IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 27, 2018, accepted May 8, 2018, date of publication June 4, 2018, date of current version July 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2836917

Neural Network-Based Uncertainty Quantification: A Survey of Methodologies and Applications

H. M. DIPU KABIR, (Student Member, IEEE), ABBAS KHOSRAVI, (Member, IEEE), MOHAMMAD ANWAR HOSEN, AND SAEID NAHAVANDI, (Senior Member, IEEE)
Institute for Intelligent Systems Research and Innovation, Deakin University, Geelong, VIC 3220, Australia
Corresponding author: H. M. Dipu Kabir (dkabir@deakin.edu.au)

ABSTRACT Uncertainty quantification plays a critical role in the process of decision making and optimization in many fields of science and engineering. The field has gained an overwhelming attention among researchers in recent years resulting in an arsenal of different methods. Probabilistic forecasting and in particular prediction intervals (PIs) are one of the techniques most widely used in the literature for uncertainty quantification. Researchers have reported studies of uncertainty quantification in critical applications such as medical diagnostics, bioinformatics, renewable energies, and power grids. The purpose of this survey paper is to comprehensively study neural network-based methods for construction of prediction intervals. It will cover how PIs are constructed, optimized, and applied for decision-making in presence of uncertainties. Also, different criteria for unbiased PI evaluation are investigated. The paper also provides some guidelines for further research in the field of neural network-based uncertainty quantification.

INDEX TERMS Prediction interval, uncertainty quantification, heteroscedastic uncertainty, neural network, forecast, time series data, regression, probability.

I. INTRODUCTION

Many engineering and scientific problems consisted of partly deterministic and partly random situations. The traditional point prediction is unable to predict the level of randomness or uncertainty. Prediction Intervals (PIs) have been extensively used in a range of applications for over 50 years [1]–[3] to quantify that uncertainty. Mostly to overcome the limitations of the point prediction, an interval prediction is widely accepted in many fields of study including economics [4], food industry [5], tourism [6], medical statistics [2], power consumption [7], even in compression algorithms [8]. Moreover, the recent installation of the large-scale renewable energy [9], [10], the rapid growth of the online auction systems [11], [12], and the design of different types of autonomous robots [13] are increasing the uncertainty and therefore, increasing the essence of the probabilistic forecasting. Fig. 1 presents the importance of PI with a rough sketch. The point prediction gives a value close to the median or the mean of the probable values of targets. Two green lines in Fig. 1 presents the PI. The width of the interval changes based on the probable values of the target. Observing the point forecast, the user is not provided by any complement information about the uncertainty of the system [14]–[16]. Therefore, interval forecasts are a popular method of uncertainty quantification.

The uncertainty in risk analysis processes is traditionally classified as follows:

1) Aleatory uncertainty (inherent randomness): The output of a system may vary slightly from time to time even for the same set of inputs. There might be day to day or year to year variation while the inputs are the same. Also, there might be no trend of such changes. Therefore these uncertainties are the inherent randomness or the aleatory uncertainty [17]. An example of such randomness is the electricity demand for a certain time. Although the temperature, the time in a day, the day in a week, humidity, wind speed is the same, the electricity demand can be slightly different [18], [19].
2) Epistemic (subjective) uncertainty: The uncertainty may also happen due to the secondary or tertiary effects, not considered during the modeling. It can be

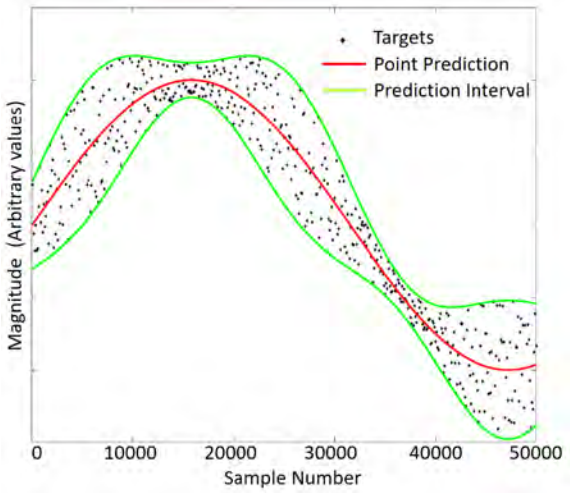

FIGURE 1. A rough sketch presenting the importance of PI. Point prediction provides a value close to mean or median and contains no information about the variation.

an effect from a phenomenon, which is unknown to the research community or can be the effect of an internal parameter of an object, not readable from outside. In such scenario, the uncertainty can be reduced by enhancing knowledge or by performing measurements. An example of such uncertainty is the strength of solids such as steel and concrete. The internal defects of a solid material are difficult to measure from the outer appearance and even from the formation process. However, the provider can provide a range of the strength parameters. With additional modeling efforts, the uncertainty parameter can be transformed from a random variable to a bounded random or pseudorandom variable using probability intervals or percentile ranges [19]–[23].

These two kinds of uncertainties may present together or separately. The point prediction is unable to provide any information about them.

The probability density function (PDF) contains exact information about the uncertainty [24]. However, It is impossible to present the probability density with a few numbers or words. The PDF may extend towards infinity along either side of the most probable region. The value of the PDF is also confusing. The PDF is commonly drawn as the bar chart or a continuous function presenting probability density of nearby regions. Then the presenter needs to provide the width of each bar or the range of the nearby region considered. Cumulative probability density function (CDF) solves that issue of confusing amplitude, as its value is ranged from zero to one. However, CDF line is also a difficult parameter to express it to the majority of the user and it requires more information to represent CDF. Probabilistic forecast [25]–[28] expresses the uncertainty with several numbers, presenting percentile distribution of probability. PIs express the uncertainty with the minimum information (three numbers) and therefore, PIs are the most understandable uncertainty quantification mechanism.

The point prediction provides a value where the statistical error is the lowest. Statistical errors of the point prediction are defined by error metrics, such as the root mean square error (RMSE) and the mean absolute percentage error (MAPE). A PI with $(1-\alpha)$ confidence level can be derived probabilistically with the following flow of information: The upper bound $(\overline{y}_i)$ is higher than $(1-\alpha/2)$ portion of the probability density function. Therefore, the upper bound can be represented as [29]:

$$P(t_i < \overline{y}_i) = 1 - \alpha/2 \tag{1}$$

where, P is the conditional probability function of target (t_i). That relation can be represented by the cumulative probability density function (CP) as follows:

$$CP(\overline{y}_i) = 1 - \alpha/2 \tag{2}$$

Taking inverse:

$$\overline{y}_i = CP^{-1}(1 - \alpha/2) \tag{3}$$

Therefore, the PI consists of the lower bound $(\underline{y}_i)$ and the upper bound $(\overline{y}_i)$ can theoretically be represented as:

$$[\underline{y}_i, \overline{y}_i] = [CP^{-1}(\alpha/2), CP^{-1}(1 - \alpha/2)] \tag{4}$$

That relation is correct for a symmetric probability distribution on both sides of its maxima. Many PI construction techniques do not follow that relation for an asymmetric probability distribution. A survey of recently proposed NN-based PI construction techniques is presented to help analysts and future researchers in understanding the relationship between PIs and the probability density function. The scope of our current study is limited to NN-based PIs.

PIs have been increasingly applied to design systems as a promising mechanism to quantify the uncertainties prevailing in operation of complex systems [30]–[33]. Interval forecasts contain a greater significance to decision-makers compared to the traditional point forecasts and therefore, can be applied confidently in a range of applications and for contingency planning. They provide more robust and meaningful information of the future inevitable uncertainty. Both PI and confidence interval (CI) are popular forms of interval forecasts. They provide an interval forecast with certain success probability, known as the confidence level (CL) [34] and the PI coverage probability (PICP) [35]. The CL and PICP are the probability that the interval will enclose the target. PI applies the Bayesian statistics, and CI applies the Frequentist statistics. The Frequentist statistics consider previous occurrences of the predicted event where Bayesian statistics considers both corresponding and at least one correlated event. Therefore, PI is more robust compared to CI. Three values are required to represent a PI: the upper bound $(\overline{y})$, the lower bound $(\underline{y})$ and the coverage probability (PICP). Therefore, the user of PI is aware of both the most probable range and the probability of enclosure [36].

In the decades 1960-2000s, many researchers identified the importance of probabilistic forecasting and developed some statistical methods to construct prediction intervals. The methods include, normal approximation [37], [38], hypothesis testing approach [39], pivotal quantity [40], [41], sufficient statistic [42], sequential search [43], Bayesian [44], bootstrap [45] and other statical models, to name a few. Interval forecasts brought a new dimension during these periods to solve some statistical problems emerging in the quality control and the business (economic prediction). For instance, PIs could provide warranty limits for the future effects of a specified number of systems based on the past data or future mean performance of a specific product and help the investor to understand the future scenario [46], [47].

At the end of the 20th century, artificial intelligence-based regression technique, namely neural network (NN) has appeared as an effective method to quantify uncertainties through the construction of PIs. PIs using NNs can be found in more diverse fields of studies. These include but not limited to manufacturing systems, chemical processes, wind power forecasting, electricity load forecasting, transportation, fatigue lifetime prediction, financial services, health-related issues, hydrological studies, baggage handling systems and food industries [5], [36], [48]–[50]. NNs not only improve the quality of the PIs but also improves the run-time computation efficiency. On the other hand, PIs also brought a new dimension for NN-based modeling and strategy developments. Such as, recently developed game strategic systems are applying the upper confidence bound (UCB) to optimize the exploration time [51]–[53].

NN-based PI construction methods can be classified into following categories:

1. NN-based Multi-step PI Construction Methods.
   a) Delta Method
   b) Bayesian Method
   c) Mean-Variance Estimation Method (MVEM)
   d) Bootstrap Method
   e) Modified Delta Method
   f) Modified MVEM Method
   g) Modified Bootstrap Method
2. NN-based Direct PI Construction Methods.
   a) Lower Upper Bound Estimation (LUBE) Method
   b) Direct Interval Forecasting by C. Wan
   c) Normalized Root-Mean-Square Width (PINRW) and Particle Swarm Optimization (PSO) by H Quan
   d) Independent Width and Penalty Factors by A Khosravi
   e) Deviation from Mid-interval Consideration by G. Marn
   f) Improvement through Optimal Aggregation by A Hosen
   g) Deviation Information-based Criterion by G Zhang

The current survey is aimed help future researchers of NN based uncertainty quantification in constructing smarter PI and smarter NN training methods. Therefore, the scope of the current survey is limited to NN-based approaches of constructing PIs. The paper is organized into various sections with the following flow of information. Section II presents a brief review of traditional NN-based multi-step PI construction methods. Recent advancement on NN-based PI construction methods is described in section III. An overview of applications of PI in different fields is provided in section IV. Section IV also presents the future possibilities of PI. Section V is the concluding section.

II. NN-BASED MULTI-STEP PI CONSTRUCTION METHODS

Four traditional methods containing multiple stages in constructing PIs have been reported in the literature to construct quality PIs using NNs. The first one is the Delta method proposed by Hwang and Ding [54]; the nonlinear regression representation of the NN-based system performed to obtain the PI. The second one is the Bayesian Method which applies Bayes' theorem to optimize the weight. Nix and Weighed [55] proposed another traditional technique, namely Mean-Variance Estimation method (MVEM) to construct PIs using two NNs. Like other traditional methods, MVEM also assumes normally distributed error probability around the average of the target, t_j and PIs can be easily formed if mean and variances are known. The fourth one is the Bootstrap method which applies resampling to form an ensemble of several NNs to construct a quality PI.

