Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2023) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

A Deep Stacked Bidirectional LSTM (SBiLSTM) Model for Petroleum Production Forecasting

B Sirisha[a], Katamouni Kranthi Chaitanya Goud[b], B.T.V Saketh Rohit[c]

[a]*Maturi Venkata Subba Rao (MVSR) Engineering College, Hyderabad and India*
[b]*Maturi Venkata Subba Rao (MVSR) Engineering College, Hyderabad and India*
[c]*Maturi Venkata Subba Rao (MVSR) Engineering College, Hyderabad and India*

Abstract

Deep learning models have demonstrated outstanding forecasting effects and are extensively used in forecasting problems in numerous scenarios. The Long short-term memory neural network (LSTM), a deep learning variant, has huge potential in forecasting the forthcoming values of given patterns using historical time series data. Lately, the forecasting task has captivated the focus of researchers in deep learning areas to handle the constraints of traditional statistical approaches. Owing to the dire need for accurate forecasting and also due to technological advancements, the collection of time series data has become more accessible, and this has paved the way for deep forecasting models. In this paper, we propose the Stacked Bi-LSTM (SBiLSTM) architecture, an adaptation of the traditional deep long-short term memory (TDLM). The approach is evaluated using two oilfield production time series. The performance of the proposed SBiLSTM model is compared with the recurrent neural network-RNNs, multi-layer RNNs, deep gated recurrent unit (DGRU), and deep long-short term memory (DLSTM). Using different measurement criteria, the empirical results show that the proposed DLSTM model outperforms other standard approaches. It is observed that the proposed SBiLSTM model acquires long and short-range interdependent features of univariate time series data without the large memory requirement.

Keywords: Recurrent Neural Networks; Deep gated recurrent unit; Deep Long Short Term Memory; Forecasting ; Multivariate time series.

1. Introduction

Oil production forecasting plays an essential role in industries related to oil and gas. Oil production forecasting majorly helps in the administration of oil reservoirs, examination of oil wells, and prediction of faults in oil plants.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: author@institute.xxx

Moreno et al. (2011) [1] stated that precise forecasting is tough and involves huge investment, and technological advancements are required to assess the oil reservoir production. The data employed to forecast petrol production is time-series in nature. The famous time series data forecasting approach is classified into two broad categories, i.e. statistical approaches and soft computing approaches. There are three major dimensions that determine the forecasting algorithm selection: 1. Time Series Data Characteristics, 2. Forecasting Accuracy, and 3. Solution Complexity Taking solution complexity and accuracy aspects into account, statistical algorithms (Decline curve analysis, ARIMA) provide promising results [2] [3]. However, when the time-series data characterization aspect is considered, soft computing methods (e.g., genetic algorithms, fuzzy, and optimization) are considered more appropriate than the statistical approach.

There is an exponential growth in the use of artificial intelligence in the oil and gas industrial domain. Time series forecasting algorithms like Genetic Algorithms (GA), Support Vector Machines (SVM), and Fuzzy Logic (FL) are extensively used in time series forecasting [4], [5]. An artificial neural network is a well-known machine learning algorithm, that can handle nonlinear datasets [6].

These feed-forward networks are reasonably good in forecasting applications, but they have several setbacks. Since the data is fed in a forwarding direction, it fails to acquire the interdependent data between past and future values [1]. In order to extract the interdependent features of time series data, recurrent neural networks (RNNs) are employed. Though RNNs are popularly used in varied time series applications, they fail to capture the inter-dependencies of long-range time series data [7]. This setback is beaten by Long Short-Term Memory (LSTM), an advancement of RNN. Experimental surveys have proved that LSTM can competently deal with static and dynamic petroleum time-series data [7]. The capability to acquire several levels of data visualisation that counterpart the categorised features of the relational tailor-made architecture is the characteristic feature of DL-deep learning. This algorithm can learn the temporal interdependency of homogeneous and heterogeneous time series data and can extract features from short and long-range input sequences [8]. In this paper, we propose the Stacked Bidirectional LSTM (SBiLSTM) architecture, an adaptation of the traditional deep long-short term memory.

