A HYBRID CNN & BiLSTM WITH AM MODEL TO PREDICT AND ESTIMATE LOAD HARMONICS AT NESTLE EAST LONDON SOUTH AFRICA.

Excellence Kuyumani[1], Thokozani Shongwe[2], and Ali Hasan[2]

[1]University of Johannesburg Faculty of Engineering and Built Environment
[2]University of Johannesburg - East Rand Campus

October 31, 2022

Abstract

Prediction of electrical power harmonics promotes development and provision of clean power supply. However, the elevated frequency combined with noise of harmonics make precise forecasting challenging and demanding. Due to the stochastic nature of harmonics occurrence and the challenge of attaining a dependable, and efficient working model, to this date research efforts have not been sufficient. Although various statistical and machine learning algorithms have shown very interesting and promising results, but work is still being done on algorithms that would produce the least possible error. This research uses hybrid convolutional neural network (CNN) and bidirectional long short term memory (BiLSTM) with Attention Mechanism (AM) model, to predict load harmonics at Nestle, a confectionary manufacturing plant in South Africa. Historical load harmonics data is used as the dataset. The results show that the hybrid deep machine learning method CNN – BiLSTM – AM has better performance compared to other five prediction models when detecting and forecasting load harmonics. The hybrid algorithm has a prediction accuracy of 92.3569 % and the lowest RMSE of 0.0000002215.

1 Introduction

The wide usage of variable speed drives (VSDs), rectifiers and inverters create non-linear loads in a power network resulting in non -sinusoidal current that is nonlinearly related to the supply voltage. The non-sinusoidal current is a periodic function, and it can be decomposed into its fundamental sine wave plus various other sine waves of harmonic frequencies. Harmonics are an integral of multiples of the fundamental frequency. Thus, non-sinusoidal current may have both odd and even harmonics. Harmonics have a negative impact on the reliability, stability and safety of a power network. They may cause power supply interruptions, abnormal grounding protection, shortens equipment life and overheating of conductors as well as equipment [1]. Both current and voltage harmonics are non-linear, non-stationary, dynamic, complex, noisy, unstable and are classified as time series. These characteristics make harmonics difficult to predict.

Machine Learning techniques model complex and non-linear problems better than statistical techniques. Timely and accurate detection of harmonics is of vital importance in maintaining reliability and efficiency in a modern manufacturing plant. Harmonics cause steady-state waveform distortion and severely degrade the power quality [2]. A reliable and accurate harmonics forecasting method is a primary step in an efficient planning of an electrical network. Power network efficiency is characterized by its reliability and system reliability is a function of cost-saving. The widespread use of electronic converters and inverters contributes to distorting both the voltage and current waveforms resulting in harmonics. Statistical and machine learning algorithms have been used in forecasting power consumption, fault prediction on machines and number of time series data, but work is still being done to find more accurate algorithms for the prediction of harmonics in an electrical power network [3], [4],[5],[6] and [7].

Six models are used to predict the load harmonics and are separately trained and tested using the historical harmonics data. The independent results are compared in order to determine the most accurate model and root mean squared error (RMSE) is used as the performance indicator. The paper's contribution is to: 1) propose an innovative method of harmonics detection and prediction based on a hybrid CNN-BiLSTM-AM model, 2) compare CNN-BiLSTM-AM model harmonics forecasting capabilities with six other deep learning methods, 3) demonstrate that the proposed CNN-BiLSTM-AM network has superior performance in detecting and forecasting of harmonics. The paper is organised as follows; Section 2 covers related work done so far in the field of study. Section 3 outlines the literature underpinning the three DL algorithms used in this paper. Results are discussed in section 4 followed by the conclusion.

2 Related works

Fourier transforms (FFT), zero-crossing method, least squared method (LSM), Prony method and Time-Domain Quasi-Synchronous Sampling method have traditionally been used in harmonics measurements [8], [9], [10], [11], [12], [13], and [14]. However, most of these methods are challenged when they are applied to big data. FFT has problems associated with spectrum leakage and picket fencing resulting from nonsynchronous sampling. These challenges are handled better by using LSM based on interpolation method resulting in improved accuracy [15]. Most of these methods extract features in the time and frequency domains and associated challenges include 1. Harmonics and many other power quality disturbances have almost similar features, thus leads to poor feature selection. 2. Features captured in the time and frequency domains do not describe harmonic pollution, thus leading to poor selection accuracy. 3. Feature extraction process demand high level understanding of harmonics characteristics leading to complex feature extraction [16].

Time series prediction techniques fall into two categories, namely traditional time series methods and forecasting methods based on machine learning. Traditional time series techniques deal with specific models to describe time series. The noisy and complex features of time series data cannot be determined using analytical equation with parameters, since the dynamic equation for time series data is either unknown or extremely complex. These ML techniques require low-end hardware [17] and [18]. The power network management, through diagnostic and prognosis, of electrical power network harmonics through several techniques is enhanced using Machine Learning (ML) and deep learning (DL) techniques that are bioinspired mathematical models [19]. ARIMA, support vector machines (SVR), logistic regression, decision trees, Random Forest etc are some of the commonly used machine learning time series methods and they perform well on small data [20].

In comparison with signal processing methods for feature extraction, DL is a data-driven technique, and has capability to solve prediction challenges associated with big data. DL algorithms have made significant contribution in solving power quality disturbance issues [21]. DL technique, a branch of ML, have the advantage of handling nonlinear, non-stationary and high-dimension big data and are also superior in data-driven prediction [19]. The most popular DL methods are recurrent neural networks (RNN), LSTM, BiLSTM, and CNN. These algorithms have densely connected neurons resulting in high learning and generalisation capabilities [22]. Deep recurrent neural network with long short term memory (DRNN-LSTM) was used to forecast an hour ahead solar panel output and the load [23] and [24].

RNN is a deep neural network (DNN) is a developed and advanced artificial neural network (ANN) that is sequence-based with the capability of learning and extracting features from inputs in time series domain. RNN is a cyclic neural network and has a special network architecture and is shown in fig 1.

Fig 1: RNN structure

The neuron output can act directly on itself as an input at the following time point such that the output of the network at this instant comes from the interaction between the input of a particular time and all the times in memory. Thus, it effectively allows signals to move forwards and backwards. It is a regression algorithm for data in the time domain and has successfully been used in wind speed prediction and current control [25], [26] and [27]. Generally, RNN algorithms maybe of the types: one input to many outputs, many inputs to many outputs, and many inputs to one output. RNN is challenged when dealing with long term dependences and leads to vanishing gradient. LSTM is a specialised RNN capable of dealing with these long term dependencies challenges and has shown significant performance when dealing with time series data, and forecasting [28]. RNN and particularly LSTM has been used in feature sequence extraction, for example transient period, as well as data classification in power applications including fault diagnosis of photovoltaic, wind turbines with multivariate time series, transmission lines, and prediction of fault location distance in two-bus line test system of 220 km [29]. LSTM is an assumption-free algorithm with the capabilities of handling complex nonlinear dynamic data in noisy and higher dimensional space. A single LSTM commonly referred to as just LSTM processes data in one forward direction while the Bi-LSTM consists of at least two LSTM layers structured in a way that one is in the forward direction and the other in the backward direction and thus improve the prediction accuracy [30].

CNN has been used extensively in image recognition and is increasingly being used in time series data prediction [7]. CNN has been applied in various electrical power systems relating to prediction and classification problems varying from harmonics, transients, islanding, instability etc. CNN automatically extract system features and is easy to train compared to other neural networks having several hidden layers. CNN can also be used for 1-D inputs like most power systems problems and relies mainly on the convolutional and dense layers rendering the pooling layer less significant, and has been used successfully in a number of power systems related classification and prediction problems [31]. Fig 2 shows CNN 1-D architecture.

