A Survey of Uncertainty Quantification in Deep Learning for Financial Time Series Prediction

Txus Blasco[a,*], Vicente García[b], J. Salvador Sánchez[a]

[a] *Institute of New Imaging Technologies, Dept. Computer Languages and Systems, Universitat Jaume I, Avinguda de Vicent Sos Baynat, s/n, Castello de la Plana, Castello, 12006, Spain*
[b] *Departamento de Ingenieria Electrica y Computacion, Division Multidisciplinaria en Ciudad Universitaria, Universidad Autonoma de Ciudad Juarez, Avenida Plutarco Elias Calles 1210, Ciudad Juarez, Mexico*

Abstract

Investors make decisions about buying and selling a financial asset based on the information available. The traditional approach in deep learning when it comes to predicting the behaviour of an asset is to take a price history, train a model, and forecast a single price in the near future. This is called the frequentist perspective. Uncertainty quantification is an alternative in which the models manage a probability distribution for the prediction. It provides investors with more information than the traditional frequentist form, so that they can consider the risk of making or not making a certain decision. Thus, we systematically review the existing literature in the period 2001–2022 on deep learning uncertainty quantification methods to predict the behaviour of financial assets, such as currencies, the stock market, cryptocurrencies, among others. The article discusses types of models, categories of financial assets, prediction characteristics, and types of uncertainty. We found that, generally speaking, the publications focus on price accuracy as a metric, although other metrics, such as trend accuracy, might be more suitable. Very few authors analyse both epistemic and aleatory uncertainty, and none analyse in depth how to unlink them.

Keywords: Deep learning, Time series, Prediction, Uncertainty quantification, Financial assets

1. Introduction

Deep learning (DL) applications have increasing relevance in almost every field, particularly finance. Along with this growing interest, it seems that most authors are participating in a model race to build the most accurate model, predicting future asset prices with the smallest error. This competition could make sense in some applications where the level of volatility, called noise, stays between low and moderate levels. However, the

*Corresponding author
Email addresses: `al419983@uji.es` (Txus Blasco), `vicente.jimenez@uacj.mx` (Vicente García), `sanchez@uji.es` (J. Salvador Sánchez)

Preprint submitted to Elsevier *June 8, 2023*

financial market presents a very high level of volatility. The amount of noise related to this type of time series is still so high that researchers are still debating whether or not the Efficient Market Theory is valid [67]. Instead of looking for a very accurate prediction, traders look for a reasonable prediction that helps them understand which way the market will move next, up or down. A successfully predicted trend is much more helpful in securing an investment than knowing, for example, if the next price will hit 29.4€instead of 29.5€. In this context, it seems reasonable to ask what is the point of trying to boost the accuracy of DL models in a scenario where the volatility of the input data is much higher than the minimum error that can be obtained from the predictions. Trying to follow that line will lead to an overfit model, inevitably falling into the bias-variance dilemma. The authors of this review propose an alternative approach to predicting financial trading series, in which the next price vector falls within a confidence interval. That range can be defined by the degree of certainty of the model and the amount of noise in the raw input data. Defining that interval means quantifying the uncertainty of both the data noise and the degree of insight gained by the DL model.

We can distinguish between two types of uncertainties in the prediction with DL models, called *aleatory* and *epistemic* [36, 56]. The total uncertainty in the prediction is the sum of both uncertainties. The first is related to the intrinsic noise in the data and is called aleatory uncertainty. Aleatory refers to the randomness inherent in price data, that is, price fluctuations or volatility confound genuine underlying trends [1]. Noise is an intrinsic part of financial data and is the main reason why predicting upcoming prices remains so complex: DL models struggle to recognize the difference between noise and a real trend. In fact, financial data like stock market prices differs from other time series in that it is characterized by a very high level of noise. The stochastic component of financial data misrepresents valuable price information and thus confounds the models. Unfortunately, financial data always has this stochastic component and it is impossible to reduce the related uncertainty simply by adding more data [44]. Therefore, the need to address and quantify this type of uncertainty is clear: traders want to distinguish between a real trend led by market dynamics and a false trend led by noise, in order to keep trading profitable. The identification of a random part is the previous and necessary step to address it. Although aleatory uncertainty cannot be eliminated, noise can and should be treated as a necessary preliminary step before training DL models. If the data is not cleaned and filtered to remove noise beforehand and raw data is ingested by the model, the prediction accuracy is negatively affected. The reason is that the model will stick to noise patterns instead of valuable real price trends [32].

The second type of uncertainty, the epistemic component, is related to the prediction model, coming from the sets of hyperparameters not chosen for the prediction. A trained DL model is defined by several parameters related to its own architecture, that is, the combination of parameters (W, b) corresponding to the neural network weights and biases, length and depth, shape (variation of length through depth), the number of iterations and batches, the optimizer or the regularization term, to name a few. Instead of just choosing an optimized combination of those parameters, one could think of a model from a different perspective, considering that not just one but many of those combinations could form a set of valid models. In this way, we do not have a matrix of weights and biases, but the model handles a probability distribution corresponding to each of the weights of the matrix (epistemic uncertainty) [44].

In 1992, Mackay [68] established the foundations of what is now called uncertainty quantification (UQ) and applied them to the approximation of functions. He claims that the Bayesian approach is not consistently better in performance than other methods for interpolating a noisy data set. However, a second level of inference is generally forgotten, which is the ability of Bayesian methods to classify alternative models to the best chosen one.

If we take this concept to finance, it translates into opportunity cost, that is, how much we are paying for not choosing the alternative to the (supposed) best operation at the right time. Big financial decisions are not made simply by knowing a single future price value recommended by a DL model (this is how a frequentist approach is defined [79]). Risk and opportunity cost [85] must also be considered. We believe the UQ approach is best suited for real financial scenarios where probability and confidence intervals can be entered, not just for a forecast point in the future, but for entire identifiable trends.

UQ is defined differently throughout the existing literature. Within the field of finance, if we continue to delve into the even more specific sub-domain of predicting the behaviour of financial and marketable assets through DL, it can be defined as *the science of measuring the amount by which a DL model it is uncertain about the benefit that can be obtained in the future*. Ideally, this amount of uncertainty would be measured with a probability distribution [104], or at least with the time-varying mean and variance [103]. UQ can be applied to a regression problem, that is, predicting future price values [39]. It can also be applied to a classification problem, where the output of a DL model is a buy/hold/sell signal [71] or the start or end of a trend [98]. It can even be applied to a ranking type prediction, for example, the prediction of momentum in the future.

The main objective of this survey is to show if the state of the art regarding UQ in DL applied to financial forecasting can leave unexplored or insufficiently explored techniques, methods or approaches that could be exploited by researchers in possible future investigations and to clarify what would be. Some other literature reviews explore UQ applied to various fields: water resources research [30], flood forecasting [33], climate modelling [90], cycle cost estimation of aerospace life [92] and others. To the best of our knowledge, this is the first survey on the subject of UQ applied to financial time series.

Hereafter, the paper is organized as follows. Section 2 describes the existing techniques commonly used to quantify uncertainty and provides the definitions and terms that will be used in the remainder of the article. Section 3 explains the methodology used in the review to analyse the existing literature, following the PRISMA approach for records published between 2001 and 2022 without time period filters. Section 4 compiles the results, while Section 5 discusses the results and provides potential directions for future research. Finally, Section 6 presents the main conclusions of this work.

2. Research methodology

PRISMA was used as the main framework to carry out the research. In addition, the following questions were defined to shed light on the current state of the art in UQ for forecasting financial assets using DL:

1. What is the state of the art related to research on UQ for DL applied to finance?
2. What are the main UQ techniques in DL used so far for financial forecasting?
3. What needs in financial forecasting are not yet covered in UQ for DL?
4. What technical challenges remain in financial forecasting using UQ in DL?
5. What approaches could potentially be explored to overcome those challenges?

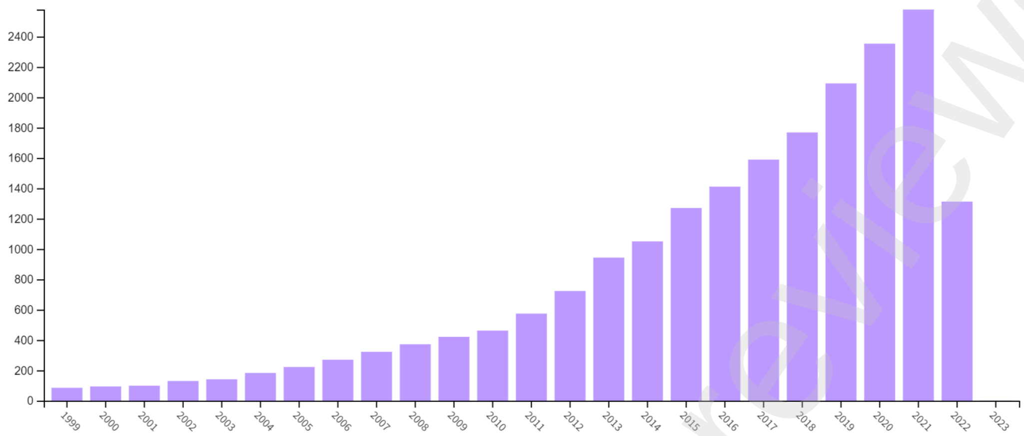

Figure 1: Number of publications per year related to the general term Quantification of uncertainty. Source: Web of Science.

2.1. Inclusion and exclusion criteria

UQ in DL is a relatively new domain. As can be seen in Figure 1, it has been found that from 2010 there is an explosion of publications that deal with the subject and explicitly mention the exact phrase "quantification of uncertainty". While some articles mentioned UQ before then, they were more sparse. In fact, the first article that applies UQ methods to a neural network is the work of Mackay [69], in which a Bayesian regularization is used.

Not all publications found use UQ intentionally, nor do they mention it explicitly. Even so, made on purpose or not, all the selected articles meet the criteria for the application of UQ techniques. For example, the most typical unintentional use of UQ methods is dropout as a regularization technique. The most likely reason why the authors might want to apply it is to decrease the probability of overfitting, since the technique causes the neural network to systematically forget data at each backpropagation step. Applying this method is equivalent to approximating the Bayesian posterior, which is often not mentioned in most references.

The three parameters considered in the publications that will be added to this work are described below (all must be met):

1. **Financial asset time series prediction and trading**: The publication should base the topic on methods or techniques for predicting the value of trading instruments (Forex, stocks, commodities and cryptocurrencies, see Figure 2). The problem to be solved can be regression, predicting the future value of the asset, confidence and return intervals or classification, predicting bullish or bearish trends or buy/sell signals. The condition for being included in this group is that the asset is negotiable, that is, that the asset can be bought and sold as a financial product for speculation.
2. **Use of DL**: The prediction must be made using DL. The classification can be done, for example, by categorizing a price matrix with a buy signal (go long) or a sell signal (go short). Regression is used to predict single price data points in the

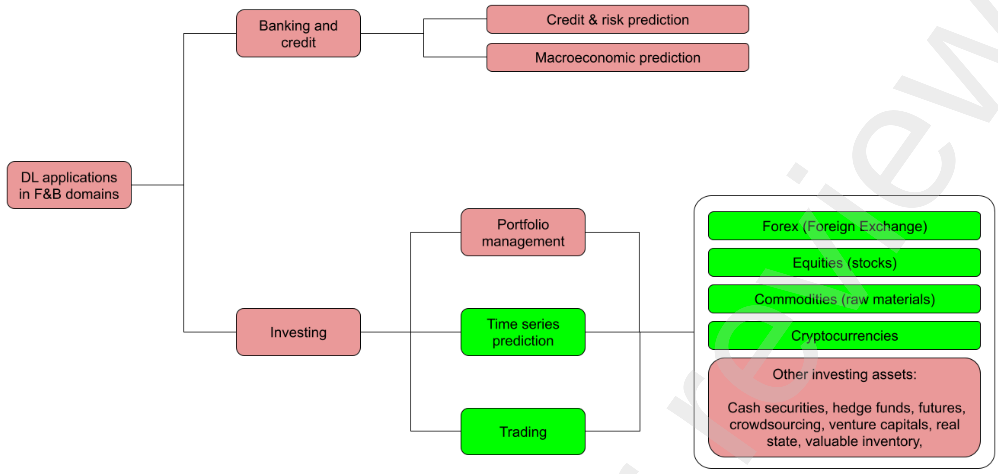

Figure 2: Type of financial and banking applications in DL. Based on Huang 2020 [40] and updated with Investopedia [25].

future, multiple step forward price data points, or price ranges. Figure 3 shows a taxonomy of the main DL methods used to quantify uncertainty in the forecast of financial time series.

3. **Use of UQ methods**: The methods used must fall within the UQ domain. Even the unintentional use of UQ methods has been considered and added to the survey. Based on the work of Abdar et al. [3], a list of UQ methods has been adapted as the main criteria to include an article. These are listed and described in Section 3, divided by type of techniques.

