Neural Network–Based Financial Volatility Forecasting: A Systematic Review

WENBO GE*, The Australian National University, Australia
POOIA LALBAKHSH*, RMIT University, Australia
LEIGH ISAI, Euler Capital, Australia
ARTEM LENSKIY, The Australian National University, Australia
HANNA SUOMINEN, The Australian National University, Australia, Data61/CSIRO, Australia, and University of Turku, Finland

Volatility forecasting is an important aspect of finance as it dictates many decisions of market players. A snapshot of state-of-the-art neural network–based financial volatility forecasting was generated by examining 35 studies, published after 2015. Several issues were identified, such as the inability for easy and meaningful comparisons, and the large gap between modern machine learning models and those applied to volatility forecasting. A shared task was proposed to evaluate state-of-the-art models, and several promising ways to bridge the gap were suggested. Finally, adequate background was provided to serve as an introduction to the field of neural network volatility forecasting.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Computing methodologies** → *Neural networks*; • **Mathematics of computing** → Time series analysis; • **Applied computing** → *Forecasting*; Economics;

Additional Key Words and Phrases: Financial volatility forecasting, machine learning, neural networks, predictive analytics

ACM Reference format:
Wenbo Ge, Pooia Lalbakhsh, Leigh Isai, Artem Lenskiy, and Hanna Suominen. 2022. Neural Network–Based Financial Volatility Forecasting: A Systematic Review. *ACM Comput. Surv.* 55, 1, Article 14 (January 2022), 30 pages.
https://doi.org/10.1145/3483596

*The majority of this study was conducted whilst working at Euler Capital.
This work was supported by Euler Capital Pty Ltd, under APR Intern Agreement (INT-0804). We would also like to thank APR.Intern and Australian National University (ANU) for their support to this research through Australian Postgraduate Research Intern (APR.Intern) program. We also gratefully acknowledge funding from the Australian Government Research Training Program (AGRTP) Domestic Scholarship for the first author's PhD studies.
Authors' addresses: W. Ge and A. Lenskiy, The Australian National University, School of Computing, 145 North Road, Acton, Canberra, ACT, 2601, Australia; emails: {wenbo.ge, artem.lenskiy}@anu.edu.au; P. Lalbakhsh, Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Monash University, 25 Exhibition Walk, Clayton VIC 3800, Australia; email: pooia.lalbakhsh@monash.edu; L. Isai, Euler Capital, 63-67 Cemetery Road, Drysdale, Victoria, 3222, Australia; email: leigh@eulercapital.com.au; H. Suominen, The Australian National University, School of Computing, 145 Science Road, Acton, Canberra, ACT, 2601, Australia, Data61/CSIRO, 108 North Road, Acton, Canberra, ACT, 2601, Australia, and University of Turku, Department of Computing, FI-20014 Turun yliopisto, Turku, 20500, Finland; email: hanna.suominen@anu.edu.au.

0360-0300/2022/01-ART14 $15.00
https://doi.org/10.1145/3483596

1 INTRODUCTION

One of the most important tasks in finance is to monitor volatility of market variables such as commodity prices, interest rates, and the variables that constitute the value of its portfolio. It is also a key factor in the pricing of many financial instruments, and is the underlying asset of many derivatives. However, like all financial forecasting and prediction (used interchangeably in this text), it is not a simple task [103]; the stylised facts of volatility [39, 50, 90], the efficient market hypothesis [87, 124], and the ephemeral nature of financial relationships [29, 37] are just a few of the many reasons why. Despite this, it is still possible to forecast volatility to some degree [103, 124].

Volatility is often referred to as latent and unobservable [103], even *ex-post*, and finding a suitable way to quantify or estimate it is a problem in and of itself, leading to many different definitions being proposed [101]. Although these are only proxies of latent volatility, they still have practical value as they provide a quantitative way of comparison and often align with market definitions [1].

In addition to the many volatility proxies, there are many methods that attempt to model, understand, and forecast volatility. One of the most widely used is the ***generalised autoregressive conditional heteroscedasticity*** (**GARCH**) model, and its family of variants [12]. In contrast to these traditional models, intelligent methods have recently gained much traction and are often nonlinear, encompassing methods such as ***machine learning*** (**ML**) and ***deep learning*** (**DL**), ***evolutionary algorithms*** (**EAs**), and fuzzy logic [20, 114]. ML and DL specifically have surged in popularity over recent years due to a flurry of successful ***neural network*** (**NN**) applications [62], a movement that is also being seen in financial volatility forecasting [20].

This systematic review looks to provide an overview of NN-based financial volatility forecasting, henceforth referred to as NN volatility forecasting. This article does not aim to be a comprehensive review of volatility forecasting in financial markets; for such a review, we direct the readers to Poon and Granger's 2003 paper [103], which provides a dated yet relevant review of financial volatility forecasting, or Sezer, Gudelek, and Ozbayoglu's 2020 publication [114], which provides a broad snapshot of state-of-the-art DL in financial time series forecasting. This study differs from the above in that the scope will be significantly narrowed to volatility forecasting with NN-based methods, from 2015 onward, thus offering a higher level of detail. To the best of our knowledge, there exists no timely and comprehensive review for volatility forecasting on financial time-series data using NNs.

Specifically, we had the following three aims: (1) to create a text that can be used as an introductory source to the field of financial volatility forecasting, (2) to provide a snapshot of the state-of-the-art in NN volatility forecasting, and (3) to identify some common issues, how these may be addressed, and some future directions.

The rest of the review is organised as follows: Section 2 provides a broad background on financial volatility, including some theory surrounding the different definitions and when these definitions may be appropriate. Background is also provided on several traditional forecasting methods, intelligent forecasting methods, as well as how the forecasting task can be defined. Section 3 outlines the methodology for the analysis and sub-analyses of this review. Section 4 presents several key results synthesised from the analysis, and Section 5 discusses these, as well as presents a few common issues observed, how these may be addressed, and a few limitations. Concluding remarks are given in Section 6.

2 BACKGROUND

Financial volatility and its forecasting have been studied to a great extent for many years, and the demand for creating a suitable model to understand and forecast volatility is only increasing as the future grows with uncertainty and the number of market players increases. Although our focus is on NN volatility forecasting, it would be valuable to provide background on volatility forecasting in general.

2.1 Defining Volatility

Due to the numerous ways in which volatility can be quantified, it must be concretely defined before any discussion can begin. Volatility is often referred to as latent and unobservable. This is because, within theory, it is an instantaneous variable that scales a stochastic Wiener process, also known as a Brownian motion. The appropriate way to quantify this would be an integral over time, resulting in the *integrated volatility* [103]. However, this is not possible due to the discrete nature of financial systems. Instead, we can look to the theory of quadratic variation [66], in which the integrated volatility is approximated by the sum of squared returns within a time window τ_1 to τ_2, given a high enough sampling frequency. This results in the ***Realised Volatility*** (**RV**) proxy [4, 6]:

$$RV = \sqrt{\sum_{t=\tau_1}^{\tau_2} r_t^2}, \tag{1}$$

where $r_t = log(P_t/P_{t-1})$ is the log return at time t, and P_t is the price at time t. When deciding on the sampling frequency, there is a tradeoff between wanting a very high frequency to approximate continuously observed frictionless prices, and wanting a lower frequency as to avoid market microstructures. Seminal works around this suggest from 5- to 15-minute intervals as a good balance [5, 6, 59, 112]; however, the decision ultimately relies on the market liquidity of the given asset.

Oftentimes, a financial asset does not meet the market liquidity condition, or high frequency data for an asset is not easily attainable; in such cases, other proxies for volatility are required. One such method is to quantify the dispersion in the daily closing prices of assets, known as the close-to-close method, or better known as ***Historic Volatility*** (**HV**) [103]. It is defined as the standard deviation (or variance) of the log return series, from the time window τ_1 to τ_2:

$$HV = \sqrt{\frac{1}{N}\sum_{t=\tau_1}^{\tau_2} \left(r_t - \mu_{\tau_1,\tau_2}\right)^2}, \tag{2}$$

where $\mu_{\tau_1,\tau_2} = \frac{1}{N}\sum_{t=\tau_1}^{\tau_2} r_t$ is the mean of the log returns within a time window, $N = \tau_2 - \tau_1$ is the number of samples within that time window, $r_t = log(P_t/P_{t-1})$ is the log return at time t, and P_t is price at time t. The log returns series is used as opposed to the daily closing prices as they often fit a Gaussian distribution [5, 55]. This is important as the standard deviation is only an appropriate and meaningful measure of dispersion when the underlying distribution of the sample is Gaussian, and it is well known that asset prices do not meet this condition. Occasionally, the returns series is used rather than the log returns as $log(1 + x) \approx x$ when x is small. In this case, the returns would be defined as $R_t = P_t/P_{t-1} - 1$, that is, $log\,(P_t/P_{t-1}) \approx P_t/P_{t-1} - 1$. There are many ways to test if the drawn data fits a Gaussian distribution; one that is commonly used is the Jarque-Bera test [65], though this may be biased for distributions with short tails, and other tests may be more appropriate, such as a Shapiro-Wilk test or a modified Cramér–von Mises test [123].

Closely related to HV are several other proxies that incorporate additional daily data, each with their own set of assumptions, advantages, and disadvantages. The extreme value method, proposed by Parkinson (1980) [100], assumes a continuous geometric Brownian motion with no drift that utilises high and low prices rather than the closing price. Another proxy proposed by Garman and Klass (1980) [46] also assumes a continuous geometric Brownian motion with no drift, but utilises open, close, high, and low prices. Since then, many other proxies that utilise the range of daily price have emerged, such as that of Rogers and Satchell (1991) [107] that allows for drift, and was empirically shown to be superior than an adjusted Garman Klass method in the presence of time varying drift [108]. Yang and Zhang (2000) [131] proposed a proxy that also assumes continuous geometric Brownian motion, but can handle drift and open price jumps. This was affirmed in simulation: all four proxies are good estimates of the true variance under a geometric Brownian motion with small drift and no opening jumps; the Parkinson and Garman Klass proxies overestimate true variance in the presence of large drift; and only Yang and Zhang's proxy is stable in the presence of large opening jumps [117]. Additionally, all four were shown to be a close approximation of the RV (15-minute sum of squared log returns) of S&P500 [117]. For a more comprehensive view and comparison of range-based proxies, we refer readers to the works of Chou et al. [24, 25]. Interestingly, there is evidence that some range-based proxies can be forecasted with less error than the traditional close-to-close method (or HV) [75]. However, some have also found that this does not hold when reasonable modifications were made, such as using an adjusted closing price rather than raw closing price for the close-to-close method [102].

