



A flexible and lightweight deep learning weather forecasting model

Gabriel Zenkner¹ · Salvador Navarro-Martinez¹

Accepted: 20 June 2023 / Published online: 1 August 2023
© The Author(s) 2023

Abstract

Numerical weather prediction is an established weather forecasting technique in which equations describing wind, temperature, pressure and humidity are solved using the current atmospheric state as input. This study examines deep learning to forecast weather given historical data from two London-based locations. Two distinct Bi-LSTM recurrent neural network models were developed in the TensorFlow deep learning framework and trained to make predictions in the next 24 and 72 h, given the past 120 h. The first trained neural network predicted temperature at Kew Gardens with a forecast accuracy of $\pm 2^\circ\text{C}$ in 73% of instances in a whole unseen year, and a root mean squared errors of 1.45°C . The second network predicted 72-h air temperature and relative humidity at Heathrow with root mean squared errors 2.26°C and 14% respectively and 80% of the temperature predictions were within $\pm 3^\circ\text{C}$ while 80% of relative humidity predictions were within $\pm 20\%$. Both networks were trained with five years of historical data, with cloud training times of over a minute (24-h network) and three minutes (72-h).

Keywords Recurrent Neural Network · Bi-LSTM · Weather Forecast

1 Introduction

Numerous sectors are heavily reliant on accurate weather forecasting including renewable energy production, energy consumption, agriculture and emergency services. Numerical weather prediction is an established weather forecasting technique in which the transport fluid equations momentum, energy and scalar transport are solved using the current atmospheric state as an input. The output is the temperature, humidity, pressure, etc. in a desired forecast length. Modelling large scale weather is notoriously difficult due to uncertain boundary conditions and the chaotic nature of the underlying fluid mechanics equations. The accuracy of numerical forecast predictions has improved steadily since the 1960s, carried mostly by the increase of computational power and turbulence modelling techniques [1]. To reduce the uncertainty of the predictions, expensive ensemble modelling is used, where simulations are run many times with small differences in initial conditions. Beyond five days, chaotic

effects become dominant and the simulations demand large computational resources and are exceedingly expensive [2]

Ensemble modelling is computationally demanding requiring numerous runs of each model with different initial conditions. To make meaningful seasonal predictions, the number of runs should be between 100 and 200 [3], increasing the cost 100-fold over deterministic approaches. Moreover, the multi-scale nature of the fluid equations and physical processes associated creates simplifications and the initial state approximation may be inaccurate [4]. Similarly, the acquisition of representative initial conditions is one of the biggest hurdles in numerical weather prediction [5]. This characterisation process becomes increasingly challenging in cities where landscape drastically affects wind and temperature behaviour. Machine learning approach can complement existing numerical weather prediction, or in some cases even substitute it, thereby reducing the enormous computational demands associated with numerical weather prediction.

The present work proposes to use historical data of weather stations to produce short-term *local* forecasts. The locality of the data and forecast simplifies the complexity of spatial correlations that exist in turbulent fluid dynamics and reduces the size and training of the network. Moreover, local data is attractive for Deep Learning, which can account for the "unpredictability" of the local conditions.

The novelty of the method resides in the use of large historical data of nearby locations, to create simple input–output network models independent of the date. The approach is

✉ Salvador Navarro-Martinez
s.navarro@imperial.ac.uk

Gabriel Zenkner
g.zenkner@imperial.ac.uk

¹ Department Mechanical Engineering, Imperial College
London, Exhibition Road, London SW7 2AZ, UK

purely data-driven, without any kind of data assimilation or hybridisation. The model is tested using historical data from two London-based locations to train a Bi-LSTM recurrent neural network to predict temperature and relative humidity.

The main contributions of this article are:

- The creation of a Deep Neural Network framework to use historical weather data to create forecasts of selected weather features over desired length.
- The development of two models to predict temperature and humidity hourly evolution over 24 and 72 h in two locations in London.
- The study of forecasting errors investigating seasonal variations and forecast length.