Some studies, even predicting economic time series, have been rejected because the data does not correspond to a financial asset. For example, Mishra and Ayyub [76] use a long short-term memory (LSTM) of Montecarlo (MC) churn to predict the net state domestic product, which is not noisy enough compared to financial markets to be included in this survey.

Other studies focused on portfolio management have been excluded because they do not solve a value prediction problem. Instead, they are working out an asset classification or a ranking. Sometimes the portfolio management problem can classify the behaviour of the value curve in the past, but it does not predict the behaviour of the asset value in the future. An example of this can be seen in the article by Park et al. [82]. However, it does not mean that the portfolio management is excluded from this survey. In the case of the paper by Lin and Blum [59], the portfolio management problem includes a time series prediction with UQ, so it is considered in the study.

Some keywords may cause confusion when meeting the inclusion criteria. For example, when searching for the terms *regularization*, *stock price prediction*, and *DL*, some studies were initially included in the selection because they have both in the title or abstract. However, they were discarded after reading the abstract and realizing that they do not use UQ techniques. Wu et al. [114] use DL for stock price prediction; however, the

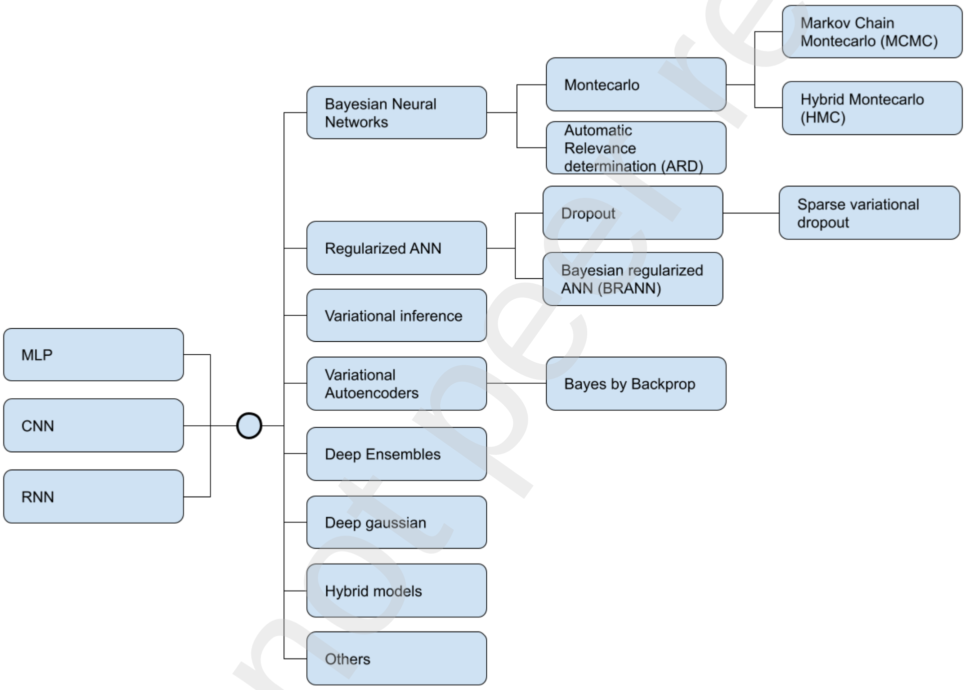

Figure 3: Taxonomy of the main DL methods used to quantify uncertainty in financial time series prediction.

term *regularization* is not used as a synonym for dropout in neural networks, but rather as a way to restrict the decision tree. Similar cases can be found in other articles [91, 117]. In addition, when searching for *neural networks Bayesian neural networks* in price prediction, a few other studies related to *Bayesian Networks* were added. However, those are not deep neural networks, but just networks, and therefore discarded after screening [11, 55, 74, 100, 115, 127, 128]. It was necessary to include the broader term *network* instead of the exact term *neural network* because some authors, possibly prompted by the (not so) obvious use the former to refer to the latter [112].

Other studies, such as the one by Alarab et al. [6], use MC-dropout DL for valid/invalid classification of Bitcoin transactions and the publication was dropped because it does not include a regression problem, Bitcoin future value prediction, trend, not a purchase/sale classification problem. On the other hand, only studies in English have been included. Therefore, works in other languages have been discarded [52, 60]. Additionally, an article by Liu and Wang [63] was removed from selected records because the single-point price prediction was made considering not only past data, but also future data. This publication was discarded as it is not a price prediction as future data in real world scenarios will never be available.

Variational autoencoder (VAE) is a widely used method for generating synthetic and resampling data sets on financial time series. All studies that do not directly use VAE for the prediction of the next value or trade have been excluded from this survey [15, 21, 97].

2.2. Search strategy

We use Scopus and Google Scholar as the main search engines in the databases. The search strings were refined until we found the most appropriate records. We chose Scopus because it is a widely recognized automatic database that shows quality results. However, the number of records chosen for selection after refinement was only 171, too low a number for our purposes. That is why we decided to use a more general search engine such as Google Scholar, in which we could add 251 more records to sift through. Due to the low initial number of records found, we decided not to narrow the range of dates. As long as a study met the requirements described in Section 2.1, it was included without date restriction. In this way, the total time range covered by this review goes from 2001 to 2022. In total, considering both the database and the search engine, we found 15,295 raw records. After refinement and removal of duplicates, 422 records were reviewed, 156 were evaluated, and 69 were added to the study. Google Scholar shows large numbers for some searches and the search was limited to the first 10 pages displayed as these are considered the most relevant. Table 1 shows a summary of the number of records found, sorted by database/search engine.

2.3. Study selection process

The main idea behind this survey is to bring together all the studies that directly represent or use uncertainty within the architecture of neural networks applied to financial time series forecasting. With this idea in mind, the search queries have been designed. A record was chosen for screening only if its title and abstract represent the aim of the study, being rejected otherwise. The selection process involves extensive reading and gathering additional information to understand the article. If the work did not meet the selection criteria, it was rejected. Throughout this process, the selection criteria were

Table 1: Search queries in two different sources.

Search Query	Scopus	Google Scholar
("Bayesian neural network") AND ("stock") OR ("forex") OR ("foreign exchange") OR ("cryptocurrency") OR ("bitcoin") OR ("financial")	24	2730
("uncertainty quantification") AND ("forex") OR ("foreign exchange") OR ("crypto*") OR ("bitcoin") OR ("stock") OR ("financ*") AND ("deep") OR ("neural network*")	14	45300
("dropout") AND ("neural network") AND ("stock") OR ("crypto*") OR ("financ*") AND ("prediction")	23	37600
("markov chain montecarlo") OR ("MCMC") AND ("neural network") AND ("stock") OR ("crypto*") OR ("financ*") AND ("prediction")	4	19300
("variational inference") AND ("neural network") AND ("stock") OR ("crypto*") OR ("financ*") AND ("prediction")	4	8040
("Bayesian active learning") AND ("bitcoin") OR ("crypto*") OR ("financ*") OR ("stock") OR ("Forex") OR ("foreign exchange")	0	5
("bayes by backprop") AND ("bitcoin") OR ("crypto*") OR ("financ*") OR ("stock") OR ("Forex") OR ("foreign exchange")	0	4
("variational autoencoder*") AND ("bitcoin") OR ("crypto*") OR ("financ*") OR ("stock") OR ("Forex") OR ("foreign exchange")	53	208
("deep gaussian") AND ("bitcoin") OR ("crypto*") OR ("financ*") OR ("stock") OR ("Forex") OR ("foreign exchange")	1	28
("laplace approximation*") AND ("bitcoin") OR ("crypto*") OR ("financ*") OR ("stock") OR ("Forex") OR ("foreign exchange")	46	49
("deep ensemble*") OR ("deep Bayesian ensemble*") OR ("Bayesian ensemble*") OR ("Bayesian deep ensemble*") AND ("bitcoin") OR ("crypto*") OR ("financ*") OR ("stock") OR ("Forex") OR ("foreign exchange")	16	46

refined to ensure that all articles met the objectives set in the study. Figure 4 shows a summary of the results of the search and selection process.

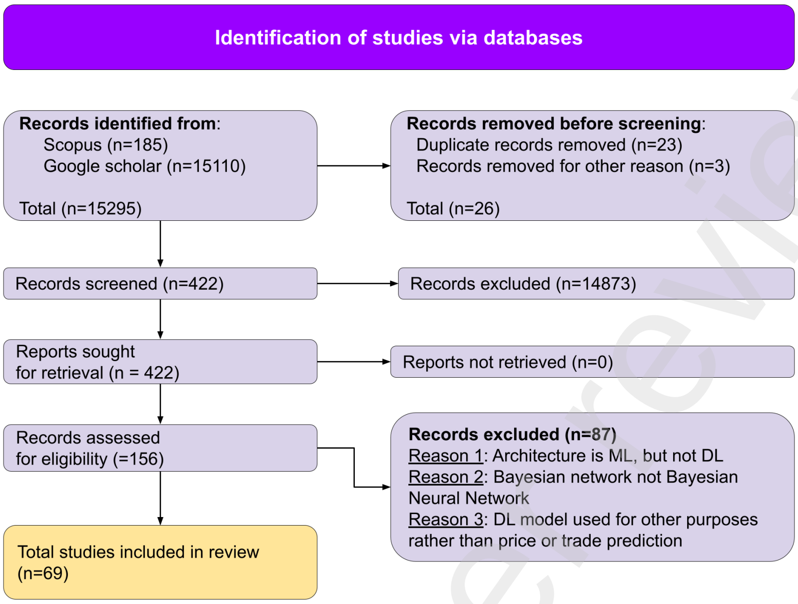

Figure 4: Results of the search and selection process, from the number of records identified in the search to the number of studies included in the survey.

3. Definitions, terminology and techniques

In this section, we will present the main techniques used in the literature for a probabilistic approach to time series forecasting, providing a brief description of each of them and some definitions. Following the proposal of Gordon Wilson and Izmailov [113], we will not differentiate between Bayesian and non-Bayesian methods to approximate the exact posterior distribution, since some of the non-Bayesian methods approximate the integral even better than the Bayesian ones. Figure 3 has been included to shed light on the main techniques that can be found in the literature.

3.1. Bayesian neural networks (BNN)

For a financial time series $D = \{X, Y\} = \{(x_i, y_i)\}_i^N$ is a training data set with historical price inputs $x_i \in \mathbb{R}$ and labelled price outputs $y_i \in \mathbb{R}$. From a frequentist point of view, a neural network tries to represent future values of a function $y = \Phi_\theta(x)$, given that it learned the behaviour of Φ in the past. Let f and l be a nonlinear transformation (activation function) and a linear transformation, respectively. Then, in the simplest architecture of a neural network, l_i is applied to W_i in each hidden layer:

$$l_0 = x, \tag{1}$$

$$l_i = s_i(W_i l_i - 1 + b_i) \forall i \in [1, n], \tag{2}$$

$$y = l_n \tag{3}$$

A stochastic neural network is a specific type of artificial neural network (ANN) in which the activation function is stochastic or the weights are stochastic. Stochastic means that those values are not single, but a set of values that make up a probability distribution [126]. A BNN would then be a stochastic neural network trained using Bayesian inference [49]. Given a traditional ANN, learning is the process of regressing a set of parameters $\theta = (W, b)$ from the training data D, where $D = (x, y)$ consists of a vector of input data values x and their corresponding values y, W is a weight matrix and b is a bias vector. From the set of parameters θ, the training process consists of optimizing a cost function, or from a probabilistic point of view, a maximum likelihood estimate, to achieve $\hat{\theta}$, a set of parameters where W and b are unique and optimal. The above is the description of a frequentist approach to DL. Viewed from another perspective, the probabilistic (Bayesian) approach states:

$$\theta \sim p(\theta) \tag{4}$$

$$y = \Phi_\theta(x) + \epsilon \tag{5}$$

where $p(\theta)$ is the prior probability associated to all possible models θ that explain y as an approximation to the real, but unknown function $\Phi_\theta(x)$, being ϵ the representation of random noise added to the previous function.

In summary, the above means that the goal here is not to find an exact set of parameters θ, but a probability distribution $p(\theta)$ given that we know D, i.e. the posterior probability $p(\theta|D)$. Applying the Bayesian rule and enforcing the independence between the model parameters and the inputs, it can be affirmed that:

$$p(\theta|D) = \frac{p(D_y|D_x, \theta)p(\theta)}{\int_\theta p(D_y|D_x, \theta')p(\theta')d\theta'} \tag{6}$$

where $p(D_y|D_x, \theta)$ is the likelihood, $\int_\theta p(D_y|D_x, \theta')p(\theta')d\theta'$ refers to evidence, and $p(\theta)$ denotes the prior probability.

Instead of using backpropagation over past data as in ANN, BNN learns the probability distribution of the weights by approximating the posterior [49]. However, in practice, the evidence is computationally expensive and is never calculated analytically. Instead, it is approximated by methods such as Markov chain Monte Carlo or variational inference which are described later.

