The International Journal of Engineering and Science (IJES)
|| Volume || 10 || Issue || 12 || Series I || Pages || PP 01-10 || 2021 ||
ISSN (e): 2319-1813 ISSN (p): 20-24-1805

Time Series Prediction of Server Workload Using Hybrid Model of Recurrent Neural Network and Genetic Algorithms

Amin Talahmeh[1], Mohammed Awad[2*],Mujahed Eleyat[2]
[1]*Department of Computer Science, Arab American University, Jenin, Palestine*
[2]*Department of Computer Systems Engineering, Arab American University, Palestine*

----------***ABSTRACT***----------
Internet users demand service from online websites by sending requests to website servers which process the requests and return responses to the clients. Many requests could be sent to a server at the same moment;therefore, predictingserver workload becomes one of the factors that affect its efficiency. In this paper, we applied three optimized hybrid models to predict the workload of the Portal server (Server 1) and web server (Server 2) of the Arab American University Palestine (AAUP).The models are called Recurrent Neural Networks combined (RNNs), Long-short term memory network (LSTM), and Nonlinear Auto-Regressive Exogenous Neural Networks (NARX) hybrid with Genetics Algorithms (GAs). The experimental results showed that the hybrid model (NARX-GAs) has a better performance than (RNNs-GAs and LSTM-GAs), while the LSTM-GAs model produces better accuracy than RNNs-GAs when used to predict the workload of Server 1, and the RNNs-GAs model produces better accuracy than LSTM-GAs in predicting Server 2 workload. These findings, which are expressed by the RMSE factor, were obtained after the proposed models were applied to the used datasets (Servers 1&2 Processor and Memory usage). Accordingly, when the proposed models (RNNs-GAs, LSTM-GAs, and NARX-GAs) were applied to Server 1 Processor dataset, the RMSE test values were 0.1003, 0.1031, and 0.0998 respectively while they were 0.1687, 0.1676, and 0.1668when applied to server 1 memory dataset. In addition, they were 0.0547, 0.0609, and 0.0417when applied to Server 2 Processor dataset and 0.1052, 0.1115, and 0.1125 when applied to Server 2 Memory dataset.This showed that the RMSE test value for NARX-GAs is slightly greater than other models which makes the NARX-GAs the best hybrid model compared to the other models that we tested.

***KEYWORDS*:** *Server workload, TimeSeries Prediction, RNNs, LSTM, NARX, Hybrid Model, GAs.*

Date of Submission: 01-12-2021 Date of Acceptance: 15-12-2021

I. INTRODUCTION

Since it had been released, people use the Internet for many needs, and this usage has recently increased significantly. Therefore, internet websites are required to use powerful hardware resources (like servers, networks, routers, switches, etc.), and advanced software resources (like operating systems, protocols, database systems, etc.) to ensure high-performance services that meet the needs of users. Internet users (clients) demand service from online websites through sending requests to website servers which process these requests and return responses to the client. Hence, many requests could be sent to the server at the same moment. As of today, the servers, especially for big companies and organizations, such as Facebook and Google, show a clear view about the server workload (SWL) in form of client requests and the server processes that are associated with these requests. For example, Google and Facebook companies receive millions of requests at the same moment, which indicates a large workload on these servers, besides the other processes on the server for receiving requests, preparing response, handle communications, send responses, etc.). Moreover, these servers should be always available and on service for any request. Therefore, anticipating server workload becomes one of the factors that affects server efficiency as required hardware and software could be bought or rented to ensure offering services with high quality and availability, maintain effective server resources such as (CPU and Memory) with minimum cost.in order to So, researchers have begun to research and develop models and technologies that can organize or predict server workloads, such as workload forecasting, workload balancing, and workload distribution. The prediction of server workload is an important task to minimize resource usage of the Central Processing Unit (CPU) and Random-Access Memory (RAM) and reduce the possibility of failures that occur due to unexpected overload, and so ensuring a continuous and effective service. Having a server workload prediction model, we can establish the thresholds to be used in different systems that are used for monitoring the performance of the server. The predicted values of server workload will help the server's

administrators to decide the suitable server capacity's needs over the time and take the preventive measures. These actions will increase server performance, increase server availability, and improve productivity. The server workload prediction is based on the workload history. Therefore, the server workload could be tracked/forecastedin the later period in the form of "Time series", where the time series of server workload means a series of workloads during a time interval, while a workload concept expresses the amount of the client's request processing load which a server will have to do during a time step (e.g. minute, hour, etc.).

There're a lot of techniques that can forecast a time series in the future, but the use of traditional statistical methods such as ARIMA models for forecasting the future server's workload may lead to generating problems because the type of data in the server workloads is nonlinear which could lead topredicting inaccurate or not optimal workload time series [1].While such problems don't appear when using the Artificial Neural Networks (ANNs / NNs) for time series forecasting as in [2], NNs are considered the most used Machine Learning (ML) technique for time series forecasting. Besides, NNs can learn unobserved relationships in the data and visualize the total knowledge of the data, then learn the network based on that knowledge. So, there are three models of NNs were used in this study; Recurrent Neural Networks (RNNs), Long-short term memory neural networks (LSTM), and Nonlinear Auto-Regressive Exogenous Neural Networks (NARX). Moreover, to ensure that the used NNs models efficiently understand, deal with, and describe this nature of data (Nonlinear), they have been augmented with one of the most efficient Genetic algorithms (GAs), which is an Evolution Strategy (ES).

