Chapter 4

Deep learning-based identification of false data injection attacks

4.1 Introduction

Real-time successful identification of FDIAs is an indispensable requirement to ensure secure and reliable grid operation. State forecasting-driven attack detection models can effectively determine the deviations of the operating states due to FDIA. This chapter showcases scalable deep neural networks based state forecasting strategy which is capable of detecting FDIAs in real-time using thresholding over the $\mathcal{L}_2$ norm of the error vector and on the rate of change of the eigenvalues of the error covariance matrix. With an optimal set of hyper-parameters, an effective, scalable, real-time state forecasting policy with minimal error indices has been showcased. A comparison between the proposed neural network models with the state-of-the-art state forecasting policies showcases its efficacy. Furthermore, the developed intrusion detection algorithm defined on the basis of the $\mathcal{L}_2$ norm of the error vector and on the error covariance matrix furnishes an effective robust, real-time attack detection scheme within the obtained measurements with high accuracy. An extensive survey on the IEEE 14-bus test bench portrays the effectiveness of the proposed real-time FDIA detection policies. Moreover, the developed approaches demonstrate a superior, real-time, robust FDIA identification scheme under varying noise and attack scenarios.

4.2 Proposed forecasting strategy

This chapter undertakes the nonlinear state estimation model followed by the effective implementation of FDIA as shown in chapter 2. Deep learning models have recently shown efficient performance in regression analysis. Such data-driven forecasting policies have shown minimal error in their performance [246]. This thesis has undertaken RMSE, MSE and MAE as the parameter metrics which can be defined as follows:

$$RMSE = \sqrt{\sum_{a=1}^{b_1} \frac{(f_a - s_a)^2}{b_1}} \tag{4.1}$$

$$MSE = \sum_{a=1}^{b_1} \frac{(f_a - s_a)^2}{b_1} \tag{4.2}$$

$$MAE = \sum_{a=1}^{b_1} \frac{|(f_a - s_a)|}{b_1} \tag{4.3}$$

where $f,\ s$ represent the predicted set of state estimates and the actual computed set of estimated states respectively. b_1 represents the total number of the predicted samples. With an effective tuning of the model hyper-parameters, it can be seen that the developed deep learning models can effectively forecast the operating states for steady-state operation of the grid with minimal RMSE, MSE, and MAE index. A comparative analysis with ARIMA based state forecasting policy along with another machine learning model (SVM) has demonstrated the superiority of the proposed approach.

4.2.1 DNN based state forecasting model

The proposed nonlinear deep neural network topology can be shown in Fig. 4.1. It can be inferred from Fig. 4.1 that the proposed nonlinear network topology is modeled with 5 hidden layers along with one input and output layer. The model can be seen to split at layer two with hidden sublayers. It can be seen that there exist two sub-hidden layers. Such hidden sublayers are finally merged into a common concatenation layer. Two dense layers succeed this common concatenation layer followed by one output layer. All the layers of the proposed nonlinear network topology are fed with the rectified linear unit (Relu) activation function which can be shown as follows:

$$y_i = Relu(w_i k_i + b_i) \tag{4.4}$$

y_i represents the predicted set of features at the i^{th} layer. k_i denotes the input features given to the model at the i^{th} layer. w_i, b_i respectively represents the weight and the bias of the network for the i^{th} layer. An optimal number of hidden layers for the nonlinear network topologies are selected to ascertain an effective state forecasting policy with minimum RMSE and MSE index as shown in table 4.1. An effective tuning of the hyper-

	Hidden & (sub hidden layers)	RMSE	MSE
DNN	3 (1)	2.3343	5.4489
	3 (2)	2.1288	4.5317
	4 (1)	2.0009	4.0036
	4 (2)	1.9651	3.8616
	5 (1)	1.7751	3.1509
	5 (2)	**1.6782**	**2.8163**
	5 (3)	1.7122	2.9316
LSTM	2 (2)	0.0045	2.0249×10^{-5}
	2 (3)	0.0015	2.25×10^{-6}
	3 (2)	0.0003	8.9999×10^{-8}
	3 (3)	0.0001	1×10^{-8}
	4 (2)	0.000097	9.409×10^{-9}
	4 (3)	0.000084	7.056×10^{-9}
	5 (2)	0.000065	4.2249×10^{-9}
	5 (3)	**0.000057**	$\mathbf{3.249} \times 10^{-9}$
	6 (2)	0.000062	3.844×10^{-9}

Table 4.1: Hyper-parameter tuning of the developed models

parameters showcases the efficacy of the proposed approach as can be seen in the results. The optimal number of such hidden layers with the Relu activation function has been adopted to yield higher performance accuracy with a reduced computational burden as shown in table 4.1.