A. DELTA METHOD

The Delta method is a strategy for constructing intervals through nonlinear regression. This method is based on the Tylor series expansion of the regression function. In this method, the NN model is linearized through optimizing a set of parameters by minimizing the error-based cost function, sum square error (SSE). The standard asymptotic theory is then applied to the developed NN model for the construction of PIs [56]. It is assumed that the uncertainty is normally distributed and homogeneous. However, in practice, the uncertainty is heterogeneous in most of the cases. Therefore, this method is not suitable for all cases [57]. An extended version of delta method has been proposed by Veaux *et al.* [58] to eliminate this limitation. In contrast to SSE, they used weight decay cost function to train the NN and improved the generalization power of the NN. In another study, Khosravi *et al.* [59] proposed a new measure for quantitative assessment of PIs for the delta method. As width and coverage probability includes in this measure, the modified version of this delta method produced better quality PIs (significantly reduced the width of PIs) compared to traditional delta method. However, still, the constructed PIs suffer from the basic constraint of this method (*i.e.* linearization). Despite this limitation, the delta technique has been used in numerous case studies [60]–[62].

In delta method, the total variance of a Markov decision process is represented as [63]:

$$\sigma_0^2 = \sigma_\epsilon^2(1 + g_0^T(\mathbf{J}^T\mathbf{J})^{-1}g_0) \tag{5}$$

where J is the NN model's Jacobian matrix, σ_ϵ^2 can be obtained from:

$$\sigma_\epsilon^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(t_j - \hat{y}_j\right)^2$$

The distribution is assumed as a Gaussian one and $(1-\alpha)\%$ PI for $\hat{y}_i$ is formulated as [54]:

$$\hat{y}_0 \pm t_{n-p}^{1-\frac{\alpha}{2}}\sigma_\epsilon\sqrt{1+g_0^T(\mathbf{J}^T\mathbf{J})^{-1}g_0} \tag{6}$$

where $t_{n-p}^{1-\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile of a cumulative distribution function (CDF) with $n-p$ degrees of freedom.

As discussed in [58], the PI constructed using (6) is reliable for large training data sets. However, the NN training process for converging the matrix J^TJ can be nearly singular for a small dataset and that may result in unreliable PIs. Moreover, termination of the training process prior to convergence to avoid overfitting reduces the number of parameters results in wide PIs.

The Weight Decay Cost Function (WDCF) in training process can solve the above problems and enhance the generalization power of NN [64]. The WDCF attempts to maintain the magnitude of NN-parameters to the lowest feasible value [65]:

$$WDCF = \lambda\ w^T\ w + SSE \tag{7}$$

The PI construction formula using WDCF is derived as [58],

$$\hat{y}_0 \pm t_{n-p}^{1-\frac{\alpha}{2}}\sigma_\epsilon\sqrt{1+g_0^T(\mathbf{J}^T\mathbf{J}+\lambda I)^{-1}(\mathbf{J}^T\mathbf{J})(\mathbf{J}^T\mathbf{J}+\lambda I)^{-1}g_0} \tag{8}$$

The insertion of λ in (8) enhances the quality and reliability of PIs, especially for circumstances that J^TJ is nearly singular.

The delta method is computationally expensive as this technique involved the complex calculation of Jacobian matrix ($\mathbf{J}$) and σ_ϵ^2. Moreover, the gradient and the Jacobian matrix calculations, and the estimation of $\sigma_{\hat{y}_0}^2$ can be potential sources of error for this technique [66]. In addition, this technique estimates a constant σ_ϵ^2 for all samples; that means the noise is assumed to be uniform. However, it is not true for all situations, the order of noise can be linked to the target magnitude or the NN input sets in several practices. Therefore, the delta method fails to produce quality PIs for some situations.

B. BAYESIAN METHOD

The Bayesian learning involves the network training for the distribution of weights. Bayes' theorem is applied to optimize the weights (such as posterior distribution) from the assumed prior distribution. The predictive distribution of network outputs is then evaluated using the posterior distribution. The Bayesian technique provides a natural framework for estimating prediction intervals as described below [67], [68],

In the Bayesian method, the total variance, σ_i^2 can be rewritten for NN training as [63]:

$$\begin{aligned}\sigma_i^2 &= \sigma_D^2 + \sigma_{w^{MP}}^2 \\ &= \frac{1}{\beta} + \nabla_{w^{MP}}^T\hat{y}_i\ \ (H^{MP})^{-1}\ \ \nabla_{w^{MP}}\hat{y}_i\end{aligned} \tag{9}$$

If the total probability distribution of the i^{th} future sample is available, an $(1-\alpha)\%$ PI can be obtained [69]:

$$\hat{y}_i \pm z^{1-\frac{\alpha}{2}}\left(\frac{1}{\beta} + \nabla_{w^{MP}}^T\hat{y}_i\ \ (H^{MP})^{-1}\ \ \nabla_{w^{MP}}\hat{y}_i\right)^{\frac{1}{2}} \tag{10}$$

where $z^{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of the corresponding normalized probability distribution function with zero average value and the unit variance. Moreover, $\nabla_{w^{MP}}^T\hat{y}_i$ is the NN output gradient with respect to its parameters, w^{MP}.

According to the literature, the generalization power of NN models using the Bayesian technique is better than other traditional techniques. However, in spite of strong mathematical foundation of the Bayesian technique, it is computationally demanding in the development stage, likely delta technique. The cost function of this method involves Hessian matrix. The calculation of Hessian matrix is cumbersome and time-consuming for large datasets and NNs. Despite this, the computational complexity is reduced in the PI formation stage as the gradient of NN output is only required.

C. MEAN-VARIANCE ESTIMATION METHOD (MVEM)

The Delta and Bayesian methods use a fixed target variance for PI formation. The MVEM calculates the target variance from a dedicated NN [63]. Therefore, this method allows more flexibility for evaluating the heteroscedastic variance of the targets. The basic concept of MVEM is shown in Figure 2. There are two NNs, The set of inputs for these NNs can be identical or diverse. There is no restriction on the dimension and composition of two NNs. Thus, NN structure may vary as per data nonlinear patterns and experimental requirements. It is assumed that the output activation function for NN_σ is exponential ($\hat{\sigma}^2$) and this consideration resulting in strictly positive variance estimation. With the accurate estimation of $y(x)$ through the NN, the PIs can be approximated with a $(1-\alpha)\%$ confidence level; presented as follows:

$$\hat{y}(x, w_y) \pm z_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2(x, w_\sigma)} \tag{11}$$

where w_σ and w_y are parameters for NN_σ and NN_y, respectively. As the variance values of the target, σ_i, are not known a priori, tradition error-based minimization techniques cannot be applied to the training of NN_σ. Usually, the maximum likelihood estimation (MLE) method is implemented for the NN training. Considering normally distributed forecasting errors, the data conditional distribution can be written as [55]:

$$P(t_i \mid x_i, NN_y, NN_\sigma) = \frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}}\ e^{-\frac{(t_i-\hat{y}_i)^2}{2\hat{\sigma}_i^2}} \tag{12}$$

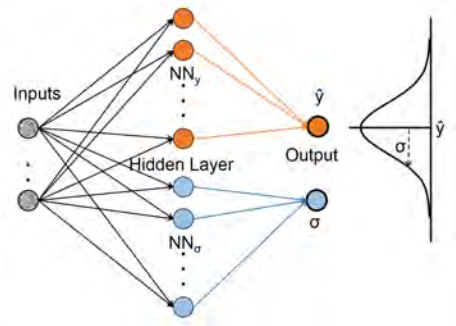

FIGURE 2. An overview on the NN-based mean-variance estimation (MVE) method for the computation of PIs.

where, i is the pattern number. Taking the natural logarithm and ignoring the constant terms of (12), the cost function for MVEM can be defined as:

$$C_{MVE} = \frac{1}{2}\sum_{i=1}^{n}[ln(\hat{\sigma}_i^2) + \frac{(t_i - \hat{y}_i)^2}{\hat{\sigma}_i^2}] \tag{13}$$

Using the cost function defined in (13), a three phase training method was introduced in [55] to optimize NNs parameters, w_σ and w_y. In the introduced method, two sets of Data (lets say, D_1 and D_2) are needed for the training NN_σ and NN_y. The brief description of the three phase optimization steps are as follows [63], [70],

- **Phase I:** Train the NN_y using D_1 dataset to estimate y_i. Traditional error-based optimization function is applied to optimize the NN parameters, w_y. D_2 dataset is used as the validation dataset to avoid overfitting problem. Please note that NN_σ is not trained in this phase.
- **Phase II:** The NN_y parameters, w_y optimized in previous phase are considered as fixed. Then, train the NN_σ using dataset 2, D_2. This time, the cost function defined in (13) is used to optimize the w_σ. NN_σ and NN_y are then applied to approximate respectively σ^2 and y_i for each sample. The optimization equation is then checked for the current set of the NN_σ weights (w_σ). Traditional gradient descent method can be used to update these weights.
- **Phase III:** In this stage, two new training sets (one for the training set and the other one for validation set) are re-sampled and applied for the simultaneous tuning of w_σ and w_y. NN_σ and NN_y is then re-trained again by the minimization of (13).

In contrast to Delta and Bayesian methods, MVEM is simple in terms of implementation. Calculation of time-consuming complex derivatives and matrixes, such as Hessian and Jacobian is not required for construction of PIs using MVEM.

As described earlier, this method assumes that NN_y accurately estimates the true mean of targets, y_i. This assumption is the main drawback of MVEM. It can be contravened in practice and leads to poor NN generalization power, and hence, the formed PIs through (11) may overestimate (or underestimate) the predefined CL $(1-\alpha)\%$ and that leads to a low PICP [63]. Another disadvantage of the MVEM is this method does not include misspecification of the model variables (either w_y or w_σ) to measure the variance that can potentially result in much narrow PIs with a degraded PICP; as discussed in [71].

D. BOOTSTRAP METHOD

Bootstrap is a resampling method that ensemble several NNs to construct quality PIs [71]–[75]. It is the most popular among traditional PI construction techniques. There are several types of bootstrap methods. They include smooth, parametric, wild, pairs, residual, Gaussian process, block bootstraps etc. C. Wan et al. investigated NN-based bootstrap PI following pairs, residual, and wild bootstrap methods [10]. The NN-based pairs bootstrap algorithm is implemented through the following steps:

1) Receive training samples.
2) Generate bootstrapped pairs by uniform sampling with replacement from the original training data.
3) Estimate the ELM for a single bootstrap dataset.
4) Repeat steps 2) - 3) to obtain bootstrap replicates.

Residual and wild bootstrap methods differ from the pairs bootstrap method mainly in sampling the residuals. According to their study, pairs bootstrap constructs the most reliable PIs.

Khosravi *et al.* [77] also applied pairs bootstrap based PIs. At the very first stage, this method resampled the original training data into B training datasets [63]. Total B NN models are then developed and the variance associated with the model misspecification, $\sigma_{\hat{y}}^2$ is estimated as presented in Fig. 3. The true regression is then estimated by taking the mean and variation of the point forecasts of all trained NN models through following equations:

$$\hat{y}_i = \frac{1}{B}\sum_{b=1}^{B}\hat{y}_i^b \tag{14}$$

$$\sigma_{\hat{y}_i}^2 = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{y}_i^b - \hat{y}_i\right)^2 \tag{15}$$

where $\hat{y}_i^b$ is the forecast of the i^{th} sample obtained through the b^{th} NN-bootstrap model.