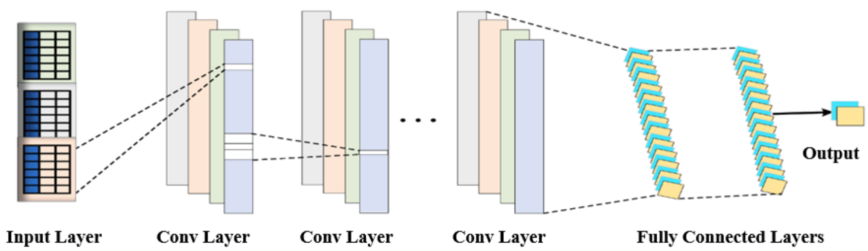


Fig 2: CNN 1-D Architecture

CNN combined with AM focuses on the specific features and leaves unnecessary features and boost the desired information, and thus enhancing the feature selectivity of the chosen model [22]. CNN-BiLSTM with Attention Mechanism (AM) has been successfully applied in a multivariate time series for two-phase flow pattern identification. AM was used to select the highest values of the small vectors as the vital features of the small vectors that could not be detected and selected by CNN-BiLSTM layers. These vectors would be combined in the n-dimensional vector to give a vector of each medium-sized vector [32]. Hybrid LSTM with AM model has been used in residential load forecasting and showed impressive results. AM is well-suited for demand side forecasting methods involving LSTM. LSTM cannot pick up inner correlation among the hidden features that have significant impact on the forecast results. Thus AM is deployed to mitigate this weakness by adaptively weighting the hidden features [33].

3 Methodology

The idea of deep learning (DL) was developed from the study of artificial neural networks (ANN), a shallow learning algorithm. DL has multi-hidden layers and can perform a layer-by-layer transformation of data features thereby extracting the most effective information from the data. DL is a data-driven algorithm, and it does not need to construct a physical model. It only needs historic data in order to extract optimal features of the power network and subsequently do tasks such as harmonics/fault diagnosis, harmonics/fault classification and prediction.

The goal is to combine harmonic data from various VSDs to determine if a VSD has produced harmonics. If so which one, then we can build a DL machine model that can forecast production of harmonics, reducing the risk of machine break down. Supervised learning typically requires iterations of building and evaluating different models. Fig 3 shows that the data is initially split into training and validation data. The process begins with the harmonics data set where the training data has a known result. The model root mean square error (RMSE) loss is checked if it is stable. If the RMSE error rate is not stable the model will retrain otherwise the model will be validated and the training ends. Once the model is trained, its performance is evaluated by typically comparing model predictions and known responses.

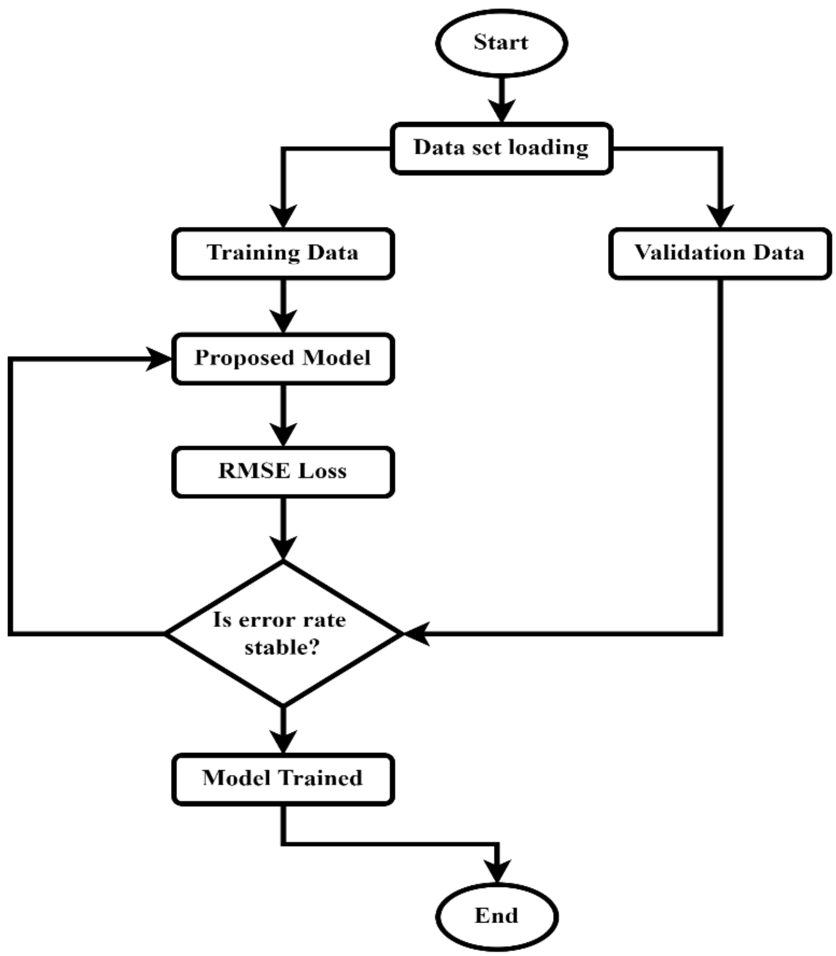
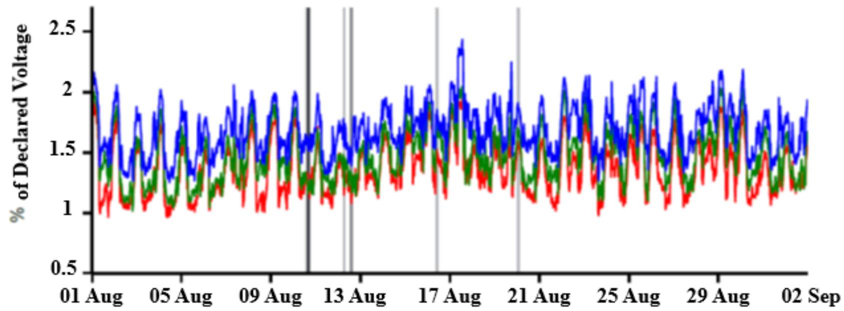


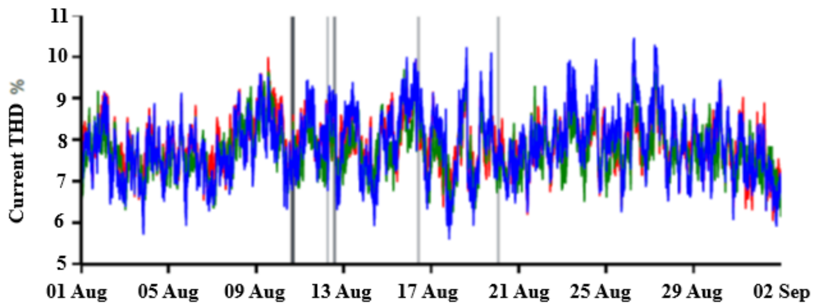
Fig 3: Flowchart

Data Analysis and problem formulation

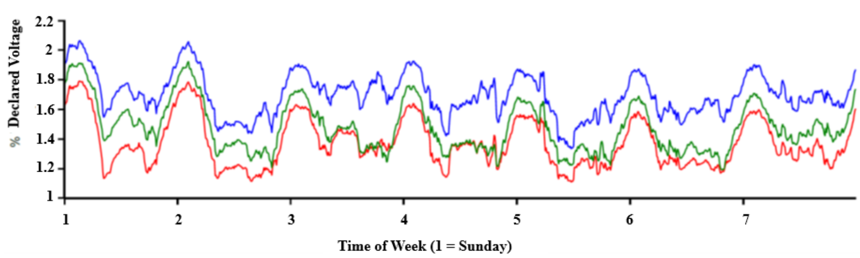
The study is conducted using historical data collected at Nestle, a chocolate manufacturing plant in East London, South Africa. The factory uses both variable speed drives (VSDs) and variable frequency drives (VFDs) for speed control of DC and AC motors and variable source inverters (VSI) in uninterruptible power supplies (UPS) units. Modern VSDs generally use pulse-width modulation voltage source inverters (PWM VSI) to control speed of three-phase induction motors. The advantages of PWD VSI are extremely good dynamics, peak current protection, instantaneous current waveform control, better accuracy, compensation of load parameter changes effect, semiconductor voltage drop compensation, and dc link and ac-side voltage changes compensation. The VSDs and VFDs have numerous advantages; to match motor and load characteristics, energy saving speed and position control, reduction of stresses as well as transients caused by ON/OFF and sudden motion operations. VSDs may decrease the quality of the alternating current (AC) to which they are connected significantly, and they pull non-sinusoidal current from the main power AC supply thus leading to the generation of high-order harmonics in the power supply. These harmonics are reduced by direct current (DC) chokes in order to maintain the levels below the International Electro-technical Commission (IEC) 62635 maximum standard measured total harmonics distortion (THD) index. The sample THD data that is used in this paper is shown in fig 4 (a) to (d). The objective of the prediction is to accurately manage both the response and reaction time delays of active power filters (APFs) designed for a manufacturing plant with largely nonlinear and stochastic loads.