The rest of the paper is structured as follows. In Section 2, the relevant literature related to the use of Machine Learning in Weather forecasting is discussed, while in Section 3, the architecture and the dataset used for testing is described. In Section 4, the results with the two models developed are

presented, while Section 5 concludes the paper and outlines future research directions.

2 Related work

Machine Learning (ML) is showing large potential in fluid mechanics [6, 7], where it can be used to model sub-grid stress [8, 9] or extract turbulent structures [10]. One of the first ML applications in weather forecasting was Schizas et al. [11] in 1991, where Artificial Neural Networks (ANN), were used to predict minimum temperatures. Similarly, Ochiai et al. [12] used ANN in 1995 to predict rainfall and snowfall. These models were able to improve the forecasting accuracy compared to statistical models [13]. However, the limited forecast of 30–180 min and difficulties in obtaining solution convergence made practical application impossible. Traditional machine learning examples include support vector machine or linear regression which are typically far less computationally

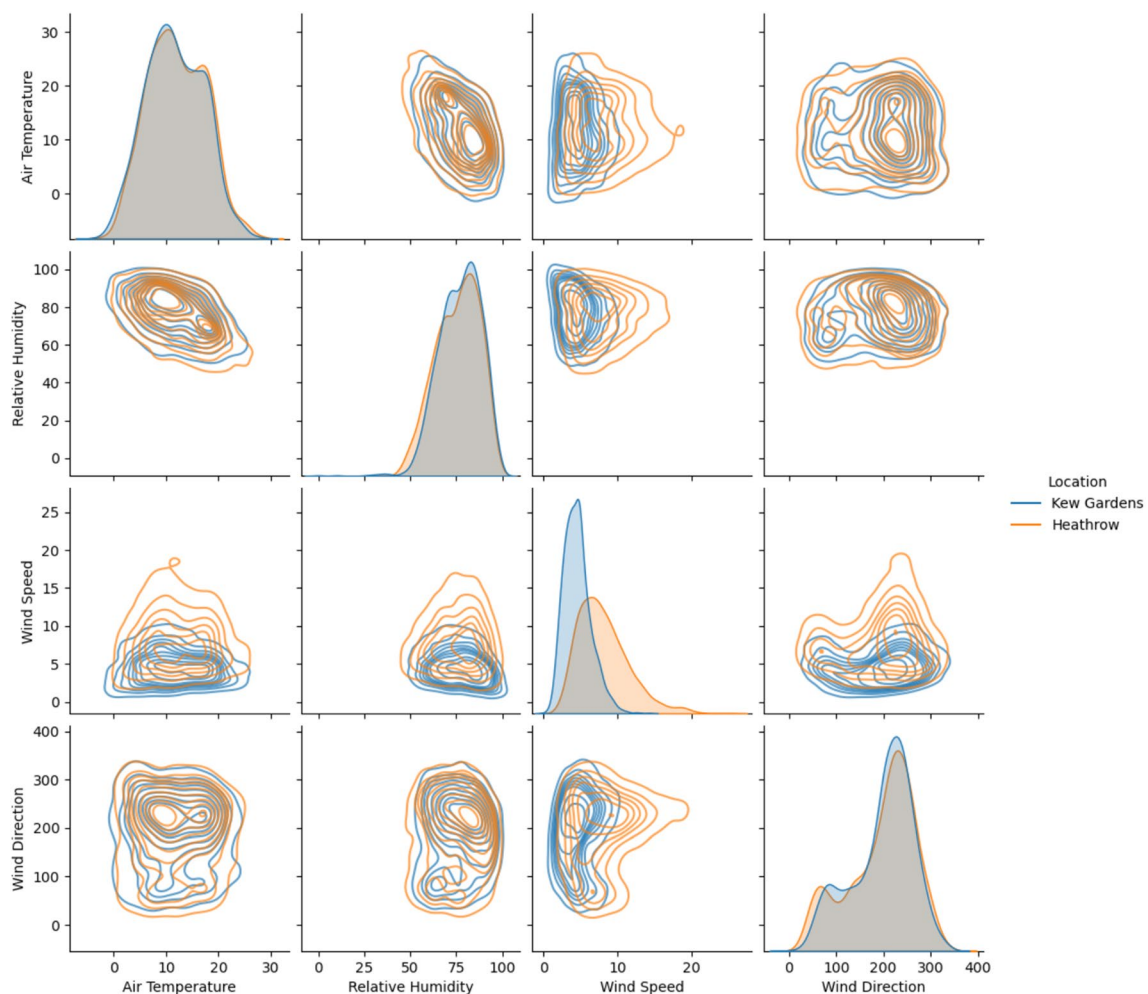


Fig. 1 Joint probability density functions of two features (off-diagonal) and single-feature probability density functions (diagonal) for the two locations

