Neurocomputing 576 (2024) 127339

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Survey Paper

A survey on uncertainty quantification in deep learning for financial time series prediction

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

[a] Institute of New Imaging Technologies, Department of Computer Languages and Systems, Universitat Jaume I, Ave. Vicent Sos Baynat s/n, 12071 Castello de la Plana, Spain

[b] Departamento de Ingenieria Electrica y Computacion, Division Multidisciplinaria en Ciudad Universitaria, Universidad Autonoma de Ciudad Juarez, Ave. Plutarco Elias Calles 1210, Ciudad Juarez, Mexico

ARTICLE INFO

Communicated by P. Zhao

Keywords:
Deep Learning
Time series
Uncertainty quantification
Prediction
Financial assets

ABSTRACT

Investors make decisions about buying and selling a financial asset based on available information. The traditional approach in Deep Learning when trying to predict the behavior of an asset is to take a price history, train a model, and forecast one single price in the near future. This is called the frequentist perspective. Uncertainty Quantification is an alternative in which models manage a probability distribution for prediction. It provides investors with more information than the traditional frequentist way, so they can consider the risk of making or not making a certain decision. We systematically reviewed the existing literature on Uncertainty Quantification methods in Deep Learning to predict the behavior of financial assets, such as foreign exchange, stock market, cryptocurrencies and others. The article discusses types of model, categories of financial assets, prediction characteristics and types of uncertainty. We found that, in general terms, references focus on price accuracy as a metric, although other metrics, such as trend accuracy, might be more appropriate. Very few authors analyze both epistemic and aleatoric uncertainty, and none analyze in depth how to decouple them. The time period analyzed includes the years 2001 to 2022.

1. Introduction

Deep Learning (DL) applications have increasing relevance in almost all fields, particularly finance. Along with this growing interest, it appears that most authors are engaging in a model race to build the most accurate, predicting future asset prices with the least error. This competition might make sense in some applications where the level of volatility, called noise, remains low to moderate. However, the financial market presents a very high level of volatility. The amount of noise related to this type of time series remains so high that researchers are still debating whether Efficient Market Theory is valid or not [1]. Instead of looking for a very accurate prediction, traders seek for a reasonable prediction that helps them understand which direction the market will move next, up or down. A successfully predicted trend is much more useful for securing an investment rather than knowing if the next price will reach €29.4 instead of €29.5. In this context, it seems reasonable to ask about the point of trying to boost the accuracy of DL models in a scenario where the volatility of the input data is much larger than the minimum error one can obtain from the predictions. Trying to follow that line will lead to an overfitted model, inevitably falling into the bias–variance dilemma. The authors of this review propose an alternative approach for predicting financial trading series, in which the next price vector is framed inside 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 noise in the data and the degree of knowledge acquired by the DL model.

We can distinguish between two types of uncertainties in prediction with DL models, called *aleatoric* and *epistemic* [2,3]. The total uncertainty in the prediction is the sum of both uncertainties. The first one is related to the intrinsic noise in data and is called aleatoric uncertainty. Aleatoric refers to inherent randomness of price data, i.e. price fluctuations or volatility confusing genuine underlying trends [4]. Noise is an intrinsic part of financial data and is the main reason why predicting next prices remains so complex: DL models struggle to recognize the difference between noise and a real trend. In fact, financial data, such as stock market prices, differ from other time series in that they are characterized by a very high level of noise. The stochastic component of financial data misrepresents valuable price information and therefore confounds models. Unfortunately, financial data always exhibit this stochastic component and it is impossible to reduce the related

* Corresponding author.
E-mail addresses: al419983@uji.es (T. Blasco), sanchez@uji.es (J.S. Sánchez), vicente.jimenez@uacj.mx (V. García).

https://doi.org/10.1016/j.neucom.2024.127339
Received 8 June 2023; Received in revised form 28 October 2023; Accepted 26 January 2024
Available online 28 January 2024

uncertainty simply by adding more data [5]. 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, to maintain trades profitable. Identification of the aleatoric part is the previous and necessary step to address it. While the aleatoric uncertainty cannot be eliminated, in fact 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 the model ingests the raw data, the prediction accuracy is negatively affected. The reason is that the model will stick to noise patterns rather than the valuable real price trends [6].

The second type of uncertainty, the epistemic component, is related to the prediction model, coming from sets of hyperparameters not chosen for 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 weights and biases of the neural network (NN), the length and depth, the shape (variation of length across the depth), the number of iterations and batches, the optimizer or the regularization term, to name a few. Instead of simply choosing one optimized combination of those parameters, one could think about a model from a different perspective, considering that not just one but many of those combinations could shape an *ensemble* of valid models. In this way, we do not have a matrix of weights and biases, but rather the model handles a probability distribution corresponding to each of the weights in the matrix (epistemic uncertainty) [5].

Already in 1992, Mackay [7] establishes the foundations of what is called today Uncertainty Quantification (UQ) and applies them to function approximation. He states that the Bayes approach is not consistently better in performance than other methods for interpolating a noisy dataset. However, a second level of inference is generally forgotten, which is the ability of Bayesian methods to rank the alternative models to the best chosen one.

If we take this concept to finance, it translates into opportunity cost, or in other words, how much we are paying for not choosing the alternative to the (supposedly) best trade at the right time. Big financial decisions are made not simply by knowing a single future price value recommended by a DL model (this is how a frequentist approach is defined [8]). It is also necessary to consider risk and opportunity cost [9]. We believe that the UQ approach is best suited for real financial scenarios where probability and confidence intervals can be introduced, not just for one predicted point in the future but for entire identifiable trends.

UQ is defined differently throughout the existing literature. Within the finance field, if we continue to approach the even more specific subdomain of predicting the behavior of financial and tradeable assets through DL, it can be defined as *the science of measuring the extent to which a DL model is uncertain about the profit that can be obtained in the future*. Ideally, this amount of uncertainty would be measured with a probability distribution [10], or at least with the mean and the variance varying over time [11]. UQ can be applied to a regression problem, that is, predicting future price values [12]. It can also be applied to a classification problem, where the output of a DL model is a buy/hold/sell signal [13] or the start or end of a trend [14]. It can even be applied to a ranking type prediction, for example predicting the momentum in the future.

The main objective of this paper is to evidence whether the state of the art related to UQ in DL applied to financial forecasting may leave unexplored or insufficiently explored techniques, methods or approaches that could be further exploited by researchers in potential future research and clarify which would be. Some other literature reviews explore UQ applied to various fields: water resources research [15], flood forecasting [16], climate modeling [17], cost estimation of the aerospace life cycle [18] and others. To the best of our knowledge, this is the first survey in the topic of UQ applied to financial time series.

Henceforward, the paper is organized as follows. Section 2 explains the methodology used in the review to analyze the existing literature, following the PRISMA approach for records found between 2001 and 2022 without filters regarding the time period. Section 3 collects and discusses the results of the analysis, provides potential approaches in future research, and describes existing techniques commonly used to quantify uncertainty. Finally, Section 4 draws the main conclusions of this study.

2. Research methodology

PRISMA has been used as the main framework to carry out the research. In addition, the following questions are defined to shed light on the current state of the art on UQ for financial asset prediction using DL:

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

The above questions shape the research and guide the search for references, while they will try to be answered in Section 3.6.

2.1. Inclusion and exclusion criteria

UQ in DL is a relatively new domain. It has been found that as of 2010, there is an explosion of documents dealing with the topic and explicitly talking about the exact phrase “uncertainty quantification”, as can be seen in Fig. 1. Although some articles mentioned UQ before that date, they were rarer. In fact, as stated by Magris et al. [19], Mackay’s paper [20] was the first work to apply UQ methods to a neural network using a Bayesian regularization.

Not all found references intentionally use UQ, nor do they explicitly talk about it. Still, whether prepared intentionally or not, all the selected papers meet the criteria for applying 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 in 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 references to be added to this review are described as follows (all of them must be met):

1. **Prediction and trading of financial asset time series**: The reference must base the topic on methods or techniques for predicting the value of trading instruments (Forex, equities, commodities and cryptocurrencies, see Fig. 2). The problem to be solved can be regression, predicting the future value of the asset, return and confidence intervals, or classification, predicting up or down trends or buy/sell signals. The condition to be included in this group is that the asset is tradeable, that is, the asset can be bought and sold as a financial product for speculation.
2. **Use of DL**: The prediction should be done using DL. Classification can be done, for example, by categorizing a series of prices with a buy (go long) signal or a sell (go short) signal. Regression is used to predict single price data points in the future, several step-ahead price data points, or price intervals. Fig. 3 shows a classification of the main DL methods used to quantify uncertainty in financial time series forecasting.

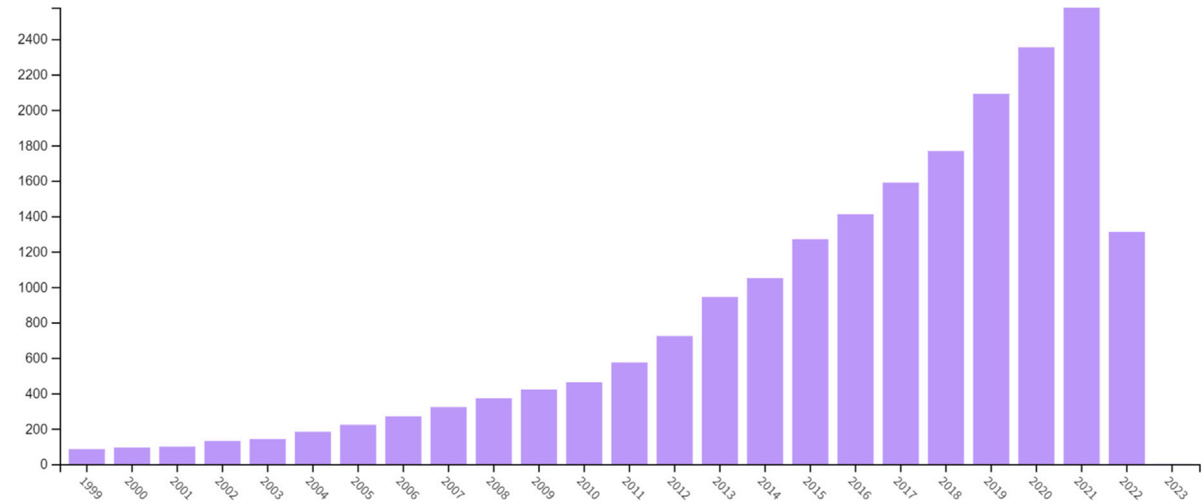

Fig. 1. Number of publications per year related to the general term Uncertainty Quantification. Source: Web of Science.

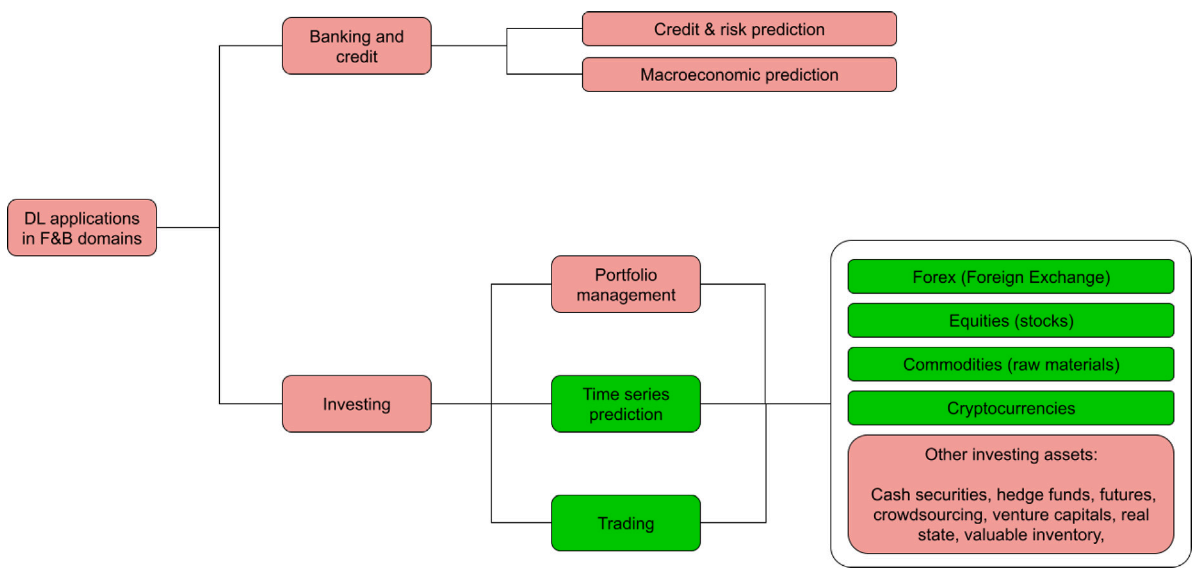

Fig. 2. Type of financial and banking applications in DL. Based on Huang 2020 [23] and updated with Investopedia [24].

3. **Use of UQ methods**: The used methods should fall within the UQ domain. Even the unintentional use of UQ methods has been considered and added to the survey. Based on [21], a list of UQ methods has been adapted as the main criterion for including an article. These are listed and described in Section 3.1, divided by type of techniques.

Some works, even those predicting economic time series, have not been considered because the data does not correspond to a financial asset. For example, Mishra and Ayyub [22] employed an MC dropout LSTM model to predict net state domestic product (NSDP). NSDP does not present enough noise, compared to financial markets, to be included in this study.

Other studies focusing on portfolio management (PM) have been excluded because they do not solve a value prediction problem. Instead, they are solving for an asset classification or ranking. Sometimes the PM problem can classify the behavior of the value curve in the past, but it does not predict the behavior of the asset value in the future (the work of Park et al. [25] is an example of this). However, this does not mean that the PM is excluded from this survey. In the case of [26], the PM problem includes a time series prediction with UQ, and is consequently considered in the present study.

Some keywords can cause confusion when it comes to meeting the inclusion criteria. For example, when searching for the terms *regularization*, *stock price prediction* and *DL*, some studies were initially included in the screening because they appeared in both the title and the abstract. However, they were discarded after reading the abstract and realizing that they do not use UQ techniques [27–29]. Similarly, some studies related to *Neural Networks* or *Bayesian Networks* in price prediction were added to the screening, but were later discarded because the used models were not deep neural networks [30–36]. Other studies, such as the paper of Alarab et al. [37], use MC-dropout DL for valid/not valid classification of Bitcoin transactions and the article was dropped because it does not include a regression problem, predicting the future value of Bitcoin, trend, nor a buy/sell classification problem.

Only works written in plain English have been included in the present study: based on this, two papers written in Turkish and Japanese [38,39] were discarded. It is also important to mention that it was necessary to include the broader term *Network* instead of *Neural Network* because some authors such as Wilkins [40] use that more general term.

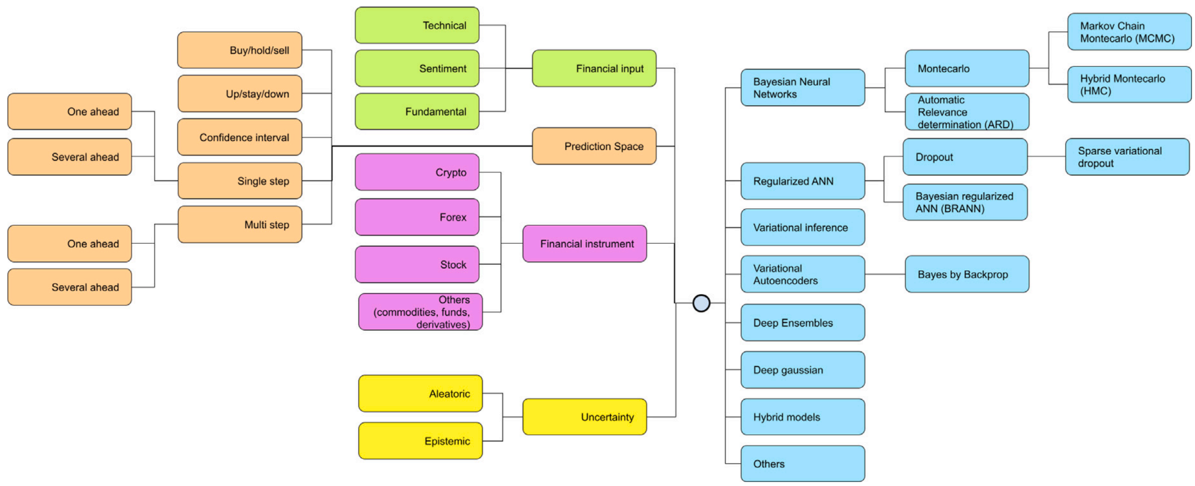

Fig. 3. Taxonomy of the main DL methods used to quantify uncertainty in financial time series forecasting (right side) and the financial analysis perspectives related to each of those DL methods.