The variation is mostly because of the random initialization of NN-parameters (w) and the effect of different datasets segmented during the data-preparation before training NNs. The variance of errors, $\sigma_{\hat{\epsilon}_i}^2$ needs to be calculated as a part of the PI construction. $\sigma_{\hat{\epsilon}}^2$ can be computed by the following equation:

$$\sigma_{\hat{\epsilon}}^2 \simeq E\{(t - \hat{y})^2\} - \sigma_{\hat{y}}^2 \tag{16}$$

With the help of (16), a set of variance squared residuals are computed:

$$r_i^2 = max\left((t_i - \hat{y}_i)^2 - \sigma_{\hat{y}_i}^2, 0\right) \tag{17}$$

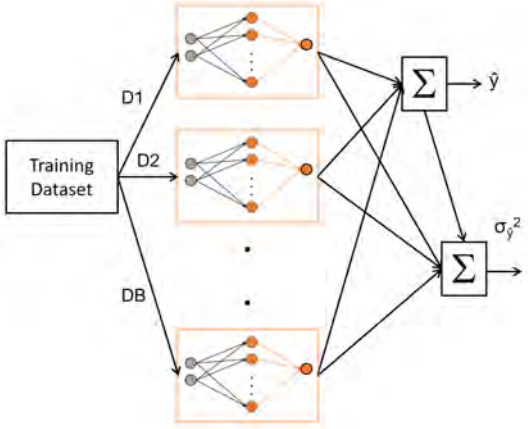

FIGURE 3. The composition of the *B* NN models applied in the bootstrap algorithm [76].

These residuals form a new dataset and linked by the set of corresponding inputs as follows:

$$D_{r^2} = \left\{(x_i, r_i^2)\right\}_{i=1}^{n} \tag{18}$$

An NN model is indirectly trained to estimate uncertainties through the estimation of $\sigma^2_{\hat{\epsilon}_i}$; therefore, through the maximization of the probability of observing samples within the range (in D_{r^2}). The method for the indirect training of this new NN is quite identical to the training process of the above mentioned MVE method. The optimization function for the NN-training is as follows:

$$C_{BS} = \frac{1}{2}\sum_{i=1}^{n}[ln(\sigma^2_{\hat{\epsilon}_i}) + \frac{r_i^2}{\sigma^2_{\hat{\epsilon}_i}}] \tag{19}$$

The NN-based node activation function is selected to be exponential with the enforcement of a positive value of $\sigma^2_{\hat{\epsilon}_i}$. The C_{BS} minimization can be performed through a variety of models, including the traditional gradient descent techniques.

While *B* NN models (assumed to be unbiased) are used for the estimation of $\sigma^2_{\hat{y}_i}$, one model is used for the estimation of $\sigma^2_{\hat{\epsilon}_i}$. Thus, the bootstrap technique is computationally more expensive compared to other techniques in its developing stage ($B + 1$ times). However, the computation time for constructing intervals (PIs) through a trained NN-bootstrap system is only limited to $B + 1$ NN point prediction latency.

The major disadvantage of the trained bootstrap system is the dependency on *B* NN-models. Often a number of these models are biased and resulting an fallacious estimation of $\sigma^2_{\hat{y}_i}$ in (15). That may result in the underestimation of the total variance resulting narrow PIs with a low PICP [6], [78].

Although the bootstrap process is conventionally known as the *bootstrap pairs*, there is another bootstrap process, named *bootstrap residuals*, which resamples the prediction residuals. Readers may read the document [72] for more detailed information.

E. EVALUATION OF TRADITIONAL PIs

Khosravi et al. evaluated and compared the performance of the traditional methods in 2011 [63]. The quality evaluation of the PI is not straightforward. The quality of a point prediction algorithm can be measured using several statistical error values (MAPE, RMSE etc.). The quality of PIs can be measured based on their sharpness (smaller width) and calibration (coverage probability). Perfect PIs have a small width and a high coverage probability.

The PI coverage probability, *PICP* provides the statistical probability of target values limited by the upper and the lower limits of PIs inclusive. The mathematical expression of *PICP* is represented as [35]:

$$PICP = \frac{1}{n}\sum_{j=1}^{n} c_j \tag{20}$$

where,

$$c_j = \begin{cases} 1, & t_j \in [\underline{y}_j, \bar{y}_j] \\ 0, & t_j \notin [\underline{y}_j, \bar{y}_j]. \end{cases}$$

Here, t_j, $\underline{y}_j$ and $\bar{y}_j$ are the actual value, lower bound, and upper bound of j^{th} sample respectively. PICP measures the reliability of constructed PIs [79]–[82].

According to (20), the *PICP* directly depends on the width of PIs. 100% *PICP* can be achieved by extending PIs from either side. However, wide PIs are of a low-quality value as they convey no information about the target. The width of the PI is also needed to be included in the PI assessment/optimization process. The PI average width (PIAW) can be defined as [63]:

$$PIAW = \frac{1}{n}\sum_{j=1}^{n}(\bar{y}_j - \underline{y}_j). \tag{21}$$

When the constructed width for a fair amount of samples is known, the PI normalized average width can be obtained as:

$$PINAW = \frac{1}{R \times n}\sum_{j=1}^{n}(\bar{y}_j - \underline{y}_j). \tag{22}$$

Where, R is the range of the underlying set of targets. *PINAW* provides a scalar value of the average width and it is often reported as a percentage.

Therefore, authors in [63] developed a novel criterion known as the coverage width-based criterion (*CWC*) presented by the following equation:

$$CWC = PINAW\{1 + \gamma(PICP)e^{\eta(\mu - PICP)}\} \tag{23}$$

where, $\gamma(PICP)$ is represented by the following equation.

$$\gamma(PICP) = \begin{cases} 1, & PICP < \mu \\ 0, & PICP \geq \mu \end{cases}$$

Here, $\eta = 50$ and $\mu = 1 - \alpha$ are two hyperparameters.

The value of the quality criterion is equal to the normalized interval width when the target coverage probability is achieved (*i.e.* $CWC = PINAW$ when $PICP \geq 1-\alpha$). Usually, width and coverage of PIs change in opposite direction. The smaller the width, the lower the coverage probability. Therefore, the best PIs have a PICP, equal or slightly higher than $\geq 1-\alpha$ with the smallest or the most optimized PINAW.

Table 1 presents the summary of the comprehensive review [63]. In [63] Khosravi *et al.* observed the best, median and standard deviation of CWCs in twelve case studies. A lower CWC indicates a better PI. Average PICP and average PINAW is also observed for twelve case studies.

TABLE 1. Performance evaluation of traditional methods.

Parameter	Delta	Bayesian	MVE	Bootstrap
Average PICP	91	92.5	94.5	95.1
Average PINAW	45	46	66	58
Average Lowest CWC	40.64	45.80	69.94	58.36
Average Median CWC	87.68	189.5	113.6	102.4
Average STD of CWC	3217	1.1e7	5.5e5	185.04

Delta and Bayesian methods generate narrow PIs with poor coverage compared to MVE and Bootstrap methods. Among MVE and Bootstrap methods, the Bootstrap method is clearly superior as it provides a higher coverage with lower width. However, it is impossible to compare Delta, Bayesian and Bootstrap methods from PICP and NMPIW values. Therefore, *CWC* is calculated for all methods and case studies. Although the Delta and Bayesian techniques have better quality PIs with lower *CWCs* for a few data sets, the standard deviation of PIs is much lower with the bootstrap method. Hence, the bootstrap method provides PIs of good qualities irrespective of the data sets compared to other traditional methods.

F. MODIFIED DELTA METHOD

The delta method described above uses an error-based optimization function, *WDCF* presented in (7) to optimize the NN-structure and parameters. This cost function tries to reduce the prediction error, instead of improving the quality of PI. To improve the quality of PI, Khosravi *et al.* integrated the PICP and PIW to develop a PI-NN model using delta method [35], [59]. In [35], Khosravi et al. integrate PICP and PINAW in the process of NN optimization. The authors tested this modified method for airport baggage handling systems with many internally connected homogeneous and heterogeneous components [35]. In contrast to *WDCF*, a new cost function was proposed in [59] covering both PICP and PINAW. The new cost function called PI-error-based cost function (PICF) is defined as:

$$PICF = CLC + e^{(PICF_{opt} - WDCF_{trad})}. \tag{24}$$

where $WDCF_{trad}$ obtained through minimization of (7). *CLC* is the coverage-length-based criterion defined as:

$$CLC = \frac{PINAW}{\sigma(PICP, \eta, \mu)}. \tag{25}$$

where $\sigma(\cdot)$ is the sigmoid function defined as:

$$\sigma(PICP, \eta, \mu) = \frac{1}{1 + e^{-\eta(PICP-\mu)}}. \tag{26}$$

where, η and μ are two regulating parameters defining where and how sharply the sigma function rises. As (24) includes *WDCF*, this method optimise in two stages. Firstly, NNs are trained using Levenberg-Marquardt training algorithm by minimizing (7). Then NNs are retrained using a different set of data through minimizing (24). The authors used simulated annealing optimization technique for the second stage. The details procedure for *PICF* development and modified delta technique can be found in [59].

G. MODIFIED MVEM METHOD

Khosravi and Nahavandi [70] extended the MVEM for the interval prediction of the wind power forecasts. In contrast to traditional MLE, they used a PI-based optimization function that developed in [83] to optimized the NN_σ. This cost function includes two key components of PIs- PINAW and PICP that are defined in (22) and (20). The inclusion of these PI indexes in NN_σ parameters optimization process, significantly improved the overall performance of the MVEM to construct quality PIs.

H. MODIFIED BOOTSTRAP METHOD

A new method to improve the quality of bootstrap PIs was introduced in [76]. Although, there is always a tradeoff between width and coverage probability of PIs, the designer of the prediction algorithm needs to ensure a certain coverage probability. Therefore, constructing a narrower PI with the same coverage probability requires a better training NN model.

A new optimization criterion is obtained empirically for the optimal training of NNs; known as the coverage width-based criterion (*CWC*):

$$\min_{NN} \; CWC = PINAW + \gamma(PICP)e^{\eta(\mu - PICP)} \tag{27}$$

where, $\gamma(PICP)$ has the same expression as equation (23), η and μ are two hyperparameters determining the position and the magnitude of the CWC jump.

Although the modified bootstrap interval method is computationally expensive during the training due to three extra steps, the execution time is the same. The modified bootstrap NN PI training steps are as follows:

- PI construction using the traditional bootstrap method.
- Initialization with the traditional bootstrap-optimized parameters.
- Optimization through the *CWC* minimization.
- Evaluation of the NN model.

Several optimization algorithms such as particle swarm optimization and the genetic algorithm can be applied for the CWC minimization and optimal training of the NN. It is shown that the quality of PIs is improved by 28% on average in 70 experiments. Fig. 4 shows the traditional optimal PIs

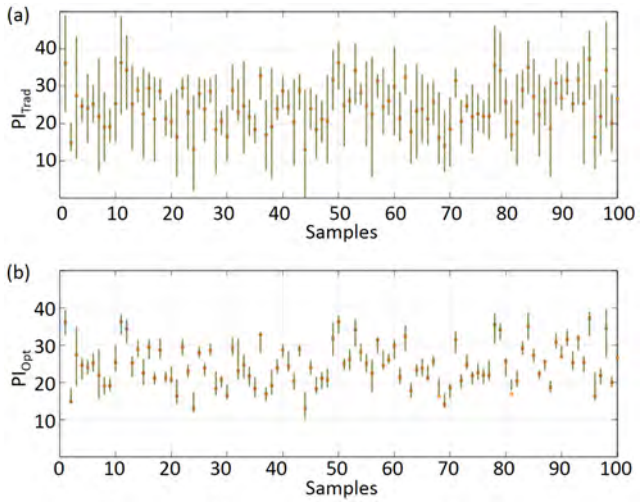

FIGURE 4. Performance of the modified bootstrap method on a typical case study. Optimized PI with narrower widths, reproduced with the permission of authors [76].

for a typical case study. The confidence level of both PIs are greater than the nominal confidence level of optimal PIs are significantly narrower than the traditional PI. Therefore, these PIs contain more meaningful information about the uncertainty of targets.

