This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIA.2022.3162186, IEEE Transactions on Industry Applications

Novel Uncertainty-aware Deep Neuroevolution Algorithm to Quantify Tidal Forecasting

Seyed Mohammad Jafar Jalali, *Member, IEEE,* Sajad Ahmadian, Md Kislu Noman, *Member, IEEE*, Abbas Khosravi, *Senior Member, IEEE,* Syed Mohammed Shamsul Islam, *Member, IEEE,* Fei Wang, *Senior Member, IEEE*, and João P. S. Catalão, *Fellow, IEEE*

***Abstract*—Tide refers to a phenomenon that causes the change of water level in oceans. Tidal level forecasting plays an important role in many real-world applications specially those related to oceanic and coastal areas. For instance, accurate forecasting of tidal level can significantly increase the vessels' safety as an excessive level of tidal makes serious problems in the movement of vessels. In this work, we propose a deep learning-based prediction interval framework in order to model the forecasting uncertainties of tidal current datasets. The proposed model develops optimum prediction intervals (PIs) focused on the deep learning-based CNN-LSTM model (CLSTM), and non-parametric approach termed as the lower upper bound estimation (LUBE) model. Moreover, we develop a novel deep neuroevolution algorithm based on a two-stage modification of the Gaining-Sharing Knowledge (GSK) optimization algorithm to optimize the architecture of the CLSTM automatically without the procedure of trial and error. This leads to a decline in the complexity raises in designing manually the deep learning architectures, as well as an enhancement in the performance of the prediction intervals. We also utilize coverage width criterion (CWC) to establish an excellent correlation appropriately between both the PI coverage probability (PICP) and PI normalized average width (PINAW). We indicate the searching efficiency and high accuracy of our proposed framework named as MGSK-CLSTM-LUBE by examining over the practical collected tidal current datasets from the Bay of Fundy, NS, Canada.**

***Index Terms*—Uncertainty quantification, Deep neuroevolution, Tidal current forecasting.**

I. Introduction

Tidal power is known as green energy source because it transmits zero greenhouse gasses. The vertical movement of tides creates a tidal current that moves massive quantities of water horizontally near offshore twice a day [1]. This tidal current produces a pattern of energy with the help of gravitational force from both the sun and the moon [2]. The marine energy from the tidal current has integrated into the electric grid. The accurate prediction of tidal current can reduce the operational and battery cost that may lead to reduce the overall cost of electricity production. Besides, this accurate forecasting can assist to monitor the oil slick movements, towing of vessels, and other activities like swimming, boating, and fishing [3].

The relationship between all forces and factors creating tides are very complicated. On December 15, 1892, Sir G. H. Darwin, presented the harmonic model of the tidal current at the Royal Society in London [4]. He stated that the sum of a number of simple harmonic waves can represent the tidal oscillation of the ocean. In 1921, Doodson [5] developed the harmonic component theory of the tidal current model with the help of Darwin's concept. In 1958, Doodson [6] proposed the least-squares estimation technique to estimate the parameters of the harmonic series. Since then, the harmonic analysis technique has been widely applied for tidal forecasting [7]. Classic harmonic-based tidal forecasting has several drawbacks and is heavily influenced by environmental noise. Following the emergence of computer systems, artificial Neural Networks (ANNs) are among the first models to predict the tidal level more accurate than the harmonic model [8]. Lee and Jeng [9] used short-term tidal records to develop an ANN model to forecast the tidal level in Taiwan. Lee et al. [10] applied back-propagation based neural networks with a descent algorithm to forecast the tidal level. In 2007, Lee et al. [11] proposed a hybrid model in a combination of neural networks (NNs) and harmonic models. The performance of four different types of NNs was analyzed in [12] for tidal level forecasting. In 2012, Remya et al. [13] proposed a genetic algorithm (GA) based approach to predict the tidal current.

Although many research works are conducted to predict the tidal data, they suffer from increasing the prediction accuracy and no sign of performance improvement in forecasting the tidal level with the uncertainty data [14]. To overcome this uncertainty issue, a transition from the deterministic prediction to the probabilistic prediction methods is required [8]. The probabilistic prediction methods replace the point-by-point forecast system by predicting intervals (PIs) with a confidence level of the forecast points. Bootstraps [15] and Mean-variance estimation technique [16] are two PIs with NNs based approaches. The tidal data is volatile and has a random characteristic. This uncertainty characteristic makes it very hard to forecast. Due to the high-volatile and random characteristics, it is also impossible to make any prediction of tidal forecasting error.

Developing a high-performance tidal level forecasting model is a major challenging task for the uncertainty behaviour of the tidal data. After applying the harmonic models, machine learning-based models, and hybrid approaches, researchers tried to solve the uncertainty behaviours of tidal by using deep learning-based (DL) approaches. These approaches are attracting different domain types of researchers due to their automatic feature extraction ability. They can easily represent distributed features from deep and complex nonlinear data. DL approaches have also shown better performance in comparison with the traditional machine learning-based approaches in the various research domains. Recently, Riazi [17] proposed a deep learning-based approach to predict the tidal in different beaches.

S.M.J. Jalali, and A. Khosravi are with the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Waurn Ponds, VIC 3216, Australia (e-mail: sjalali@deakin.edu.au and abbas.khosravi@deakin.edu.au).

S. Ahmadian is with the Faculty of Information Technology, Kermanshah University of Technology, Kermanshah, Iran (e-mail: s.ahmadian239@gmail.com).

M.K Noman and SMS Islam are with the School of Science, Edith Cowan University, Joondalup 6027, Australia (e-mails: md.k.noman@gmail.com and syed.islam@ecu.edu.au).

F. Wang is with the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, 102206, China (e-mail: feiwang@ncepu.edu.cn).

J. P. S. Catalão is with University of Porto Faculty of Engineering, 112048 Porto Portugal 4200-465 (e-mail: catalao@fe.up.pt).