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	45 300
("dropout") AND ("neural network") AND ("stock") OR ("crypto*") OR ("financ*") AND ("prediction")	23	37 600
("markov chain montecarlo") OR ("MCMC") AND ("neural network") AND ("stock") OR ("crypto*") OR ("financ*") AND ("prediction")	4	19 300
("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

An article was also removed [41] from the selected records because the single point price prediction was made not only considering past data but also future data. This reference was discarded because it does not predict prices, since future data in real scenarios will never be available.

Variational Autoencoders (VAE) is a method widely used to generate synthetic datasets and resampling in financial time series. All studies not directly using VAE for the prediction of the next value or trading have been excluded from this survey [42–45].

2.2. Information sources

In this study, the Scopus database and the Google Scholar search engine were used. Given that there are very few articles that emerge from the search, no time limitation was applied.

2.3. Search strategy

The search strings were refined until the most appropriate records were found. This survey chose Scopus because it is a widely recognized automatic database that shows quality results. However, the number of records considered for screening after refinement was only 171, a number too low for our purpose. That is why we decided to use a more general search engine such as Google Scholar, in which we could add 251 more records for screening. Due to the initial low number of records found, we decided not to limit the date range. To the extent a study meets the requirements described in 2.1, it was included with no 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 screened, 156 were assessed, and 69 were added to the study. Since Google Scholar shows huge numbers for some searches, the search was limited to the first 10 pages shown, as these are considered the most relevant. Table 1 shows a summary of the number of records found, ordered by database/search engine.

2.4. Study selection process

The main idea behind this survey is to bring together all studies that directly represent or use uncertainty within the neural network architecture applied to financial time series prediction. Search queries have been designed with this idea in mind. The record was chosen

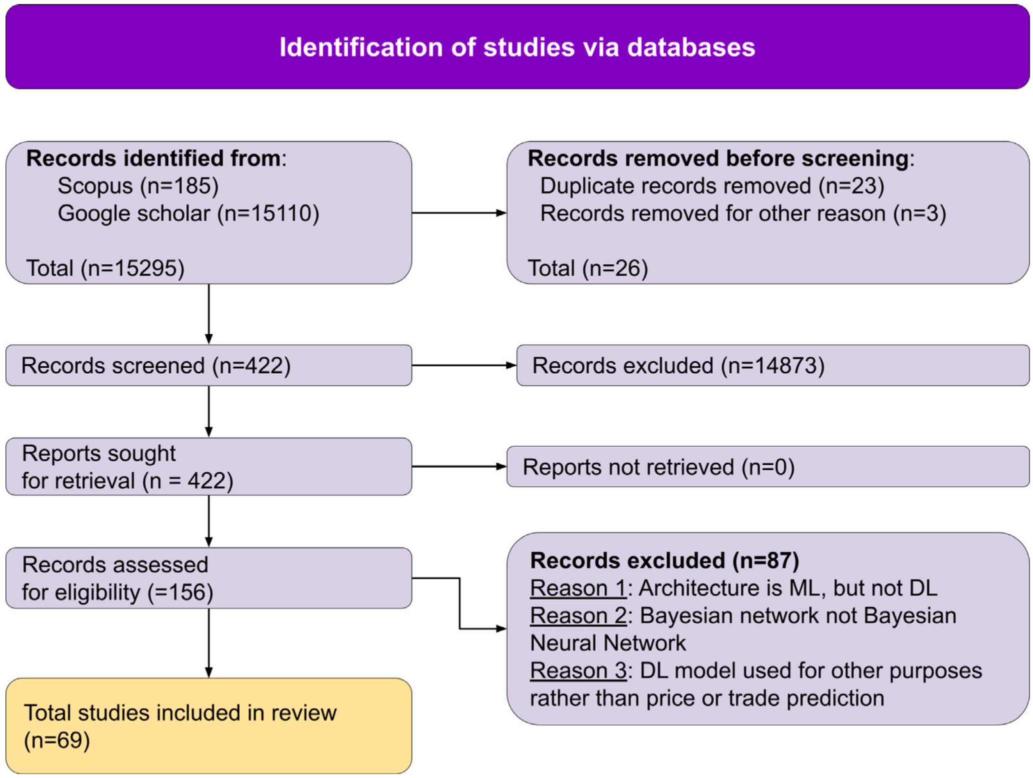

Fig. 4. Results of the search and selection process, from the number of records identified in the search to the number of studies included in the review.

for screening only if the title and abstract represented the purpose of the study. The screening process involves deep 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 was refined to ensure that all articles met the objectives established in the study. Fig. 4 shows a summary of the results of the search and selection process.

3. Results and discussion

This section lists the most representative studies of the survey. The order presented in Table 2 is the same in which the studies were found and therefore included. It is not uncommon that studies listed in Table 2 do not explicitly describe all the parameters defining the models they used. For example, the number of predicted steps in the future is not always specified, as in the case of [46]. Possibly, a single-step prediction could have been deducted, but instead, and to keep caution, the key "unspecified" has been added. The records have been divided into five categories based on the parameters considered most relevant: *model*, *type of asset*, *financial input (analysis)*, *prediction space* and *epistemic/aleatoric*.

3.1. Analysis by model

The category *model* includes the type of DL models used to represent uncertainty. They should be part of at least one of the categories described in Section 2.1. In the following subsections, we list the main techniques used in the literature found for a probabilistic approach to time series prediction and briefly explain them and provide some definitions. Following Wilson and Izmailov [107], we will not make a difference between Bayesian and non-Bayesian methods for approximating the exact posterior distribution, since some of the non-Bayesian methods approximate the integral even better than the Bayesian ones. Fig. 3 has been included to shed light on the set of main techniques that can be found in the literature.

3.1.1. Bayesian neural networks (BNN)

For a financial time series, $D = \{X, Y\} = \{(x_i, y_i)\}_i^N$ is a training dataset with historical price inputs $x_i \in \mathbb{R}$ and labeled price outputs $y_i \in \mathbb{R}$. From a frequentist point of view, a neural network attempts to represent future values of a function $y = \Phi_\theta(x)$, given that it has learned the behavior of Φ in the past. Let f and l be a non-linear transformation (activation function) and a linear transformation, respectively. Then, in the simplest architecture definition of a neural network, l_i is applied to W_i at 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 rather an ensemble of values that make up a probability distribution [108]. A BNN would then be a stochastic neural network trained using Bayesian inference [109]. 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 an array of input 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 Estimation (MAP), 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 establishes

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

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

where $p(\theta)$ is the prior probability associated with all possible models θ that explain y as an approximation to the real, but unknown

Table 2
List of all studies included in the survey.

Ref.	Model	Asset	Analysis	Prediction	Uncertainty
[47]	BNN	CRY	T	SS	EPI
[46]	Bayesian regularized NN	STO	T	U	EPI
[48]	BNN + MCMC	STO	T	MVMS	EPI
[49]	Stochastic ANN + LSTM	CRY	T,S	SS	EPI
[50]	Hybrid model BST + LSTM	STO	T,S	TR	ALE
[51]	NARX BRANN	STO	T	SVMS	EPI
[12]	Bayesian LSTM	STO	T	SS	ALE
[52]	dropout Bayesian CNN	STO	T	SS	EPI
[53]	Levenberg–Marquardt BRANN	STO	T	SS	EPI
[54]	BNN + ARD and BNN + HMC	STO	T	SS	EPI
[13]	SVDBNN	STO	T	B/S	EPI
[55]	SSA-EWSVM-RNN-GPR DE	STO	T	CONF	EPI
[56]	VMD-AE-RNN-LSTM DE	STO	T	CONF	EPI
[57]	Dropout CNN	STO	T	TR	EPI
[58]	RBF, BNN and ARIMA DE	FOR	T,F	SVMS	EPI
[59]	BRANN	STO	T	MVMS	EPI
[60]	GBNN	STO	T	SVMS	EPI
[61]	Dropout CNN	STO	T	B/S	EPI
[62]	BRANN	FOR	T	B/S	EPI
[63]	Variational Autoencoders	FOR	T	CONF	EPI
[64]	BRANN	STO	T	U	EPI
[11]	Interval adapted LSTM AE	STO	T	SVMS, CONF	ALE
[65]	MC Dropout MLP	STO	T	SS	EPI
[66]	Dropout MLP	STO	T	SS	EPI
[67]	BRANN with NARX	STO	T	SS	EPI
[68]	BRANN	STO	T	SS	EPI
[69]	BRANN	STO	T	SS	EPI
[70]	BRANN	STO	T	SS	EPI
[71]	BRANN	CRY	T	U	EPI
[72]	CAED-TCN	STO	T	MSMV, CONF	EPI
[19]	Backprop BRANN	STO	T	TR	EPI
[14]	MCMC BRANN	STO	T	TR	EPI
[73]	VAE	STO	T	SS	EPI
[74]	Attention based VAE-LSTM	STO	T	TR	EPI
[75]	VAE-GCN-LSTM	STO	T	SS, TR	EPI
[40]	MDN	STO	T	SS	ALE
[76]	MC Dropout LSTM	STO	T, F	CONF	EPI
[77]	Noise quantification MLP	STO	T	SS	ALE
[78]	MDN	STO	T, S	CONF	ALE
[79]	LSTM + MC Dropout	CRY	T, S	S	EPI
[80]	BRANN via Variational Inference	STO	T	TR	EPI
[81]	VAE	STO	T	TR	EPI
[82]	VAE	CRY	T, S	TR	EPI
[83]	VAE RNN	STO	T, S	SS	ALE
[84]	VAE + VI Neural Network	STO	T, S	SS	EPI
[85]	BNN	CRY	T	SS	EPI
[26]	BNN	CRY	T	SVMS	EPI
[86]	recurrent BNN	STO	T	MVMS	EPI
[87]	BNN	STO	T	SS	EPI
[88]	Stochastic ANN	CRY	T, S	SS, B/S	EPI
[89]	dropout RNN	CRY	T	SS	EPI
[90]	dropout ANN	STO	T	SS	EPI
[91]	Dropout LSTM	STO	T	SS	EPI
[92]	Dropout LSTM	STO	T	SS	EPI
[93]	Dropout CNN+ LSTM	CRY	T	SS	EPI
[94]	Dropout LSTM	STO	T	SS	EPI
[95]	Dropout LSTM	STO	T	SS	EPI
[96]	Dropout MLP	STO	S	SVMS	EPI
[97]	Dropout MLP	STO	S	SVMS	EPI
[98]	NSVM (VI)	STO	T	SS	EPI
[99]	VAE-LSTM with dropout	STO	T	MSMV	ALE
[100]	dropout VAE-GRU	FUT	T	SS	ALE
[101]	VAE-LSTM	STO	T	SVMS	ALE
[102]	PAE-XNMF	STO	T	SS	ALE
[103]	Heteroskedastic DGP	STO	T	U	EPI
[10]	BPNN, LSTM, GPR, Lasso, BILSTM ensembles	STO	T	CONF	EPI
[104]	Deep ensembles LigthGBM	FUT	T	SS	EPI
[105]	A2C, DDPG, PPO ensembles	STO	T	B/S	EPI
[106]	RNN, GRU, LSTM ensembles	IND	T	SS	EPI

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: aleatoric uncertainty analysis EPI: epistemic uncertainty analysis U: unspecified.

function $\Phi_\theta(x)$, being ϵ the representation of the random noise added to the previous function.

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

- $p(D_y|D_x,\theta)$: likelihood
- $p(\theta)$: prior probability
- $\int_\theta p(D_y|D_x,\theta')p(\theta')d\theta'$: evidence

Instead of using backpropagation over past data as in ANN, BNN learns the probability distribution of the weights by approximating the posterior [109]. However, in practice the evidence is computationally expensive and is never calculated analytically. Instead, it is approximated by methods such as Markov Chain Montecarlo or Variational Inference that are described in the following sections.

BNNs have two main applications in financial time series: avoiding overfitting [46] and quantifying uncertainty. In a sense, both are closely related. The fact of having a stochastic weight matrix makes them good at generalizing, a desired characteristic in very noisy time series as financial ones. The difference between each application arises at the time of inference. Authors who employ BNNs as a way to generalize use the mean in the predicted posterior, while authors who want to quantify uncertainty use both the mean and the variance.

Yan et al. [87] apply BNNs to predict the closing price of the Shanghai Stock Index. As they state, the posterior probability distribution $P(h|X)$ of the neural network refines the prior distribution $P(\theta)$ upon receiving input data. In Bayesian training, the objective function is based on the likelihood of the sample data and the weight adjustments use the prior distribution of weights and threshold. This combines input data to modify the posterior distribution of weights and threshold. The Bayesian network parameters reference this posterior distribution, potentially improving prediction accuracy. Authors like Yan's team who use raw BNNs need a regularization method to overcome overfitting problems. In addition to the Mean Squared Error function E_D, in BNNs, the regularization term E_w is added and the error function remains $M_W = \alpha E_D + \beta E_W$, where α and β are hyperparameters that control the probability distribution of other parameters. The authors improve the generalization ability of the BNN using a Particle Swarm Optimization (PSO) algorithm. They claim that the BNN-PSO is reliable in predicting the Shanghai Index.

In the context of predictive uncertainty, BNN is a very general way of naming a neural network that uses Bayes theory to construct a posterior distribution. In the following sections, we will go deeper in different methods that make neural networks have probability distributions at prediction time, rather than single points, and how they apply to financial data.

3.1.2. Montecarlo (MC) dropout

MC dropout consists of randomly disconnecting different pairs of neurons during backpropagation for training and testing. This has two effects: on the one hand, it prevents overfitting and on the other, it approximates the posterior probability described above. In fact, [110] states that a neural network with arbitrary depth, without non-linearities and with a dropout approximates the Bayesian posterior probability. The objective function using 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. However, to mention a drawback, this technique needs to be complemented with others, since it cannot capture the uncertainty of the data by itself [111].

Chauhan et al. [76] explain that MC-dropout can be applied at inference time, based on a previous work from Gal et al. [110]. The authors use three financial time series data for each time step: fundamental features (income statement, cash flow and balance sheet), momentum features (calculated over 1, 3, 6 and 9 months), and auxiliary features (company's short interest, industry group and size category). They build experiments with two models, MLP and LSTM. Both are trained to meet the mean and the variance as targets, and both have prediction modeling where they have two neurons as output, the predicted values of mean and variance respectively. At prediction time, dropout is activated to produce different outputs each time, and the model is run 10 times to generate a sample of that size. Therefore, the total uncertainty is given by the sum of the predicted variance (coming directly from the data) and estimated epistemic uncertainty coming from sampling the predicted mean.

3.1.3. 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 stationary Markov Chain can construct a different but known distribution $f(x)$, 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. Typically $f(x)$ is chosen as an easy or convenient distribution, such as a normal or binomial distribution. If the chosen distribution is normal one, then the MCMC method is called the Metropolis method. In a more generalized definition, the sampling distribution can be a convenient distribution, different from the normal one, and the method in this case is called Metropolis–Hastings [109,112].