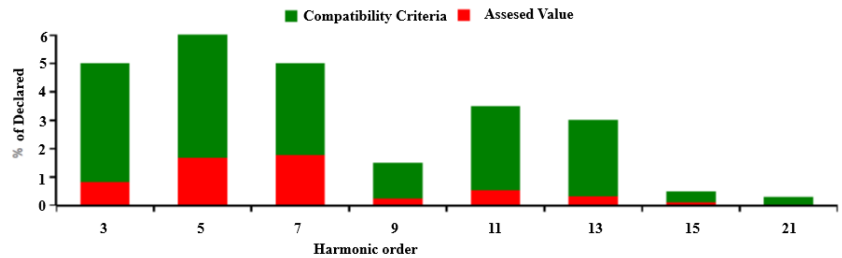
Declared Voltage reading



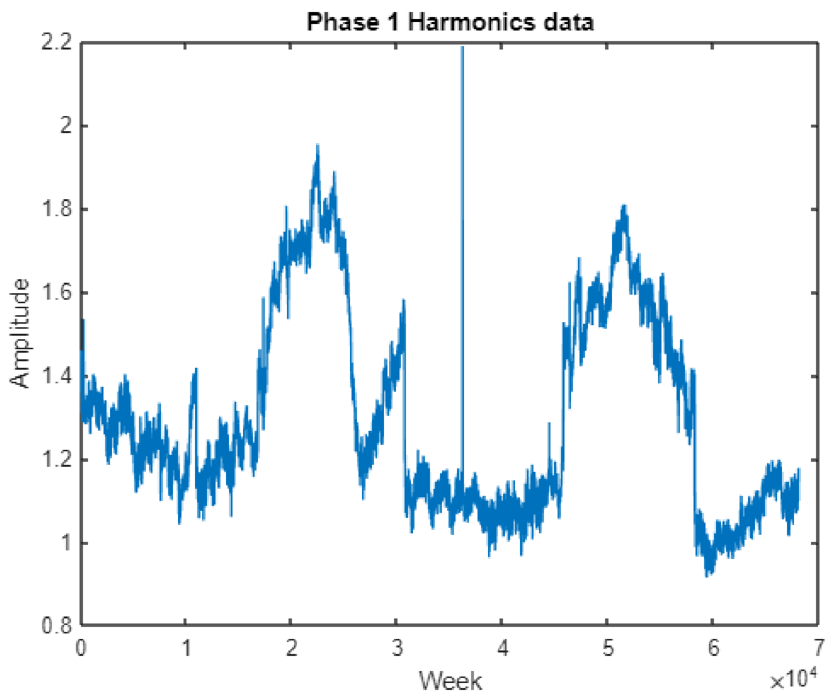
10 min Current THD reading

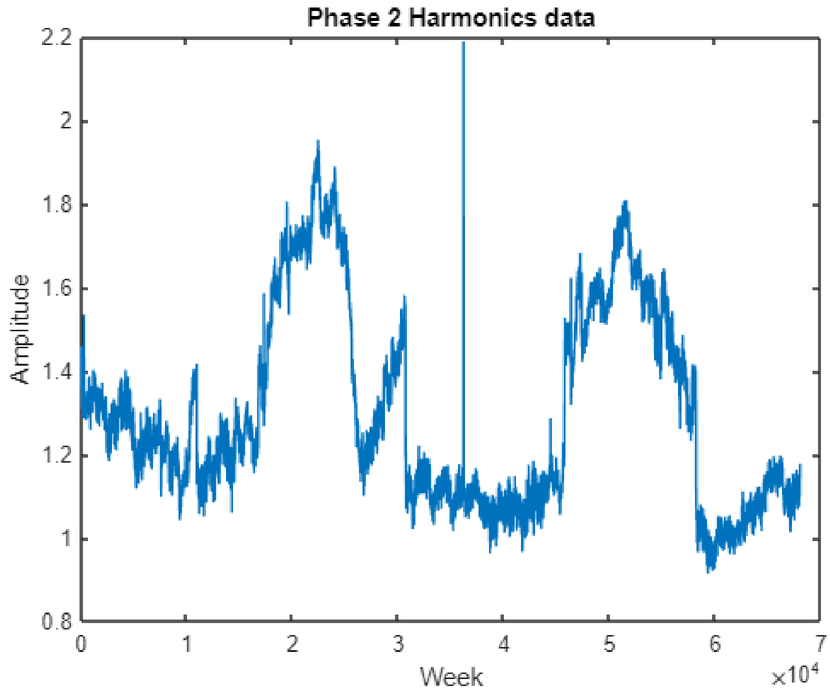


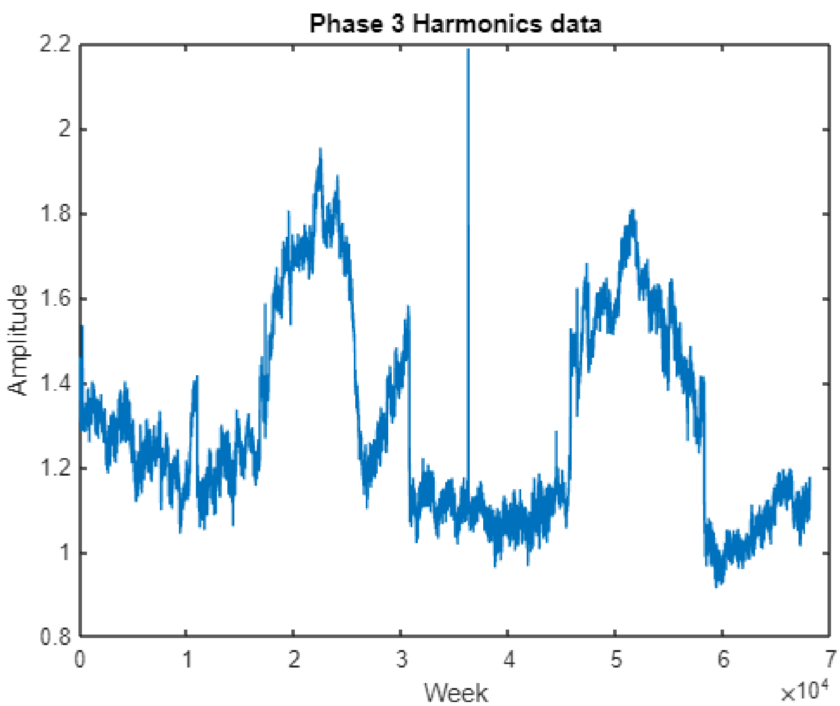
Voltage THD Profile – a typical week



Harmonics assessment







Phase harmonics

Fig 4: Measured harmonics

3.2 Deep Learning Algorithms

3.2.1 Bi-directional LSTM

BiLSTM architecture is best explained using the LSTM architecture and as the fundamental building block. LSTM was developed to improve RNN weaknesses in dealing with long term dependences. The RNN model structure has a return loop that uses prior information efficiently. However, RNN model has limited memory and does not learn well long term dependences and as a result that leads to vanishing gradient. The vanishing gradient phenomenon is when the input information or the gradient passes through several layers, it vanishes and wash out when it reaches the end. The consequence of this is that it makes the RNN struggle to capture long-term dependencies and makes the RNN training much harder as the training algorithm assigns smaller values to the weights. When this happens the RNN algorithm stops learning. Another challenge associated with this phenomenon is "exploding gradients". Exploding gradients refers to situations where the input information or gradient passes through multiple layers and by the time it reaches the end the gradient is very large making the training of RNN challenging. This challenge makes the training algorithm to assign bigger values to the weights. In mitigation the gradient can be truncated. The gradient is mathematically expressed as a partial derivative of the output of a function with respect to its input. Thus, the gradient determines how much the output changes with respect to the changes affecting the input. Gated Recurrent Unit (GRU) shown in fig 5 is used to mitigate the RNN gradient problem.

