to fault tolerant BLS networks.

This paper studies the effect of coexistence of weight/node failures on BLS [13]. It first examines the behaviour of the original BLS model under the coexistence of several fault/noise, including open weight fault on output weights, multiplicative noise on the nodes, and multiplicative noise on the output weights. To compensate the degradation effect from the aforementioned noise/fault, an objective function for enhancing the fault aware performance of BLS is developed. The objective function contains a fault aware regularizer term. The learning algorithm for minimizing the fault aware objective function is proposed. The proposed algorithm is called fault aware BLS (FABLS) algorithm. Several real life datasets from the UCI repository [21] are used to verify the effectiveness of the proposed FABLS algorithm. From the simulation results, our proposed FABLS is superior to the comparison BLS algorithms.

In summary, our major contributions are as follows.

- This paper investigates the effect of the coexistence of several fault/noise on the performance of BLS network.
- A fault aware objective function is proposed. The objective function contains a fault aware regularizer term which mitigates the effects of weight/node failures.
- Based on the developed fault aware objective function, the fault aware BLS (FABLS) algorithm is developed.
- From the simulation results, our proposed FABLS is superior to the comparison BLS algorithms.

The paper is presented as follows. Section II discusses related works, Section III presents the background on the traditional broad learning system (BLS) network and model. Section IV provides the concept of coexistence of a weight failure situation and the weight failure models. The fault-aware objective function and the learning algorithm are developed in Section V. Section VI uses several real life datasets to demonstrate the superiority of the proposed FABLS algorithm. Finally, Section VII provides conclusion remarks. Table 1 lists some key notations used in this paper.

### II. RELATED WORKS

**A. FAULT TOLERANT IN ANNS**

One fascinating feature or characteristic of ANNs is fault tolerance. However, if proper learning steps are not adopted, a trained ANN usually has poor fault tolerant ability. In the last two decades, various results on fault-tolerant ability of ANNs were reported. For instance, the common node fault situations are stuck-at-zero and stuck-at-one [6], [22]–[24]. For stuck-at-zero, the output of a faulty node is stuck at zero. For stuck-at-one, the output of a faulty node is stuck at one. Another kind of “stuck-at” fault is “stuck-at-any-values”, in which the output of a faulty node is stuck at a value in between the output range of the activation function. This stuck-at-any-values situation was analyzed in [25]. Apart from node fault/failure, connection weights in a ANN may also have fault or failure [26]–[28]. In the open weight fault, the connection weight between two nodes is disconnected. In the weight noise case, due to the technology limitation, there are some precision errors/noise in the realization. The common used model to describe the behaviour of the weight noise is the multiplicative weight noise model [2], [3], [26]. There are many methods to handle node/weight failures. Some approaches inject random failure during training [22], [24], [25]. The demerit of this approach is that a large number of training cycles is required to achieve some certain performance. Otherwise, the statistical information of the fault cannot be expressed. Furthermore, another common approach for protecting fault/failure is replication approach

<table>
<thead>
<tr>
<th>Sym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>Input vector to the BLS network $x \in \mathbb{R}^{m}$</td>
</tr>
<tr>
<td>$y$</td>
<td>Augmented input vector: $y = [x^T, \mathbf{1}]^T$</td>
</tr>
<tr>
<td>$n$</td>
<td>Num. of groups of feature mapped nodes.</td>
</tr>
<tr>
<td>$f_i$</td>
<td>Num. of nodes in the $i$-th group of feature mapped nodes.</td>
</tr>
<tr>
<td>$f$</td>
<td>Num. of feature mapped nodes $\sum_{i=1}^{n} f_i$.</td>
</tr>
<tr>
<td>$\Psi_i$</td>
<td>Weight matrix of the $i$-th group of feature mapped nodes.</td>
</tr>
<tr>
<td>$g_i$</td>
<td>Output vector of the $i$-th group of feature mapped nodes: $g_i = \Psi_i x$.</td>
</tr>
<tr>
<td>$\mathbf{g}$</td>
<td>The collection of all feature mapped node outputs: $\mathbf{g} = [\mathbf{g}_1^T, ..., \mathbf{g}_n^T]^T$.</td>
</tr>
<tr>
<td>$q$</td>
<td>Augmented vector of $g$: $q = [q^T, \mathbf{1}]^T$.</td>
</tr>
<tr>
<td>$m$</td>
<td>Num. of groups of enhancement nodes.</td>
</tr>
<tr>
<td>$e_j$</td>
<td>Num. of nodes in the $j$-th group of enhancement nodes.</td>
</tr>
<tr>
<td>$\mathbf{e}$</td>
<td>Num. of enhancement nodes $\sum_{j=1}^{m} e_j$.</td>
</tr>
<tr>
<td>$\mathbf{W}_j$</td>
<td>Weight matrix of the $j$-th group of enhancement nodes.</td>
</tr>
<tr>
<td>$h_j$</td>
<td>Output vector of the $i$-th group of enhancement nodes: $h_j = \xi(\mathbf{W}_j q)$.</td>
</tr>
<tr>
<td>$\xi(.)$</td>
<td>Activation function of enhancement nodes.</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Outputs of all the enhancement nodes: $\eta = [\eta_1, ..., \eta_m]^T$.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Output of the BLS network: $\alpha = [\eta^T \beta^T]$.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Output weight vector: $\beta = [\beta_1, ..., \beta_f]^T$.</td>
</tr>
<tr>
<td>$\beta_k$</td>
<td>Duplication labelling for $\beta_k$.</td>
</tr>
<tr>
<td>$\beta_{h,j}$</td>
<td>Duplication labelling for $\beta_k$.</td>
</tr>
<tr>
<td>$z_{x_k}$</td>
<td>Training input: $z_{x_k} = [z_{x_k,1}, ..., z_{x_k,p}]$ and $\mathbf{x}$ and $\mathbf{y}$.</td>
</tr>
<tr>
<td>$\mathbf{y}_k$</td>
<td>Training output: $k = 1, ..., N$.</td>
</tr>
<tr>
<td>$\mathbf{X}$</td>
<td>Collection of all training outputs.</td>
</tr>
<tr>
<td>$\mathbf{Z}$</td>
<td>Training input data matrix.</td>
</tr>
<tr>
<td>$\mathbf{Z}_i$</td>
<td>The $i$-th training data feature matrix: $\mathbf{Z}_i = \mathbf{X} \Psi_i$.</td>
</tr>
<tr>
<td>$\mathbf{H}_j$</td>
<td>The $j$-th training data enhancement matrix.</td>
</tr>
<tr>
<td>$\mathbf{H} = [\mathbf{H}_1 \cdots \mathbf{H}_m]$</td>
<td>Collection of all training data enhancement matrices.</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Open fault rate of output weights is opened.</td>
</tr>
<tr>
<td>$\sigma_d^2$</td>
<td>Noise variance of output weights.</td>
</tr>
<tr>
<td>$\sigma_f^2$</td>
<td>Noise variance of feature mapped nodes.</td>
</tr>
<tr>
<td>$\sigma_w^2$</td>
<td>Noise variance of input weights of enhancement nodes.</td>
</tr>
</tbody>
</table>
These nodes are arranged in a broad sense. In a BLS network, the connection weights between the input to the enhancement node layer and another layer, called enhancement node layer. The enhancement nodes receive the inputs from the feature nodes. In the original BLS learning, the connection weights between the input to the enhancement node are generated in a random manner.

