CNN-BiLSTM based GAN for Anamoly Detection from Multivariate Time Series Data

Sumit Kumar Singh
School of Computer Science and Electronic Engineering, University of Essex
Colchester, United Kingdom
ss20727@essex.ac.uk

Mohammad Hossein Anisi
School of Computer Science and Electronic Engineering, University of Essex
Colchester, United Kingdom
m.anisi@essex.ac.uk

Simon Clough
Soil Moisture Sense
Leiston, Suffolk
simonclough@soilmoisturesense.com

Tim Blyth
Soil Moisture Sense
Leiston, Suffolk
timblyth@soilmoisturesense.com

Delaram Jarchi
School of Computer Science and Electronic Engineering, University of Essex
Colchester, United Kingdom
delaram.jarchi@essex.ac.uk

Abstract*— Continuous recording of sensor data for monitoring applications does require detection of data patterns which deviate from normal condition. Detection of such events is necessary for implementing early preventive methods to improve overall system performance and potentially identify sensor failure causes. Recently, deep learning techniques based on generative models such as generative adversarial network (GAN) are proposed for anomaly detection from multiple time-series data. In this research, a variant deep learning method-based GAN is proposed for anomaly detection from multivariate time series data. Based on our proposed approach, the generator block consists of both CNN and BiLSTM blocks whilst the discriminator uses BiLSTM. To evaluate the performance of our new approach, multiple recordings from soil moisture measurement system are used to compare our proposed framework to the state-of-the-art techniques. Our proposed CNN-BiLSTM based GAN model presents an improved performance for the soil moisture recordings.

***Keywords*—Anomaly detection; GAN; BiLSTM; CNN; multivariate time series data**

I. Introduction

The increase in applications of Internet of Things (IoT) have generated enormous amount of time series data. These data points might capture irregular patterns or values due to sensor failure, lack of network communication or interruption of environmental factors. These irregular data points are termed as anomalies, detection of such anomalies provides an opportunity to diagnose and fix the primary issue before it can affect the quality of sequential data. Anomaly detection is also used to flag the possibility of sensor failure or network interruption that might cause such irregularity in the time series data. Detection of such behavioral patterns within a set time frame is essential as irregularity in the time series might behave like a normal pattern and thereafter can affect the time series prediction models.

Various algorithms of anomaly detection are classified into supervised, semi-supervised and unsupervised techniques [1]. When there are associated labels for the training and testing time series data, supervised anomaly detection algorithms are employed. The training set with limited data labels requires a semi-supervised way of detection. Moreover, several imputation algorithms for time series data are based on unsupervised machine learning model as the anomalies are not labelled and the data is exclusively marked based on its natural attributes. An extensive review on different types of anomalies is introduced by [2], where the outliers are classified as point, contextual and collective anomalies. The first type of outliers is those which are entirely different from the dataset. Data points which deviate from a populated area are classified as global anomalies. When the incoming values are differently marked from the historical values for the same context, it is termed as contextual anomaly. Values which are irregular with respect of overall dataset is termed as collective outliers. They are noticed in small data packets within a sequential collection.

Long short-term memory (LSTM) and convolution neural networks (CNN) are amalgamated which enable sequential learning to incorporate a temporal spatial sequence for better detection of cardiac arrhythmias [3]. In another study, stacked CNN-BiLSTM architecture is used for electrocardiography (ECG) analysis where CNN is employed for feature extraction whilst the BiLSTM is used for ECG waveform classification [4]. CNN creates a feature vector which can be used as input timestep for the architecture of BiLSTM. Generative adversarial networks (GAN) based on the fundamental principal of LSTM with variational autoencoders (LSTM based VAE-GAN) is introduced for anomaly detection application [5]. This method enables an accurate and fast mapping of real time data to latent space for detection of anomalies in the timeseries data. A GAN based anomaly detection algorithm is introduced for multivariate time series data in [6]. The method employs LSTM model with 100 hidden units for both generator and discriminator where there is slightly less model complexity for the discriminator. Autoencoders are also deployed for detecting anomalies in the data [7]. Vanilla LSTM has been used in [8] to predict the sequence based on learnt sequences in the data.