From different thinking arises ***Implied Volatility*** (**IV**), in which the volatility is backwards calculated from an option price via some option pricing model, such as the infamous Black-Scholes model [91]. There is no closed-form inverse for the Black-Scholes option implied volatility, and are often evaluated through numerical solvers, such as the Newton-Raphson method [73, 91], or through a closed-form approximation [15, 119]. Because the prices of options are set by the market, and due to the efficient market hypothesis, IV can be thought of as an efficient expectation of future volatility by the market for a specific asset within a certain time period, that is, a good forecast of future volatility. It is also the definition underlying major volatility indices, such as the *Chicago Board Options Exchange's* ***Volatility Index*** (**VIX**) [1], which is the 30-day expected volatility resulting from the bid and ask prices of options whose expiry period lies between 23 and 37 days. It has also been shown to hold explanatory power in forecasting other volatility proxies, more so than past values of those proxies themselves [18, 27]. Despite its widespread use, the IV has several flaws. One such flaw is known as volatility smile, in which options that are identical in all but strike price will result in different levels of IV. This is smallest with at-the-money options, and increases as the options become increasingly in-the-money or out-of-the-money [3], meaning that the volatility of an asset will be different depending on what option is used. Another flaw is that option pricing models typically assume constant volatility throughout the period, which is generally never the case [91]. Additionally, it is empirically observed that IV is typically higher than other measures of volatility, such as the RV or HV, for the same period [27, 30, 43].

2.2 Forecasting Methods

There exists a plethora of methods for which to forecast, model, and understand volatility; one of which is NNs, which has recently gained significant popularity due to numerous successful applications in other fields. The field of NNs and ML is extremely broad, and there are many options for readers to gain a deeper understanding [9, 49]; this text only provides a brief overview of NNs, with emphasis in the context of financial volatility forecasting. NNs are often referred to as a universal function approximator [58] as they have the ability to learn any arbitrary nonlinear mapping f from input $\mathbf{X}$ to output y; $y = f(\mathbf{X})$ [129]. In the context of financial volatility forecasting, the mapping may be represented as

$$\widehat{y}_{t+1} = f(\mathbf{X}_t), \tag{3}$$

where $\widehat{y}_{t+1}$ may be the forecasted volatility proxy for the next time period, and inputs $\mathbf{X}_t$ may be a matrix of observations (e.g., previous returns $\mathbf{X}_t = [r_t, r_{t-1}, r_{t-2}, \dots]^T$). This mapping is learned through some optimisation method (typically back-propagation), in conjunction with some loss function and sampled data, in hopes that the learned map can generalise to out-of-sample instances.

The complexity of this mapping is bounded by the structure and quantity of computational nodes, also known as neurons. Increasing the complexity of this mapping is not always a good thing, as it may lead to over-fitting of the in-sample data, thus failing to generalise to out-of-sample instances, and is one of the reasons why the complexity of the network is often restricted [10].

Roughly, NNs can be categorised into three broad groups, each of which is designed for different applications and leverage different aspects of the data. First, ***multi-layer perceptrons*** (**MLPs**) do not naturally leverage any structure within the data and can be thought of as the simplest, most traditional category of NNs, with the most history and largest body of research. It is extremely flexible and can be formulated in different ways to exhibit different properties, such as with a ***nonlinear autoregressive*** (**NAR**) framework, to enforce an autoregressive property to the nonlinear mapping in which the forecast is a function of its previously observed values (e.g., $\widehat{y}_{t+1} = f([y_t, y_{t-1}, \ldots, y_{t-m}]^T))$ [26, 77]. This can be extended to include exogenous variables (a NARX framework), thus incorporating more information [17, 83]. The learning method can be altered, like that of an ***extreme learning machine*** (**ELM**) to learn significantly faster while maintaining relatively good generalisation performance [60, 98]. The non-linearity could be altered, like that of a *radial basis function* (RBF) network, to capture different kinds of nonlinearity within the data more easily [16, 88]. The loss could be defined differently, like that of a ***quantile regression NN*** (**QRNN**) allowing the network to estimate the conditional median and other quantiles, rather than just the conditional mean [104, 122].

Second, ***recurrent neural networks*** (**RNNs**) were originally designed to analyse sequential data and exploit the sequence in which data appears, a natural fit for time series forecasting. The RNN builds a hidden state as it recursively parses through the input sequence, retaining useful information from previous elements, often referred to as the networks memory. However, the vanilla RNN can typically only hold short-term information [69], which has resulted in several extensions, such as the ***long short-term memory*** (**LSTM**) [57] and ***gated recurrent unit*** (**GRU**) [28]. These are widely used and combat this problem by using gates to allow the network to remember, update, and forget information. One important consideration is how much of the series should be exposed to the model; a recurrent network can parse the entire history of the time series or just a small window of it, a key differentiating point when compared to other NN architectures. The former allows the network to build context from the entire series, whilst the latter can only build context from the small time window.

Third, ***convolutional neural networks*** (**CNNs**), originally based on the visual cortex [45], were designed for image tasks, and leverages spatial information in the data. Independent units search over the entire image in small sections, looking for the presence of a feature to create a feature map. Thus, translational invariance is built into the model, meaning the presence of a feature is detected regardless of its location in the input data, which is not the case with the MLP or RNN. This process is hierarchically repeated on the feature maps until the final layer aggregates all the extracted information to perform a prediction. Due to the highly hierarchical structure, a well-trained CNN can be thought of as a feature extractor, as the lower levels of the network often detect base features that are common to all images, whilst the upper levels are conditioned for the specific task [74]. There are many ways in which time series data can be used with a CNN. The first is to replace the standard two-dimensional convolutions with one-dimensional convolutions. This simply means the spatial information is leveraged in one dimension: time. Some successful examples of this are WaveNet [125] and ***temporal convolutional networks*** (**TCNs**) [78]. It is also possible to convert the time series into an image, such as through a recurrence plot [36], ***Gramian Angular Fields*** (**GAFs**) [128], or ***Markov Transition Fields*** (**MTFs**) [19], and then use traditional two-dimensional convolutions.

In terms of volatility forecasting models, the traditional GARCH family of models are infamous due to their theoretical underpinnings and popularity. The GARCH model is an extension of the ***autoregressive conditional heteroscedasticity*** (**ARCH**) model [38], belonging to the ***autoregressive*** (**AR**) universe of models. As the name suggests, an ARCH(q) model forecasts future volatility, conditioned on previous observations, and can be represented as

$$\epsilon_t = \sigma_t z_t, \tag{4}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2, \tag{5}$$

where ϵ_t are the residuals (or innovations), z_t is white noise, σ_t is the conditional variance, α_0 and α_i are learned coefficients, and q is the number of lagged square residuals to include (or the order of the ARCH model). This can be extended to the GARCH(p, q) model by assuming the error variances follow an ***autoregressive moving average*** (**ARMA**) model, which can be expressed as

$$\sigma_t^2 = w + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \tag{6}$$

where w, α_i, and β_i are learned coefficients, and p is the number of lagged conditional variances to include (or order of the model). Since the conception of the GARCH(p, q) model, there have been many advancements that address the models inability to capture several stylised facts of volatility [39]. Exponential, threshold, and Glosten-Jagannathan-Runkle versions (EGARCH, TGARCH, GJR-GARCH) allow for asymmetric dependencies in volatility. Integrated and fractionally integrated versions (IGARCH, FIGARCH) address volatility persistence, where an observed shock in the volatility series seems to impact future volatility over a long horizon. Despite the countless other versions of the GARCH model, some have found that a GARCH(1, 1) forecasting model outperforms all other GARCH variants (including ARCH) for foreign exchange volatility. However, this was not seen for stock market volatility, where a GARCH(1, 1) was outperformed by a variant (though was still significantly better than the ARCH model) [53]. Others have found that the ARCH model is superior to GARCH for country indices when estimating ***value at risk*** (**VaR**) [99].

There are also many other models belonging to the AR universe that can be used for forecasting. One example that has been used in literature is the ***conditional autoregressive range*** (**CARR**) models [23, 48], a generalisation of GARCH models with the extreme value theory to allow forecasting of range-based volatility proxies from range-based inputs. Another example is the ***smooth transition autoregressive*** (**STAR**) family of models, which uses different AR models in different regimes with a smooth transition function between them [8, 21], or the ***heterogeneous autoregressive*** (**HAR**) models, inspired by the Heterogeneous Market Hypothesis, a relatively simple model that considers volatilities realised over different interval sizes [7, 31]. A final example is the ***moving average*** (**MA**) family of models, such as ARMA [14, 132] and ***exponentially weighted moving average*** (**EWMA**) [63, 68]. All these models are unified in that the forecasts they produce are conditioned on past values, a key property in all AR models, however, they all differ in exact formulation, and thus have different attributes.

A hybridised NN forecasting method can also be fashioned by combining the previously mentioned AR methods with a NN, each compensating for the others flaws. There are also many other methods to hybridise NNs with, such as methods derived from fuzzy logic [33, 121] or chaos theory [84]. Fuzzy models allow better handling of imprecise and non-numerical information with differing degrees of truth by allocating inputs into fuzzy sets, which can increase robustness to uncertainty. One way to incorporate this with a NN is through an ***adaptive network-based fuzzy inference system*** (ANFIS) architecture [33, 64]. It provides the advantage of combining the rules in the rule base of fuzzy theory to describe the complex relationships, and the learning ability of NN to adjust the membership functions and rule base. Chaos theory models the apparent disorder within a system that nevertheless obeys specific rules and is highly sensitive to initial conditions. Although there does not seem to be solid support for the presence of chaos in economics in the past years [40], some recent developments in testing have found empirical evidence of chaotic behaviour in volatility, specifically volatility indices [84].

There are also a number of other methods that can be used in conjunction with NNs, but were not considered hybridisations for the purposes of this review. These included methods such as transforming/preprocessing the input data to make information more readily available for the forecasting model, such as ***Principal Component Analysis*** (**PCA**), for dimensionality reduction [121, 130], ***Discrete Wavelet Transform*** (**DWT**), for basis function dependant signal decomposition [115, 132], and ***Empirical Mode Decomposition*** (**EMD**), to decompose signals into ***Intrinsic Mode Functions*** (**IMFs**) [61, 133]. Evolutionary computing can also be used as an alternative parameter search and optimisation method to train the weights of the NN, as opposed to the standard first order back propagation.

2.3 Defining the Forecasting Task

Given that the NN is a universal approximator that attempts to learn the mapping f in $\widehat{y}_{t+1} = f(\mathbf{X}_t)$, the *task* defines what $\widehat{y}_{t+1}$ and $\mathbf{X}_t$ are. This is an important factor in the success of a model; the inputs $\mathbf{X}_t$ must contain enough information to predict the output $\widehat{y}_{t+1}$. Note that information is not used interchangeably with data: data is simply a set of observations or samples, whilst information is more abstract; noise can produce an infinite amount of data, but will contain almost no information. The input can be constructed in many ways, the simplest of which is through a univariate approach, where the input contains only one variable (but possibly multiple lags of it). This will lead to a simple and parsimonious model, but may not perform well as the inputs may not hold enough information to predict the output effectively. An example can be seen in Figure 1, in which a NAR MLP network forecasts volatility 1 day ahead from the previous five observed values of volatility. Additional information can be introduced to the model through more variables, either derived from the same underlying asset or other assets. This can better describe the movements of both the underlying asset and the market in general, but may mean facing the curse of dimensionality, in which high dimensional inputs result in samples becoming exponentially more sparse. This means all samples will become approximately equidistant, resulting in the network not being able to discover useful clusters, ultimately leading to over-fitting and poor out-of-sample performance. The output can also be constructed in many ways, not only in terms of what volatility proxy to use, but also in terms of when the window of forecast starts and the length of this window (specifically, τ_1 and τ_2 in Equations (1) and (2)). Different forecasting windows require different information from the inputs. Whether a set of inputs holds enough information to forecast an output is particularly difficult to evaluate, as typical measures, such as correlation or information criterion, are linear and have trouble capturing how well a nonlinear NN mapping will perform.