As shown in Fig. 1, the output layer collects the information from the feature node layer and enhancement node layer. In other words, the outputs of feature nodes and enhancement nodes are concatenated in a wide manner and are fed to the output layer. The output weights could be obtained by a least-mean square technique based on Moore-Penrose’s generalized inverse [55], [56]. The overall training computational cost in BLS is not excessive and is much lower than that of other classical training algorithms, such as meta-heuristic techniques (differential evolution [57], particle swarm optimization [58], and gradient descent methods [59].

As the BLS framework provides efficient ways for constructing a flat structured ANN, and a BLS network owns the universal approximation ability [13], [14], the BLS framework has attracted increasing attentions. Many variants of BLS models and algorithm were proposed to enhance its performance [16], [60], [61]. Besides, many applications of BLS were reported in the last several years.

In [16], a hybrid model of BLS and fuzzy concept was proposed, in which the feature nodes are replaced with some fuzzy subsystems. In [18], the BLS system is applied to fault diagnosis in rotor systems. In [62], a BLS model was proposed for object detection using the event data obtained from event cameras. In this model, the incremental learning concept was also used for constructing the resultant BLS network. In [63], the BLS concept was used for processing hyperspectral images. Apartment from applications, there are modifications on its learning algorithm [19], [20]. In [19], the weighted BLS (WBLS) algorithm was proposed for compensating the effect of noise and outliers. In [20], the correntropy criterion based BLS (CBLS) was proposed for handling outlier data based on the maximum correntropy criterion (MCC). Although the BLS concept is a quite promising framework, the original BLS and other variants focus on the faultless situation only, where a BLS network is assumed to be perfectly implemented. To the best of our knowledge, there are not many literatures related to fault tolerant BLS networks. This paper studies the behaviour of original BLS which are concurrently affected by node noise, weight noise, and open weight fault. Also, a fault aware objective function is derived. Furthermore, based on the proposed objective function, a fault

---

**FIGURE 1.** A typical structure of a BLS network.

---

[36x-1815]In replication approach, a number of networks from a trained network are replicated. Hence, the overall output is equal to the median of the outputs of the replicated networks. The disadvantage of this approach is that it requires extra resources. Besides, a median circuit is more complicated than the traditional linear neuron.

There are other approaches which focus on weight noise/fault or combination of two failures on a ANN. For instance, the works in [26]–[28] investigated the effect of failure situation on the performance of faulty radial basis function (RBF). Parara and Catala in [32] showed how a fault tolerant RBF network can be obtained based on weight decay regularizer [33]. Furthermore, from the perspective of model sensitivity, an approach called explicit regularizer to attain a multi-layer perceptron (MLP) [34]–[36] or RBF [37], [38] was proposed. This approach can handle or tolerate multiplicative weight noise. Recently, failure tolerance ability of the extreme learning machine (ELM) were studied [12], [39]–[42].

Additionally, in [10], fault tolerant ability of a convolutional neural network (CNN) was investigated. Besides, in [43] fault tolerant methods based on diversifying learning were proposed to improve deep neural networks (DNNs) dependability. Moreover, in [44] Xu et al. studied the feasibility of developing fault-tolerant deep learning systems based on model redundancy.

Similarly, recent literatures indicate that soft errors are unavoidable in modern electronic systems, from edge computing devices to supercomputers [45], [46] due to multiple factors [47] such as device aging, wear devices, and high-energy cosmic radiation. In [48], an algorithm for improving fault tolerance of CNNs was proposed.

Furthermore, Leung et al. [49] investigated the effect of open weight faults and multiplicative noise on the performance of bidirectional associative memories (BAM). Moreover, in [50] literature review on fault and noise tolerance in neural networks is discussed. For fault tolerant systems, principles and concepts, interested readers are referred to [51]–[54].

**B. BLS**

The BLS concept [13] is a recent technique for broad connection of hidden nodes of a neural network. As shown in Fig. 1, these nodes are arranged in a broad sense. In a BLS network, there are two types of hidden nodes, feature mapped nodes, and enhancement nodes.

The feature mapped nodes form a layer, namely feature mapped node layer. The input connection weights of these nodes are constructed based on auto-encoder learning. That means, the feature mapped node layer is designed for feature extraction. The outputs of the feature mapped layer are connected to the output layer and another layer, calling enhancement node layer. The enhancement nodes receive the inputs from the feature nodes. In the original BLS learning, the connection weights between the input to the enhancement node are generated in a random manner.

As shown in Fig. 1, the output layer collects the information from the feature node layer and enhancement node layer. In other words, the outputs of feature nodes and enhancement nodes are concatenated in a wide manner and are fed to the output layer. The output weights could be obtained by a least-mean square technique based on Moore-Penrose’s generalized inverse [55], [56]. The overall training computational cost in BLS is not excessive and is much lower than that of other classical training algorithms, such as meta-heuristic techniques (differential evolution [57], particle swarm optimization [58], and gradient descent methods [59].

As the BLS framework provides efficient ways for constructing a flat structured ANN, and a BLS network owns the universal approximation ability [13], [14], the BLS framework has attracted increasing attentions. Many variants of BLS models and algorithm were proposed to enhance its performance [16], [60], [61]. Besides, many applications of BLS were reported in the last several years.

In [16], a hybrid model of BLS and fuzzy concept was proposed, in which the feature nodes are replaced with some fuzzy subsystems. In [18], the BLS system is applied to fault diagnosis in rotor systems. In [62], a BLS model was proposed for object detection using the event data obtained from event cameras. In this model, the incremental learning concept was also used for constructing the resultant BLS network. In [63], the BLS concept was used for processing hyperspectral images. Apartment from applications, there are modifications on its learning algorithm [19], [20]. In [19], the weighted BLS (WBLS) algorithm was proposed for compensating the effect of noise and outliers. In [20], the correntropy criterion based BLS (CBLS) was proposed for handling outlier data based on the maximum correntropy criterion (MCC).
tolerant BLS (FABLS) is proposed to mitigate the effect of network failures.

**III. BLS NETWORKS AND NETWORK FAILURE**

**A. OPERATION OF BLS NETWORKS**

This subsection gives an introduction to the operation of a BLS network. Consider that a BLS network is designed for solving nonlinear regression problem. Let \( x \in \mathbb{R}^D \) and \( o \in \mathbb{R} \), respectively, be the input data and output of a BLS network. First, for clear and easy presentation, This paper denotes the input \( x \) augmented with 1 as \( \chi = [x^T, 1]^T \).