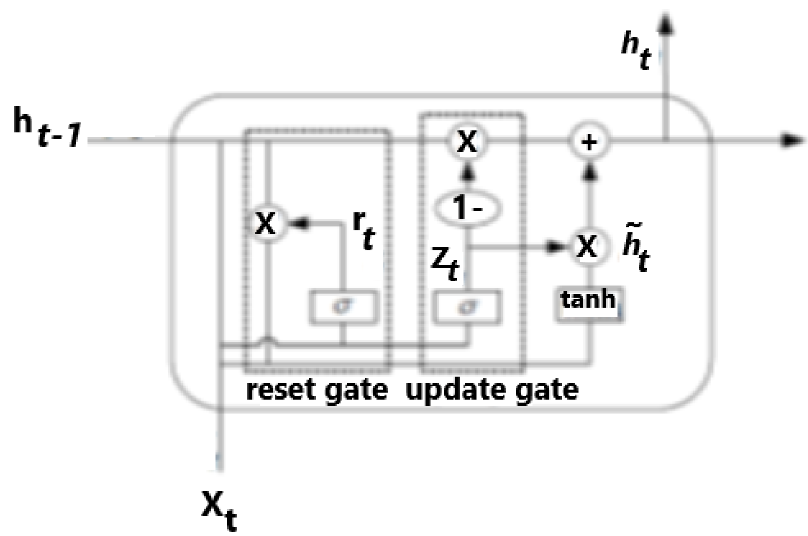


Fig 5: GRU structure

The input is x_t; reset gate isr_t; h_t,$h_{t-1}$is hidden layer state output at the current and previous time; the update gate is z_t;tanh is activation function;σ is a sigmoid function; $\tilde{h}_t$ is the active state of a hidden layer at the current time. The GRU has been used in a hybrid model with AM in order to improve its performance [34]. LSTM was developed based on GRU and is explained below.

LSTM was developed to deal with RNN limitations [35]. The LSTM model has memory cells that are used to remember long-term historical data and an LSTM cell architecture is shown in fig 5. The data in an LSTM cell is regulated through three gate mechanisms namely the input gate i_t, forget gate f_t and output gate o_t. The start of the memory cell is performed through point-wise multiplication and sigmoid function operations. The input information x_t at the current state together with h_{t-1}form the hidden state of the previous LSTM cell/layer are fed into all the gates and transitioned by the sigmoid and tanh functions. The forget gate determines data that must be kept or rejected, and its output is anywhere between zero and one. If the forget output is leaning towards zero, then it means the information has been ignored and if the information is close to one then the data has been kept.

Hosted file

`image11.emf` available at https://authorea.com/users/518360/articles/592543-a-hybrid-cnn-bilstm-with-am-model-to-predict-and-estimate-load-harmonics-at-nestle-east-london-south-africa

Fig 5: LSTM cell architecture

The forget operation can be expressed mathematically as

$f_t = \delta(W_f.[\,h_{t-1},\ x_t] + \ b_i)(1)$

Where δ is the sigmoid activation function, W_f is the weight and b_i is the bias. The vectorial summation ofh_{t-1} and x_t is fed into all gates through the sigmoid and tanh activation functions. The input gate decides which data to update. The input gate is expressed as

$i_t = \delta(W_i.[\,h_{t-1},\ x_t] + \ b_i)(2)$

The next step is to feed the current input state x_t and hidden state h_{t-1}into the tanh function. At this stage the cell state is determined, and the resulting value update the cell state, and is expressed by the equation.

$\tilde{C}_t = tanh(W_c.[\,h_{t-1},\ x_t] + \ b_c)(3)$

Where tanh is the activation function. The new memory stateC_t and new hidden stateh_t are passed on to the next state. C_t is determined by the equation.

$C_t = \ f_t \odot C_{t-1} + \ i_t \odot \tilde{C}_t(4)$

$o_t = \delta(W_o.[\,h_{t-1},\ p_t] + \ b_o)(5)$

$h_t = \ o_t \odot \tanh(C_t) \quad (6)$

A single LSTM layer processes data in one direction while the Bi-LSTM has two LSTM layers. One layer processes data in the forward direction and the other layer in the backward direction as shown in in fig 6. The LSTM layer in the forward direction receives the input past data information and the reverse LSTM layer obtains the input sequences' future information, then the output in both hidden layers is combined. The hidden state of Bi-LSTM h_t at time t contains forward$\overrightarrow{h_t}$ and reverse $\overleftarrow{h_t}$ and is expressed by

$h_t = \ \overrightarrow{h_t} \bigoplus \overleftarrow{h_t}\ (7)$

Where $\bigoplus$ is summation on the forward and reverse components. Bi-LSTM has better performance than LSTM and RNN because it uses both preceding and subsequent information.

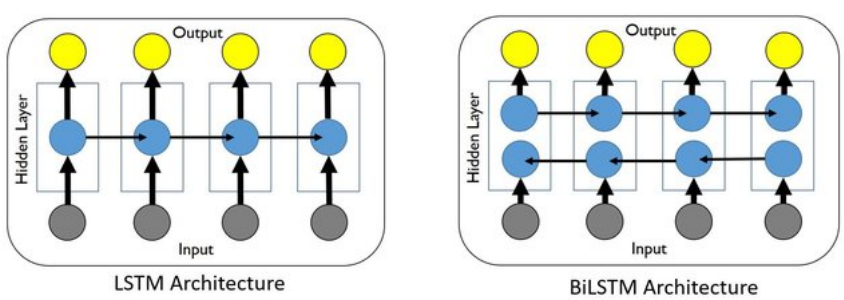


Fig 6: LSTM and Bi-LSTM architecture compared

Convolutional neural network

CNN is a deep learning model that has predominantly been used in image recognition and is increasingly being used in time series data prediction [20], [36] and [37]. CNN is usually used in modelling complex nonlinear systems that handle multi-dimensionality aspects. For example, an image I , a 2D convolution manipulation is expressed by

$$H_F\left(k,i,j\right) = \sum_{u=U}^{U}\sum_{v=V}^{V} W\left(k,u,v\right).\mathrm{I}\left(i+u,\ j+v\right) + \ b_k$$

where *(i, j)* is the pixel coordinate vector, *I(i + u, j + v)* is the image intensity at *(i + u, j + v)* , W is weight and$b$ is bias. W and b are learned during training phase. $H_F\left(i,j\right)$ represents the convolution kernel feature at position *(i, j)* and k is number of convolution kernels. During the training phase the input layers are updated with the back propagation learning process which results in internal covariant transformation. Batch normalisation is used to mitigate this problem. A non-linear down sampling process, maximum pooling, is used to decrease the data dimension at the same time retaining the invariability of image translation, expansion and rotation [38].