3 METHODS

3.1 Literature Searching and Screening

This review followed PRISMA guidelines for systematic reviews [94]. SCOPUS, Web of Science, and the IEEE Xplore Digital Library databases were searched, using the query of *(financ*) AND (volatilit*) AND ("machine learning" OR "deep learning" OR "neural network") AND (predict* OR forecast*)*, tailored specifically for the search engine of each database. Additionally, the results were restricted to publications from 2015 onward in order to assess the most recent methods, as

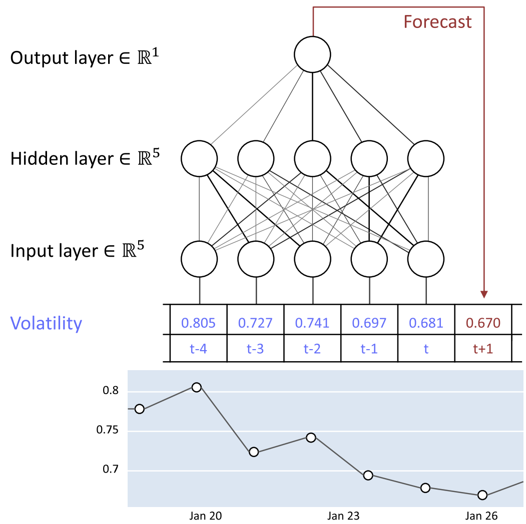

Fig. 1. Example of the NAR framework.

around this time marks a paradigm shift in which DL began setting records in several international competitions, gaining widespread recognition and use [111]. Retrieved articles were screened for duplicates and independently assessed for inclusion by two of the authors. Any conflicts that arose were discussed, and another author was consulted if it could not be resolved.

Six inclusion criteria were developed and applied in this review:

(1) The article had to attempt to forecast or predict volatility. This excluded articles that only forecasted price, but included articles that go further and use that price to forecast volatility.
(2) The input to the volatility prediction system had to be derived from a traded financial asset, excluding inputs such as news articles or social media (e.g., tweets).
(3) The forecasted volatility (output) had to be on a traded financial asset. This excluded cryptocurrencies, electricity, and foreign currencies, but included assets such as crude oil and gold.
(4) The methods had to apply a NN, or some variation of it, which excluded methods that solely rely on other ML techniques, such as ***Support Vector Machines*** (**SVMs**), ***k-nearest-neighbours*** (**kNNs**), and ***random forests*** (**RFs**).
(5) A NN method could not only appear as a benchmark method, as the original article for that method most likely already falls under this review.
(6) The article had to include an out-of-sample evaluation, to ensure validity of outcomes.

3.2 Literature Analyses

A sub-analysis was performed on the frequency of use of different volatility proxies in the reviewed literature by categorising them into four groups: historical volatility, realised volatility, implied volatility, and others (a combination of range-based, jump-based, and GARCH-based volatility proxies). In the cases that it was unclear which volatility proxy was used, an educated guess was made based on other factors of the paper, such as how the inputs and outputs were structured.

The NN forecasting methods and their hybridisations were also tabulated in order to compare frequency of use. These methods were categorised as either "Pure" if only a NN method was used, or "Hybrid" if the NN was combined with another method. Further categorisation of the NN model

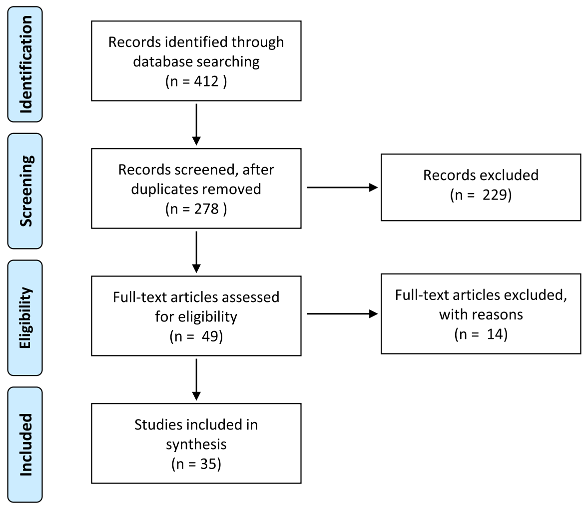

Fig. 2. PRISMA flow diagram.

was made into three broad classes: "MLP," "RNN," and "CNN." The presence of an MLP in conjunction with the RNN or CNN classes did not justify including it as an MLP category; instead, it sat solely in the other class. The hybridisation method was also tabulated into the categories: AR methods (encompassing GARCH, CARR, STAR, HAR, and MA models), fuzzy methods, and methods drawn from chaos theory. Although the use of data preprocessing/transformation or evolutionary computing for network optimisation were not considered as a form of hybridisation, they were also tabulated.

The input structure was also categorised as univariate or multivariate, depending on if the network takes in one variable or multiple variables. Note that some papers were categorised in both as they have different versions of a model with different inputs. Also, the presence of lagged values of a single variable was still categorised as univariate. In addition to this, the input structure was also categorised into single or multi-asset, depending on what the input variables were derived from. Note that if the input structure was multi-asset, it was also multivariate. The exact input variables, along with which assets they were derived from, and number of lags were also tabulated.

Additionally, the assets of the forecasted volatility were also catalogued, along with the window of forecasted volatility, that is, τ_1 and τ_2 in Equations (1) and (2). Any asset that had only been used once, was grouped into either "single use indices," "single use stocks," or "single use volatility indices." Additionally, metals (gold, silver, copper), oils (crude oil, natural gas, heating oil), and notes and bonds, were also grouped.

4 RESULTS

The literature search generated 412 results (as of August 2020), which was reduced to 278 after duplicates were removed. This was further reduced to 49 based on the title and abstract filtered against the inclusion criteria, and was finalised to 35 upon reading the full articles, shown in Figure 2.

4.1 Volatility Definition

The most commonly used volatility proxy was HV, in 25 out of 35 studies, followed by RV, in 6 of 35 (Figure 3). The next most frequently used were the other volatility proxies, consisting of

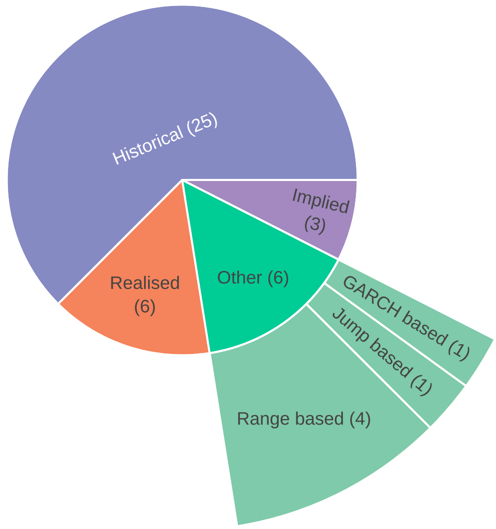

Fig. 3. Type of volatility definitions used in the reviewed literature. The number of studies is indicated in parentheses.

range-based proxies (4 of 35), jump-based proxies (1 of 35), and GARCH-based proxies (1 of 35). The least used volatility proxy was IV, in 3 of the 35 reviewed articles. Of these 3 articles, 2 of them forecasted the value of a volatility index, rather than the implied volatility itself. It should be noted that the sum of the number of studies exceeds 35, as some studies have used more than one volatility proxy, as shown in Appendix A, Table 2.

4.2 NN Methods

A purely NN approach was the most common in literature, occurring in 21 of 35 studies (Figure 4). The remaining 14 studies took a hybridisation approach by introducing an additional method in conjunction with the NN. Of the network categories, the MLP was the most common archetype, used in all but one of the hybrid models (13 out of 14), and over half of the pure models (12 out of 21). The RNN was the second most frequently used, in 9 of the purely NN models, and once as a hybrid, in conjunction with an AR method (specifically with GARCH and EWMA). The CNN was only used in one study, and only alongside an RNN. Of the hybridisation methods, AR hybridisation occurred in 12 of the 14 hybridised NN studies, 10 of which used models from the GARCH family, 2 studies that used MA methods (ARMA and EWMA), 1 used an HAR model, and 1 study that used an LSTAR model. Fuzzy models were used in five studies as a hybridisation method, mostly through an ANFIS or ANFIS-like architecture, whilst chaos theory models were only used once. A total of six studies used an EA, twice for network weight initialisation, and four times for network training (Appendix A, Table 1).

4.3 Task Definition

The majority of studies (26 of 35) derive all input variables from a single asset, whilst only 7 of 35 derive inputs from multiple different assets (Figure 5). Of the single-asset representations, 11 studies used a univariate input, 13 used a multivariate input to the model, and 2 studies used both a univariate and multivariate input. The input structure was not available in 2 of the studies reviewed. Of the actual input variables used, returns and volatilities were the most common, used in 19 and

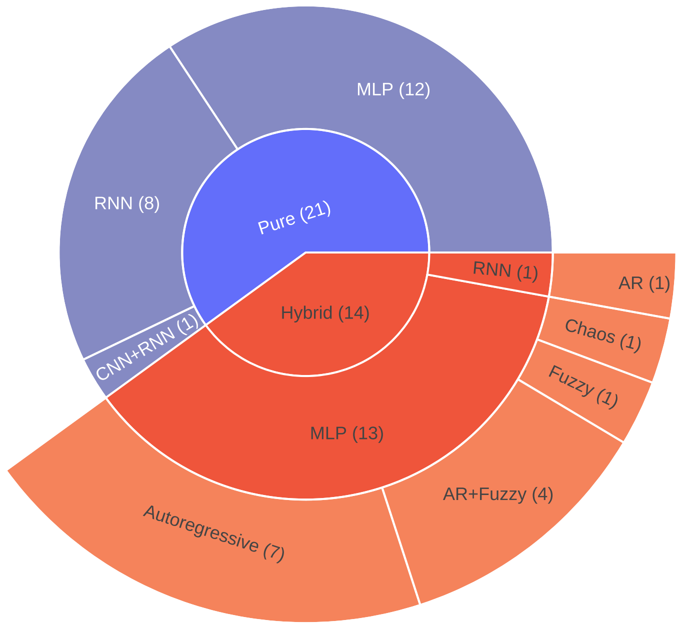

Fig. 4. Volatility forecasting methods used in the reviewed literature. The number of studies is indicated in parentheses.

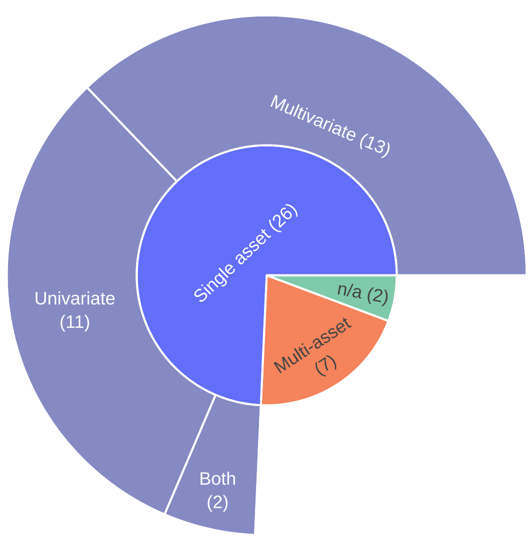

Fig. 5. Input structure of the tasks identified in the reviewed literature. The number of studies is indicated in parentheses.