1) **FEATURE MAPPED NODES**

In the BLS network, the feature mapped nodes are grouped into \( n \) groups. Each group is used to extract different features. The \( i \)-th group has \( f_i \) feature mapped nodes. Hence, the total number of feature mapped nodes is

\[
f = \sum_{i=1}^{n} f_i.
\] (1)

For the \( i \)-th group, there is a learned projection matrix, given by

\[
\Psi_i = \begin{pmatrix} \psi_{i,1,1} & \cdots & \psi_{i,1,(D+1)} \\ \vdots & \ddots & \vdots \\ \psi_{i,f_i,1} & \cdots & \psi_{i,f_i,(D+1)} \end{pmatrix},
\] (2)

where \( \Psi_i \in \mathbb{R}^{f_i \times (D+1)} \). The feature mapped nodes are employed to explore the hidden features of the input data. The mapped features \( \boldsymbol{g}_i \) of the \( i \)-th group are obtained by projecting the input data with the matrix \( \Psi_i \). They are given by

\[
\boldsymbol{g}_i = [g_{i,1}, \cdots, g_{i,f_i}]^T = \Psi_i \chi \quad \forall \ i = 1, \cdots, n.
\] (3)

where \( g_{i,u} \) is the \( u \)-th feature of the \( i \)-th group, \( i = 1, \cdots, n \), and \( u = 1, \cdots, f_i \).

The construction procedure of \( \Psi_i \)'s is established via sparse optimization step. This can be achieved in several ways, such as solving sparse optimization problem based on the alternating direction method of multipliers (ADMM) [64] algorithm. The construction procedure of \( \Psi_i \)'s is presented in the next subsection III-B1. It should be noticed that a nonlinear operation on \( \boldsymbol{g}_i \)'s can be used. In the classical BLS framework, no nonlinear operation is applied on \( \boldsymbol{g}_i \)'s. This paper follows the classical BLS framework and does not apply nonlinear operation on \( \boldsymbol{g}_i \)'s.

All the outputs from the feature mapped nodes are collected together as

\[
\boldsymbol{g} = [\boldsymbol{g}_1^T, \cdots, \boldsymbol{g}_n^T]^T \in \mathbb{R}^f.
\] (4)

The augmented vector \( \boldsymbol{q} \) of \( \boldsymbol{g} \) is denoted as

\[
\boldsymbol{q} = [\boldsymbol{g}^T, 1]^T \in \mathbb{R}^{f+1}.
\] (5)

2) **ENHANCEMENT NODES**

The BLS network has \( m \) groups of enhancement nodes. In the BLS framework, the \( j \)-th group of enhancement nodes has \( e_j \) nodes. The total number of enhancement nodes in the BLS network is given by

\[
e = \sum_{j=1}^{m} e_j.
\] (6)

Besides, the output of \( j \)-th group of enhancement nodes is given by

\[
\boldsymbol{h}_j = [h_{j,1}, \cdots, h_{j,e_j}]^T = \xi(W_jq).
\]

where \( j = 1, \cdots, m \) and \( W_j \) is a randomly generated. The elements of \( W_j \) are denoted as

\[
W_j = \begin{pmatrix} w_{j,1,1} & \cdots & w_{j,1,f+1} \\ \vdots & \ddots & \vdots \\ w_{j,e_j,1} & \cdots & w_{j,e_j,f+1} \end{pmatrix}.
\] (7)

It should be noticed that \( \xi(\cdot) \) is the activation function for enhancement nodes. Generally, each group of enhancement nodes can have its own activation function. In the original BLS algorithm, hyperbolic tangent is employed as the activation function for all the enhancement nodes. This paper uses hyperbolic tangent as the activation function for all enhancement nodes. Let \( \eta \) be the collection of all the enhancement node outputs, given by

\[
\eta = [\boldsymbol{h}_1^T, \cdots, \boldsymbol{h}_m^T]^T \in \mathbb{R}^e.
\] (8)

3) **NETWORK OUTPUT**

For a given input vector \( x \), the output of the network is

\[
o = [\boldsymbol{g}^T, \eta^T]^T \beta,
\] (9)

where \( \beta \) is the output weight vector. The number of elements in \( \beta \) is equal to \( f + e \). Hence, its components are given by

\[
\beta = [\beta_1, \cdots, \beta_{f+e}]^T.
\] (10)

**B. CONSTRUCTION OF WEIGHT MATRICES AND VECTORS**

Consider that there are \( N \) training pairs in the training set \( \mathcal{D}_{train} = \{(x_k, y_k) : k = 1, \cdots, N\} \), where \( x_k = [x_{k,1}, \cdots, x_{k,D}]^T \) is the \( k \)-th training input vector and \( y_k \) is the corresponding target output. First, the training data matrix is formed by packing all the input \( x_k \)'s together. The augmented data matrix denoted as \( X \) is given by

\[
X = \begin{pmatrix} x_1^T & 1 \\ \vdots & \vdots \\ x_N^T & 1 \end{pmatrix}.
\] (11)
1) CONSTRUCTION OF THE PROJECTION MATRIX: \( \Psi_i \)
For each group of feature mapped nodes, one vital thing is how to construct the projection matrix, \( \Psi_i \). The technique described here is based on the procedures in [13], [60], [65]. In the BLS, a random matrix \( P_i \in \mathbb{R}^{(D+1) \times n} \) is generated first for each group of feature mapped nodes. Afterwards, a random projection data matrix \( Q_i \) is obtained, given by

\[
Q_i = X P_i
\]  
(12)

The projection matrix \( \Psi_i \) is the solution of the sparse approximation problem given by

\[
\min_{\Psi_i} \| Q_i \Psi_i - X \|_F^2 + \rho \| \Psi_i \|_1
\]  
(13)

where in (13), the term \( \rho \| \Psi_i \|_1 \) is to enforce the solution of (13) to be a sparse matrix. Also, \( \rho \) is a regularization parameter for sparse regularization.

2) CONSTRUCTION OF THE WEIGHT MATRICES OF THE ENHANCEMENT NODES: \( W_j \)
The construction step of \( W_j \)'s is a bit simple. The traditional BLS algorithm and other variants randomly generate the weight matrices for each group of the enhancement nodes. The articles [13], [60], [65] give the details. This paper follows this random generation concept.