In this paper, we integrate deep learning approaches for robust anomaly detection from soil moisture measurements. Our framework uses CNN and BiLSTM in a GAN architecture to exploit both temporal and spatial patterns for anomaly detection. Our motivation behind our proposed model is to handle and process long time series data to achieve better performance for anomaly detection. BiLSTMs have the power to learn the relation between complex hierarchical features within high dimension of input data which increases the accuracy of prediction of next time series value for multivariate timeseries data. Additionally, CNN is employed for extraction of key features form the dataset. GAN based methods are widely used for detection of anomalies by training the discriminator block where the reconstruction loss is calculated between the real and predicted value. Therefore, combination

979-8-3503-3959-8/23/$31.00 ©2023 IEEE

of such complex architecture helps in accurate detection of anomaly in multivariate time series data which are partially correlated.

The remainder of the paper is as follows. In Section II, the proposed method is explained. This section briefly explains the basic deep learning algorithms used as core algorithms in our proposed model. Experimental results are provided in Section III. Section IV concludes the paper by highlighting the advantages and potential improvements of the proposed model for future studies.

II. Methodology

We propose a CNN-BiLSTM based GAN model for predicting the anomalies in multivariate time series data. Our model consists of a generator (G) and a discriminator (D) block. The generator has a 1D CNN module for extraction of features from long time series data and a BiLSTM module for generating sequential signals from noise. The output of the generator is sent to the discriminator block (which is a BiLSTM module) along with the sliced raw data; it flags the fake signals and produces an output as $D(x) \in \{0, 1\}$. During the training phase, we start by training the discriminator such that it can produce maximum accuracy for assigning accurate labels for generated and realistic data sequence. Thereafter, we train the generator to minimize $log(1\text{-}D(G(x)))$. The fundamental training function is described in Equation 1:

$$min(G_\theta)min(D_\emptyset)L_{\theta,\emptyset} = \frac{1}{n}\sum_{i=1}^{n}(logD_\emptyset(\chi_i) + (log(1 - D_\emptyset(G_\theta(x_i))))) \quad (1)$$

where number of points for each sequence is represented by n and $\theta, \emptyset$ are the generator and discriminator parameters.

We have used CNN for feature extraction in the generator block as it does not have any recurrent connection or forget gates, in contrary to the GRU or LSTM blocks. Therefore, the training of generator block is faster for long sequential time series data. CNN has proved to achieve high performance for generation of fake sequence which enables the discriminator to learn the features correctly during the training phase. CNN and LSTM are being used by several studies for detection of anomaly in time series data [3,4], however, we are using GAN based CNN and BiLSTM networks, which increases the precision of prediction of sequential data and therefore enabling an accurate detection of anomaly from long term time series data.

A. 1D-CNN module

CNN has been used for extraction of features from time series data and thereafter help the generator block to produce alike sequential series:

$$y_j^i = \partial\left(\sum_{m=1}^{N_{i-1}} conv(u_{m,j}^i, v_m^{i-1}) + b_j^i\right) \quad (2)$$

Where the m^{th} feature is represented as v_m^{i-1} and N_{i-1} stands for the number of features in $(i-1)^{th}$ layer, j^{th} feature of i^{th} layer is shown by y_j^i. The convolutional kernel is represented by $u_{m,j}^i$, $conv()$ stands for *1D* convolutional function with no padding, b_j^i is the bias for j^{th} feature map in i^{th} layer and ∂ stands for the activation function which is ReLU.

B. BiLSTM

To overcome the issue of vanishing gradient problem by recurrent neural networks (RNN), the special type of RNN was designed, i.e., LSTM. It has a memory cell to store the continuous sequence of data that is useful for processing the long time series data. A sequential processing of two LSTM layers is used to design a BiLSTM model where one of the layers is used for feed forward and the other for back propagation. Thus, it becomes more effective than simple LSTM.