SBiLSTM has been submitted for the purpose of forecasting petroleum production. The proposed architecture includes a stack of varied BiLSTM layers, each of which comprises numerous neural network NNs. The approach is evaluated using three oilfield production time-series data. The prime contribution of the paper is;

1. Identification of optimal number of stacked BI-LSTM layers required, to build efficient forecasting model specific to oil production application.
2. Identification of number of neurons in each stacked BI-LSTM layers required.
3. Identification of optimal number of epochs which helps the model to learn better.

The performance of the proposed SBiLSTM model is compared with the recurrent neural network- RNNs, multi-layer RNNs, and deep gated recurrent unit (DGRU), and deep long-short term memory (DLSTM). Using different measurement criteria, the empirical results show that the proposed DLSTM model outperforms other standard approaches. It is observed that the proposed SBiLSTM model acquires long and short-range interdependent features of uni-variate time series data without the large memory requirement. The paper is prepared as follows. Section 2 will give a detailed review of related works in time series forecasting. Section 3 contains a detailed description of the proposed Stacked Bidirectional LSTM (SBiLSTM) model. setup and results are shown in Section 4 while the conclusion and future work are in Section 5.

2. Related Work

Nathan Sesti et al. (2021)[9] put forward "Integrating LSTMs and GNNs for COVID-19 Forecasting," The author integrated LSTMs and graph neural networks (GNN) to predict the spread of COVID 19. The authors embed the graphsage graph convolution operator instead of linear transformers within the gates of LSTM, creating graphlstm. The model was implemented on a dataset consisting of daily COVID-19 cases between January 24th 2020 and May 19th 2021 from 37 European countries. The model was evaluated using the mean absolute scaled error metric (MASE). In this paper MASE is reported 0.27 per country. According to the authors, this model could prove useful to decision-makers in taking preemptive steps by providing them with knowledge about the future spread of the pandemic. Rajat

Palil (2021) [10] used machine learning and deep learning techniques for time series analysis and stock price forecasting. He compared three different machine learning models: ARIMA, ARIMAX, and LSTM, and found out that ARIMAX gave him the best results. The dataset that he implemented the model on is taken from Kaggle, and the dataset contains stock prices of 50 different companies. The evaluation metrics he considered are RMSE and MAE. For the ARIMA model, the RMSE score is 2110.63 and the MAE score is 1575.35. For the LSTM model, the RMSE score is 4623.99, the MAE score is 4523.34, and for the ARIMAX model, the RMSE score is 235.27, and the MAE score is 165.85, outperforming LSTM and ARIMAX. Tae-Woong Yoo et al. (2020)[11] developed a seasonal LSTM model for time series forecasting of agricultural product sales. The dataset used for the proposed model is extracted from the database of sales of over 3000 items sold in local food retail stores in Korea. The 5 items were chosen from a dataset that had the least number of missing dates. The author has used multiple models like PROPHET, SARIMA, and LSTM to compare it to their proposed model, which is SLSTM. The performance metric that the author is using is NMAE(normalized mean absolute error), and SLSTM has shown the least amount of error compared to other models. The NMAE of SLSTM is below 0.20 whereas other models like LSTM and Prophet show 0.26 and 0.29 respectively. Similar to stock market prime forecasting, Mohammed Mudassir et al. (2020) [12] put forth machine learning approaches for bitcoin price forecasting. They applied ANN, SANN, SVM, and LSTM to the dataset. The data was scraped from a bitcoin information website and it consists of 700n different features based on technical indicators. The models were evaluated with RMSE, MAE, and MAPE metrics. All models performed satisfactorily, with LSTM outperforming all the other models. The models showed up to 65% accuracy for the next day's forecast and 62-64% accuracy for the seventh-ninetieth day forecast. Neo Wu et al. (2020) [13] have proposed deep transformers for time series forecasting of influenza cases. They compared their transformer model to deep learning models like ARIMA and LSTM and used RMSE and Pearson correlation as evaluation metrics. The Pearson correlation for ARIMA, LSTM, and TRANSFORMER is 0.769, 0.924, and 0.928, respectively. The RMSE scores for ARIMA, LSTM, and TRANSFORMER are 1.020, 0.807, and 0.588, respectively. The transformer clearly outperformed the traditional deep learning models. For forecasting, the power load Jie Du et al. (2020)[14] employed Bilstm with the attention model. The data consists of the power load of two companies on a single day. MAPE is used as a performance metric, and BiLSTM outperformed other models. MAPE for BiLSTM with attention was 1.1526, whereas for BiLSTM it was 1.4087.