It must be noted that a similar tuning of hyper-parameters are also achieved taking MAE index into account. Furthermore, it can be also inferred from table 4.1 that as the number of hidden layers are further increased than layer 5 for both the DNN and LSTM model, saturation of model performance is seen.

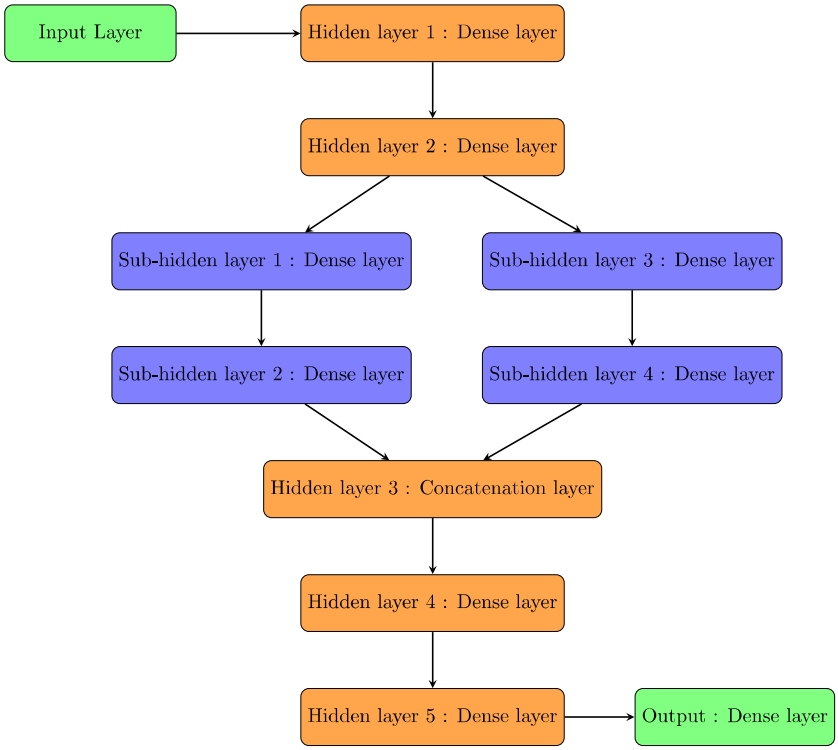

Figure 4.1: Proposed Neural Network Structure

4.2.2 LSTM based state forecasting model

LSTM topologies can be defined as specialized recurrent neural network structures useful in learning complex temporal patterns in the training dataset. It can inherently retain temporal information of the data for a period of discrete time steps. It can read, store, and erase data from its memory cell. It is made achievable with the aid of three gates like the input gate $[i'_1(t)]$, forget $[f'_1(t)]$ and the output gate $[o'_1(t)]$ respectively. Forget gate is essential in determining whether information must be retained for the current cell state or not. Sigmoid activation function has been adopted for the forget gate $[f'_1(t)]$, input gate $[i'_1(t)]$ and the output gate $[o'_1(t)]$. $w_{1(.)}$, $h'_1(t)$, $x'_1(t)$ and $b_{1(.)}$ respectively represent the weights, cell output and cell input along with the scalar bias. The input gate is present to store the input values within the cell state. $g'_1(t)$ denotes the new candidate solution, $c'_1(t)$ represents the new state of the cell. The output gate is present to supply the stored information to the upcoming cell. This is useful in dealing with the non-linear relationship within the measurements followed by their long-term dependencies. For each

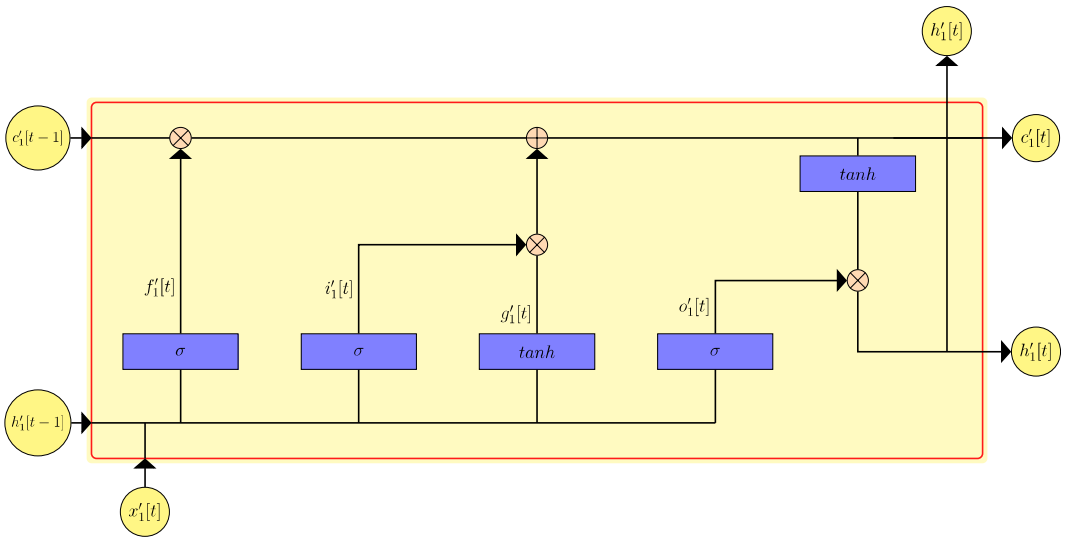

Figure 4.2: LSTM module

LSTM module, [., .] denotes the concatenation operation.