20 studies, respectively (Appendix A, Table 2). Additionally, input transformation occurred in 6 studies, 2 of which used **variational mode decomposition (VMD)**, 2 of which used Wavelet decomposition, 1 used PCA, and 1 used Cholesky decomposition.

There was a wide spread of assets for which the volatility was predicted on, the most common being S&P500, forecasted in 12 studies, followed by oils in 7 studies, and then metals in 6 studies

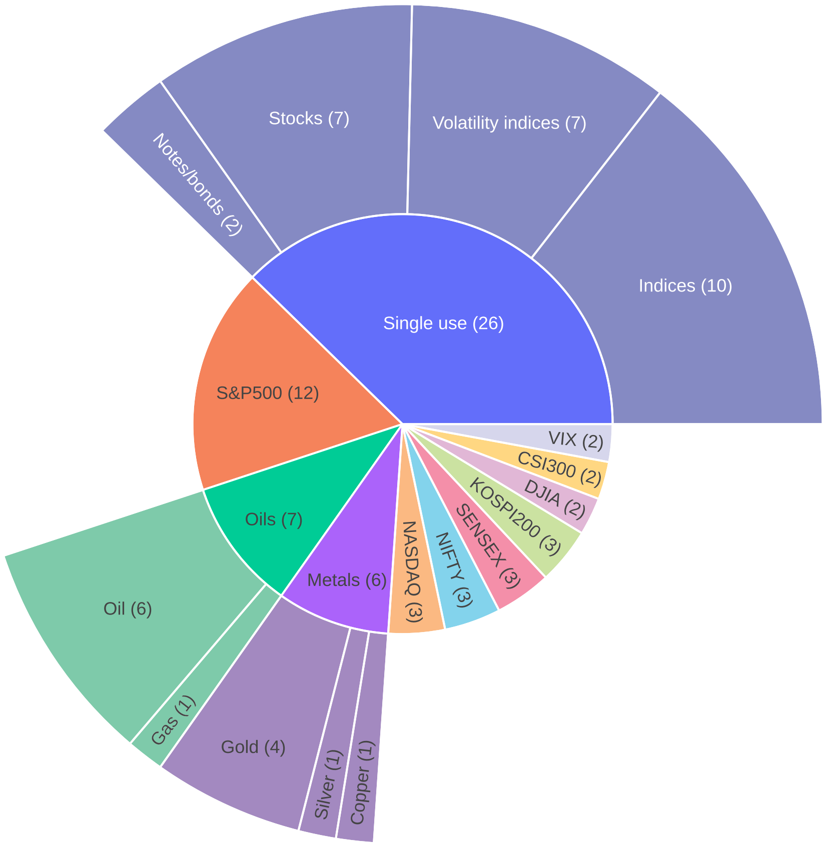

Fig. 6. Assets for which volatility was forecasted in the reviewed literature. The number of studies is indicated in parentheses.

(Figure 6). It can also be seen that the interest and application was global, with indices from the United States, India, Korea, China, Japan, UK, German, French, Dutch, Swiss, and Hungary (Appendix A, Table 2). Note again that the total number of assets was greater than 35 as some studies used multiple.

5 DISCUSSION

It is not surprising that HV was the most frequently forecasted volatility proxy, as it is used in many risk valuation methods [35] and technical analyses [13], and given its simplicity and intuition. Following this was RV, and although it may be the best theoretical approximation of integrated volatility, there are several limiting factors, such as requiring high frequency data and an asset that has high market liquidity, which makes it less appealing. Given that IV requires options data, it is not surprising it was used the least frequently; options data is not easily attainable, and using a volatility index means being confined in terms of underlying asset and forecast window. Of the network archetypes, MLP was the most commonly used for both the purely NN models and the hybridised models, likely as it is extremely flexible, very well studied, and the simplest to use. The RNN, being a natural fit for time series prediction tasks, was used the second most frequently; however, was only hybridised once, possibly as it is less approachable than the MLP.

The CNN was only used once, presumably because it was originally designed to leverage spatial information and is typically used for image data. In terms of hybridisation methods, AR models were the most common, likely to supplement the inability of the MLP to hold memory regarding historical observations in the series. It is also interesting to note that EA was only used for the MLP networks, likely because the weights of the RNN are a lot more unstable with regard to their performance [113].

As the field of financial volatility forecasting grows, so too will the number of related publications, methods, and models. This increases the need for simple yet meaningful comparisons between the forecasting performance of different models, something that does not currently exist. This is because the forecasting task (what defines the inputs and outputs) across different experiments and models are too heterogeneous. More specifically, task heterogeneity exists due to different volatility proxies, underlying assets, and forecasting windows. Although there exist methods to make meaningful performance comparisons between models, such as the ***Superior Predictive Ability*** (**SPA**) test [52] or the ***Model Confidence Set*** (**MCS**) [54], they are far from simple as they require the competing models to be implemented, which may take a considerable amount of time. Moreover, this does not address the issue of task heterogeneity. One solution that has been successful in other fields is a *shared task*: a specific question posed to the research community with the aim of evaluating the state-of-the-art, in which the outputs and (not necessarily) the inputs are shared [22, 42]. Given that the same error metric is used (i.e., mean squared error), this allows for a direct comparison of results between publications without the need to implement competing models, therefore allowing synthesis and meta-analyses in spite of experimental heterogeneity [47, 120]. Without a shared task, even if the research community works on the same general problem, the nuances of the problems will vary to the point where the approaches would not be comparable, leading to slower overall development [42]. This is what has happened within the field of financial volatility forecasting: although the general goal is shared (to forecast volatility), the task itself is not. Shared tasks have had much success in the other fields of ML, such as MNIST and ImageNet in computer vision [79, 110] and Penn TreeBank and WikiText-2 in NLP [22, 89, 92, 93]. An example of a shared task for financial volatility may be to forecast next month's HV of S&P500 for a 2-year test period: 1st Jan. 2017–31st Dec. 2018, using any data prior to the start of the test period for training, with the performance reported as mean squared error. This would also accommodate variable input and model complexity by including as much additional information as desired (assuming the information is known before forecasting). By not constraining the inputs, this shared task may also provide an avenue for determining if certain pieces of information increase forecasting power; if a model utilised sentiment analysis from news articles in addition to time series financial information, and can perform the shared task with a lower error than other models that only utilised time series financial information, a reasonable conclusion is that the information contained in the news articles gave the model a predictive edge.

When comparing state-of-the-art methods in ML to financial volatility forecasting, there seems to be a clear disparity in the NN models used. That is, financial volatility forecasting models seem drastically out of date. While some level of delay is to be expected due to the speed of information propagation and the time required to perform and publish experiments, the gap observed still seems far too large: only one of the reviewed papers applies DL to financial volatility forecasting. This begs the question: why? One scenario is that DL was already implemented and performed poorly, thus was not published. If this were the case, it is a blunder by the research community to not have published such results, as they are important and allow us to be more efficient in our exploration [109]. Another scenario is that they have been implemented and performed well, but were kept private to retain any profit making potential. In this case, it is not the responsibility of academia to keep these models private, and thus we should proceed by disregarding this scenario. A final scenario is that DL methods have truly not yet been implemented. In all three scenarios, the best way to proceed is the investigation and publication of DL models for volatility forecasting. There are many ways to approach this, and bridges to close the gap between ML and financial models: firstly, we suspect there are large gains to be had by utilising state-of-the-art models for image recognition and ***Natural Language Processing*** (**NLP**) that have been modified for time series data, such as the TCN or WaveNet (modified image recognition models) [78, 125], or the ***Temporal Fusion Transformer*** (**TFT**) (modified NLP models) [81, 82]. Additionally, we believe there is much potential in forecasting the distribution of volatility, rather than the exact value of volatility, which can be achieved through stochastic bootstrapping [44], stochastic NNs [2], or Gaussian process NNs [80]. Incorporating the attention mechanism may also prove beneficial, allowing the model to apply an attention mask to ignore data that may be considered uninformative [126]. Lastly, due to the ephemeral nature of financial markets, the application of continual learning seems fitting [116], as creating accurate forecasts is a constant battle between the model trying to learn new patterns to maintain an accurate volatility forecast, and the market, composed of many individuals, trying to discover patterns to exploit for profit. Moving forward, we should not only draw from ML, but also leverage aspects of the data that are specialised to financial time series to further increase performance of forecasting models. One simple example is making the model aware of the trading day through a simple encoding. Because there are only five trading days in a week (Monday to Friday), a forecast made for Monday (from Friday) should have a lower degree of confidence than a forecast made for Tuesday (from Monday). Current models do not take this into account and views all next-day forecasts equally. The approaches listed here are not intended to be comprehensive, but merely a few that we feel are promising; there are many other areas of ***artificial intelligence*** (**AI**) research that has shown success and any of them may be a good place to start.

Another way to increase the forecasting performance of a predictive model is to modify the task, rather than the model. That is, utilising a different set of inputs to forecast the output, one that contains more information and requires a less complex mapping. A commonly observed experimental design choice was to derive all input variables from a single asset, and whilst this may lead to a more simple and parsimonious model, a multivariate-multi-asset approach would be able to offer more information in terms of both the underlying asset and the market in general, and thus likely a better forecasting model. This can be seen in Kristjanpoller and Minutolo (2015) [70], which forecasted the volatility of gold spot prices using gold spot returns, several market indices (DJIA and FTSE100), exchange rates (EUR/USD and YEN/USD), and other assets (price of oils). Another example can be seen in Kim and Won (2018) [68], in which the volatility of KOSPI200 is forecasted with additional inputs from treasury and corporate bond rates, gold prices, and crude oil prices. Empirical evidence seems to show that the inclusion of these additional variables derived from other assets tend to increase out-of-sample performance as compared to when they are not present [70]. However, it is not as simple as including all possible variables as there exists a tradeoff between too many exogenous variables and the curse of dimensionality, leading to reduced out-of-sample performance [71, 72]. Not only this, the capacity of the model also needs to be considered, as a simple model fed with too much information will not be able to utilise it, whilst a complex model fed with too little information may get lost with a weak signal in its complex loss landscape.

Several issues were encountered in the synthesis of this literature review. One major issue seen in several studies was the inadequate reporting of experimental details, which not only affects the reproducibility of results, but also makes meaningful comparisons much more difficult. There also seems to be confusion in certain terminology, exaggerating this issue. Certain papers use implied volatility to mean the future volatility that is implied from the past data, and while this is valid under the English language, should be refrained from as it may be confused with the option price implied volatility proxy. In the same way, historical and realised volatility have been used to refer to volatility derived from observed/in-sample/historic data, whilst other papers have (correctly) used them to refer to specific volatility proxies. Both of these could be addressed by making all related works open source, something that should be common practice, but unfortunately is not. Additionally limitations include only reviewing a subset of all literature due to not searching every database possible, thus missing some studies. The presence of publication bias also affects the results of this literature review, as papers with negative results may have not been published, meaning that the results presented here may not be representative of all research being done on volatility forecasting with NNs. A final limitation is that because this review targets studies in which the NN method is the main contribution, the majority of these papers will report the NN method outperforming the benchmarks. This may lead to the conclusion that NNs are the best volatility forecasting model, a result that likely would not be seen if the focus of the review was targeted at other methods.