He and Chandra [48] apply this method to the prediction of multi-step prices in the stock market (3M, China Spacesat Company Limited, Common Wealth Bank of Australia and Daimler AG). A vector of historical closing prices $[x_1,\dots,x_n]$ is used to predict the next closing prices $[x_{n+1},\dots,x_m]$. This approach is called *technical analysis*, since it only takes into account the history of prices and not major economic events or investor sentiment. The model used by the authors utilizes a neural network architecture with five output neurons, one for each price step. Uncertainty in the prediction is managed by running the model 30 times, each time with a different weight initialization. To quantify the uncertainty, the RMSE is calculated from the sample of size 30, and also the 95% confidence interval.

A particular use case of MCMC is that of Back et al. [66] to predict open, close, high and low prices in 78 futures markets. They treat not only the model weights as probabilistic variables but also other hyperparameters such as the number of features, hidden dimensions, dropout rate, epochs, and activation function. The authors compared MCMC with variational inference and dropout. It is concluded that MCMC, unlike the other two methods, simultaneously offers regularization and generates predictive uncertainties associated with the prediction error. These uncertainties contributed to improving a trading strategy. However, this method does not fully converge to the desired true posterior distribution, so more research is needed on this aspect.

3.1.4. Variational inference (VI)

MCMC methods are effective if the size of datasets is moderate since they do not scale well. For large datasets, Variational Inference (VI) is preferred [113]. Instead of sampling from the exact posterior distribution, VI uses optimization to find the best performer of a family Q of approximate densities, each of which attempts to minimize the Kullback–Leibler (KL) divergence to the exact posterior

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

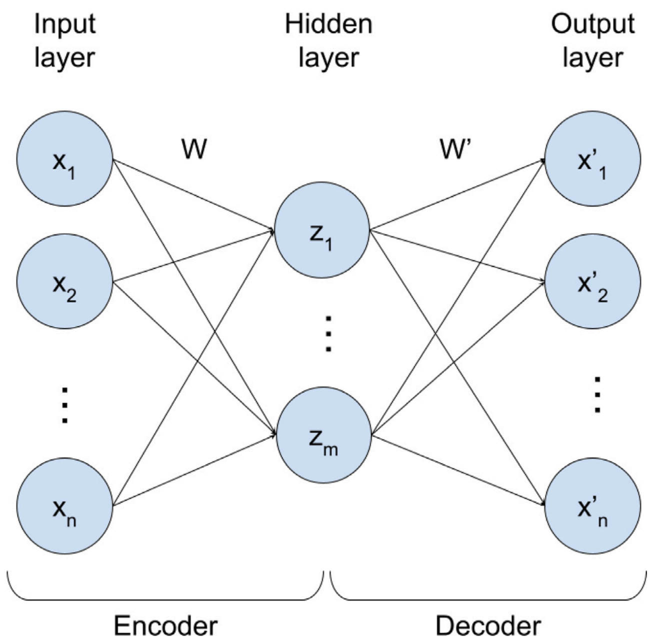

Fig. 5. Basic architecture of an AE.

then the posterior is approximated using the best member of the family Q, $q^*(\cdot)$. Because the family of densities Q can be chosen, it should be made as flexible as possible to capture $p(x|z)$ closely, but not too complex as to make optimization difficult. [114] states 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 VI approach: expectation propagation [115], belief propagation [116] or Laplace approximation.

In the article by Cocco et al. [85], two cryptocurrency series are studied using VI: BTC/USD and ETH/USD. The authors compare the performance of a BNN alone and an SVR + BNN with other models. In the first case, the BNN takes the market signals (SMA, EMA, MOM, RSI and MACD) as a feature vector to predict the price $t+n$. In the second case, the SVR takes the market signals to predict the market signals at $t+n$, and the BNN ingests those predictions to predict the price at $t+n$. n evaluates to $n=1$, $n=10$, and $n=20$. The posterior $p(\theta|D)$ is estimated by maximizing the ELBO through the reparameterization gradient.

3.1.5. Variational autoencoders (VAE)

Variational Bayes (VB), also called Neural Variational Inference, is a way to optimize an approximation of the posterior probability, which has already been stated as analytically intractable [117]. An autoencoder (AE) is a specific type of DL model that consists of two components: the encoder and the decoder. Fig. 5 shows the basic structure of an AE. The goal of the encoder is to map a high-dimensional input vector $x=\{x_1,x_2,\dots,x_n\}$ into a low dimensional latent output vector $z=\{z_1,z_2,\dots,z_m\}$ 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 $x'=\{x'_1,x'_2,\dots,x'_n\}$ 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 corresponding decoder. The decoder's parameters are defined by an $n\times m$ matrix of weights W' and a bias vector $b'\in R^n$.

The training objective of the AE is to minimize the reconstruction error between x and x' [118]. In that way, a learning representation for a high-dimensional distribution can be converted into a simpler VI problem [119]. 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 lower bound of evidence 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–Leibler divergence.

Haq et al. [84] construct a VAE network to encode the latent variables Z and decode the price movements y from a variable space of time X. However, price movements here are not used as the desired final output. Instead, they are used as an intermediate result to infer a more valuable prediction: market signals that will help make decisions about buying, holding, or selling stocks. The market signal module is then used as input to an attention mechanism that allows the neural network to focus on relevant features. The market signal extractor is implemented as a recurrent neural network with gated recurrent unit cells to extract information from price curves and decode the market signal. The encoder in this case is the mean and variance of the market signals, which are then used as input to the decoder. Following the VAE principle, prior and posterior are fit using neural approximation and reparameterization [117,120]. The final output is a vector of importance-weighted targets related to what market signals should be considered for buying and selling.

Gunduz [74] predicts Istanbul Stock Exchange stocks using an attention-based LSTM that predicts the price rise and fall of the next hour. In this paper, a VAE was used as a UQ method to reduce the dimensionality of the feature vector that feeds the LSTM model. When authors use UQ in the intermediate process to obtain a price output, as in this case, the prediction that emerges out from the model latently includes uncertainty, even if it is not visible in the prediction. The VAE in this article reduces a feature vector of dimension 65 to 15. The probabilistic approach is solved in a latent space when using VAEs and therefore the benefit of taking uncertainty into account becomes blurred for traders, since they cannot see the risk of trusting the prediction.

VAEs are included in this study because they are extensively recognized UQ methods in the existing literature. However, they might not be the most suitable UQs for financial use if a trader wants to take risk into account. A possible breakthrough to this application with VAEs would be to apply Montecarlo methods during the decoding phase to sample the posterior and propagate the uncertainty from the latent space to the prediction space. However, as for the best of our knowledge, this approach is still untested and remains as a potential way of future exploration.

3.1.6. Bayesian active learning (BAL)

Active Learning (AL) methods can be understood as a mechanism to improve the way models are trained. AL 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 computing time for training [121]. 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. Nevertheless, not all data points within the distribution provide the most information. Therefore, we can let the AL mechanism decide which points the model has the most uncertainty and pick them. Then, an *oracle* (as the standard literature calls it) provides the label to the chosen and most uncertain data points [122]. 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, no article has yet applied BAL to a financial problem like the one defined in this study.

3.1.7. Bayes by backprop (BBB)

Blundell first proposed BBB in his paper *Weight uncertainty in neural networks* [123]. The idea behind it lies again in the approximation of the posterior distribution previously defined in Eq. (6). Once again we use KL divergence in Eq. (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 leads us to the ELBO (recall Eq. (12)). Taking advantage of Blundell's first proposition that claims that "under certain conditions, the derivative of an expectation can be expressed as the expectation of a derivative", we generalize the Gaussian reparameterization trick [124]. To turn intractability into something tractable, 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

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

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

Although this technique has been used as a way to quantify uncertainty, at the time of writing this paper, and to our knowledge, no author has applied the technique to financial time series forecasting.

3.1.8. Deep Gaussian processes (DGP)

Uncertainty in DL can be expressed as a function of the existence of infinite possible models that explain an observed data pattern. When defining BNN in Section 3.1.1, it has been said that the uncertainty can be modeled around a series of neural network weights w through the prior $p(w)$ to arrive at a definition of the posterior $p(y|w)$. That Bayesian perspective implies that a constraint is applied on the infinite parameters w that a neural network can take to explain the observed data, leading to Eq. (6). Let us consider another approach to defining a Gaussian process [125]. 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

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

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 0 mean 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 basis functions ϕ, one can choose a $K = k(\cdot,\cdot)$ such that $\varphi(\cdot)$ is infinite dimensional. 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:

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

The DGP method is used in [55] to predict 95% confidence intervals of the Nikei 225 Index and the Hang Seng Index. The radial basis is at the core of the kernel function and its width ranges from 0.01 to 10000. In this case, DGP is applied as a module added at the end of a point prediction (driven by a recurrent neural network). The DGP ingests the point prediction and the confidence interval for training is calculated via MWP_{95} and MC_{95}.

3.1.9. Laplace approximations (LA)

The Laplace approximation uses the first-order Taylor series to fit a Gaussian distribution around the Maximum A Posteriori (MAP) as the center, defined as θ^*, to approximate the posterior distribution

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

To the best of our knowledge, as of writing this survey, no record has applied the Laplace approximation to financial time series prediction.

3.1.10. Deep ensembles (DE)

Deep ensembles can be of two kinds. The first is to retrain the same DL model multiple times with different initial conditions and then take the average of all those diverse behaviors [107]. These are called *boosting methods* and are adjusted sequentially. The second type of DE is called randomization-based because it uses decision trees, as is the case with random forests. As a result of both approaches, one has a committee of models that hopefully focuses on different basins of the true posterior representing the entire distribution by sampling. As the number of samples increases, the group of models collapses into a single model [126].

In other cases, like in the article [10], the models included in the ensemble do not share the same architecture. The ensemble is composed of MLP, LSTM, GPR, Lasso and BILSTM models. They are applied to the prediction of daily and weekly crude oil price. The uncertainty is calculated as a combination of an upper and lower bound for both upper and daily crude oil prices at confidence levels of 90% and 99%. To train each of the ensemble models, the article analyzes the density distribution stable (SDF), among others. With that statistical data, the upper and lower bounds in crude oil price are calculated as

$$[L_\alpha, U_\alpha] = [\tilde{T} - SDF_{\alpha/2}\sqrt{\text{var}(e)}, \tilde{T} + SDF_{\alpha/2}\sqrt{\text{var}(e)}], \tag{29}$$

where $\tilde{T}$ is the predicted value and $var(e)$ is the variance of the fitting error.

3.1.11. Conclusions on the type of model

Conclusions on the analysis by model type are defined as follows:

- Applying directly the Bayesian regularization to shape the posterior distribution in BNNs can be computationally expensive. Consequently, some methods such as VI, MC or MCMC arise to sample the posterior instead.

- Very few authors consider relating financial risk to predicted confidence intervals, as they are more interested in improving accuracy.
- Some authors, however, consider this perspective, extending the concept of predictive uncertainty to annual earnings and adding a risk consideration.
- Some methods such as MC or VI allow for the representation of predictive uncertainty, for example by obtaining the mean and variance and converting them into confidence intervals. Some other methods, such as VAE, use UQ in a latent space, which is not visible at the end of the prediction unless a different technique is applied. A method to take the uncertainty from the latent space in which a VAE operates and translate it into predictive uncertainty remains unexplored.

3.2. Analysis by financial asset

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 part of ownership in a company. Forex is the short name for Foreign Exchange Market, where investors trade currencies. A similar concept is that of cryptocurrency, with the difference that these are virtual currencies secured by cryptography and impossible to counterfeit.

3.2.1. Cryptomarket

The cryptocurrency markets are known by its high intrinsic volatility and high dependence on social sentiment [47,49] (50% of authors in this study predicting cryptocurrencies are adding twitter sentiment or google trends in the feature vector). Also, due to the small market size, big traders known as crypto whales can influence the direction the market takes. In such scenario, where trends are difficult to identify, and the uncertainty is high, one could expect that UQ acquires high relevance. Also in such volatile markets there is an opportunity to analyze the paper of uncertainty generated by data, compared to the one coming from the model, which is disentangling epistemic and aleatoric uncertainties. However, authors investigating the prediction of cryptocurrencies do not make UQ the central investigation matter of their models. Although Jang et al. [47] use a BNN their motivation is to reduce the occurrence of overfitting. They apply confidence intervals to quantify the uncertainty, but they do not mention what the coverage probability is, how it was calculated nor do they make a deeper analysis of uncertainty in cryptomarkets.

Livieris et al. [89] use a RNN with one hidden 12-neuron layer to predict the next day price for Bitcoin, Litecoin, Ripple, Ethereum, and CCi30 index. They study the model behavior with a daily close price input window of 7, 14 and 21. The model implements dropout at training and prediction time, which is the equivalent of sampling from the posterior distribution, as we have seen already. Given that they have a posterior approximation, they did have the opportunity to add confidence intervals, although they prefer to just calculate the distribution's mean only. As many other authors, they are interested in the predicted mean performance and reducing overfitting, instead of exploiting the full capabilities of the uncertainty quantification technique they use: dropout.

In an article written by Kalariya et al. [88], they propose a technique to predict next prices for Bitcoin, Litecoin and Ethereum. The financial features used to train the algorithm include Blockchain operations, open and highest day price, and sentiments like Tweets and Google Trends. It is based on perturbing the activation function during training and prediction, using a factor γ that moves the learned predicted price from its mean and that is multiplied by the magnitude of the price movement in the current day. With this mechanism, the authors are sampling a posterior. Interestingly, they use the predicted sample price, instead of the distribution mean at prediction time. They are interested on evaluating investment returns, instead of price prediction uncertainty, so they build a sequential algorithm to trade those cryptocurrencies. No probabilistic analysis or conclusions are thrown in this case.

Lin and Blum [26] are aware of the need for uncertainty quantification approach to measure the risk related to investments on seven cryptocurrencies. They build a 3-stage algorithm to predict hourly price changes, model short-term market patterns and make portfolio investment decisions. The first stage uses a recurrent Bayesian NN (BRNN), the second one uses a GAN and the third one uses a reinforcement model. The BNN ingests 60 financial technical indicators including open, close, high and low hourly historical data and predicts prices eight hours ahead. Again, the authors predicting cryptocurrencies seem not to be attracted to the idea of exploiting the power of a Bayesian neural network. They only consider means, but not confidence intervals, to quantify uncertainty. Those output prices are provided to the GAN as inputs. The GAN is built in an uncommon and interesting manner, in which the generator builds price patterns (simulated historical prices) and the discriminator chose the most common one. This approach is equivalent to choose the "mean" pattern from a sample of patterns. In this article, the sample is not composed of single prices, but instead it is made of price vectors, each one representing a pattern.

3.2.2. Foreign exchange market (Forex)

The article written by Pandey et al. [58] investigates the use of two UQ methods applied to the prediction of Forex, more precisely EUR/USD, GBP/USD and JPY/USD. They build an ensemble of models, one of which is a Bayesian NN. The approach taken for choosing the feature vector is to add fundamental and technical analysis, but not sentiment. The output of the BNN is the expectation (mean) of the exchange rate, with no variance. The output of the ensemble including the BNN is a weighted exchange rate average. The authors choose to average the outputs, however, they do not measure the prediction variability coming out from the ensemble of models.

Hassanniakalager et al. [62] are using a BNN on open, close, high and low EUR/USD, GBP/USD and JPY/USD Forex daily prices to predict the mean and variance of the next day. They do not use the variance to make buy/hold/sell decisions. Instead they use the predicted mean as an input for a series of trading strategies (Moving Averages, Support and Resistance, Channel Breakout and others). An interesting further analysis as a potential continuation of this work would be to study how the sparsity in the variance can modify the trading behavior of the trading module, for example making decisions on how much risk it is willing to accept when opening a new trading position.