Fig 7 shows CNN architecture consisting of stacking up the convolutional, pooling and fully connected layers. The convolution layer has filters used to automatically extract features from the input matrix. The filters perform convolution processes through weight sharing and thus reduce the complexity of the computational operations and increases the network performance [39]. The pooling layer performs the down sampling process, process of reducing the feature dimensions and avoiding overfitting. The fully connected layer is used to learn the nonlinear combinations of features and produces the output [40]. CNN algorithms have problems in learning relationships between time series data. A combination of RNN and CNN is used to mitigate this challenge.

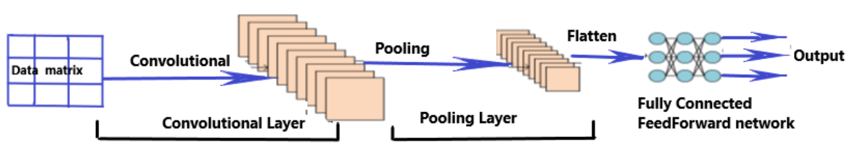


Fig 7: CNN structure

Attention mechanism

Attention Mechanism is used to upgrade and make desired features outstanding and is inspired by the human visual system. When the human vision system observes a scene, it focuses on detail as needed. Thus, it does not observe the whole scene from start to finish. Therefore, AM selectively focuses on important feature information, ignores unnecessary feature information and enhance the desired feature information. AM is predominately used in image captioning and is increasingly applied to other fields including time series data, like in this research. AM process is illustrated in fig 8.

Hosted file

`image14.emf` available at https://authorea.com/users/518360/articles/592543-a-hybrid-cnn-bilstm-with-am-model-to-predict-and-estimate-load-harmonics-at-nestle-east-london-south-africa

Fig 8: Attention Mechanism process

AM process is based on weight allocation determining the key and important information through distribution of higher weights. It has three stages and is usually used after the CNN and Bi-LSTM networks to focus sharply on features that contribute greatly to the desired output. As a result, the model performance is improved. The attention function can be defined as a mapping from a Query to a sequence of Key-Value. Similarity between the Query and each Key is computed in the first stage using the mathematical expression.

$s_t = tanh(W_t h_t + \ b_h)$ (8)

Where s_t is the attention score, W_t is the weight,h_t is the input vector and b_h is the bias. At stage two, the score obtained in stage one is normalized and softmax function is used to convert the attention score using the mathematical expression.

$a_t = \ \frac{e^{(s_t)}}{\sum_t e^{(s_t)}}$ (9)

The final attention score is computed at stage 3 using the mathematical expression

$s_t = \ \sum_t a_t h_t$ (10)

3.2.4 CNN-Bi-LSTM-Attention Model

Due to the stochastic nature of harmonics, the CNN and Bi-LSTM networks are used to fully extract the spatial and temporal features of the harmonics and AM is used to focus on the features that have strong correlation with harmonics making the prediction performance better. Firstly, the harmonics data is pre-processed. Secondly the CNN is used to capture the spatial dependences on the harmonics in order to reduce the dimension of the harmonics data so that effective spatial information is used. Thirdly the temporal dependencies are captured using Bi-LSTM. Fourthly the AM is used to highlight important harmonics features in order to achieve satisfactory prediction performance. Bi-LSTM has a good forecasting accuracy, that would capture the future and past harmonics data simultaneously. Reverse relationship of the data is considered in order to predict the long term and short-term production of stochastic harmonics efficiently. Fig 9 shows the model flowchart and layout. The input vectors$w_1 w_2 \ldots\ldots\ldots w_{n-1} w_n$ are fed into the CNN layer, the output is the input to the BiLSTM connected to the attention mechanism that is finally connected to the output.

Hosted file

`image15.emf` available at https://authorea.com/users/518360/articles/592543-a-hybrid-cnn-bilstm-with-am-model-to-predict-and-estimate-load-harmonics-at-nestle-east-london-south-africa

Hosted file

`image16.emf` available at https://authorea.com/users/518360/articles/592543-a-hybrid-cnn-bilstm-with-am-model-to-predict-and-estimate-load-harmonics-at-nestle-east-london-south-africa

Fig 9: CNN-Bi-LSTM-Attention model flowchart and layout

Hybrid combination of CNN-LSTM

The main motivation in using the hybrid models on time-series data is that CNN models are good at extracting the important features and is also able to filter out noise that maybe present in the input and LSTM models are used to capture the sequence pattern information in the input data. Although CNN models are employed to extract patterns of local trend and similar patterns in other sections of the time-series data, they are not normally good at dealing with long temporal dependencies. This CNN weakness is taken care of by LSTM as well as dealing with temporal correlations using attributes in the training data [4]. The hybrid model exploits strengths of both deep learning models resulting in improved forecasting accuracy.

Performance Indicators

Widely used deep learning algorithms performance indicators are root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and mean square error (MSE), and they are given in equations (11) to (14). Although all these performance indicators are used for the same purpose, there is no consensus on the most appropriate for model errors amongst researchers.

$$RMSE = \left[\frac{1}{n}\ \sum_{i=1}^{n} \left(P_i -\ O_i\right)^2\right]^{\frac{1}{2}} (11)$$

$$MAE = \ \frac{1}{n}\ \sum_{i=1}^{n} \left|P_i -\ O_i\right| (12)$$

$$MAPE = \ \frac{1}{n}\ \sum_{i=1}^{n} \left|\frac{O_i -\ P_i}{O_i}\right|\ x\ 100\ \% (13)$$

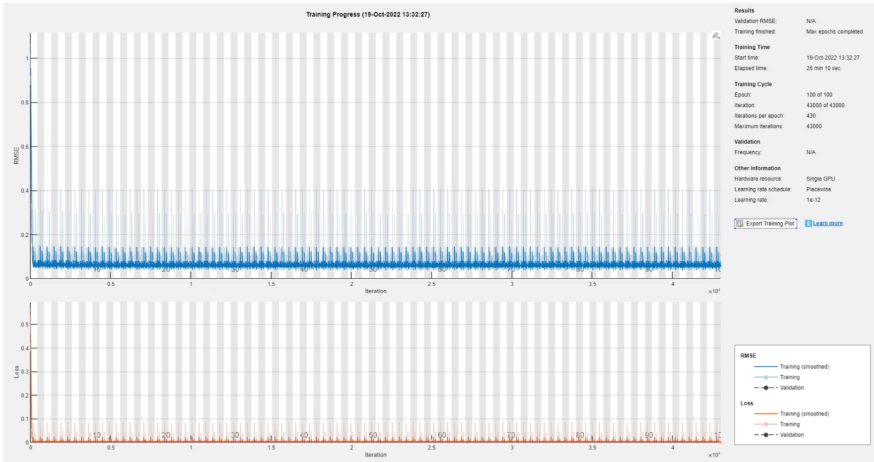
$$MSE = \ \frac{1}{n}\ \sum_{i=1}^{n} \left(O_i -\ P_i\right)^2 (14)$$

where n is the number of data points, O_i is the observed values and P_i are the predicted values. RMSE is used to measure the model's error in predicting quantitative data and establishes the standard deviation of the prediction errors and as a result it uses variance to assign more weight to errors with larger absolute values as compared to errors with smaller absolute values. Thus, it is effective in assessing the model's error data and is concentrated around the line of best fit. When the samples are large enough, and error distribution is Gaussian, RMSE tends to be more appropriate to use. RMSE avoids the use of absolute values and is preferred in mathematical computations. MAE is an indicator of areas between the predicted value and the observed value, it assigns the same weights to all errors. MAPE determines the prediction accuracy in model forecasting. MSE is a risk function related to the expected value of the squared error loss and is an average of the squared value of the difference between the predicted outcome and the actual outcome. The closer the MSE value to zero the better the quality and superiority of the model prediction capabilities. The selection of the most superior performance varies and depends on different researchers and the task they are working on. RMSE is used in this paper since it is the most appropriate and has better results.