0093-9994 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: b-on: UNIVERSIDADE DO PORTO. Downloaded on March 25,2022 at 13:22:11 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIA.2022.3162186, IEEE Transactions on Industry Applications

One of the main shortcomings that deep learning algorithms face is the variability of the number of hyperparameters in these models, which sometimes reaches more than 10 values of hyperparameters in a deep learning architecture. These hyperparameters often suffer from a large number of real and integer value numbers that each time the deep learning algorithm is implemented on a real-world dataset, it is necessary to obtain the best values of these hyperparameters to obtain the best possible accuracy. This strategy is very time-consuming and requires a lot of knowledge on how to use the best hyperparameter combination for a network architecture. Therefore, methods such as deep neuroevolution (DNE) are needed to automatically find the best possible combination of hyperparameters that also achieve the highest performance of the deep neural networks. In general, DNE is known as automated neural architecture search technique which is an effective method to tune accurately the hyperparameters of deep neural networks [18]–[20]. DNE technique utilizes Evolutionary Algorithms (EAs) to optimize the hyperparameters and search for an accurate architecture of Deep Neural Networks (DNNs) [21]. This technique has many benefits compared to gradient-based back-propagation algorithms [22]. Convolutional neural networks (CNNs) and long short term memory (LSTM) networks are two popular deep neural network models that are utilized in various applications such as image classification [23]–[25] and speech recognition [26]. CNNs can automatically extract the features from the input directly during the training process. Although CNNs have shown better performance and can save a lot of human effort in feature extraction, selecting the appropriate model dept and optimize many hyperparameters in its architecture is a highly time-consuming and complex task. LSTM networks are a modification of Recurrent Neural Networks (RNNs) that have been applied as a solution to the time series forecasting problem. The LSTM based networks are more perfect because of the network architecture of LSTM units and the forget gates added to the networks. These added gates can record past features information during training and can select useless features to ignore from training data [27].

To deal with such challenges related to the tidal datasets, this study presents a probabilistic approach centered on the lower-upper bound estimation (LUBE) method and deep CNN-LSTM models (CLSTM). The LUBE strategy is a non-parametric technique for developing effective PIs without allowing any assumptions about prediction error. In other words, we have proposed a deep learning-based CLSTM model for quantifying the uncertainty forecasting of the tidal data points. In addition, a new evolutionary algorithm is proposed by improving the original version of Gaining-Sharing Knowledge (GSK) algorithm. Then, the proposed evolutionary algorithm is utilized to achieve the optimal values of hyperparameters of the CLSTM model. This leads to significantly progress the efficiency of the proposed forecasting model. Therefore, the novelty of this research paper lies in the automatic search for architecture as well as the search for a set of hyperparameters by an improved evolutionary algorithm. To the best of our knowledge, this paper is the first attempt to develop a forecasting model for tidal level based on an automatic architecture search of hyperparameters of CNN and LSTM models using an improved evolutionary algorithm.

The major contributions of this research paper are as follows:

1) A new evolutionary algorithm is proposed based on GSK algorithm by incorporating two positive modifications in the original version of GSK. The main advantage of the improved GSK method is to speed-up the convergence process and also better performing the search methodology.
2) A new uncertainty-aware based forecasting approach is proposed for the first time for the tidal data points based on a combination of CNN and LSTM models in which their hyperparameters are optimized by the proposed evolutionary algorithm.
3) Numerous experiments are conducted on the tidal data points time series to show the superiority of the proposed forecasting model and the results indicate that the proposed model significantly outperforms other tidal forecasting models.

The rest of the paper is organized as follows: Section II describes the proposed forecasting model in detail. The analysis of experimental results is presented in Section III. Finally, the conclusion of this paper is highlighted in Section IV.

II. Methodology

A. LUBE-based CLSTM model

LUBE is known as an appropriate non-parametric method, which can offer the necessary of PIs. Generally, LUBE is built based on neural networks to calculate the upper and lower bounds from the forecasting data. The neural networks have as many inputs as the training data. However, the number of outputs are two for stabilising the upper and lower bounds.

As the deep CLSTM models are such kinds of neural network (NN) family, in the training period, no lower and upper bounds exist in the network. Instead, these lower and upper bounds are replaced by a fitness function called Coverage Width Criterion (CWC). This fitness function connects the PI bandwidth with its confidence level. The coverage width criterion establishes the optimal PIs with the help of two effective parameters such as A. PI Confidence level and B. PI Normalized Average Width.

A. PI Confidence Level: The percentage of probability points are determined by the PI confidence level. The lower and upper bounds of PI cover the PI confidence level. It is also called the PI coverage probability (PICP) function. A PI with a higher coverage probability shows higher probability targets. PICP can be formulated as follows:

$$PICP = \frac{1}{N}\sum_{i=1}^{N} c_i \tag{1}$$

$$c_i = \begin{cases} 1; & y_i \in [L_i, U_i] \\ 0; & y_i \notin [L_i, U_i] \end{cases} \tag{2}$$

where N denotes to the number of samples, and $c_i = 1$ if $y_i \in [L(X_i), U(X_i)]$, else $c_i = 0$. Besides, $L(X_i)$ and $U(X_i)$ respectively represent the lower and upper boundaries of the ith PI. If the value of PICP is far less than it's own nominal value, the first conclusion is that the constructed PIs are completely untrustworthy. This metric has been indicated almost in all PI-related studies as an indicator of how well the PIs are designed and constructed [28]–[30].

The deep CLSTM is trained with PICP in such a way that the output PIs will cover the forecast target with higher probability or equal to the confidence level of $(1-a)\%$. A PI with lower PICP than the confidence level is unacceptable and will be discarded from the training.

B. PI Normalized Average Width: A PI with a higher PICP than the confidence level can cover more forecasting points. But, it should be considered for forecasting targets. Alternately, we should consider PI width to train the CLSTM. Because PI with wide bound

0093-9994 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: b-on: UNIVERSIDADE DO PORTO. Downloaded on March 25,2022 at 13:22:11 UTC from IEEE Xplore. Restrictions apply.