3) CONSTRUCTION OF OUTPUT WEIGHT VECTOR
This section gives the procedure to construct the output weight vector \( \beta \). Given the projection matrix \( \Psi_i \) s \( \forall i = 1, \cdots, n \) of the feature mapped nodes and the training data matrix \( X \), the \( i \)-th training data feature matrix for all training samples is given by

\[
Z_i = \begin{pmatrix} z_{i,1}^T \\ \vdots \\ z_{i,n}^T \end{pmatrix} = X \Psi_i^T,
\]  
(14)

where \( z_{i,k} = [z_{i,k,1}, \cdots, z_{i,k,f}]^T \), and

\[
z_{i,k,u} = \sum_{l=1}^D \Psi_{i,u,l} x_{k,l} + \Psi_{i,u,D+1}.
\]  
(15)

Let \( Z \) be the collection of all training data feature matrices, given by

\[
Z = [Z_1, \cdots, Z_n]
\]  
(16)

In this way, \( Z \) is an \( N \times f \) matrix, denoted as

\[
Z = \begin{pmatrix} Z_{1,1} & \cdots & Z_{1,f} \\ \vdots & \ddots & \vdots \\ Z_{N,1} & \cdots & Z_{N,f} \end{pmatrix} = \begin{pmatrix} Z_1^T \\ \vdots \\ Z_N^T \end{pmatrix},
\]  
(17)

where the \( k \)-th row vector \( Z_k^T \) of \( Z \) contains the inputs of the enhancement nodes (the outputs of the feature mapped nodes) for the \( k \)-th training input vector \( x_k \). In order to handle input biases, an one vector is added into \( Z \), given by

\[
Z' = \begin{pmatrix} Z_1^T & 1 \\ \vdots & \vdots \\ Z_N^T & 1 \end{pmatrix} = \begin{pmatrix} Z_1'^T \\ \vdots \\ Z_N'^T \end{pmatrix},
\]  
(18)

Furthermore, given \( Z' \), the enhancement node outputs of the \( j \)-th enhancement group for all training data are given by

\[
H_j = \xi(Z'W_j^T) = \begin{pmatrix} h_{j,1}^T \\ \vdots \\ h_{j,N}^T \end{pmatrix},
\]  
(19)

for \( j = 1, \cdots, m \), where

\[
h_{j,k} = [h_{j,k,1}, \cdots, h_{j,k,e}]^T,
\]  
(20)

and

\[
h_{j,k,v} = \xi(\sum_{r=1}^{f+e} W_{j,v,r} Z'_{k,r}).
\]  
(21)

Packing all the enhancement node outputs together, a \( N \times (\sum_{j=1}^m e) \) matrix \( H \) is obtained:

\[
H = [H_1, \cdots, H_m].
\]  
(22)

Define \( A = [Z|H] \). The output weight vector \( \beta \) can be calculated based on least square techniques

\[
\arg\min_{\beta} : \| A \beta - y \|_2^2 + q \| \beta \|_2^2,
\]  
(23)

where \( y = [y_1, \cdots, y_N]^T \) is the collection of all training outputs, and \( q \) is a regularization parameter.

IV. MULTIPLICATIVE NOISE AND OPEN WEIGHT FAULT IN BLS
In the digital realization of a well trained neural network, the finite precision representation of the weights leads to multiplicative weight noise [3], [4]. Also, using analog implementation, the precision errors are usually specified based on the percentage of error. Hence, the weight deviation is proportional to the magnitude of the weight value.

The output weights \( \beta_l \)’s (\( l = 1, \cdots, f + e \)) can be decomposed into two sets. One set contains the output weights related to the feature mapped nodes. Another set contains the output weights related to the enhancement nodes.

For easy of reading, this paper introduces a duplication labelling for the output weights based on the feature mapped node set and the enhancement node set. The duplication labelling for \( \beta_l \)'s is as follows.

- The symbol \( \beta_{g,i,u} \) denotes the \( u \)-th output weight in the \( i \)-th group of feature mapped nodes, where \( i = 1, \cdots, n \) and \( u = l, \cdots, f_l \).
- The symbol \( \beta_{h,j,v} \) denotes the \( v \)-th output weight in the \( j \)-th group of feature mapped nodes, where \( j = 1, \cdots, m \) and \( v = l, \cdots, e_j \).
A. MULTIPlicative NOISE IN OUTPUT WEIGHTS $\beta$

This paper uses the following model to describe the behaviour of of multiplicative noise on output weights $\beta_i$’s, given by

$$\tilde{\beta}_{g,i,u} = (1 + \gamma_{g,i,u})\beta_{g,i,u},$$

$$\tilde{\beta}_{h,j,v} = (1 + \gamma_{h,j,v})\beta_{h,j,v},$$

where $\tilde{\beta}_{g,i,u}$ and $\tilde{\beta}_{h,j,v}$ are implemented values of $\beta_{g,i,u}$ and $\beta_{h,j,v}$, respectively. The variables $\gamma_{g,i,u}$’s and $\gamma_{h,j,v}$’s are the noise factors that describe the deviation percentages. This paper assumes that they are independent and identically distributed (iid) zero mean random variables with variance equal to $\sigma^2_\beta$.

B. OPEN WEIGHT FAULT IN OUTPUT WEIGHTS $\beta$

Due to hardware imperfection, open weight faults [2], [4], [6] may happen. For the BLS case, feature mapped nodes and enhancement nodes may be disconnected to the output layer. For open weight fault, an implemented output weight can be described as

$$\tilde{\beta}_{g,i,u} = \zeta_{g,i,u} \beta_{g,i,u}, \quad \text{and} \quad \tilde{\beta}_{h,j,v} = \zeta_{h,j,v} \beta_{h,j,v},$$

where $\zeta_{g,i,u}$’s and $\zeta_{h,j,v}$ are open weight factors that describes whether the output weights are opened or not. If an output weight is opened, $\zeta = 0$. Otherwise, $\zeta = 1$. In this paper, the open weight factors are modelled as iid binary random variables. Hence, their probability mass functions are given by

$$\text{Prob}(\zeta) = \begin{cases} p_\beta, & \zeta = 0, \\ 1 - p_\beta, & \zeta = 1, \end{cases}$$

where $p_\beta$ is the fault rate of the output weights.

C. CONCURRENT WEIGHTS IN OUTPUT WEIGHTS $\beta$

When concurrent weight failures (multiplicative weight noise and open weight fault) occur on the output weights, the output weights can be described as

$$\tilde{\beta}_{g,i,u} = (1 + \gamma_{g,i,u})\zeta_{g,i,u} \beta_{g,i,u},$$

$$\tilde{\beta}_{h,j,v} = (1 + \gamma_{h,j,v})\zeta_{h,j,v} \beta_{h,j,v}.$$  

From (28) and (29), when an output weight is opened, its weight value is equal to 0. When the output weight is not opened, its weight value is affected by weight noise only.

D. NOISE IN FEATURE MAPPED NODES

Feature nodes can be affected by multiplicative noise. For easy of reading, denote the current input vector is the $k$-th training vector $x_k$. That means, the augmented input is $x_k = [x_k^T, 1]^T$. Under the multiplicative noise, the feature node output is given by

$$\tilde{z}_{i,k,u} = (1 + \alpha_{i,u})\zeta_{i,k,u}$$

$$= (1 + \alpha_{i,u})\sum_{i=1}^D \psi_{i,u,i} x_{k,i} + \psi_{i,u,D+1}).$$

where $i = 1, \ldots, n$, $u = 1, \ldots, f_i$ and $t = 1, \ldots, D + 1$. In (30), $\alpha_{i,u}$’s are noise factors that describe the percentage of deviation from the feature node output $z_{i,k,u}$. This paper assumes that these noise factors are iid zero mean random variables with variance equal to $\sigma^2_\alpha$. From (30), the noisy version of the training data feature matrices are denoted as $\tilde{Z} = [\tilde{Z}_1, \ldots, \tilde{Z}_n]$.