III. NN-BASED DIRECT PI CONSTRUCTION METHODS

Although PIs have been applied for uncertainty quantification from the nineteenth century, recent advancements of NN open new opportunities for generating better quality PIs. NN is a data-based technique and does not require the detailed knowledge of the system for mapping the relationship between their inputs and outputs. Most cases, NNs outperform their traditional rivals for a wide range of applications [84]. Therefore, numerous studies on NN applications have been reported in the literature, including PI-based forecasts. Researchers have invented NN based novel PI construction techniques, those can construct PI through single NN. Moreover, their training does not require any theoretical intermediate variable.

A. LOWER UPPER BOUND ESTIMATION (LUBE) METHOD

Although the modified bootstrap method constructs the best quality PIs among the modified traditional methods [76], the computational overhead of that method is quite high. In that method, a number of NNs are trained with different initialization values for achieving the probability distribution of target. A smoother probability distribution requires a larger number of NNs result in a greater computational overhead. Moreover, NN-structure and initialization values need to change to achieve the optimum NN-structure. Also, the Bootstrap PIs cannot produce a good PICP and PINAW tradeoff for any distribution of targets due to the assumption of a Gaussian probability distribution. The final NN of the Bootstrap method computes the variation among initial NNs and returns a PI considering a Gaussian probability distribution. However, that assumption is not true in all situations.

When NNs are constructed through a direct training process without any assumption on the distribution, they can provide a smarter PI for any distribution of targets [85]. Fig. 5 explains the situation with a rough sketch. Point prediction and the level of uncertainty change over time. However, the distribution of probability is log normal. Ordinary NNs consider the Gaussian distribution and construct PIs presented by black lines. The LUBE NN is optimized for any arbitrary distribution of targets and provides a smarter PI, presented by green lines. Therefore, an improved algorithm is developed that directly calculates the lower and the upper bounds through trained NNs. CWC is used as the optimization function for the NN optimization. Fig. 6 presents the value of CWC over iterations. Fig. 7 presents the structure of the LUBE NN. The NN-based LUBE interval is trained with the cost function presented in equation (28).

$$\min_{NN} \; CWC = PINAW\{1 + \gamma(PICP)e^{\eta(\mu - PICP)}\} \qquad (28)$$

The LUBE NN training steps are as follows [83]:

- The available data is split randomly into the training, cross-validation and test sets at first.
- A NN (as shown in Fig. 7) is constructed and initialized with random values. The value of the cost function (equation 28) is calculated with the initial values as CWC_{Opt}. The NN parameters are also assumed as the optimum one (W_{Opt}).
- Cooling temperature is updated as the first step of the simulated annealing based training loop.
- A new set of NN weights (W_{new}) is generated through the random perturbation of the current NN.
- PIs are constructed with the new NN parameters (W_{new}).
- The cost function is updated along with the NN parameters when required.

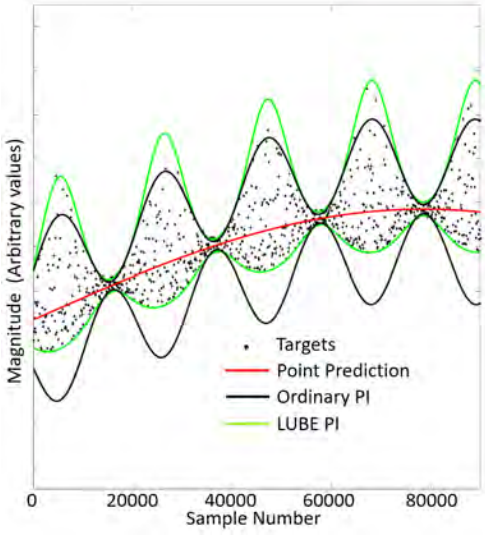

FIGURE 5. A rough sketch presenting the advantage of LUBE method. The probability distribution of target can be a non-gaussian one. The direct construction of PI through the NN training can consider an unknown probability distribution.

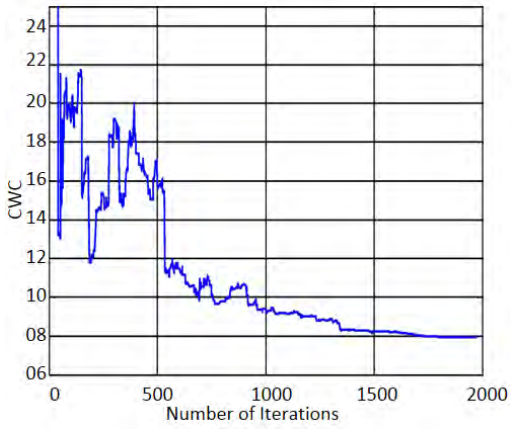

FIGURE 6. Improvement of the NN through the minimization of *CWC* values through random initialization and iterations [86].

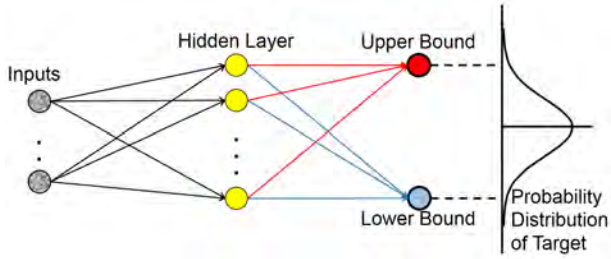

FIGURE 7. NN-structure for the LUBE Method, reproduced with the permission of authors [83].

$$CWC_{Opt} = min\{CWC_{New}, CWC_{Opt}\}$$
$$W_{Opt} = W(CWC_{Opt})$$

- The training loop ends when the maximum number of iteration is reached, or no improvement is achieved from recent iterations, or the temperature becomes lower than the threshold.
- The trained NN is tested on the test samples.

The *CWC* values at different iterations are recorded and reported by Hosen *et al.* [86] while deriving PIs for the control of polystyrene polymerization reactor; shown in Fig. 6. The LUBE method can also be optimized using other optimization algorithms such as the genetic algorithm [87] and particle swarm optimization [88].

Table 2 presents execution times and CWC values. Among the traditional methods, Bootstrap has the lowest execution time. Computation time is further reduced by more than one order with the LUBE method compared to the Bootstrap method. The *CWC* is also improved on average with the LUBE method. Moreover, all of the *CWC* values of the LUBE method is lower than 100. However, the initial version of the LUBE cost-function had several limitations. Therefore, several new cost functions are proposed by several groups.

B. DIRECT INTERVAL FORECASTING BY CAN WAN

Can Wan et al. also proposed an NN based direct interval forecasting method with a slightly different cost function [89]–[91]. Their cost function also contains coverage dependent and width dependent components. The coverage dependent component is known as, the average coverage error (ACE) is defined by the following equation:

$$ACE = PICP - PINC \tag{29}$$

Where $PINC = 1 - \alpha$ is the PI nominal coverage. Width of the PI for j^{th} sample is defined as:

$$PIW_j = \bar{y}_j - \underline{y}_j \tag{30}$$

Can Wan et al. defined an interval score for the development of cost function and for accessing the quality of PI. The interval score for j^{th} sample and the average interval score is defined by following equations:

$$S_j = \begin{cases} -2\alpha \times PIW_j - 4(\underline{y}_j - t_i), & t_i < \underline{y}_j \\ -2\alpha \times PIW_j & \\ -2\alpha \times PIW_j - 4(t_i - \bar{y}_j), & t_i > \bar{y}_j \end{cases} \tag{31}$$

$$S_{AV} = \frac{1}{n}\sum_{j=1}^{n} S_j \tag{32}$$

Their proposed cost function is as follows:

$$\min_{NN} \ \gamma|ACE| + \lambda|S_{AV}| \tag{33}$$

Where γ and λ are two fitting parameters, defined as importance weights. Both of them are assigned to one to provide equal importance towards the reliability and the sharpness.

The algorithm developed for the direct interval forecasting of wind power has all benefits of the LUBE method. As the intervals are calculated using a single NN, the execution time becomes much shorter compared to traditional methods. Moreover, the NN is optimized for any arbitrary distribution of target. Therefore, the PIs become narrower with the maintenance of equal PICP.

TABLE 2. Improvement through the LUBE method.

Attribute	Case Study	LUBE	Bootstrap	Delta	Bayesian
Execution Time	1	3ms	51ms	3562ms	4195ms
	2	3ms	54ms	3159ms	4761ms
	3	3ms	51ms	3181ms	3508ms
	4	3ms	55ms	3199ms	3498ms
	5	3ms	56ms	5924ms	7818ms
	6	4ms	56ms	3886ms	5826ms
	7	4ms	51ms	3892ms	5813ms
	8	3ms	51ms	2500ms	3002ms
	9	4ms	50ms	2029ms	3265ms
	10	3ms	51ms	2368ms	2779ms
CWC	1	90	>200	120	>200
	2	25	30	15	70
	3	45	>200	70	55
	4	35	65	60	70
	5	35	55	35	35
	6	70	>200	75	190
	7	75	>200	60	65
	8	45	90	>200	50
	9	30	35	>200	>200
	10	45	50	>200	>200

However, the direct construction proposed by Wan *et al.* [89] penalizes both high PICP and low PICP with equal weight and the most optimized situation is $PICP = PINC$. Therefore, the NN often converges to a slightly low PICP (PICP < PINC). Moreover, the width is not normalized with the range. Therefore, the optimization can potentially vary when the signal amplitude is changed.

C. NORMALIZED ROOT-MEAN-SQUARE WIDTH (PINRW) AND PARTICLE SWARM OPTIMIZATION (PSO) BY H QUAN

Quan et al. proposed an improved LUBE method through the particle swarm optimization (PSO) and cost function improvement [88]. The cost function of the LUBE method contains PINAW. Quan et al. applied the prediction interval normalized root-mean-square width (PINRW) instead of PINAW. The PINRW is defined by the following equation:

$$PINRW = \frac{1}{R}\sqrt{\frac{1}{n}\sum_{j=1}^{n}(\bar{y}_j - \underline{y}_j)^2} \tag{34}$$

Therefore, the improved cost function becomes as follows:

$$\min_{NN} \; CWC = PINRW\{1 + \gamma(PICP)e^{\eta(\mu - PICP)}\} \tag{35}$$

Their work demonstrates six case studies, consisted of four real-world and two synthetic ones, are investigated to validate the PSO-based improved LUBE PI construction method. A five-fold cross-validation technique is applied to obtain the optimal structure of NNs. Both of the computation complexity reduction and PI quality enhancement are achieved [88].