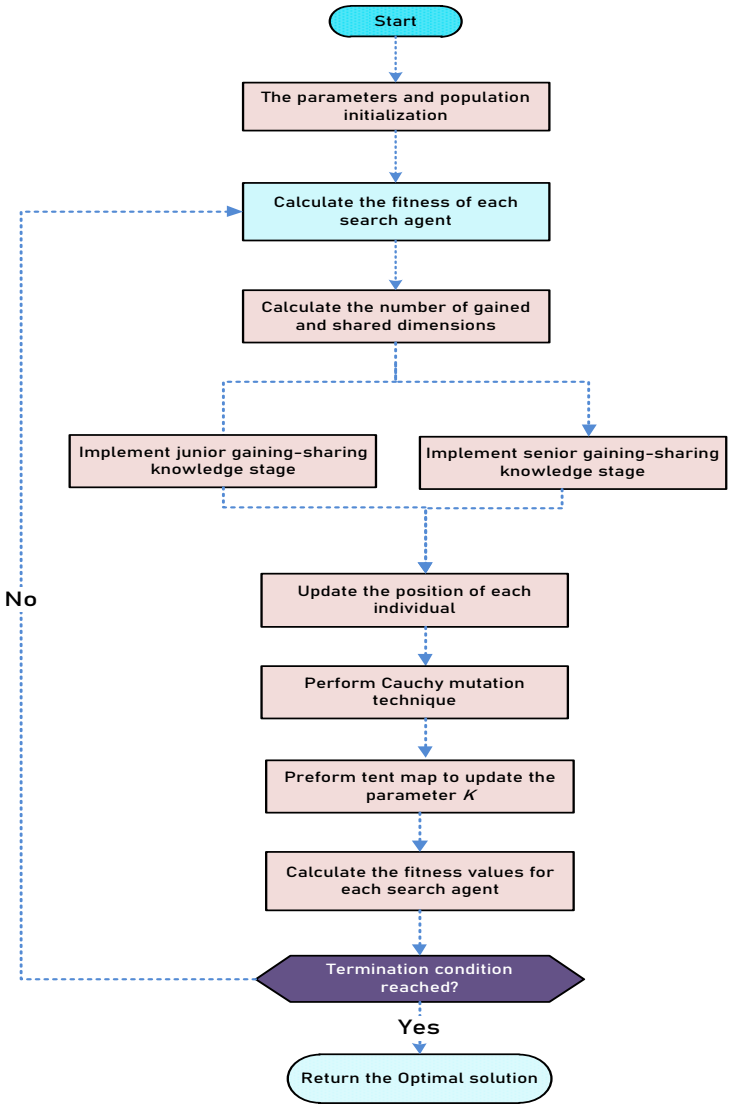

Fig. 1: The flowchart of the novel MGSK algorithm.

TABLE I: The valuations list of CLSTM hyperparameters during the DNE procedure.

Expression	Valuation List
$N_c = \psi$	$\overline{N_c(\psi)}_{\psi\in[1,8]} \in [1, 2, 3, ...]$
$fN_{1<i<N_c} = 2^{\psi+2}$	$\overline{fN(\psi)}_{\psi\in[1,6]} \in [8, 16, 32, \cdots]$
$R_{\text{dropout}} = (\psi + 3) \times 0.05$	$\overline{R_{\text{dropout}}(\psi)}_{\psi\in[1,8]} \in [0.2, 0.25, \cdots]$
$S_{\text{batch}} = 10 \times \psi$	$\overline{S_{\text{batch}}(\gamma)}_{\gamma E[1,10]} \in [10, 20, \cdots]$
$L_{\text{rate}} = 0.001 + 0.005 \times (\psi - 1)$	$\overline{L_{\text{rate}}(\psi)}_{\psi\in[1,21]} \in [0.001, 0.006, \cdots]$
uN_i	$\overline{N_c(\psi)}_{\psi\in[1,300]} \in [1, 2, 3, ...]$

does not provide the necessary information about the forecast data. Another criterion called PI normalized average width (PINAW) is used in the CLSTM training to avoid extra growth of intervals. Mathematically, the PINAW can be defined as:

$$PINAW = \frac{1}{NR}\sum_{i=1}^{N}(U_i - L_i) \tag{3}$$

R denotes the scope of the underlying targets in use which is used for PI normalisation.

C. Coverage Width Criterion: By considering the PINAW and PICP functions, we can construct the optimal PI through the CWC fitness function. The PICP and PINAW can behave contradictory

Algorithm 1 The pseudo-code of the proposed deep neuroevolution-based uncertainty quantification model for tidal current datasets.

Input: P (Population size), k_f (Knowledge factor), k_r (Knowledge ratio), and GEN (Maximum number of generations).
Output: Predicted uncertainty outcome.
Begin algorithm:
Partition the tidal datasets into training Tr, test Te and validation Val sets;
Generate a random initial population X_i (i=1,2,..., P);
Set g=1;
while ($g < GEN$) **do**
    Set a CLSTM model for each solution based on its hyperparameter values;
    Train the CLSTM model of each solution based on training set Tr;
    Measure the fitness of population using Eq.(4) as the predicted uncertainty of CLSTM model obtained by the validation set Val;
    Calculate the number of Gained and Shared dimensions of both phases using Eqs.(3) and (4);
    //Perform Junior gaining-sharing knowledge phase;
    //Perform Senior gaining-sharing knowledge phase;
    if fitness(x_i^{new}) $\leq$ fitness(x_i^{old}) **then**
        $x_i^{old} = x_i^{new}$, fitness(x_i^{old}) = fitness(x_i^{new})
    end if
    if fitness(x_i^{new}) $\leq$ fitness(x_{best}^{Global}) **then**
        $x_{best}^{Global} = x_i^{new}$, fitness(x_{best}^{Global}) = fitness(x_i^{new})
    end if
    Employ the CM operator using Eq.(11);
    Update the parameter k using Eq.(12);
    Set g=g+1;
end while
Set an optimal CLSTM model according to the obtained hyperparameter values by the best solution;
Output the predicted uncertainty of tidal datasets in the test set Te by CLSTM algorithm;
End algorithm

so that an accurate formulation is required to prepare a settled solution for the final PI. A hyperparameter η is applied in the CWC to extend the difference between the PICP and the preferred confidence level $(1-a)$. Finally, the PI with low confidence level is penalized and dropped from the PI candidate sets. Mathematically, CWC can be defined as:

$$\text{CWC} = \text{PINAW}\left(1 + \gamma(\text{PICP})e^{-\eta(\text{PICP}-(1-\alpha))}\right) \tag{4}$$