A particularly interesting approach is that of Kim [63], who uses a Variational Autoencoder on Forex to smooth the south Korean Won KRN/USD. The author is making an important assumption: the smoothed curve is the real curve without the intrinsic data noise, or at least an approximated curve to the real one. In other words, the aleatoric uncertainty is deleted before a deep neural network ensemble is fed. If the smoothed curve can be considered as the approximated real curve, the estimated uncertainty at the exit after the prediction module approximately corresponds to the epistemic uncertainty. This method is considered as one of the many existing ones to disentangle both types of uncertainty.

3.2.3. Stock market

Stocks is the biggest portion of the analyzed markets. None of the authors exploring this market is explicitly mentioning a particular behavior of it, different to the rest of market types, related to the quantification of uncertainty. All articles in the study, except those mentioned in the above sections for Forex and cyrptocurrencies, are working with stocks. In order not to repeat records, those other records will be analyzed in the rest of sections.

3.2.4. Conclusions on financial asset

As a summary of the analysis based on the financial market we can draw the following conclusions:

- The cryptocurrency market is characterized by comparatively high volatility. Despite this, no article analyzes how aleatoric uncertainty is affecting the results, because no author disentangles both types of uncertainty.
- Authors who predict cryptocurrencies are the most prone to consider sentiment analysis (50%), and those who predict Forex markets are the least likely (none of them). Among the authors studied who focus on stocks, 13% apply sentiment analysis. This may be because Forex markets are affected by long-term variables, such as political decisions, long-term traders, and big size of markets. On the other hand, cryptocurrencies are made out of many small traders that can be very easily affected by breaking news. How aleatoric uncertainty can be modeled based on these markets needs further exploration, as no author goes into depth on that topic.
- Many authors who perform predictions on stock market disentangle UQ into epistemic and aleatoric uncertainties. This evidence is greatly affected by the much larger number of records focused on this market, compared to Forex or cryptomarkets.
- No author explores how different markets react to UQ and makes comparisons between them. This remains a potential field to explore in the future.

3.3. Analysis by financial input

The *analysis method* comprises the fundamental, technical and sentiment subcategories. Investors perform fundamental analysis (FA) if they only analyze 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 company management style. Technical analysis (TA) means estimating the future value of an asset only by looking at past values or patterns of values. Sentiment analysis (SA) focuses on extracting information from investor's will, expressed through text such as news, articles, social networks and blogs. With Natural Language Processing (NLP), SA is gaining popularity and is being used more frequently in investing. FA, TA and SA can be used alone or in combination.

3.3.1. Fundamental analysis

In the article by Pandey et al. [58], fundamental data is added into the feature vector. The inputs include interest rate, inflation rate, country account deficits, public debts and short and long term moving average of prices. It makes sense to add the fundamental inputs, given that Forex markets are highly dependent on macroeconomic variables. However, although the implemented predictor is a deep ensemble, the authors choose to get a weighted average, but no confidence intervals to quantify the uncertainty. The "committee machine" applies deterministic weights to the weighted average: although they are selected from a probability method, but they do not specify which one or how it was performed. A potential field to explore this model in the future is to let the weights of each model as a Gaussian distribution in a way such as to have a sum of distributions with the objective to quantify the uncertainty.

Chauhan et al. [76] proposed an investment strategy based on predicting future fundamentals, more precisely the EBIT, as opposed to most of the articles which make final price predictions. Financial reports are a type of data more difficult to acquire, compare to prices and sometimes those are only available every year or twice a year. Financial reports depend directly on the performance of the company, instead on investors desire to buy or sell stocks. The fact that the EBIT depends on how well the company is effectively performing makes the EBIT to have a larger signal to noise ratio, compared to final price. The authors use dropout at prediction time to build a posterior distribution around the expected EBIT value.

3.3.2. Sentiment analysis

Hájek and Boháčová [97] quantified the uncertainty of sentiments when applying the term-weighting scheme tf–idf to the word type frequency in financial 10-K filing forms and feed a dropout regularized NN with those distributions. The output of the NN is a distribution characterized by the average and standard deviation sampled from running 10-fold cross-validation experiments. The authors use the model accuracy mean and standard deviation to compare the models between them. This article uses the UQ to make decisions over the model adequateness to the predictions purpose. Surprisingly, although some authors calculate confidence intervals or at least standard deviations in the predicted values, not many use the uncertainty they quantify to make financial decisions.

Another interesting point of view is proposed in the article by Xing et al. [83]. The literature generally applies a UQ perspective to a technical approach (predictive uncertainty in price data). This article analyzes the sentiment volatility instead. Authors extend a Variational Recurrent NN (VRNN) applied to stock return and volatility forecasting with the integration of social media sentiments. The goal is to learn the joint probability interactions between volatility returns and positive-negative sentiments in social media in a bidirectional way (new prices are changing sentiment and sentiments are updating prices). The difference of the proposed model with a VRNN is that the latent variables z are no longer autoregressive Gaussian distributions, but instead takes into consideration past both sentiments and price history. The final model is used for next single step prediction and outperforms state of the art VRNN and neural stochastic volatility model.

In the article by Jay et al. [49], tweet volume and Google Trends spikes are incorporated into a feature vector, together with Blockchain data and technical analysis coming from past Bitcoin History. They process together this input data to feed MLP and LSTM models. Each neuron implements a mechanism similar to the one in Kalariya [88] that stochastically moves the activation value form its current value. This is equivalent to estimating the posterior mean, if the stochastic mechanism can be approximated to a Gaussian distribution. Although they are indeed sampling the approximate posterior, they however do not include a way to recover the standard deviation from it and thus the uncertainty quantification is lost at prediction time.

Ray et al. [50] employed a Bayesian Structural Time (BST) series model in which Twitter news from Governments, Business dailies, news agencies and financial portals are included. Given a feature vector X of regressors, two elements (positive and negative) are considered in order to represent sentiments. To catch non-linear trends in the rise and fall of the stock market, they added a LSTM deep learning model fed by the residuals of the BST. The posterior distribution is calculated via Bayes Theorem and the posterior is estimated using MCMC. Although the authors are able to quantify uncertainty and they represent confidence intervals, in the final validation using MAPE, they only use the predicted mean value.

3.3.3. Technical analysis

The rest of articles in this study are based on technical analysis. For this reason, they will be analyzed in the rest of sections.

3.3.4. Conclusions on financial input

The following points can be stated to summarize the analysis by financial input:

- Most authors perform a technical analysis, while very few do fundamental or sentiment analysis.
- Those who include fundamental or sentiment along with technical analysis include those parameters together in the feature vector.
- Only Hajek [97] analyzes the aleatoric uncertainty from the sentiment analysis.
- No author analyzes aleatoric uncertainty from a fundamental analysis perspective.

3.4. Analysis by prediction space

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's output considers a rise, fall or stay in the price. An even more interesting space from a UQ perspective are confidence intervals, associated with the probability of the predictor variable is within the predicted range.

3.4.1. Single value multi-step ahead

Primasiwi et al. [51] acquire open, close, high and low data from the Indonesian Stock Exchange to feed a NARX and a NN model and they use Bayesian regularization with the aim to have improved generalization abilities, but they do not exploit the UQ power of it. Instead of predicting the next day price, they predict the future 5th day, with no uncertainty associated to it.

3.4.2. Multi-value multi-step-ahead

It is rare in the investigated literature that an author applies a multi-step-ahead strategy and uncertainty quantification in the same study. The author Alghamdi et al. [72] use both strategies but separated. They run 1000 times a temporal encoder–decoder that uses a hybrid CNN-LSTM using MC dropout during inference time to predict the next single step close price on Apple and Amazon stocks. To quantify uncertainty they use 90% and 95% confidence intervals extracted from the sample of size 1000. Unfortunately the authors do not estimate uncertainty in the multi-step approach.

In other study made by Dixon [86], they develop a similar strategy, with Bayesian RNN and variational inference to predict 1 to 5 ahead close prices from IBM stocks. In this case, opposed to Alghamdi, they apply UQ to the multi-step prediction, including the close price sequence inside 90% and 95% coverage probabilities. This is the only study combining both multi-step as prediction space and uncertainty quantification together.

3.4.3. Trend (up/down/stay)

Maeda et al. [57] proposed a UQ-based framework to trade Tokyo Stock Exchange in which the explanatory variable is the predicted next price, and the objective variable is an up, down or stay trend related to that price. A dropout convolutional NN is used to extract patterns form five Stock price curves from TYO (Tokyo Index) and predict the next stock price. The dropout is performed during backpropagation but also during the forward pass to infer the posterior price distribution. 100 predictions predict the shape of the same sample and the objective variable is calculated as $l_{\text{pred}} = \arg\max_i \sum_k y_{k,i}$, where l_{pred} is the predicted trend. The reliability of such method is low, and thus the authors propose to add UQ to the framework. Adding a threshold of probability dramatically improves the precision score, and thus the prediction reliability. The conclusion then is that considering uncertainty make stock prediction more reliable and precise.

In a later work by the same authors [13] they used CNN to extract Orderbook features as sharp ratio distributions and a LSTM to extract price features running in parallel. Both outputs are merged in a feature vector feeding a sequential set of sparse variational layers. The architecture gives a vector of buy/hold/sell based on a score belonging to the interval $[-1, 1]$. Values close to -1 means that the price will fall next and a sell order is necessary. The opposite is valid for 1 and a buy order. If the score is near 0, then a hold order is performed. The output from the final layer is achieved by taking the mean value of the output distribution.

Magris et al. [19] employed a Bayesian version of a Bilinear NN to predict up, down or stay trends in tick-by-tick mid price of 5 stocks in the NASDAQ Stock Exchange, extracted from Limit-Order Book (LOB). Instead of using SGD (Stochastic Gradient Descent), RMSProp, Adam or AdaGrad, the authors use Variational Online Gauss–Newton (VOGN) to optimize the best weight distributions, and then they use the mean of them to get the predicted buy/hold/sell classes. The uncertainty is predicted performing 50 forward passes to estimate the predictive posterior. The interpretation of the results is complex, because three (non Gaussian) probability distributions, corresponding to each one of the classes, are overlapped. However, this representation of the results is much richer than only having a single point predicted class, with no information of its related uncertainty (and the related information of not chosen classes).

The focus in the article by Zhang et al. [61] is put on showcasing how uncertainty can be beneficial to predict buy/hold/sell signals on 5 stocks from the London Stock Exchange. One of the strategies they evaluate is called *Bayesian trading*, and is based on the idea of predictive entropy $\mathbb{H} = -\sum_i (p_i \cdot \log(p_i))$. The study uses the MC technique with a sample of size 100 to estimate the posterior probability of each buy, hold or sell order, and p_i is a vector with the three probabilities. They enter a trade if $\max(\bar{p}_t) > \alpha$ and exit the trade if $\mathbb{H} < \beta_2$. The uncertainty is used to upsize the position to $1.5x\mu$ if $\mathbb{H} < \beta_1$, downsize the position to $0.5x\mu$ if $\mathbb{H} > \beta_2$ or keep the original size μ if $\beta_1 < \mathbb{H} < \beta_2$. α is a probability threshold above which the trade is done; β_1 and β_2 are entropy thresholds to magnify or decrease the trade position. Thresholds are decided by grid search.

Another Bayesian approach to predict price direction on Apple, Alphabet, Facebook, Microsoft and AMZN stocks is explored by Lind et al. [80]. They sample 1000 times the prediction of a MLP to get the probability distribution of an uptrend and another 1000 for a downtrend to build a histogram that they after approximate to a Gaussian.

Hou X. et al. [75] proposed a VAE model to learn latent fundamental features from firms in the stock market and then a hybrid CNN-LSTM model to build the relationships among firms as a graph. They name their model as spatial–temporal to include the distances of edges between two nodes in a graph and also to account for sequence of events happening at the same time or with a delay between two related stocks. To include enough financial data, they operate in the range of minute trading. The prediction problem is hybrid, regression and classification, accounting for two values: the next minute price and a signal of the next price going up or down as a label in the integer range $[0, 1]$. The UQ in this article is applied to the distance between two stocks using a Variational Autoencoder. Although they apply the fully Bayesian approach, finally they only stay with the mean provided by the model and they reject to keep uncertainty measures like standard deviation. A further probabilistic analysis can be potentially made in the future if an author accounts for uncertainty in the graph edge length, which could be propagated to a probabilistic approach to the rise and fall in prices.

3.4.4. Trading order (buy/hold/sell)

Skabar [14] named the trend prediction as *direction-of-change forecasting*. That name relates to buy/hold/sell signals in the same way as Magris et al. and the other authors predicting trends. In this article, the author predicts the trend of daily close values of the Australian All Ordinaries financial index. Related to the problem of results interpretation described with Magris, Skabar states that using the Metropolis algorithm for Montecarlo integration leads to many candidates being unwittingly rejected in the sampling chain. This is due to strong correlations in the posterior probability distribution of weights when there is a decrease in $p(w/D)$. This effect is relaxed in the article by using the Hybrid Montecarlo approach with Gibbs sampling. The author runs 1000 prediction iterations to shape the posterior and then chooses 100 among them to get the probability of an upward trend in the index. Finally the probabilities of an upward trend where averaged over the 100 samples. The author's conclusion on using Bayesian methods is that they present a superior performance due to its integrative nature.

Each individual weight vector has a bias, but the integration over them reduces the bias.

Yang et al. [105] used deep ensembles in a novel way from 30 Dow Jones stocks. Instead of sampling price predictions from a group of models (which can be defined as parallel sampling), they implement an inception algorithm that chooses the best performer of the ensemble, which is equivalent to sequential sampling. The prediction is presented as a reinforced classification algorithm whose actions (buy/hold/sell) are rewarded or penalized based on the gains or losses it generates. This way of dealing with uncertainty can be seen as an adaptive alternative to deep ensembles, in which the posterior distribution is formed by a group of sample generators, but only the best performer is used each time.

3.4.5. Confidence interval

Wang et al. published a second work [56] similar to the one analyzed above [55], in which a deep ensemble is sampled to predict prices from stock market (S&P 500, DJI and Nasdaq). The difference is that in this article they include a feature extraction module using Variational Mode Decomposition (VMD) techniques. They implement a 95% confidence interval extracted after applying a Gaussian Process to the output of an LSTM point predictor module.

A different perspective on predicting confidence intervals is studied by Wang B. et al. [11]. They predict a sequence of $t + 1$ future oil prices using an autoencoder. The model implicitly learns the Brent oil price probability distribution without explicitly working with it. They state that aleatoric oil price do not follow a Gaussian distribution, and their model is consistent with non-Gaussian noise. The prediction interval learnt is inspired from another article by Pearce et al. [127]. During testing, every Prediction Interval Coverage Probability (PICP) was greater than 0.9.

The conclusions drawn from the perspective of prediction space analysis after reviewing the literature can be summarized as follows:

- Confidence intervals are the natural way of quantifying the uncertainty. The typical coverage probability that the authors are considering is 90% and 95%. A potential field of study that no authors have explored is what are the possibilities of having a down or up trend inside a coverage probability interval. If the interval size is so big as to allow for the next price to be above or below the current price, then the confidence interval is not providing a useful representation. A potential future field of study would be to find coverage probabilities that maximize the certainty about the trend, and not only the certainty about the next (mean) price.
- Predicting up/down/stay trends is equivalent to predicting buy/sell/hold trading orders. They are both related: a predicted up trend is a signal to buy the asset, a predicted down trend is a signal to sell the asset, and stay trends means that no action needs to be taken (hold). The difference is that authors who predict trends work with final price explanatory variables, while authors who work with trade orders predict investment returns. The second form is richer from a financial point of view because the explanatory variable to be maximized is the liquidity of the annual account, which is a high-order KPI in financial companies, understandable by all members. Both perspectives can benefit from UQ, although the second allows organizations to directly establish best and worst case scenarios.
- Only one article is combining multi-step prediction with uncertainty quantification. From a financial point of view, a multi-step prediction space combined with a coverage probability is the richer representation of all. It combines the risk estimation of the confidence intervals with a trend prediction. This view is especially valuable for trader, who look not only for the next price, but for a trend to start their trading opening.
- The farther we predict a price in the future, the more unsure the model is about its value. No authors have deepen into the variation of UQ related to price predictions beyond single-step. This topic remains as a potential field of study related to UQ when a multi-step perspective is adopted.