demanding than neural networks and have been investigated as forecasting candidates. For example, Ma et al. [14] deployed a traditional machine learning model known as XGBoost, which are comprised of gradient boosted decision trees, to predict air temperature and humidity over a 3-h period with resulting root mean square errors (RMSE) of temperature of 1.77 °C. Despite the relatively good result of traditional machine learning approaches, there are several reasons why a deep learning approach is preferred for weather prediction. Traditional algorithms are unable to model non-linearity, which is essential in predicting the evolution of the weather [15, 16]. Similarly, Shao et al. [17] reported that statistical and traditional ML techniques are not well-suited for complex wind forecasting and attribute this need to the turbulent and chaotic behaviour of wind. Recent efforts have focused on using Support Vector Machines and variations for short term series forecasting and classification of non-linear data and time series [18–20]. Deep Learning (DL) leverages the growing volume and accessibility of data. While traditional machine learning models reach a point beyond which additional training data no longer improves model performance, deep learning models have been observed to benefit from the increase in data [21]. DL networks have been increasingly used in time series forecasting in several applications, examples include finance [22], sugarcane yield prediction [23] and power load forecasting [24] among others. DL has the potential to significantly improve the accuracy of weather forecasting and its applications increased exponentially. Bauer et al. [4] showed that their Convolutional Neural Network (CNN) ensemble forecasting model can predict anomalies such as Hurricane Irma. Weyn et al. [25] increased the accuracy of weather prediction by applying ensemble modelling of separate CNN models, each with different starting conditions and sets of weights. Roy et al. [26] evaluated a multilayer perceptron, a long short-term memory (LSTM) model and a hybrid CNN/LSTM model and concludes that models with more complex architectures in general improve performance, while Ravuri et al. [27] demonstrated that their neural network model can

Table 1 Architecture of the Bi-LSTM used in Model A, which includes the number and type of layers and the number of nodes in each layer

Layer	Type	Value	Shape	Parameters
Input	-	-	(120×6)	0
Hidden	Bi-LSTM	Tanh activation function	(32×512)	538,624
Hidden	Dropout	0.25	(32×512)	0
Output	Linear	-	(32×6)	3,078
Total				541,702

predict precipitation more accurately in 89% of instances compared to existing weather prediction techniques. Hewage et al. [13] report that their ML models predict weather conditions 12 h into the future with higher accuracy than conventional weather forecasting.

Neural networks have been identified as being particularly promising in precipitation forecasting. A MetNet model developed at Google [28] was shown to predict precipitation accurately over the course of eight hours. In this hybrid approach, several models were used at different stages including LSTMs and CNNs. Despite its good performance, the model requires large volumes of data. An improvement was obtained by Met-Net2 [29], outperforming up to 12 h state-of-the-art weather models operating in the Continental United States. Fu et al. [30], upon evaluating many neural network architectures, settled on a combined Bidirectional-LSTM (Bi-LSTM) and a one-dimensional CNN to predict ground air temperature, relative humidity and wind speed over seven days. They used weather station data from ten weather stations in Beijing and the final model contained more than a million nodes. Despite its size and complexity, the quantitative performance relative to the local weather observations was uncertain. The latest trends among others, include the use of hybrid LSTM/GAN [31] to predict cloud movement, LSTM/CNN for drought forecast [32]. Wind forecasting is of great importance in wind power and load estimations and DL has been recently applied [33–36]. Most of the applications focused on short term which sped up prediction by up to 24 h.

