A CNN–LSTM model for gold price time-series forecasting

Ioannis E. Livieris, Emmanuel Pintelas & Panagiotis Pintelas

Neural Computing and Applications

ISSN 0941-0643

Neural Comput & Applic
DOI 10.1007/s00521-020-04867-x

ONLINE FIRST

Springer

Neural Computing and Applications
https://doi.org/10.1007/s00521-020-04867-x

S.I. : EMERGING APPLICATIONS OF DEEP LEARNING AND SPIKING ANN

A CNN–LSTM model for gold price time-series forecasting

Ioannis E. Livieris[1] · Emmanuel Pintelas[1] · Panagiotis Pintelas[1]

Received: 21 November 2019 / Accepted: 14 March 2020

Abstract

Gold price volatilities have a significant impact on many financial activities of the world. The development of a reliable prediction model could offer insights in gold price fluctuations, behavior and dynamics and ultimately could provide the opportunity of gaining significant profits. In this work, we propose a new deep learning forecasting model for the accurate prediction of gold price and movement. The proposed model exploits the ability of convolutional layers for extracting useful knowledge and learning the internal representation of time-series data as well as the effectiveness of long short-term memory (LSTM) layers for identifying short-term and long-term dependencies. We conducted a series of experiments and evaluated the proposed model against state-of-the-art deep learning and machine learning models. The preliminary experimental analysis illustrated that the utilization of LSTM layers along with additional convolutional layers could provide a significant boost in increasing the forecasting performance.

Keywords Deep learning · Convolutional neural networks · Time series · Gold price forecasting

1 Introduction

The price of gold constitutes a significant part of the economic and financial state of banks and stock markets. Fluctuations in gold prices have always increased the risk of investment, while the causes of these fluctuations are exceptionally intricate and the trend of gold price is impacted by numerous factors. Even when the financial crisis swept across the entire world in 2008, gold price remained at a high point [3, 5, 29]. Historically, the supporting capacity and sufficient liquidity of gold have attracted an enormous number of financial specialists. Notice that a potential fractional change in gold price may result in huge profit/benefit or on the other hand significant investment and economic losses. Additionally, gold mining companies are highly affected by the gold price fluctuations since the cost of a mining project may be nonprofitable if the future price of gold drops significantly. As a result, the prediction of gold behavior can assist and support financial investors and central banks for making proper decisions for their investment policies and mitigate potential risks. Nevertheless, accurate gold price prediction is generally considered, by its nature, a significantly complex and challenging task.

Research on gold price and movement forecasting and studies on its influence factors were performed for decades, and numerous approaches have been proposed. Classic time-series techniques such as multilinear regression and the well-known Auto-Regressive Integrated Moving Average (ARIMA) have been applied for gold price prediction problem [2, 12, 20]. Besides the classic econometric and time-series approaches, various machine learning methods are utilized to mining the inner complexity of gold price [10, 17, 18, 21, 28]. Nevertheless, the statistical methods usually require assumptions such as stationarity and linear correlation between historical data, while the more sophisticated machine learning methods seem to fail to identify and capture the nonlinear and complex behavior of gold price time series. As a result, all these methods cannot guarantee the development of a reliable and robust forecasting model.

✉ Ioannis E. Livieris
livieris@upatras.gr

Emmanuel Pintelas
ece6835@upnet.gr

Panagiotis Pintelas
pintelas@math.upatras.gr

[1] Department of Mathematics, University of Patras, 265-00 Patras, Greece

Published online: 13 April 2020

In recent years, deep learning methods and techniques have been successfully applied in a variety of real-world challenging prediction problems, including time-series forecasting [1, 17, 31, 32]. They constitute the appropriate methodology to deal with the noisy and chaotic nature of time-series forecasting problem and lead to more accurate predictions. Long short-term memory (LSTM) networks and convolutional neural networks (CNNs) are probably the most popular, efficient and widely used deep learning techniques [11]. The basic idea of the utilization of these models on time-series problems is that LSTM models may efficiently capture sequence pattern information, due to their special architecture design, while CNN models may filter out the noise of the input data and extract more valuable features which would be more useful for the final prediction model. However, standard CNNs are well suited to address spatial autocorrelation data, they are not usually adapted to correctly manage complex and long temporal dependencies [4], while in contrast LSTM networks although they are tailored to cope with temporal correlations, they exploit only the features provided in the training set. Therefore, a time-series model which exploits the benefits of both deep learning techniques could improve the prediction performance.

The main objective of this research is to contribute on the accurate prediction of gold price and movement. For this purpose, we propose a new forecasting model which is based on the principle idea of exploiting the advantages of deep learning techniques. More specifically, the proposed model predicts gold's price value by exploiting the capability of convolutional layers for extracting useful knowledge and learning the internal representation of the time-series data as well as the effectiveness of LSTM layers for identifying short-term and long-term dependencies. Furthermore, the proposed model has the ability to predict the gold price movement direction (increase or decrease) on the next day. In more detail, by analyzing previous gold prices, the model predicts the price on the next day (regression) and also predicts whether the price on the next day will increase or decrease (classification) with respect to today's gold price. It is worth noticing that the information of predicting the gold's movement is of high significance for gold investors and central banks. The proposed model was evaluated against state-of-the-art deep learning and machine learning models for the prediction of the gold's daily price and movement. Our experiments demonstrate that although LSTM models constitute a wide and efficient choice for gold price time series, their utilization along with additional convolutional layers could provide a significant boost in increasing the forecasting performance.