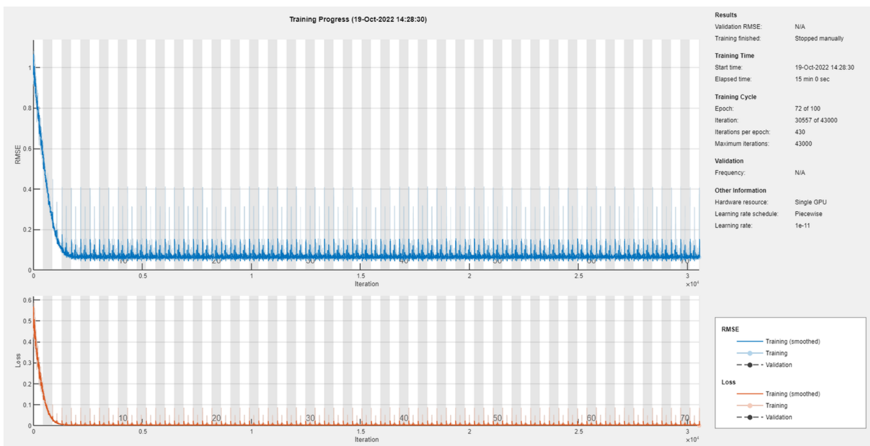
Results and Discussions

The same dataset is applied on five other algorithms and the results are compared with the proposed hybrid model. These models are CNN, LSTM, BiLSTM, CNN-LSTM, CNN-BiLSTM and CNN-BiLSTM-AM, and their prediction performance are shown in table 1. The proposed CNN-BiLSTM-AM hybrid model has the best effect and the minimum prediction error. CNN-Bi-LSTM-AM model performance in terms of RMSE and loss is shown in fig 10. For the RMSE graph the y-axis is $x\ 10^{16}$ and the Loss graph y-axis is $x\ 10^{33}$. The network training cycle had 430 epochs and completed 43 000 iterations with a piecewise learning rate of 1exp(-12), and (a) shows 100 % training complete while (b) is 70 % complete. Fig 11 shows CNN-BiLSTM RMSE and Loss curves. The network training cycle had 430 epochs and completed 43 000 iterations with a piecewise learning rate of 1exp(-12), and (a) shows 100 % training complete while (b) is 70 % complete. RMSE and Loss starts to fall at epoch 3.

Both RMSE and Loss progressively decreased as the iterations increased. An indication measure of how a model can predict the expected outcome is called a loss function. A loss function is a commonly used metric to evaluate a model's misclassification rate i.e., the proportion of incorrect predictions. The deep learning network learns by means of the loss function. Loss function tends to be large when the predictions deviate significantly from the actual results. Optimisation algorithms are used to ensure that the loss function learns to minimise the error in the prediction process. In general loss functions are classified into two major categories namely regression losses and classification losses. In our work the loss function is a regression loss because it is used to predict continuous values.

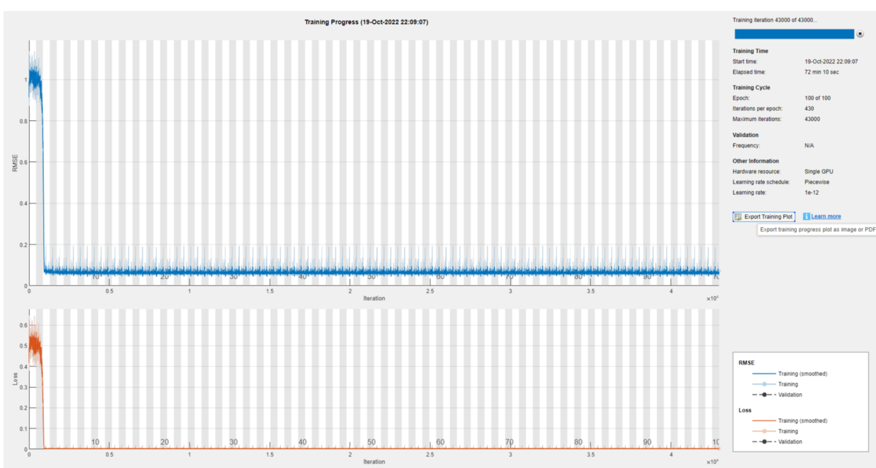


(a)

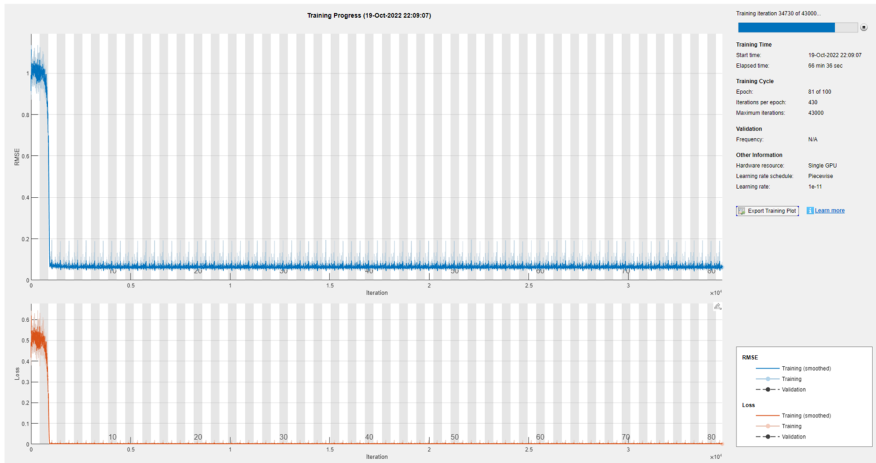


(b)

Fig 10: CNN-BiLSTM-AM - RMSE and Loss



(a)

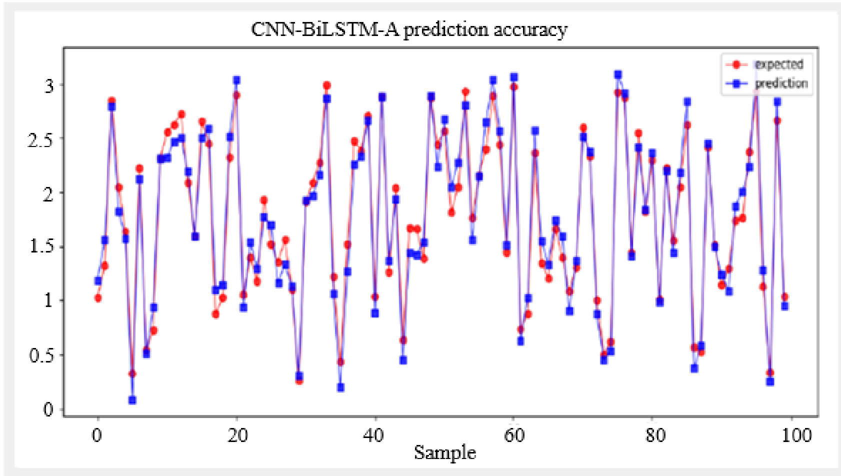


(b)