The recent literatures shows that DL applications in weather forecast are accelerating, with large-scale forecasts using CNN-variant architectures and LSTM dominating point forecast. However, there are clearly several research bottlenecks associated with short-term forecasting. Most applications have been in wind-farm sites with "simple" weather patterns, while urban environments are more complex to predict as the turbulence content of the signal is larger. Moreover, there is a deterioration

Table 2 Parameters used in Model A including number of epochs and optimiser settings

Parameter	Value
Context Length	120 h
Gradient Optimisation	Adaptive moment estimation (ADAM)
Learning rate	0.001
Model optimised metric	Mean squared error
Performance metric	Root mean squared error
Epochs	2
Batch size	32
Runtime size	78 s
Train, validate, test ratios	0.7, 0.15 and 0.15

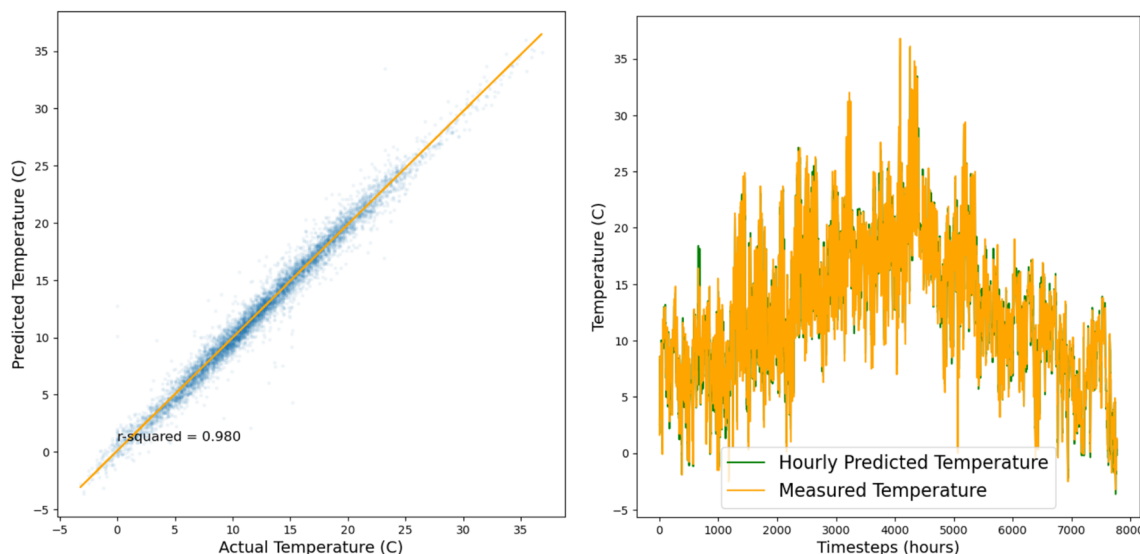


Fig. 2 Comparison between predicted and measured temperature at Kew Gardens using the forecast length of one hour and a context length of 120 h. Scatter plot (left), one-year predictions (right)

of predictions after several hours and there is not an optimal forecast length, which seems to depend on application.

3 Methodology and data processing

LSTMs are applied frequently in sequential problems as they address the issue of loss of long-term memory [37]. The Bi-LSTM recurrent neural network builds upon the LSTM structure. In a Bi-LSTM model a duplicate layer is produced, where sequential information flows in chronological order through the first layer while the duplicate layer is used for the same sequential information, but in reversed order. This provides the model with far more context as key information at both the start and end of the sequence is available.

The training data is openly available by the Met Office from two London weather observation stations: Kew Gardens (51.482, -0.294) and Heathrow (51.479, -0.451). The data was

extracted from the Centre for Environmental Data Analysis [38] and contains weather information from 2015–2021 with dozens of hourly weather parameters, hereinafter referred to as *features* for consistency. However, not all features are available for all weather stations and so the selection was limited to six unique features (three per weather station). The features of particular interest are air temperature, relative humidity and wind speed at both Heathrow and Kew Gardens, see Fig. 1.

With the features selected, the dataset is normalised. This is performed by using the mean and standard deviation for each feature. The mean and standard deviation are calculated from the training dataset, as including data from the validation and test sets and may result in overfitting [39].