The remainder of this paper is organized as follows: Section 1 presents a brief survey of recent studies, regarding the application of advanced machine learning techniques in gold price forecasting. Section 3 presents in detail the proposed deep learning model. Section 4 presents the data collection and data preparation processes. Section 5 presents a series of numerical experiments. Section 6 summarizes the findings of our research and provides some outlines for future research.

2 Literature review

During the last decade, the development of machine learning and deep learning methodologies for predicting gold's price prediction and movement has gained popularity in the scientific and industrial community. These methodologies provided some useful findings and outcomes about the price behavior. Shafiee and Topal [27] conducted a comprehensive research in which they briefly presented the chronicles of advances and developments for gold price prediction. They also attempted to gain insights on gold price prediction by identifying the factors which affect gold time series as well as its possible relation with crude oil price. In the sequel, we briefly report some rewarding studies which present a number of approaches for forecasting gold price and movement.

Makridou et al. [21] proposed a gold price forecasting model utilizing an Adaptive Neural Fuzzy Inference System (ANFIS). The proposed model was evaluated against an auto-regressive model, an artificial neural network and an ARIMA model which constitute traditional time-series prediction models as well as the "buy & hold" strategy. Based on the reported good performance of the proposed model, the authors highlighted the potential of neuro-fuzzy-based modeling for forecasting gold price.

Dubey [10] proposed two time-series gold price prediction models based on an ANFIS and a support vector model, respectively. The former was designed utilizing subtractive clustering and grid partition algorithms, while the latter was developed using epsilon support vector regression algorithm. Their data were collected from Perth Min, which constitutes the official bullion of Australia, and contained 2206 trading days from Jan 2007 to Apr 2015. Their experiments showed that the support vector model reported better prediction ability and performance.

Liu and Li [18] proposed a new framework based on random forest for monthly gold price prediction. Their proposed model exploited a number of financial factors such as the crude oil price, USD index, the US consumer price index, Dow Jones industrial average, Standard & Poor's 500 index, the prices of US ten year bond futures and the Hang Seng index. These factors affect the prediction ability of gold price. They conducted extensive numerical experiments based on monthly real-world data which were collected from the Web sites of World Bank

Group, US Federal Reserve and International Monetary Fund as well as from the IHS Global Insight Inc., ranging from Jan 1988 to Mar 2015. Their results demonstrated the strong forecasting ability and accuracy of random forest model. Additionally, the authors stated that Dow Jones industrial average and Standard & Poor's 500 index were the most significant factors for gold price prediction.

Wen et al. [30] proposed an interesting approach for studying the external events on gold's price volatility features and influencing factors based on a combination of Chow test along with iterative cumulative sum of squares test. Furthermore, they proposed an ensemble forecasting model for predicting the monthly gold price which exploits the predictions of an artificial neural network and a support vector model. The authors utilized historical data from the London Gold PM Fixing, containing 582 monthly gold prices from Jan 1968 to Jun 2016. Their results revealed that their proposed model attained considerably better performance compared to traditional single models.

Sami and Junero [28] attempted to predict future daily gold rates based on 22 market variables utilizing artificial neural networks and linear regression. The data used in their research were collected from Jan 2005 to Sep 2016 containing Standard & Poor's 500 index, New York Stock Exchange, US Bond rates, oil price and Euro-USD exchange rate. Based on their experiments, the authors stated that their proposed forecasting models can accurately predict the daily gold rates and could be utilized from financial investors and central banks for assistance in their policy and decision making.

Salis et al. [25] presented an extensive study for predicting the fluctuation in daily gold price utilizing artificial neural networks and LSTM models. They presented some discouraging and disappointing experimental results and stated that the accurate gold price prediction is a significantly complex and complicated task due to the large number of random influencing factors. Thus, they concluded that trading experts' advice should be taken into consideration instead of a decision support system.

Jianwei et al. [15] introduced a novel model called ICA-GRUNN which is based on the integration of a separation technique called independent component analysis (ICA) along with a Gate Recurrent Unit Neural Network (GRUNN) model. The authors stated that the gold prices are affected by stochastic factors, cyclic recurrent elements and long-term trend which constituted their motivation. Along this line, their proposed model initially applies ICA to identify the hidden influence factors from data. Then, GRUNN is used to exploit each independent component (IC) and the final gold price prediction emerges as the combination of forecasting results of all ICs. For their experimental analysis, they collected data from monthly gold prices ranging from Jan 1979 to Dec 2017 and their proposed ICA-GRUNN was evaluated against the state-of-the-art benchmark models ARIMA, radial basis function (RBF) neural network and LSTM. The numerical experiments illustrated that ICA-GRUNN reported the best forecasting ability and performance.

Nevertheless, none of the mentioned research studies considered CNN-based prediction models and their combinations with other deep learning techniques for gold price prediction and movement. Our research contribution focuses on exploiting the ability of convolutional layers of learning the internal representation of the gold price data and the effectiveness of LSTM layers for identifying short-term and long-term dependencies. Furthermore, unlike the previous studies, we provide detailed performance evaluation of various deep learning models for both regression and classification problems.