In this research, the server workload refers to Arab American University, Palestine (AAUP) servers' workload (AAUP Portal and Web servers as a case study), and the main objective is to apply the hybrid models that consist of the RNNs and GAsto improve AAUP servers' productivity, stability, and increase the performance of the servers with minimum costs and administrators' efforts.The rest of the paper is organized as follows. In section 2, we present the background concepts. Section 3 presents the related works whilesection 4 discussesthe proposed methodology and applied models. Section 5presents the results obtained based on the applied models and discussion.The conclusions and future work are presented in section 6.

II. BACKGROUND

The server workload is a concept that expresses the amount of the client's request processing load which a server will have to do during a time step (e.g. minute, hour, etc.).In addition, the below equation describes a server workload for a one-time step that comprise the summation of the seconds' workloads values in the time step:

$$WL = \sum_{i=1}^{n} Li\text{SW} \qquad (1)$$

Where *n* is the number of seconds in the time step, *iis the* element number(from *1, 2 … n),* and *Li* is the server workload at the i^{th} second.

A timeseries is a set of observations (Oi), each of them has been sampled at a specific time (Ti), which are demonstratedas a sequence of discrete-time data [3]. Thus, the time series can be described as a series of values that are indexed in the time order. Therefore, to make a time series forecasting model, there's a need for the information from the past periods, and the dependency along these historical data, which will give rise to models capable of predicting future observations.Time series forecasting is considered one of the most valuable data science applications, and it can be demonstrated in a regular interval (i.e. hourly, daily, weekly, monthly, quarterly, etc.). In general, time-series techniques can predict anything that has a continuous change over a time period, using the data collected from the past. Therefore, the main goal in time series analysis is to forecast future behavior based on patterns that exist in the past data. Theoretically, as in [4], time series is a set of vectors *y(t)* where it represents the time elapsed and can be t = 0, 1, 2 …, while *y(t)* can be a random variable. There are a lot of forecasting methodologies that employ the past data to forecast future observations, such as Hidden Markov Models (HMMs) [5][6], Autoregressive Integrated Moving Average (ARIMA)[7][8], Space Vector Regression (SVR) [9] [10], and Artificial Neural Networks (ANNs) [11].Server workload in time series means a series of workloads during a time interval, as shown below:

$$SWL_{ts} = \{L_{t1}, L_{t2}, L_{t3} \ldots L_{tn}\} \quad (2)$$

Where *n* is the number of time series steps, and Lt_n is the server workload at the nth time step.The principle of Artificial Neural Networks (ANNs) is similar to the human neural network's principle, including the nerve cells and links that connect the cells. Where these cells in ANNs are called artificial neurons (nodes), and their links are characterized by a value called weight. The purpose is to develop a method to perform different, computational, and complex tasks more quickly than traditional methods. Practically, an ANNs application could be an application for detection, recognition, classification, or clustering of the patterns, or prediction of the future observations of these patterns. Therefore, with the vario6us ANNs applications, a lot of benefits were

obtained. As shown in [11], the forecasting of crops and animal yield is helpful in agricultural development.Mathematically, an ANNs is a function $f_i(x)$ that's calculated by using the following general formula:

$$f_i(x) = \varphi(\sum_{i=1}^{n} w_i * x_i \qquad (3)$$

x: the input data represented as (x_1, x_2…, x_n), *w*: weights represented as ($w_{i,1}$, $w_{i,2}$,…, $w_{i,n}$),φ: the transfer (activation) function [12].One of the ANNs classifications is network topology-based classification, where the network can be a single layer, multilayer, or recurrent (networks that allow previous outputs to be used as inputs while having hidden states). The general structure of the NNs is illustrated in figure 1.

In this study, and as mentioned in the introduction section, there are three AI hybrid models produced via combining the Recurrence Neural Networks and Optimization Algorithm. Because the recurrence topology is essential with the sequential data since it contains important information about what's coming next, the prediction model must be aware of all previous observations to predict the next series of observations that are based on each other. Thus, a nonlinear time series that depends on knowledge of the previous information and computations can be effectively processed by Recurrence Neural Networks [14], due to the existence of feedback loops in RNNs architecture. The Optimization Algorithm or Evolution Strategy (ES) that has been combined with the used Recurrence Neural Networks (RNNs, LSTM, and NARX)in the 1960s and was described by John Holland as Genetic Algorithms (GAs) [15].It aims to find the optimal solution from a group of possible solutions for a specific problem [16]. Thus, combining a Recurrence Neural Networks (either simple RNNs, and LSTM or dynamic NARX), with Genetics Algorithms (GAs) to forecast future time series which's based on the non-linear type of data, is well appropriate to avoid the problems that could be generated due to using traditional methods such as the back-propagation method for training [17] It is also likely to get better results as shown in [18] [19], which can be done in many various cases. One of these cases, that's used in this study, is letting GAs determine the optimal weights as in [20, 40], for the applied neural networks (RNNs, LSTM, and NARX), to optimize these neural network performances.

For the study datasets, the datasets refer to the AAUP servers (Portal and Web) resources workloads where server resource, in this study refers to the Processor or Memory.In other words, theworkload refers to the usage of a server resource (CPU, or RAM) as an attempt to accomplish any phase of the request life-cycle in the server during a time-period. Therefore, there are four datasets that the proposed models use (Portal-Processor dataset, Portal-Memory dataset, Web-Processor dataset, and Web-Memory dataset). Also, the sampling process was carried out during the summer semester of (2019 / 2020), with a sampling rate of1 sample per hour.