An elementary framework of GAN has a generator block which takes random latent space as its input and produces unreal time series data and thereafter passes the sequence to the discriminator block which is employed for detecting the fake data series. In our proposed model for anomaly detection, we have used the generator and discriminator block for detection of anomalies in the time series data. Generator loss is being calculated by comparing the generated sample with the live time series data from the test sample based on real time mapping. The generator after its training is able to produce realistic samples as it is being mapped from random latent space to real data space which have a normal data distribution. It is trained for sufficient iterations such that the loss is very small compared to the real time series data for training and any prediction of time series data can be confidently considered as true value. Therefore, any significant change in real-time data can be compared with the predicted value (which is treated as

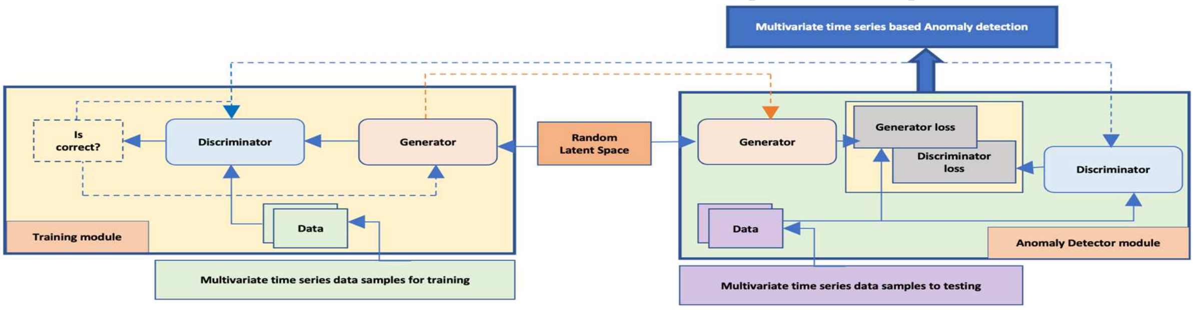

Fig. 1. Pictorial illustration of the proposed model (CNN-BiLSTM GAN)

true value) and flagged as an anomaly. On the other hand, discriminator loss is calculated by classifying live test data based on the learnt weight and biases of the discriminator module. Thereafter, both the losses (generator and discriminator loss) are being combined to check potential imputations in the data. Higher value of generator loss flags a potential anomaly in the data, while a higher value of discriminator loss shows that the test samples are grouped with true labels. Therefore, cross entropy loss is calculated to be consistent with both the losses and higher value increases the chance of anomaly in the data. The anomaly loss is mapped to the original data to calculate the detection loss in a sub-sequence, this decreases the chance of any false alarms in the data. Our proposed method is illustrated in Figure 1.

C. *Hyperparameter tuning*

Rolling window technique is used to split the train and test data, where the model is trained on the data within the window and tested on consecutive point on the timeseries. We have employed this method as it allows the model to capture the temporal dependencies of the data. Bayesian optimization framework along with the upper confidence bound (UCB) is used as an acquisition function for fine tuning the hyperparameters. The model is trained with a learning rate of 0.001, optimizer = 'Adam' and loss function = 'categorical_cross_entropy'. Leaky-ReLU is used as the activation function in the model and layer-based normalization is applied to fasten and support the learning rate by lowering the loss function. Activity regularization layers can be used to learn weak features in the time series data as it upgrades the input layer which is depending on the devised cost function. RNN is inclined towards overfitting and requires notable modification in the training dataset therefore we have added dropout layers during the training period which decrease the possibility of overfitting.

III. Experimental Results

The proposed model is trained and evaluated using soil moisture data recordings from multiple sensors. Multiple time series data are used as input to train the proposed model and evaluate their performance against state-of-art deep learning-based techniques for anomaly detection.