3 Proposed model

The contribution of this research is the development of a prediction model for forecasting the gold price and movement utilizing and exploiting advanced deep learning techniques.

Convolutional layers are characterized by their ability to extract useful knowledge and learn the internal representation of time-series data, while LSTM networks are effective for identifying short-term and long-term dependencies. The principle idea of our proposed model is to efficiently combine the advantages of these deep leaning techniques.

To this end, our proposed model, named CNN–LSTM, consists of two main components: The first component consists of convolutional and pooling layers in which complicated mathematical operations are performed to develop features of the input data, while the second component exploits the generated features by LSTM and dense layers.

In the sequel, we present a brief description of the convolutional, pooling and LSTM layers which constitute the core of the proposed model.

3.1 Convolutional and pooling layers

Convolutional and pooling layers [23] are specially designed data preprocessing layers which have the task to filter the input data and extract useful information which will be used as an input usually on a fully connected network layer.

More specifically, the convolutional layers apply convolution operation between the raw input data and convolution kernels producing new feature values. The input data must have structured matrix form, since this technique was originally intended for extracting features from image datasets [16]. The convolution kernel (filter) can be considered as a tiny window (comparing to the input matrix) which contains coefficient values into a matrix form. This window "slides" all over the input matrix applying convolution operation on each subregion (patch) that this specified window "meets" across the input matrix. The result of all these operations is a convolved matrix which represents a feature value specified by the coefficient values and the dimension size of the applied filter. By applying different convolution kernels on the input data, multiple convolved features can be generated, which are usually more useful than the original initial features of the input data, therefore enhancing the model's performance.

The convolutional layers are usually followed by a nonlinear activation function (e.g., a rectified linear unit) and then a pooling layer. A pooling layer is a subsampling technique which extracts certain values from the convolved features and produces a lower dimension matrix. By a similar procedure, as with the performed operations on the convolutional layer, the pooling layer utilizes a small sliding window which takes as input the values of each patch of the convolved features and outputs one new value which is specified by an operation that the pooling layer is defined to do. For example, max pooling and average pooling calculate the maximum and the average value of each patch's values. As a result, the pooling layer produces new matrices which can be considered as summarized versions of the convolved features that the convolutional layer produced. The pooling operation can help the system to be more robust since small changes of the input will not change the pooled output values.

3.2 LSTM layers

LSTM neural networks [14] are a special type of recurrent neural networks (RNNs) which possess the ability to learn long-term dependencies through the utilization of feedback connections.

Classical RNNs attempt to solve the problem of feedforward neural networks, called "*lack of memory*," which is responsible for exhibiting poor performance on sequences and time-series problems. These models utilize cyclic connections on their hidden layer in order to acquire short-term memory and being able to capture information from time-series and sequences data. Nevertheless, RNNs suffer from the famous vanishing gradient problem which restricts the model to learn long-range dependencies. Therefore, LSTM come to solve this problem by storing useful information on memory cells and vanishing unnecessary information, thus achieving, in general, a better performance than a classical RNN.

Each LSTM unit is composed of a memory cell and three main gates: input, output and forget. By this structure, the LSTM manages to create a controlled information flow by deciding which information must "forget" and which has to "remember," therefore managing to learn long-term dependencies. In more detail, the input gate i_t along with a second gate c_t^*, controls the new information which is stored into the memory state c_t at time t. The forget gate f_t controls the past information which must be vanished or must be kept on the memory cell at time $t-1$, while the output gate o_t controls which information could be utilized for the output of the memory cell. Summarizing, Eqs. (2)–(5) briefly describe the operations performed by an LSTM unit.

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i), \tag{1}$$

$$f_t = \sigma(U_g x_t + W_g h_{t-1} + b_g), \tag{2}$$

$$c_t^* = \tanh(U_c x_t + W_c h_{t-1} + b_c), \tag{3}$$

$$c_t = g_t \odot c_{t-1} + i_t \odot c_t^*, \tag{4}$$

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o), \tag{5}$$

where x_t denotes the input, W_* and U_* are weight matrices, b_* are the vectors of bias term, σ is the sigmoid function, and the operator $\odot$ denotes component-wise multiplication. Finally, the hidden state h_t which constitutes the output of the memory cell is calculated by

$$h_t = o_t \odot \tanh(c_t). \tag{6}$$

Notice that in case several LSTM layers are stacked together, both the memory state c_t and the hidden state h_t of each LSTM layer are forwarded as inputs to the next LSTM layer.

3.3 CNN–LSTM model

In our implementation, we utilized versions of the proposed model. The first one CNN–LSTM$_1$, consists of two convolutional layers of 32 and 64 filters of size (2,), respectively, followed by a pooling layer, an LSTM layer and an output layer of one neuron. The second one, called CNN–LSTM$_2$ consists of two convolutional layers of 64 and 128 filters of size (2,), respectively, followed by a max pooling layer with size (2,), a LSTM layers of 200 units, a dense layer of 32 neurons and an output layer of one neuron. An overview of the proposed CNN–LSTM models architecture is depicted in Figs. 1 and 2.