III. LITERATURE REVIEW

During the last years, a few researchers focused on the server workload prediction in different fields using intelligent methods that were applied on the workload datasets collected from servers of commercial websites, IT companies, airport data centers, Internet clouds, etc. More specifically, some researchers tried to predict the incoming workload on the university servers, as we did in this work where we proposed an intelligent methodology to predict the incoming workload on AAUP servers. In [21] the researchers show that the RNNs are a short-term time sequence predictor that's better than the traditional neural networks. On the other hand, the LSTM could be used for the long-term time sequence tasks, which is proved in [22] where the researchers compared both of the black hole and backpropagation learning algorithms with the Long Short-Term Memory (LSTM) Network for forecasting the cloud datacenters workload.Their work showed that the LSTM prediction results had higher accuracy in predictions than both the learning algorithms. Besides, in [23] [24] the study demonstrated the LSTM power to forecast the future non-linear server workload time series with better and more accurate results against the Autoregressive Integrated Moving Average (ARIMA) model, which's statistical and traditional.

Another Neural network that could be used for time series forecasting, because it’s a powerful prediction tool, is the Nonlinear Autoregressive Neural Networks (NARX) which outperformed other traditional prediction models such as Adaptive Neuro-Fuzzy Inference System (ANFIS), Autoregressive Integrated Moving Average (ARIMA), and Support Vector Regression (SVR) in Google cloud systems workload prediction [25]. Also, as shown in [26], when the authors compared the Time Delay Neural Networks (TDNN) and (NARX) to predict the workload of the webmail server for the Palestine Polytechnic University (PPU), they conclude that NARX provides the best result of the prediction than TDNN. In [27], the authors compared the NARX and set of other Neural Networks which are: Multilayer Perceptron (MLP), Elman (RNNs), Cascade-Neural Network (CCNNs), and Pattern Recognition Neural Networks (PRNNs) to forecast the future server workload (defined as HTTP requests), where the best prediction accuracy had been obtained by the NARX model.In [28], the authors proved that updating some parameters in the dynamic neural networks, which are used to predict the server

workload time series, could reflect good results. Their experiments showed that certain values of the time delay parameter used with LSTM increased the prediction accuracy greatly. Moreover, a new orientation is shown in [29], where LSTM and bidirectional-LSTM (BLSTM) have been combined to form a hybrid model with associative learning. which returns highly accurate prediction results for the server resource workload. In the same context, a hybrid model has been proposed in [30] which includes an RNNs and "Autoencoder" and it achieved improvements that reflected positively on the prediction results.

As a new forecasting mechanism, a few researchers have tried using the evolution algorithms (EAs) with the artificial neural networks as an attempt to get more accurate results with minimum time and utilization of resources. In [31], the proposed prediction model was a combination of the Feed-Forward Neural Networks (FFNNs), and self-adaptive differential evolution (SaDE) algorithm, and it has been compared with the backpropagation (BP) based prediction model which showed that the SaDE algorithm outperforms the backpropagation learning algorithm. In [32], the researchers used the Genetic Algorithms for optimizing the Elman network to forecast future load and their experimental results show that the "Elman-GAs" outperforms the traditional linear and nonlinear load prediction models, such as Back-Propagation (BP) and Moving-Average (MA). In [33], the optimization algorithm (SaDE) used in [31] has been improved, and the results showed that the prediction model based on the improved optimized algorithm (MSaDE) predicts the cloud workloads with higher prediction accuracy than the other algorithms such as self-Adaptive Differential Evolution (SaDE) and Back-Propagation (BP).

IV. PROPOSED METHODOLOGY AND APPLIED MODELS

In this study, three recurrence neural networks (Recurrent Neural Networks (RNNs), Long-short term memory network (LSTM), and Nonlinear Auto-Regressive Exogenous Neural Networks (NARX)) are used and combined with the Evolutionary Strategy (ES) to forecast AAUP server's workload time series.Each one of them has its own specifications which allowed proposing three hybrid models and comparing between them to select the best prediction model that can forecast AAUP servers' workloads with highly accurate results.RNNs are found to overcome the shortcoming of other NNs because the RNNs' neuron is recurrent.It has two inputs, one of them is the current input, and the other is the output of the previous neuron that represents the previous state. This state is stored in a short-term memory to be used with the current input for getting the output value, which represents the new state. Briefly, the RNNs neurons' decision of the hidden layer (h) is affected by the layer (h -1) decision. In other words, the current input depends on collecting all previous inputs of recurrent neurons. Therefore, there are correlations between the current, next, and previous time steps in RNNs [34]. This is similar to the human decision-making which depends on all previous and present data to decide to do or not, and how to do it [35].The following equations show how the new state can be calculated:

$$h_t = f(h_{t-1}, x_t) \qquad (4)$$

Where h_t is the new state, h_{t-1} is the previous state, x_t is the current input, and $f(h, x)$ is the activation function. The *tanh activation function"* is used for calculating the new state (h_t), so:

$$h(t) = \varphi_{\tanh}(W_{hh} * h_{t-1} + W_{xh} * x_t + b \qquad (5)$$

Where w_{hh} is the weights between the Hidden-Hidden layers for the previous state, w_{xh} is the weights between the Input-Hidden layers, $\sigma_{tanh}$is the tanh activation function, and b is the bias.
The output state (final state) can be calculated as:

$$y(t) = W_{hy} * h_t + b \qquad (6)$$

Where $w_{hy}$are the weights between the Hidden-Output layers, and b is the bias. The general structure of the RNNs illustrated in figure 1.