During the NN training period, $\gamma(\text{PICP}) = 1$. From Eq. (4), the CWC will concentrate on decreasing the PINAW criterion as long as the PICP value is equivalent or higher than the confidence level $(1 - \alpha)$. During the optimization process, the following condition is considered to avoid singleton in the CWC:

$$PINAW > 0 \tag{5}$$

B. The Modified Deep Neuroevolution Strategy

A CWC function specified in Eq.(4) as the fitness function needs to be optimized throughout the PI development in order to optimize the deep CLSTM hyperparameters. The hyperparameters including number of convolutional layers (N_c), number of filters in each convolutional layer ($fN_{i,i\in N_c}$), dropout rate (R_{dropout}), batch size (S_{batch}), learning rate ($Lrate$), and neural unites in the LSTM layer (uN_i) with their aforementioned values in Table I have the most impact in the performance of CLSTM architecture. Thus, we evolve these six hyperparameters during the optimization procedure for obtaining the automated architecture with the least effort as well as the best performance of CLSTM model. As this problem is indeed a linearized constrained optimization problem, we need to address it effectively using a robust algorithm. As a consequence, in this study, we develop a modified version of Gaining-Sharing Knowledge (GSK) optimization algorithm to handle this problem as discussed below.

The GSK draws on the theory of knowledge learning and sharing throughout the lifespan of the human. It takes place on the basis of two essential phases, the first level is known as the beginning-intermediate or the junior phase, and the second phase is known

0093-9994 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: b-on: UNIVERSIDADE DO PORTO. Downloaded on March 25,2022 at 13:22:11 UTC from IEEE Xplore. Restrictions apply.

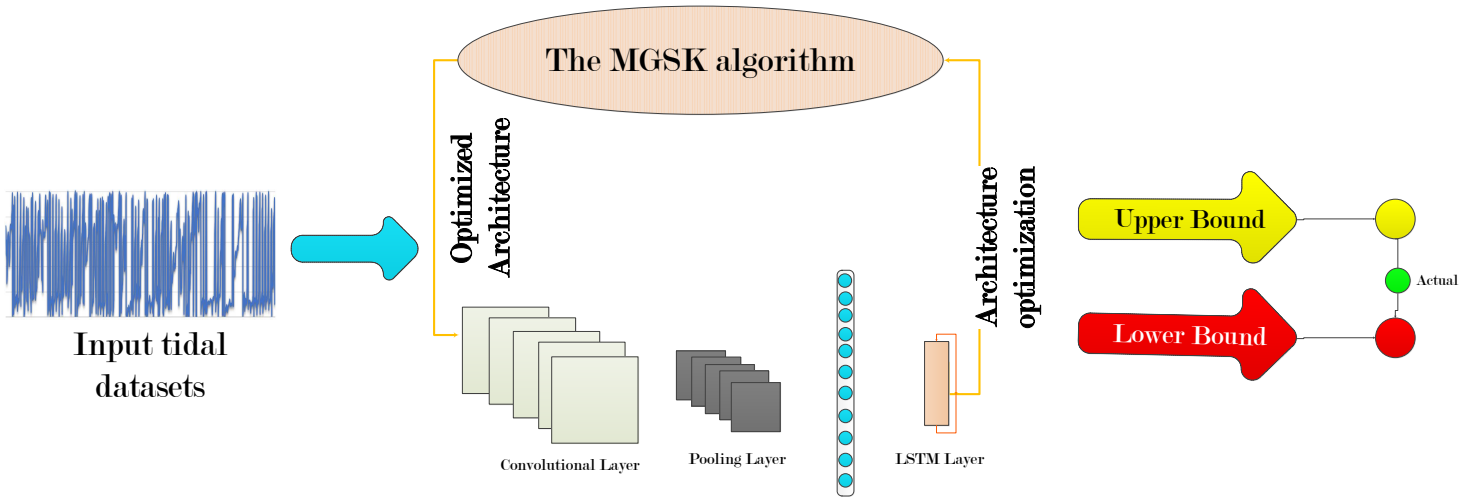

Fig. 2: The schematic diagram of the proposed DNE-based uncertainty quantification (MGSK-CLSTM-LUBE) model.

as the intermediate or senior level. The two stages are defined respectively in the following.

Firstly, the above-mentioned idea of junior phase is clarified in the following mathematical formalism:
Consider $x_i, i = 1, 2, 3, \ldots, N$ is a population of individuals in which N individuals remain in such population and x_i indicates $x_{i1}, x_{i2}, \ldots, x_{iD}$, in which D denotes to the quantity of fields of expertise and $f_i, i = 1, 2, 3, \ldots, N$ implies each corresponding fitness value.

The number of dimensions (D) at the beginning of the search process is determined on the basis of the concept of gaining-sharing knowledge depending on the aforementioned non-linear phrase termed as the experiencing equation.

$$D\left(juniorphase\right) = \left(problemsize\right) \times \left(1 - \frac{G}{GEN}\right)^{k} \tag{6}$$

where k represents a positive knowledge cost number. G is the generation number and GEN is indicative of maximum number of generations.

$$D\left(seniorphase\right) = problemsize - D\left(juniorphase\right) \tag{7}$$

As the original work reveals, we regard the k parameter as equal to 2.