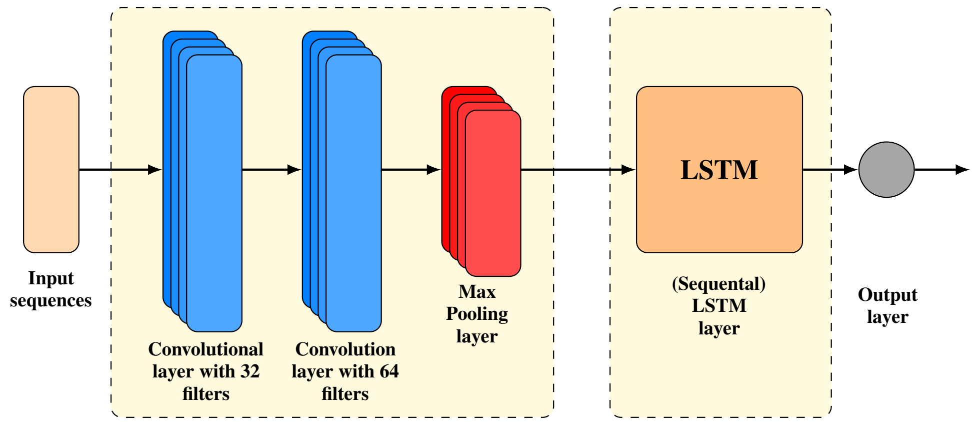

Fig. 1 Proposed CNN–LSTM$_1$ model architecture with two convolutional layers, a pooling layer, a LSTM layer and an output layer

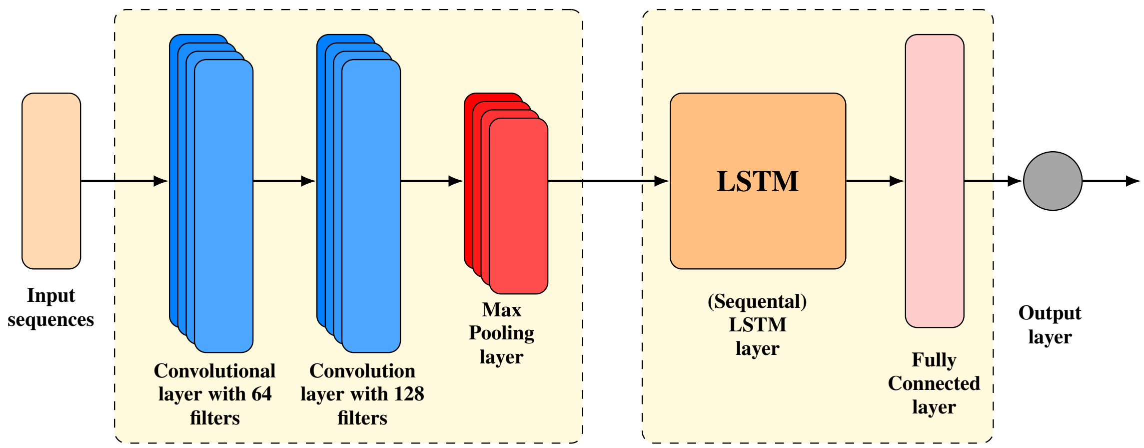

Fig. 2 Proposed CNN–LSTM$_2$ model architecture with two convolutional layers, a pooling layer, a LSTM layer, a fully connected layer and an output layer

4 Data

The data utilized in this research concern the daily gold prices in USD from Jan 2014 to Apr 2018 which were obtained from http://finance.yahoo.com Web site. Table 1 presents the descriptive statistics including the measures: minimum, mean, maximum, median, standard deviation (SD), skewness and kurtosis for describing the nature of the distribution, while Fig. 3 illustrates the daily gold prices.

Table 1 Descriptive statistics for gold daily prices

Statistic	Value
Minimum	100.50
Mean	118.4784
Maximum	133.10
Median	119.325
SD	6.8776
Skewness	− 0.5509
Kurtosis	2.6991

The data were divided into training set and testing set. The training set consists of daily gold prices from Jan 2014 to Dec 2017 (4 years) which ensures a substantial amount of data for training and covers a wide range of long- and short-term trends. The testing set contains daily prices from Jan 2018 to Apr 2018 (4 months) which ensures the evaluation of the forecasting models will perform a considerable amount of unseen "*out-of-sample*" data. Finally, it is worth mentioning that both training and testing data were transformed utilizing a natural logarithm (ln) for homogenizing the variability and stability of the patterns and reducing exponential trend.

5 Experimental results

In this section, we evaluate the performance of the proposed CNN–LSTM model against that of LSTM models (with one and two LSTM layers) and that of the state-of-the-art machine learning models: support vector regression (SVR) [9] and multilayer feed-forward neural network (FFNN) [6]. Notice that the hyper-parameters of all