6 SUMMARY AND CONCLUSION

This text reviewed the literature for NN-based financial volatility forecasting, analysing a total of 35 studies published after the year 2015, in order to (1) create a text that can be used as an introductory source for NN volatility forecasting, (2) provide a snapshot of the state-of-the-art in NN volatility forecasting, and (3) identify some common issues, how these may be addressed, and some promising future perspectives.

We have provided a background on several important and commonly used definitions of volatility, along with a general overview of NNs, the broad types of NNs, and how the forecasting task may be defined. We also provide a background on the more traditional yet extremely popular GARCH models, as well as several other models in the AR family, fuzzy logic models, and chaos theory, all of which could be used in combination with a NN.

A snapshot of the state-of-the-art in NN volatility forecasting was created, in which we identified that HV (close-to-close method) was the most frequently used proxy, as opposed to others that may hold more theoretical backing, such as IV or RV. We also identified the relatively simple and well-studied MLP to be the most popular NN structure, used in all but one hybridised method, and over half the pure models. The RNN followed, with the CNN being used only once. Of the hybridisation methods, AR models were by far the most popular, with GARCH models being the biggest contributor, likely to add an element of memory and autoregression to the NN models. When defining the input variables, the most common approach was to derive multiple variables from a single asset, which may sufficiently capture the movements of that asset, but will not be able to sufficiently capture the movements of the broader market. We also observed that the most common asset for which volatility was forecasted was S&P500, which is not surprising due to its liquidity and accessability.

As more work is published in this field, the current inability to make easy and meaningful comparisons across publications becomes an increasingly pressing issue. Right now, heterogeneity in experimental design means that a comparison can only be made by comparing how a model performs relative to a common benchmark, like the GARCH (1, 1) model (though the benchmark model may be implemented differently, and heterogeneity may still persist), or by implementing the competing models for the same task (which may take significant work). This problem may be solved by introducing a shared task, unifying the efforts of the research community. The example shared task we have outlined is simple and the data is easily accessible with many training and test samples, but would allow the research community to make meaningful comparisons of model performance across publications without needing to implement competing models, and would also allow meta-analysis and results synthesis. However, the difficulty is not in designing a shared task, it is in the widespread adoption and standardisation of one. We believe this is the most pressing issue for the research field, and is where significant effort should be placed. This would not only be the responsibility of the researchers, but the reviewers too. In many fields of ML, such as computer vision, it is nearly impossible to introduce a novel concept, idea, or model, without reporting performance on some standard shared task, like the MNIST or ImageNet. Deeply ingraining a shared task in the field of financial volatility forecasting should be the next key goal in the future.

Another interesting observation is the large gap between the NN models in ML research, and the NN models applied to financial volatility forecasting. We identify a variety of reasons why this might be the case, and ultimately conclude that the best way to proceed is to bridge the gap, providing several approaches we feel might be promising, such as utilising state-of-the-art models adapted for time series data, forecasting a distribution to reflect confidence in the prediction, and also incorporating an attention mechanism into the model. Moving past this, we should look to specialise models by leveraging the structure and information only found in financial time series data. By doing so, this could move financial volatility forecasting into a distinct AI field of its own, similar to how computer vision and NLP have evolved. We also highlight the importance of deriving input variables from multiple assets, thus introducing more information that may yield better performance.

Financial volatility forecasting will always be an important task. It continues to grow, and with the increase in AI research and popularity, is becoming more accessible and exciting than ever. However, financial markets will also continue to evolve, exposing new relationships to exploit, which will eventually burn out to make way for new hidden relationships to uncover, constantly creating new opportunities and inspiring new avenues of research.

7 DATA AVAILABILITY

All data, extended results, tables, figures, and analysis methods can be found openly available online at https://github.com/Wenbo-G/NN-volatility-forecasting-review.

APPENDIX

A SUPPLEMENTARY TABLES

Table 1. Full Methods Results

Paper	NN method	Hybrid method	Transformation	Evolutionary algorithm
Dash, 2015 [34]	MLP (FLNN)	GARCH, Fuzzy	-	Differential harmony search, optimisation
Bucci, 2020 [17]	RNN	-	Cholesky decomposition	-
Barunik, 2016 [7]	MLP	HAR	-	-
Mostafa, 2015 [97]	MLP	-	-	-
Khan, 2017 [67]	MLP	-	-	-
Kumar, 2015 [75]	MLP	-	-	-
Pyo, 2018 [105]	MLP	-	-	-
Petnehazi, 2019 [102]	RNN (LSTM)	-	-	-
Nayak, 2020 [98]	MLP (ELM)	-	-	Chemical reaction optimisation, initialisation
Pradeepkumar, 2017 [104]	MLP (Quantile regression NN)	-	-	Particle swarm optimisation, optimisation
Marcek, 2018 [88]	MLP (Gaussian RBF)	-	-	-
Zhu, 2019 [133]	RNN (Bi-GRU)	-	VMD	-
Kim, 2018 [68]	RNN (LSTM)	GARCH, EWMA	-	-
Bildirici, 2015 [8]	MLP (default, RBF)	GARCH, LSTAR	-	-
Kristjanpoller, 2016 [71]	MLP	GARCH	-	-
Ramos-Perez, 2019 [106]	MLP (stacked ANN)	-	-	-
Kristjanpoller, 2015 [70]	MLP	GARCH	-	-
Vidal, 2020 [127]	CNN (VGG16), RNN (LSTM)	-	-	-

(Continued)

Table 1. Continued

Paper	NN method	Hybrid method	Transformation	Evolutionary algorithm
Das, 2016 [32]	MLP	Fuzzy, GARCH	-	-
Hajiabotorabi, 2019 [51]	RNN	-	Wavelet transform	-
Stefani, 2017 [118]	MLP	-	-	-
Fan, 2020 [41]	RNN (GRU)	-	-	-
Zhai, 2020 [132]	MLP	ARMA	Wavelet transform	-
Monfared, 2015 [95]	MLP	GARCH	-	-
Liu, 2019 [85]	RNN (LSTM)	-	-	-
Moon, 2019 [96]	RNN (LSTM)	-	-	-
Bisoi, 2015 [11]	MLP (FLNN and dynamic NN)	-	-	Differential evolution, initialisation
Kristjanpoller, 2017 [72]	MLP	GARCH	-	-
Harish, 2015 [56]	MLP	-	-	-
Luo, 2018 [86]	RNN	-	-	-
Tan, 2017 [121]	MLP (ANFIS)	Fuzzy	PCA	Fruit fly optimisation, optimisation
Lahmiri, 2016 [76]	MLP (General regression NN)	-	VMD	-
Dash, 2016 [33]	MLP (FLNN)	Fuzzy, GARCH	-	Differential evolution, optimisation
Geng, 2016 [48]	MLP (ANFIS)	Fuzzy, GARCH (CARRX)	-	-
Litimi, 2019 [84]	MLP	Chaos theory	-	-

The evolutionary algorithm column indicates the algorithm, and the purpose it was used for (that is, either optimisation or initialisation of the network).

Table 2. Full Task Results[1]

Paper	Volatility definition	Data[2]	Input structure	Input window (lag)[3]	Output window[4]
Dash, 2015 [34]	Historical volatility	**SENSEX, NIFTY**	Single asset, multivariate (log returns, volatility)	2	1–nearest option expiry
Bucci, 2020 [17]	Realised volatility	**S&P500, 10 yr treasury note futures, 1m treasury bond futures**	Multi-asset, multivariate (Cholesky factors, dividend price, price earning ratio, equity market return, FF 3 factors, 1 mo T-bill rate, term spread difference between log and short-term bond yields, inflation rate, industrial production growth, all lagged 1 period)	all (1)	1–21 (month)
Barunik, 2016 [7]	Realised volatility, other (jump based)	**Crude oil, heating oil, natural gas**	Single asset, univariate (volatility)	22	1, 1–5, 1–10
Mostafa, 2015 [97]	Implied volatility	**FTSE100**	Single asset, multivariate (moneyness (index and strike), time to expiry, historical volatility)	1	1
Khan, 2017 [67]	Historical volatility	**Crude oil (futures), S&P500**	Multi-asset, multivariate (S&P500 price, volatility of T value, weekly crude oil future prices, average of last 6 months returns crude oil future price, implied volatility of crude oil price)	1	1–126 (prev. 6 months)
Kumar, 2015 [75]	Historical volatility, other (range based)	**S&P500**	Single asset, univariate (volatility)	n/a	1–10
Pyo, 2018 [105]	Historical volatility (n/a)	**KOSPI200**	Single asset, univariate (volatility)	10 (months)	1–21 (month)
Petnehazi, 2019 [102]	Historical volatility, other (range based)	**DJIA (all constituents)**	Single asset, univariate (volatility proxy)	all (10)	1–21

(Continued)

Table 2. Continued

Paper	Volatility definition	Data[2]	Input structure	Input window (lag)[3]	Output window[4]
Nayak, 2020 [98]	Historical volatility	**SENSEX**	Single asset, univariate (sigmoid-normalised close price)	6	1–nearest option expiry
Pradeepkumar, 2017 [104]	Historical volatility	**Gold, crude oil, S&P500, NSE**	n/a	n/a	1–252
Marcek, 2018 [88]	Historical volatility (n/a)	**BUX**	Single asset, univariate (log returns)	2	1–n/a
Zhu, 2019 [133]	Historical volatility	**RUL9 (rubber)**	Single asset, multivariate (open, high, low, close, volume, VMD decomposed log returns, VMD decomposed 7-day volatility)	all (n/a)	1–7, 1–30
Kim, 2018 [68]	Historical volatility	**KOSPI200**, 3 yr Korea Treasury Bond, 3 yr AA grade corporate bond, gold, crude oil	Multi-asset, multivariate (KOSPI price, KOSPI log return, KOSPI differential rate of return at −1 time, 3 yr KTB interest rate, 3 yr AA grade corporate bonds interest rate, crude oil price, gold price, parameters estimated from GARCH, parameters estimated from EGARCH, parameters estimated from EWMA)	all (22)	1–n/a
Bildirici, 2015 [8]	Historical volatility (n/a)	**Brent crude oil**	n/a	5	1–2, 1–10, 1–40
Kristjanpoller, 2016 [71]	Historical volatility	**Oil (spot and futures)**, DJIA, EUR exchange rate, FTSE100, JPY exchange rate	Multi-asset, multivariate (GARCH predictions, square log returns, DJIA log returns, EUR exchange rate variation, FTSE log returns, JPY exchange rate variation)	252	1–14, 1–21, 1–28
Ramos-Perez, 2019 [106]	Historical volatility	**S&P500**	Single asset, multivariate (volatility (n=5), RF outputs, SVM outputs, GB outputs)	30	1–5

(Continued)