$$f'_1(t) = \sigma(w_{1f}[h'_1(t-1), x'_1(t)] + b_{1f}) \tag{4.5}$$

$$c'_1(t) = f'_1(t) \odot c'_1(t-1) + i'_1(t) \odot g'_1(t) \tag{4.6}$$

$$i'_1(t) = \sigma(w_{1i}[h'_1(t-1), x'_1(t)] + b_{1i}) \tag{4.7}$$

$$g'_1(t) = tanh(w_{1g}[h'_1(t-1), x'_1(t)] + b_{1g}) \tag{4.8}$$

$$o'_1(t) = \sigma(w_{1o}[h'_1(t-1), x'_1(t)] + b_{1o}) \tag{4.9}$$

$$h'_1(t) = o'_1(t) \odot tanh(c'_1(t)) \tag{4.10}$$

$\odot$ represents the element-wise product operation. Fig. 4.2 indicates a single cell structure of an LSTM module while the proposed LSTM model is shown as per Fig. 4.3

The proposed nonlinear architectures have been trained for 500 epochs with adam optimizer. An initial learning rate of 0.001 has been adopted for the proposed nonlinear architectures. The current estimated set of states along with the historical set of estimated states is aggregated to formulate the dataset used for model training. For an effective determination of model performance, 70% of this dataset is fed to the nonlinear architectures for training while the remaining 30% is kept for model testing. For effective model validation, 10% of the training dataset is used while training the model. Such nonlinear models encompass drop-out regularisation [247] to prevent model over-fitting. Drop-out regularisation can be adopted at each sub-hidden layer also, but it hardly showcases any performance enhancement.

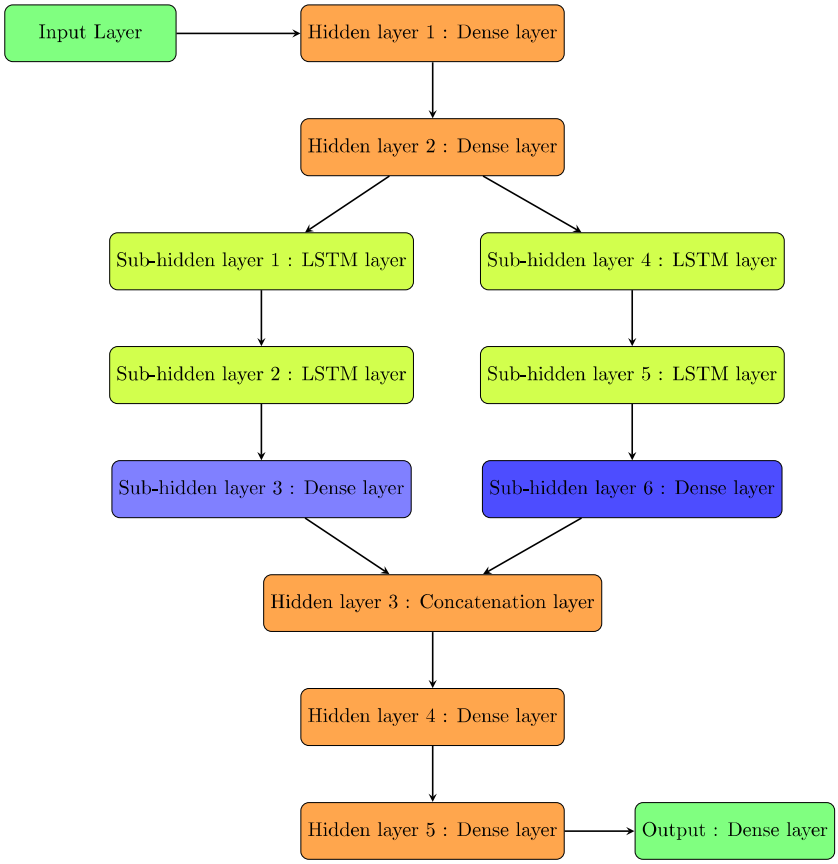

Figure 4.3: Proposed deep learning structure with LSTM modules

The following subsections demonstrate the two developed anomaly detection algorithms within the set of estimated states using the error generated due to the forecasted and the estimated set of state variables.