For LSTM networks, The RNNs are using the back-propagation methodology for training over time. Also, the traditional RNNs have short-term memory blocks that make updating the state of a neuron's depends only on the state of the previous neuron. This means that short-term dependencies could cause one of the gradient problems: the vanishing problem which occurs when the gradient (error) tends to zero and the exploding problem where the gradient (error) tends to be infinite [36]. Thus, the RNN's training may be terminated without reaching the best solution. To overcome these issues, the authors in [37] proposed a new version of RNNs called Long Short-Term Memory network (LSTM) which includes 'memory cells' that can save the information in memory for long periods. The memory cell is a set of gates which are (forget, input, and output gates) that control the flow of information in the cell. The input gate controls the flow of input activations

into the memory cell and the output gate controls the output flow of cell activations into the rest of the network. Later, the forget gate was added to the memory block [38], which is used to retain the relevant information and forget the non-relevant to the memory cell.

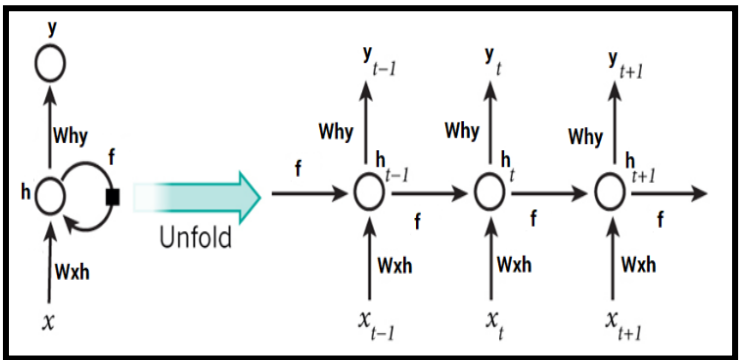

Figure. 1. The basic structure of an RNNs [34]

NARX is a recurrent dynamic network and it has feedback connections. Moreover, NARX has some advantages over other Neural Networks (NNs).For example, NARX is more effective than others in learning, and it generalizes better. Besides, NARX proved that it is better at learning long time dependencies than traditionalRNNs.In addition, it has been shown in [39] that NARX is a good predictor for nonlinear time series because it employs its memory ability which contains the past values of predicted or true-time series to predict the future value of the time series.The general used procedure is illustrated in figure 2, where the input data presents a time series for the AAUP server resource workload (Server 1 CPU, Server 1 Memory, Server 2 CPU, or Server 2 Memory).It includes 1880 samples that were extracted over 7 weeks (the summer semester) and are passed to a pre-processing step, where the data will be smoothed, normalized, and divided to (70%) training datasets and (30%) testing datasets. After that, the training and testing data will be used to check the performance evaluation of the three applied hybrid models (RNNs-GAs, LSTM-GAs, and NARX-GAs) via the root mean square error (RMSE)

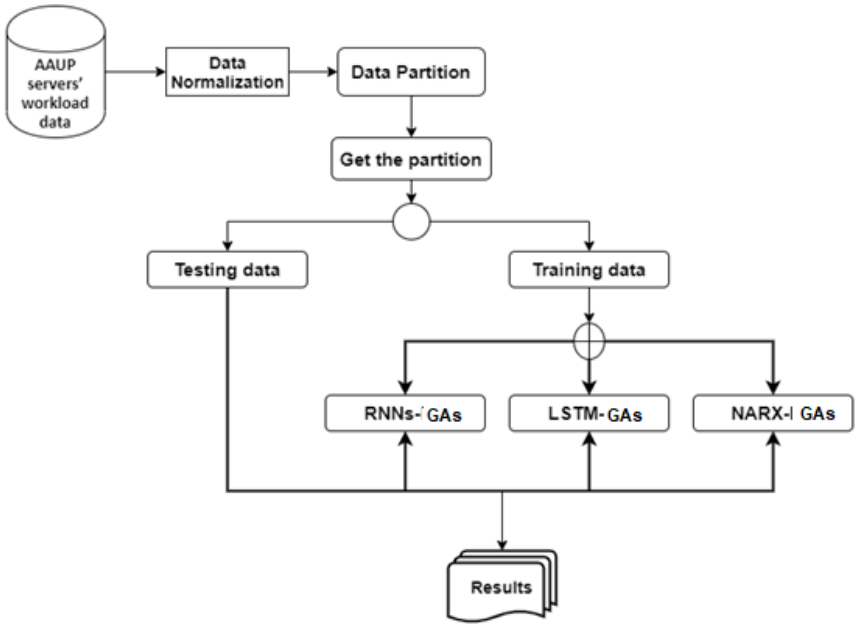

Figure 2. The general procedure of the methodology

The root means square error (RMSE) is calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum(Target - Output)^2}$$

Where *n* is the total number of input data, *Target* is the objective value, and *Output* is the forecasted /predicted value.

The proposed hybrid models to forecast the AAUP servers' resources workloads of the next week (8^{th} week) depends on the following stages:

- Initialize and configure the neural network (RNNs, LSTM, or NARX) based on a prepared dataset (server resource workload).
- Start evolution strategy to optimize the neural network weights, where the updating procedure of weights using GAs includes:
  - Input the prepared data (training data), the number of generations, and population size.
  - Apply the basic operations of the GAs (create and evaluate the initial population, repeat these operations: selection, crossover, mutation, and next population creation for next-generation, until maximum iteration reached).
  - Create a new population and evaluate it by the fitness function.
  - Output the weights-based GAs to be used as new weights for the neural network.

  The following figure illustrates the general steps of the prediction process that combine RNNs and GAs.