Table 2. Continued

Paper	Volatility definition	Data[2]	Input structure	Input window (lag)[3]	Output window[4]
Kristjanpoller, 2015 [70]	Historical volatility	**Gold (spot and futures)**, DJIA, EUR exchange rate, FTSE100, JPY exchange rate, oil	Multi-asset, multivariate (GARCH predictions, square log returns, DJIA log returns, EUR exchange rate variation, FTSE log returns, JPY exchange rate variation, daily price variation of oil)	252	1–14, 1–21, 1–28
Vidal, 2020 [127]	Historical volatility	**Gold (spot)**	Single asset, univariate (squared log returns)	all (4) for LSTM, 80 for CNN (an image is composed of 80 lags)	1–14
Das, 2016 [32]	Realised volatility (intraday)	**STI (all constituents)**	Single asset, multivariate (outputs of GARCH, EGARCH, GJR-GARCH, which use innovations, log variance, asymmetric shock, leverage effect, intraday volatility indicator (which is realised volatility, realised range, realised power variation, bipower variation))	5	1
Hajiabotorabi, 2019 [51]	Historical volatility	**S&P500, NASDAQ, DJIA, NYSE**	Single asset, multivariate (volatilities, log returns, forecast volatilities, wavelet decomposed volatilities)	3 (n/a if days or months)	1–22 (month)
Stefani, 2017 [118]	Historical volatility, other (range based), other (GARCH as filter)	**CAC40 (all constituents)**	Single asset, univariate (primary volatility proxy), multivariate (primary volatility proxy, additional volatility proxies, volume)	2, 5	1–2, 1–5, 1–8, 1–10, 1–12

(Continued)

Table 2. Continued

Paper	Volatility definition	Data[2]	Input structure	Input window (lag)[3]	Output window[4]
Fan, 2020 [41]	Historical volatility (n/a)	**S&P500, NASDAQ100, Nikkei225**	Single asset, multivariate (returns to the power of 1, 2, 3, 4)	20, 40, 60	1–n/a, 2–n/a, 3–n/a, 4–n/a, 5–n/a
Zhai, 2020 [132]	Realised volatility (intraday)	**CSI300**	Single asset, univariate (wavelet decomposed volatility)	4	1–5, 1–20, 1–100, 100–200, 260–360, 400–500
Monfared, 2015 [95]	Realised volatility	**S&P500 (five randomly selected: Allstate Corp., Walt Disney, WW Grainger, HP Company, Brown-Forman Corp.)**	Single asset, univariate (error from GARCH (1,1) model)	3	1–n/a
Liu, 2019 [85]	Realised volatility	**S&P500, AAPL**	Single asset, multivariate (volatility, log returns)	1, 2, 3, 4, 5	1–n/a, 3–n/a
Moon, 2019 [96]	Historical volatility (of stock price)	**S&P500, NASDAQ, DAX, KOSPI200, IPC**	Single asset, univariate (volatility), multivariate (volatility, open, high, low, volume, returns, MA(m1), MA(m2), MA(m3), EMA(m1), EMA(m2), EMA(m3))	all (n/a)	n/a
Bisoi, 2015 [11]	Historical volatility	**NIFTY, RIL**	Single asset, multivariate (volatility, log returns, SMA, EMA, MACD)	5	1—4, 3—2, 7–2, 15–10, 30–25
Kristjanpoller, 2017 [72]	Historical volatility	**Gold, silver, copper**, SZSE, FTSE100, SENSEX, US-EURO exchange, US-YEN exchange, oil	Multi-asset, multivariate (squared log returns, GARCH prediction, SZSE log returns, DJIA log returns, Euro, FTSE log returns, SENSEX log returns, Yen, oil price variation)	252	1–14, 1–21, 1–28
Harish, 2015 [56]	Historical volatility	**S&P500**	Single asset, multivariate (log returns, volatility)	Default 2	1–n/a

(Continued)

Table 2. Continued

Paper	Volatility definition	Data[2]	Input structure	Input window (lag)[3]	Output window[4]
Luo, 2018 [86]	Historical volatility (n/a)	**China A shares (all 162 constituents)**	Single asset, univariate (log returns)	all (n/a)	1–n/a
Tan, 2017 [121]	Historical volatility	**600827.SH, 600115.SH, 600018.SH**, SSE composite index	Multi-asset, multivariate (principal components of SSE composite index (close, high, low, volume, range of daily fluctuations, swing, price earnings ratio, price to book ratio, price to sales ratio, price cash flow ratio, dividend yield ratio) and the interested stock (close, high, low, volume, range of daily fluctuations, swing, price earnings ratio, price to book ratio, price to sales ratio, price cash flow ratio, dividend yield ratio, BOLL, BIAS, DMA, EXPMA, MACD, RSI, KDJ, stage high, stage low, up days, down days))	n/a	1–21 (month)
Lahmiri, 2016 [76]	Implied volatility	**VIX**	Single asset, multivariate (VMD decomposed VIX)	n/a	n/a
Dash, 2016 [33]	Historical volatility	**SENSEX, S&P500, NIFTY**	Single asset, multivariate (log returns, volatility)	2	1–nearest option expiry, 2–nearest option expiry, 3–nearest option expiry, 4–nearest option expiry, 5–nearest option expiry, 10–nearest option expiry

(Continued)

Table 2. Continued

Paper	Volatility definition	Data[2]	Input structure	Input window (lag)[3]	Output window[4]
Geng, 2016 [48]	Other (range based)	**CSI300, HSI**	Single asset, multivariate (log asset price, conditional range, log return, volume)	1	1
Litimi, 2019 [84]	Implied volatility	**VIX, VSTOXX, JNIV, VFTSE, VDAX, VCAC, VAEX, VSMI**	Single asset, univariate (volatility)	2-48	1

[1]n/a indicates that the relevant information is not available.

[2]Bold text refers to the variables which were fed into the forecasting model and for which volatility was predicted, i.e., the input and output variables. The non-bold text refers to variables that were only fed into the forecasting model, i.e., only used as input.

[3]The input window, or input lag, is defined as the number of previous timesteps required as input to the model to make a prediction. Note that for an RNN, this is labelled as "all" as it likely needs all previous observations to build context before making a prediction. A bracketed number should also appear if the input window is labelled "all"; this bracketed value refers to how many lags are present as input at a given timestep. As an example, at time t, the RNN will take in $\mathbf{X}_t$ as input, and would have taken all $[\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{t-1}]$ to build context. However, the input may be made up of observations from previous timesteps, i.e., $\mathbf{X}_t = [r_t, r_{t-1}, r_{t-2}]$. In this case, the input window would be labelled "all (3)."

[4]The output window notation is "start of window–end of window." As an example, "1–21" means the volatility forecasted was for the next trading month, the window starting 1 day into the future and ending 21 days into the future.

REFERENCES

[1] Chicago Board Options Exchange. 2019. Cboe Volatility Index [White Paper].

[2] Hans-Martin Adorf and Mark D. Johnston. 1990. A discrete stochastic neural network algorithm for constraint satisfaction problems. In *1990 IJCNN International Joint Conference on Neural Networks*. IEEE, 917–924, vol. 3. https://doi.org/10.1109/IJCNN.1990.137951

[3] Leif Andersen and Jesper Andreasen. 2000. Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing. *Review of Derivatives Research* 4, 3 (Oct. 2000), 231–262. https://doi.org/10.1023/A:1011354913068

[4] Torben G. Andersen and Tim Bollerslev. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 4 (Nov. 1998), 885. https://doi.org/10.2307/2527343

[5] Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 1 (July 2001), 43–76. https://doi.org/10.1016/S0304-405X(01)00055-1

[6] Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2001. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association.* 96, 453 (March 2001), 42–55. https://doi.org/10.1198/016214501750332965

[7] Jozef Baruník and Tomáš Křehlík. 2016. Combining high frequency data with non-linear models for forecasting energy market volatility. *Expert Systems with Applications* 55 (Aug. 2016), 222–242. https://doi.org/10.1016/j.eswa.2016.02.008

[8] Melike Bildirici and Özgür Ersin. 2015. Forecasting volatility in oil prices with a class of nonlinear volatility models: Smooth transition RBF and MLP neural networks augmented GARCH approach. *Petroleum Science* 12, 3 (Aug. 2015), 534–552. https://doi.org/10.1007/s12182-015-0035-8

[9] Christopher Bishop. 2006. *Pattern Recognition and Machine Learning*. Springer-Verlag, New York.

[10] Christopher M. Bishop. 1995. *Neural Networks for Pattern Recognition.* Oxford University Press, Inc.

[11] Ranjeeta Bisoi and Pradiptaishore Kishore Dash. 2015. Prediction of financial time series and its volatility using a hybrid dynamic neural network trained by sliding mode algorithm and differential evolution. *International Journal of Information and Decision Sciences* 7, 2 (2015), 166. https://doi.org/10.1504/IJIDS.2015.068757

[12] Tim Bollerslev. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 3 (April 1986), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

[13] John Bollinger. 1992. Using Bollinger bands. *Stocks & Commodities* 10, 2 (Feb. 1992), 11.

[14] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. 2015. *Time Series Analysis: Forecasting and Control.* John Wiley & Sons.

[15] Menachem Brenner and Marti G. Subrahmanyam. 1988. A simple formula to compute the implied standard deviation. *Financial Analysts Journal* 44, 5 (1988), 80–83.

[16] D. Broomhead and D. Lowe. 1988. Multivariable functional interpolation and adaptive networks. *Complex Systems* 2, 3 (1988), 321–355.

[17] Andrea Bucci. 2020. Cholesky–ANN models for predicting multivariate realized volatility. *Journal of Forecasting* 39, 6 (2020), 865–876. https://doi.org/10.1002/for.2664

[18] Thomas Busch, Bent Jesper Christensen, and Morten Ørregaard Nielsen. 2011. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics* 160, 1 (Jan. 2011), 48–57. https://doi.org/10.1016/j.jeconom.2010.03.014

[19] Andriana S. L. O. Campanharo, M. Irmak Sirer, R. Dean Malmgren, Fernando M. Ramos, and Luís A. Nunes Amaral. 2011. Duality between time series and networks. *PLoS One* 6, 8 (Aug. 2011), e23378. https://doi.org/10.1371/journal.pone.0023378

[20] Rodolfo C. Cavalcante, Rodrigo C. Brasileiro, Victor L. F. Souza, Jarley P. Nobrega, and Adriano L. I. Oliveira. 2016. Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications* 55 (Aug. 2016), 194–211. https://doi.org/10.1016/j.eswa.2016.02.006

[21] Kim Suk Chan and Howell Tong. 1986. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 3 (1986), 179–190. https://doi.org/10.1111/j.1467-9892.1986.tb00501.x

[22] Wendy W. Chapman, Prakash M. Nadkarni, Lynette Hirschman, Leonard W. D'Avolio, Guergana K. Savova, and Ozlem Uzuner. 2011. Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative solutions. *Journal of the American Medical Informatics Association (JAMIA)* 18, 5 (2011), 540–543. https://doi.org/10.1136/amiajnl-2011-000465

[23] Ray Yeutien Chou. 2005. Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking* 37, 3 (2005), 561–582.