4.3 Proposed anomaly detection scheme - I

With an effective training of the proposed neural network models with minimal error indices, an effective state forecasting strategy can be demonstrated. Moreover, it can be also deployed for an efficient FDIA detection policy. The proposed FDIA detection scheme with such trained neural network and machine learning models can be defined as per Fig. 4.4. The following criteria is adopted to define the presence of FDIA within the raw measurements.

$$\varepsilon = \begin{cases} 1 & \Delta > RMSE + \delta \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

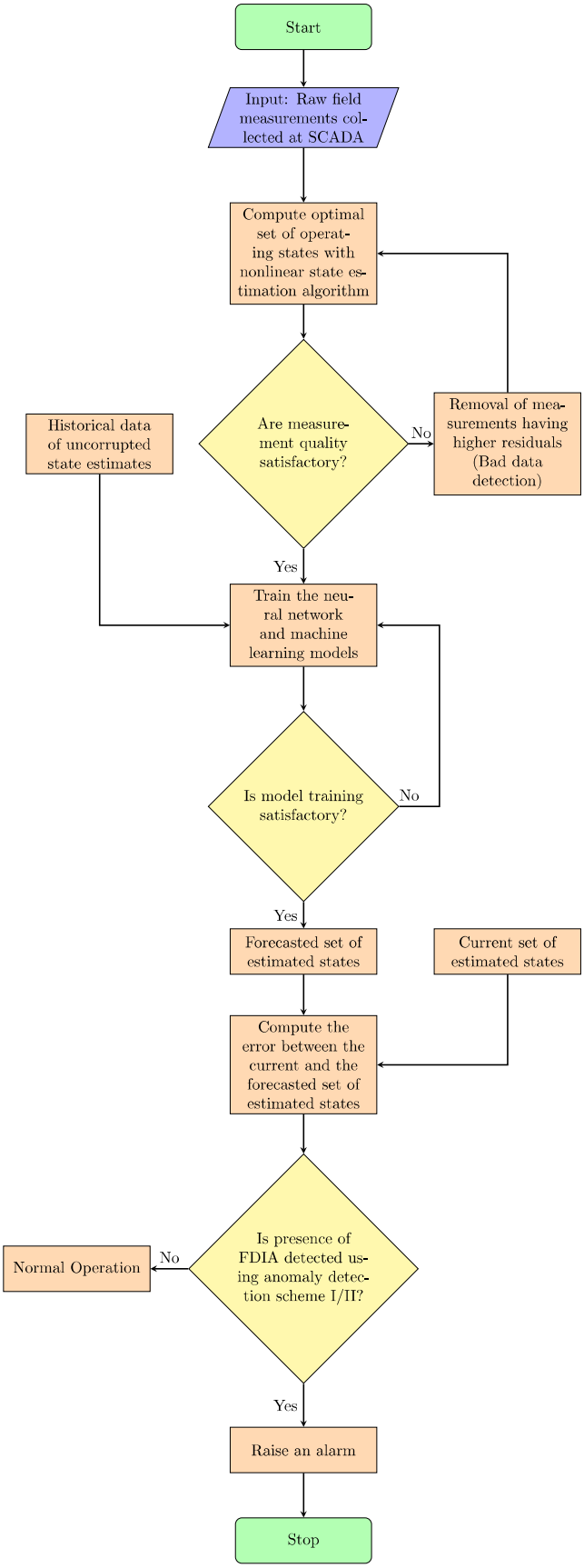

Figure 4.4: Proposed FDIA detection strategy using anomaly detection schemes

where, Δ represents the $\mathcal{L}_2$ norm of the deviation between the current set of state estimates as derived from the EMS module and the forecasted set of state estimates as derived from the proposed neural network. It can be seen that the operator can detect the presence of FDIA if the detection parameter ε is set to one. A normal operating condition of the grid prevails when the detection parameter is zero for the current set of estimated states. δ represents an uncertainty parameter as can be defined from the operator experience. It is seen to be very small as the model shows enhanced forecasting accuracy. It can be mentioned here that such detection parameters can be modeled by taking the MSE index into account as well. FDIA detection under contingency conditions is kept for future research. It can be seen from (4.11) that models with the least RMSE index tend to generate a better bound on the detection parameter ε.

Fig 4.4 demonstrates the effective FDIA presence detection strategy under a steady-state operating scenario. Under system dynamic conditions, the aforesaid strategy can be implemented by making the following changes:

- Primarily, under small disturbances or when the load changes are not significant, the operator may define the thresholding parameter based on the uncertainty parameter (δ) defined on operator experience. This indirectly leads to thresholding on the error vector defined due to the forecasted and the current set of state estimates. Although this strategy may take care of small-scale system dynamics, operator experience for defining δ plays a crucial role in defining the FDIA presence detection model. The benefit of this approach is that the pre-trained neural networks can be deployed directly without needing any further modifications.