3.2. Montecarlo dropout

Montecarlo dropout consists of randomly disconnecting different pairs of neurons during backpropagation for training and testing. This prevents overfitting and approximates the posterior probability described above. In fact, Gal and Ghahramani [23] state that a neural network with arbitrary depth, without nonlinearities, and with a dropout approximates Bayesian posterior probability. The objective function that uses the L_2 regularization is defined as:

$$\mathcal{L}_{dropout} := \frac{1}{N}\sum_{i=1}^{N} E(y_i, \hat{y_i}) + \lambda \sum_{l=1}^{L}(||W_l||_2^2 + ||b_l||_2^2) \tag{7}$$

MC dropout has the advantage of low computational cost and simplicity; although, to mention one drawback, this technique needs others to complement it, since by itself it cannot capture the uncertainty of the data [94].

3.3. Markov chain Montecarlo (MCMC)

MCMC is a set of algorithms for sampling from a probability distribution. The objective is to approximate the unknown distribution of weights of a neural network corresponding to the posterior probability $p(w)$. The intuition behind MCMC is that a different but known distribution $f(x)$ can be built using a stationary Markov chain such that $p(w)$ can be sampled from $f(x)$, where $[S_0, S_i]$ is the burn-in initial sequence of the Markov chain and $[S_{i+1}, S_n]$ is the final stationary sequence of the Markov chain for a Markov chain of length n. Usually, $f(x)$ is chosen as an easy or convenient distribution, such as a normal or binomial distribution. If the chosen distribution is normal, then the MCMC method is called the Metropolis method. In a more general definition, the sampling distribution can be convenient, different from normal, and the method in this case is called Metropolis-Hastings [17, 49].

3.4. Variational inference (VI)

MCMC methods are effective if the size of the data sets is moderate, since they do not scale well. For large data sets, variational inference [9] is preferred. Rather than sampling from the exact posterior distribution, variational inference uses optimization to search for the best performance of a family $\mathcal{Q}$ of approximate densities, each of which attempts to minimize the Kullback-Liebler (KL) divergence to the exact posterior

$$q^*(z) = \arg\min KL(q(z)||p(z|x))q(z) \in \mathcal{Q}, \tag{8}$$

then the posterior is approximated using the best member of the family $\mathcal{Q}$, $q^*(\cdot)$. Because $\mathcal{Q}$ can be chosen, it should be made as flexible as possible to capture $p(x|z)$ closely, but not too complex to make optimization difficult. Wainwright and Jordan [102] state that any method that uses optimization to approximate a probability density can be called *variational inference*. If this rule is followed, then some other procedures become part of the variational inference approach: expectation propagation [75], belief propagation [121] or Laplace approximation.

3.5. Variational autoencoder

Variational Bayesian (VB), also called neural variational inference, is a way of optimizing an approximation of the posterior probability, which has already been described as analytically intractable [54]. An autoencoder (AE) is a specific type of DL model that consists of two components, the encoder and the decoder. Figure 5 shows the basic scheme of an autoencoder. The goal of the encoder is to map a high-dimensional input vector $x = \{x_1, x_2, ..., x_n\}$ into a low-dimensional latent output vector $z = \{z_1, z_2, ..., z_m\}$ using a function f such that

$$z = f(x) = S_f(Wx + b), \tag{9}$$

where S_f is the activation function. The parameters that define the encoder are a $m \times n$ matrix of weights w and the bias vector $b \in R^m$.

The decoder reconstructs the latent vector z back to an output vector $x' = \{x'_1, x'_2, ..., x'_n\}$ using a function g such that

$$x' = g(z) = S_g(W'z + b'), \tag{10}$$

where S_g is the activation function of the decoder. The decoder parameters are defined by a $n \times m$ matrix of weights and a bias vector $b' \in R^n$. The training objective of the autoencoder is to minimize the reconstruction error between x and x' [123]. In this way, a learning representation for a high-dimensional distribution can be converted into a simpler variational inference problem [29]. Given a sample space defined by a variable x and a latent sample space given by a latent variable z, then the probability distribution $P_\theta(x)$ can be written as

$$p_\theta(x) = \int_z p_\theta(x|z)p(z). \tag{11}$$

The evidence lower bound is defined as

$$\log p_\theta(x) = \mathbb{E}_{q_{\phi(z|x)}}[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)||p(x)), \tag{12}$$

where $q_\phi(z|x)$ is the encoder, $p_\theta(x|z)$ is the decoder, ϕ and θ are their respective parameters and D_{KL} is the Kullback-Liebler divergence.

3.6. Bayesian active learning (BAL)

Active learning methods can be understood as a mechanism to improve the way models are trained. The active learning algorithm decides which data values are most relevant, and as a consequence, it can be used when there is a large amount of data to reduce the computation time for training [95]. If we now recall the Bayesian approach, we have a large amount of data in a probability distribution that describes the posterior to model the uncertainty of the weights. However, not all data points within the distribution provide the greatest amount of information. Then we can let the active learning mechanism decide about which points the model has more uncertainty and choose them. Then, an *oracle* (as the standard literature calls it), provides the label to the chosen and most uncertain data points [24]. In our finance case, since it is a regression problem, the labels are numbers. Also in a buy/sell approach, the regression becomes a classification problem where the labels are *buy*, *sell*, or *hold*. As we will see in the following sections, to the best of our knowledge, there are still no publications that apply Bayesian active learning to a financial problem as defined in this survey.

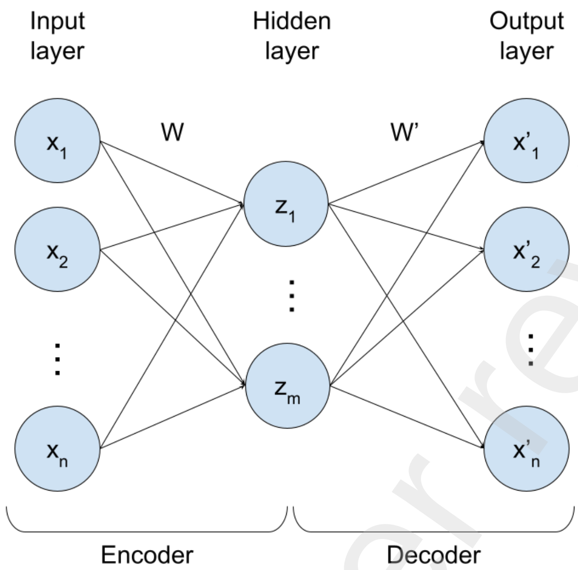

Figure 5: Basic architecture of autoencoders.

3.7. Bayes by backprop (BBB)

The idea behind Bayes by backprop lies again in the approximation of the posterior distribution defined previously in Equation 6 [10]. Once again we make use of KL divergence of Equation 8, but we experience another intractability given by the integral

$$KL[q_\theta(w|D)||p(w|D)] = \int q_\theta(w|D) log \frac{q_\theta(w|D)}{p(w|D)} dw, \tag{13}$$

which brings us to the evidence lower bound (Equation 12). Taking advantage of Blundell's first proposition that "under certain conditions, the derivative of an expectation can be expressed as the expectation of a derivative", we are generalizing the Gaussian reparameterization trick [80]. To turn intractability into something treatable, we follow the steps

$$\text{sample from } \epsilon \sim N(0, 1) \tag{14}$$

and set Θ as

$$\Theta = w = \mu + \sigma \cdot \epsilon \tag{15}$$

where μ is the mean and σ is the standard deviation of the Gaussian distribution. Finally, one can calculate the unbiased MC-gradients w.r.t μ and σ for the expected loss and optimize the variational parameters according to

$$\Delta\mu = \frac{\partial\Theta}{\partial w} + \frac{\partial\Theta}{\partial\mu}, \tag{16}$$

$$\Delta\sigma = \frac{\partial\Theta}{\partial w}\frac{\epsilon}{\sigma} + \frac{\partial\Theta}{\partial\sigma}. \tag{17}$$

3.8. Deep Gaussian processes (DGP)

The uncertainty in DL can be expressed based on the existence of infinite possible models that explain an observed data pattern. In defining BNN in Section 3.1, it has been said that uncertainty can be modelled around a series of neural network weights w via the prior $p(w)$ to arrive at a definition of the posterior $p(y|w)$. This Bayesian perspective implies that a restriction is applied on the infinite parameters w that a neural network can take to explain the observed data, and that leads to Equation 6. Let us consider another approach to define a Gaussian process [22]. For simplicity, consider a linear regression

$$f(x) = w^T\varphi(x) = f(x) + \epsilon, \tag{18}$$

where ϵ is white noise and φ are basis functions defining f:

$$\varphi(x) = (\varphi_1(x), ..., \varphi_D(x))^T. \tag{19}$$

In this simple financial regression with white noise, the observed prices are defined as a Gaussian distribution:

$$p(y|w, X, \lambda) = \mathcal{N}(\phi w, \lambda I), \tag{20}$$

where ϕ is defined as

$$\phi := \phi(x) = \begin{pmatrix} \varphi_1(x_1) & \cdots & \varphi_1(x_D) \\ \vdots & \ddots & \vdots \\ \varphi_D(x_1) & \cdots & \varphi_D(x_D) \end{pmatrix}. \tag{21}$$

Instead of defining the prior $p(w) = \mathcal{N}(0, S)$ as a Gaussian with mean 0 and covariance S, let us define a latent function $f = \phi w$ such that

$$p(f) = \mathcal{N}(0, K) \tag{22}$$

is Gaussian and

$$K = cov(f) = \mathbb{E}[\phi w w^T \phi^T] = \phi S \phi^T. \tag{23}$$

Instead of defining the group of base functions ϕ, one can choose $K = k(\cdot, \cdot)$ such that $\varphi(\cdot)$ is of infinite dimension. In this case, f can be defined as a Gaussian process (GP):

$$f(x) \sim GP(m(x), k(x, x'))\forall x, x', \tag{24}$$

where $m(x)$ is the mean of f

$$m(x) = \mathbb{E}[f(x)] \tag{25}$$

and $k(x, x')$ is a kernel function

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]. \tag{26}$$

For example, k can be chosen to be defined as follows (Gaussian kernel) to make φ infinite-dimensional:

$$k(x, x') = e^{-\frac{||x-x'||^2}{2\sigma^2}}. \tag{27}$$

3.9. Laplace approximations (LA)

The Laplace approximation uses the first-order Taylor series to fit a Gaussian distribution around the maximum a posteriori as the centre, defined as θ^*, to approximate the posterior distribution.

$$p(\theta|D) \approx p(\theta^*)exp[\frac{1}{2}(\theta - \theta^*)^T] \tag{28}$$

3.10. Deep ensembles

There are two types of deep ensembles: one consists of retraining the same DL model several times with different initial conditions and then taking the average of all these diverse behaviours [113]. These are called *boosting methods* and are adjusted sequentially. The second type of deep ensembles is called randomization-based because they use some scheme based on decision trees, for example, random forests. As a result of both approaches, one has a committee of models that hopefully focuses on different basins of the true posterior that represent the entire distribution through sampling. As the number of samples increases, the group of models collapses into a single model [57].

4. Results

4.1. Study categorization

In this section all the studies included in the survey are reported. The order presented in Table 2 is the same in which the studies were found and therefore included. The records have been divided into the five categories that have been considered most significant: *model*, *type of asset*, *analysis method*, *prediction space* and *epistemic/aleatory*. The *model* category includes the type of DL models used to represent uncertainty. They must fall into at least one of the categories described in Section 2.1.

The *type of asset* indicates which category the asset belongs to: stocks, Forex, or cryptocurrencies. A stock is also called *equity* and represents a small portion of ownership in a company. Forex is the abbreviated name for the foreign exchange market, where investors trade currencies. A similar concept is that of cryptocurrency, with the difference that it is virtual currencies secured by cryptography and impossible to counterfeit.

The *analysis method* comprises the subcategories fundamental, technical and sentiment. Investors perform fundamental analysis if they only look at economic and financial factors to estimate the value of an asset. Examples of fundamental factors are company balance sheets, gross domestic product, unemployment rate, politics or business management style. Technical analysis means estimating the future value of an asset solely by looking at past values or patterns of values. Sentiment analysis focuses on extracting information from the investor's will, expressed through text such as news, articles, social networks and blogs. With Natural Language Processing, sentiment analysis is gaining popularity and is being used more often in investments. These three analysis methods can be used alone or in combination.

The *prediction space* is the shape and characteristics of the output variable of the DL technique. If the output variable is just a value, then the space is a single-step prediction, and a multi-step prediction if the output variable is a vector of values. Furthermore, the prediction space can be positive, negative and neutral if the predictor output considers a rise, fall or stay in price. An even more interesting space from a UQ perspective are confidence intervals, associated with the probability that the variable of prediction is within the predicted range.