[24] Ray Yeutien Chou, Hengchih Chou, and Nathan Liu. 2010. Range volatility models and their applications in finance. In *Handbook of Quantitative Finance and Risk Management*, Cheng-Few Lee, Alice C. Lee, and John Lee (Eds.). Springer US, Boston, MA, 1273–1281. https://doi.org/10.1007/978-0-387-77117-5_83

[25] Ray Yeutien Chou, Hengchih Chou, and Nathan Liu. 2015. Range volatility: A review of models and empirical studies. In *Handbook of Financial Econometrics and Statistics*, Cheng-Few Lee and John C. Lee (Eds.). Springer, New York, NY, 2029–2050. https://doi.org/10.1007/978-1-4614-7750-1_74

[26] Tommy W. S. Chow and Chi-Tat Leung. 1996. Neural network based short-term load forecasting using weather compensation. *IEEE Transactions on Power Systems* 11, 4 (Nov. 1996), 1736–1742. https://doi.org/10.1109/59.544636

[27] Bent Jesper Christensen and Nagpurnanand R. Prabhala. 1998. The relation between implied and realized volatility. *Journal of Financial Economics* 50, 2 (Nov. 1998), 125–150. https://doi.org/10.1016/S0304-405X(98)00034-8

[28] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*. arXiv:1412.3555.

[29] Richard Clarida, Jordi Gali, and Mark Gertler. 2000. Monetary policy rules and macroeconomic stability: Evidence and some theory. *The Quarterly Journal of Economics* 115, 1 (2000), 147–180.

[30] Charles J. Corrado and Thomas W. Miller, Jr. 2005. The forecast quality of CBOE implied volatility indexes. *Journal of Futures Markets* 25, 4 (2005), 339–373. https://doi.org/10.1002/fut.20148

[31] Fulvio Corsi. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 2 (March 2009), 174–196. https://doi.org/10.1093/jjfinec/nbp001

[32] Ron Tor Das, Kai Keng Ang, and Chai Quek. 2016. ieRSPOP: A novel incremental rough set-based pseudo outer-product with ensemble learning. *Applied Soft Computing* 46 (Sept. 2016), 170–186. https://doi.org/10.1016/j.asoc.2016.04.015

[33] Rajashree Dash and Pradiptaishore Kishore Dash. 2016. An evolutionary hybrid fuzzy computationally efficient EGARCH model for volatility prediction. *Applied Soft Computing* 45 (Aug. 2016), 40–60. https://doi.org/10.1016/j.asoc.2016.04.014

[34] Rajashree Dash, Pradiptaishore Kishore Dash, and Ranjeeta Bisoi. 2015. A differential harmony search based hybrid interval type2 fuzzy EGARCH model for stock market volatility prediction. *International Journal of Approximate Reasoning* 59 (April 2015), 81–104. https://doi.org/10.1016/j.ijar.2015.02.001

[35] Darrell Duffie and Jun Pan. 1997. An overview of value at risk. *The Journal of Derivatives* 4, 3 (Feb. 1997), 7–49. https://doi.org/10.3905/jod.1997.407971

[36] Jean-Pierre Eckmann, Sylvie Oliffson Kamphorst, and David Ruelle. 1987. Recurrence plots of dynamical systems. *Europhysics Letters (EPL)* 4, 9 (Nov. 1987), 973–977. https://doi.org/10.1209/0295-5075/4/9/004

[37] Sebastian Edwards, Javier Gómez Biscarri, and Fernando Pérez de Gracia. 2003. Stock market cycles, financial liberalization and volatility. *Journal of International Money and Finance* 22, 7 (Dec. 2003), 925–955. https://doi.org/10.1016/j.jimonfin.2003.09.011

[38] Robert F. Engle. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50, 4 (July 1982), 987–1008. https://doi.org/10.2307/1912773

[39] Robert F. Engle and Andrew J. Patton. 2007. 2 - What good is a volatility model?*. In *Forecasting Volatility in the Financial Markets* (3rd. ed.), John Knight and Stephen Satchell (Eds.). Butterworth-Heinemann, Oxford, 47–63. https://doi.org/10.1016/B978-075066942-9.50004-2

[40] Marisa Faggini. 2014. Chaotic time series analysis in economics: Balance and perspectives. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 24, 4 (Dec. 2014), 042101. https://doi.org/10.1063/1.4903797

[41] Xiangru Fan, Xiaoqian Wei, Di Wang, Wen Zhang, and Wu Qi. 2020. Multi-step prediction of financial asset return volatility using parsimonious autoregressive sequential model. In *Mining Data for Financial Applications*, Lecture Notes in Computer Science, Valerio Bitetta, Ilaria Bordino, Andrea Ferretti, Francesco Gullo, Stefano Pascolutti, and Giovanni Ponti (Eds.). Springer International Publishing, Cham, 107–121. https://doi.org/10.1007/978-3-030-37720-5_9

[42] Michele Filannino and Özlem Uzuner. 2018. Advancing the state of the art in clinical natural language processing through shared tasks. *Yearbook of Medical Informatics* 27, 1 (Aug. 2018), 184–192. https://doi.org/10.1055/s-0038-1667079

[43] Jeff Fleming. 1998. The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance* 5, 4 (Oct. 1998), 317–345. https://doi.org/10.1016/S0927-5398(98)00002-4

[44] Jürgen Franke and Michael H. Neumann. 2000. Bootstrapping neural networks. *Neural Computation* 12, 8 (Aug. 2000), 1929–1949. https://doi.org/10.1162/089976600300015204

[45] Kunihiko Fukushima. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36, 4 (April 1980), 193–202. https://doi.org/10.1007/BF00344251

[46] Mark B. Garman and Michael J. Klass. 1980. On the estimation of security price volatilities from historical data. *The Journal of Business* 53, 1 (1980), 67–78.

[47] Wenbo Ge, Christian J. Lueck, Deborah Apthorp, and Hanna Suominen. 2020. Which features of postural sway are effective in distinguishing Parkinson's disease from controls? A systematic review. *Brain and Behavior* (Nov. 2020), e01929. https://doi.org/10.1002/brb3.1929

[48] Liyan Geng and Zhanfu Zhang. 2016. Application of ANFIS-based CARRX model to stock volatility forecasting. *International Journal of Simulation Systems, Science & Technology* 17, 22 (Jan. 2016), 7.

[49] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. The MIT Press.

[50] Dominique M. Guillaume, Michel M. Dacorogna, Rakhal R. Davé, Ulrich A. Müller, Richard B. Olsen, and Olivier V. Pictet. 1997. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics* 1, 2 (April 1997), 95–129. https://doi.org/10.1007/s007800050018

[51] Zeinab Hajiabotorabi, Aliyeh Kazemi, Faramarz Famil Samavati, and Farid Mohammad Maalek Ghaini. 2019. Improving DWT-RNN model via B-spline wavelet multiresolution to forecast a high-frequency time series. *Expert Systems with Applications* 138 (Dec. 2019), 112842. https://doi.org/10.1016/j.eswa.2019.112842

[52] Peter Reinhard Hansen. 2005. A test for superior predictive ability. *Journal of Business & Economic Statistics* 23, 4 (2005), 365–380.

[53] Peter R. Hansen and Asger Lunde. 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 7 (2005), 873–889. https://doi.org/10.1002/jae.800

[54] Peter R. Hansen, Asger Lunde, and James M. Nason. 2011. The model confidence set. *Econometrica* 79, 2 (2011), 453–497.

[55] Mary R. Hardy. 2001. A regime-switching model of long-term stock returns. *North American Actuarial Journal* 5, 2 (April 2001), 41–53. https://doi.org/10.1080/10920277.2001.10595984

[56] Vijayalaxmi Harish, Chandrashekara S. Adiga, H. G. Joshi, and S. V. Harish. 2015. Volatility prediction model for option pricing: A soft computing approach. *International Journal of Soft Computing* 10, 6 (2015), 391–399. https://doi.org/10.3923/ijscomp.2015.391.399

[57] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (Nov. 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[58] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 5 (Jan. 1989), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

[59] David A. Hsieh. 1991. Chaos and nonlinear dynamics: Application to financial markets. *The Journal of Finance* 46, 5 (1991), 1839–1877. https://doi.org/10.1111/j.1540-6261.1991.tb04646.x

[60] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. 2006. Extreme learning machine: Theory and applications. *Neurocomputing* 70, 1 (Dec. 2006), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

[61] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454, 1971 (March 1998), 903–995. https://doi.org/10.1098/rspa.1998.0193

[62] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery* 33, 4 (July 2019), 917–963. https://doi.org/10.1007/s10618-019-00619-1

[63] Jacques Longerstaey and Martin Spencer. 1996. RiskMetrics Technical Document, 4th ed.

[64] Jyh-Shing Roger Jang. 1993. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* 23, 3 (May 1993), 665–685. https://doi.org/10.1109/21.256541

[65] Carlos M. Jarque and Anil K. Bera. 1987. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique* 55, 2 (1987), 163–172. https://doi.org/10.2307/1403192

[66] Ioannis Karatzas and Steven Shreve. 1998. *Brownian Motion and Stochastic Calculus* (2nd ed.). Springer-Verlag, New York. https://doi.org/10.1007/978-1-4612-0949-2

[67] Saqib Khan, Hamed Shafiee Hasanabadi, and Rene Mayorga. 2017. Determining the relationship between speculative activity and crude oil price volatility, using artificial neural networks. In *2017 International Conference on Information and Communication Technologies (ICICT'17)*. IEEE, 138–144. https://doi.org/10.1109/ICICT.2017.8320178

[68] Ha Young Kim and Chang Hyun Won. 2018. Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications* 103 (Aug. 2018), 25–37. https://doi.org/10.1016/j.eswa.2018.03.002

[69] John F. Kolen and Stefan C. Kremer. 2001. Gradient flow in recurrent nets: The difficulty of learning LongTerm dependencies. In *A Field Guide to Dynamical Recurrent Networks*. IEEE, 237–243. https://doi.org/10.1109/9780470544037.ch14

[70] Werner Kristjanpoller and Marcel C. Minutolo. 2015. Gold price volatility: A forecasting approach using the artificial neural network–GARCH model. *Expert Systems with Applications* 42, 20 (Nov. 2015), 7245–7251. https://doi.org/10.1016/j.eswa.2015.04.058