- Secondly, if large-scale dynamics are persistent, then firstly a dynamic state estimation strategy like the Kalman filter etc. should be deployed instead of the nonlinear state estimation strategy. This will lead to a set of state estimates which can be used to train the nonlinear neural networks. Once trained under the aforesaid circumstances, the proposed detection strategy as shown in Fig. 4.4 will hold again. The drawback of such an approach is that dynamic state estimators should be incorporated and the high computational time due to re-training of the neural network models should be addressed. Although the proposed nonlinear neural networks demonstrate a training time in the order of μ seconds, an online training based on

the current set of state estimates can be adopted. This leads to tackling the FDIA detection problem under large-scale system dynamics.

4.4 Proposed anomaly detection scheme - II

An advanced anomaly detection scheme undertaking the rate of change of the eigenvalues of the error covariance matrix is demonstrated in this section. The proposed anomaly detection algorithm undertakes the error vector developed due to the estimated and the forecasted operating states at SCADA as shown:

$$e(t) = \hat{x}_{for}(t) - \hat{x}_{est}(t) \tag{4.12}$$

where, $\hat{x}_{for}(t) \in \mathcal{R}^n$ represents the forecasted set of estimated states as derived from the scalable, nonlinear neural network architectures and $\hat{x}_{est}(t) \in \mathcal{R}^n$ denotes the set of estimated states for the current time step t as retrieved from the state estimation algorithm within the EMS module in SCADA. $e(t) \in \mathcal{R}^n$ represents the error vector. The primary objective of the anomaly detection scheme lies in identifying the rate of change of the eigen values of the error covariance matrix as shown:

$$\frac{d\lambda}{dt} = x_1^T(t)\frac{dE(t)}{dt}x_1(t) \tag{4.13}$$

$$\text{where } \frac{dE}{dt} \simeq \frac{E(t) - E(t-\delta t)}{\delta t} \tag{4.14}$$

where, $x_1(t) \in \mathcal{R}^n$ represents the eigenvector for the error covariance matrix $E(t) \in \mathcal{R}^{n \times n}$ developed during the current sampling time t. As $E(t)$ and $E(t)^T$ are positive semidefinite symmetric covariance matrices, hence they have similar eigenvalues and eigenvectors. The rate of change of the eigenvalues between two successive time intervals (t) and $(t - \delta t)$ (for a very small δt) can be represented by (4.13) and (4.14). As SCADA samples the measurements at an order of a few hundred Hz, hence δt can be determined in the order of a few milliseconds. The proposed FDIA detection scheme undertaking the rate of change of the eigenvalues can be defined as:

$$\varepsilon_1 = \begin{cases} 1 & \frac{d\lambda}{dt} > RMSE + \delta_1 \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

where, ε_1 represents the detection criterion which is set to 1 denoting the presence of FDIA within the acquired set of measurements if the rate of change of eigenvalues $(\frac{d\lambda}{dt})$

for a particular time instant t overshoots a predefined threshold as shown in (4.15). δ_1 represents a very small positive numeric constant ($\simeq 10^{-5}$) that depends on operator experience. Such a parameter is inherently very small as the nonlinear state forecasting model demonstrates an enhanced forecasting accuracy.

Proof. To demonstrate the proof of (4.13), the followings can be inferred from [239]:

$$E(t)x_1(t) = \lambda(t)x_1(t) \tag{4.16}$$

$$x_1(t)^T E(t) = \lambda(t)x_1(t)^T \tag{4.17}$$

where, $\lambda(t)$ represents the eigen value of $E(t)$ for a particular sampling time t. For an effective normalisation of the eigen vectors $x_1(t)$ and $x_1(t)^T$, the followings can be inferred:

$$x_1(t)^T x_1(t) = 1 \tag{4.18}$$

As $E(t)$ and $E(t)^T$ have the same eigen values and eigen vectors, hence (4.16) and (4.17) holds. It can be inferred from (4.16) - (4.18) that:

$$\lambda(t) = x_1(t)^T E(t) x_1(t) \tag{4.19}$$

Differentiating both sides with respect to t gives:

$$\frac{d\lambda}{dt} = \frac{dx_1(t)^T}{dt}E(t)x_1(t) + x_1(t)^T\frac{dE}{dt}x_1(t) + x_1(t)^T E(t)\frac{dx_1(t)}{dt} \tag{4.20}$$

$$= \frac{dx_1(t)^T}{dt}\lambda(t)x_1(t) + x_1(t)^T\frac{dE}{dt}x_1(t) + \lambda(t)x_1(t)^T\frac{dx_1(t)}{dt} \tag{4.21}$$

$$= \lambda(t)[\frac{dx_1(t)^T}{dt}x_1(t) + x_1(t)^T\frac{dx_1(t)}{dt}] + x_1(t)^T\frac{dE}{dt}x_1(t) \tag{4.22}$$

$$= \lambda(t)\frac{d}{dt}[x_1(t)^T x_1(t)] + x_1(t)^T\frac{dE}{dt}x_1(t) \tag{4.23}$$

$$\text{As } x_1(t)^T x_1(t) = 1, \text{ hence : } \frac{d\lambda}{dt} = x_1(t)^T\frac{dE}{dt}x_1(t) \tag{4.24}$$