Ala sagheer et al.(2021)[15] implemented DLSTM for petroleum production forecasting. They implemented the model on two different case studies of Indian and Chinese oil and the metric they used was RMSE. DLSTM surpassed other models with RMSE scores of 0.209 and 0.025 for both the case studies. Peng Chen et al.(2018)[16] used Seasonal Sarima for time series forecasting of temperatures. The data used is the monthly mean temperature of Nanjing from 1951 to 2017 whereas temperatures from 2015 to 2017 are being used as training data. MSE is used as a metric, and the MSE for forecast values is 0.89 which is relatively low. Salvatore Carta et al.(2018) [17] executed ARIMA to predict the sale prices of Amazon e-commerce products. The authors considered many parameters like Google trends, manufacturers, and sentiments to build the model to forecast the sale price of goods. The authors used MAPE as their performance metric for their model. This model can help customers know when and where to buy a product at the right price. 1 shows the comparison of various deep learning models employed in forecasting time series data.

3. Implementation of Proposed Deep Stacked Bidirectional LSTM-(SBiLSTM)

It is widely demonstrated that to effectively enhance the performance of a neural network, the depth of the neural network has to be increased. Inspired by the remarkable learning capabilities of deep recurrent networks, a stacked Bi-LSTM has been proposed. The proposed architecture is used for forecasting uni-variate time series applications. Several layers of Bi-LSTM cells are stacked consecutively in a recurrent fashion. Stacking several layers of Bi-LSTM in a hierarchical fashion will help to unscramble various dissimilarities and dimensions of input time series data. These dissimilar data representations are combined in the upper layers. It is established that such stacked bidirectional LSTM architectures are superior in generalised representation for complex uni-variate time series data. The bidirectional LSTM consists of 2 layers, i.e., a forward layer and a backward layer. The forward layer is a layer where the input and respective weights travel in the forward direction, and the backward layer is a layer where the input and respective weights travel in the backward direction. Figure 1 shows the Deep Stacked Bidirectional LSTM-(SBiLSTM) architecture of the recurrent network. From figure1 it is observed that X (t) represents the input signal at time-stamp

Table 1. Quantitative Assessment: Comparison of LSTM model Accuracy

Author	Proposed Year	Forecasting Model	Application	Dataset Employed	Eval. Metrics	Ref.
Nathan Sesti et al.	2021	LSTM+GNN	COVID 19 SPREAD FORECASTING	COVID 19 Covid testing and deaths	MASE (0.27 per country)	[9]
Rajat Patil	2021	ARIMAX	Stock Price Forecasting	Stock Market Prices	RMSE (235.27) and MAE (165.85)	[10]
Tae-Woong Yoo and Il-Seok Oh	2020	Seasonal LSTM	Forecasting Of Agriculture Product Sales	Agriculture Product Sales	NMAE(0.19)	[11]
Mohammed Mudassir et al.	2020	LSTM models	Bitcoin Price Forecasting	Bitcoin Features And Price Data	Accuracy-(65%)	[12]
NEO WU et.al	2020	Deep Transformer	Influenza Cases Forecasting	Influenza Prevalence Case	RMSE(0.588)	[13]
Jie Du et.al	2020	BILSTM with ATTENTION	Power Load Forecating	Power Load	MAPE(1.1496)	[14]
Alaa Sagheer et al	2018	DLSTM	Petroleum Production Forecasting	Petroleum Production	RMSE (0.219) AND RMSPE (3.124)	[15]
Peng Chen et al	2018	SARIMA	Temperature Forecasting	Monthly Mean Temperature	MSE(0.84)	[16]
Salvatore Carta et al	2018	ARIMA	Forecasting Ecommerce Product Prices	E-Commerce Product Details	MAPE(8.54)	[17]

t. The signal X (t) is fed as the input to the first bidirectional LSTM cell together with the preceding hidden state $B^{(1)}_{t-1}$. B (1) represents the hidden state and 1 represents the hidden state at layer 1. The first hidden state, B(1) at time t, is calculated in this way: **Step 1:-** The first bidirectional LSTM will select the data that is going to be allowed from one cell state to another cell state. This is achieved using the forget gate $f_g(t)$, where U^{fg} is the input weights, W^{fg} is recurrent weights, and Bias- b_{fg}

$$f_g(t) = \sigma(x_t U^{fg} + B_{t-1} W^{fg} + b_{fg}) \tag{1}$$

Step 2:- In this step, the input gate decides if the new data should be preserved. Input gate is represented by $i_g(t)$, where U^{ig} is the input weights, W^{ig} is recurrent weights, and Bias-b_{ig}. The tanh layer generates a vector of new values N'_t. The old cell state N(t-1) is updated into the newly generated cell state N(t). The output gate decides the part of the cell state that proceeds as output.