The category most associated with the definition of UQ is *epistemic/aleatory*. The uncertainty related to the DL model is called epistemic, that is, the variability in the data associated with the many different configurations that a model could take. You could choose a DL model with three, four, or ten hidden layers for the same data. Furthermore, the width of each layer can vary, or even the matrix of weights in a neural network can take an infinite number of different configurations. All this uncertainty makes the output variable more or less accurate. Epistemic uncertainty can be adjusted to become smaller (i.e. by narrowing the price probability distribution) if a better model can be found. By contrast, aleatory uncertainty is related to the intrinsic volatility of the data and is irreducible, which means that financial data always comes with noise. And by default, the noise level in the price of assets is high. This definition of randomness in the data expresses the concept of a latent price curve that the model can never know about, except through noise distorting its pure form. Interestingly, the idea of a pure price pattern that will never be known directly contradicts the Efficient Market theory.

Table 2: Studies included in the survey.

Ref.	Model	Asset	Analysis	Prediction	Uncertainty
[46]	BNN	CRY	T	SS	EPI
[28]	Bayesian regularized NN	STO	T	U	EPI
[13]	BNN + MCMC	STO	T	MVMS	EPI
[48]	Stochastic ANN + LSTM	CRY	T,S	SS	EPI
[88]	Hybrid model BST + LSTM	STO	T,S	TR	ALE
[86]	NARX BRANN	STO	T	SVMS	EPI
[39]	Bayesian LSTM	STO	T	SS	ALE
[70]	dropout Bayesian CNN	STO	T	SS	EPI

Continued on next page

Table 2 – *Continued from previous page*

Ref.	Model	Asset	Analysis	Prediction	Uncertainty
[101]	Levenberg-Marquardt BRANN	STO	T	SS	EPI
[84]	BNN + ARD and BNN + HMC	STO	T	SS	EPI
[71]	SVDBNN	STO	T	B/S	EPI
[106]	SSA-EWSVM-RNN-GPR DE	STO	T	CONF	EPI
[107]	VMD-AE-RNN-LSTM DE	STO	T	CONF	EPI
[72]	Dropout CNN	STO	T	TR	EPI
[81]	RBF, BNN and ARIMA DE	FOR	T	SVMS	EPI
[38]	BRANN	STO	T	MVMS	EPI
[47]	GBNN	STO	T	SVMS	EPI
[125]	Dropout CNN	STO	T	B/S	EPI
[35]	BRANN	FOR	T	B/S	EPI
[53]	Variational Autoencoders	FOR	T	CONF	EPI
[93]	BRANN	STO	T	U	EPI
[103]	Interval adapted LSTM AE	STO	T	SVMS, CONF	ALE
[12]	MC Dropout MLP	STO	T	SS	EPI
[8]	Dropout MLP	STO	T	SS	EPI
[5]	BRANN with NARX	STO	T	SS	EPI
[122]	BRANN	STO	T	SS	EPI
[27]	BRANN	STO	T	SS	EPI
[99]	BRANN	STO	T	SS	EPI
[87]	BRANN	CRY	T	U	EPI
[7]	CAED-TCN	STO	T	MSMV, CONF	EPI
[73]	Backprop BRANN	STO	T	TR	EPI
[98]	MCMC BRANN	STO	T	TR	EPI
[108]	VAE	STO	T	SS	EPI
[31]	Attention based VAE-LSTM	STO	T	TR	EPI
[37]	VAE-GCN-LSTM	STO	T	SS, TR	EPI
[112]	MDN	STO	T	SS	ALE
[14]	MC Dropout LSTM	STO	T, F	CONF	EPI
[96]	Noise quantification MLP	STO	T	SS	ALE
[111]	MDN	STO	T, S	CONF	ALE

Continued on next page

Table 2 – *Continued from previous page*

Ref.	Model	Asset	Analysis	Prediction	Uncertainty
[16]	LSTM + MC Dropout	CRY	T, S	S	EPI
[61]	BRANN via Variational Inference	STO	T	TR	EPI
[118]	VAE	STO	T	TR	EPI
[26]	VAE	CRY	T, S	TR	EPI
[116]	VAE RNN	STO	T, S	SS	ALE
[34]	VAE + VI Neural Network	STO	T, S	SS	EPI
[19]	BNN	CRY	T	SS	EPI
[59]	BNN	CRY	T	SVMS	EPI
[20]	recurrent BNN	STO	T	MVMS	EPI
[119]	BNN	STO	T	SS	EPI
[50]	Stochastic ANN	CRY	T, S	SS, B/S	EPI
[65]	dropout RNN	CRY	T	SS	EPI
[83]	dropout ANN	STO	T	SS	EPI
[89]	Dropout LSTM	STO	T	SS	EPI
[64]	Dropout LSTM	STO	T	SS	EPI
[124]	Dropout CNN+ LSTM	CRY	T	SS	EPI
[78]	Dropout LSTM	STO	T	SS	EPI
[110]	Dropout LSTM	STO	T	SS	EPI
[42]	Dropout MLP	STO	S	SVMS	EPI
[43]	Dropout MLP	STO	S	SVMS	EPI
[66]	NSVM (VI)	STO	T	SS	EPI
[18]	VAE-LSTM with dropout	STO	T	MSMV	ALE
[58]	dropout VAE-GRU	FUT	T	SS	ALE
[4]	VAE-LSTM	STO	T	SVMS	ALE
[77]	PAE-XNMF	STO	T	SS	ALE
[109]	Heteroskedastic DGP	STO	T	U	EPI
[104]	BPNN, LSTM, GPR, Lasso, BILSTM ensembles	STO	T	CONF	EPI
[105]	Deep ensembles LigthGBM	FUT	T	SS	EPI
[120]	A2C, DDPG, PPO ensembles	STO	T	B/S	EPI
[51]	RNN, GRU, LSTM ensembles	IND	T	SS	EPI

Continued on next page

Table 2 – *Continued from previous page*

Ref.	Model	Asset	Analysis	Prediction	Uncertainty
	T: technical analysis S: sentiment analysis F: fundamental analysis CRY: cryptocurrencies STO: Stock market FOR: Forex market FUT: Futures market IND: indexes market SS: Single step prediction MVMS: Multi-value multi-step ahead prediction SVMS: Single value multi-step ahead prediction CONF: Confidence interval B/S: buy, sell or hold signals TR: up, down or stay trends ALE: aleatory uncertainty analysis EPI: epistemic uncertainty analysis U: unspecified				

4.1.1. Records focused on aleatory uncertainty

Ray et al. [88] applied MCMC to sample the posterior distribution of forecast prices in a hybrid LSTM model with Bayesian structural time series (LSTM-BST). The authors included a sentiment analysis on Twitter and applied a confidence interval that they averaged to show the most likely value for the next price. Huang et al. [39] implemented a Bayes-LSTM model, where Bayes optimization was used for the dynamic selection of the window size. This work can be considered as a particular case of UQ, in which there is a probability distribution of ways in which the input data can be divided. Thus, the authors recover the optimal division of the most probable value. One can think of that model as an ensemble of different models with different prediction capabilities. However, in this case, the authors only focus on the optimal value and not a confidence interval. This is an interesting case of aleatory uncertainty analysis, because the historical window used for the prediction is dynamic and depends on the variance and changing periods of the highly volatile stock market.

In the work of Wang et al. [103], it is stated that the quantification methods of uncertainty in DL have some limitations. Instead, an approach that does not require a prior assumption is proposed, a hybrid loss function is proposed that improves accuracy and stability, and finally, training is done using backpropagation. As a result, although the model provides a confidence interval, it does not require probability distributions and is easier to train. The uncertainty range in this case is derived directly from the volatility of the data, rather than from the probabilistic variations of the model hyperparameters. Wilkins et al. [112] uses mixture density networks (MDN) to explicitly quantify uncertainty in large and complex data sets. The model is claimed to be superior in performance to probabilistic backpropagation, MC dropout, and deep ensembles. It directly addresses aleatory uncertainty by applying softplus to predict the current standard deviation and a regressor network to predict the mean. Another work by the same authors [111], infinite mixture density networks (iMDN) and Gaussian networks (GN) are implemented to predict both the price and the uncertainty of various actions. The authors state that these methods outperform the reference methods. In a similar way to the previous paper, the work captures the aleatory uncertainty through the regression of the mean and the standard deviation using iMDN.

Shen and Song [96] proposed a way to directly quantify noise in data sets in which a neural network classifies the amount of noise. That amount is then passed as a multiplier to the loss function to linearly ignore noise-bound data and pay attention to clean data. The authors claim a 5% increase in annual returns when investing in the stock market.

The framework proposed in the article by Choudhury et al. [18] implements a regularized VAE dropout framework to remove noise and a stacked LSTM autoencoder to predict the next step stock price. Li et al. [58] established a multi-modal VAE approach to extract high-level deep structures from unsupervised futures price data.

4.1.2. Records focused on epistemic uncertainty

Jay et al. citeJAY-2020 described how to implement a stochastic LSTM in which activation functions are applied with a stochastic mechanism, instead of the traditional probabilistic weight matrix approach. This is done because the probabilistic activation strategy can be interpreted as random changes in features. On the contrary, the authors state that the probabilistic weights strategy might not be ideal because feature detection can become noisy as the network evolves and this can cause dependencies to be forgotten. Jang and Lee [46] modelled a cross-validated Bayesian regularized neural network (BRANN) to predict the next step of Bitcoin, using Blockchain structures as extracted features. It also includes the quantification of logarithmic volatility, stating that the predictive performance of BRANN is superior to other reference methods.

To the best of our knowledge, the article by Gençay and Qi [28] is the first work that explicitly describes the BRANN mechanism and explains its ideas in depth. It is based on the work of Mackay in 1992 [68] although he called the methodology *Bayesian interpolation*. Gençay and QI applied a BRANN model to the S&P 500 index and compared the model with other regularization methods, such as early stopping and bagging, and concluded that bagging presents more accurate results.

Chandra [13] made explicit use of UQ methods to predict the price five steps ahead with a 95% confidence interval. He implemented Langevin gradients with parallel tempering MCMC and compared the results with a forward NN with Adam's optimizer and stochastic gradient descent. In the article by Primasiwi et al. [86], a non-linear autoregression with exogenous input (NARX) version of a BRANN was used to predict stock prices, indicating that the NARX model is more accurate than the multi-layer perceptron (MLP). Maeda et al. [70] tested a Bayesian convolutional neural network (CNN) that uses variational dropout as a method to perform Bayesian inference against other non-Bayesian CNN and demonstrates better accuracy and higher final capital. Ticknor [101] combined a BRANN model with a Levenberg-Marquardt algorithm to forecast the movement of stock prices, although its goal was to reduce the potential for overfitting and find local minima.

Pires and Marwala [84] also attempted to reduce the overfitting problem (without uncertainty quantification search) when they implemented a Bayesian neural network to predict upcoming stock market prices and sample the subsequent one with automatic relevance determination (ARD) and hybrid Monte Carlo (HMC). However, Maeda et al. [71] made conscious use of predictive uncertainty when using variational decay as a Bayesian approximation and optimizing BNN parameters with stochastic gradient variational Bayesian to make buy/sell decisions. Wang et al. [106] proposed a deep nonlinear ensemble framework for stock index forecasting and Gaussian regression (GPR) process to obtain interval forecasting. In a subsequent work [107], the same authors build a multi-scale ensemble model based on a LSTM neural network again using GPR for the construction of confidence intervals.

Maeda et al. [72] proposed a Bayesian CNN method to predict short-term stock price trends using stock order data. They used a CNN to learn the series characteristics of the stock order data and a sparse variational dropout technique to enable Bayesian inference (SVDBNN). Instead of the traditional one-step spatial prediction, they apply a trend classification approach (ascending/descending classifications). The app has been trading on the Tokyo Stock Exchange and has proven more efficient and safer investments than alternatives.

Pandey and Jagadev [81] described a deep ensemble system consisting of a nonlinear parallel combination of several neural networks, each of which predicts the rate of change. The combination of the three models provides better results than each one separately. Housein et al. [38] compared the performance of three autoregressive nonlinear NNs trained using Bayesian regularization, stochastic gradient descent, and Levenberg-Marquardt training mechanisms. Jang and Lee [47] created a generative Bayesian neural network to overcome the problem of traditional BNN that does not consistently estimate price at the extremes of time series due to missing data in those regions. Zhang et al. [125] explained the implementation of a hybrid model made of a CNN to filter limit order books, another CNN to capture non-linear price patterns, and an LSTM module to capture long-term patterns. Variational dropout applies to approximate Bayesian inference.

Hassanniakalager et al. [35] implemented a BRANN to make buy/sell/hold decisions based on Forex price data. They discovered that the Bayesian application increases yields 5% more than other reference techniques. Kim [53] used VAE to obtain noise-free series from original series to train deep neural networks. In the study, a collaborative learning approach is taken to design confidence bands from which to extract the mean and standard deviation. Selvamuthu et al. [93] He disassembled a BRANN and compared its accuracy with other training mechanisms such as Levenberg-Marquardt and Scale Conjugate Gradient. BRANN provides the most accurate results. Calvo-Pardo et al. [12] conducted a literature review on machine learning for empirical finance, presenting some examples of BRANN models to reduce overfitting.