This concludes the proof of (4.13) which shows the rate of change of the eigenvalues of the error covariance matrix. □

In general, the developed FDIA detection strategies adopt a deterministic threshold over the set of state estimates. With successful attack vector injections within the raw measurements, a definitive deviation in the set of estimated states is observed. The operator reports a successful identification of an attack within the raw measurements when such aforesaid deviations overshoot the deterministic threshold.

4.5 Results

This section primarily showcases a perfect FDIA implementation on the IEEE 14-bus system as shown in chapter 2 followed by its effective identification. It must be noted here that the $\mathcal{L}_2$ norm of the formulated attack vector is initially defined within the unit ball. RMSE has been adopted as the loss function while training the models. Figs. 4.5 and 4.6 represent the minimization of this loss function while tuning the hyper-parameters of the MLP and the nonlinear LSTM model.

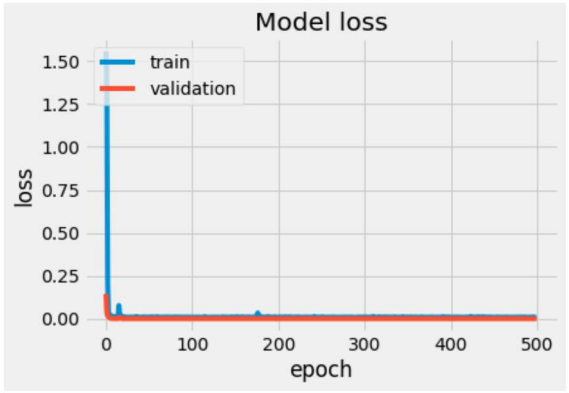

Figure 4.5: Training & Validation loss of MLP model

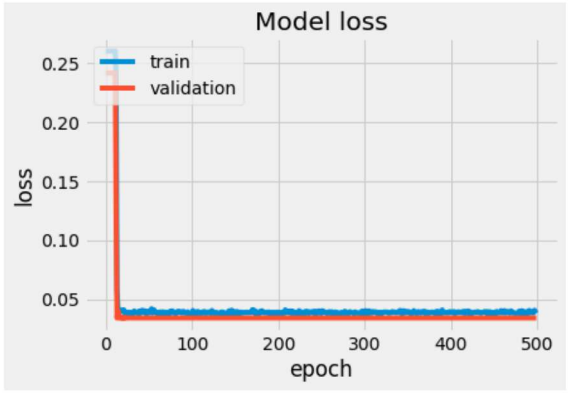

Figure 4.6: Training & Validation loss of LSTM model

Moreover, it can be clearly seen from Fig. 4.7 that the nonlinear LSTM model can efficiently predict the estimated states (voltage magnitude and their respective phase angles) efficiently with a forecasting range of about 200 samples each. Additionally, it shows a better forecasting performance than the other undertaken models like ARIMA and SVM.

Sl. No	Forecasting Model	RMSE	MSE	MAE
1	Deep Neural Network (LSTM)	0.000057	3.249×10^{-9}	0.00872×10^{-3}
2	Deep Neural Network (MLP)	1.6782	2.8163	2.2234
3	Support Vector Machine	1.7103	2.9251	2.6612
4	ARIMA	3.3245	11.0523	8.8876

Table 4.2: Comparative analysis of forecasting models

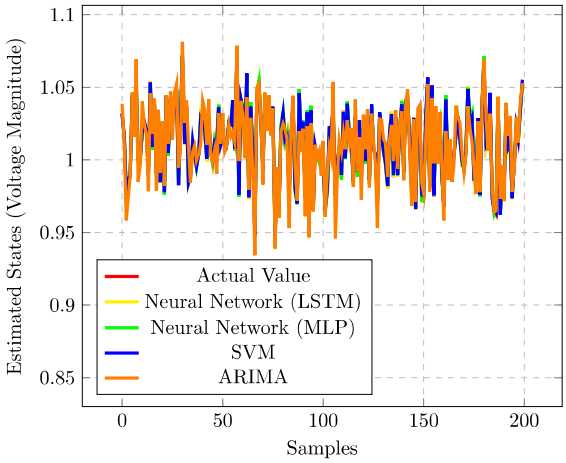

Figure 4.7: State forecasting performance of Voltage magnitude

Therefore, it is observed from Figs. 4.7 and 4.8 that the suggested nonlinear LSTM structure showcase an efficient forecasting of the estimated states i.e. voltage magnitudes and their respective phase angle up to nearly 14-time steps.