D. INDEPENDENT WIDTH AND PENALTY FACTORS BY A KHOSRAVI

The penalty factor ($\gamma(PICP)$) was in a multiplicative manner with the width factor (PINAW), as shown in (23). During 2013 and early 2014 several researchers doubted the effectiveness of the CWC based optimization [92], [93]. One major issue was the optimization at zero width (PINAW = 0). Because the minimum or the most optimized value of CWC is zero. During mid-2014, Khosravi et. al. resolved most of those issues by modifying the CWC equation [94]; as presented at (27). According to that closure discussion, CWC provides a strictly proper score. However, the CWC based optimization can produce too wide PIs in some samples in rare situations and the optimization fails once in twelve cases on average. While the NN is trained three times with different random initialization, at least one NN usually provide quality PIs. That improved CWC cost function is applied in various studies in later years [76].

E. DEVIATION FROM MID-INTERVAL CONSIDERATION BY G. MARN

L G. Marn et al. presented type-2 fuzzy system based wind power generation and load forecasting systems [95]. Their cost function is as follows:

$$\min_{NN} \; \beta_1 PINAW + \beta_2 ||e||^2 + exp[-\eta\big(PICP - (1-\alpha)\big)] \tag{36}$$

where β_1 and β_2 are weighting factors. They also perform normalization of PINAW and $||e||^2$. The parameter $||e||$ is the error quantity defined by the following equation:

$$||e|| = \sqrt{\sum_{j=1}^{n}\left|t_j - \frac{\bar{y}_j + \underline{y}_j}{2}\right|^2} \tag{37}$$

where t_j is the value of j^{th} sample. However, that cost function penalizes both higher and lower PICP and the adjustment of β values are highly dependent on the nature of data. Moreover, $||e||$ is not normalized with the number of samples and therefore, β values need to be tuned for different sample numbers.

F. IMPROVEMENT THROUGH OPTIMAL AGGREGATION BY A HOSEN

Constructed PIs quality significantly differs from one data-set to data-set because of the instability of optimized models. Therefore, Hosen et al. proposed an optimal system to develop PI-based NN to enhance the performance of LUBE PIs; through maintaining a good PICP with a reasonable width. In order to combine PIs, the weighted average aggregation technique is employed, wherein two optimization techniques, known as, Genetic Algorithm (GA) and Simulated Annealing (SA), are applied to optimally adjusting weights by minimizing a PI-based optimization function, CWC. The optimized CWC values are reduced by 22%, 18%, and 78% respectively for the first, the second, and the third case studies. That technique also enhances the PI performance by 3%-4% compared to the simple averaging aggregation method in all tested cases [96].

G. DEVIATION INFORMATION-BASED CRITERION BY G ZHANG

G Zhang et al tried to avoid an exponential cost function. Therefore, the derived one cost function named deviation information-based criterion (DIC) [97], shown as follows:

$$DIC = PINAW + \gamma(PICP) \cdot pun \tag{38}$$

provided that,

$$pun = \sigma \sum_{j=1}^{N_L}(\underline{y}_j - t_j) + \sigma \sum_{j=1}^{N_U}(t_j - \bar{y}_j)$$

where σ is the penalty factor, N_L and N_U are the numbers of times, the target becomes lower than the lower bound and the number of times the target becomes higher than the upper bound respectively. σ is set to $1/\alpha$.The parameter *pun* presents the deviation information. However, the value of *pun* may become very large for a large dataset, as authors did not normalize it with the number of samples. Moreover, the cost function is discontinuous at ($PICP = 1 - \alpha$).

IV. APPLICATIONS OF PIs

Despite the practical importance, PI-based forecasting techniques have received little attention by the investors or companies [31], [34]. Major reasons for their unpopularity are as follows:

1) There is no widely accepted method of calculating PIs.
2) Theoretical PIs are not easy to evaluate for many economic models.
3) Methods of construction of empirically-obtained PIs are not widely understood.
4) Although a number of procedures for calculating PIs have been proposed, their verification and comparison process is unknown to many users.
5) The recent NN based PI construction method is still debatable due to the initialization dependent training and effectiveness issues.

Besides those limitations in understanding and difficulties in the training process, a well-trained NN provides state of the art performance in constructing PIs. Many researchers have structured and trained efficient NNs for different applications.

A. FINANCIAL INSTITUTIONS

The interval forecast with a certain coverage is widely used in financial institutions [98], [99]. For example, currency exchange rate changes dynamically over time and monetary organizations need to reserve a certain amount of foreign currency for their customers demand [100], [101]. The exchange rate may vary by several percentages over a day. Therefore, they need to exchange currencies by a large amount when the exchange rate seems profitable. Therefore, monetary organizations need to choose a suitable time for the exchange. As the exchange rate changes quite frequently and in a high order, several monetary organizations are using the interval forecast [4], [30]. Financial institutions are keeping a certain amount of currency based on widths of PIs to maintain a good quality of service. Financial institutions need to predict both of the demand and price of foreign currencies. Similarly, PIs are used in price prediction of different assets [102].

B. FOOD INDUSTRY

The food industry needs to produce, deliver, and preserve foods based on the variable demand, storage cost, and quality degradation over time. The demand also varies over time and price. Such as, the demand may increase with the increase of price for prestigious commodities. However, the demand decreases with the increase in price for basic goods those have alternatives. While any large industry fails to meet the food requirement the price increases, other firms grow, consumers switch to other foods and that affects the future selling. Therefore, the industry needs to meet the demand. Also, overproduction or preservation can potentially cause the waste of food and money. An interval forecast presenting the upper and the lower bounds is effective in production and storage capacity planning. Thus, the prediction interval is widely used in the prediction of the food consumption [5] and the quality degradation [103].

C. LOAD FORECASTING

The level of uncertainty of the electricity consumption has increased due to the installation of new loads in many parts of the world. Moreover, the climate change has added additional uncertainties and it is almost impossible to develop analytical equations to predict the power consumption. As the neural network (NN) can be trained to predict without knowing the detailed physical understanding of the phenomenon of a multivariable system [104] and the point forecast is not sufficient for predicting such uncertain system [105], the NN based prediction interval has recently brought the attention of the researchers. For example, Hao et al. develop a prediction interval-based load forecasting technique and verified the technique with the historical electrical market datasets from Texas (TX) (USA), Ottawa (OTT) (Canada) and Singapore (SG). One-week leading demand forecasting is performed utilizing NN-based PIs and PIs are found to be a superior tool to quantify uncertainties of the power systems and compared with the point forecasts [106].

The number of neurons in a neural network is highly dependent on the level of uncertainty. Therefore, the size of the optimum neural network varies from application to application. While the level of uncertainty is higher more neurons are required. An NN-based prediction system with inadequate neurons results in a significantly high error due to the underfitting. Also, the excess of neuron results in overfitting; increases computational complexity and invited high order polynomials. High order polynomials result in a good match with the training data-sets and a high error in few samples of the verification and application datasets. It is currently impossible to predict the optimum number of neurons required for the optimal performance with any signal without trial and error method.

Recently Quan et. al. presents the NN-optimization process for the load curves of the TX, USA [106]. The optimum NN-size is selected based on the median PINAWs. Their NN-structure consists of two hidden layers (1 & 2) and the number of neurons in those hidden layers are $n1$ and $n2$. Fig. 8 presents the median PINAWs on verification data for different hidden layer sizes. According to their experiment, $n1 = n2 = 7$ provides the optimized output on the verification data.

Fig. 9 presents the weekly load profile and constructed PIs of Singapore during March 2127, 2011 [9]. That figure also expresses how PIs encloses the target signal in 95% or $(1-\alpha)$ cases. The upper bound is always higher than the lower bound. The interval between the upper and the lower bounds is the prediction interval and it changes along with its width from sample to sample. In less than 5% or α cases, the target value can be higher than the upper limit or lower than the lower limit. While drawing a PI curve like Fig. 9, the NN is first trained with the training dataset. Then the prediction interval of each verification sample is calculated through the

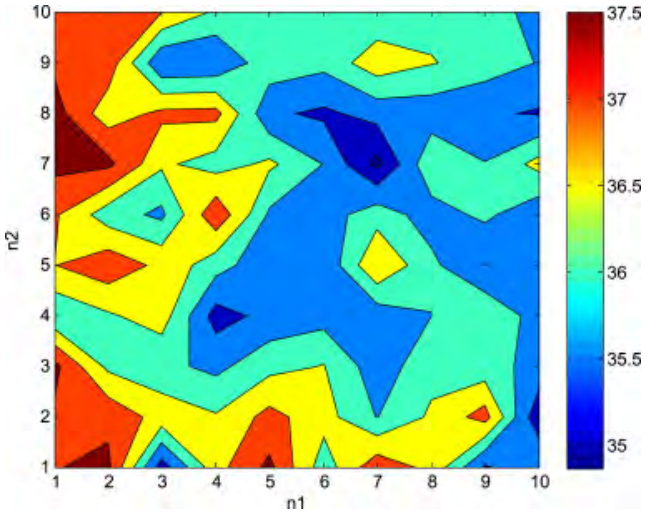

FIGURE 8. Finding the optimal NN neuron size for the TX load curves while applying LUBE method; reproduced with the consent of authors [106].

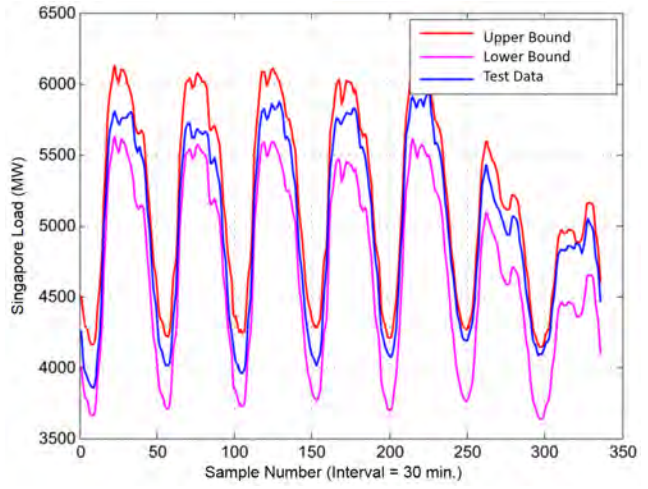

FIGURE 9. Weekly load profile and constructed PIs of Singapore during March 21-27, 2011; reproduced with the consent of authors [9].

trained NN and recent previous samples. Finally, the lower bound, the upper bound and the target signal is drawn in the same plot [107].

D. FORECASTING OF THE RENEWABLES

One of the popular application of the uncertainty quantification is the power grid management. The degree of uncertainty in the power systems has increased due to the increased installation of renewable power generation sources in recent decades [108]. With that increased uncertainty, the statistical error value of the point prediction becomes larger [29], [109]–[111]. Moreover many countries are planning to switch on the renewable sources, such as the wind and the solar energy. Maintaining a stable grid with a large amount of renewable generation is challenging and that opens a broader scope of research and many researchers are customizing algorithms for wind and solar generation [107], [112]–[115].