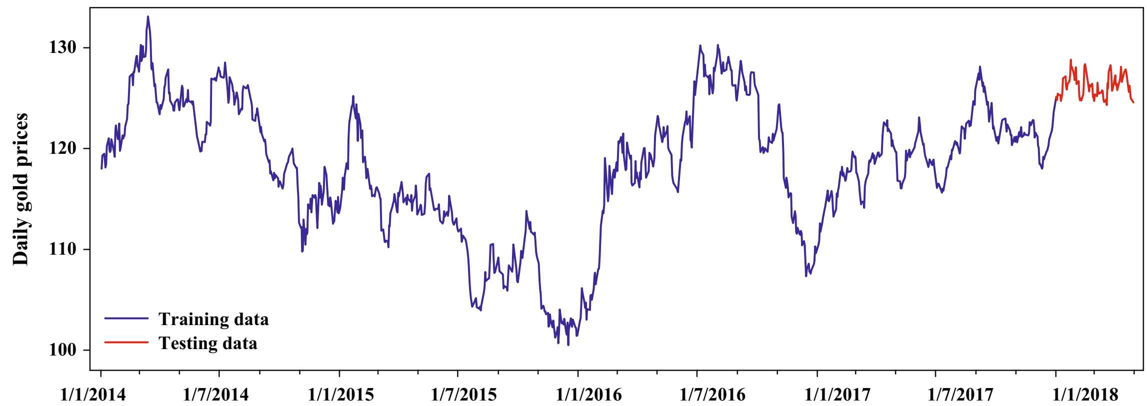

Fig. 3 Daily gold price trend from January 2014 to April 2018

compared models as well as the neural networks' topologies were optimized and are summarized in Table 2. The implementation code was written in Python 3.4 on a PC (Intel(R) Core(TM) i7-6700HQ CPU 2.6 GHz, 16 Gbyte RAM) running Windows 10 operating system. The deep learning models were implemented using Keras library [13] and Theano as backend, while the machine learning models were implemented using Scikit-learn library [22].

All LSTM and CNN–LSTM models were trained for 50 epochs with adaptive moment estimation (ADAM) with a batch size equal to 128, using a mean-squared loss function. ADAM algorithm ensures that the learning steps, during the training process, are scale invariant relative to the parameter gradients. Moreover, for ensuring that no features are dropped out during convolution operations, we apply the same padding.

From a forecasting aspect, the forecasting horizon is crucial for the prediction accuracy of an intelligent model. The forecasting horizon stands for the number of daily prices which are taken into consideration by a forecasting model for predicting the next daily price. More specifically, in case the forecasting horizon is equal to 9 means that the model takes into account prices collected from 9 days and for predicting the price on the 10th day. In this study, we utilized three different values for the forecasting horizon, i.e., 4, 6 and 9 days which presented the best overall performance regarding all forecasting models.

The regression performance of all evaluated models was measured by mean absolute error (MAE) and root-mean-square error (RMSE) which are, respectively, defined by

Table 2 Parameter specification of all forecasting models

Model	Description
SVR	Kernel = RBF, $C = 1.0$, Tolerance $= 10^{-3}$
FFNN	1 hidden layer with 3, 3 and 5 neurons for $F = 4, 6$ and 9, respectively
$LSTM_1$	LSTM layer with 100 units
$LSTM_2$	LSTM layer with 200 units
$LSTM_3$	LSTM layer with 100 units
	LSTM layer with 50 units
$LSTM_4$	LSTM layer with 100 units
	LSTM layer with 100 units
	Fully connected layer with 32 neurons
CNN–$LSTM_1$	Convolutional layer with 32 filters of size (2,)
	Convolutional layer with 64 filters of size (2,)
	Max pooling layer with size (2,)
	LSTM layer with 100 units
CNN–$LSTM_2$	Convolutional layer with 64 filters of size (2,)
	Convolutional layer with 128 filters of size (2,)
	Max pooling layer with size (2,)
	LSTM layer with 200 units
	Fully connected layer with 32 neurons

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|p_i - a_i| \quad \text{and} \quad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(p_i - a_i)^2}$$

where n is the number of predictions, while a_i and p_i are the actual value and the predicted value for i-instance, respectively.

Additionally, we measured the performance of each forecasting model, regarding the classification problem of predicting whether the gold price would increase or decrease on the next day. In more detail, by analyzing F previous gold daily prices including today's price, the model predicts the price on the next day and also predicts whether the price on the next day will increase or decrease with respect to today's gold price. For this binary classification problem, four performance metrics were used: accuracy (Acc), area under the curve (AUC), sensitivity (Sen) and specificity (Spe).

Tables 3, 4 and 5 present the performance of the proposed CNN–LSTM model against the state-of-the-art regression models, relative to forecasting horizon 4, 6 and 9, respectively. Firstly, it is worth mentioning that both CNN–LSTM_1 and CNN–LSTM_2 reported the best overall performance, regarding all values of the forecasting horizon.

Regarding the gold price prediction problem, CNN–LSTM_2 considerably outperformed all forecasting models, presenting the lowest MAE and RMSE score, followed by CNN–LSTM_1. More specifically, CNN–LSTM_2 exhibited 0.0079, 0.0082 and 0.0089 MAE score for forecasting horizon equal to 4, 6 and 9, respectively, while CNN–LSTM_1 exhibited 0.0099, 0.0094 and 0.0095 in the same situations. Moreover, CNN–LSTM_2 reported 0.0082, 0.0095 and 0.01 MAE score, while CNN–LSTM_1 reported 0.0127, 0.011 and 0.0117 for forecasting horizon 4, 6 and 9, respectively. Relative to the deep learning models, LSTM_2 and LSTM_4 presented the best regression performance, outperforming both LSTM_1 and LSTM_3. More specifically, LSTM_2 presented 0.0136, 0.0148 and 0.0158 MAE score and 0.0172, 0.0184 and 0.0196 RMSE score, for forecasting horizon 4, 6 and 9, while LSTM_4 presented 0.0128, 0.0138 and 0.0141 MAE score and 0.0162, 0.0174 and 0.0178 RMSE score, in the same situations. Furthermore, FFNN reported similar performance with the LSTM models, while SVR reported the worst performance.