Table 4.2 indicates that the suggested nonlinear LSTM structure has minimal performance parameters (RMSE, MSE, and MAE) in contrast to other advanced state forecasting schemes namely MLP, SVM, and ARIMA, hence demonstrating a superior state forecasting policy. In comparison to the demonstrated nonlinear MLP model, the suggested robust, nonlinear neural network model encompassing LSTM modules demonstrates the least performance parameters with a subsequent enhancement in prediction.

4.5.1 FDIA identification

The proposed schemes for detection of FDIA as described in section 4.3 and 4.4 considers the IEEE 14-bus system as the test bench. It can be inferred from Fig. 4.9 that the proposed nonlinear LSTM model demonstrates a superior detection of FDIA of more

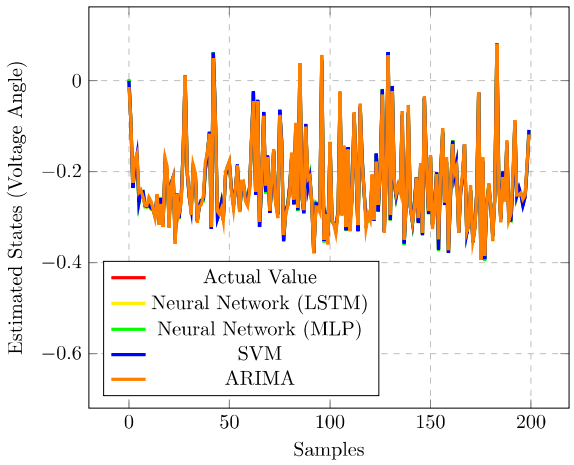

Figure 4.8: State forecasting performance of Voltage angle

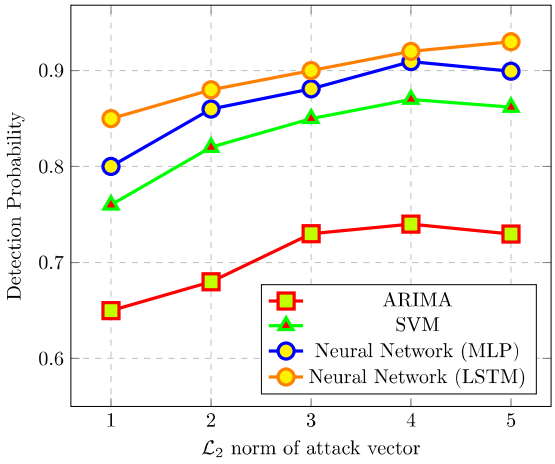

Figure 4.9: FDIA detection performance of the undertaken models for anomaly detection Scheme - I

than 92.5% by incorporating the anomaly detection scheme - I.

Furthermore, from Fig. 4.10 it can also be inferred that the recommended FDIA identification scheme - II can strategically detect the presence of attacks within the acquired measurements with an accuracy higher than 95%. Furthermore, it can be concluded from Figs. 4.9 and 4.10 that the proposed nonlinear LSTM model detects the presence of FDIA with a probability of nearly 92.5% and 95% for an attack vector having an $\mathcal{L}_2$ norm of strength equal to five times the unit ball in comparison to the nonlinear MLP model whose detection probability is furnished 91% with the effective implementation of the aforesaid anomaly detection algorithms. With a higher strength ($\mathcal{L}_2$ norm) of attack vectors as shown in Figs. 4.9 and 4.10, a higher FDIA identification has been achieved.

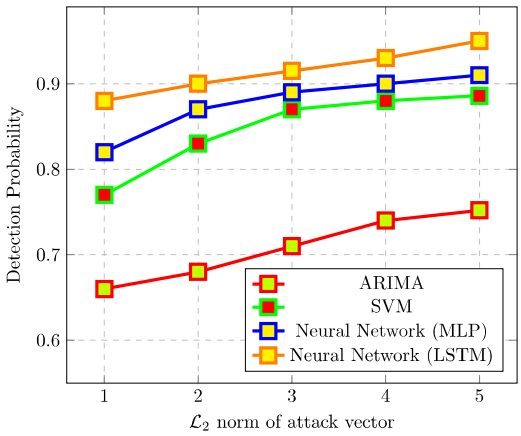

Figure 4.10: FDIA detection performance of the undertaken models for anomaly detection Scheme - II

The primary reason for such an enhanced detection probability can be outlined as that with a lower strength of attack vector formulation, noise within the raw measurements due to communication channels, meter failure, etc. cannot be effectively distinguished from attack.