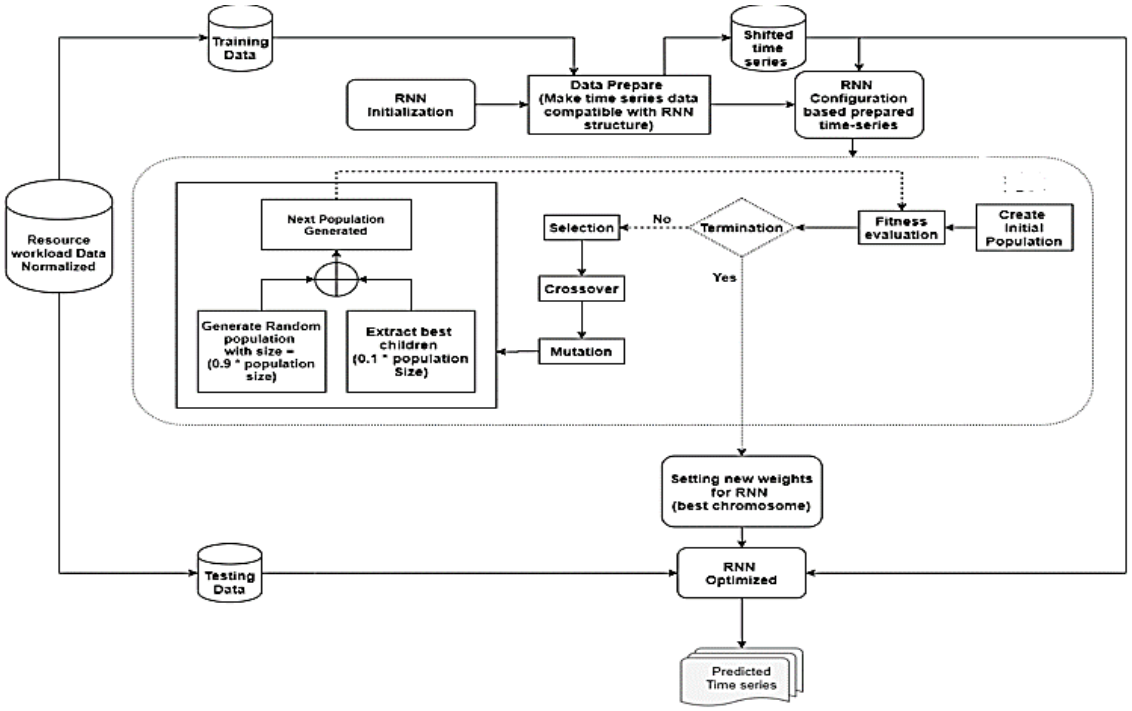

Figure 3. RNNs-GAsStructure

- Set the optimal weights for the neural network (NNs optimized).
- Compute the forecasted outcome and RMSE using training data ($Predicated_{train}$, and $RMSE_{train}$).
- Compute the forecasted outcome and RMSE using testing data ($Predicated_{test}$, and $RMSE_{test}$).
- Evaluate the performance of the applied hybrid models using the outcome $RMSE_{test}$ values to find the best model for forecasting AAUP servers' workloads.
- Predict the next week's AAUP servers' workloads time-series for each resource by computing the mean of the $Predicated_{train}$valuesfor the past 7 weeks (apply for all datasets).

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, to predict and find the AAUP servers' workload for the next time series, there are historical workloads collected by the AAUP computer center relay on our request. The workload datasets of the AAUP servers (Portal and Web) were obtained from two resources (CPU and RAM) throughout the summer semester (from 29-June-2019 to 15-September-2019) at the rate of one sample per hour, which produced 1880 samples. The sample has two formats which are percentageand decimal value. Accordingly, the server processor dataset sample is represented via the percentage format (e.g. 2.57%), while the server memory dataset sample is represented by the decimal value format (e.g. 6.8 MByte). Both of theercentage and decimal values reflect how much the client's request processing has consumed from the total resource's utilization during a specific period.

The results were obtained for the three applied hybrid models (RNNs- GAs, LSTM-GAs, and NARX-GAs) in three stages. The first stage includes determining the general characteristics of the Evolution Strategy (ES) through a set of try-error experiments for obtaining the best GAs parameters (crossover and mutation probabilities, population size, generations, and the best group of the GAs operators which are selection, crossover, and mutation) that will give the best performance. In this stage, we used the RNNs and the server 1 (portal server)' processor workload dataset. Then, based on the obtained general GAs characteristics, the first stage involved applying the RNNs-GAs hybrid model on each dataset while the second stage involved applying the LSTM-GAs hybrid model on each dataset. Then, the third stage involved applying the NARX-GAs hybrid model on each dataset. Also, the number of neurons in each stage has been changed in each dataset so that it starts at 5 neurons and increases incrementally to 30 neurons by adding 5 neurons at a time. Table 1 shows the general characteristics that have been obtained and used on our hybrid models.

Table 1: General characteristics of applied hybrid models on the four datasets.

Parameter	**Server 1 Processor**	**Server 1 Memory**	**Server 2 Processor**	**Server 2 Memory**
Number ofgenerations	500			
Population Size	120			
Measure of performance	RMSE			
Selection Function	Uniform			
Mutation Function	Uniform			
Crossover Function	Two-points			
Mutation Constant	2			
Crossover and Mutation probabilities pair	{0.8, 0.1}			

After the proposed hybrid models have been applied to the study datasets (server 1 processor, server 1 memory, server 2 processor, and server 2 memory), table 2 demonstrates the best $RMSE_{test}$ values that are obtained based on the best number of neurons selected for each model to determine which of the proposed models is considered the best prediction model for AAUP servers' workloads

Table 2: The RMSE test values for each applied proposed model

Proposed Model	Server 1 Processor Workload	Server 1 Memory Workload	Server 2 Processor Workload	Server 2 Memory Workload
RNNs- GAs	0.1003	0.1687	0.0547	0.1052
LSTM- GAs	0.1031	0.1676	0.0609	0.1115
NARX- GAs	0.0998	0.1668	0.0417	0.1125