$$i_g(t) = \sigma(x_t U^{ig} + B_{t-1} W^{ig} + b_{ig}); \qquad N_t^{'} = tanh(x_t U^n + B_{t-1} W^n + b_n) \qquad N_t = N_{t-1} \otimes f_g(t) \oplus f_g(t) \otimes N_t^{'} \tag{2}$$

The figure 2 shows the proposed deep stacked bidirectional LSTM-(SBiLSTM) architecture, which is composed of four layers of bi-directional LSTM, in which the first layer is composed of 256 LSTM cells, the second layer composed

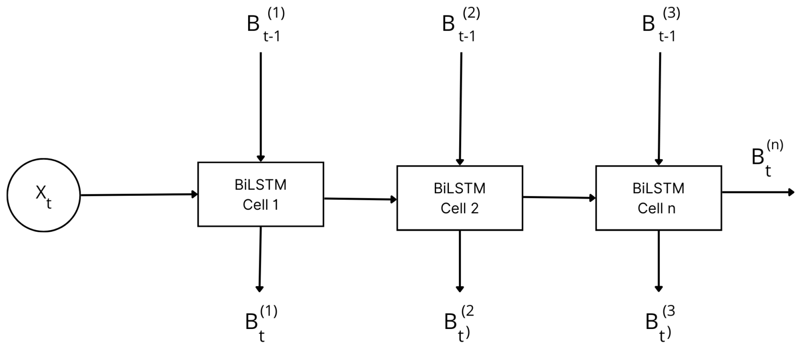

Fig. 1. The Deep Stacked Bidirectional LSTM-(SBiLSTM) architecture of recurrent network

of 128 LSTM cells, the third layer composed of 64 LSTM cells, and the final layer composed of 32 LSTM cells. The output of this layer is passed through an activation functionrelu from which the output passes into a dense layer. The first advantage of this architecture is that each hidden layer will process a part of the required task and pass it to the subsequent layer until the last layer accumulated will provide the desired output. The second advantage is that the architecture will allow the hidden layers at each stage to function at different time periods. These two benefits will aid in the forecasting of complex uni-variate time series data with short and long time lags.

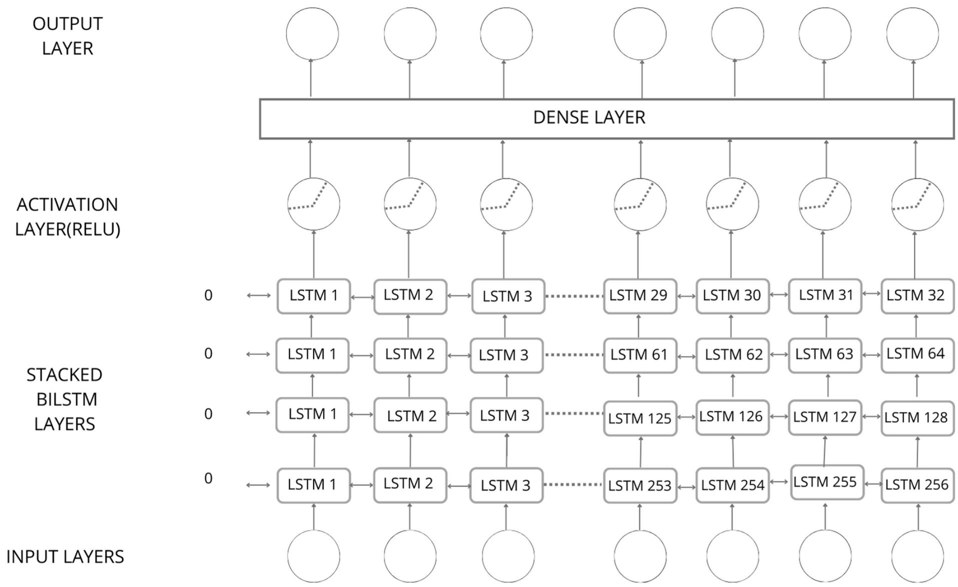

Fig. 2. Proposed Deep Stacked Bidirectional LSTM-(SBiLSTM) Architecture

4. Experimental Results and Discussion

In this paper, we have compared the proposed deep-stacked bidirectional LSTM model against four different state-of-the-art approaches using two real-time oil field datasets.