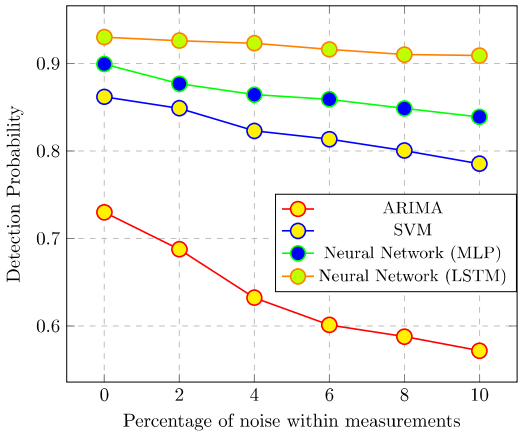

Figure 4.11: Variation in FDIA detection performance of the undertaken models for anomaly detection Scheme - I

A variation in the FDIA identification probability due to the presence of noise within the measurements can be seen as per Figs. 4.11 and 4.12 respectively. Here, Gaussian noise has been considered with standard deviation and mean of 1 and 0 respectively within the acquired set of measurements. It can be seen from Figs. 4.9 and 4.10 that the anomaly detection scheme -II which is based on the error covariance matrix demonstrates a higher

FDIA detection probability.

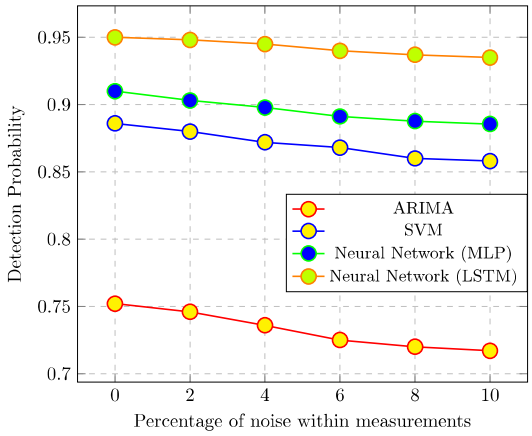

Figure 4.12: Variation in FDIA detection performance of the undertaken models for anomaly detection Scheme - II

It is observed that with an increase in noise, the detection probability of FDIA diminishes. It can also be concluded from Figs. 4.11 and 4.12 that the developed nonlinear LSTM structure demonstrates a minimum deviation in the FDIA detection probability under varying noise margins, hence can be seen as a robust FDIA detector. Furthermore, it can be seen from Fig. 4.11 that ARIMA demonstrates a significant deviation in the detection probability for the developed anomaly detection scheme-I under presence of noise.

4.5.2 Computational performance

The real-time performance of the proposed robust, nonlinear deep learning models is presented in this sub-section. It can be seen that the ARIMA model has the least training and testing times while giving a poor forecasting performance yielding to a slackly bound on the detection parameter. Although it shows the lowest detection accuracy with respect to the other considered models, nevertheless FDIA recognition can be performed quickly. Further observations suggest that SVM has a higher testing time than the nonlinear MLP and ARIMA for identifying FDIA effectively. Table 4.3 shows that the proposed nonlinear LSTM model has the highest time for training and testing to identify FDIA, but shows the least performance parameters, hence signifying a superior attack detection scheme. It can be said that all the models considered in this research for state forecasting and FDIA

detection in the smart grid can be easily executed in real-time as their computational burden lies in the range of μs while the sampling period of SCADA is approximately 100Hz. For large-scale power grids, the proposed attack detection strategies imbibing

Sl. No	Forecasting Model	Training Time (μs)	Testing Time (μs)
1	Deep Neural Network (LSTM)	770.34	558.77
2	Deep Neural Network (MLP)	200.45	110.34
3	Support Vector Machine	480.01	346.02
4	ARIMA	117.22	61.23

Table 4.3: Computational burden of the forecasting models

the nonlinear neural networks can also be implemented. The primary constraints for deploying them for large networks can be demonstrated as:

- For large-scale networks, primarily the training data for the nonlinear neural networks i.e. the time-series data of the state estimation algorithm is required. This large set of training data on the other hand may take a longer time for training the nonlinear neural networks. This will lead to a higher computational burden.
- Moreover, for large networks as training time for the neural networks is high, online training of the networks under varying load scenarios, system restoration etc. is not possible. This leads to a purely offline training scheme, thus losing real-time performance.

4.6 Summary

Although by implementing the proposed state forecasting-based FDIA detection scheme, efficient identification of the presence of attacks within the raw measurements can be furnished, still it can be seen that the operator faces difficulty in determining their respective locations of intrusions. Hence, to mitigate this issue and also to furnish a real-time robust FDIA identification scheme, the following chapter presents advanced deep learning models which can be effectively deployed as a real-time FDIA location and presence detector.