The training, validation and test datasets are split up in fractions of 0.7, 0.15 and 0.15 respectively with the chronological sequence of the data maintained. This corresponds to a sample size of 36,825, 7,891 and 7,892 observations respectively.

Two networks were created, **Model A**, to forecast 24 h and **Model B** to predict 72 h. The same dataset with the

Fig. 3 Comparison between predicted and measured temperature at Kew Gardens using the 24-h and 1-h temperature predictions

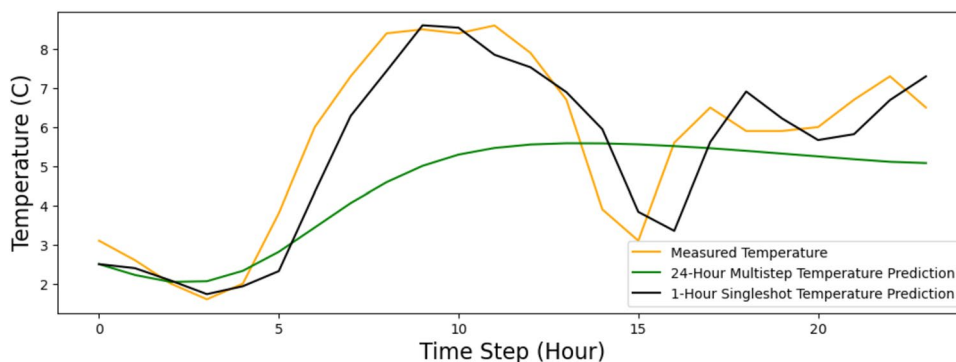


Table 3 Root mean squared error (RMSE), mean average error (MAE) and maximum error between hourly and 24-h temperature predictions in Fig. 3

	RMSE [°C]	MAE [°C]	Max. Error [°C]
Single timestep	0.86	0.63	2.19
Multi-timestep	1.74	1.33	4.76

same split ratio for training, validation, and testing was used in both models. However, Model B is deeper, with a denser Bi-LSTM with more cells and an additional Feed Forward neural network (FNN) in the second hidden layer. Model B was trained on the same dataset with the same split ratio for training, validation, and testing.

The architecture of Model A is characterised in Table 1 and determines the number of calculations performed. The input layer shape is defined by the length of the context and the number of features. The hidden layer shape is defined by the batch size and number of Bi-LSTM units; 256 forward and 256 backward units. A batch size of 32 results in 1,151 observations per batch from a total of 36,825 training observations with any difference subtracted from the final batch. Finally, the output layer shape is defined by

the number of features and batch size. The total number of parameters to be trained in the model is the sum of those in the hidden layer and output layer, totalling 541,702.

A dropout layer is included to minimise the impact of overfitting by randomly setting the weight of 25% of the units in the hidden layer to zero. Dropout is a well-established technique in neural network modelling to overcome overfitting and is considered a more practical approach than regularisation, which is a common approach to reduce overfitting in traditional machine learning problems (Table 2) [40].

The training process was performed using Jupyter Notebook within a Google Colaboratory environment. The complete runtime was 78 s after which predictions could be made within 10 s. The maximum memory usage during training was less than 16 GB. The entire test dataset corresponds to roughly one year of data in 2020 (while training is 2015–2019). The model uses 120 measured hourly data as input and the output is the desired forecast hours. A benefit of having a context length greater than the forecast length is that some measured data will always be used in making the prediction. However, the returns are diminished as the temporal gap between the measured data and forecast increases. A model with a larger context of 240 h capture the data trend but failed to express the peaks and troughs accurately. The approach was first tested by doing