E. NOISE IN ENHANCEMENT NODES

Also, enhancement nodes can be affected by multiplicative noise. This paper considers that the input weights of the enhancement nodes have precision errors under implementation. Hence, under the multiplicative noise model, the affected input weight, $\tilde{w}_{j,v,r}$ is given by

$$\tilde{w}_{j,v,r} = (1 + \delta_{j,v,r}) w_{j,v,r},$$

where $j = 1, \ldots, m$, $v = 1, \ldots, e_j$, and $r = 1, \ldots, f + 1$. $\delta_{j,v,r}$’s are noise factors that describe the percentage of deviation in input weights. Given the current input is $x_k$, with the aids of Taylor series, the output $h_{j,k,v}$ of an enhancement node can be modelled as

$$\tilde{h}_{j,k,v} = h_{j,k,v} + \sum_{r=1}^{f+1} \delta_{j,v,r} w_{j,v,r} \frac{\partial h_{j,k,v}}{\partial w_{j,v,r}}.$$  

This paper assumes that $\delta_{j,v,r}$’s are iid zero mean random variables with variance equal to $\sigma^2_\delta$.

It should be noticed from (7) and (16),

$$h_{j,k,v} = 1 + \sum_{r=1}^{f+1} w_{j,v,r} Z'_{k,r}.$$  

When the hyperbolic tangent is used as the activation,

$$\frac{\partial h_{j,k,v}}{\partial Z_{k,r}} = Z'_{k,r}(1 - h^2_{j,k,v}).$$

Hence, from (19), (32), and (34), the noise version of enhancement node outputs for all training inputs are denoted as $\tilde{H} = [\tilde{H}_1, \ldots, \tilde{H}_m]$.

F. NOISE STATISTICS OF THE WEIGHTED INPUTS TO THE OUTPUT NODES

This section presents the statistical properties of the weighted inputs to the output node, given by $(\tilde{z}_{i,k,u} \tilde{\beta}_{g,i,u})$ and $(\tilde{h}_{j,k,v} \tilde{\beta}_{h,j,v})$. From the statistical properties of $\gamma_{g,i,u}$’s, $\gamma_{h,j,v}$’s, and $\alpha_{i,u}$’s, the statistics of $(\tilde{z}_{i,k,u} \tilde{\beta}_{g,i,u})$ are given by

$$\langle \tilde{z}_{i,k,u} \tilde{\beta}_{g,i,u} \rangle = (1 - p_\beta)\zeta_{i,k,u} \beta_{g,i,u}$$

$$\langle \tilde{z}^2_{i,k,u} \tilde{\beta}^2_{g,i,u} \rangle = (1 - p_\beta)(1 + \sigma^2_\beta)(1 + \sigma^2_\alpha) \times \zeta^2_{i,k,u} \beta^2_{g,i,u}.\)
In addition, the statistics of \( \tilde{h}_{j,k,v} \tilde{\beta}_{h,j,v} \) are given by

\[
\begin{align*}
\left( \tilde{h}_{j,k,v} \tilde{\beta}_{h,j,v} \right) &= (1 - p_{\beta})\hat{h}_{j,k,v} \hat{\beta}_{h,j,v} \\
\left( \tilde{h}_{j,k,v}^2 \tilde{\beta}_{h,j,v}^2 \right) &= (1 - p_{\beta})(1 + \sigma_{\beta}^2)\hat{h}_{j,k,v}^2 + \sigma_{\beta}^2(1 - \hat{h}_{j,k,v}^2) \sum_{l=1}^{f+1} w_{j,v}^2(Z_{k,v}^2).
\end{align*}
\] (38)

It should be noticed that \( \tilde{h}_{j,k,v} \tilde{\beta}_{h,j,v} \) and \( \tilde{h}_{j',k,v}^2 \tilde{\beta}_{h',j',v}^2 \) are uncorrelated for \( j \neq j' \) or \( v \neq v' \). Also, \( (\tilde{z}_{i,k,u} \tilde{\beta}_{g,i,u}) \) and \( (\tilde{h}_{j,k,v} \tilde{\beta}_{h,j,v}) \) are uncorrelated.

V. THE PROPOSED FAULT-AWARE OBJECTIVE FUNCTION AND LEARNING ALGORITHM

In the noise/fault situation, this paper proposes to minimize the training set error for a particular noise/fault pattern. First, the training set error for a particular fault pattern is considered. With the fault model (24)–(26), and (28)–(32), the training set error for a particular noise/fault pattern is given by

\[
\hat{e} = \frac{1}{N} \sum_{k=1}^{N} \left( \sum_{i=1}^{m} \tilde{z}_{i,k,u} \tilde{\beta}_{g,i,u} + \sum_{j=1}^{n} \tilde{h}_{j,k,v} \tilde{\beta}_{h,j,v} - y_k \right)^2
= \| \hat{A}\hat{\beta} - y \|_2^2.
\] (39)

where \( \hat{A} = [\hat{Z} | \hat{H}] \).

With (35)–(38) and some manipulations, the average training set error of noisy networks over all possible noise/fault is the given by

\[
J = \langle \hat{e} \rangle = \frac{1}{N} \left\| \hat{A}\hat{\beta} - y \right\|_2^2
= \frac{1 - p_{\beta}}{N} \left\| \hat{A}\hat{\beta} - y \right\|_2^2
+ \frac{1 - p_{\beta}}{N} \left( p_{\beta} + \sigma_{\beta}^2 \right) \text{diag}(\hat{A}^T \hat{A}) \beta
+ \frac{1 - p_{\beta}}{N} \left( -p_{\beta} \hat{A}^T \hat{A} \beta \right)
+ \frac{1 - p_{\beta}}{N} \left( (1 + \sigma_{\beta}^2) \text{diag}(\hat{S}) \beta \right)
+ \frac{p_{\beta}}{N} \| y \|_2^2
\] (40)

where \( \text{diag}(\cdot) \) is to return a diagonal matrix from the main diagonal of a square matrix. Note that since the squared error stated in (39) is a convex function of the output weight vector \( \beta \), the average squared error stated in (40) is a convex function too.