The order of uncertainty is much higher in the renewable generations compared to the uncertainties in the power consumption. A certain flow of wind is required for the power generation at the wind turbine. While the flow of wind drops, the wind firm stops generating electricity. Similarly, the electricity generated from the solar panels are highly dependent on the position of clouds. Moreover, the fuel-based generator can manage certain fluctuation and utility companies have certain backup generation and storage units. When the renewable generation becomes more uncertain, utility companies need to take more precautions. Therefore, NN based prediction intervals are a better option compared to the point prediction [116]–[121]

E. ELECTRICITY PRICE FORECASTING

Based on the renewable generation and the consumption, fossil fuel based generators need to provide the electricity. A constant electricity price brings a large difference between the peak hour and the off-peak hour consumption. Thus, the price of the electricity is changed over time to insist the users to use more electricity during the off-peak hours [122]–[125]. Especially many appliances, such as washing machine, water pumps, dryers etc. and charging appliances can be used during the off-peak hour. Although several appliances have noise issues while using at night, those appliances can be used during the semi-off-peak hours, such as early morning. Moreover, almost all industries need pumps, freezers and the requirement of production vary over the year. When the production requirement is low, the machines are operated during off-peak hours only. Therefore, industries need the price prediction to reduce their electricity consumption cost and the NN based PIs are a good choice for the price prediction due to their optimized performance in terms of width and accuracy [126]. For instance, Khosravi *et al.* [127] employ the bootstrap and delta techniques for the calculation of intervals for the uncertainty prediction. The PIs confidence level is varied from 50% to 90% for investigating the effect. According to their experiment with the three-month Australian electricity price data, constructed prediction intervals efficiently quantify the uncertainties compared to the NN-based point forecasting [128].

F. REACTOR CONTROLLER DESIGN

Many reaction processes are extremely complex in nature. Various chemical reaction controllers need accurate models for proper control [129]. The designer needs to learn the underlying mechanism to derive an analytical model. However, the NN-based model formation does not require the underlying knowledge about the mechanism. Also, many reaction controllers exhibit a significant amount of uncertainty and the point prediction is followed by a large error value (RMSE, MAPE etc.). Such as the point prediction of the polystyrene polymerization reactor faces 22% mean

absolute percentage error (MAPE) on average and the prediction interval is developed for the control of that reactor. Therefore the NN based PI is applied to controller design [86].

G. MEDICAL APPLICATIONS

PIs are widely used in predicting health conditions and detecting diseases. A Mayr et al. applied PIs for the analysis of body mass index (BMI) data [130]. H Nishiura applied PIs for early detection of Nosocomial Outbreaks [131]. Interval forecasts are also applied for the prediction of stroke risk and retinal degradation through the retinal image analysis [132], [133]. These analyses are recently applying NN-based interval formation and decision making [134].

H. OTHER APPLICATIONS

The NN based PIs are extensively designed and implemented in the above-mentioned topics. There are other myriad applications of the PIs. For example, Kim *et al.*[6] predicted the tourist arrival in Australia and Hong Kong as the interval forecast. Also, PIs are effective in the prediction of dynamic pricing. In fact, both of the demand and price is required to predict. The service provider or the seller needs to know the demand and customers need for the proper management and price setting. Usually, a user lacks information about the other users and he needs to take the decision based on the time series analysis. Gong et al. calculated 1-minute ahead prediction interval for CPU usage of the Google cluster through the time series analysis [11]. Lu and Viljanen [61] obtained PI for predicting the indoor temperature and the relative humidity. Taormina and Chau [135] proposed PIs for streamflow prediction.

V. FUTURE OF PI

A. NEW DIRECT METHODS

Reasons for the unpopularity of PIs arise from the methods to construct the PIs. This is because traditional PI construction methods rely on the strong assumptions (*e.g.* distribution of the data) and the models are itself linear [136] and the absence of robust cost function for direct PI construction. It is impossible to conclude which method is suitable for a specific system, especially for a complex system [137]. Researchers may find a cost function for obtaining a better quality PI considering these issues in near future. Moreover, researchers may derive customized cost functions for different applications.

B. NEW APPLICATIONS

With the increased popularity of PI over time, researchers may apply PIs to new fields of studies. Besides that, researchers need to think differently for PIs than the point prediction. For example, the difference between energy demand and the renewable generation is important for the scheduling of fossil-fuel based power plants. In point prediction, their values can be subtracted to gain the required fossil-fuel generation. However, demand PIs and renewable PIs of 95% PICP generate a much wider PI with higher PICP. The width and the PICP depend on relative magnitudes of signals and level of uncertainties in each signal. Therefore, the user needs to construct PIs directly for the required fossil-fuel generation.

C. NOVEL TRAINING ALGORITHMS

In near future, researchers may apply new learning algorithms for training NNs for generating optimal PIs. Novel algorithms can be deep learning, deep bayesian learning, dropouts [138]–[140]. These techniques have already achieved state of the art performance in many NN based systems in critical situations. We also observed that the PI coverage probability is low for rare and critical situations [141]. In future, researchers may apply those revolutionized training mechanisms for constructing better PIs.

D. STANDARDISE DATASETS

The most challenging part of the performance evaluation of any novel PI construction training is the lack of standardized data. Different NN may work well for different datasets. Even the same cost function may construct different NNs for different initializations and the size of NN. Standardised datasets are available in various fields of recognition and classification. In future, the PI research community may decide a standardize dataset and range of NN-size for fair comparisons.

Besides these theoretical qualities and optimization issues, the directly constructed NNs are constructing PIs for any arbitrary distributions. As a result, PIs are becoming smarter. However, users do not have any information to guess the most probable region of the target. Traditional PIs are constructed using mean and variance estimations. Therefore, the most probable region of target stays in the middle of the interval. When a smart interval is formed for a logarithmic distribution, the most probable region of the target stays near one bound of the interval and the user remains unaware of it. In future, researchers may design systems for representing the uncertainty by constructing at least two PIs of different α or constructing both PIs and point predictions [142].

VI. CONCLUSION

The prediction is one of the oldest tasks in this world. The point prediction is the most widely used and understandable form of prediction. However, the point prediction does not convey any information about the uncertainty. Therefore, probabilistic forecasting is becoming popular to predict uncertainties in emerging engineering problems with high uncertainty. The NN based approaches for the quantification of uncertainty is relatively new. Without a thorough discussion, it is impossible to know advantages and disadvantages of different NN based PI construction techniques. The paper provides detailed research activities on NN-based PI construction concerning motives of each work. Our current survey may help analysts in knowing recently available popular

NN based PI construction techniques. They may select an appropriate PI construction technique for their application. The work may also help future searchers in developing novel algorithms for constructing NN-based PIs and investigating new applications.

REFERENCES

[1] G. C. Chow, "Tests of equality between sets of coefficients in two linear regressions," *Econometrica, J. Econ. Soc.*, vol. 28, no. 3, pp. 591–605, 1960.

[2] R. Henderson, M. Jones, and J. Stare, "Accuracy of point predictions in survival analysis," *Statist. Med.*, vol. 20, no. 20, pp. 3083–3096, 2001.

[3] S. Maskey, V. Guinot, and R. K. Price, "Treatment of precipitation uncertainty in rainfall-runoff modelling: A fuzzy set approach," *Adv. Water Resour.*, vol. 27, no. 9, pp. 889–898, 2004.

[4] J. Longerstaey and M. Spencer, *RiskMetricstm—Technical Document*. New York, NY, USA: Morgan Guaranty Trust Company, 1996.

[5] D. M. Heien and C. R. Wessells, "The demand for dairy products: Structure, prediction, and decomposition," *Amer. J. Agricult. Econ.*, vol. 70, no. 2, pp. 219–228, 1988.

[6] J. H. Kim, K. Wong, G. Athanasopoulos, and S. Liu, "Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals," *Int. J. Forecasting*, vol. 27, no. 3, pp. 887–901, 2011.

[7] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Feb. 2001.

[8] S. D. Stearns, "Arithmetic coding in lossless waveform compression," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1874–1879, Aug. 1995.

[9] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 303–315, Feb. 2014.

[10] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.

[11] Z. Gong, X. Gu, and J. Wilkes, "PRESS: Predictive elastic resource scaling for cloud systems," in *Proc. Int. Conf. Netw. Service Manage. (CNSM)*, 2010, pp. 9–16.

[12] A. S. Sabyasachi, H. M. D. Kabir, A. M. Abdelmoniem, and S. K. Mondal, "A resilient auction framework for deadline-aware jobs in cloud spot market," in *Proc. IEEE 36th Symp. Reliable Distrib. Syst.*, Sep. 2017, pp. 1–3.

[13] S.-M. Lee and H. Myung, "Cooperative particle swarm optimization-based predictive controller for multi-robot formation," *J. Inst. Control, Robot. Syst.*, vol. 19, no. 5, pp. 429–434, 2013. [Online]. Available: http://www.koreascience.or.kr/article/ArticleFullRecord.jsp?cn=JOJDCV_2013_v19n5_429, doi: 10.5302/J.ICROS.2013.13.9007.

[14] F. S. Mjalli, S. Al-Asheh, and H. Alfadala, "Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance," *J. Environ. Manage.*, vol. 83, no. 3, pp. 329–338, 2007.

[15] K. Shahbaz, F. Mjalli, M. A. Hashim, and I. M. AlNashef, "Prediction of deep eutectic solvents densities at different temperatures," *Thermochim. Acta*, vol. 515, nos. 1–2, pp. 67–72, 2011.

[16] B. Nandram and J. Yin, "A nonparametric Bayesian prediction interval for a finite population mean," *J. Stat. Comput. Simul.*, vol. 86, no. 16, pp. 3141–3157, 2016.

[17] M. A. Hosen, "Prediction interval-based modelling and control of nonlinear processes," Ph.D. dissertation, Centre Intell. Syst. Res., Deakin Univ., Victoria, Australia, 2015.

[18] M. Alam, J. P. Parmigiani, and J. J. Kruzic, "An experimental assessment of methods to predict crack deflection at an interface," *Eng. Fracture Mech.*, vol. 181, pp. 116–129, Aug. 2017.

[19] M. Rana, I. Koprinska, and V. G. Agelidis, "Forecasting solar power generated by grid connected pv systems using ensembles of neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2015, pp. 1–8.

[20] H. M. D. Kabir, Z. Ahmed, R. Kariyadan, L. Zhang, and M. Chan, "Modeling of fringe current for semiconductor-extended organic TFTs," in *Proc. IEEE Int. Conf. Electron Devices Solid-State Circuits (EDSSC)*, Aug. 2016, pp. 177–180.

[21] M. Alam, B. Grimm, and J. P. Parmigiani, "Effect of incident angle on crack propagation at interfaces," *Eng. Fract. Mech.*, vol. 162, pp. 155–163, Aug. 2016.

[22] M. Rana, I. Koprinska, and A. Troncoso, "Forecasting hourly electricity load profile using neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2014, pp. 824–831.

[23] H. M. D. Kabir, M. A. Hosen, S. Nahavandi, and A. Khosravi, "Prediction interval with examples of similar pattern and prediction strength," in *Proc. IEEE 30th Can. Conf. Elect. Comput. Eng. (CCECE)*, Apr./May 2017, pp. 1–4.

[24] M. P. Clements, "Evaluating the bank of England density forecasts of inflation," *Econ. J.*, vol. 114, no. 498, pp. 844–866, 2004.

[25] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004.

[26] A. U. Haque, M. H. Nehrir, and P. Mandal, "A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1663–1672, Jul. 2014.

[27] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 763–770, Mar. 2016.

[28] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2767–2778, Jul. 2017.

[29] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845–1856, Nov. 2010.

[30] P. F. Christoffersen, "Evaluating interval forecasts," *Int. Econ. Rev.*, vol. 39, no. 4, pp. 841–862, 1998.

[31] J. K. Patel, "Prediction intervals—A review," *Commun. Statist.-Theory Methods*, vol. 18, no. 7, pp. 2293–2465, 1989.

[32] M. Valipour, "Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms," *Meteorological Appl.*, vol. 23, no. 1, pp. 91–100, 2016.