3.5. Analysis by uncertainty type

The category most associated with the definition of UQ is *epistemic/aleatoric*. 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. Statistically speaking and applied to the field of ML (Machine Learning), epistemic uncertainty is the variability of weight distributions or activation functions in a neural network. From a practical perspective, a ML model has several hyperparameters that define the model itself. One could choose a DL model with 3, 4 or 10 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 infinity number of different configurations. All this uncertainty makes the output variable more or less accurate. Epistemic uncertainty can be adjusted to reduce (i.e., narrow the probability distribution of prices) if a better model can be found. In contrast, aleatoric uncertainty is related to the intrinsic variability of noise in the data and is irreducible, meaning that financial data always comes with noise. And by default, the noise level in asset price is high. This definition of randomness in the data expresses the concept of a latent price curve that the model can never know, except through the noise that dirties its pure shape. Interestingly, the idea of a pure price pattern that will never be known directly contradicts the Efficient Market theory.

3.5.1. Aleatoric uncertainty

Following the definition of Valdenegro [128], based on the predictive uncertainty of the model, both epistemic and aleatoric uncertainties can be defined. Considering variational or ensemble model approaches in which the model generates multiple samples, each defined by the mean and variance, epistemic uncertainty is the variance of the means and aleatoric uncertainty is the mean of the variances. This definition implies that the model should learn the mean and variance from a price dataset. This perspective is adopted in articles such as that of Chauhan et al. [76], as we have seen previously. However, the methodology to meet the mean and variance labels is rarely found in the literature studied in this survey. The rest of the authors quantify only epistemic uncertainty, leaving aside aleatoric uncertainty.

Wilkins et al. [40] used Mixture Density Networks (MDN) to explicitly quantify stock market uncertainty for Fox, Warner Bros, Netflix, Disney, Amazon, and Comcast. They state that their model is capable of predicting epistemic and aleatoric uncertainty at the same time, by learning the mean and standard deviation from the closing price datasets. However, they do not explain how they do it nor do they show a disentanglement between both uncertainties. It is unknown if they use an approach similar to Valdenegro [128].

Huang et al. [12] implemented a Bayes-LSTM model run n times on different dataset samples drawn from two stocks in the Chinese Stock Market: Shanghai Index and Shenzen Index. They acquire six variables of daily features: opening, closing, high and low prices, volume and value. They predict the value of the next day's closing price. This article implements a special case of uncertainty quantification and different from the other methods in the survey. The initial price dataset starts in 1990 and ends in 2016 with daily prices. The dataset is divided into an arbitrary number of samples. The same LSTM model is trained with each different sample to have as many models as samples. They use an optimization algorithm to find the best number of samples, which is the number of samples that maximizes the set of accuracies in each model. One can think of that model as an ensemble of different models with different prediction capabilities. However, in this case, the authors are only interested in the optimal value and not in a confidence interval. This is an interesting case of aleatoric uncertainty analysis because the history window used for prediction is dynamic and depends on the variance and changing periods of the very volatile stock market.

In the work of Shen et al. [77], the authors state that most classification and regression tasks focus on improving performance, but UQ studies have not made much progress. Furthermore, among those studies that apply a UQ perspective, most consider only model uncertainty, but very few consider stochasticity caused by data noise. In fact one of the hardest questions to ask in financial trading is *what is noise?*. Traders ignore data from the price curve because they consider irrelevant to define what a price trend is. The choice of distinguishing between noise, which should be ignored, and trends, which should prompt action, is subjective and varies based on the individual interpreting the data. The article attempts to solve this problem by defining a neural network that classifies each price time window as clean or noisy using a variable $\sigma \in [0, 1]$. They then build a second neural network that predicts the next day's price from stocks in the NYSE. The second is trained with a loss function that depends on the amount of noise decided by the first. They extract opening, closing, high and low prices and trading volume for each stock and use technical indicators such as momentum and volatility as a feature vector. The model takes uncertainty into account by ingesting the mean and standard deviation and also by predicting a distribution. The authors use the model and measure the effectiveness through the annual cumulative returns, depending on the threshold set to σ. The annual profit increases by 5% compared to a framework without noise consideration.

In the article by Abrishami et al. [101], a Variational Autoencoder was used to denoise the original data features, composed of closing, opening, high, low prices and volume. The extracted data is considered high-frequency trading as prices are extracted every minute. The framework is able to predict 7-minute closing prices several steps ahead from 12 Nasdaq stock prices. However, the paper does not perform an explicit UQ analysis, since the mean and variance of the VAE remain in the latent space Z. Inherently, a model that can eliminate noise from actual price data and reveal the underlying *true* curve quantifies aleatoric noise by the same amount of noise it eliminates. Unfortunately, the authors do not perform a deeper probabilistic analysis after denoising to quantify the predictive uncertainty. Roy et al. take a similar approach in a later work [99], clearly influenced by Abrishami's team. The difference between both records is that the second takes into account top 100 Nasdaq stocks and they extend the experiment with different input and output window sizes.

The opposite strategy was implemented in the paper by Li et al. [100]. They ingest raw price data from S&P 500 (open, close, high, low and volume) and add various levels of noise to make a VAE more robust against noise. As in the case of Roy et al. Li. et al. do not look further into the probabilistic perspective of the predictive uncertainty related to the additional noise they are adding to the input data.

Montesdeoca et al. [102] proposed a framework in which a VAE is fed by 16 exogenous financial variables related to the FTSE100. The feature vector contains UK macroeconomic factors, other stock market indices and Forex rates and is defined by a matrix $V \in \mathbb{R}^{m \times n}$ with m dimensions and n data points. The goal is to approximate that space to a richer latent space that admits exogenous market factors such that $V \approx W_1 H_1 + W_2 H_2$, where $W_2 \in \mathbb{R}^{m \times r_2}$ is a matrix with constant elements produced from some external data source. In this case, the external source is a second VAE fed by input data noise. $W_1 \in \mathbb{R}^{m \times r_1}$, $H_1 \in \mathbb{R}^{r_1 \times n}$ and $H_2 \in \mathbb{R}^{r_2 \times n}$ are matrices that the first VAE needs to find. The dimensions r_1 and r_2 are the selected subspace data size and the exogenous size respectively. In summary, the authors treat the real price curve as one of many possibilities that could be found in a financial scenario. This means that the real curve is treated as just a sample of a probability distribution predicted by the model. This is equivalent to estimating noise directly from the input data, or in other words, quantifying the uncertainty of the data, which can be richer that estimating the predictive uncertainty of the data (aleatoric).

3.5.2. Epistemic uncertainty

The rest of the records not mentioned in the previous subsection focus on epistemic uncertainty and we will not review them again in this subsection.

3.5.3. Conclusions on the uncertainty type analysis

- Most authors focus their research only on epistemic uncertainty and not on aleatoric uncertainty or both.
- No records have been found among the records studied that predict epistemic uncertainty as a variance of means or aleatoric uncertainty as a mean of variances, as described by Valdenegro [128].
- In general, authors use predictive uncertainty to improve accuracy metrics, but very few authors perform an in-depth analysis on the financial implications of predictive uncertainty. For example, how should an investor interpret a confidence interval? Ideally, the actual price will fluctuate between the confidence interval with a probability similar to the coverage probability. Sometimes it will be close to the upper bound sometimes close to the lower bound. When should an investor place an order? What are the implications of placing an order based on position within the interval?
- Also, derived from the previous point, what is the maximum level of uncertainty that an investor should accept to place or not to place an order and what are the implications for the financing of an investment from a probabilistic point of view?
- Another unexplored area is what are the implications of placing orders making a difference between epistemic and aleatoric uncertainties? Can we trust the prediction and place an order if one of the uncertainties is high and the other is low? Nothing is said about the threshold of acceptable uncertainty during investment. All of these questions could potentially be explored in subsequent analyses.

3.6. Discussion

This section will discuss the results related to the research questions formulated in Section 2.

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

Table 3 lists the journals, proceedings and publishers of each one of the records described in Table 2, corresponding to articles studying UQ for DL applied to finance. 25 studies have been published as proceedings coming from congresses or conferences, 42 studies have been published in journals and finally 2 records have been published as reports for master thesis or doctoral thesis.

As can be seen in Fig. 6, the number of publications related to UQ applied to business economics has been growing and exploding since 2010. The peak is found in 2019, with a total of 389 publications. Despite the explosion, it appears that the popularity of the term applied to economics has then decreased, possibly due to a migration of research towards more popular areas such as medicine, microbiology or infectious diseases after the global Covid-19 pandemics. This effect has been seen in the Web of Science searching for those other topics: the amount of research in those other areas has grown considerably since then. In fact, the general term *Uncertainty Quantification*, applied to all possible domains, continues to gain popularity and has 2578 publications in 2021 (see Fig. 1).

China, USA and India are the three countries publishing the most papers on the subject, with 16, 12 and 10 articles, respectively. Fig. 7 shows the complete list of countries and their respective number of publications.

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

Table 3
List of all studies ordered by publisher, journal or conference paper.

Ref.	Journal/Conference	Publisher
[47]	Access	IEEE
[46]	Transactions on neural network	IEEE
[48]	Plos One	Public Library of science
[49]	Access	IEEE
[50]	Transactions on computational social systems	IEEE
[51]	Conference ICTS 2019	IEEE
[12]	Conference ICMLC 2018	Association for Computing Machinery
[52]	Conference IIAI-AAI 2019	IEEE
[53]	Expert systems with applications	Elsevier
[54]	Arxiv	Arxiv
[13]	International Journal of Smart Computing and Artificial Intelligence	IIAI
[55]	Cognitive Computation	Springer
[56]	Applied soft computing	Elsevier
[57]	Conference AISC 2020	Elsevier
[58]	Journal of King Saud University - Computer and Information Sciences	Elsevier
[59]	Neural computing and applications	Springer
[60]	Quantitative finance	Taylor and Francis Online
[61]	Conference Neurips 2018	Arxiv
[62]	Journal of empirical finance	Elsevier
[63]	SSRN	Elsevier
[64]	Financial Innovation	Springer
[11]	Neurocomputing	Science Direct
[65]	Journal of Risk and financial Management	MDPI
[66]	Plos One	Public library of science
[67]	Conference WCECS 2017	IAENG
[68]	International Journal of Performability Engineering	Totem
[69]	Conference NCCS 2019	Springer
[70]	Conference ISKE 2012	Springer
[71]	Conference SCI 2018	Springer
[72]	Conference IJCNN 2021	IEEEXplore
[19]	Arxiv	Arxiv
[14]	Advances in Electrical Engineering and Computational Science	Springer
[73]	Academic Commons	University of Columbia
[74]	Financial Innovation	Springer
[75]	IEEE/CAA Journal of Automatica Sinica	IEEEXplore
[40]	Arxiv	Arxiv
[76]	Conference PMLR 2020	Arxiv
[77]	Conference ICAI 2018	Springer
[78]	Conference PMLR 2015	JMLR
[79]	Repository	Pontificia Universidad Católica de Chile
[80]	Repository	Uppsala Universitet
[81]	Conference 56th Annual Meeting of the Association for Computational Linguistics	ACL
[82]	Conference ICMLA 2021	IEEEXplore
[83]	Knowledge-based systems	Science Direct
[84]	Expert Systems with Applications	Science Direct
[85]	Computer science	PeerJ
[26]	Conference Spring Simulation 2020	IEEEXplore
[86]	Technometrics	TANDF
[87]	International Journal of Production Research	TANDF
[88]	Mathematics	MDPI
[89]	Evolving Systems	Springer
[90]	Conference ICDAM	Springer
[91]	Conference CEC 2020	IEEEXplore
[92]	Digital Signal Processing	Science Direct
[93]	Conference BlockSys 2022	Springer
[94]	Conference EANN 2019	Springer
[95]	Conference DCABES 2018	IEEEXplore
[96]	Neural Computing and Applications	Springer
[97]	Neural Computing and Applications	Springer
[98]	2018 Conference AAAI 2018	Arxiv
[99]	AI Communications	IOS Press
[100]	Access	IEEEXplore
[101]	Conference ICTAI 2019	IEEEXplore
[102]	Conference ISPA 2019	IEEEXplore
[103]	Conference PMLR 2016	JMLR
[10]	Applied Soft Computing Journal	Elsevier
[104]	Digital Signal Processing	Elsevier
[105]	Conference ICONIP 2020	Elsevier
[106]	Applied Sciences	MDPI

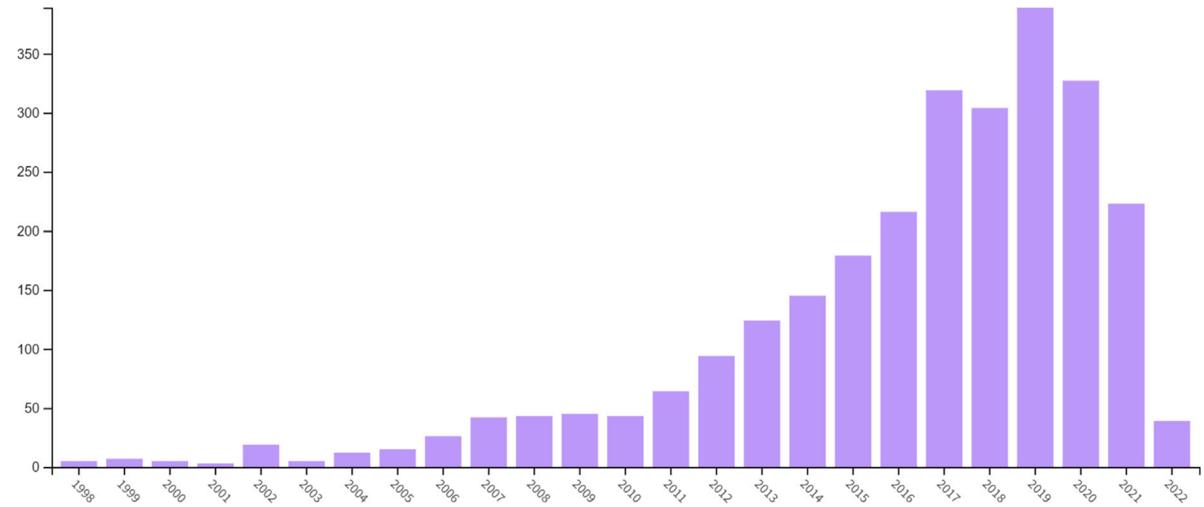

Fig. 6. Number of publications per year related to UQ applied to business economics. Source: Web of Science.