In junior level, each individual is organized in the ascending order as x_{best} , ... ,x_{i-1} ,x_i ,x_{i+1} , ... , x_{worst} as per their cost function. After this, each individual is assembled to obtain a gaining knowledge source by two different individuals: the better (x_{i-1}), and the worst (x_{i+1}) than the current individual (x_i). The algorithm also randomly generates a source of another individual for sharing knowledge procedure.

As each individual is grouped in a higher order, they are classified into three divisions in the senior knowledge sharing level, namely the best, the better, and the worst based on the corresponding fitness function. Besides, in each of the individual throughout the senior scenario, two variables are randomly chosen from the maximum and downside of the current population, whereas one of the third variables is randomly assigned from the middle individual to form the sharing component.

Although GSK shows its great ability in solving several real-world optimization problems, we modify it to ameliorate the shortage of population diversity, the disparity between exploitation and exploration phases, and the GSK's convergence speed. To this end, a modified version of GSK algorithm named MGSK is developed by incorporating two evolutionary operators in the main search process of GSK algorithm. The flowchart of the proposed MGSK algorithm is shown in Fig. 1. In the following, we explain the two modifications considered in the original version of GSK algorithm according to the two utilized evolutionary operators.

-First modification:

Cauchy mutation (CM) is an evolutionary operator that helps increasing the diversify the population and consequently enhancing the exploration of the search space. This operator is based on a single-dimensional probability density function specified by:

$$f(x) = \frac{1}{\pi}\frac{t}{x^2 + t^2}, \quad t > 0 \;\&\; -\infty < x < \infty \tag{8}$$

where t represents a scalar parameter. The following is the corresponding Cauchy distribution function:

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x}{t}\right) \tag{9}$$

The mutation operator affects the search agent's population and allows them in bypassing the local minimum. The usage of this operator in the GSK is described by:

$$W_j = \frac{\left(\sum_{i=1}^{P} x_{ij}\right)}{P} \tag{10}$$

where W_j denotes to a weight matrix, x_{ij} is the ith position, and the population size is signified as P by the ith search agent. The following formula is given by the Cauchy mutation operator:

$$x'_j = x_j + W_j \cdot M \tag{11}$$

where M represents a random value.

-Second modification:

The parameter k serves an important role in the convergence speed of GSK algorithm. In order to improve the convergence speed of the GSK, we tune this parameter with tent map as an efficient chaotic evolutionary operator for harmonizing and escaping from the local solutions. The following formula represents the tent map:

$$k = \begin{cases} \frac{k}{0.7} & k < 0.7 \\ \frac{10}{3}(1-k) & k \geq 0.7 \end{cases} \tag{12}$$

0093-9994 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: b-on: UNIVERSIDADE DO PORTO. Downloaded on March 25,2022 at 13:22:11 UTC from IEEE Xplore. Restrictions apply.

It is worth noting that the exploration and exploitation are two important phases in evolutionary algorithms which can significantly impact the ability of these algorithms in finding optimal solutions among the search space. The main challenge is to make a balance between the exploration and exploitation phases leading to an improvement in the performance of evolutionary algorithms. The main purpose of the proposed MGSK algorithm is to enhance the quality of exploration and exploitation phases in the original version of GSK algorithm. For this purpose, we use Cauchy mutation operator to update the positions of the solutions in GSK algorithm. This operator is applied in each iteration to obtain a new position for each solution by updating its previous value in each dimension according to Eq.(11). The main advantage of the Cauchy mutation operator is to increase the diversity of searched areas which allow the optimization process to search further areas in the search space. Also, increasing the diversity of searched areas results in reducing the probability of falling into local optima as well as increasing the probability of finding global optima. On the other hand, we use the tent map operator in Eq.(12) to tune the value of the parameter k used as a positive knowledge cost number in Eq.(6). The main advantage of the tent map operator is to speed-up the convergence of the GSK algorithm as well as escape from the local solutions. According to the two modifications of the proposed MGSK algorithm, we can claim that our proposed evolutionary algorithm can outperform the original version of GSK algorithm by obtaining more effective solutions for the optimization problem. We will prove this claim experimentally in the experiments section.

The proposed MGSK algorithm is used to optimize the hyperparameters of deep CLSTM model. In other words, we aim to propose an effective deep neuroevolution-based uncertainty quantification framework for tidal current forecasting named MGSK-CLSTM-LUBE based on deep CLSTM model in which the hyperparameters of deep CLSTM model are determined automatically through the proposed MGSK algorithm. To this end, the values of the CLSTM's hyperparameters are considered to define the solutions in the optimization process. Then, the proposed MGSK algorithm is performed to obtain the best optimal solution containing the values of the CLSTM's hyperparameters. Finally, the optimized deep CLSTM model is used to predict the uncertainty of tidal values. The pseudo-code of the proposed MGSK-CLSTM-LUBE method is represented in Algorithm 1.

In Fig. 2, the schematic diagram of our optimization-based hybrid algorithm is demonstrated. As can be seen from this figure, input tidal datasets are fed to the deep CNN-LSTM model optimized by the proposed MGSK model to calculate the uncertainty of tidal datasets.

III. Experimental Results

In this section, experimental data from the Bay of Fundy, NS, Canada are used to assess the efficiency of the proposed deep neuroevolution-based uncertainty model. From September to October 2008, the tidal dataset is processed for 10 minutes of instant readings. Including both tidal current speed and direction, the numerical simulations are carried out to demonstrate the strong functionality of our proposed model. Tidal current dataset is separated among three training, validation and test categories, with a 50%, 40%, and 10% of the actual tidal data, respectively.

To measure the performance of our proposed algorithm in calculating the uncertainty of tidal datasets, we compete our model with nine high-powerful deep learning-based probabilistic algorithms.

TABLE II: The simulation results of the proposed model vs other competitive deep learning models for tidal current speed datasets.