Finally, it is worth noticing that all models reported the best MAE and RMSE scores for forecasting horizon equal to 4 and that the regression performance of all models slightly worsens as the value of the forecasting horizon increases.

Regarding the binary classification problem, of predicting whether the price on the following day will increase (Up) or decrease (Down) with respect to today's gold price, CNN–LSTM_1 exhibited the best performance, followed by CNN–LSTM_2 and LSTM_3. More specifically, for

Table 3 Performance comparison based on MAE, RMSE, accuracy, AUC, sensitivity and specificity of the proposed CNN–LSTM models against traditional regression models for forecasting horizon equal to 4

Model	MAE	RMSE	ACC (%)	AUC	SEN	SPE
SVR	0.0567	0.0554	48.68	0.476	0.316	0.658
FFNN	0.0149	0.0190	47.37	0.489	0.421	0.526
LSTM_1	0.0222	0.0272	52.63	0.526	0.500	0.553
LSTM_2	0.0136	0.0172	50.00	0.500	0.500	0.500
LSTM_3	0.0242	0.0236	53.55	0.538	0.553	0.526
LSTM_4	0.0128	0.0162	51.32	0.513	0.579	0.447
CNN–LSTM_1	0.0099	0.0127	55.26	0.553	0.553	0.553
CNN–LSTM_2	0.0079	0.0082	51.58	0.519	0.553	0.519

Table 4 Performance comparison based on MAE, RMSE, accuracy, AUC, sensitivity and specificity of the proposed CNN–LSTM models against traditional regression models for forecasting horizon equal to 6

Model	MAE	RMSE	ACC (%)	AUC	SEN	SPE
SVR	0.0571	0.0576	48.68	0.476	0.316	0.658
FFNN	0.0242	0.0255	50.00	0.501	0.474	0.526
LSTM_1	0.0205	0.0259	52.63	0.526	0.468	0.584
LSTM_2	0.0148	0.0184	50.19	0.522	0.495	0.549
LSTM_3	0.0251	0.0246	53.95	0.540	0.526	0.553
LSTM_4	0.0138	0.0174	52.63	0.526	0.500	0.553
CNN–LSTM_1	0.0094	0.0110	56.81	0.577	0.577	0.558
CNN–LSTM_2	0.0082	0.0095	55.53	0.555	0.579	0.500

Table 5 Performance comparison based on MAE, RMSE, accuracy, AUC, sensitivity and specificity of the proposed CNN–LSTM models against traditional regression models for forecasting horizon equal to 9

Model	MAE	RMSE	ACC (%)	AUC	SEN	SPE
SVR	0.0571	0.0579	48.68	0.476	0.316	0.658
FFNN	0.0287	0.0301	46.05	0.461	0.368	0.553
LSTM_1	0.0200	0.0254	52.37	0.524	0.440	0.608
LSTM_2	0.0158	0.0196	51.89	0.509	0.326	0.732
LSTM_3	0.0194	0.0243	52.63	0.526	0.474	0.579
LSTM_4	0.0141	0.0178	51.62	0.508	0.400	0.679
CNN–LSTM_1	0.0095	0.0117	55.26	0.553	0.500	0.618
CNN–LSTM_2	0.0089	0.0100	54.21	0.542	0.474	0.632

forecasting horizon equal to 4, CNN–LSTM$_1$ and CNN–LSTM$_2$ reported 55.26% and 51.58%, respectively, while LSTM$_3$ reported 53.55%. For forecasting horizon equal to 6, CNN–LSTM$_1$ and CNN–LSTM$_2$ reported 56.81% and 55.53%, respectively, while LSTM$_3$ reported 53.95%. For forecasting horizon equal to 9, CNN–LSTM$_1$ and CNN–LSTM$_2$ reported 55.26% and 54.21%, respectively, while LSTM$_3$ reported 52.63%. Regarding AUC metric, CNN–LSTM$_1$ presented the best performance, considerably outperforming all forecasting models. More analytically, CNN–LSTM$_1$ exhibited 0.533, 0.577 and 0.533 AUC score for forecasting horizon equal to 4, 6 and 9, respectively, while CNN–LSTM$_2$ exhibited 0.519, 0.555 and 0.542 AUC score and LSTM$_3$ exhibited 0.538, 0.54 and 0.526, in the same situations. Additionally, it is worth mentioning that both CNN–LSTM$_1$ and CNN–LSTM$_2$ demonstrated considerably better and more stable performance compared to LSTM models regarding their performance on the trade-off between sensitivity and specificity performance metrics.

As a result, based on the previous analysis, we conclude that CNN–LSTM$_1$ is preferable for predicting the gold price movement on the next day and that the classification performance of CNN–LSTM$_2$ is comparable to LSTM models for forecasting horizon 4 and superior for forecasting horizon 6 and 9.

Summarizing it is worth mentioning that based on the previous analysis we are able to conclude that although LSTM models constitute a popular and robust choice for gold price time series, their usage along with convolutional layers could provide a boost in the development of an efficient forecasting model. Regarding the two versions of the proposed model: CNN–LSTM$_1$ exhibited the best performance for the prediction of gold's price increase or decrease, while CNN–LSTM$_2$ considerably outperformed all state-of-the-art time-series models for forecasting gold price, reporting the best MAE and RMSE performance. Furthermore, it is worth noticing that all models reported the best performance for forecasting horizon equal to 6.