Back and Keith [8] studied uncertainty estimates given by a Bayesian framework, represented by Monte Carlo dropout, variational inference, and MCMC. They compare Bayesian with non-Bayesian methods and conclude that Bayesians generalize better. Al-Shayea [5] compared Levenberg-Marquart, scaled conjugate gradient, and Bayesian regularization using NARX in Dow Jones index forecasting, and concluded that the Levenberg-Marquardt algorithm achieves the best performance. In the article by Hou and Liu [122], the prediction of the Chinese Bao Gang was analysed using the Bayesian regularization optimization algorithm. The goal of the Bayesian framework was to reduce overfitting. The Bayesian framework predicted better than traditional backpropagation algorithms.

Garg [27] applied a Bayesian neural network for gold price prediction because it was found to be suitable for time series data and does not depend on any historical features. However, the frame was not compared to any other model. Garg concluded by stating that BRANN's accuracy was very high. Similarly, Sun et al. [99] proposed a BRNN because it can predict new data to a greater extent, compared to the traditional backpropagated NN. Also in the work by Priya and Garg [87], the accuracy of a BRANN model was compared with that of a Levenberg-Marquardt in the prediction of various cryptocurrencies, showing that BRANN had a lower mean percentage error.

Alghamdi et al. [7] developed the architecture of an encoder-decoder model capable of receiving exogenous factors with an attention mechanism. To account for uncertainty, the model displays the Monte Carlo dropout between each CNN layer. For each stock price value given as input, the model predicts a thousand one-step forward forecasts. This makes it possible to generate confidence intervals of 90% and 95%, which represent the degree of ignorance of the model. Magris et al. MAGRIS-2022 proposed a Bayesian version of a temporal attention augmented bilinear neural network and compared its performance with non-Bayesian versions. They implement a subsequent dropout approach via Montecarlo applied to limit order books.

Skabar [98] studied the performance of a MCMC trained through the Metropolis Hastings algorithm and stated that it has superior performance compared to an MLP, in addition to reduced overfitting. Wang [108] explored a daily performance prediction model with variational autoencoders to find minimum variance portfolios. In the same work you can find a study on the uncertainty related to the time to execute a limit order in time, as well as the optimal stop, using VAE to make those decisions. Gunduz [31] used VAE to choose the features that will feed an LSTM price prediction model in Borsa Istanbul. This model was compared with Support Vector Machines and LightGBM. The attention-based VAE-LSTM excelled in accuracy compared to the other reference models.

In the work of Hou et al. [37], the authors proposed a VAE discriminant to know the latent characteristics of lesser dimension among the company's actions. A later hybrid GCN-LSTM system modeled graphical interactions between companies, as well as price fluctuations on the timeline. Such a model outperformed other benchmark models that predict stock prices. Chauhan et al. [14] followed an investment strategy in which current prices are related to future fundamentals, rather than current reports. To this end, the authors implemented uncertainty estimates from both neural heteroscedastic regressions and a dropout-based heuristic. In Cheuque's master's thesis [16], the MC drop in an LSTM model was used to predict the next Bitcoin price. The author also included an analysis of Twitter messages as a way to predict the Bitcoin price, comparing the accuracy of just feeding the model with price data. The conclusion was that sentiment analysis helps drive prediction beyond time.

Lind et al. [61] used a BNN to perform explicit quantification of uncertainty. VI was applied to train the BNN, using Bayes from backprop to do the reparameterization trick. Xu and Cohen [118] studied a hybrid approach where they learned from price signals and sentiment to predict the next stock price. A VAE was built as a prediction model. Garg et al. [26] also used a VAE model to test a dataset in which Twitter messages were related to price movements in the top 12 cryptocurrencies. Xing et al. [116] modelled the interaction of stock market sentiment with its volatility by means of a VAE-based recurrent neural network. Haq et al. [34] compared support vector machine (SVM), linear regression and random forest as models to extract a feature vector of NASDAQ parameters, and then a VAE model was implemented for next price step prediction.

Coco et al. [19] analysed the performance of a BRANN model to predict the price of Bitcoin. The predictive distribution was sampled from a mixed distribution. The predictive performance of BNN is inferior to that achieved by support vector regression (SVR) + feed forward NN or SVR + LSTM NN. In the work of Lin and Blum [59], a BRANN framework was implemented to capture future price trends, and as a result, it was observed that it can increase the dividend yield by 43% compared to the CNN and MLP frameworks. Dixon [20] created a recursive BNN to predict the values of IBM stock prices in multiple steps and exponentially smoothed. The posterior distribution was approximated by variational inference. They demonstrated how 95% confidence intervals can be applied to a recurrent network by integrating them into a Bayesian framework. Yan et al. [119] proposes a framework in which a BRANN model is coupled in a series circuit with a particle swarm optimization (PSO) algorithm that continues to update the weight matrix. This framework reduced the MAPE error from 0.85% to 0.77%.

Kalariya et al. [50] proposed a stochastic neural network in which stochasticity was applied to activation functions instead of the elements of a weight matrix to predict cryptocurrencies. They discussed some investment strategies like buy and hold, various oscillators or Bollinger bands. Stochastic neural networks outperformed traditional strategies. Livieris et al. [65] explored stock price prediction and claimed that using an MC dropout-based ANN improves performance because the model explores regions of weight spaces that traditional training algorithms do not reach. Similarly, Patel et al. [83] discovered how a dropout regularization approach is demonstrating a positive consistency delta of 2.11% compared to the same ANN without dropout and stochastic gradient descent, applied to stock market forecasting.

The approach proposed by Rokhsatyazdi et al. [89] performed better than SARIMA, ETS and NAIVE in stock market prediction using a dropout regularized LSTM model. Liu et al. [62] explained how one can use a LSTM model whose hyperparameters are optimized with PSO and the weights are dropout regularized to predict the next stock price. A regularized churn hybrid CNN-LSTM model was explored to predict Bitcoin and Ethereum, showing that it outperforms SVR, LSTM, CNN, MLP, and ARIMA, among other models. Similarly, the articles by Naik and Mohan[78] and Wang et al. [110] analysed the performance of defection regularized LSTM models to predict the next stock market price. Both concluded that the regularized model works better than other reference methods without regularization. Two papers [42, 43] proposed similar approaches in which a regularized multilayer dropout perceptron predicts the next stock market price based on company financial reports, concluding that there is a strong correlation between abnormal returns and textual information from annual reports. returns.

The focus of Livieris's work [66] was to reformulate time-based volatility in stochastic neural networks. The authors call it the neural stochastic volatility model (NSVM), which consists of a hybrid MLP + recurrent neural network (RNN) scheme in which the RNN module finds data patterns and the MLP predicts the mean and variance through variational inference. Abrahami et al. [4] addressed the very high frequency trading level and created a VAE-LSTM autoencoder to denoise the data and a stacked LSTM autoencoder to predict the next seven minute tick price of various Actions. In the article by Montesdeoca et al. [77], the authors took a particular approach when using a VAE model, so they used a non-negative probabilistic matrix refactoring with exogenous information to drive the VAE in financial data modelling. Wang et al. [109] used a deep Gaussian process to sequentially infer predictions from data by estimating states recursively and updating the model with the feature that more complex and non-linear dependencies can be learned compared to traditional Gaussian processes.

Wang et al. [104] focused on quantifying the price of crude oil by building confidence intervals. To do so, a deep set made of a backpropagated NN, LSTM, GPR, and Lasso operator was constructed. [105] [105] [105] was also proposed. In this work, LightGBM was used to integrate the set into a final price result. The framework showed better accuracy than other reference models such as MLP, LSTM, and others. Yang et al. [120] implemented another deep set approach that integrates proximal policy optimization (PPO), actor advantage critic (A2C) and deep deterministic policy gradient (DDPG) agents to learn equity trading strategy , which shows that the integrated model outperformed the individual components. Finally, Kamal et al. [51] predicted the Baltic Dry index with a deep ensemble made of GRU, LSTM and RNN models, showing that it outperforms ARIMA, MLP, RNN, LSTM tested separately.

It is not uncommon for the studies listed in Table 2 not to explicitly describe all the parameters that define the models they used. For example, the number of steps predicted in the future is not always specified, as in the case of the work by Gençay and Qi [28]. Possibly, a single-step prediction could have been inferred, but instead, and out of caution, the "unspecified" key has been added.

Table 3 lists the journals, proceedings, and publishers of each of the records described in Table 2. In summary, 25 studies were published as proceedings in congresses or conferences, 42 studies in journals and 2 records as master's thesis or Ph.D. thesis reports.

Table 3: Studies ordered by publisher, journal or conference paper.

Ref.	Journal / Conference	Publisher
[46]	Access	IEEE
[28]	Transactions on neural network	IEEE
[13]	Plos One	Public Library of science
[48]	Access	IEEE
[88]	Transactions on computational social systems	IEEE
[86]	Conference ICTS 2019	IEEE
[39]	Conference ICMLC 2018	Association for Computing Machinery
[70]	Conference IIAI-AAI 2019	IEEE
[101]	Expert systems with applications	Elsevier
[84]	Arxiv	Arxiv
[71]	International Journal of Smart Computing and Artificial Intelligence	IIAI
[106]	Cognitive Computation	Springer
[107]	Applied soft computing	Elsevier
[72]	Conference AISC 2020	Elsevier
[81]	Journal of King Saud University - Computer and Information Sciences	Elsevier
[38]	Neural computing and applications	Springer
[47]	Quantitative finance	Taylor and Francis Online
[125]	Conference Neurips 2018	Arxiv
[35]	Journal of empirical finance	Elsevier
[53]	SSRN	Elsevier
[93]	Financial Innovation	Springer
[103]	Neurocomputing	Science Direct
[12]	Journal of Risk and financial Management	MDPI
[8]	Plos One	Public library of science
[5]	Conference WCECS 2017	IAENG

Continued on next page

Table 3 – *Continued from previous page*

Ref.	Journal / Conference	Publisher
[122]	International Journal of Performability Engineering	Totem
[27]	Conference NCCS 2019	Springer
[99]	Conference ISKE 2012	Springer
[87]	Conference SCI 2018	Springer
[7]	Conference IJCNN 2021	IEEEXplore
[73]	Arxiv	Arxiv
[98]	Advances in Electrical Engineering and Computational Science	Springer
[108]	Academic Commons	University of Columbia
[31]	Financial Innovation	Springer
[37]	IEEE/CAA Journal of Automatica Sinica	IEEEXplore
[112]	Arxiv	Arxiv
[14]	Conference PMLR 2020	Arxiv
[96]	Conference ICAI 2018	Springer
[111]	Conference PMLR 2015	JMLR
[16]	Repository	Pontificia Universidad Católica de Chile
[61]	Repository	Uppsala Universitet
[118]	Conference 56th Annual Meeting of the Association for Computational Linguistics	ACL
[26]	Conference ICMLA 2021	IEEEXplore
[116]	Knowledge-based systems	Science Direct
[34]	Expert Systems with Applications	Science Direct
[19]	Computer science	PeerJ
[59]	Conference Spring Simmulation 2020	IEEEXplore
[20]	Technometricss	TANDF
[119]	International Journal of Production Research	TANDF
[50]	Mathmatics	MDPI
[65]	Evolving Systems	Springer
[83]	Conference ICDAM	Springer
[89]	Conference CEC 2020	IEEEXplore
[64]	Digital Signal Processing	Science Direct
[124]	Conference BlockSys 2022	Springer
[78]	Conference EANN 2019	Springer
[110]	Conference DCABES 2018	IEEEXplore
[42]	Neural Cmputing and Applications	Springer
[43]	Neural Cmputing and Applications	Springer
[66]	2018 Conference AAAI 2018	Arxiv
[18]	AI Communications	IOS Press
[58]	Access	IEEEXplore

Continued on next page

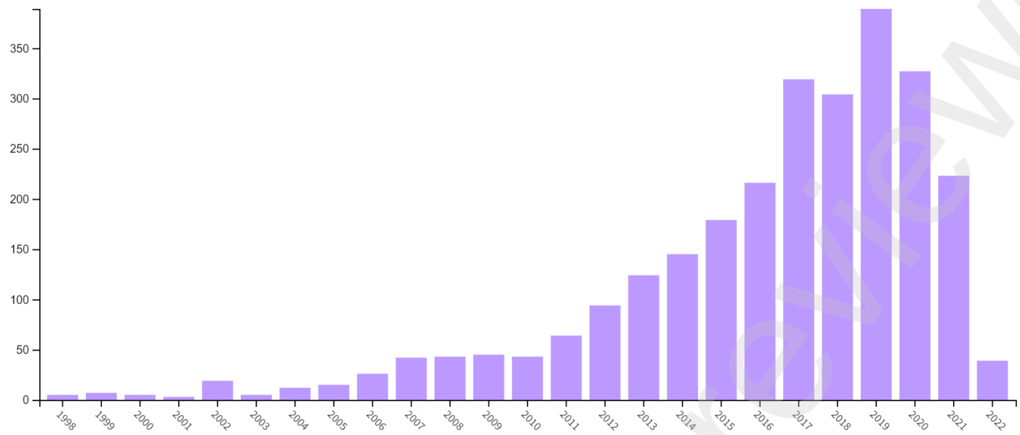

Figure 6: Number of publications per year related to UQ applied to business economics. Source: Web of Science.