Model	PICP	PINAW	CWC
TDNN	84.8743	39.2823	0.2341
GRU	85.5301	38.6321	0.2239
LSTM	86.1795	37.6423	0.2208
CNN-LSTM	88.5102	37.5213	0.2054
GA-CNN-GRU	88.7339	37.3901	0.1875
GA-CNN-LSTM	89.1206	36.9802	0.1866
PSO-CNN-LSTM	89.2045	36.7014	0.1834
DE-CNN-LSTM	91.3449	34.0939	0.1709
GSK-CNN-LSTM	92.0893	31.8038	0.1641
CMGSK-CNN-LSTM	92.4541	31.4292	0.1625
CGSK-CNN-LSTM	92.8977	30.8956	0.1598
MGSK-CNN-LSTM	**96.3563**	**28.6801**	**0.1509**

TABLE III: The simulation results of the proposed model vs other competitive deep learning models for tidal current direction datasets.

Model	PICP	PINAW	CWC
TDNN	80.99739	41.46123	0.5152
GRU	82.89291	40.48892	0.4735
LSTM	83.08569	40.09688	0.3501
CNN-LSTM	83.27367	37.70943	0.2463
GA-CNN-GRU	83.83852	37.65725	0.2685
GA-CNN-LSTM	84.45945	36.82787	0.1486
PSO-CNN-LSTM	86.95083	37.14851	0.1911
DE-CNN-LSTM	87.5905	35.50966	0.1769
GSK-CNN-LSTM	89.84966	34.74806	0.1605
CMGSK-CNN-LSTM	90.05432	34.13375	0.1587
CGSK-CNN-LSTM	90.67143	33.56339	0.1523
MGSK-CNN-LSTM	**92.98591**	**29.53225**	**0.1428**

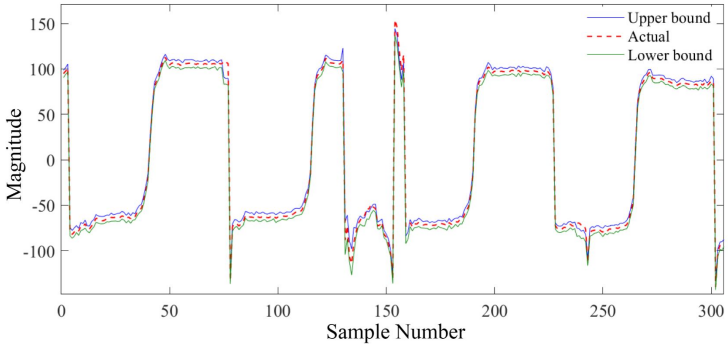

Fig. 3: Optimal PIs for tidal current speed with our proposed algorithm (90% confidence level).

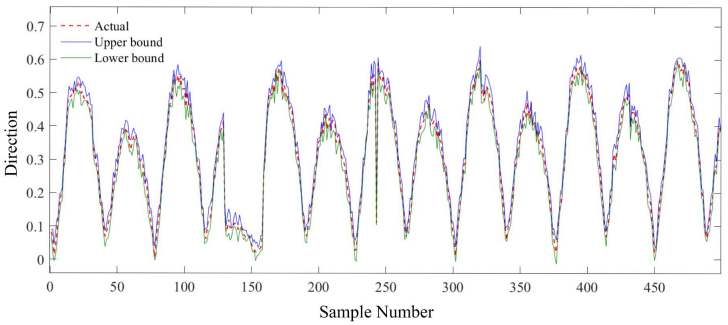

Fig. 4: Optimal PIs for tidal current direction with our proposed algorithm (90% confidence level).

0093-9994 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: b-on: UNIVERSIDADE DO PORTO. Downloaded on March 25,2022 at 13:22:11 UTC from IEEE Xplore. Restrictions apply.

Among these nine deep learning methods, four of them belong to the powerful deep learning algorithms benchmarked to measure the ability to solve time series problems, which are: traditional deep neural network (TDNN), gated recurrent unit (GRU), LSTM and CLSTM. We also use robust evolutionary algorithms such as genetic algorithm, particle swarm optimization, differential evolution, standard version of GSK algorithm, and two models based on the modifications applied to the GSK including CMGSK-CNN-LSTM (Cauchy mutation-based model) and CGSK-CNN-LSTM (Chaotic tent map based model) to measure the searching ability of the MGSK algorithm in order to quantify the forecasting uncertainty. The population size is set to 40 and the maximum iteration value is fixed to 30 for the MGSK and other evolutionary algorithms for having the fair comparison among the benchmarked models. In the following, we discuss the ability of our proposed probabilistic model to measure the uncertainty of the current speed and direction of the Canadian tidal datasets.

Regarding the current speed forecasting, in Table II, it can be seen the experimental results related to three performance indicators including PICP, PINAW, and CWC. Among the deep learning-based benchmarked algorithms, the algorithms that did not use optimization in their training process have weaker performance among the three PICP, PINAW and CWC indices. For instance, the lower PICP (82.15459), higher PINAW (42.86452) and higher CWC (0.323344) are devoted to TDNN algorithm (82.15459). As can be seen from the results of this table, the best performance among the three uncertainty quantification indices (PICP equal to 94.10344, PINAW equal to 30.01873 and CWC equal to 0.184105) has our proposed algorithm, in which its closest follower is the standard version of the GSK algorithm.

In Fig. 3, the forecasting performance outcomes of the tidal current speed for our proposed MGSK model is represented. In this figure, the actual tidal speed signal is shown by red color, whereas the upper bound and lower bound forecasting data points are represented by blue and green colors. As this figre clearly shows, the forecasting bounds can cover accurately and efficiently the forecasting acutal data points with the optimal PICP and PINAW of 94.10344 and 30.01873, respectively. Based on this figure, it can be well inferred that our proposed probabilistic framework can accurately track the variations in tidal current speed. Fig. 4 demonstrates the PI prediction results where data observations for the actual current tidal direction are observed by red. Based on these observations, the tidal current rate can efficiently be followed by the prediction intervals.