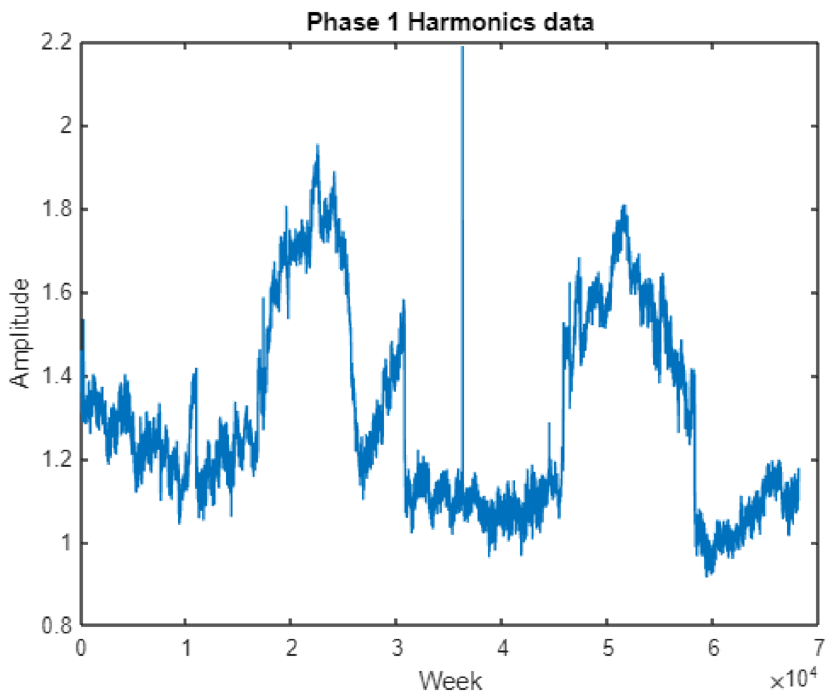
Fig 11: CNN-BiLSTM – RMSE and Loss

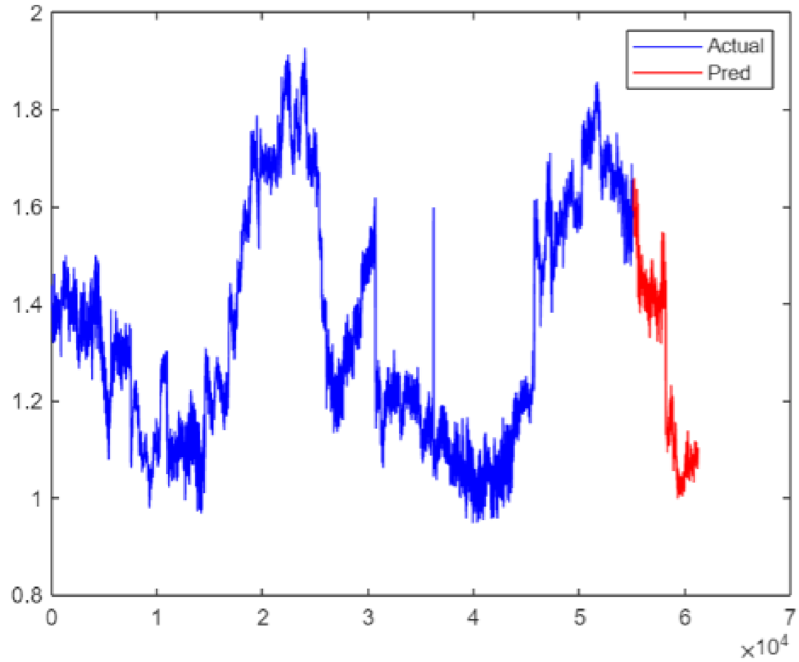
The trained network is tested by forecasting multiple harmonics in the future. The network would predict time steps one at a time, then it updates the network state at each prediction i.e., the previous prediction is used as an input function for the current prediction. The harmonics data that is used is large and Intel Core i7 CPU is used. Intel Core i7 processor is an industry-leading CPU in terms of its performance for discrete-level graphics and AI acceleration and as a result predictions were computed faster, and RMSE is calculated.

The results are in line with the expectations. CNN has a fast-training time and is superior prediction when dealing with images instead of time series data. LSTM prediction performance on time series data is better than CNN prediction performance on time series data. BiLSTM prediction performance is better than LSTM since it extracts information from both forward and reverse directions. CNN-BiLSTM prediction is superior to CNN-LSTM. These hybrid models combine the advantages of either model resulting in faster training time and better performance. The introduction of the attention mechanism to CNN-BiLSTM model, the focus of this paper, further improves the prediction accuracy as shown in fig 12 (a) and (c) and (b) is used to show that the red trace in (c) follows harmonics trace in (b). Expected and predicted values are close; thus, the model prediction accuracy is superior. The proposed hybrid model achieved excellent results in the prediction of harmonics.


CNN-BiLSTM-A prediction accuracy
expected
prediction
Sample

(a)


Phase 1 Harmonics data
Amplitude
Week
×10^4



(b) (c)

Fig 12: CNN-BiLSTM-A model predication accuracy

The superiority of the proposed model is further shown in table 1 where it is compared to the other five models. The table compares model prediction accuracy and RMSE.

Table 1: Comparison of six methods prediction accuracy and performance evaluation indexes

Model	Accuracy %	RMSE
CNN	82.5741	0.0298311012
LSTM	83.6513	0.0221400014
BiLSTM	84.3241	0.0013522131
CNN-LSTM	86.3792	0.0000014917
CNN-BiLSTM	87.4689	0.0000011352
CNN-BiLSTM-AM	92.3569	0.0000002215

Conclusion

This work proposes a deep learning CNN-BiLSTM-A hybrid model for harmonics prediction. The historical data was collected at a chocolate manufacturing factory at East London South Africa and the main source of harmonics is variable speed drives. The prediction methodology has the following steps: The dataset is divided into training data and test data, and the data is apportioned as 80 % and 20 % respectively. The proposed CNN-BiLSTM-A model prediction performance is compared with five other algorithms. The hybrid model exploits strengths of both deep learning models resulting in improved forecasting accuracy RMSE is used to measure the model prediction performance. Based on the experimental results, the proposed hybrid CNN-BiLSTM-A model produces superior results compared to the other models. The superiority of the individual algorithms and the attention mechanism all put together greatly improved the prediction accuracy and lowers the RMSE.

References

[1] H. Deng, Y. Gao, X. Chen, Y. Zhang, Q. Wu, and H. Zhao, "Harmonic Analysis of Power Grid Based on FFT Algorithm," *Proc. - 2020 IEEE Int. Conf. Smart Cloud, SmartCloud 2020* , no. Vmd, pp. 161–164, 2020, doi: 10.1109/SmartCloud49737.2020.00038.

[2] P. Mancuso, V. Piccialli, and A. M. Sudoso, "A machine learning approach for forecasting hierarchical time series," *Expert Syst. Appl.* , vol. 182, no. June 2020, p. 115102, 2021, doi: 10.1016/j.eswa.2021.115102.

[3] F. Chen, J. Chen, and B. Chen, "A Fault Prediction Of Equipment Based On CNN-LSTM Network *," pp. 537–541, 2019, doi: 10.1109/ICEI.2019.00101.

[4] M. Alhussein, K. Aurangzeb, and S. Member, "Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3028281.

[5] Z. Tan and P. Pan, "Network Fault Prediction Based on CNN-LSTM Hybrid Neural Network," 2019, doi: 10.1109/CISCE.2019.00113.

[6] S. Motepe, A. N. Hasan, B. Twala, and R. Stopforth, "Power Distribution Networks Load Forecasting Using Deep Belief Networks : The South African," *2019 IEEE Jordan Int. Jt. Conf. Electr. Eng. Inf. Technol.* , no. 2, pp. 507–512, 2019.

[7] A. Gellert, U. Fiore, A. Florea, R. Chis, and F. Palmieri, "Forecasting Electricity Consumption and Production in Smart Homes through Statistical Methods," *Sustain. Cities Soc.* , vol. 76, no. October 2021, p. 103426, 2022, doi: 10.1016/j.scs.2021.103426.

[8] C. I. Garcia, F. Grasso, A. Luchetta, M. C. Piccirilli, L. Paolucci, and G. Talluri, "applied sciences A Comparison of Power Quality Disturbance Detection and Classification Methods Using CNN ," 2020, doi: 10.3390/app10196755.

[9] S. Khan, S. Nagar, M. Meena, and B. Singh, "Comparison between SPWM and OHSW technique for harmonic elimination in 15 level multilevel inverter," *IEEE Int. Conf. Information, Commun. Instrum. Control. ICICIC 2017* , vol. 2018-Janua, no. 1, pp. 1–5, 2018, doi: 10.1109/ICOMICON.2017.8279088.