The matrix \( \hat{S} \) is a block diagonal matrix, given by

\[
\hat{S} = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}
\] (41)

\[
S_1 = \sigma_{\hat{e}}^2 (\hat{Z})^T \hat{Z}
\] (42)

Define \( \mathcal{R} \) as

\[
\mathcal{R} = (p_{\beta} + \sigma_{\beta}^2) \text{diag}(\hat{A}^T \hat{A}) + (1 + \sigma_{\beta}^2) \text{diag}(\hat{S}) - p_{\beta} \hat{A}^T \hat{A}.
\] (46)

With (46), the average training set error, stated in (40), can be rewritten as

\[
J = \frac{1 - p_{\beta}}{N} \| \hat{A}\hat{\beta} - y \|_2^2 + \frac{1 - p_{\beta}}{N} \| \hat{A}^T \hat{A} \beta \|_2^2
+ \frac{1 - p_{\beta}}{N} \| \text{diag}(\hat{S}) \|_2^2 + \frac{p_{\beta}}{N} \| y \|_2^2.
\] (47)

Since the last term in (47) is independent of the output weight vector, the objective function can be rewritten as

\[
J = \| \hat{A}\hat{\beta} - y \|_2^2 + \beta^T \mathcal{R} \beta.
\] (48)

In (48), it can be seen that the term \( \beta^T \mathcal{R} \beta \) can be considered as regularization term for handling the noise/fault in the BLS network.

Since (40) is a convex function of \( \beta \), (48) is a convex function too. Hence, the optimal solution can be obtained by setting the gradient of (48) to zero. The proposed FABLS solution is

\[
\hat{\beta} = (\mathcal{R} + \hat{A}^T \hat{A})^{-1} \hat{A}^T y.
\] (49)

The FABLS algorithm is summarized in Algorithm 1.

Remark: One advantage of the proposed FABLS algorithm is that there is no tuning parameter. In the original BLS (see (23)) and other variants, such as WBLs [19] and CBLS [20], there is a tuning parameter \( \rho \). The good values of \( \rho \) are chosen by trial and error.

VI. EXPERIMENTS

A. EXPERIMENTAL SETTINGS AND DATASETS

In this section, the proposed algorithm FABLS is compared with the original BLS algorithm, WBLs [19] and CBLS [20]. The WBLs algorithm was designed to handle both noisy data and outliers. Similarly, CBLS is a robust approach based on correntropy criterion. The CBLS algorithm has an anti-noise agent in its objective function. Hence, these comparison algorithms, original BLS, WBLs, and CBLS are best suitable for comparison with the proposed FABLS under network failure situations.

In the comparison, several real life datasets from the UCI machine learning repository [21] are used. They are CTslice,
Algorithm 1 The Training Process of FABLS Algorithm

Require: Training data matrix $X$ according to (11) and the corresponding target $y$, number of groups of feature nodes $n$, number of groups of enhancement nodes $m$. Construct the projection matrices $\Psi_j$'s based on (13). Construct the weight matrices $W_j$'s based on Section III-B1.

Output: Output weight $\beta^*$

1: Compute $Z_i$ using (13)-(14), where $i = 1, \cdots, n$.
2: Set $Z = [Z_1, \cdots, Z_n]$.
3: Compute $H_j$ using (18) and (19), where $j = 1, \cdots, m$.
5: Compute $S$ and $R$ using (41) and (46) respectively.
6: Compute $\beta^*$ according to (49).

TABLE 2. Details of the datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of samples</th>
<th>Number of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTlice</td>
<td>53500</td>
<td>385</td>
</tr>
<tr>
<td>Airfoil Self Noise</td>
<td>1530</td>
<td>7</td>
</tr>
<tr>
<td>Building Energy</td>
<td>4208</td>
<td>14</td>
</tr>
<tr>
<td>Abalone</td>
<td>4177</td>
<td>8</td>
</tr>
<tr>
<td>CCPP</td>
<td>9566</td>
<td>4</td>
</tr>
<tr>
<td>Coil2000</td>
<td>5822</td>
<td>85</td>
</tr>
</tbody>
</table>

TABLE 3. The data division over the 5 fold.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Subset 1</th>
<th>Subset 2</th>
<th>Subset 3</th>
<th>Subset 4</th>
<th>Subset 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTlice</td>
<td>10700</td>
<td>10700</td>
<td>10700</td>
<td>10700</td>
<td>10700</td>
</tr>
<tr>
<td>ASN</td>
<td>301</td>
<td>301</td>
<td>301</td>
<td>301</td>
<td>299</td>
</tr>
<tr>
<td>Building Energy</td>
<td>842</td>
<td>842</td>
<td>842</td>
<td>842</td>
<td>840</td>
</tr>
<tr>
<td>Abalone</td>
<td>835</td>
<td>835</td>
<td>835</td>
<td>835</td>
<td>837</td>
</tr>
<tr>
<td>CCPP</td>
<td>1914</td>
<td>1914</td>
<td>1914</td>
<td>1914</td>
<td>1912</td>
</tr>
<tr>
<td>Coil2000</td>
<td>1164</td>
<td>1164</td>
<td>1164</td>
<td>1164</td>
<td>1166</td>
</tr>
</tbody>
</table>

TABLE 4. various tested $\varphi$ values for BLS, CBLS and WBLS algorithms.

<table>
<thead>
<tr>
<th>Data-set</th>
<th>$\varphi \in (0.01, 0.05, 0.07, 0.35, 0.5, 0.6, 0.9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTlice</td>
<td>$\varphi \in (0.001, 0.02, 0.05, 0.07, 0.09, 0.10, 0.2)$</td>
</tr>
<tr>
<td>ASN</td>
<td>$\varphi \in {0.002, 0.03, 0.06, 0.08, 0.2, 0.4, 0.6}$</td>
</tr>
<tr>
<td>Building Energy</td>
<td>$\varphi \in {0.002, 0.006, 0.009, 0.0, 0.4, 0.7, 0.85}$</td>
</tr>
<tr>
<td>Abalone</td>
<td>$\varphi \in {0.01, 0.05, 0.08, 0.1, 0.25, 0.35, 0.8}$</td>
</tr>
<tr>
<td>CCPP</td>
<td>$\varphi \in {0.01, 0.25, 0.35, 0.4, 0.6}$</td>
</tr>
</tbody>
</table>

The original BLS algorithm, WBLS algorithm and CBLS algorithm have a regularizer term $\varphi$ that can mitigate the effect of fault in the network. In the original BLS algorithm and other variants, there is a tuning regularization parameter $\varphi$. The value of $\varphi$ affects the performance of resultant BLS networks. There is no analytical way of selecting the suitable value of $\varphi$. In this experiment, various $\varphi$ values are tested for each dataset, and then the best $\varphi$ value is found based on the test set. Table 4 summarizes the various tested $\varphi$ values.

In the BLS concept, the feature nodes of a BLS network uses linear activation function to process the input data. The problem with this approach is that the output values of feature nodes could be very large. This behaviour can affect the performance of the original BLS algorithm, if the output value is forwarded to enhance node directly, without any additional process. In the traditional BLS, the commonly used solution is that after getting the output values of feature nodes, the values are mapped or normalized to range $[0, 1]$. This approach can improve the performance of the original BLS algorithm. In this paper, for fair comparison, this normalization concept is used for the original BLS algorithm, WBLS algorithm, CBLS algorithm and the proposed FABLS algorithm.
Because when noise levels are small, the advantage of our age test MSE values of the four algorithms are similar. It is algorithms are listed in tabular form in Table 5.