Table 3 – *Continued from previous page*

Ref.	Journal / Conference	Publisher
[4]	Conference ICTAI 2019	IEEEXplore
[77]	Conference ISPA 2019	IEEEXplore
[109]	Conference PMLR 2016	JMLR
[104]	Applied Soft Computing Journal	Elsevier
[105]	Digital Signal Processing	Elsevier
[120]	Conference ICONIP 2020	Elsevier
[51]	Applied Sciences	MDPI

5. Discussion

In this section, the results related to the research questions formulated in Section 2 are discussed.

What is the state of the art related to the research in UQ for DL applied to finance?

As can be seen in Figure 6, the number of publications related to UQ applied to business economics has been growing very significantly since 2010. The peak can be found in 2019, with 389 publications. Despite this huge increase, the popularity of the term as applied to economics appears to have declined since then, possibly due to a migration of research to more popular areas such as medicine, microbiology, or infectious diseases after the global Covid-19 pandemic. This effect has been seen on the Web of Science looking for those other topics: the number of investigations in those other areas has grown considerably since then. In fact, the term UQ, applied to all possible domains, continues to increase in popularity reaching 2578 publications in 2021 (see Figure 1).

China, the USA and India are the three countries that publish the most papers on the subject, with 16, 12 and 10 articles, respectively. Figure 7 shows the complete list of countries and their respective number of publications. On the other hand, there seems

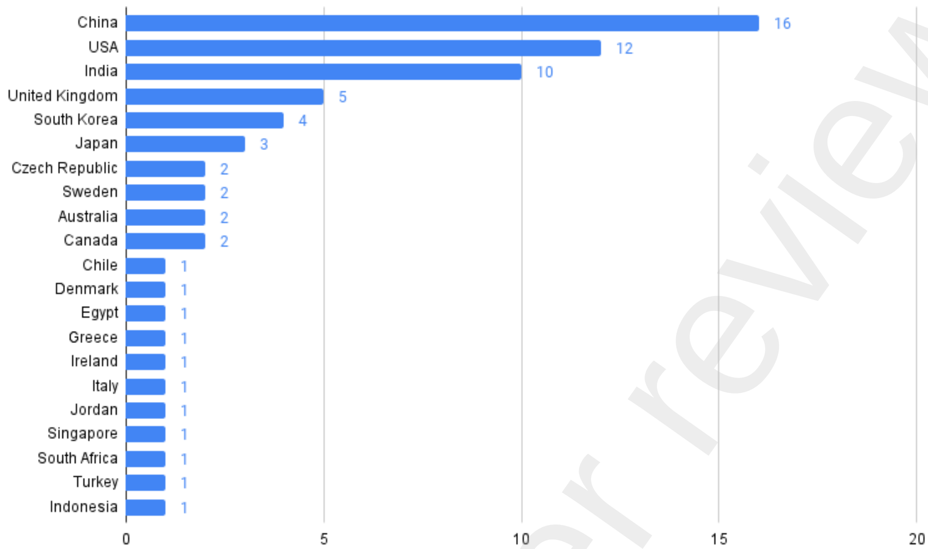

Figure 7: Number of publications by country.

to be a preference for stock price prediction, as can be seen in Figure 8; this effect may also be influenced by the specific existence of the keyword *stock* in search queries.

Furthermore, it can be seen in Table 2 how researchers focus mainly on epistemic uncertainty (57 records), and only 12 are based on aleatory uncertainty. No records have been found that explicitly use epistemic and aleatory uncertainty in combination. As summarized in Figure 9, a large portion of the records (57) look for patterns in historical data (technical analysis), seven combine a technical and sentiment analysis approach, three of them only a sentiment analysis and two mix a technical and fundamental approach. The authors of this survey did not find any papers that use all three methodologies, technical, sentiment and fundamental at the same time, and this could inspire researchers to explore this avenue in the future.

Something that could potentially come as a priori surprise is that most of the articles focus on a one-step approach (Figure 10), although they use UQ methods to predict, and can take advantage of interval prediction spatial representations. Authors are already benefiting from more informationally efficient models, i.e. UQ methods, over frequentists, and have a distribution of information at hand, rather than a single point. One might wonder why the authors settle for a mean value instead of using a more robust prediction space, since they have already computed the space. In other words: why use optimization methods after finding the uncertainty space? There are possibly many answers to those questions. First of all, some authors, as we discussed earlier, are not intentionally using UQ methods, but using UQ as a way to generalize, as in the case of dropout. Another reason could be that not many authors use confidence intervals in the prediction (only eight), which restricts the number of records to compare their work. In any case, it

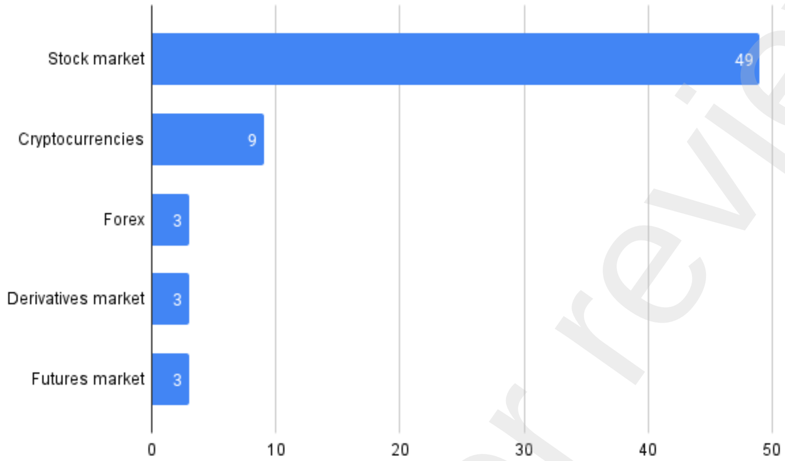

Figure 8: Number of publications by type of financial asset.

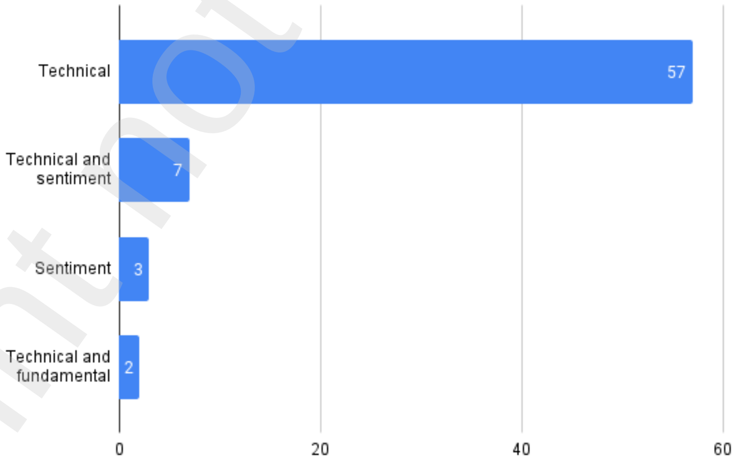

Figure 9: Number of publications by type of analysis.

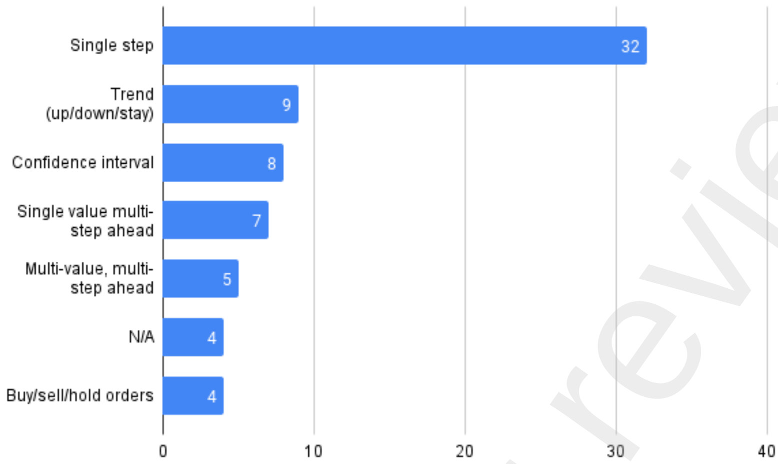

Figure 10: Number of publications by type of predicted space.

appears that interval forecasting is hardly explored in financial time series and could be another potential avenue of investigation for future work.

What are the main UQ techniques in DL used so far for financial prediction?

The main techniques of UQ in DL have been described in Section 3: Bayesian neural networks and some associated methods to approximate the true posterior, i.e. Montecarlo dropout or Markov Chain Montecarlo or variational inference. Also variational autoencoders, Bayesian active learning, Bayes by backprop, deep Gaussian processes, Laplace approximation or deep ensembles. We have seen that Bayesian active learning and Bayes by backprop have not been sufficiently explored and no records of these have been found, which may be an interesting field for future research.

What needs in financial forecasting are not yet covered in UQ for DL?

The initial hypothesis that triggered the idea of carrying out this work is that financial investment needs information (the most probable future value) and quantifying the variance related to that information to minimize risk. If the variance is wide and the probability distribution is flat, then the most likely future price is not really valuable. In these situations, an investor should decide to hold trades until the variance is low and the probability distribution is thinner. A similar situation would be found if the predictive probability distribution were modal, which is synonymous with having peaks and valleys. The investor would probably be confused and not know what decision to make. The authors of this work believe that further research is needed to understand the implications of the shape of the variance in order to relate the shapes to investment decisions and profit.

An overwhelming majority of studies focus on the prediction of the single next step (34 out of 69). Although this approach might make sense on a large time scale, such as weeks or months, it does not represent actual needs of investment. As stated earlier, financial time series have a strong component of noise, so predicting a single next point does not add much value if the delta price is less than the variance, because all it represents is noise. A more valuable approach would be to predict a trend or even represent a vector of future values from which to infer a trend. However, only eight articles predict trends and five work with multi-valued prediction vectors. A better approach would be to relate buy/sell/hold positions to UQ. Although four studies focus on the buy/sell/hold approach, none see a clear relationship with UQ. Consequently, we believe that there are many opportunities to extend the work related to the prediction vector.

A good investor considers all three sources of information available: technical, fundamental, and sentiment. This survey has shown how researchers are highly attracted to technical analysis (57 records relate to purely technical approaches), but not so much to sentiment or fundamentals (see figure 9). Only two publications use the fundamental approach, combined with the technical one. Ten records focus on sentiment analysis combined with a technical approach. Here we find an interesting research opportunity, combining different UQ methods, one for each of the three approaches. Some questions remain for this future research, such as what is the optimal weight of each approach in the final prediction and under what conditions or applied to what assets.

The foreign exchange market is by far the largest investment market, the volume of which is 700 billion US dollars per day. This is compared and contrasted with smaller markets such as stocks (200 billion US dollars) or futures (30 billion US dollars) [45]. However, only three records related to the prediction of international currencies have been found. The Forex market offers some advantages in some areas, such as the amount of time that the Forex market is open for trading. From the Monday open in Australia to the Friday close in New York, there are five days of uninterrupted trading 24 hours a day. That positively impacts the amount of data available. Also, there are no commissions and transaction costs are low. Due to the size of the market, it is always liquid and trades are executed instantly [2]. This market is less popular for researchers, but shows potential for further investigation.

Epistemic uncertainty is related to the lack of knowledge in the model. The studies found focus on the variability related to the weights of the neural networks, but other parameters could also be brought to light in the evaluation. A researcher is deciding, as empirically as possible, the architecture of the network, but that decision is leaving aside other possible configurations that are part of the epistemic uncertainty. Gençay [28] considers exactly this aspect when studying different network architecture configurations and potentially more could be carried out in future studies, possibly varying the length and width of the neural network, choosing different activation functions, different training iterations or batches.

One of the most important aspects that the authors of this work have identified is that all studies are based on a limited amount of data and are trained accordingly, although real-world and professional settings use up-to-date transmission models throughout their operation. Real models continually adapt training parameters to new and unseen data, which certainly has an impact on scaling the training data. Possible research on streaming data could explore an unsupervised approach to price prediction.

Shen and Song in their article [96] say that "too many studies focus on model uncertainty, but too few focus on data uncertainty". The typical study that includes uncertainty in the prediction mixes both types of uncertainty in such a way that they cannot be differentiated. Hüllermeier [44] describes this as a necessary discrimination between a predicted probability score and its related prediction uncertainty. One, for example, can be very sure that tomorrow it will not rain with a 10% probability, but very unsure that the probability of rain next week is 50%. This uncertainty comes from a lack of knowledge, the opposite of the probability confidence interval that comes from noise in the data. Indeed, future studies may further explore the combination of epistemic and aleatory uncertainty.