For all datasets, our findings revealed that the applied hybrid models (RNNs-GAs and LSTM-GAs) can't achieve good forecast results, while the last applied hybrid model (NARX- GAs) can achieve very good forecast results over others, except for the server 2 memory dataset where the RNNs-ES had an RMSE test value is less slightly than the RMSE test value of the NARX-GAs. As a result,thisindicates that the performance of the proposed hybrid model (NARX-GAs) has outperformed other proposed hybrid models (RNNs-GAs and LSTM-GAs) to forecast the server's workload time series.The behavior of RMSE errors of all applied hybrid models for each dataset (Server 1 Processor, Server 1 Memory, Server 2 Processor, and Server 2 Memory) is illustrated in the following figures

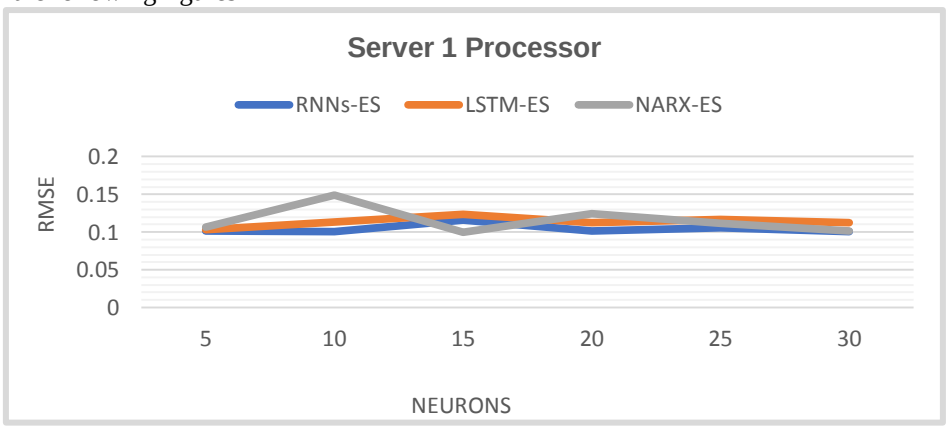

Figure 4: RMSE-Test Errors behavior for the three applied Models (Server 1 Processor workload).

As we notice in figure 4, the best testing of RMSE for Server 1 Processor is when the number of neurons equals (10 and 30) using RNNs-GAswith RMSE test = 0.1003, 5 using LSTM-GAswith RMSE test = 0.1031, and 15 using NARX-GAswith RMSE test = 0.0998.

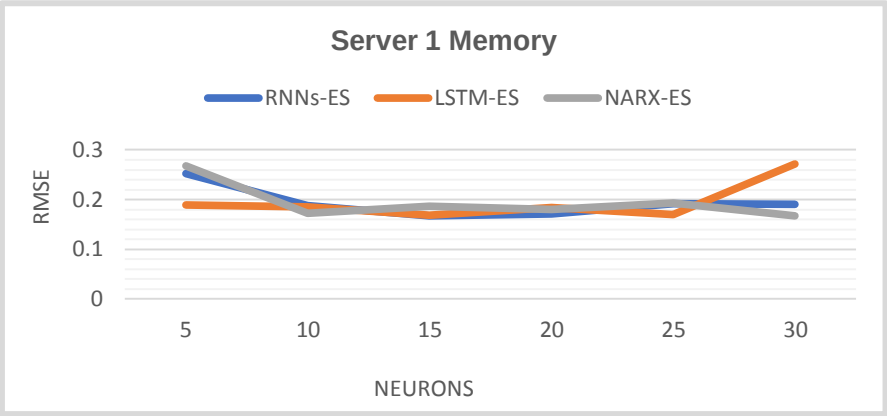

Figure 5: RMSE-Test Errors behavior for the three applied Models
(Server 1 Memory workload).

As shown in figure 5, the best prediction results for Server 2 Memory is when the number of neurons equals 15 using RNNs-GAswith RMSE test = 0.1687, 15 neurons using LSTM-GAswith RMSE test = 0.1676, and 30 neurons uing NARX-GAswith RMSE test = 0.1668.

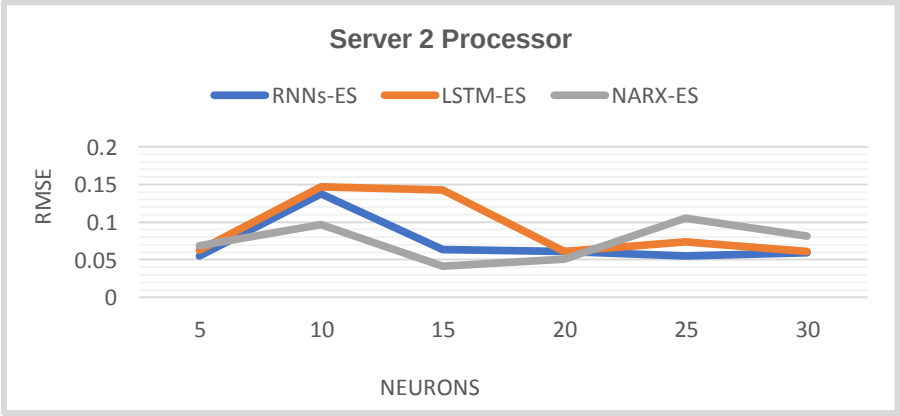

Figure 6: RMSE-Test Errors behavior for the three applied Models
(Server 2 Processor workload).

As shown in figure 6, the best prediction results for Server 2 Processor is when the number of neurons equals 5 using RNNS-GAswith RMSE test = 0.0547, 20 neurons using LSTM-GAswith RMSE test = 0.0609, and 15 using NARX-GAswith RMSE test = 0.0417.