Fig. 5 indicates the convergence profiles of the proposed MGSK in comparison to other evolutionary techniques over 30 iterations based on the CWC as the fitness function. As can be observed by Fig. 5, the MGSK algorithm can effectively converge in finding the best solution. These findings indicate that the proposed MGSK is preferable to other optimization evolutionary techniques based on the fitness metric (CWC). On the other hand, the violin diagrams in Fig. 6 show the six main hyperparameters of CLSTM model in the optimization process by the proposed MGSK algorithm. It should also be noted that this figure is for quantitative analysis of CLSTM hyperparameters that all of these ones are specified in details and are depicted based on the values of the hyperparameters in each run of the proposed algorithm. As these diagrams show, our proposed algorithm has selected the values as close as possible to the minimum interval limit of the desired hyperparameters, which computationally obtain the most minimum possible computational volume to determine the optimal deep CLSTM architecture automatically without the trial and error procedure.

We also investigate the forecasting ability of our proposed model over the tidal current direction time series signals. The simulation results raised by the proposed model and other various algorithms are shown in Table III. The experimental results demonstrate the proposed method's strong capacity to maximize the CWC objective. An essential consideration derived from this table shows that the fact that the optimisation of CWC can efficiently lead to higher PICPs and lower PINAWs towards more optimal PIs. In addition, CWC optimization can further enhance the quality of PIs, although it is not a direct correlation.

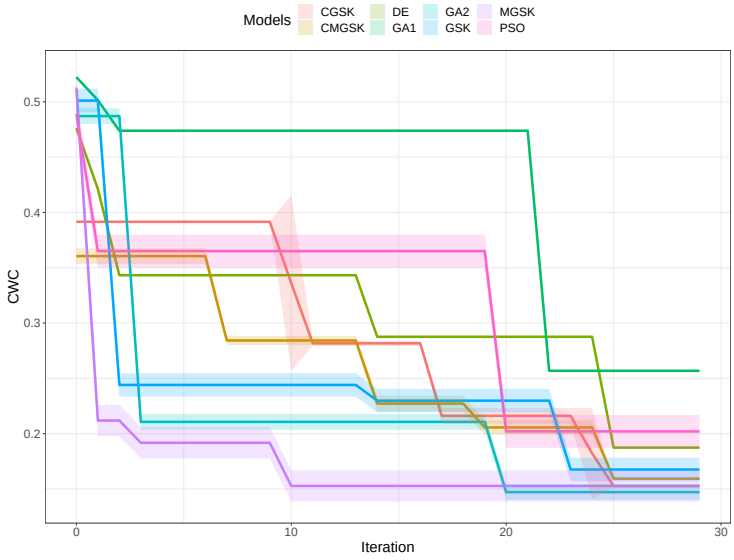

Fig. 5: The convergence profiles of different evolutionary algorithms.

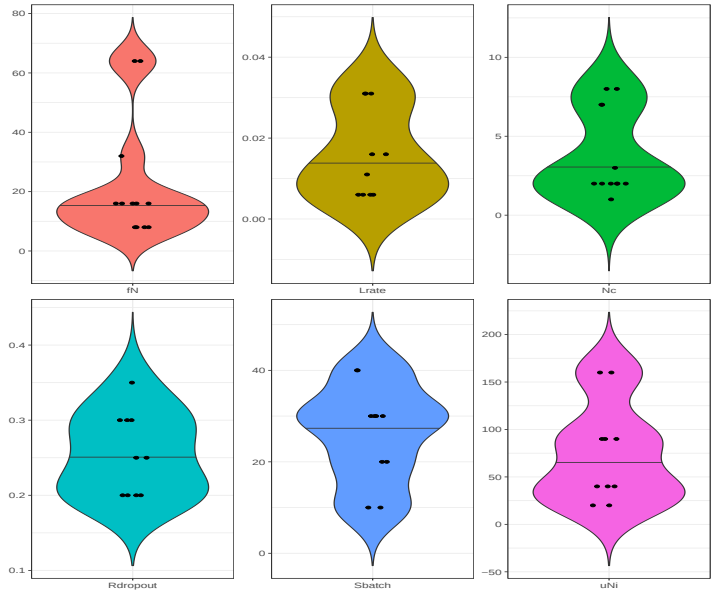

Fig. 6: The violin plots of various evolved CLSTM hyperparameters. It should be noted that this figure examines the evolved hyperparameters, where the horizontal axis represents the type of hyperparameters and the vertical axis represents the value of each hyperparameter.

IV. Conclusion

In this work, we proposed a powerful stochastic deep learning-based hybrid framework centered on deep CLSTM and LUBE

0093-9994 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: b-on: UNIVERSIDADE DO PORTO. Downloaded on March 25,2022 at 13:22:11 UTC from IEEE Xplore. Restrictions apply.

using optimal PIs in order to forecast the uncertainty of the tidal current speed and direction data points. We also develop a novel powerful deep neuroevolution strategy named as MGSK to optimize the architectures of CLSTMs automatically in order to construct the lower and upper bounds efficiently. The numerical simulations of the Bay of Fundy, NS, Canada tidal current data indicate that the proposed approach can tackle the uncertainty prediction accurately. Furthermore, the optimal PIs can adopt fluctuations in tidal current speed and direction, ensuring that the acceptable level of confidence is reached while also reducing the average width of PIs.

REFERENCES

[1] S. M. J. Jalali, M. Khodayar, S. Ahmadian, M. K. Noman, A. Khosravi, S. M. S. Islam, F. Wang, and J. P. Catalão, "A new uncertainty-aware deep neuroevolution model for quantifying tidal prediction," in *2021 IEEE Industry Applications Society Annual Meeting (IAS)*. IEEE, 2021, pp. 1–6.

[2] S. E. B. Elghali, M. E. H. Benbouzid, and J. F. Charpentier, "Marine tidal current electric power generation technology: State of the art and current status," in *2007 IEEE International Electric Machines Drives Conference*, vol. 2, 2007, pp. 1407–1412.