[33] S. J. Mason and G. M. Mimmack, "Comparison of some statistical methods of probabilistic forecasting of ENSO," *J. Climate*, vol. 15, no. 1, pp. 8–29, 2002.

[34] C. Chatfield, "Calculating interval forecasts," *J. Bus. Econ. Statist.*, vol. 11, no. 2, pp. 121–135, 1993.

[35] A. Khosravi, S. Nahavandi, and D. Creighton, "A prediction interval-based approach to determine optimal structures of neural network metamodels," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2377–2387, 2010.

[36] D. L. Shrestha and D. P. Solomatine, "Machine learning approaches for estimation of prediction interval for the model output," *Neural Netw.*, vol. 19, no. 2, pp. 225–235, 2006.

[37] F. Proschan, "Confidence and tolerance intervals for the normal distribution," *J. Amer. Stat. Assoc.*, vol. 48, no. 263, pp. 550–564, 1953.

[38] P. Vit, "Interval prediction for a poisson process," *Biometrika*, vol. 60, no. 3, pp. 667–668, 1973.

[39] D. A. Fraser and I. Guttman, "Tolerance regions," *Ann. Math. Statist.*, vol. 27, no. 1, pp. 162–179, 1956.

[40] G. J. Hahn, "Factors for calculating two-sided prediction intervals for samples from a normal distribution," *J. Amer. Stat. Assoc.*, vol. 64, no. 327, pp. 878–888, 1969.

[41] M. A. Shayib, A. M. Awad, and A. M. Dawargeh, "Large sample prediction intervals for a future sample mean: A comparative study," *J. Stat. Comput. Simul.*, vol. 24, nos. 3–4, pp. 255–270, 1986.

[42] G. D. Faulkenberry, "A method of obtaining prediction intervals," *J. Amer. Stat. Assoc.*, vol. 68, no. 342, pp. 433–435, 1973.

[43] J. Robinson, "Sequential choice of an optimal dose: A prediction intervals approach," *Biometrika*, vol. 65, no. 1, pp. 75–78, 1978.

[44] H. A. Sartawi and M. S. Abu-Salih, "Bayesian prediction bounds for the burr type x model," *Commun. Statist.-Theory Methods*, vol. 20, no. 7, pp. 2307–2330, 1991.

[45] J. Felsenstein, "Confidence limits on phylogenies: An approach using the bootstrap," *Evolution*, vol. 39, no. 4, pp. 783–791, 1985.

[46] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, no. 1, pp. 1–26, 1979.

[47] R. A. Stine, "Bootstrap prediction intervals for regression," *J. Amer. Stat. Assoc.*, vol. 80, no. 392, pp. 1026–1031, 1985.

[48] D. L. Harnett and J. L. Murphy, *Introductory Statistical Analysis*. Boston, MA, USA: Addison-Wesley, Jul. 1975. [Online]. Available: https://www.amazon.com/Introductory-Statistical-Analysis-Donald-Harnett/dp/0201027496 and http://www.worldcat.org/title/introductory-statistical-analysis/oclc/5798763?page=citation
[49] R. J. Wonnacott, "Trade and investment in a hub-and-spoke system versus a free trade area," *World Economy*, vol. 19, no. 3, pp. 237–252, 1996.
[50] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, Apr. 2014.
[51] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
[52] M. Bellemare, S. Srinivasan, G. Ostrovski, T. Schaul, D. Saxton, and R. Munos, "Unifying count-based exploration and intrinsic motivation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1471–1479.
[53] P. Auer, "Using confidence bounds for exploitation-exploration tradeoffs," *J. Mach. Learn. Res.*, vol. 3, pp. 397–422, Nov. 2003.
[54] J. T. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," *J. Amer. Statist. Assoc.*, vol. 92, no. 438, pp. 748–757, Jun. 1997.
[55] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 1, Jul. 1994, pp. 55–60.
[56] G. Seber and C. Wild, *Nonlinear Regression*. New York, NY, USA: Wiley, 1989.
[57] A. A. Ding and X. He, "Backpropagation of pseudo-errors: Neural networks that are adaptive to heterogeneous noise," *IEEE Trans. Neural Netw.*, vol. 14, no. 2, pp. 253–262, Mar. 2003.
[58] R. D. de Veaux, J. Schumi, J. Schweinsberg, and L. H. Ungar, "Prediction intervals for neural networks via nonlinear regression," *Technometrics*, vol. 40, no. 4, pp. 273–282, Nov. 1998.
[59] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1496–1503, Aug. 2010.
[60] S. L. Ho, M. Xie, L. C. Tang, K. Xu, and T. N. Goh, "Neural network modeling with confidence bounds: A case study on the solder paste deposition process," *IEEE Trans. Electron. Packag. Manuf.*, vol. 24, no. 4, pp. 323–332, Oct. 2001.
[61] T. Lu and M. Viljanen, "Prediction of indoor temperature and relative humidity using neural network models: Model comparison," *Neural Comput. Appl.*, vol. 18, no. 4, pp. 345–357, 2009.
[62] A. Khosravi, S. Nahavandi, and D. Creighton, "Improving prediction interval quality: A genetic algorithm-based method applied to neural networks," in *Proc. 16th Int. Conf. Neural Inf. Process. (ICONIP)*, Bangkok, Thailand, C. S. Leung, M. Lee, and J. H. Chan, Eds. Berlin, Germany: Springer, Dec. 2009, pp. 141–149.
[63] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Trans. Neural Netw.*, vol. 22, no. 9, pp. 1341–1356, Sep. 2011.
[64] L. Yang, T. Kavli, M. Carlin, S. Clausen, and P. F. M. De Groot, "An evaluation of confidence bound estimation methods for neural networks," in *Advances in Computational Intelligence and Learning*. Dordrecht, The Netherlands: Springer, 2002, pp. 71–84. [Online]. Available: https://link.springer.com/chapter/10.1007/978-94-010-0324-7_5
[65] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[66] R. Tibshirani, "A comparison of some error estimates for neural network models," *Neural Comput.*, vol. 8, no. 1, pp. 152–163, 1996.
[67] L. H. Ungar, R. D. De Veaux, and E. Rosengarten, "Estimating prediction intervals for artificial neural networks," in *Proc. 9th Yale Workshop Adapt. Learn. Syst.*, 1996, pp. 1–6.
[68] K. S. Kasiviswanathan and K. P. Sudheer, "Comparison of methods used for quantifying prediction interval in artificial neural network hydrologic models," *Model. Earth Syst. Environ.*, vol. 2, no. 1, pp. 1–11, 2016.
[69] A. Khosravi, "Construction and optimisation of prediction intervals for neural network models," Ph.D. dissertation, Centre Mater. Fibre Innov. Inst. Technol. Res. Innov., Deakin Univ., Geelong, VIC, Australia, 2010.
[70] A. Khosravi and S. Nahavandi, "An optimized mean variance estimation method for uncertainty quantification of wind power forecasts," *Int. J. Elect. Power Energy Syst.*, vol. 61, pp. 446–454, Oct. 2014.
[71] R. Dybowski and S. J. Roberts, "Confidence intervals and prediction intervals for feed-forward neural networks," in *Clinical Applications of Artificial Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2001, pp. 298–326.
[72] E. Zio, "A study of the bootstrap method for estimating the accuracy of artificial neural networks in predicting nuclear transient processes," *IEEE Trans. Nucl. Sci.*, vol. 53, no. 3, pp. 1460–1478, Jun. 2006.
[73] F. Giordano, M. La Rocca, and C. Perna, "Forecasting nonlinear time series with neural network sieve bootstrap," *Comput. Statist. Data Anal.*, vol. 51, no. 8, pp. 3871–3884, 2007.
[74] J. G. Carney, P. Cunningham, and U. Bhagwan, "Confidence and prediction intervals for neural network ensembles," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 2, 1999, pp. 1215–1218.
[75] R. Errouissi, J. Cardenas-Barrera, J. Meng, E. Castillo-Guerra, X. Gong, and L. Chang, "Bootstrap prediction interval estimation for wind speed forecasting," in *Proc. Energy Convers. Congr. Expo. (ECCE)*, Sep. 2015, pp. 1919–1924.
[76] A. Khosravi, S. Nahavandi, D. Srinivasan, and R. Khosravi, "Constructing optimal prediction intervals by using neural networks and bootstrap method," *IEEE Trans. Neural Netw. Lear. Syst.*, vol. 26, no. 8, pp. 1810–1815, Aug. 2015.
[77] T. Heskes, "Practical confidence and prediction intervals," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 176–182.
[78] S. Kumar and A. N. Srivistava, "Bootstrap prediction intervals in non-parametric regression with applications to anomaly detection," in *Proc. 18th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Beijing, China, Aug. 2012. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=20130014367
[79] C. P. I. J. van Hinsbergen, J. W. C. van Lint, and H. J. van Zuylen, "Bayesian committee of neural networks to predict travel times with confidence intervals," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 5, pp. 498–509, 2009.
[80] J. H. Zhao, Z. Y. Dong, Z. Xu, and K. P. Wong, "A statistical approach for interval forecasting of the electricity price," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 267–276, May 2008.
[81] S. G. Pierce, K. Worden, and A. Bezazi, "Uncertainty analysis of a neural network used for fatigue lifetime prediction," *Mech. Syst. Signal Process.*, vol. 22, no. 6, pp. 1395–1411, 2008.
[82] C. Zhang, H. Wei, L. Xie, Y. Shen, and K. Zhang, "Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework," *Neurocomputing*, vol. 205, pp. 53–63, Sep. 2016.
[83] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Trans. Neural Netw.*, vol. 22, no. 3, pp. 337–346, Mar. 2011.
[84] M. Paliwal and U. A. Kumar, "Neural networks and statistical techniques: A review of applications," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 2–17, 2009.
[85] Y. Chu, M. Li, H. T. C. Pedro, and C. F. M. Coimbra, "Real-time prediction intervals for intra-hour DNI forecasts," *Renew. Energy*, vol. 83, pp. 234–244, Nov. 2015.
[86] M. A. Hosen, A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction interval-based neural network modelling of polystyrene polymerization reactor—A new perspective of data-based modelling," *Chem. Eng. Res. Des.*, vol. 92, no. 11, pp. 2041–2051, 2014.
[87] A. Khosravi, S. Nahavandi, and D. Creighton, "Improving prediction interval quality: A genetic algorithm-based method applied to neural networks," in *Proc. Int. Conf. Neural Inf. Process.*, 2009, pp. 141–149.
[88] H. Quan, D. Srinivasan, and A. Khosravi, "Particle swarm optimization for construction of neural network-based prediction intervals," *Neurocomputing*, vol. 127, pp. 172–180, Mar. 2014.
[89] C. Wan, Z. Xu, and P. Pinson, "Direct interval forecasting of wind power," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4877–4878, Nov. 2013.
[90] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Optimal prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1166–1174, May 2014.
[91] W. Wu, K. Chen, Y. Qiao, and Z. Lu, "Probabilistic short-term wind power forecasting based on deep neural networks," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, 2016, pp. 1–8.
[92] P. Pinson and J. Tastu, "Discussion of 'prediction intervals for short-term wind farm generation forecasts' and 'combined nonparametric prediction intervals for wind power generation,'" *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 1019–1020, Jul. 2014.

[93] C. Wan, Z. Xu, J. Østergaard, Z. Y. Dong, and K. P. Wong, "Discussion of 'combined nonparametric prediction intervals for wind power generation,'" *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, p. 1021, Jul. 2014.