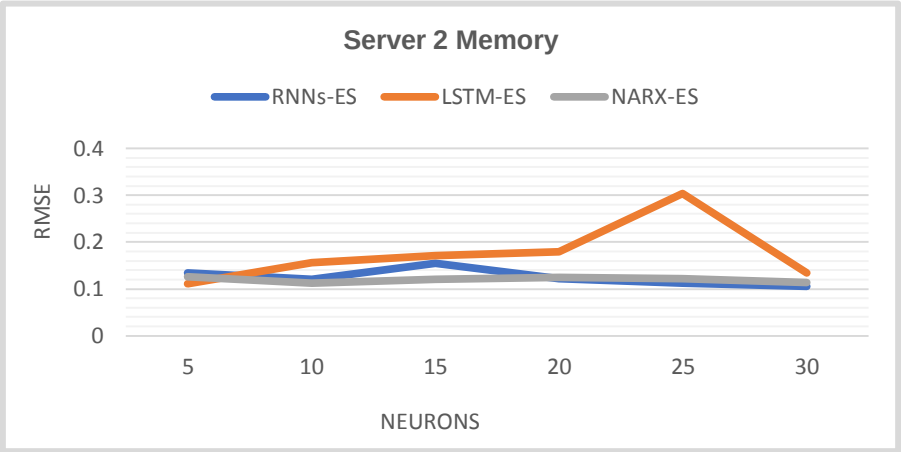

Figure 7: RMSE-Test Errors behavior for the three applied Models
(Server 2 Memory workload).

As shown in figure 7, the best prediction results for Server 2 Memory is when the number of neurons equals 30 using RNNS-GAsS with RMSE test = 0.1052, 5 neurons using LSTM-GAswith RMSE test = 0.1115, and 10 using NARX-GAswith RMSE test = 0.1125.As shown in the illustrated figures for all servers' resources, the behavior of each one of the datasets, that we have applied our models on, is as follows.Server 1 processor behavior using the RNNs-GAs model is close to the behavior of the LSTM-GAsmodel. For server 1 memory behavior using NARX-GAs, it has outperformed the behavior of RNNs-GAs and LSTM-GAsat the 10 and 30 neurons. Server 2 processor behavior using NARX-GAshas outperformed the behavior that results using RNNS-GAsand LSTM-GAs, except when it's applied with 5, 25, and 30 neurons. For the server 2 memory, the behavior of NARX-GAswas better than LSTM-GAs except when it's applied with 5 hidden neurons.

VI. CONCLUSION

The paper analyzed the effects of intelligent hybrid systems using artificial neural networks (RNNs, LSTM, and NARX) integrated with the optimization algorithm (OA) to predict the AAUP servers' resources workloads for the next week. Prediction is achieved depending on the patterns of the historical datasets, where the genetic algorithm (GAs) is used to optimize the optimal weights for all used artificial neural networks (RNNs, LSTM, and NARX). Then, using the RMSE values, the NARX-ES model has been selected as a best-proposed hybrid model that produces the best results for forecasting the AAUP servers' workloads, and the AAUP servers' resources workloads for the next week. Moreover, our study can help keeping the server resources abilities for long periods, reducing server administrators' management efforts and the need for the servers' periodic maintenance. Thus, the servers will provide high-quality services, with a full utilization of their resources at any time and with any workload size. In the future, the researchers intend to enhance and generalize the best-proposed model (NARX-GAs) through an application on additional servers' workload time series and benchmark datasets.

REFERENCES

[1]. Debusschere, V., & Bacha, S. (2013, February). Neural networks for web server workload forecasting. In 2013 IEEE International Conference on Industrial Technology (ICIT) (pp. 1152-1156). IEEE.

[2]. Shoaib, Y., & Das, O. (2015). Modeling Website Workload Using Neural Networks. arXiv preprint arXiv:1507.07204.

[3]. Brockwell, P. J., Brockwell, P. J., Davis, R. A., & Davis, R. A. (2016). Introduction to time series and forecasting. springer.

[4]. Awad, M. (2017). Forecasting of chaotic time series using RBF neural networks optimized by genetic algorithms. Int. Arab J. Inf. Technol., 14(6), 826-834.

[5]. Salameh, O., Awad, M., &AbuAlrub, F. (2019). A Markovian model for Internet of Things applications. International Journal of Computer Networks, 113-124.

[6]. Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77(2), 257-285

[7]. Tran, V. G., Debusschere, V., & Bacha, S. (2012). Hourly server workload forecasting up to 168 hours ahead using Seasonal ARIMA model. 2012 IEEE International Conference on Industrial Technology.

[8]. Calheiros, R. N., Masoumi, E., Ranjan, R., &Buyya, R. (2015). Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications' QoS. IEEE Transactions on Cloud Computing, 3(4), 449–458.

[9]. Lo, J.-H. (2011). A study of applying ARIMA and SVM model to software reliability prediction. 2011 International Conference on Uncertainty Reasoning and Knowledge Engineering.

[10]. Zaman, M., Ahmadi, A., & Makris, Y. (2015, July). Workload characterization and prediction: A pathway to reliable multi-core systems. In 2015 IEEE 21st International On-Line Testing Symposium (IOLTS) (pp. 116-121). IEEE.

[11]. HAMDAN, I., AWAD, M., & SABBAH, W. (2018). Short-Term Forecasting Of Weather Conditions In Palestine Using Artificial Neural Networks. Journal of Theoretical & Applied Information Technology, 96(9).