When the open weight fault level is fixed at \( p_\beta = 0.05 \), the CBLS algorithm, as shown in the first row and the first column subfigures in Figs 2-4. For CTslice, when the noise level is \( \sigma_2^w = \sigma_2^z = \sigma_2^0 = 0.09 \), \( \{p_\beta = 0.01, \sigma_2^w = \sigma_2^z = \sigma_2^0 = 0.09\} \). Also, three open fault rates are considered, and they are: \( \{p_\beta = 0.01, \sigma_2^w = \sigma_2^z = \sigma_2^0 = 0.25\} \).

In our experiment, \( n = 10 \) and \( f_i = 10 \), where \( i = 1, \ldots, n \). With this setting, there are 100 feature mapped nodes. For the enhancement nodes, \( m = 1 \) and \( d_0 = e = 100 \). Hence, there are 100 enhancement nodes. With the setting, there are 200 hidden nodes in the BLS network.

### B. PERFORMANCE COMPARISON

This subsection presents the performance of the comparison algorithms in term of average test set MSE of noisy networks. The 5-fold evaluation strategy is used in the experiments. For each dataset and each fold, 2000 faulty networks are generated. For an easy and quick view of the result, Figs. 2–4 summarize the average test set MSE of noisy networks and the corresponding standard deviation over the 5 fold evaluation. Each figure shows the results for an open fault rate. Besides, the numerical values of the average test set MSE and the corresponding standard deviation of the comparison algorithms are listed in tabular form in Table 5.

From Figs. 2–4, when the noise levels are small, the average test MSE values of the four algorithms are similar. It is because when noise levels are small, the advantage of our proposed FABLS is not fully demonstrated. However, the average test set MSE values of the proposed FABLS algorithm are still small those of the original BLS algorithm, WBLS algorithm, and CBLS algorithm. The improvements of the proposed FABLS are more significant when the noise levels is large.

For example, for the ASN dataset, at low noise level of \( \{p_\beta = 0.01, \sigma_2^w = \sigma_2^z = \sigma_2^0 = 0.01\} \), the average test set MSE of the original BLS is 0.156109, the average test set MSE of the WBLS jumps to 0.082206. However, the average test set MSE of FABLS jumps to 0.082964. Also, the average test set MSE of the WBLS jumps to 0.018603. For the proposed FABLS algorithm, it has an average test set MSE of 0.016880 only.

In a more complicated dataset, namely CTslice, the performance of our proposed FABLS algorithm is clearly superior to the BLS algorithm, the WBLS algorithm and the CBLS algorithm, as shown in the first row and the first column subfigures in Figs 2–4. For CTslice, when the noise level is set to \( p_\beta = 0.01, \sigma_2^w = \sigma_2^z = \sigma_2^0 = 0.01 \), the average test set MSE of the original BLS algorithm is 0.156109. Besides, the
TABLE 5. Average test set MSE and Standard Deviation of BLS, WBLS, CBLS and FABLS. There are 200 nodes in the network. The 5-fold evaluation strategy is used. For each dataset and each fold, 2,000 faulty networks are generated.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Noise Level</th>
<th>p_β</th>
<th>σ^2_β = σ^2_ε = σ^2_γ</th>
<th>AVG MSE</th>
<th>SD</th>
<th>AVG MSE</th>
<th>SD</th>
<th>AVG MSE</th>
<th>SD</th>
<th>AVG MSE</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building Energy</td>
<td>0.01</td>
<td>0.01</td>
<td>0.04183</td>
<td>0.00229</td>
<td>0.00408</td>
<td>0.00232</td>
<td>0.00418</td>
<td>0.00229</td>
<td>0.00408</td>
<td>0.00232</td>
<td></td>
</tr>
<tr>
<td>CCPP &amp; CTSlice</td>
<td>0.05</td>
<td>0.05</td>
<td>0.00819</td>
<td>0.00297</td>
<td>0.00822</td>
<td>0.00300</td>
<td>0.00819</td>
<td>0.00297</td>
<td>0.00822</td>
<td>0.00300</td>
<td></td>
</tr>
<tr>
<td>Abalone</td>
<td>0.10</td>
<td>0.10</td>
<td>0.00775</td>
<td>0.00314</td>
<td>0.00775</td>
<td>0.00314</td>
<td>0.00775</td>
<td>0.00314</td>
<td>0.00775</td>
<td>0.00314</td>
<td></td>
</tr>
<tr>
<td>Coll2000</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00609</td>
<td>0.00231</td>
<td>0.00625</td>
<td>0.00234</td>
<td>0.00609</td>
<td>0.00231</td>
<td>0.00625</td>
<td>0.00234</td>
<td></td>
</tr>
<tr>
<td>ASN</td>
<td>0.05</td>
<td>0.05</td>
<td>0.00282</td>
<td>0.00168</td>
<td>0.00285</td>
<td>0.00171</td>
<td>0.00282</td>
<td>0.00168</td>
<td>0.00285</td>
<td>0.00171</td>
<td></td>
</tr>
</tbody>
</table>

For other datasets, similar results are obtained. Hence, our proposed FABLS algorithm outperforms the original BLS algorithm, the WBLS algorithm and the CBLS algorithm under the concurrent noise situation.

C. PAIRED T-TEST

As explained earlier in Subsection VI-B, from Table 5, in terms of average test set MSE, the proposed FABLS algorithm are better than the comparison algorithms. However, there is still a key question: if the improvements of using our FABLS are statistically significant. Hence, the paired t-test analysis is performed. It is a standard method to investigate whether the mean difference between two sets of samples is statistically significant.
TABLE 6. The paired t-test result between FABLS and BLS. There are 200 nodes in the network. The 5-fold evaluation strategy is used. For each dataset and each fold, 2,000 faulty networks are generated. The critical t-value is 2.132.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Noise Level</th>
<th>AVG MSE Improvement</th>
<th>t-value</th>
<th>Confident interval of improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building Energy</td>
<td>0.01</td>
<td>0.00140</td>
<td>7.32</td>
<td>[0.00086-0.00193]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.00213</td>
<td>7.55</td>
<td>[0.00163-0.00219]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.00579</td>
<td>24.77</td>
<td>[0.05146-0.06446]</td>
</tr>
<tr>
<td>CPP</td>
<td>0.01</td>
<td>0.00124</td>
<td>5.91</td>
<td>[0.00066-0.00183]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.00184</td>
<td>21.06</td>
<td>[0.01650-0.02093]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.00571</td>
<td>24.46</td>
<td>[0.05070-0.06368]</td>
</tr>
<tr>
<td>Coil2000</td>
<td>0.01</td>
<td>0.00071</td>
<td>16.97</td>
<td>[0.00116-0.02130]</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.00424</td>
<td>24.09</td>
<td>[0.02166-0.02732]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.00591</td>
<td>26.92</td>
<td>[0.03509-0.06530]</td>
</tr>
<tr>
<td>Abalone</td>
<td>0.01</td>
<td>0.00034</td>
<td>16.49</td>
<td>[0.00028-0.00040]</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.00039</td>
<td>44.83</td>
<td>[0.00039-0.00349]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.00096</td>
<td>46.42</td>
<td>[0.00097-0.01556]</td>
</tr>
<tr>
<td>ASN</td>
<td>0.01</td>
<td>0.00025</td>
<td>14.90</td>
<td>[0.00175-0.02255]</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.00037</td>
<td>21.54</td>
<td>[0.00329-0.00546]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.00080</td>
<td>27.83</td>
<td>[0.00320-0.00880]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.00070</td>
<td>33.59</td>
<td>[0.00444-0.00995]</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.00105</td>
<td>37.95</td>
<td>[0.00974-0.01218]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.00143</td>
<td>40.68</td>
<td>[0.01332-0.01527]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.13233</td>
<td>75.59</td>
<td>[0.12747-0.13719]</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>1.36863</td>
<td>173.57</td>
<td>[1.34673-1.39052]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>4.17775</td>
<td>153.99</td>
<td>[4.1462-4.24112]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>1.14347</td>
<td>24.49</td>
<td>[0.13444-0.20531]</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>1.19772</td>
<td>126.02</td>
<td>[1.17134-1.12411]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>3.49150</td>
<td>137.46</td>
<td>[3.42098-3.56202]</td>
</tr>
</tbody>
</table>