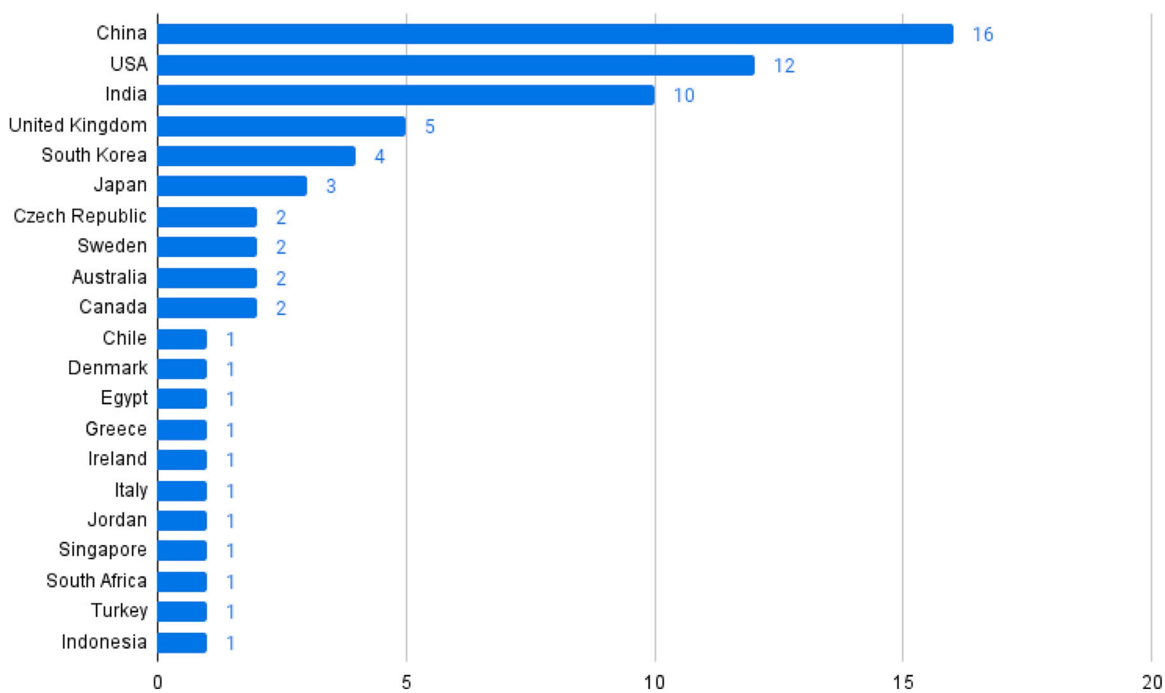

Fig. 7. Number of records by country.

Furthermore, one can observe in Table 2 how researchers focus mainly on epistemic uncertainty (57 records), while only 12 are based on aleatoric uncertainty. No records have been found that explicitly use the combination of epistemic and aleatoric uncertainty. As summarized in Fig. 9, a large portion of records (57) look for patterns in historical data (technical analysis), seven combine technical and sentiment analysis approaches, three use only a sentiment analysis and two combine technical and fundamental approaches. The authors of this survey did not find any articles that use all three methodologies (technical, sentiment and fundamental) at the same time, and this could inspire researchers to explore this path.

Something that could potentially and a priori be a surprise is that most of the papers focus on a single step approach (Fig. 10), although they use UQ methods to predict, and can take advantage of interval prediction space representations. The authors are already taking advantage of more informationally efficient models, i.e., UQ methods, rather than frequentist ones, and have in-hand distribution of information, rather than a single point. One might wonder why the authors settle for a mean value instead of using a more robust prediction space, given that they have already calculated the space. In other words, why use optimization methods after finding the uncertainty space? There are possibly many answers to those questions. Firstly, some authors, as we stated before, do not intentionally use UQ methods, but rather use it as a way to generalize, as in the case of dropout. Another reason could be that not many other authors use confidence intervals in prediction (only eight), which restricts the number of records to compare their works. In any case, it appears that interval prediction is barely explored in financial time series and could be a potential avenue for future work.

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

These have been defined in Section 3.1: Bayesian neural networks and some associated methods for approximating the true posterior, i.e., Montecarlo dropout or Markov chain Montecarlo or variational inference. Also variational autoencoders, Bayesian active learning, Bayes

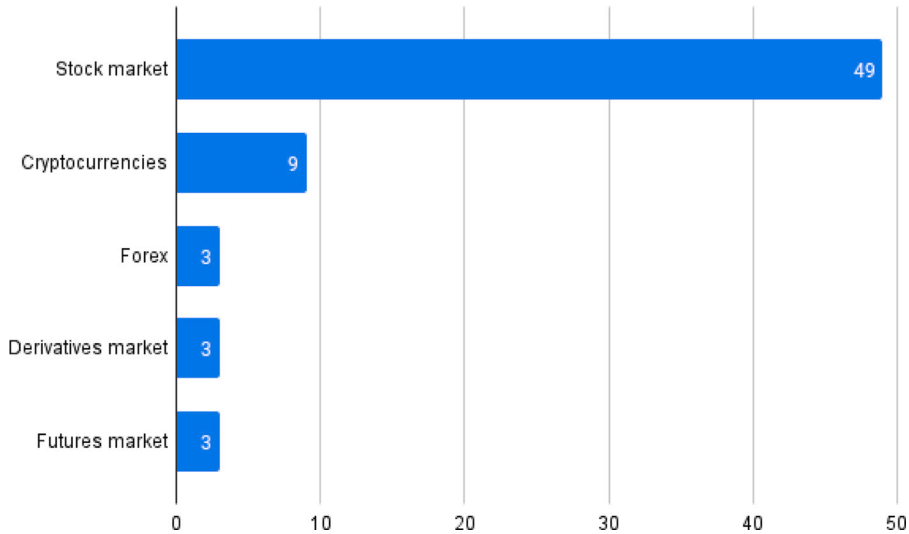

Fig. 8. Number of records by type of financial asset.

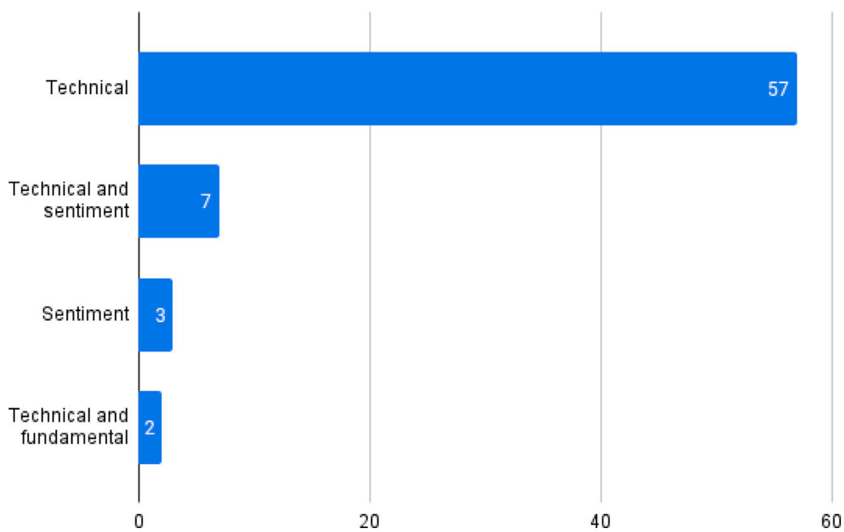

Fig. 9. Number of records by type of analysis.

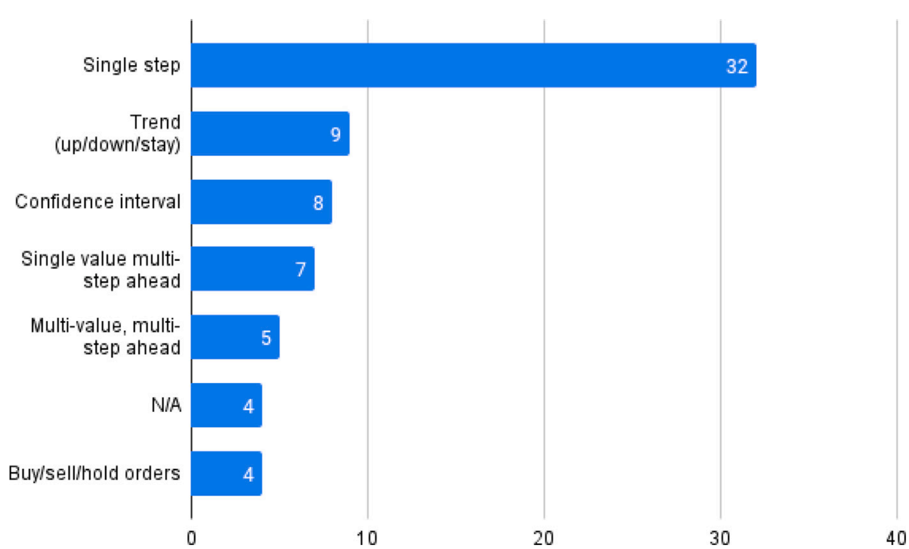

Fig. 10. Number of records by type of predicted space.

by backprop, deep Gaussian processes, Laplace approximation or deep ensembles. We have seen that Bayesian active learning and Bayes by backprop are not sufficiently explored techniques and no records of them have been found, which can be a potential field for future research.

What financial forecasting needs are still not met at UQ in DL?

The initial hypothesis that triggered the idea of tailoring this survey 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-like, then the most probable future price is not really valuable. In this situation an investor should decide to hold operations until the variance is low and the probability distribution is thinner. A similar situation would be found if the predictive probability distribution is modal, which is synonym to having peaks and valleys. The investor would probably be confused and not know what decision to make. The authors of this review consider that a further research is necessary to understand the implications of the variance shape in order to relate shapes with investment decisions and profit.

An overwhelming majority of studies focus on next single step prediction (34 out of 69). Although this approach could make sense on a big time scale, such as weeks or months, it does not represent the actual needs of investing. As it was said before, the financial time series present a strong noise component, and thus the next single point prediction does not supply much value if the delta price is smaller than the variance, because all that it is representing 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. Nevertheless, only eight articles predict trends and five work with multi-valued prediction vectors. A better approach would be to relate buy/sell/hold positions with UQ. Although four studies center on buy/sell/hold approach, none sew a clear relationship with UQ. Consequently, we consider that there are many opportunities for taking further the state of the art related to the prediction vector, including spinning around UQ.

A good investor considers the three available sources of information: technical, fundamental and sentiments. This review has shown how researchers are very attracted to technical analyses (57 records relate to pure technical approaches), but not that much to sentiment nor fundamental (see Fig. 9). Only two records use the fundamental approach, combined with the technical one. Ten records focus on sentiment analysis combined with a technical approach. We find here an exciting 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 is, by a vast difference, the biggest investment market which volume is 700 billion US dollars per day. This compares and contrasts with smaller markets like stocks (200 billion US dollars) or futures(30 billion US dollars) [129]. However, only three records have been found related to the prediction of international currencies. The Forex market supplies a few advantages in some areas, like the amount of time the Forex market is open for trading. From Monday opening in Australia to Friday closing in New York, there are five days of non-stop trading 24 hours a day. That positively impacts the amount of available data. Also, there are no commissions and the transaction costs are low. Because of the market size, it is always liquid and transactions are executed instantly [130]. This market is less popular for researchers, but it shows a potential to be investigated further.

The epistemic uncertainty is related to the lack of knowledge in the model. The found studies focus on the variability related to the neural network's weights, but other parameters could also be put under the light of evaluation. A researcher decides, as empirically as possible, the architecture of the network, but that decision leaves aside other possible configurations that are part of the epistemic uncertainty. Gençay [46] considers exactly this aspect when studying different network architecture configurations and it could potentially be carried out further in future studies, possibly varying the length and width of the neural net, choosing different activation functions, different training iterations or batches.

One of the more important aspects that the authors of this review have identified is that all the studies are based on a limited amount of data and trained accordingly, although the real and professional scenarios use updated streaming models along their functioning. The real models continuously adapt the training parameters to new and unseen data, which certainly impacts in scaling the training data. A possible research on data streaming could be exploring an unsupervised approach to price prediction.

As Shen [77] states: "Too many studies are focused on model uncertainty, but very 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 [5] ably describes this as a necessary discrimination between a predicted probability score and its related prediction uncertainty. One for example, can be very confident that tomorrow will not rain with a 10% probability, but very unsure that the probability of rain the next week will be 50%. This uncertainty comes from the lack of knowledge, as the opposite of the probability confidence interval, coming from noise in data. Indeed, future studies can explore further the combination of epistemic and aleatoric uncertainty.

What technical challenges remain in financial prediction using UQ in 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 especially sensitive to external and not completely known factors [11–13,52,55,72,81]. Magris et al. [19] consider the inability of ML methods to address uncertainties in financial applications as a major drawback and a major concern in econometrics. Due to this dynamic behavior, the authors believe that point estimation is not optimal for predicting financial time series. Instead, UQ is a much better approach to predict trends, offering plenty of opportunities to dig deeper. Wilkins [40] states that some UQ methods have been proposed in the past years; however, they fail when applied to large and noisy datasets. In fact, there is a limit to reducing the total uncertainty, defined by the aleatoric uncertainty. Even if we hypothetically managed to reduce epistemic uncertainty to zero, there will always be uncertainty intrinsic to the data. Researchers typically try to extract the function hidden beneath all the stochasticity; however, it is very common to find in the literature that the chosen feature space does not capture all the information it could from stochastic data. As a consequence, the prediction of a target, i.e., price, trend change, or buy/sell orders, is not optimal [40].

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

What approaches could potentially be explored to overcome those challenges?

Some approaches have already been proposed previously, which can be summarized as follows:

1. Increase the efficiency of extracting information from the feature space.
2. Explore a more meaningful feature space that includes fundamental, sentiment, and technical information combined.
3. Develop newer methods to represent uncertainty in a richer way.
4. Delve into trend prediction as an alternative to point estimation.

5. Consider alternative markets to stocks such as Forex, cryptocurrencies, futures or derivatives.
6. Explore newer methods to approximate the true posterior distribution.

4. Conclusions

In this survey, a PRISMA approach was followed to review 69 records on UQ in DL applied to financial time series forecasting. We have analyzed some of the most distinctive aspects of such studies: the type of asset, the techniques used, the forecast space, the analysis method and the epistemic versus aleatoric 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 foreign exchange market and finding a better way to address aleatoric uncertainty. As a conclusion of this survey, we can state that there is a lot of room for future research on UQ for financial time series prediction.

CRediT authorship contribution statement

Txus Blasco: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **J. Salvador Sánchez:** Supervision, Validation, Writing – review & editing. **Vicente García:** Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] T. Ma, P. Fraser-Mackenzie, M. Sung, A. Kansara, J. Johnson, Are the least successful traders those most likely to exit the market? A survival analysis contribution to the efficient market debate, European J. Oper. Res. 299 (1) (2022) 330–345, http://dx.doi.org/10.1016/j.ejor.2021.08.050.

[2] S.C. Hora, Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management, Reliab. Eng. Syst. Saf. 54 (1996) 217–223, http://dx.doi.org/10.1016/S0951-8320(96)00077-4.

[3] A.D. Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter? Struct. Saf. 31 (2009) 105–112, http://dx.doi.org/10.1016/J.STRUSAFE.2008.06.020.

[4] W. Kenton, Noise, 2022, https://www.investopedia.com/terms/n/noise.asp.

[5] E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Mach. Learn. 10 (3) (2021) 457–506, http://dx.doi.org/10.1007/s10994-021-05946-3.

[6] S. Gupta, A. Gupta, Dealing with noise problem in machine learning datasets: A systematic review, Procedia Comput. Sci. 161 (2019) 466–474, http://dx.doi.org/10.1016/J.PROCS.2019.11.146.

[7] D.J.C. Mackay, Bayesian interpolation, in: Maximum Entropy and Bayesian Methods. Fundamental Theories of Physics, vol. 50, Springer, 1992, pp. 39–66, http://dx.doi.org/10.1007/978-94-017-2219-3-3.

[8] M. Nardino, F.F. Silva, T. Olivoto, W.S. Barros, C.G. de Carvalho, V.S. Signorini, H.C. Mezzomo, C.R. Casagrande, A.P. Rolfs, Meta-analysis of the experimental coefficient of variation in wheat using the Bayesian, Sci. Agric. 80 (2023) e20210190, http://dx.doi.org/10.1590/1678-992X-2021-0190.

[9] J. Pixley, The use of risk in understanding financial decisions and institutional uncertainty, J. Socio-Econ. 39 (2) (2010) 209–222, http://dx.doi.org/10.1016/j.socec.2009.12.002.

[10] J. Wang, T. Niu, P. Du, W. Yang, Ensemble probabilistic prediction approach for modeling uncertainty in crude oil price, Appl. Soft Comput. 95 (2020) 106509, http://dx.doi.org/10.1016/J.ASOC.2020.106509.