Dataset 1: Huabei China Oil Field Production Data

TThe dataset consists of raw time series data gathered from the 1st block of the Huabei-China oilfield. The dataset is composed of 227 oil production data, 180 time series observations are used to build the training model, and the leftover 47 data observations are used to test the trained forecasting model. The quantitative evaluation of single-RNN, Multi-RNN, DGRU,DLSTM and proposed BiLSTM is presented sepeartely in Tables 2,3,4,5, 6 respectively. Figure 3 depicts the relationship between the original oil field production data and their forecasted data. Table 7 shows an overall comparison of varied models. For evaluation, root mean square error (RMSE) and root mean square percentage error (RMSPE) are employed. It is observed from table 6 that three layers are optimal to build the forecast model. Experiments were conducted to identify the number of neurons used in each stacked layer and neuron combinations over the stack. It is empirically reported that (5,4,2) neuron combination produces minimum error.After identification of layers and neurons, we have experimented with no of epochs which helps the model to learn better. It is observed that model learns better at 800 epochs. From table 7 we can infer that proposed stacked Bi-LSTM performs better compared to other variants of LSTM.

Table 2. Quantitative Assessment: Findings of Single-RNN.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	[6]	450	4	0.262	3.858
1	[8]	674	4	0.244	3.529
1	[7]	1000	5	0.267	3.79

Table 3. Quantitative Assessment: Findings of Multi-RNN.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
2	[1,3]	800	5	0.273	3.898
2	[5,4]	710	5	0.236	3.329
3	[6,6,5]	600	5	0.248	3.506
3	[5,4,5]	1650	5	0.232	3.293
2	[3,3]	796	5	0.228	3.292

Table 4. Quantitative Assessment: Findings of Deep-GRU.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	[5]	1700	2	0.251	3.554
2	[5,4]	1072	6	0.286	4.044
2	[6,4]	1500	6	0.298	4.245
3	[5,4,2]	500	6	0.225	3.169

Table 5. Quantitative Assessment:Findings of DLSTM.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	[3]	965	5	0.236	3.322
2	[3,1]	794	5	0.231	3.295
3	[4,3,1]	820	5	0.226	3.225
3	[5,4,2]	800	5	0.242	3.146
1	[4]	953	5	0.242	3.456

Table 6. Quantitative Assessment: Results of Proposed Stacked Bidirectional LSTM

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	[5]	974	5	0.325	4.705
2	[5,3]	770	5	0.236	3.373
3	[5,4,2]	800	5	0.192	2.767

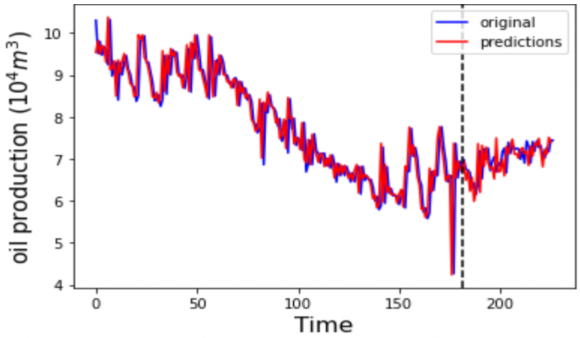

Fig. 3. Production data v.s. prediction using Proposed Stacked Bidirectional LSTM

Table 7. Quantitative Assessment: Overall comparison among RNN, DGRU, and DLSTM using dataset of case study 1

Forecasting models	RMSE	RMSPE
Single-RNN	0.244	3.529
Multi-RNN	0.228.	3.292
Deep-GRU	0.225	3.169
Deep-LSTM	0.226	3.225
Proposed Stacked-Bi-LSTM	0.192	2.767