are statistically significant or not. For the two tailed test with 5 trials and 95% confident level, the critical t-value is 2.132.

1) PAIRED T-TEST BETWEEN FABLS AND BLS

Table 6 summarizes the paired t-test result between the proposed FABLS algorithm and the original BLS algorithm. From the table, all the t-values are greater than the critical t-value, i.e., 2.132. Hence there is a strong evidence to conclude that the improvements are statistically significant. For instance, in the Coil2000 dataset with noise/fault level equal to $p_{\beta} = 0.1, \sigma_{\beta}^2 = \sigma_{\alpha}^2 = \sigma_{\gamma}^2 = 0.09$, the t-value is 13.12, which is greater than 2.132. Furthermore, the average MSE improvement of using FABLS is 0.053712 and the corresponding confident interval of average improvement is [0.042344-0.065081]. In addition, the confident interval does not include zero. In the table, similar situations are found in other noise levels and datasets. With these results, there is a strong evidence that the FABLS is better than BLS. In fact, from the t-values at the large noise levels, the confidence
TABLE 8. The paired t-test result of FABLS and CBLS. There are 200 nodes in the network. The 5-fold evaluation strategy is used. For each dataset and each fold, 2,000 faulty networks are generated.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Noise Level</th>
<th>AVG MSE Improvement</th>
<th>t-value</th>
<th>Confident interval of improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building Energy</td>
<td>0.01</td>
<td>0.00138</td>
<td>7.30</td>
<td>[0.00086-0.00191]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.00122</td>
<td>5.84</td>
<td>[0.00040-0.00180]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.00101</td>
<td>4.47</td>
<td>[0.00030-0.00155]</td>
</tr>
<tr>
<td>CCPP</td>
<td>0.01</td>
<td>0.00004</td>
<td>16.53</td>
<td>[0.00028-0.00080]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.00099</td>
<td>46.45</td>
<td>[0.00035-0.00105]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.00079</td>
<td>27.79</td>
<td>[0.00018-0.00098]</td>
</tr>
<tr>
<td>CTSlice</td>
<td>0.01</td>
<td>0.13228</td>
<td>7.57</td>
<td>[0.12743-0.13714]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.13682</td>
<td>173.70</td>
<td>[1.34635-1.39009]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.13774</td>
<td>182.32</td>
<td>[1.40502-1.44386]</td>
</tr>
<tr>
<td>Abalone</td>
<td>0.01</td>
<td>0.00080</td>
<td>3.12</td>
<td>[0.00008-0.00152]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.00095</td>
<td>4.13</td>
<td>[0.00040-0.00137]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.00101</td>
<td>4.47</td>
<td>[0.00030-0.00155]</td>
</tr>
<tr>
<td>Coil2000</td>
<td>0.01</td>
<td>0.00536</td>
<td>7.22</td>
<td>[0.00433-0.00729]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.04573</td>
<td>12.83</td>
<td>[0.04353-0.05563]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.13700</td>
<td>13.17</td>
<td>[0.10812-0.16588]</td>
</tr>
<tr>
<td>ASN</td>
<td>0.01</td>
<td>0.000172</td>
<td>7.37</td>
<td>[0.000107-0.00237]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.000056</td>
<td>5.87</td>
<td>[0.000000-0.00292]</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.000261</td>
<td>4.47</td>
<td>[0.000184-0.002652]</td>
</tr>
</tbody>
</table>

The t-value is 11.50, which is greater than 2.132. Furthermore, the average improvement of using FABLS is 0.015640 and the corresponding confident interval of average improvement does not include zero. In the table, similar situations are found in other noise/fault levels and datasets. With these results, there is a strong evidence that the FABLS is better than WBLIS.

3) PAIRED-T TEST BETWEEN FABLS AND CBLS

Similar to the other comparison algorithms, Table 8 summarizes the paired t-test results between the proposed FABLS algorithm and the CBLS algorithm.

As shown in Table 8, all confident intervals of the average improvements of using FABLS do not include the zero, and all the t-values are greater than the critical t-value. For instance, in the CTSlice dataset with noise level equal to $[\beta_\text{p} = 0.01, \sigma_\beta^2 = \sigma_\gamma^2 = \sigma_\omega^2 = 0.09]$, the t-value is 173, which is much greater than 2.132. Furthermore, the average MSE improvement of using FABLS is 1.368225 and the corresponding confident interval of average improvement $[1.36435 - 1.39009]$. In the table, similar situations are found in other noise levels and datasets. With these results, there is a strong evidence that the FABLS is better than CBLS.

VII. CONCLUSION

This paper first analyzed the effects of open weight fault and multiplicative noise in the BLS network. Four kinds of noise/fault are considered. They are open weight fault in output weights, multiplicative noise in output weights, multiplicative noise in feature nodes, and multiplicative noise in input weights of enhancement nodes. From the analyzed result, a fault aware objective function for BLS networks is then developed. The objective function contains a fault aware regularizer term which is capable of mitigating the effect of aforementioned fault/noise. Besides, there is no the regularizer parameter in our approach. On the other hand, in the traditional BLS, WBLIS and CBLS, there is is tuning regularization parameter.

Based on the proposed objective function, a fault-aware learning algorithm was developed to train a BLS network. The simulation experiments show that the proposed FABLS algorithm outperforms the original BLS algorithm, the WBLIS algorithm and the CBLS algorithm. Also, the paired t-test results show that the improvements of the proposed FABLS algorithm are statistically significant.

REFERENCES