[12]. García-Pedrajas, N., Hervás-Martínez, C., & Muñoz-Pérez, J. (2003). COVNET: a cooperative coevolutionary model for evolving artificial neural net4works. IEEE Transactions on Neural Networks, 14(3), 575-596 [24]

[13]. Awad, M., & Zaid-Alkelani, M. (2019). Prediction of Water Demand Using Artificial Neural Networks Models and Statistical Model. International Journal of Intelligent Systems and Applications, 11(9), 40.

[14]. Salehinejad, H., Sankar, S., Barfett, J., Colak, E., &Valaee, S. (2017). Recent advances in recurrent neural networks. arXiv preprint arXiv:1801.01078.

[15]. Awad, M. (2014). Using genetic algorithms to optimize wavelet neural networks parameters for function approximation. Int J Comput Sci, (11), 1.

[16]. McCall, J. (2005). Genetic algorithms for modeling and optimization. Journal of Computational and Applied Mathematics, 184(1), 205-222.

[17]. Sexton, R. S., & Gupta, J. N. D. (2000). Comparative evaluation of genetic algorithm and backpropagation for training neural networks. Information Sciences, 129(1-4), 45–59

[18]. Cai, X., Zhang, N., Venayagamoorthy, G. K., & Wunsch, D. C. (2007). Time series prediction with recurrent neural networks trained by a hybrid PSO–EA algorithm. Neurocomputing, 70(13-15), 2342–2353.

[19]. Blanco, A., Delgado, M., &Pegalajar, M. C. (2001). A real-coded genetic algorithm for training recurrent neural networks. Neural Networks, 14(1), 93–105.

[20]. Jadav, K., & Panchal, M. (2012). Optimizing weights of artificial neural networks using genetic algorithms. Int J Adv Res Comput Sci Electron Eng, 1(10), 47-51.

[21]. R2.36: Zhang, W., Li, B., Zhao, D., Gong, F., & Lu, Q. (2016). Workload Prediction for Cloud Cluster Using a Recurrent Neural Network. 2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI).

[22]. Kumar, J., Goomer, R., & Singh, A. K. (2018). Long short-term memory recurrent neural network (lstm-rnn) based workload forecasting model for cloud datacenters. Procedia Computer Science, 125, 676-682.

[23]. Sudhakar, C., Kumar, A. R., Siddartha, N., & Reddy, S. V. (2018). Workload Prediction using ARIMA Statistical Model and Long Short-Term Memory Recurrent Neural Networks. 2018 International Conference on Computing, Power and Communication Technologies (GUCON).

[24]. Janardhanan, D., & Barrett, E. (2017). CPU workload forecasting of machines in data centers using LSTM recurrent neural networks and ARIMA models. 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST).

[25]. Rasheduzzaman, M., Islam, M. A., & Rahman, R. M. (2014). Workload prediction on google cluster trace. International Journal of Grid and High-Performance Computing (IJGHPC), 6(3), 34-52.

[26]. Aljabari, G., & Tamimi, H. (2012). Server Load Prediction Based on Dynamic Neural Networks. Students Innovation Conference, Palestine Polytechnic University.

[27]. Tran, V. G., Debusschere, V., & Bacha, S. (2013). Neural networks for web server workload forecasting. 2013 IEEE International Conference on Industrial Technology (ICIT).

[28]. Pukach, P., &Hladun, V. (2018). Using dynamic neural networks for server load prediction. Computational linguistics and intelligent systems (2), 2018, 157-160.

[29]. Kumar, S., Muthiyan, N., Gupta, S., D., D. A., & Nigam, A. (2018). Association Learning based Hybrid Model for Cloud Workload Prediction. 2018 International Joint Conference on Neural Networks (IJCNN).

[30]. Nguyen, H. M., Woo, S., Im, J., Jun, T., & Kim, D. (2016). A Workload Prediction Approach Using Models Stacking Based on Recurrent Neural Network and Autoencoder. 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS).

[31]. Kumar, J., & Singh, A. K. (2018). Workload prediction in cloud using artificial neural network and adaptive differential evolution. Future Generation Computer Systems, 81, 41–52.

[32]. Xu, D. Y., Yang, S. L., & Liu, R. P. (2013). A mixture of HMM, GA, and Elman network for load prediction in cloud-oriented data centers. Journal of Zhejiang University SCIENCE C, 14(11), 845-858.

[33]. Attia, M. A., Arafa, M., Sallam, E. A., & Fahmy, M. M. (2019). Application of an Enhanced Self-adapting Differential Evolution Algorithm to Workload Prediction in Cloud Computing.

[34]. D. BRITZ, (2015). "Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs" From Url: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

[35]. Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., &Soman, K. P. (2018). NSE stock market prediction using deep-learning models. Procedia computer science, 132, 1351-1362.

[36]. R2.17: Chung, J., Gulcehre, C., Cho, K., &Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling

[37]. S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, 1997.

[38]. R2.19: F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," Journal of Machine Learning Research, vol. 3, pp. 115–143, Mar. 2003.

[39]. DiPietro, R., Rupprecht, C., Navab, N., & Hager, G. D. (2017). Analyzing and exploiting NARX recurrent neural networks for long-term dependencies. arXiv preprint arXiv:1702.07805.

[40]. Awad, M. (2018). Optimizing the Topology and Learning Parameters of Hierarchical RBF Neural Networks Using Genetic Algorithms. International Journal of Applied Engineering Research, 13(10), 8278-8285.

Amin Talahmeh, et. al. "Time Series Prediction of Server Workload Using Hybrid Model of Recurrent Neural Network and Genetic Algorithms." *The International Journal of Engineering and Science (IJES)*, 10(12), (2021): pp. 01-10.