What technical challenges remain in financial forecasting using UQ for DL?

The technical challenge most mentioned by almost all authors is to achieve valuable predictions, considering that the data available in financial time series are extremely noisy, non-stationary, have a poor signal-to-noise ratio, and are especially sensitive to external and not completely known factors [7, 39, 70, 71, 103, 106, 118]. Magris et al. [73] considers that the inability of DL methods to cope with uncertainties in financial applications is a major drawback and major concern in econometrics. Due to this dynamic behaviour, the authors believe that the point estimate is not optimal for predicting financial time series. Instead, UQ is a much better approach to trend forecasting, offering plenty of opportunity to explore further. Wilkins [112] states that some UQ methods have been proposed in recent years; however, they fail when applied to large and noisy data sets. Indeed, there is a limit to decrease the total uncertainty, defined by the aleatory uncertainty. Even if we hypothetically reduce epistemic uncertainty to zero, there will always be uncertainty intrinsic to the data. Researchers usually try to extract the function hidden under all the stochasticity, however, it is very common to find in the literature that the chosen feature space does not capture all the information that it could from the stochastic data. As a consequence, the prediction of a target, i.e. price, trend change or buy/sell orders, is not optimal [112].

Furthermore, the true posterior distribution in a Bayesian neural network cannot be determined analytically (without an unacceptable amount of computing power). In fact, almost all of the probabilistic methods for UQ prediction mentioned in this paper are based on approximating the posterior distribution using much less computing power, in exchange for giving up some accuracy.

What approaches could potentially be explored to overcome those challenges?

Some of the possible approaches to overcome those challenges have already been mentioned above, which can be summarized in the following list:

- Increase the efficiency of extracting information from the feature space.
- Explore a more meaningful feature space that includes fundamental, opinion and technical information in a combined way.
- Develop new methods to represent uncertainty in a richer way.
- Delve into trend forecasting as an alternative to point estimation.
- Consider alternative markets to stocks such as Forex, cryptocurrencies, futures or derivatives.

- Explore newer methods for approximating the true posterior distribution.

6. Conclusion

In this survey, a PRISMA approach was followed to review 69 records on UQ for DL applied to financial forecasting time series. We have analysed the most distinctive aspects of those studies such as the type of asset, the techniques used, the forecast space, the method of analysis and the epistemic vs. aleatory approaches. We have seen that there are potential areas that are not sufficiently explored in the literature, such as combining fundamental, sentiment and technical analysis, further exploring the application to the Forex market, and finding a better way to address aleatory uncertainty. As a conclusion of this survey, we can state that there is a lot of room for future research at UQ for financial time series forecasting.

References

[1] Investopedia - noise, 2022 - https://www.investopedia.com/terms/n/noise.asp.

[2] Preschool - school of pipsology, 2022 - https://www.babypips.com/learn/forex/preschool.

[3] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. 11 2020.

[4] Soheila Abrishami, Michael Turek, Ahana Roy Choudhury, and Piyush Kumar. Enhancing profit by predicting stock prices using deep neural networks. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2019-November:1551–1556, 11 2019.

[5] Qeethara Al-Shayea. Neural networks to predict stock market price. 10 2017.

[6] Ismail Alarab, Simant Prakoonwit, and Mohamed Ikbal Nacer. Illustrative discussion of mc-dropout in general dataset: Uncertainty estimation in bitcoin. *Neural Processing Letters*, 53, 04 2021.

[7] Dhaifallah Alghamdi, Faris Alotaibi, and Jayant Rajgopal. A novel hybrid deep learning model for stock price forecasting. *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021.

[8] Alexander Back and William Keith. Bayesian neural networks for financial asset forecasting. Master's thesis, KTH, Mathematical Statistics, 2019.

[9] David M Blei, Alp Kucukelbir, and Jon D Mcauliffe. Variational inference: A review for statisticians. 2018.

[10] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, Wierstra@google Com, and Google Deepmind. Weight uncertainty in neural networks. 2015.

[11] by Parisa Tabassum and Supervisor Mahbub Alam Majumdar. Stock price forecasting using bayesian network, 2018.

[12] Hector F. Calvo-Pardo, Tullio Mancini, and Jose Olmo. Neural network models for empirical finance. *Journal of Risk and Financial Management*, 13(11), 2020.

[13] He Y Chandra R. Bayesian neural networks for stock price forecasting before and during covid-19 pandemic. *Plos One*, July 2021.

[14] Lakshay Chauhan, John Alberg, and Zachary C Lipton. Uncertainty-aware lookahead factor models for quantitative investing. 2020.

[15] Hsin-Yi Chen and Szu-Hao Huang. Generating a trading strategy in the financial market from sensitive expert data based on the privacy-preserving generative adversarial imitation network. *Neurocomputing*, 500:616–631, 8 2022.

[16] Cheuque. Bitcoin price prediction trough stimulus analysis: On the footprints of twitter's crypto-influencers, 2021.

[17] Siddhartha Chib and ; Edward Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49:327–335, 1995.

[18] Ahana Roy Choudhury, Soheila Abrishami, Michael Turek, and Piyush Kumar. Enhancing profit from stock transactions using neural networks. *AI Communications*, 33:75–92, 2020.

[19] Luisanna Cocco, Roberto Tonelli, and Michele Marchesi. Predictions of bitcoin prices through machine learning based frameworks. *PeerJ Computer Science*, 7:e413, 03 2021.

[20] Matthew Dixon. Industrial forecasting with exponentially smoothed recurrent neural networks. *Technometrics*, 64:114–124, 2022.

[21] Mihai Dogariu, Liviu daniel Ştefan, Bogdan Andrei Boteanu, Claudiu Lamba, Bomi Kim, Liviu-Daniel Ştefan, Bogdan Ionescu, L d Ştefan, B A Boteanu, C Lamba, and B Kim. Generation of realistic synthetic financial time-series. *ACM Trans. Multimedia Comput. Commun. Appl*, 18:96, 2022.

[22] Maurizio Filippone. Deep gaussian processes. Bayes Group, 8 2019.

[23] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *Proceedings of The 33rd International Conference on Machine Learning*, 06 2015.

[24] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. 2017.

[25] Akilesh Ganti, Gordon Scott, and Yarilet Pérez. Asset class definition, 2022.

[26] Amish Garg, Tanav Shah, Vinay Kumar Jain, and Raksha Sharma. Cryptop12: A dataset for cryptocurrency price movement prediction from tweets and historical prices. 2021.

[27] Shruti Garg. Forecasting of gold prices using bayesian regularization neural network. 2020.

[28] Ramazan Gençay and Min Qi. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *Neural Networks, IEEE Transactions on*, 12:726 – 734, 08 2001.

[29] Partha Ghosh, Mehdi S M Sajjadi, Antonio Vergari, Michael Black, and Bernhard Schölkopf. From variational to deterministic autoencoders. 2020.

[30] Joseph H A Guillaume, Casey Helgeson, Sondoss Elsawah, Anthony J Jakeman, and Matti Kummu. Toward best practice framing of uncertainty in scientific publications: A review of water resources research abstracts. *Water resources research.*, 53(8):6744–6762, 2017-08.

[31] Hakan Gunduz. An efficient stock market prediction model using hybrid feature reduction method based on variational autoencoders and recursive feature elimination. *Financial Innovation*, 7, 12 2021.

[32] Shivani Gupta and Atul Gupta. Dealing with noise problem in machine learning data-sets: A systematic review. *Procedia Computer Science*, 161:466–474, 1 2019.

[33] Shasha Han and Paulin Coulibaly. Bayesian flood forecasting methods: A review. *Journal of hydrology*, 551:340–351, 2017-08.

[34] Anwar Ul Haq, Adnan Zeb, Zhenfeng Lei, and Defu Zhang. Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems with Applications*, 168:114444, 4 2021.

[35] Arman Hassanniakalager, Georgios Sermpinis, and Charalampos Stasinakis. Trading the foreign exchange market with technical analysis and bayesian statistics. *Journal of Empirical Finance*, 63:230–251, 2021.

[36] Stephen C. Hora. Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. *Reliability Engineering and System Safety*, 54:217–223, 11 1996.

[37] Xiurui Hou, Kai Wang, Cheng Zhong, and Zhi Wei. St-trader: A spatial-temporal deep neural network for modeling stock market movement. *IEEE/CAA Journal of Automatica Sinica*, 8:1015–1024, 5 2021.

[38] Essam Houssein, Mahmoud Dirar, Kashif Hussain, and Waleed Makram Mohamed. Assess deep learning models for egyptian exchange prediction using nonlinear artificial neural networks. *Neural Computing and Applications*, 09 2020.

[39] Biao Huang, Qiao Ding, Guo-Zi Sun, and Huakang Li. Stock prediction based on bayesian-lstm. pages 128–133, 02 2018.

[40] Jian Huang, Junyi Chai, and Stella Cho. Deep learning in finance and banking: A literature review and classification. 2020.

[41] Yang Huang, Duen Ren Liu, Shin Jye Lee, Chia Hao Hsu, and Yang Guang Liu. A boosting resampling method for regression based on a conditional variational autoencoder. *Information Sciences*, 590:90–105, 4 2022.

[42] Petr Hájek. Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. 66:35–65, 2011.

[43] Petr Hájek and Jana Boháčová. Predicting abnormal bank stock returns using textual analysis of annual reports – a neural network approach. *Communications in Computer and Information*

Science, 629:67–78, 2016.
[44] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. 10 2019.
[45] iMarc. Foreign exchange market size, share, trends, analysis 2022-2027, 2021.
[46] Huisu Jang and Jaewook Lee. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, 6:5427–5437, 2018.
[47] Huisu Jang and Jaewook Lee. Generative bayesian neural network model for risk-neutral pricing of american index options. *Quantitative Finance*, 19:1–17, 11 2018.
[48] Patel Jay, Vasu Kalariya, Pushpendra Parmar, Sudeep Tanwar, Neeraj Kumar, and Mamoun Alazab. Stochastic neural networks for cryptocurrency price prediction. *IEEE Access*, 8:82804–82818, 2020.
[49] Laurent Valentin Jospin, Wray Buntine, Farid Boussaid, Hamid Laga, and Mohammed Bennamoun. Hands-on bayesian neural networks – a tutorial for deep learning users. 7 2020.
[50] Vasu Kalariya, Haitao Li, June Feng, Ben Niu, Shihua Fu, Pushpendra Parmar, Patel Jay, Sudeep Tanwar, Maria Simona Raboaca, Fayez Alqahtani, Amr Tolba, and Bogdan-Constantin Neagu. Stochastic neural networks-based algorithmic trading for the cryptocurrency market. 2022.
[51] Imam Mustafa Kamal, Hyerim Bae, Sim Sunghyun, and Heesung Yun. Dern: Deep ensemble learning model for shortand long-term prediction of baltic dry index. *Applied Sciences (Switzerland)*, 10, 2 2020.
[52] Zeynep Hilal Kilimci. Financial sentiment analysis with deep ensemble models (dems) for stock market prediction. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35:635–650, 2020.
[53] Soohyon Kim. Macroeconomic and financial market analyses and predictions through deep learning. *SSRN Electronic Journal*, 01 2020.
[54] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. 2014.
[55] Eisuke KITA, Masaaki Harada, and Takao Mizuno. Application of bayesian network to stock price prediction. *Artificial Intelligence Research*, 1:171, 9 2012.
[56] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31:105–112, 3 2009.
[57] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. 12 2016.
[58] Hui Li, Yunpeng Cui, Yunpeng Cui, Shuo Wang, Shuo Wang, Juan Liu, Juan Liu, Jinyuan Qin, and Yilin Yang. Multivariate financial time-series prediction with certified robustness. *IEEE Access*, 8:109133–109143, 2020.
[59] Yen Chih Lin and Jeremy Blum. Feature transformation and simulation of short term price variability in reinforcement learning for portfolio management. Institute of Electrical and Electronics Engineers Inc., 5 2020.
[60] Wen Yulian Lin Peiguang, Zhou Jiaqian. Sconv: A financial market trend forecast method based on emotional analysis. *Journal of Computer Research and Development*, 57(8):1769, 2020.
[61] Gustav Lind, Aliaksandra Kupreyeva, and Andreas Lindgren. (un)certain asset price prediction using bayesian neural networks. 2021.
[62] Bingqing Liu, Ivan Kiskin, and Stephen Roberts. An overview of gaussian process regression for volatility forecasting. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 681–686, Feb 2020.
[63] Changwei Liu and Hao Wang. Stock recommendation based on depth brnn and bi-lstm. *Proceedings - 2019 4th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2019*, pages 751–755, 10 2019.
[64] Hui Liu and Zhihao Long. An improved deep learning model for predicting stock market price time series. *Digital Signal Processing*, 102:102741, 2020.
[65] Ioannis E. Livieris, Stavros Stavroyiannis, Emmanuel Pintelas, Theodore Kotsilieris, and Panagiotis Pintelas. A dropout weight-constrained recurrent neural network model for forecasting the price of major cryptocurrencies and cci30 index. *Evolving Systems*, 13:85–100, 2 2022.
[66] Rui Luo, Weinan Zhang, Xiaojun Xu, and Jun Wang. A neural stochastic volatility model. pages 6401–6408, 2018.
[67] T. Ma, P.A.F. Fraser-Mackenzie, M. Sung, A.P. Kansara, and J.E.V. Johnson. Are the least successful traders those most likely to exit the market? a survival analysis contribution to the efficient market debate. *European Journal of Operational Research*, 299(1):330 – 345, 2022.
[68] David J C Mackay. Bayesian interpolation. 1992.
[69] David J C Mackay. Probable networks and plausible predictions-a review of practical bayesian