[11] B. Wang, T. Li, Z. Yan, G. Zhang, J. Lu, DeepPIPE: A distribution-free uncertainty quantification approach for time series forecasting, Neurocomputing 397 (2020) 11–19, http://dx.doi.org/10.1016/j.neucom.2020.01.111.

[12] B. Huang, Q. Ding, G.-Z. Sun, H. Li, Stock prediction based on Bayesian-LSTM, in: ACM International Conference Proceeding Series, Association of Computer Machinery, Macau, China, 2018, pp. 128–133, http://dx.doi.org/10.1145/3195106.3195170.

[13] I. Maeda, H. Matsushima, H. Sakaji, K. Izumi, D. deGraw, A. Kato, M. Kitano, Predictive uncertainty in neural network-based financial market forecasting, Int. J. Smart Comput. Artif. Intell. 5 (1) (2021) 1–18, http://dx.doi.org/10.52731/ijscai.v5.i1.541.

[14] A.A. Skabar, Direction-of-change financial time series forecasting using neural networks: A Bayesian approach, in: Advances in Electrical Engineering and Computational Science, Springer Netherlands, 2009, pp. 515–524, http://dx.doi.org/10.1007/978-90-481-2311-744.

[15] J.H.A. Guillaume, C. Helgeson, S. Elsawah, A.J. Jakeman, M. Kummu, Toward best practice framing of uncertainty in scientific publications: A review of water resources research abstracts, Water Resour. Res. 53 (8) (2017) 6744–6762, http://dx.doi.org/10.1002/2017WR020609.

[16] S. Han, P. Coulibaly, Bayesian flood forecasting methods: A review, J. Hydrol. 551 (2017) 340–351, http://dx.doi.org/10.1016/j.jhydrol.2017.06.004.

[17] S. Sacchelli, S. Fabbrizzi, S. Menghini, Climate change effects and adaptation strategies in the wine sector: a quantitative literature review, Wine Econ. Policy 5 (2) (2016) 114–126, http://dx.doi.org/10.1016/j.wep.2016.08.001.

[18] O. Schwabe, E. Shehab, J. Erkoyuncu, Uncertainty quantification metrics for whole product life cycle cost estimates in aerospace innovation, Prog. Aerosp. Sci. 77 (2015) 1–24.

[19] M. Magris, M. Shabani, A. Iosifidis, Bayesian bilinear neural network for predicting the mid-price dynamics in limit-order book markets, J. Forecast. 42 (6) (2022) 1407–1428, http://dx.doi.org/10.1002/for.2955.

[20] D.J.C. Mackay, Probable networks and plausible predictions-a review of practical Bayesian methods for supervised neural networks, Netw., Comput. Neural Syst. 6 (1995) 469–505, http://dx.doi.org/10.1088/0954-898X-6-3-011.

[21] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U.R. Acharya, V. Makarenkov, S. Nahavandi, A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Inf. Fusion 76 (2021) 243–297, http://dx.doi.org/10.1016/j.inffus.2021.05.008.

[22] S. Mishra, B.M. Ayyub, Predicting regional economic activity using artificial intelligence (AI) methods: Case study with Indian states reliability and risk, IASSI- Quaterly 37 (2018) 330–361.

[23] J. Huang, J. Chai, S. Cho, Deep learning in finance and banking: A literature review and classification, Front. Bus. Res. China 14 (2020) 1–24, http://dx.doi.org/10.1186/s11782-020-00082-6.

[24] A. Ganti, G. Scott, Y. Pérez, Asset class definition, 2022, https://www.investopedia.com/terms/a/assetclasses.asp.

[25] K. Park, H.-G. Jung, T.-S. Eom, S.-W. Lee, Uncertainty-aware portfolio management with risk-sensitive multiagent network, IEEE Trans. Neural Netw. Learn. Syst. (2022) 1–14, http://dx.doi.org/10.1109/TNNLS.2022.3174642.

[26] Y.C. Lin, J. Blum, Feature transformation and simulation of short term price variability in reinforcement learning for portfolio management, in: Proceedings of the 2020 Spring Simulation Conference, SpringSim 2020, Institute of Electrical and Electronics Engineers Inc., online, 2020, pp. 1–10, http://dx.doi.org/10.22360/SpringSim.2020.AIS.002.

[27] A.S. Saud, S. Shakya, Evaluation of weight decay regularization techniques for stock price prediction using gated recurrent unit network, Nepal J. Sci. Technol. 19 (2021) 9–15, http://dx.doi.org/10.3126/njst.v20i1.39379.

[28] W. Wu, Y. Wang, J. Fu, J. Yan, Z. Wang, J. Liu, W. Wang, X. Wang, Preliminary study on interpreting stock price forecasting based on tree regularization of GRU, in: International Conference of Pioneering Computer Scientists, Engineers and Educators, Vol. 1059, ICPCSEE, Springer Verlag, 2019, pp. 476–487, http://dx.doi.org/10.1007/978-981-15-0121-0-37.

[29] F. Xu, S. Tan, Deep learning with multiple scale attention and direction regularization for asset price prediction, Expert Syst. Appl. 186 (2021) 115796, http://dx.doi.org/10.1016/j.eswa.2021.115796.

[30] P. Tabassum, M. Halder, Stock Price Forecasting Using Bayesian Network (Ph.D. thesis), Department of Computer Science and Engineering, BRAC University, Bangladesh, 2018.

[31] E. KI.T.A., M. Harada, T. Mizuno, Application of Bayesian network to stock price prediction, Artif. Intell. Res. 1 (2012) 171–184, http://dx.doi.org/10.5430/air.v1n2p171.

[32] L.S. Malagrino, N.T. Roman, A.M. Monteiro, Forecasting stock market index daily direction: A Bayesian network approach, Expert Syst. Appl. 105 (2018) 11–22, http://dx.doi.org/10.1016/j.eswa.2018.03.039.

[33] C.S. Tan, C.K. Yong, Y.H. Tay, Modeling financial ratios of Malaysian plantation stocks using Bayesian networks, in: 2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology, STUDENT, 1, Kuala Lumpur, Malaysia, 2012, pp. 7–12, http://dx.doi.org/10.1109/STUDENT.2012.6408352.

[34] Q. Xiao, C. Chaoqin, Z. Li, Time series prediction using dynamic Bayesian network, Optik 135 (2017) 98–103, http://dx.doi.org/10.1016/j.ijleo.2017.01.073.

[35] Y. Zuo, M. Harada, T. Mizuno, E. Kita, Bayesian network based prediction algorithm of stock price return, in: Proceedings of the 4th International Conference on Intelligent Decision Technologies - IDT, Vol. 1, Springer, Zilina, Slovakia, 2012, pp. 397–406.

[36] Y. Zuo, E. Kita, Up/Down analysis of stock index by using Bayesian network, Eng. Manag. Res. 1 (2012) 46–52, http://dx.doi.org/10.5539/emr.v1n2p46.

[37] I. Alarab, S. Prakoonwit, M.I. Nacer, Illustrative discussion of MC-dropout in general dataset: Uncertainty estimation in bitcoin, Neural Process. Lett. 53 (2021) 1001–1011, http://dx.doi.org/10.1007/s11063-021-10424-x.

[38] Z.H. Kilimci, Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction, J. Fac. Eng. Archit. Gazi Univ. 35 (2020) 635–650, http://dx.doi.org/10.17341/gazimmfd.501551.

[39] W.Y. Lin Peiguang, SCONV: A financial market trend forecast method based on emotional analysis, J. Comput. Res. Devel. 57 (8) (2020) 1769–1778, http://dx.doi.org/10.7544/issn1000-1239.2020.20200494.

[40] N. Wilkins, M. Johnson, I. Nwogu, Regression with uncertainty quantification in large scale complex data, in: 2022 IEEE International Conference on Systems, Man, and Cybernetics, SMC, Prague, Czech Republic, 2022, pp. 827–833, http://dx.doi.org/10.1109/SMC53654.2022.9945367.

[41] C. Liu, H. Wang, Stock recommendation based on depth BRNN and bi-LSTM, in: Proceedings - 4th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2019, IEEE, Hohhot, China, 2019, pp. 751–755, http://dx.doi.org/10.1109/ICMCCE48743.2019.00173.

[42] H.-Y. Chen, S.-H. Huang, Generating a trading strategy in the financial market from sensitive expert data based on the privacy-preserving generative adversarial imitation network, Neurocomputing 500 (2022) 616–631, http://dx.doi.org/10.1016/J.NEUCOM.2022.05.039.

[43] M. Dogariu, L. daniel Ştefan, B.A. Boteanu, C. Lamba, B. Kim, L.-D. Ştefan, B. Ionescu, L. d Ştefan, B.A. Boteanu, C. Lamba, B. Kim, Generation of realistic synthetic financial time-series, ACM Trans. Multimed. Comput. Commun. Appl. 18 (2022) 1–27, http://dx.doi.org/10.1145/3501305.

[44] Y. Huang, D.R. Liu, S.J. Lee, C.H. Hsu, Y.G. Liu, A boosting resampling method for regression based on a conditional variational autoencoder, Inform. Sci. 590 (2022) 90–105, http://dx.doi.org/10.1016/J.INS.2021.12.100.

[45] A. Singh, T. Ogunfunmi, An overview of variational autoencoders for source separation, finance, and bio-signal applications, Entropy 24 (55) (2022) 1–55, http://dx.doi.org/10.3390/e24010055.

[46] R. Gençay, M. Qi, Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging, IEEE Trans. Neural Netw. 12 (2001) 726–734, http://dx.doi.org/10.1109/72.935086.

[47] H. Jang, J. Lee, An empirical study on modeling and prediction of bitcoin prices with Bayesian neural networks based on blockchain information, IEEE Access 6 (2018) 5427–5437, http://dx.doi.org/10.1109/ACCESS.2017.2779181.

[48] H.Y. Chandra R, Bayesian neural networks for stock price forecasting before and during COVID-19 pandemic, PLoS One 16 (2021) 1–32, http://dx.doi.org/10.1371/journal.pone.0253217.

[49] P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, M. Alazab, Stochastic neural networks for cryptocurrency price prediction, IEEE Access 8 (2020) 82804–82818, http://dx.doi.org/10.1109/ACCESS.2020.2990659.

[50] P. Ray, B. Ganguli, A. 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 Trans. Comput. Soc. Syst. 8 (5) (2021) 1153–1162, http://dx.doi.org/10.1109/TCSS.2021.3073964.

[51] C. Primasiwi, R. Sarno, K.R. Sungkono, C.S. Wahyuni, Stock composite prediction using nonlinear autoregression with exogenous input (NARX), in: 2019 12th International Conference on Information Communication Technology and System, ICTS, Surabaya, Indonesia, 2019, pp. 43–48, http://dx.doi.org/10.1109/ICTS.2019.8850956.

[52] I. Maeda, H. Matsushima, H. Sakaji, K. Izumi, D. deGraw, A. Kato, M. Kitano, Effectiveness of uncertainty consideration in neural-network-based financial forecasting, in: 2019 8th International Congress on Advanced Applied Informatics, IIAI-AAI, Toyama, Japan, 2019, pp. 673–678, http://dx.doi.org/10.1109/IIAI-AAI.2019.00139.

[53] J.L. Ticknor, A Bayesian regularized artificial neural network for stock market forecasting, Expert Syst. Appl. 40 (14) (2013) 5501–5506, http://dx.doi.org/10.1016/j.eswa.2013.04.013.

[54] M. Pires, T. Marwala, Option pricing using Bayesian neural networks, 2007, http://dx.doi.org/10.48550/arXiv.0705.1680, ArXiv.

[55] J. Wang, L. Feng, Y. Li, J. He, C. Feng, Deep nonlinear ensemble framework for stock index forecasting and uncertainty analysis, Cogn. Comput. 13 (2021) 1–19, http://dx.doi.org/10.1007/s12559-021-09961-3.

[56] J. Wang, J. He, C. Feng, L. Feng, Y. 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, Appl. Soft Comput. 113 (2021) 1–17, http://dx.doi.org/10.1016/j.asoc.2021.107898.

[57] I. Maeda, H. Matsushima, H. Sakaji, K. Izumi, D. deGraw, H. Tomioka, A. Kato, M. Kitano, Learning uncertainty in market trend forecast using Bayesian neural networks, in: E. Bucciarelli, S.-H. Chen, J.M. Corchado (Eds.), Decision Economics: Complexity of Decisions and Decisions for Complexity, Springer International Publishing, Cham, 2020, pp. 210–218.

[58] T.N. Pandey, A.K. Jagadev, S. Dehuri, S.-B. Cho, A novel committee machine and reviews of neural network and statistical models for currency exchange rate prediction: An experimental analysis, J. King Saud Univ. - Comput. Inform. Sci. 32 (9) (2020) 987–999, http://dx.doi.org/10.1016/j.jksuci.2018.02.010.

[59] E. Houssein, M. Dirar, K. Hussain, W. Makram Mohamed, Assess deep learning models for Egyptian exchange prediction using nonlinear artificial neural networks, Neural Comput. Appl. (2020) 5965–5987, http://dx.doi.org/10.1007/s00521-020-05374-9.

[60] H. Jang, J. Lee, Generative Bayesian neural network model for risk-neutral pricing of American index options, Quant. Finance 19 (2018) 1–17, http://dx.doi.org/10.1080/14697688.2018.1490807.

[61] Z. Zhang, S. Zohren, S. Roberts, BDLOB: Bayesian deep convolutional neural networks for limit order books, in: Neurips-2018, Montreal, Canada, 2018, pp. 1–6.

[62] A. Hassanniakalager, G. Sermpinis, C. Stasinakis, Trading the foreign exchange market with technical analysis and Bayesian statistics, J. Empir. Financ. 63 (2021) 230–251, http://dx.doi.org/10.1016/j.jempfin.2021.07.006.

[63] S. Kim, Macroeconomic and financial market analyses and predictions through deep learning, SSRN Electron. J. (2020) 1–47, http://dx.doi.org/10.2139/ssrn.3684936.

[64] D. Selvamuthu, V. Kumar, A. Mishra, Indian stock market prediction using artificial neural networks on tick data, Financ. Innov. 5 (2019) 5–16, http://dx.doi.org/10.1186/s40854-019-0131-7.

[65] H.F. Calvo-Pardo, T. Mancini, J. Olmo, Neural network models for empirical finance, J. Risk Financ. Manag. 13 (11) (2020) 1–22, http://dx.doi.org/10.3390/jrfm13110265.

[66] A. Back, W. Keith, Bayesian Neural Networks for Financial Asset Forecasting (Master's thesis), KTH, Mathematical Statistics, Stockholm, 2019, pp. 1–82.

[67] Q. Al-Shayea, Neural networks to predict stock market price, in: World Congress on Engineering and Computer Science WCECS, Vol. 1, Newswood Limited, San Francisco, USA, 2017, pp. 1000–1007.

[68] Y. Hou, B. Xie, H. Liu, Bayesian regularization neural network model for stock time series prediction, Int. J. Perform. Eng. 15 (12) (2019) 3271–3278, http://dx.doi.org/10.23940/ijpe.19.12.p19.32713278.

[69] S. Garg, Forecasting of Gold Prices using Bayesian Regularization Neural Network, in: Nanoelectronics, Circuits and Communication Systems. Lecture Notes in Electrical Engineering, vol. 692, Springer, Singapore, 2020, pp. 147–153.

[70] Q. Sun, W.G. Che, H.L. Wang, Bayesian regularization BP neural network model for the stock price prediction, Found. Appl. Intell. Syst. Adv. Intell. Syst. Comput. 213 (2014) 521–531, http://dx.doi.org/10.1007/978-3-642-37829-4-45.

[71] A. Priya, S. Garg, A comparison of prediction capabilities of Bayesian regularization and levenberg–marquardt training algorithms for cryptocurrencies, Smart Intell. Comput. Appl. 1 (2020) 657–664, http://dx.doi.org/10.1007/978-981-13-9282-5-62.