In the sequel, we compare the classification performance of the forecasting models CNN–LSTM$_1$ and CNN–LSTM$_2$ against that of LSTM$_3$ and LSTM$_4$, based on the confusion matrices. Notice that the confusion matrix provides a deeper insight to the classification performance of each model since it presents additional information about classes which are commonly mislabeled one as another [19]. Each column of a confusion matrix represents the instances in an actual class, while each row represents the instances in an predicted class.

Tables 6, 7 and 8 present the confusion matrices of CNN–LSTM$_1$, CNN–LSTM$_2$, LSTM$_3$ and LSTM$_4$ models, relative to forecasting horizon 4, 6 and 9, respectively. Firstly, it is worth noticing that the forecasting model CNN–LSTM$_1$ presented considerably better trade-off between true-positive and true-negative rate, compared to all models, regarding all values of the forecasting horizon. More specifically, for forecasting horizon 4, CNN–LSTM$_1$ and LSTM$_3$ presented the best overall performance in terms of accuracy and balance between true-positive and true-negative rate. In contrast, LSTM$_4$ and CNN–LSTM$_4$

Table 6 Confusion matrices of LSTM$_3$, LSTM$_4$, CNN–LSTM$_1$ and CNN–LSTM$_2$ for forecasting horizon equal to 4

	Down	Up		Down	Up
Up	20	18	Down	17	21
Down	17	21	Up	16	22
LSTM$_3$			LSTM$_4$		
	Down	Up		Down	Up
Down	21	17	Down	19	19
Up	17	21	Up	17	21
CNN–LSTM$_1$			CNN–LSTM$_2$		

Table 7 Confusion matrices of LSTM$_3$, LSTM$_4$, CNN–LSTM$_1$ and CNN–LSTM$_2$ for forecasting horizon equal to 6

	Down	Up		Down	Up
Down	21	17	Down	19	19
Up	18	20	Up	17	21
LSTM$_3$			LSTM$_4$		
	Down	Up		Down	Up
Down	22	16	Down	23	15
Up	15	23	Up	19	19
CNN–LSTM$_1$			CNN–LSTM$_2$		

Table 8 Confusion matrices of LSTM$_3$, LSTM$_4$, CNN–LSTM$_1$ and CNN–LSTM$_2$ for forecasting horizon equal to 9

	Down	Up		Down	Up
Down	22	16	Down	24	14
Up	20	18	Up	24	14
LSTM$_3$			LSTM$_4$		
	Down	Up		Down	Up
Down	24	14	Down	23	15
Up	19	19	Up	20	18
CNN–LSTM$_1$			CNN–LSTM$_2$		

have high ratio of false positives ("down" price movement incorrectly identified as "up") meaning that these models are not reliable for predicting price movement increases. For forecasting horizon 6, both CNN–$LSTM_1$ and CNN–$LSTM_2$ reported the best true-negative rate, while CNN–$LSTM_1$ and $LSTM_4$ reported the best true-positive rate. For forecasting horizon 9, CNN–$LSTM_1$ exhibited not only the highest prediction accuracy but also most significantly the best trade-off between true-positive and true-negative rate. In contrast, $LSTM_3$, $LSTM_4$ and CNN–$LSTM_2$ reported high ratio of false positives ("up" price movement incorrectly identified as "down") which implies that these forecasting models are unreliable for predicting price movement.

6 Conclusions

In this work, we proposed a new forecasting model, called CNN–LSTM, for the prediction of gold price and movement. Two versions of the proposed CNN–LSTM model were evaluated against state-of-the-art deep learning and machine learning forecasting models, each having two convolutional layers with different number of filters. The first one reported the best forecasting performance for regression problems, reporting the lowest MAE and RMSE performance, while the second one exhibited the best performance for the classification problem of predicting the gold movement, outperforming traditional time-series models for price movements.

Conclusively, we point out that our experimental analysis indicated that although LSTM models constitute a widely accepted and efficient choice for gold price time series, their utilization along with additional convolutional layers provides a significant boost in increasing the forecasting performance. Finally, it is worth mentioning that due to the sensitivity of various hyper-parameters of the proposed CNN–LSTM and its high complexity, it is possible that additional optimized configuration and mostly feature engineering could further improve the forecasting ability.

It is worth mentioning that the proposed framework can be easily extended to cover the wider scientific area of time-series forecasting applications such as stock market predictions, cryptocurrency price prediction, weather forecasting, earthquake prediction, without the requirement of any extra modifications or additional constraints. In more detail, the proposed framework performs an efficient preprocessing step in order to exploit the internal representation of the times series, through the utilization of convolutional layers. In the sequel, the generated features are exploited by the LSTM and dense layers in order to identify short-term and long-term dependencies in the times series and provide an accurate prediction.