A. *Dataset*

In this paper a large-scale sequential time series data based on soil moisture recordings has been used. The data is being provided by the Soil Moisture Sense, Ltd, Leiston, UK. The system consists of various sensors such as soil moisture sensor, gauge to measure irrigation applied and run off, temperature, vapor pressure deficit (VPD), relative humidity (RH), potential of hydrogen (PH) and electric conductivity (EC) sensors. These systems are mounted in tunnels where soft fruits are grown. Due to the potentially rapid changes in the environment within the tunnels, it is important that the data is accurate so sound decision making can be made to maintain the optimal growing conditions. Therefore, it is critical to detect any anomalies in the data. Decisions based on anomalous data could have a significant effect on crop yield. The dataset consists of 17 features for 416 such auto systems recorded over a period of 3 years. With the time instance of 157,680 time series data points for each system. Figure 2 shows the historical trend of various sensor recordings for a system mounted in a tunnel, which is continuously recording data throughout the year at the frequency of 10 minutes.

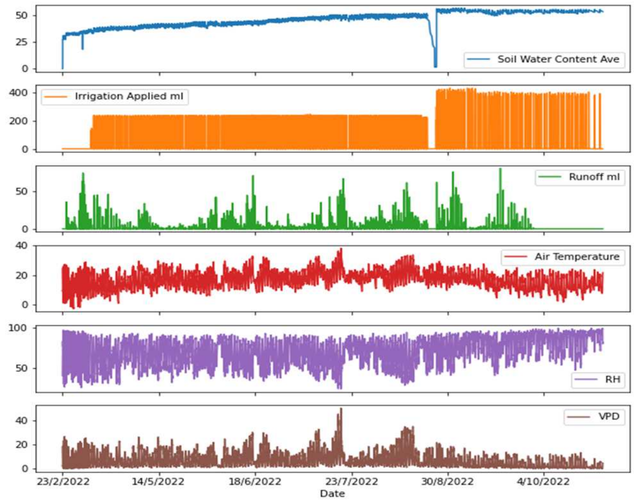

Fig. 2. Historical trend of various sensors for a system mounted in a tunnel

B. *Evaluation Study*

Various evaluation metrics such as root mean square error (RMSE), mean absolute error (MAE), cross entropy error (CEE) and coefficient of determination (R^2) are calculated. The proposed model outperforms against LSTM, BiLSTM, VAE-GAN and MAD-GAN. These methods are the state-of-the-art techniques for detection of anomaly in time series data. Our proposed method presents improved performance across all evaluation metrics using multivariate the time series data for soil moisture recordings. The computing platform used for the experiment is Intel i7 with 16gigabytes of RAM and Nvidia GeForce GPU. Ubuntu 20.04 is used as operating system. The proposed model is implemented, trained, and evaluated using Python 3. We have imported TensorFlow, NumPy, pandas and matplotlib for executing all required tasks. The performance evaluation parameters that are used to check the accuracy of anomaly detection in a time series data are presented in the following equations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i^{Predicted} - Y_i^{Actual}\right| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4}$$

Where, x_i is predicted value at i^{th} term and y_i is the actual value at ith term in a sequence of n values.

$$CEE = -(ylog(p) + (1-y)log(1-p)) \tag{5}$$

Where, y is the binary indicator of class and p is the predicted probability.

$$R^2 = \frac{\left(\sum_{i=1}^{n}(x_i - \hat{x})(y_i - \hat{y})\right)^2}{\sum_{i=1}^{n}(x_i - \hat{x})^2 \sum_{i=1}^{n}(y_i - \hat{y})^2} \text{ of} \tag{6}$$

A lower value of MAE, RMSE, and CEE shows improved performance while a higher value of R^2 is expected for an improved performance.

The proposed GAN model based on CNN and BiLSTM for anomaly detection outperforms LSTM, BiLSTM, VAE-GAN and MAD-GAN by evaluation on the soil moisture dataset. Table 1 illustrates a comparative evaluation of state-of-art techniques against our proposed method for detection of anomaly in long sequence time series data. Graphical representation of the performance of proposed method against the state-of-art techniques is represented in Figure 3.