[72] D. Alghamdi, F. Alotaibi, J. Rajgopal, A novel hybrid deep learning model for stock price forecasting, in: 2021 International Joint Conference on Neural Networks, IJCNN, IEEE, 2021, pp. 1–8, http://dx.doi.org/10.1109/IJCNN52387.2021.9533553, Online.

[73] M. Wang, Essays on the applications of machine learning in financial markets, 2021, http://dx.doi.org/10.7916/d8-qrva-sy62.

[74] H. Gunduz, An efficient stock market prediction model using hybrid feature reduction method based on variational autoencoders and recursive feature elimination, Financ. Innov. 7 (2021) 1–24, http://dx.doi.org/10.1186/s40854-021-00243-3.

[75] X. Hou, K. Wang, C. Zhong, Z. Wei, ST-trader: A spatial-temporal deep neural network for modeling stock market movement, J. Autom. Sin. 8 (2021) 1015–1024, http://dx.doi.org/10.1109/JAS.2021.1003976.

[76] L. Chauhan, J. Alberg, Z.C. Lipton, Uncertainty-aware lookahead factor models for quantitative investing, in: Proceedings of Machine Learning Research, Vol. 119, PMLR, 2020, pp. 1–11, Online.

[77] B. Shen, B. Song, Uncertainty Estimation for Strong-Noise Data, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11013 LNAI, Springer Verlag, 2018, pp. 447–454, http://dx.doi.org/10.1007/978-3-319-97310-4-51.

[78] N. Wilkins, Efficient Distribution Regression (Ph.D. thesis), Rochester Institute of Technology, New York, USA, 2019.

[79] G. Cheuque, Bitcoin Price Prediction Through Stimulus Analysis: On the Footprints of Twitter's Crypto-Influencers, Pontificia Universidad Católica de Chile, 2021, pp. 1–94.

[80] G. Lind, A. Kupreyeva, A. Lindgren, (Un)certain asset price prediction using Bayesian neural networks, Intitutionen Inf. (2021) 1–38.

[81] Y. Xu, S.B. Cohen, Stock movement prediction from tweets and historical prices, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Vol. 1, Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 1970–1979.

[82] A. Garg, T. Shah, V.K. Jain, R. Sharma, CrypTop12: A dataset for cryptocurrency price movement prediction from tweets and historical prices, in: 20th International Conference on Machine Learning and Applications, ICMLA, IEEE, Pasadena, California, USA, 2021, pp. 379–384, http://dx.doi.org/10.1109/ICMLA52953.2021.00065.

[83] F.Z. Xing, E. Cambria, Y. Zhang, Sentiment-aware volatility forecasting, Knowl.-Based Syst. 176 (2019) 68–76, http://dx.doi.org/10.1016/j.knosys.2019.03.029.

[84] A.U. Haq, A. Zeb, Z. Lei, D. Zhang, Forecasting daily stock trend using multi-filter feature selection and deep learning, Expert Syst. Appl. 168 (2021) 1–8, http://dx.doi.org/10.1016/J.ESWA.2020.114444.

[85] L. Cocco, R. Tonelli, M. Marchesi, Predictions of bitcoin prices through machine learning based frameworks, PeerJ Comput. Sci. 7 (2021) 1–23, http://dx.doi.org/10.7717/peerj-cs.413.

[86] M. Dixon, Industrial forecasting with exponentially smoothed recurrent neural networks, Technometrics 64 (2022) 114–124, http://dx.doi.org/10.1080/00401706.2021.1921035.

[87] D. Yan, Q. Zhou, J. Wang, N. Zhang, Bayesian regularisation neural network based on artificial intelligence optimisation, Int. J. Prod. Res. 55 (2017) 2266–2287, http://dx.doi.org/10.1080/00207543.2016.1237785.

[88] V. Kalariya, H. Li, J. Feng, B. Niu, S. Fu, P. Parmar, P. Jay, S. Tanwar, M.S. Raboaca, F. Alqahtani, A. Tolba, B.-C. Neagu, Stochastic neural networks-based algorithmic trading for the cryptocurrency market, Mathematics 10 (2022) http://dx.doi.org/10.3390/math10091456.

[89] I.E. Livieris, S. Stavroyiannis, E. Pintelas, T. Kotsilieris, P. Pintelas, A dropout weight-constrained recurrent neural network model for forecasting the price of major cryptocurrencies and CCi30 index, Evol. Syst. 13 (2022) 85–100, http://dx.doi.org/10.1007/S12530-020-09361-2/TABLES/12.

[90] H.R. Patel, A.M. Patel, H.A. Patel, S.M. Parikh, Hyperparameter tune for neural network to improve accuracy of stock market prediction, in: Lecture Notes on Data Engineering and Communications Technologies. Proceedings of Data Analytics and Management, Vol. 90, Springer Science and Business Media Deutschland GmbH, 2022, pp. 65–76, http://dx.doi.org/10.1007/978-981-16-6289-8-7.

[91] E. Rokhsatyazdi, S. Rahnamayan, H. Amirinia, S. Ahmed, Optimizing LSTM based network for forecasting stock market, in: 2020 IEEE Congress on Evolutionary Computation, CEC, Glasgow, UK, 2020, pp. 1–7, http://dx.doi.org/10.1109/CEC48606.2020.9185545.

[92] H. Liu, Z. Long, An improved deep learning model for predicting stock market price time series, Digit. Signal Process. 102 (2020) 102741, http://dx.doi.org/10.1016/j.dsp.2020.102741.

[93] Z. Zhang, J. Zhou, Y. Song, H.-N. Dai, Cryptocurrencies price prediction using weighted memory multi-channels, in: International Conference on Blockchain and Trustworthy Systems, in: Communications in Computer and Information Science, vol. 1679, Springer, Singapore, 2020, pp. 502–516.

[94] N. Naik, B.R. Mohan, Study of stock return predictions using recurrent neural networks with LSTM, Commun. Comput. Inf. Sci. 1000 (2019) 453–459, http://dx.doi.org/10.1007/978-3-030-20257-6-39/TABLES/2.

[95] Y. Wang, Y. Liu, M. Wang, R. Liu, LSTM model optimization on stock price forecasting, in: Proceedings - 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science, DCABES 2018, Institute of Electrical and Electronics Engineers Inc., Wuxi, China, 2018, pp. 173–177, http://dx.doi.org/10.1109/DCABES.2018.00052.

[96] P. Hájek, Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns, Neural Comput. Appl. 66 (2011) 35–65, http://dx.doi.org/10.1111/j.1540-6261.2010.01625.x.

[97] P. Hájek, J. Boháčová, Predicting abnormal bank stock returns using textual analysis of annual reports – A neural network approach, Commun. Comput. Inf. Sci. 629 (2016) 67–78, http://dx.doi.org/10.1007/978-3-319-44188-7-5.

[98] R. Luo, W. Zhang, X. Xu, J. Wang, A neural stochastic volatility model, in: Proceedings on the 32nd AAAI Conference on Artificial Intelligence, Vol. 9, AAAI 2018, New Orleans, Louisiana, USA, 2018, pp. 6401–6408.

[99] A.R. Choudhury, S. Abrishami, M. Turek, P. Kumar, Enhancing profit from stock transactions using neural networks, AI Commun. 33 (2020) 75–92, http://dx.doi.org/10.3233/AIC-200629.

[100] H. Li, Y. Cui, Y. Cui, S. Wang, S. Wang, J. Liu, J. Liu, J. Qin, Y. Yang, Multivariate financial time-series prediction with certified robustness, IEEE Access 8 (2020) 109133–109143, http://dx.doi.org/10.1109/ACCESS.2020.3001287.

[101] S. Abrishami, M. Turek, A.R. Choudhury, P. Kumar, Enhancing profit by predicting stock prices using deep neural networks, in: Proceedings - International Conference on Tools with Artificial Intelligence, Vol. 2019-November, ICTAI, IEEE Computer Society, Portland, Oregon, USA, 2019, pp. 1551–1556, http://dx.doi.org/10.1109/ICTAI.2019.00223.

[102] L. Montesdeoca, S. Squires, M. Niranjan, Variational autoencoder for non-negative matrix factorization with exogenous inputs applied to financial data modelling, in: International Symposium on Image and Signal Processing and Analysis, Vol. 2019-September, ISPA, IEEE Computer Society, 2019, pp. 312–317, http://dx.doi.org/10.1109/ISPA.2019.8868930.

[103] Y. Wang, M. Brubaker, B. Chaib-Draa, R. Urtasun, Sequential inference for deep Gaussian process, in: A. Gretton, C.C. Robert (Eds.), Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, in: Proceedings of Machine Learning Research, vol. 51, PMLR, Cadiz, Spain, 2016, pp. 694–703.

[104] J. Wang, Y. Chen, S. Zhu, W. Xu, Depth feature extraction-based deep ensemble learning framework for high frequency futures price forecasting, Digit. Signal Process. 127 (2022) 1–16, http://dx.doi.org/10.1016/J.DSP.2022.103567.

[105] H. Yang, X.-Y. Liu, S. Zhong, W. Ai, R. Bellevue, W.A. Walid, A. Walid, Deep reinforcement learning for automated stock trading: An ensemble strategy, in: ICAIF'20: Proceedings of the First ACM International Conference on AI in Finance, Vol. 20, 2020, pp. 1–8, http://dx.doi.org/10.1145/3383455.3422540.

[106] I.M. Kamal, H. Bae, S. Sunghyun, H. Yun, DERN: Deep ensemble learning model for shortand long-term prediction of baltic dry index, Appl. Sci. 10 (2020) 1–16, http://dx.doi.org/10.3390/app10041504.

[107] A.G. Wilson, P. Izmailov, Deep ensembles as approximate Bayesian inference - Deep ensembles as approximate Bayesian inference, 2021, https://cims.nyu.edu/~andrewgw/deepensembles/.

[108] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms, in: Chapman & Hall/CRC Machine Learning & Pattern Recognition, Routledge, Taylor & Francis Group, Boca Raton, Florida, USA, 2012.

[109] L.V. Jospin, W. Buntine, F. Boussaid, H. Laga, M. Bennamoun, Hands-on Bayesian neural networks – a tutorial for deep learning users, IEEE Comput. Intell. Mag. 17 (2) (2022) 29–48, http://dx.doi.org/10.1109/MCI.2022.3155327.

[110] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, in: Proceedings of the 33rd International Conference on Machine Learning, Vol. 48, Association for Computing Machinery, New York, USA, 2015, pp. 1–12.

[111] C. Serpell, I. Araya, C. Valle, H. Allende, Probabilistic forecasting using Monte Carlo dropout neural networks, in: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2019, in: Lecture Notes in Computer Science, vol. 11896, La Havana, Cuba, 2019, pp. 1–11, http://dx.doi.org/10.1007/978-3-030-33904-3-36.

[112] S. Chib, E. Greenberg, Understanding the Metropolis-Hastings algorithm, Amer. Statist. 49 (1995) 327–335, http://dx.doi.org/10.2307/2684568.

[113] D.M. Blei, A. Kucukelbir, J.D. Mcauliffe, Variational inference: A review for statisticians, J. Amer. Statist. Assoc. 112 (518) (2017) 859–877, http://dx.doi.org/10.1080/01621459.2017.1285773.

[114] M.J. Wainwright, M.I. Jordan, Graphical models, exponential families, and variational inference, Found. Trends Mach. Learn. 1 (2008) 1–305, http://dx.doi.org/10.1561/2200000001.

[115] T.P. Minka, Expectation propagation for approximate Bayesian inference, in: UAI'01: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, Association for Computing Machinery, Washington, USA, 2001, pp. 362–369.

[116] J.S. Yedidia, W. Freeman, Y. Weiss, Generalized belief propagation, in: T. Leen, T. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems, Vol. 13, NIPS 2000, MIT Press, Denver, Colorado, USA, 2000, pp. 1–7.

[117] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, 2014, pp. 1–14, http://dx.doi.org/10.48550/arXiv.1312.6114, ArXiv.

[118] G. Zhang, Y. Liu, X. Jin, A survey of autoencoder-based recommender systems, Front. Comput. Sci. 14 (2020) 430–450, http://dx.doi.org/10.1007/s11704-018-8052-6.

[119] P. Ghosh, M.S.M. Sajjadi, A. Vergari, M. Black, B. Schölkopf, From variational to deterministic autoencoders, CoRR, 2020, pp. 1–25.

[120] D.J. Rezende, S. Mohamed, D. Wierstra, Stochastic backpropagation and approximate inference in deep generative models, in: E.P. Xing, T. Jebara (Eds.), Proceedings of the 31st International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 32, PMLR, Bejing, China, 2014, pp. 1278–1286.

[121] B. Settles, Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 6, Carnegie Mellon. School of Computer Science, 2012, pp. 1–114.

[122] Y. Gal, R. Islam, Z. Ghahramani, Deep Bayesian active learning with image data, in: Proceedings of the 34th International Conference on Machine Learning, Vol. 70, Association for Computing Machinery, Sydney, Australia, 2017, pp. 1–10.

[123] C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight uncertainty in neural networks, in: Proceedings of the 32nd International Conference on Machine Learning, Vol. 37, Journal of Machine Learning Research, JMLR, Lille, France, 2015, pp. 1–10.

[124] M. Opper, C. Archambeau, The variational gaussian approximation revisited, Neural Comput. 21 (2009) 786–792, http://dx.doi.org/10.1162/neco.2008.08-07-592.

[125] M. Filippone, Deep Gaussian processes, in: Deep Bayes 2019, Bayes Group, Moscow, 2019, pp. 1–82.

[126] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, 2016, pp. 1–15, ArXiv.

[127] T. Pearce, A. Brintrup, M. Zaki, A. Neely, High-quality prediction intervals for deep learning: A distribution-free, ensembled approach, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 80, PMLR, Stockholm, Sweden, 2018, pp. 4075–4084.

[128] M. Valdenegro-Toro, D. Mori, A deeper look into aleatoric and epistemic uncertainty disentanglement, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, New Orleans, Louisiana, USA, 2022, pp. 1508–1516.

[129] iMarc, Foreign exchange market size, share, trends, analysis 2022-2027, 2021, https://www.imarcgroup.com/foreign-exchange-market.

[130] School of pipsology, 2022, https://www.babypips.com/learn/forex/preschool#why-trade-forex.

Txus Blasco received a B. Eng degree in industrial engineering, specialty in turbomachines from the Universitat Politècnica de València, Spain, in 2012, a M.Sc. Degree in aerospace propulsion from École Centrale de Lyon, France, in 2010 and is currently following studies to obtain a Ph.D degree in computer science from the Universidad Jaume I, Castellón, Spain. He develops a role as research and development project manager at the company Amaris. His research interest includes uncertainty quantification applied to the domain of financial time series prediction.

Prof. Sánchez is Full Professor at the Department of Computer Languages and Systems at the Universitat Jaume I of Castelló. He received a BSc. in Computer Science from the Universidad Politécnica de Valencia in 1990 and a Ph.D. in Computer Science Engineering from Universitat Jaume I in 1998. He has participated in more than 30 R&D projects related to machine learning, and authored or co-authored about 200 publications. Currently, he is the leader of the Pattern Analysis and Learning group and coordinator of the Ph.D. Program in Computer Science at the Universitat Jaume I. He was the former President of the Spanish Association for Pattern Recognition and Image Analysis.

Vicente García received a B.Eng. degree in computer systems from the Technological Institute of Villahermosa, México, in 2000, a M.Sc. degree in computer science from the Technological Institute of Toluca, México, in 2002, and a Ph.D. degree in computer science from the Universitat Jaume I, Castellon, Spain, in 2010. From 2010 to 2013, he was a research fellow with the Institute of New Imaging Technologies, Castellón, Spain. Since 2014 is a Full Professor of Computer Science at the Universidad Autónoma de Ciudad Juárez. His research interest includes classification, performance evaluation metrics, data mining, and medical imaging.