4.1. Dataset 2: Cambay Rift Basin, India Oil Field Production Data

The dataset consists of raw time series data gathered from Cambay Rift Basin, India oilfields. The dataset is composed of six years of oil production time series data from 20042009, i.e., sixty three months. The Rift basin is a very narrow, drawn-out rift chasm, spreading from Surat to Sanchor. In this project, data is gathered continuously from five oil wells. The data is extremely non-linear in nature. Seventy percent of the data is used to train the forecasting model, and the remaining thirty percent is used to test the model. The quantitative evaluation of thesingle-RNN, Multi-RNN, DGRU,DLSTM and proposed BiLSTM is presented separately in Tables 8,9,10,11, 12 respectively. The relationship between the original oil field production data and their forecasted data is shown in Figure 4. Table 13 shows an overall comparison of varied models. For evaluation, root mean square error (RMSE) and root mean square percentage error(RMSPE) are employed.

It is observed from the table 13 proposed stacked Bi-LSTM performs better than other variants of LSTM.From the emperical results we can infer that stacking several layers of bidirectional LSTM improves the setbacks of shallow recurrent neural network (RNN) architectures.

Table 8. Quantitative Assessment: Findings of Single-RNN.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	3	1420	2	0.0302	4.173
2	3	1420	2	0.0302	4.173
3	4,2	1703	1	0.0337	4.642

It is observed from Table 4,5, 6, 8,9 and 10 the performance of DGRU (for dataset 1 and 2 ,Refer Table4 & 10), DLSTM(for dataset-1 Refer Table5), Single RNN (for dataset-2 Refer Table8), Multi RNN(for dataset-2 Refer

Table 9. Quantitative Assessment: Findings of Multi-RNN.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	2	1420	2	0.0298	4.173
2	2	1420	2	0.0298	4.173
3	3,1	1703	1	0.0337	4.642

Table 10. Quantitative Assessment: Findings of Deep-GRU.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	5	700	2	0.031	4.448
2	5	700	2	0.031	4.448
3	4,2	432	4	0.032	4.489

Table 11. Quantitative Assessment:Findings of DLSTM.

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	5	1900	3	0.028	3.916
2	5	1900	3	0.028	3.916
3	2,2	2000	1	0.029	4.083

Table 12. Quantitative Assessment: Results of Proposed Stacked Bidirectional LSTM

No. of Layers	Neurons	Epochs	Lag	RMSE	RMSPE
1	5	1900	3	0.029	4.071
2	5	1900	3	0.029	4.071
3	3,3	1920	1	0.026	3.923

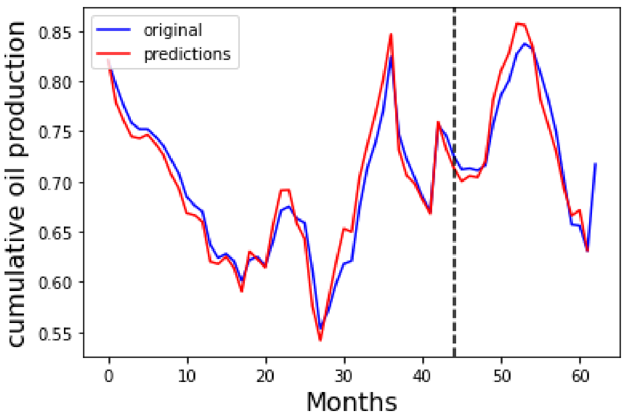

Fig. 4. Production data v.s. prediction using Proposed Stacked Bidirectional LSTM

Table9) and proposed SBiLSTM(Dataset-1 Refer Table6) are accomplished using **three layers of LSTM** whereas in table 2,3, 11 and 12 **two layers of LSTM gives better performance.**.Overall, comparison tables 7 and 13 show that the RMSE and RMSPE of proposed SBiLSTM are the lowest among other models. Though SBiLSTM is optimum compared to other reference models, it demonstrates that there is a variation in the model performance based on the setting of hyper parameters (Number of Layers).