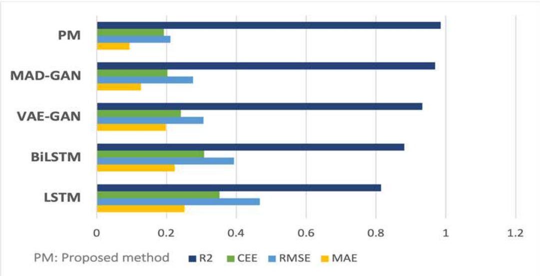

Fig. 3. Graphical representation of performance evaluation metrics for various state-of-art techniques for anomaly detection

TABLE 1. COMPARATIVE EVALUATION OF DIFFERENT DEEP LEARNING BASED ANOMALY DETECTION ALGORITHMS

Model	MAE	RMSE	CEE	R^2
LSTM	0.251	0.467	0.352	0.814
BiLSTM	0.223	0.393	0.308	0.881
VAE-GAN	0.198	0.306	0.241	0.933
MAD-GAN	0.127	0.276	0.203	0.969
CNN-BiLSTM GAN (Proposed)	**0.094**	**0.211**	**0.192**	**0.985**

IV. CONCLUSION

A new approach method for anomaly detection is proposed in this paper. The method is based on the fundamentals of GAN, where the generator block consists of CNN and BiLSTM modules for generation of "fake" data series and the discriminator block is based on BiLSTM for differentiating the real sequence from the generate sequence. The anomaly detector module is generating accurate sequence of time series data from random latent space and calculate generator loss against the real sequential data, similarly discriminator loss is being calculated when the discriminator is able to predict labels for the raw test data. Thereafter cross entropy error is calculated to detect a potential anomaly in the sequential data. The proposed method outperforms all the evaluation parameters when compared against state-of-art techniques such as LSTM, BiLSTM, VAE-GAN and MAD-GAN. In future studies, we aim to further optimize the hyperparameter to achieve more enhance results for our proposed method incorporating new datasets related to soil moisture monitoring.

ACKNOWLEDGEMENT

This research was funded by a Knowledge Transfer Partnership (KTP) from Innovate UK (Partnership No: 13298), between Soil Moisture Sense and the University of Essex '.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection for Discrete Sequences : A Survey," vol. 24, no. 5, pp. 1–16, 2010.

[2] A. Feroze, A. Daud, T. Amjad, and M. K. Hayat, "Group Anomaly Detection: Past Notions, Present Insights, and Future Prospects," SN Comput. Sci., vol. 2, no. 3, pp. 1–27, 2021, doi: 10.1007/s42979-021-00

[3] P. Warrick and M. N. Homsi, "Cardiac Arrhythmia Detection from ECG Combining Convolutional and Long Short-Term Memory Networks," vol. 44, pp. 1–4, 2017, doi: 10.22489/CinC.2017.161-460.

[4] S. Nurmaini et al., "Beat-to-Beat Electrocardiogram Waveform Classification Based on a Stacked Convolutional and Bidirectional Long Short-Term Memory," IEEE Access, vol. 9, pp. 92600–92613, 2021, doi: 10.1109/ACCESS.2021.3092631.

[5] Z. Niu, K. Yu, and X. Wu, "LSTM-based VAE-GAN for time-series anomaly detection," Sensors (Switzerland), vol. 20, no. 13, pp. 1–12, 2020, doi: 10.3390/s20133738.

[6] D. Li, D. Chen, L. Shi, B. Jin, J. Goh, and S.-K. Ng, "MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11730 LNCS, pp. 703–716, 1 2019.

[7] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. Part F1296. New York, NY, USA: Association for Computing Machinery, 8 2017, pp. 665–674. doi: https://dl.acm.org/doi/10.1145/3097983.3098052

[8] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long Short Term Memory Networks for Anomaly Detection in Time Series," vol. 89. Presses universitaires de Louvain, 2015, pp. 8