[10] H. Gupta, A. Yadav, and S. Maurya, "Multi carrier PWM and selective harmonic elimination technique for cascade multilevel inverter," *Proceeding IEEE - 2nd Int. Conf. Adv. Electr. Electron. Information, Commun. Bio-Informatics, IEEE - AEEICB 2016* , pp. 98–102, 2016, doi: 10.1109/AEEICB.2016.7538405.

[11] Z. Wang, Y. Li, and H. Zhang, "Harmonic detection method based on windowed double-spectrum interpolation FFT," *Proc. - 2019 Int. Conf. Robot. Intell. Syst. ICRIS 2019* , no. 4, pp. 425–427, 2019, doi: 10.1109/ICRIS.2019.00111.

[12] H. Wen, J. Zhang, W. Yao, and L. Tang, "FFT-Based Amplitude Estimation of Power Distribution Systems Signal Distorted by Harmonics and Noise," *IEEE Trans. Ind. Informatics* , vol. 14, no. 4, pp. 1447–1455, 2018, doi: 10.1109/TII.2017.2749682.

[13] Z. Shuai, J. Zhang, L. Tang, Z. Teng, and H. Wen, "Frequency Shifting and Filtering Algorithm for Power System Harmonic Estimation,"*IEEE Trans. Ind. Informatics* , vol. 15, no. 3, pp. 1554–1565, 2019, doi: 10.1109/TII.2018.2844191.

[14] H. Zhao, T. Jin, S. Wang, and L. Sun, "A Real-Time Selective Harmonic Elimination Based on a Transient-Free Inner Closed-Loop Control for Cascaded Multilevel Inverters," *IEEE Trans. Power Electron.* , vol. 31, no. 2, pp. 1000–1014, 2016, doi: 10.1109/TPEL.2015.2413898.

[15] R. Aljendy, "Harmonic Analysis in Distribution Systems Using a Multi-Step Prediction with NARX," pp. 2545–2550, 2020.

[16] J. Liu, Y. Shao, X. Qin, and X. Lu, "Inter-harmonics Parameter Detection Based on Interpolation FFT and Multiple Signal Classification Algorithm," *Proc. 31st Chinese Control Decis. Conf. CCDC 2019* , no. 1, pp. 4691–4696, 2019, doi: 10.1109/CCDC.2019.8833276.

[17] U. E. Akpudo, "electronics A CEEMDAN-Assisted Deep Learning Model for the RUL Estimation of Solenoid Pumps," 2021.

[18] K. Wang *et al.* , "Time Series Classification via Enhanced Temporal Representation Learning," *2021 IEEE 6th Int. Conf. Big Data Anal. ICBDA 2021* , pp. 188–192, 2021, doi: 10.1109/ICBDA51983.2021.9403177.

[19] J. A. Wang, M. Yang, J. Wang, M. Yang, and J. Wang, "Adaptability of Financial Time Series Based on BiLSTM ScienceDirect Adaptability of Financial Time Series Prediction Based on Adaptability of Financial BiLSTM Time Series a Prediction Based on Adaptability of Financial Time Series Wang * Prediction Based ," *Procedia Comput. Sci.* , vol. 199, pp. 18–25, 2022, doi: 10.1016/j.procs.2022.01.003.

[20] K. Masita, "Deep Learning in Object Detection : a Review," no. October, 2020, doi: 10.1109/icABCD49160.2020.9183866.

[21] Y. Deng, L. Wang, H. Jia, X. Tong, and F. Li, "A Sequence-to-Sequence Deep Learning Architecture Based on Bidirectional GRU for Type Recognition and Time Location of Combined Power Quality Disturbance," vol. 15, no. 8, pp. 4481–4493, 2019.

[22] P. Kavianpour, M. Kavianpour, and A. Ramezani, "A CNN-BiLSTM Model with Attention Mechanism for Earthquake Prediction," pp. 1–13.

[23] M. Khodayar *et al.* , "Deep Learning in Power Systems Research :," vol. 7, no. 2, pp. 209–220, 2021, doi: 10.17775/CSEEJPES.2020.02700.

[24] A. Ali and A. N. Hasan, "Optimization of PV Model using Fuzzy- Neural Network for DC-DC converter Systems," *2018 9th Int. Renew. Energy Congr.* , no. Irec, pp. 1–6, 2018, doi: 10.1109/IREC.2018.8362552.

[25] T. Yang, H. Wang, S. Aziz, H. Jiang, and J. Peng, "A Novel Method of Wind Speed Prediction by Peephole LSTM," pp. 364–369, 2018.

[26] E. M. Kuyunani, "Improving voltage harmonics forecasting at a wind farm using deep learning techniques," 2021.

[27] Q. Y. Xuefen Wang1, 2, Hongbin Wang3, Qingyi Wang1, 2,+, Min Ding1, 2, "Current Controller Based on Recurrent Neural Network for Grid connected converter," *Proc. 37th Chinese Control Conf.* , pp. 7564–7569, 2018.

[28] C. C. License and D. Version, "Microgrid Energy Management System With Embedded Deep Learning Forecaster and Combined Optimizer," 2020, doi: 10.1109/ACCESS.2020.3036131.

[29] S. Belagoune, N. Bali, A. Bakdi, B. Baadji, and K. Atif, "Deep learning through LSTM classification and regression for transmission line fault detection , diagnosis and location in large-scale multi-machine power systems," *Measurement* , vol. 177, no. March, p. 109330, 2021, doi: 10.1016/j.measurement.2021.109330.

[30] S. Siami-namini and N. Tavakoli, "The Performance of LSTM and BiLSTM in Forecasting Time Series," pp. 3285–3292, 2019.

[31] N. Bhusal, M. Gautam, R. Mani, and M. Benidris, "International Journal of Electrical Power and Energy Systems Coordinated data falsification attack detection in the domain of distributed generation using deep learning," *Int. J. Electr. Power Energy Syst.* , vol. 134, no. October 2020, p. 107345, 2022, doi: 10.1016/j.ijepes.2021.107345.

[32] L. OuYang, N. Jin, and W. Ren, "A new deep neural network framework with multivariate time series for two-phase flow pattern identification," *Expert Syst. Appl.* , vol. 205, no. March, p. 117704, 2022, doi:

10.1016/j.eswa.2022.117704.

[33] H. Zang *et al.* , "Residential load forecasting based on LSTM fusing self-attention mechanism with pooling," *Energy* , vol. 229, p. 120682, 2021, doi: 10.1016/j.energy.2021.120682.

[34] X. Zhao, L. Huang, and Y. Nie, "Temperature Prediction Based on Integrated Deep Learning and Attention Mechanism," *2021 IEEE 6th Int. Conf. Comput. Commun. Syst. ICCCS 2021* , pp. 163–167, 2021, doi: 10.1109/ICCCS52626.2021.9449176.

[35] S. U. P. Sone, J. J. Lehtomäki, and Z. Khan, "Enterprise Network Data : Analysis and Methods," vol. 1, no. June, 2020.

[36] A. P. Kogekar, R. Nayak, and U. C. Pati, "A CNN-BiLSTM-SVR based Deep Hybrid Model for Water Quality Forecasting of the River Ganga," 2021.

[37] T. Kim and S. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy* , vol. 182, pp. 72–81, 2019, doi: 10.1016/j.energy.2019.05.230.

[38] Z. Lyu, X. Jia, Y. Yang, K. Hu, F. Zhang, and G. Wang, "A comprehensive investigation of LSTM-CNN deep learning model for fast detection of combustion instability," *Fuel* , vol. 303, no. March, p. 121300, 2021, doi: 10.1016/j.fuel.2021.121300.

[39] T. Li, M. Hua, and X. U. Wu, "A Hybrid CNN-LSTM Model for Forecasting," pp. 26933–26940, 2020.

[40] S. Tian, "A Hybrid CNN-LSTM Model for Aircraft 4D Trajectory Prediction," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3010963.