[94] A. Khosravi and S. Nahavandi, "Closure to the discussion of 'prediction intervals for short-term wind farm generation forecasts' and 'combined nonparametric prediction intervals for wind power generation' and the discussion of 'combined nonparametric prediction intervals for wind power generation,'" *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 1022–1023, Jul. 2014.

[95] L. G. Marín, F. Valencia, and D. Sáez, "Prediction interval based on type-2 fuzzy systems for wind power generation and loads in microgrid control design," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ)*, Jul. 2016, pp. 328–335.

[96] M. A. Hosen, A. Khosravi, S. Nahavandi, and D. Creighton, "Improving the quality of prediction intervals through optimal aggregation," *IEEE Trans. Ind. Electron.*, vol. 62, no. 7, pp. 4420–4429, Jul. 2015.

[97] G. Zhang, Y. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H. H. C. Iu, "An advanced approach for construction of optimal wind power prediction intervals," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2706–2715, Sep. 2015.

[98] F.-M. Tseng and Y.-C. Hu, "Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 1846–1853, 2010.

[99] Y.-Q. Zhang and X. Wan, "Statistical fuzzy interval neural networks for currency exchange rate time series prediction," *Appl. Soft Comput.*, vol. 7, no. 4, pp. 1149–1156, Aug. 2007.

[100] D. Peramunetilleke and R. K. Wong, "Currency exchange rate forecasting from news headlines," *Austral. Comput. Sci. Commun.*, vol. 24, no. 2, pp. 131–139, 2002.

[101] M. P. Clements and N. Taylor, "Evaluating interval forecasts of high-frequency financial data," *J. Appl. Econ.*, vol. 18, no. 4, pp. 445–456, 2003.

[102] J. Arroyo, R. Espínola, and C. Maté, "Different approaches to forecast interval time series: A comparison in finance," *Comput. Econ.*, vol. 37, no. 2, pp. 169–191, 2011.

[103] E. Cohen and I. Saguy, "Statistical evaluation of Arrhenius model and its applicability in prediction of food quality losses," *J. Food Process. Preservation*, vol. 9, no. 4, pp. 273–290, 1985.

[104] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras, "Artificial intelligence in short term electric load forecasting: A state-of-the-art survey for the researcher," *Energy Convers. Manage.*, vol. 44, no. 9, pp. 1525–1534, 2003.

[105] V. Almeida and J. Gama, "Prediction intervals for electric load forecast: Evaluation for different profiles," in *Proc. 18th Int. Conf. Intell. Syst. Appl. Power Syst. (ISAP)*, 2015, pp. 1–6.

[106] H. Quan, D. Srinivasan, and A. Khosravi, "Uncertainty handling using neural network-based prediction intervals for electrical load forecasting," *Energy*, vol. 73, pp. 916–925, Aug. 2014.

[107] A. U. Haque, P. Mandal, J. Meng, and M. Negnevitsky, "Wind speed forecast model for wind farm based on a hybrid machine learning algorithm," *Int. J. Sustain. Energy*, vol. 34, no. 1, pp. 38–51, 2015.

[108] M. A. Abdullah, K. M. Muttaqi, and A. P. Agalgaonkar, "Sustainable energy system design with distributed renewable resources considering economic, environmental and uncertainty aspects," *Renew. Energy*, vol. 78, pp. 165–172, Jun. 2015.

[109] Y. Wang, Z. Zhou, A. Botterud, and K. Zhang, "Optimal wind power uncertainty intervals for electricity market operation," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 199–210, Jan. 2018.

[110] C. Wan, J. Wang, J. Lin, Y. Song, and Z. Y. Dong, "Nonparametric prediction intervals of wind power via linear programming," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1074–1076, Jan. 2018.

[111] N. Mazzi and P. Pinson, "Wind power in electricity markets and the value of forecasting," in *Renewable Energy Forecasting*. Amsterdam, The Netherlands: Elsevier, 2017, pp. 259–278.

[112] A. U. Haque, M. H. Nehrir, and P. Mandal, "Solar PV power generation forecast using a hybrid intelligent approach," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PES)*, Jul. 2013, pp. 1–5.

[113] S. Islam and N. Mithulananthan, "PV based EV charging at universities using supplied historical PV output ramp," *Renew. Energy*, vol. 118, pp. 306–327, Apr. 2018.

[114] S. Islam, N. Mithulananthan, and D. Q. Hung, "A day-ahead forecasting model for probabilistic EV charging loads at business premises," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 741–753, Apr. 2018.

[115] D. AlHakeem, P. Mandal, A. U. Haque, A. Yona, T. Senjyu, and T.-L. Tseng, "A new strategy to quantify uncertainties of wavelet-GRNN-PSO based solar PV power forecasts using bootstrap confidence intervals," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2015, pp. 1–5.

[116] R. Blonbou, "Very short-term wind power forecasting with neural networks and adaptive Bayesian learning," *Renew. Energy*, vol. 36, no. 3, pp. 1118–1124, 2011.

[117] E. Pratheepraj, A. Abraham, S. N. Deepa, and V. Yuvaraj, "Very short term wind power forecasting using PSO-neural network hybrid system," in *Advances in Computing and Communications*. Berlin, Germany: Springer, 2011, pp. 503–511. [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-642-22720-2_53, doi: 10.1007/978-3-642-22720-2.

[118] G. N. Kariniotakis, G. S. Stavrakakis, and E. F. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE Trans. Energy Convers.*, vol. 11, no. 4, pp. 762–767, Dec. 1996.

[119] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, 2009.

[120] E. Cadenas and W. E. Rivera, "Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks," *Renew. Energy*, vol. 34, no. 1, pp. 274–278, 2009.

[121] I. M. Galván, J. M. Valls, A. Cervantes, and R. Aler, "Multi-objective evolutionary optimization of prediction intervals for solar energy forecasting with neural networks," *Inf. Sci.*, vol. 418, pp. 363–382, Dec. 2017.

[122] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, vol. 403. Hoboken, NJ, USA: Wiley, 2007.

[123] J. Nowotarski and R. Weron, "Computing electricity spot price prediction intervals using quantile regression and forecast averaging," *Comput. Statist.*, vol. 30, no. 3, pp. 791–803, 2015.

[124] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation for electricity price and demand using support vector machines," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2014, pp. 3995–4002.

[125] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *Int. J. Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.

[126] J. Nowotarski, J. Tomczyk, and R. Weron, "Robust estimation and forecasting of the long-term seasonal component of electricity spot prices," *Energy Econ.*, vol. 39, pp. 13–27, Sep. 2013.

[127] A. Khosravi, S. Nahavandi, and D. Creighton, "Quantifying uncertainties of neural network-based electricity price forecasts," *Appl. Energy*, vol. 112, pp. 120–129, Dec. 2013.

[128] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using PSO-tuned support vector machines," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 322–331, Apr. 2015.

[129] R. A. M. Noor, Z. Ahmad, M. M. Don, and M. H. Uzir, "Modelling and control of different types of polymerization processes using neural networks technique: A review," *Can. J. Chem. Eng.*, vol. 88, no. 6, pp. 1065–1084, 2010.

[130] A. Mayr, T. Hothorn, and N. Fenske, "Prediction intervals for future bmi values of individual children—A non-parametric approach by quantile boosting," *BMC Med. Res. Methodol.*, vol. 12, no. 1, p. 6, 2012.

[131] H. Nishiura, "Early detection of nosocomial outbreaks caused by rare pathogens: A case study employing score prediction interval," *Osong Public Health Res. Perspect.*, vol. 3, no. 3, pp. 121–127, 2012.

[132] K. McGeechan *et al.*, "Prediction of incident stroke events based on retinal vessel caliber: A systematic review and individual-participant meta-analysis," *Amer. J. Epidemiol.*, vol. 170, no. 11, pp. 1323–1332, 2009.

[133] R. M. Hansen *et al.*, "Retinal degeneration in children: Dark adapted visual threshold and arteriolar diameter," *Vis. Res.*, vol. 48, no. 3, pp. 325–331, 2008.

[134] B. Zee *et al.*, "Stroke risk assessment for the community by automatic retinal image analysis using fundus photograph," *Qual. Primary Care*, vol. 24, no. 3, pp. 114–124, 2016.

[135] R. Taormina and K.-W. Chau, "ANN-based interval forecasting of streamflow discharges using the LUBE method and MOFIPS," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 429–440, Oct. 2015.

[136] C. Chatfield, "Prediction intervals for time-series forecasting," in *Principles of Forecasting*. Boston, MA, USA: Springer, 2001, pp. 475–494. [Online]. Available: https://link.springer.com/chapter/10.1007/978-0-306-47630-3_21

[137] S. C. Wheelwright, S. G. Makridakis, and R. J. Hyndman, *Forecasting: Methods and Applications*. Hoboken, NJ, USA: Wiley, 1998.
[138] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
[139] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[140] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[141] H. M. D. Kabir, A. Khosravi, A. Hosen, and S. Nahavandi, "Partial adversarial training for prediction interval," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2018, pp. 1–6.
[142] K. F. Wallis, "Chi-squared tests of interval and density forecasts, and the bank of England's fan charts," *Int. J. Forecasting*, vol. 19, no. 2, pp. 165–175, 2003.

H. M. DIPU KABIR (S'10) received the B.Sc. degree from the Department of EEE, Bangladesh University of Engineering and Technology, in 2011, and the M.Phil. degree from the Department of ECE, The Hong Kong University of Science and Technology, in 2016. He is currently pursuing the Ph.D. degree with Deakin University. He was with the Samsung Bangladesh R&D Center as a Software Engineer from 2011 to 2013.

ABBAS KHOSRAVI (GS'07–M'10) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2002, the M.Sc. degree in electrical engineering from the Amirkabir University of Technology, Tehran, in 2005, and the Ph.D. degree from Deakin University, Geelong, VIC, Australia, in 2010.

From 2006 to 2007, he was with the eXiT Group, University of Girona, Spain, conducting research in the field of artificial intelligence. He is currently a Research Fellow with the Institute for Intelligent Systems Research and Innovation, Deakin University. His primary research interests include computational intelligence-based data mining, modeling, and optimization.

Dr. Khosravi received the Alfred Deakin Postdoctoral Research Fellowship in 2011.

MOHAMMAD ANWAR HOSEN received the B.Sc. degree (Hons.) in chemical engineering from the Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2004, the M.Eng.Sc. degree in chemical engineering from the University of Malaya, Kuala Lumpur, Malaysia, in 2010, and the Ph.D. Degree from Deakin University, Waurn Ponds, Australia, in 2015. He is currently an Industry Post-Doctoral Research Fellow with the School of Engineering, Deakin University.

His current research interests include the development and application of AI techniques for the modeling and control of nonlinear process, analytical modeling, and discrete event simulation.

SAEID NAHAVANDI (M'92–SM'07) received the Ph.D. degree from Durham University, U.K., in 1991. He is currently an Alfred Deakin Professor, a Pro Vice-Chancellor (Defence Technologies), the Chair of engineering, and the Director of the Institute for Intelligent Systems Research and Innovation, Deakin University. He has published over 750 scientific papers in various international journals and conferences. His research interests include modeling of complex systems, robotics, and haptics.

He is a fellow of Engineers Australia and the Institution of Engineering and Technology.

He is the Co-Editor-in-Chief of the IEEE Systems Journal, and an Associate Editor of the IEEE/ASME Transactions on Mechatronicsand the IEEE Transactions on Systems, Man and Cybernetics: Systems. He is an IEEE Access Editorial Board Member, and IEEE Press Editorial Board.