[3] H. H. H. Aly and M. E. El-Hawary, "A proposed ann and flsm hybrid model for tidal current magnitude and direction forecasting," *IEEE Journal of Oceanic Engineering*, vol. 39, no. 1, pp. 26–31, 2014.

[4] A. Kavousi-Fard, "A hybrid accurate model for tidal current prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, pp. 112–118, 2017.

[5] A.Doodson, "The harmonic development of the tide-generating potential," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 100, no. 704, pp. 305–329, dec 1921. [Online]. Available: https://doi.org/10.1098%2Frspa.1921.0088

[6] A. Doodson, *The Analysis and Prediction of Tides in Shallow Water*, ser. International hydrographic review. Special publication. International Hydrographic Bureau, 1957. [Online]. Available: https://books.google.com.au/books?id=cnv6QwAACAAJ

[7] K. E. Leffler and D. A. Jay, "Enhancing tidal harmonic analysis: Robust (hybrid l1/l2) solutions," *Continental Shelf Research*, vol. 29, no. 1, pp. 78–88, 2009, physics of Estuaries and Coastal Seas: Papers from the PECS 2006 Conference. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278434308001842

[8] A. Kavousi-Fard, "A novel probabilistic method to model the uncertainty of tidal prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 828–833, 2017.

[9] T. Lee and D. Jeng, "Application of artificial neural networks in tide-forecasting," *Ocean Engineering*, vol. 29, no. 9, pp. 1003–1022, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801801000683

[10] T.-L. Lee, "Back-propagation neural network for long-term tidal predictions," *Ocean Engineering*, vol. 31, no. 2, pp. 225–238, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002980180300115X

[11] T.-L. Lee, O. Makarynskyy, and C.-C. Shao, "A Combined Harmonic Analysis–Artificial Neural Network Methodology for Tidal Predictions," *Journal of Coastal Research*, vol. 2007, no. 233, pp. 764 – 770, 2007. [Online]. Available: https://doi.org/10.2112/05-0492.1

[12] S. Liang, M. Li, and Z. Sun, "Prediction models for tidal level including strong meteorologic effects using a neural network," *Ocean Engineering*, vol. 35, no. 7, pp. 666–675, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801808000024

[13] P. Remya, R. Kumar, and S. Basu, "Forecasting tidal currents from tidal levels using genetic algorithm," *Ocean Engineering*, vol. 40, pp. 62–68, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801811002691

[14] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.

[15] B. Kappen and S. Gielen, *Neural Networks: Best Practice in Europe*, pp. 1–226. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9789814529020

[16] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, vol. 1, 1994, pp. 55–60 vol.1.

[17] A. Riazi, "Accurate tide level estimation: A deep learning approach," *Ocean Engineering*, vol. 198, p. 107013, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801820300901

[18] S. M. J. Jalali, G. J. Osorio, S. Ahmadian, M. Lotfi, V. Campos, M. Shafie-khah, A. Khosravi, and J. P. S. Catalao, "A new hybrid deep neural architectural search based ensemble reinforcement learning strategy for wind power forecasting," *IEEE Transactions on Industry Applications*, pp. 1–1, 2021.

[19] S. M. J. Jalali, S. Ahmadian, A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "Automated deep cnn-lstm architecture design for solar irradiance forecasting," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, pp. 54–65, 2022.

[20] S. M. J. Jalali, S. Ahmadian, A. Khosravi, M. Shafie-khah, S. Nahavandi, and J. P. S. Catalão, "A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting," *IEEE Transactions on Industrial Informatics*, vol. 17, pp. 8243–8253, 2021.

[21] J. Long, S. Zhang, and C. Li, "Evolving deep echo state networks for intelligent fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4928–4937, 2020.

[22] R. Miikkulainen, J. Liang, E. Meyerson, A. Rawal, D. Fink, O. Francon, B. Raju, H. Shahrzad, A. Navruzyan, N. Duffy, and B. Hodjat, "Chapter 15 - evolving deep neural networks," in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, R. Kozma, C. Alippi, Y. Choe, and F. C. Morabito, Eds. Academic Press, 2019, pp. 293–312. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128154809000153

[23] Y. Hua, L. Mou, and X. X. Zhu, "Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional lstm network for multi-label aerial image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 149, pp. 188–199, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619300243

[24] S. Ahmadian, S. M. J. Jalali, S. M. S. Islam, A. Khosravi, E. Fazli, and S. Nahavandi, "A novel deep neuroevolution-based image classification method to diagnose coronavirus disease (covid-19)," *Computers in Biology and Medicine*, vol. 139, p. 104994, 2021.

[25] S. M. J. Jalali, M. Ahmadian, S. Ahmadian, A. Khosravi, M. Alazab, and S. Nahavandi, "An oppositional-cauchy based gsk evolutionary algorithm with a novel deep ensemble reinforcement learning strategy for covid-19 diagnosis," *Applied Soft Computing*, vol. 111, p. 107675, 2021.

[26] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proceedings of the 31st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, E. P. Xing and T. Jebara, Eds., vol. 32, no. 2. Bejing, China: PMLR, 22–24 Jun 2014, pp. 1764–1772. [Online]. Available: http://proceedings.mlr.press/v32/graves14.html

[27] C. H. Yang, C. H. Wu, and C. M. Hsieh, "Long short-term memory recurrent neural network for tidal level forecasting," *IEEE Access*, vol. 8, pp. 159 389–159 401, 2020.

[28] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE transactions on neural networks*, vol. 22, no. 3, pp. 337–346, 2010.

[29] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 2, pp. 303–315, 2013.

[30] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using pso-tuned support vector machines," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 2, pp. 322–331, 2015.

0093-9994 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: b-on: UNIVERSIDADE DO PORTO. Downloaded on March 25,2022 at 13:22:11 UTC from IEEE Xplore. Restrictions apply.