methods for supervised neural networks. 1995.
[70] Iwao Maeda, Hiroyasu Matsushima, Hiroki Sakaji, Kiyoshi Izumi, David deGraw, Atsuo Kato, and Michiharu Kitano. Effectiveness of uncertainty consideration in neural-network-based financial forecasting. 07 2019.
[71] Iwao Maeda, Hiroyasu Matsushima, Hiroki Sakaji, Kiyoshi Izumi, David deGraw, Atsuo Kato, and Michiharu Kitano. Predictive uncertainty in neural network-based financial market forecasting. *International Journal of Smart Computing and Artificial Intelligence*, 5(1):1–18, 2021.
[72] Iwao Maeda, Hiroyasu Matsushima, Hiroki Sakaji, Kiyoshi Izumi, David deGraw, Hirokazu Tomioka, Atsuo Kato, and Michiharu Kitano. Learning uncertainty in market trend forecast using bayesian neural networks. In Edgardo Bucciarelli, Shu-Heng Chen, and Juan Manuel Corchado, editors, *Decision Economics: Complexity of Decisions and Decisions for Complexity*, pages 210–218, Cham, 2020. Springer International Publishing.
[73] Martin Magris, Mostafa Shabani, and Alexandros Iosifidis. Bayesian bilinear neural network for predicting the mid-price dynamics in limit-order book markets. 3 2022.
[74] Luciana S. Malagrino, Norton T. Roman, and Ana M. Monteiro. Forecasting stock market index daily direction: A bayesian network approach. *Expert Systems with Applications*, 105:11–22, 2018.
[75] Thomas P Minka. Expectation propagation for approximate bayesian inference. 2001.
[76] Saurabh Mishra and Bilal M Ayyub. Predicting regional economic activity using artificial intelligence (ai) methods: Case study with indian states reliability and risk, 2018.
[77] Luis Montesdeoca, Steven Squires, and Mahesan Niranjan. Variational autoencoder for non-negative matrix factorization with exogenous inputs applied to financial data modelling. *International Symposium on Image and Signal Processing and Analysis, ISPA*, 2019-September:312–317, 9 2019.
[78] Nagaraj Naik and Biju R. Mohan. Study of stock return predictions using recurrent neural networks with lstm. *Communications in Computer and Information Science*, 1000:453–459, 2019.
[79] Maicon Nardino, Fabyano Fonseca Silva, Tiago Olivoto, Willian Silva Barros, Chainheny Gomes de Carvalho, Victor Silva Signorini, Henrique Caletti Mezzomo, Cleiton Renato Casagrande, and Av PH Rolfs. Meta-analysis of the experimental coefficient of variation in wheat using the bayesian. *Sci. Agric. v*, 80:2023, 2021.
[80] Manfred Opper and Cédric Archambeau. The variational gaussian approximation revisited, 3 2009.
[81] Trilok Nath Pandey, Alok Kumar Jagadev, Satchidananda Dehuri, and Sung-Bae Cho. A novel committee machine and reviews of neural network and statistical models for currency exchange rate prediction: An experimental analysis. *Journal of King Saud University - Computer and Information Sciences*, 32(9):987–999, 2020.
[82] Kidon Park, Hong-Gyu Jung, Tae-San Eom, and Seong-Whan Lee. Uncertainty-aware portfolio management with risk-sensitive multiagent network. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–14, 5 2022.
[83] Hiral R. Patel, Ajay M. Patel, Hiral A. Patel, and Satyen M. Parikh. Hyperparameter tune for neural network to improve accuracy of stock market prediction. volume 90, pages 65–76. Springer Science and Business Media Deutschland GmbH, 2022.
[84] Michael Pires and Tshilidzi Marwala. Option pricing using bayesian neural networks. 06 2007.
[85] Jocelyn Pixley. The use of risk in understanding financial decisions and institutional uncertainty. *The Journal of Socio-Economics*, 39(2):209–222, 2010.
[86] Claudia Primasiwi, Riyanarto Sarno, Kelly Rossa Sungkono, and Cahyaningtyas Sekar Wahyuni. Stock composite prediction using nonlinear autoregression with exogenous input (narx). In *2019 12th International Conference on Information Communication Technology and System (ICTS)*, pages 43–48, July 2019.
[87] Anu Priya and Shruti Garg. A comparison of prediction capabilities of bayesian regularization and levenberg–marquardt training algorithms for cryptocurrencies. volume 159, pages 657–664. Springer, 2020.
[88] Paramita Ray, Bhaswati Ganguli, and Amlan Chakrabarti. A hybrid approach of bayesian structural time series with lstm to identify the influence of news sentiment on short-term forecasting of stock price. *IEEE Transactions on Computational Social Systems*, 8(5):1153–1162, Oct 2021.
[89] Ehsan Rokhsatyazdi, Shahryar Rahnamayan, Hossein Amirinia, and Sakib Ahmed. Optimizing lstm based network for forecasting stock market. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–7, July 2020.
[90] Sandro Sacchelli, Sara Fabbrizzi, and Silvio Menghini. Climate change effects and adaptation strategies in the wine sector: a quantitative literature review. *Wine economics and policy.*,

5(2):114–126, 2016-12.

[91] Arjun Singh Saud and Subarna Shakya. Evaluation of weight decay regularization techniques for stock price prediction using gated recurrent unit network. *Nepal Journal of Science and Technology*, 19:9–15, 10 2021.

[92] O Schwabe, E Shehab, and J Erkoyuncu. Uncertainty quantification metrics for whole product life cycle cost estimates in aerospace innovation. *Progress in aerospace sciences*, 77:1–24, 2015-08.

[93] Dharmaraja Selvamuthu, Vineet Kumar, and Abhishek Mishra. Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation*, 5, 12 2019.

[94] Cristián Serpell, Ignacio Araya, Carlos Valle, and Héctor Allende. Probabilistic forecasting using monte carlo dropout neural networks. 2019.

[95] Burr Settles. Active learning. synthesis lectures on artificial intelligence and machine learning, 2012.

[96] Bin Shen and Binheng Song. Uncertainty estimation for strong-noise data. volume 11013 LNAI, pages 447–454. Springer Verlag, 2018.

[97] Aman Singh and Tokunbo Ogunfunmi. An overview of variational autoencoders for source separation, finance, and bio-signal applications. *Entropy : an international and interdisciplinary journal of entropy and information studies.*, 24(1), 2022-1-28.

[98] Andrew A Skabar. Direction-of-change financial time series forecasting using neural networks: A bayesian approach. *S.I. Ao and L. Gelman (Eds.),Advances in Electrical Engineering and ComputationalScience, Lecture Notes in Electrical Engineering.*, 2009.

[99] Qi Sun, Wen Gang Che, and Hong Liang Wang. Bayesian regularization bp neural network model for the stock price prediction. volume 213, pages 521–531. Springer Verlag, 2014.

[100] Chon Sern Tan, Chin Khian Yang, and Yang Haur Tay. Modeling financial ratios of malaysian plantation stocks using bayesian networks. IEEE, 2012.

[101] Jonathan L. Ticknor. A bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14):5501–5506, 2013.

[102] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1:1–305, 2008.

[103] Bin Wang, Tianrui Li, Zheng Yan, Guangquan Zhang, and Jie Lu. Deeppipe: A distribution-free uncertainty quantification approach for time series forecasting. *Neurocomputing*, 397:11–19, 2020.

[104] Jianzhou Wang, Tong Niu, Pei Du, and Wendong Yang. Ensemble probabilistic prediction approach for modeling uncertainty in crude oil price. *Applied Soft Computing*, 95:106509, 10 2020.

[105] Jujie Wang, Yu Chen, Shuzhou Zhu, and Wenjie Xu. Depth feature extraction-based deep ensemble learning framework for high frequency futures price forecasting. *Digital Signal Processing*, 127:103567, 7 2022.

[106] Jujie Wang, Liu Feng, Yang Li, Junjie He, and Chunchen Feng. Deep nonlinear ensemble framework for stock index forecasting and uncertainty analysis. *Cognitive Computation*, 13:1–19, 11 2021.

[107] Jujie Wang, Junjie He, Chunchen Feng, Liu Feng, and Yang Li. Stock index prediction and uncertainty analysis using multi-scale nonlinear ensemble paradigm of optimal feature extraction, two-stage deep learning and gaussian process regression. *Applied Soft Computing*, 113:107898, 2021.

[108] Muye Wang. Essays on the applications of machine learning in financial markets, 2021.

[109] Yali Wang, Marcus Brubaker, Brahim Chaib-Draa, and Raquel Urtasun. Sequential inference for deep gaussian process. pages 694–703, 2016.

[110] Yifeng Wang, Yuying Liu, Meiqing Wang, and Rong Liu. Lstm model optimization on stock price forecasting. pages 173–177. Institute of Electrical and Electronics Engineers Inc., 12 2018.

[111] Nicholas Wilkins. Efficient distribution regression, 2019.

[112] Nicholas Wilkins, Michael Johnson, and Ifeoma Nwogu. Regression with uncertainty quantification in large scale complex data. 12 2019.

[113] Andrew Gordon Wilson and Pavel Izmailov. Deep ensembles as approximate bayesian inference — deep ensembles as approximate bayesian inference, 2021.

[114] Wenjun Wu, Yue Wang, Jiaxiu Fu, Jiayi Yan, Zhiping Wang, Jiaqing Liu, Weifan Wang, and Xiuli Wang. Preliminary study on interpreting stock price forecasting based on tree regularization of gru. volume 1059, pages 476–487. Springer Verlag, 2019.

[115] Qinkun Xiao, Chu Chaoqin, and Zhao Li. Time series prediction using dynamic bayesian network. *Optik*, 135:98–103, 2017.

[116] Frank Z Xing, Erik Cambria, and Yue Zhang. Sentiment-aware volatility forecasting. *Knowledge-Based Systems*, 176:68–76, 2019.

[117] Fucui Xu and Shan Tan. Deep learning with multiple scale attention and direction regularization

for asset price prediction. *Expert Systems with Applications*, 186, 12 2021.
[118] Yumo Xu and Shay B Cohen. Stock movement prediction from tweets and historical prices. pages 1970–1979. Association for Computational Linguistics, 2018.
[119] Dingqi Yan, Qi Zhou, Jianzhou Wang, and Na Zhang. Bayesian regularisation neural network based on artificial intelligence optimisation. *International Journal of Production Research*, 55:2266–2287, 4 2017.
[120] Hongyang Yang, Xiao-Yang Liu, Shan Zhong, Wormpex Ai, Research Bellevue, Washington Anwar Walid, and Anwar Walid. Deep reinforcement learning for automated stock trading: An ensemble strategy deep reinforcement learning, markov decision process, automated stock trading, ensemble strategy, actor-critic framework acm reference format. 20, 2020.
[121] Jonathan S Yedidia, Merl 201 Broadway, William T Freeman, and Abstract Yair Weiss. Generalized belief propagation. 2001.
[122] Bin Xie Yue Hou and Heng Liu. Bayesian regularization neural network model for stock time series prediction. *International Journal of Performability Engineering*, 15(12):3271, 2019.
[123] Guijuan Zhang, Yang Liu, and Xiaoning Jin. A survey of autoencoder-based recommender systems. *Front. Comput. Sci*, 2020:430–450, 2020.
[124] Zhuorui Zhang, Junhao Zhou, Yanan Song, and Hong-Ning Dai. Cryptocurrencies price prediction using weighted memory multi-channels. In Zibin Zheng, Hong-Ning Dai, Xiaodong Fu, and Benhui Chen, editors, *Blockchain and Trustworthy Systems*, pages 502–516, Singapore, 2020. Springer Singapore.
[125] Zihao Zhang, Stefan Zohren, and Stephen Roberts. 12 2018.
[126] Zhi-Hua. Zhou. *Ensemble Methods: Foundations and Algorithms.* CHAPMAN & HALL/CRC MACHINE LEA. Taylor & Francis, 2012.
[127] Yi Zuo, Masaaki Harada, Takao Mizuno, and Eisuke Kita. Bayesian network based prediction algorithm of stock price return, 2012.
[128] Yi Zuo and Eisuke Kita. Up/down analysis of stock index by using bayesian network. *Engineering Management Research*, 1, 7 2012.