[71] Werner Kristjanpoller and Marcel C. Minutolo. 2016. Forecasting volatility of oil price using an artificial neural network-GARCH model. *Expert Systems with Applications* 65 (Dec. 2016), 233–241. https://doi.org/10.1016/j.eswa.2016.08.045
[72] Werner Kristjanpoller R. and Esteban Hernández P. 2017. Volatility of main metals forecasted by a hybrid ANN-GARCH model with regressors. *Expert Systems with Applications* 84 (Oct. 2017), 290–300. https://doi.org/10.1016/j.eswa.2017.05.024
[73] Mark Kritzman. 1991. What practitioners need to know about estimating volatility. Part 1. *Financial Analysts Journal* 47, 4 (1991), 22–25.
[74] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 1097–1105.
[75] Hemanth Kumar P. and S. Basavaraj Patil. 2015. Estimation forecasting of volatility using ARIMA, ARFIMA and neural network based techniques. In *2015 IEEE International Advance Computing Conference (IACC'15)*. IEEE, 992–997. https://doi.org/10.1109/IADCC.2015.7154853
[76] Salim Lahmiri. 2016. A variational mode decomposition approach for analysis and forecasting of economic and financial time series. *Expert Systems with Applications* 55 (Aug. 2016), 268–273. https://doi.org/10.1016/j.eswa.2016.02.025
[77] Alan S. Lapedes and Robert Farber. 1987. Nonlinear signal processing using neural networks: Prediction and system modelling. In *IEEE International Conference on Neural Networks*.
[78] Colin Lea, Michael D. Flynn, Rene Vidal, Austin Reiter, and Gregory D. Hager. 2017. Temporal convolutional networks for action segmentation and detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)*. IEEE, 1003–1012. https://doi.org/10.1109/CVPR.2017.113
[79] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 11 (Nov. 1998), 2278–2324. https://doi.org/10.1109/5.726791
[80] Jaehoon Lee, Yasaman Bahri, Roman Novak, Samuel S. Schoenholz, Jeffrey Pennington, and Jascha Sohl-Dickstein. 2018. Deep neural networks as gGaussian processes. In *6th International Conference on Learning Representations (ICLR'18)*. arXiv:1711.00165.
[81] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. 2019. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc.
[82] Bryan Lim, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister. 2020. Temporal fusion transformers for interpretable multi-horizon time series forecasting. arXiv:1912.09363 [cs, stat].
[83] Tsungnan Lin, Bill G. Horne, Peter Tino, and C. Lee Giles. 1996. Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks* 7, 6 (Nov. 1996), 1329–1338. https://doi.org/10.1109/72.548162
[84] Houda Litimi, Ahmed BenSaida, Lotfi Belkacem, and Oussama Abdallah. 2018. Chaotic behavior in financial market volatility. *Journal of Risk* 21, 3 (2018), 27–53. https://doi.org/10.21314/JOR.2018.400
[85] Yang Liu. 2019. Novel volatility forecasting using deep learning–long short term memory recurrent neural networks. *Expert Systems with Applications* 132 (Oct. 2019), 99–109. https://doi.org/10.1016/j.eswa.2019.04.038
[86] Rui Luo, Weinan Zhang, Xiaojun Xu, and Jun Wang. 2018. A neural stochastic volatility model. In *AAAI*. AAAI Press.
[87] Burton G. Malkiel. 1989. Efficient market hypothesis. In *Finance*, John Eatwell, Murray Milgate, and Peter Newman (Eds.). Palgrave Macmillan UK, London, 127–134. https://doi.org/10.1007/978-1-349-20213-3_13
[88] Dusan Marcek. 2018. Forecasting of financial data: A novel fuzzy logic neural network based on error-correction concept and statistics. *Complex & Intelligent Systems* 4, 2 (June 2018), 95–104. https://doi.org/10.1007/s40747-017-0056-6
[89] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The penn treebank. *Computational Linguistics* 19, 2 (1993), 313–330.
[90] Philippe Masset. 2011. *Volatility Stylized Facts*. SSRN Scholarly Paper ID 1804070. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.1804070
[91] Stewart Mayhew. 1995. Implied volatility. *Financial Analysts Journal* 51, 4 (1995), 8–20. https://doi.org/10.2469/faj.v51.n4.1916
[92] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. 2018. The natural language decathlon: Multitask learning as question answering. arXiv:1806.08730 [cs, stat].
[93] Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. 2016. Pointer sentinel mixture models. arXiv:1609.07843 [cs]
[94] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* 6, 7 (July 2009), e1000097. https://doi.org/10.1371/journal.pmed.1000097

[95] Soheil Almasi Monfared and David Enke. 2015. Noise canceling in volatility forecasting using an adaptive neural network filter. *Procedia Computer Science* 61 (2015), 80–84. https://doi.org/10.1016/j.procs.2015.09.155

[96] Kyoung-Sook Moon and Hongjoong Kim. 2019. Performance of deep learning in prediction of stock market volatility. *Economic Computation and Economic Cybernetics Studies and Research* 53, 2/2019 (June 2019), 77–92. https://doi.org/10.24818/18423264/53.2.19.05

[97] Fahed Mostafa, Tharam Dillon, and Elizabeth Chang. 2015. Computational intelligence approach to capturing the implied volatility. In *Artificial Intelligence in Theory and Practice IV*, Tharam Dillon (Ed.). Vol. 465. Springer International Publishing, Cham, 85–97. https://doi.org/10.1007/978-3-319-25261-2_8

[98] Sarat Chandra Nayak and Bijan Bihari Misra. 2020. Extreme learning with chemical reaction optimization for stock volatility prediction. *Financial Innovation* 6, 1 (Feb. 2020), 16. https://doi.org/10.1186/s40854-020-00177-2

[99] Mehmet Orhan and Bülent Köksal. 2012. A comparison of GARCH models for VaR estimation. *Expert Systems with Applications* 39, 3 (Feb. 2012), 3582–3592. https://doi.org/10.1016/j.eswa.2011.09.048

[100] Michael Parkinson. 1980. The extreme value method for estimating the variance of the rate of return. *The Journal of Business* 53, 1 (1980), 61–65.

[101] Andrew J. Patton. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 1 (Jan. 2011), 246–256. https://doi.org/10.1016/j.jeconom.2010.03.034

[102] Gábor Petneházi and József Gáll. 2019. Exploring the predictability of range-based volatility estimators using recurrent neural networks. *Intelligent Systems in Accounting, Finance and Management* 26, 3 (2019), 109–116. https://doi.org/10.1002/isaf.1455

[103] Ser-Huang Poon and Clive W. J. Granger. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41, 2 (2003), 478–539.

[104] Dadabada Pradeepkumar and Vadlamani Ravi. 2017. Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. *Applied Soft Computing* 58 (Sept. 2017), 35–52. https://doi.org/10.1016/j.asoc.2017.04.014

[105] Sujin Pyo and Jaewook Lee. 2018. Exploiting the low-risk anomaly using machine learning to enhance the Black–Litterman framework: Evidence from South Korea. *Pacific-Basin Finance Journal* 51 (Oct. 2018), 1–12. https://doi.org/10.1016/j.pacfin.2018.06.002

[106] Eduardo Ramos-Pérez, Pablo J. Alonso-González, and José Javier Núñez-Velázquez. 2019. Forecasting volatility with a stacked model based on a hybridized artificial neural network. *Expert Systems with Applications* 129 (Sept. 2019), 1–9. https://doi.org/10.1016/j.eswa.2019.03.046

[107] Leonard Christopher Gordon Rogers and Stephen E. Satchell. 1991. Estimating variance from high, low and closing prices. *The Annals of Applied Probability* 1, 4 (1991), 504–512.

[108] Leonard Christopher Gordon Rogers, Stephen E. Satchell, and Y. Yoon. 1994. Estimating the volatility of stock prices: A comparison of methods that use high and low prices. *Applied Financial Economics* 4, 3 (June 1994), 241–247. https://doi.org/10.1080/758526905

[109] Hannah R. Rothstein, Alexander J. Sutton, and Michael Borenstein. 2006. Publication bias in meta-analysis. In *Publication Bias in Meta-Analysis*. John Wiley & Sons, Ltd, 1–7.

[110] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (Dec. 2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y

[111] Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural Networks* 61 (Jan. 2015), 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

[112] G. William Schwert. 1990. Stock market volatility. *Financial Analysts Journal* 46, 3 (1990), 23–34.

[113] Biswa Sengupta and Karl J. Friston. 2018. How robust are deep neural networks?. In *Proceedings of ICML (Workshop on Theoretical Foundations and Applications of Deep Generative Models)*.

[114] Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing* 90 (May 2020), 106181. https://doi.org/10.1016/j.asoc.2020.106181

[115] Mark J. Shensa. 1992. The discrete wavelet transform: Wedding the a trous and mallat algorithms. *IEEE Transactions on Signal Processing* 40, 10 (Oct. 1992), 2464–2482. https://doi.org/10.1109/78.157290

[116] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. 2017. Continual learning with deep generative replay. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 2990–2999.

[117] Jinghong Shu and Jin E. Zhang. 2006. Testing range estimators of historical volatility. *Journal of Futures Markets* 26, 3 (2006), 297–313. https://doi.org/10.1002/fut.20197

[118] Jacopo De Stefani, Olivier Caelen, Dalila Hattab, and Gianluca Bontempi. 2017. Machine learning for multi-step ahead forecasting of volatility proxies. In *Workshop on MIning DAta for Financial Applications (MIDAS@PKDD/ECML)*. 17–28.

[119] Dan Stefanica and Rados Radoicic. 2017. An explicit implied volatility formula. *International Journal of Theoretical and Applied Finance* 20, 7 (Jan. 2017). https://doi.org/10.2139/ssrn.2908494

[120] Hanna Suominen, Liadh Kelly, and Lorraine Goeuriot. 2018. Scholarly influence of the conference and labs of the evaluation forum eHealth initiative: Review and bibliometric study of the 2012 to 2017 outcomes. *JMIR Research Protocols* 7, 7 (July 2018), e10961. https://doi.org/10.2196/10961

[121] Lijun Tan, Shiheng Wang, and Ke Wang. 2017. A new adaptive network-based fuzzy inference system with adaptive adjustment rules for stock market volatility forecasting. *Information Processing Letters* 127 (Nov. 2017), 32–36. https://doi.org/10.1016/j.ipl.2017.06.012

[122] James W. Taylor. 2000. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting* 19, 4 (2000), 299–311. https://doi.org/10.1002/1099-131X(200007)19:4<299::AID-FOR775>3.0.CO;2-V

[123] Thorsten Thadewald and Herbert Büning. 2007. Jarque–Bera test and its competitors for testing normality—A power comparison. *Journal of Applied Statistics* 34, 1 (Jan. 2007), 87–105. https://doi.org/10.1080/02664760600994539

[124] Allan Timmermann and Clive W. J. Granger. 2004. Efficient market hypothesis and forecasting. *International Journal of Forecasting* 20, 1 (Jan. 2004), 15–27. https://doi.org/10.1016/S0169-2070(03)00012-8

[125] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A generative model for raw audio. arXiv:1609.03499 [cs].

[126] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, 6000–6010.

[127] Andrés Vidal and Werner Kristjanpoller. 2020. Gold volatility prediction using a CNN-LSTM approach. *Expert Systems with Applications* 157 (Nov. 2020), 113481. https://doi.org/10.1016/j.eswa.2020.113481

[128] Zhiguang Wang and Tim Oates. 2015. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In *Workshops at the 29th AAAI Conference on Artificial Intelligence*. 8.

[129] Halbert White. 1990. Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks* 3, 5 (Jan. 1990), 535–549. https://doi.org/10.1016/0893-6080(90)90004-5

[130] Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2, 1 (1987), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

[131] Dennis Yang and Qiang Zhang. 2000. Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business* 73, 3 (2000), 477–492. https://doi.org/10.1086/209650

[132] Jia Zhai, Yi Cao, and Xiaoquan Liu. 2020. A neural network enhanced volatility component model. *Quantitative Finance* 20, 5 (May 2020), 783–797. https://doi.org/10.1080/14697688.2019.1711148

[133] Qing Zhu, Fan Zhang, Shan Liu, Yiqiong Wu, and Lin Wang. 2019. A hybrid VMD–BiGRU model for rubber futures time series forecasting. *Applied Soft Computing* 84 (Nov. 2019), 105739. https://doi.org/10.1016/j.asoc.2019.105739

Received December 2020; revised May 2021; accepted August 2021