Table 13. Quantitative Assessment: Overall comparison among RNN, DGRU, and DLSTM using dataset of case study 1

Forecasting models	RMSE	RMSPE
Single-RNN	0.0301	4.15
Multi-RNN	0.0296	4.10
Deep-GRU	0.031	4.2
Deep-LSTM	0.028	3.926
Proposed Stacked-Bi-LSTM	0.025	3.913

5. Conclusion

In this paper, we developed a promising prediction model that can be used for the majority of time series forecasting problems. The proposed forecasting model exhibits promising results for most univariate time series forecasting problems. The focus of this paper is majorly on petroleum data forecasting applications. The proposed approach is a deep bidirectional stacked architecture of an LSTM network, denoted as SBiLSTM. It has been demonstrated experimentally that stacking several layers of bidirectional LSTM improves the setbacks of shallow recurrent neural network (RNN) architectures. The average test accuracy of the proposed forecating model is increased by 17%,16% and 14% over RNN, M-RNN and D-GRU models across the two datasets. It is observed that the model is robust in forecasting univariate, long-range dependent, heterogeneous, and complex time series data. Moreover, the model effectively depicts the non-linearity between inputs and outputs. In particular, in datasets 1 and 2 employed in this paper, the proposed SBiLSTM model exceeds its counterparts like deep RNNs and GRUs. The SBiLSTM model's precise forecasting performance qualifies it for use in nonlinear univariate time-series forecasting problems in the oil industry. Our future work is to investigate the effectiveness of SBiLSTM for multivariate time series applications.

References

[1] Moreno J.J.M.(2011) "Artificial neural networks applied to forecasting time series. " *Int. J. Psicothema* . **23(2)** : 322329.

[2] Fetkovich M.J., Vienot M.E., Bradley M.D., Kiesow U.G.(1987) "Decline curve analysis using type curves: Case histories " *Int. J. SPE Form Eval* . **2(4)** : 637656.

[3] George E.P.B., Gwilym M.J., Gregory C.R.(2015) "Time series analysis: Forecasting and control"

[4] Braswell G.(2013) "Artificial Intelligence Comes of Age in Oil and Gas" *Int. Journal of Petroleum Technology* . **65(1)** : 5057.

[5] Mohaghegh S.D.(2005) "Recent developments in application of artificial intelligence in petroleum engineering" *Int. Journal of Petroleum Technology* . **57(4)** : 8691.

[6] Zhang G., Patuwo B.E., Hu M.Y.(1998) "Forecasting with artificial neural networks : The state of the art " *Int. J. Forecasting* . **14(1)** : 3562.

[7] Hochreiter S., Bengio Y., Frasconi P.(2011) "Gradient flow in recurrent nets : the difficulty of learning long-term dependencies "

[8] Goodfellow I., Bengio Y., Courville A.(2016) "Deep learning"

[9] Sesti, Nathan and Garau Luis, Juan Jose and Crawley, Edward and Cameron, Bruce.(2021) "Integrating LSTMs and GNNs for COVID-19 Forecasting" *Int. J. ICML.*

[10] Rajat Patil .(2021) "Time Series Analysis and Stock Price Forecasting using Machine Learning Techniques"

[11] Tae-Woong Yoo , Il-Seok Oh .(2020) "Time Series Forecasting of Agricultural Products Sales Volumes Based on Seasonal Long Short-Term Memory " *Int. J. Appl. Sci* . **10(22)**

[12] Mohammed Mudassir ,Shada Bennbaia ,Devrim Unal ,Mohammad Hammoudeh (2020) "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach " *Int. J. NCA Springer* .

[13] Neo Wu, Bradley Green, Xue Ben, Shawn O'Banion.(2020) "Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case" *Int. ICML* .

[14] Du, Jie and Cheng, Yingying and Zhou, Quan and Zhang, Jiaming and Zhang, Xiaoyong and Li, Gang. (2020) "Power Load Forecasting Using BiLSTM-Attention" *Int. J OP Conference Series: Earth and Environmental Science* .

[15] Sagheer, Alaa and Kotb, Mostafa(2018) "Time Series Forecasting of Petroleum Production using Deep LSTM Recurrent Networks" *Int. J. Neurocomputing* .

[16] Chen, Peng and Niu, Aichen and Liu, Duanyang and Jiang, Wei and Ma, Bin(2018) "Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing" *Int. J. IOP Conference Series: Materials Science and Engineering.*

[17] Saia, Roberto and Carta, Salvatore and Reforgiato Recupero, Diego and Pili, Alessio and Medda, Andrea.(2018) "Forecasting E-Commerce Products Prices by Combining an Autoregressive Integrated Moving Average (ARIMA) Model and Google Trends Data" *Int. J. Future Internet*