Our future work is concentrated on improving the accuracy of the proposed model by exploiting other factors such as major stock market indices as in [10, 18, 28, 30] and utilizing higher-frequency data. Furthermore, another interesting and promising idea is to evaluate the performance of spiking neural networks (SNNs) [7, 8, 24] for forecasting gold price and movement. SNNs are a new type of artificial neural networks which are more similar to brain's neural networks than the traditional artificial neural networks. In these models, the units utilize a binary output in contrast to the continuous output of neural networks which seems to be a more powerful computational unit than traditional artificial neurons. Nevertheless, the process of training of SNN seems to be a very challenging task since the activation function is by nature nondifferentiable, which prevents the utilization of the classical backpropagation algorithm [26].

Finally, a major point which has to be addressed in future research effort is that in certain times of global instability we experience outliers in the values of the gold. To address this problem an intelligent system might be developed which would probably need an anomaly detection framework, utilizing unsupervised algorithms in order to "catch" outliers or other rare signals which could indicate gold instability.

Compliance with ethical standards

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

References

1. Ai Y, Li Z, Gan M, Zhang Y, Yu D, Chen W, Ju Y (2019) A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. Neural Comput Appl 31(5):1665–1677
2. Askari M, Askari H (2011) Time series grey system prediction-based models: gold price forecasting. Trends Appl Sci Res 6(11):1287–1292
3. Baur DG, McDermott TK (2010) Is gold a safe haven? International evidence. J Bank Finance 34(8):1886–1898
4. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828
5. Choudhry SS, Hassan T, Shabi S (2015) Relationship between gold and stock markets during the global financial crisis: evidence from nonlinear causality tests. Int Rev Financ Anal 41:247–256
6. Daniel G (2013) Principles of artificial neural networks, vol 7. World Scientific, Singapore

7. Demertzis K, Iliadis L, Anezakis VD (2017) A deep spiking machine-hearing system for the case of invasive fish species. In: 2017 IEEE International conference on innovations in intelligent systems and applications (INISTA), IEEE, pp 23–28
8. Demertzis K, Iliadis L, Bougoudis I (2019) Gryphon: a semi-supervised anomaly detection system based on one-class evolving spiking neural network. Neural Comput Appl. https://doi.org/10.1007/s00521-019-04363-x
9. Deng N, Tian Y, Zhang C (2012) Support vector machines: optimization based theory, algorithms, and extensions. Chapman and Hall/CRC, Boca Raton
10. Dubey AD (2016) Gold price prediction using support vector regression and ANFIS models. In: 2016 International conference on computer communication and informatics (ICCCI), IEEE, pp 1–6
11. Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller PA (2019) Deep learning for time series classification: a review. Data Min Knowl Disc 33(4):917–963
12. Guha B, Bandyopadhyay G (2016) Gold price forecasting using ARIMA model. J Adv Manag Sci. https://doi.org/10.12720/joams.4.2.117-121
13. Gulli A, Pal S (2017) Deep learning with Keras. Packt Publishing Ltd, Birmingham
14. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
15. Jianwei E, Ye J, Jin H (2019) A novel hybrid model on the prediction of time series and its application for the gold price analysis and forecasting. Phys A 527:1–14
16. Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet: classification with deep convolutional neural networks. In: Advances in neural information processing systems, IEEE, pp 1097–1105
17. Li J, Dai Q, Ye R (2018) A novel double incremental learning algorithm for time series prediction. Neural Comput Appl 31(10):6055–77
18. Liu D, Li Z (2017) Gold price forecasting and related influence factors analysis based on random forest. In: Proceedings of the 10th international conference on management science and engineering management, Springer, pp 711–723
19. Livieris IE (2020) An advanced active set L-BFGS algorithm for training weight-constrained neural networks. Neural Comput Appl. https://doi.org/10.1007/s00521-019-04689-6
20. Liping X, Mingzhi L (2011) Short-term analysis and prediction of gold price based on ARIMA model. Finance Econ 1
21. Makridou G, Atsalakis GS, Zopounidis C, Andriosopoulos K (2013) Gold price forecasting with a neuro-fuzzy-based inference. Int J Financ Eng Risk Manag 1(1):35–54
22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
23. Rawat W, Wang Z (2017) Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput 29(9):2352–2449
24. Reid D, Jaafar HA, Hissam T (2014) Financial time series prediction using spiking neural networks. PloS One 9(8):e103656
25. Salis VE, Kumari A, Singh A (2019) Prediction of gold stock market using hybrid approach. In: Emerging research in electronics, computer science and technology, Springer, pp 803–812
26. Schliebs S, Kasabov N (2013) Evolving spiking neural network: a survey. Evol Syst 4(2):87–98
27. Shafiee S, Topal E (2010) An overview of global gold market and gold price forecasting. Resour Policy 35(3):178–189
28. ur Sami I (2017) Predicting future gold rates using machine learning approach. Int J Adv Comput Sci Appl 8(12):92–99
29. Wang GJ, Xie C, Jiang ZQ, Stanley HE (2016) Extreme risk spillover effects in world gold markets and the global financial crisis. Int Rev Econ Finance 46:55–77
30. Wen F, Yang X, Gong X, Lai KK (2017) Multi-scale volatility feature analysis and prediction of gold price. Int J Inf Technol Decis Mak 16(01):205–223
31. Zheng J, Fu X, Zhang G (2019) Research on exchange rate forecasting based on deep belief network. Neural Comput Appl 31(1):573–582
32. Zou W, Xia Y (2019) Back propagation bidirectional extreme learning machine for traffic flow time series prediction. Neural Comput Appl 31:7401–7414