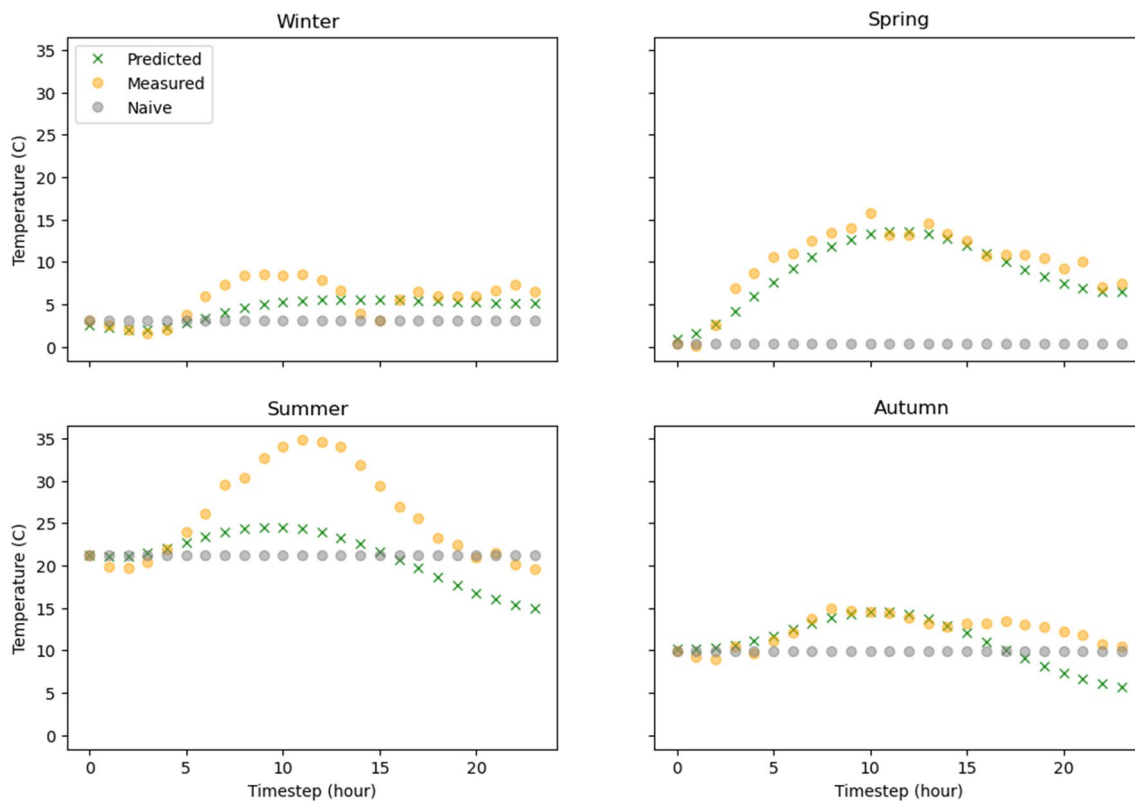


Fig. 4 24-h forecast of the air temperature at Kew Gardens during four days in different seasons

Table 4 Root mean squared error (RMSE), mean average error (MAE) and maximum errors for the 24-h temperature prediction (Fig. 4), values in parentheses are normalised RMSE

	RMSE [°C]	RMSE Naïve [°C]	MAE [°C]	Max. Error [°C]
Summer	1.33 (0.91)	3.30 (2.27)	1.74 (1.20)	4.76 (3.27)
Autumn	1.12 (0.77)	10.4 (7.15)	1.36 (0.93)	2.39 (1.64)
Winter	1.64 (1.13)	7.30 (5.02)	2.03 (1.40)	4.01 (2.76)
Spring	1.73 (1.19)	3.00 (2.06)	2.25 (1.55)	4.24 (2.91)
Mean	1.45	6.00	1.84	3.85
Std. dev	0.244	3.06	0.333	0.886

a single-hour forecast (see Fig. 2). This process is repeated across the entire test dataset and 7,772 single-hour predictions are generated. The root mean, mean absolute and maximum errors were 0.89°C , 0.62°C and 12.81°C respectively.

4 Results

4.1 24-h temperature forecast

To predict 24-h, a comparison was initially made between the *single-step* (predict 24 h in one step) and *multi-step* prediction models to assess the impact of error propagation, see Fig. 3. Table 3 shows that the multi-step model prediction

error according to all three metrics is approximately twice as large as the single-step error.

To quantify how well our 24-h model generalises to different time periods and seasons, four prediction windows spaced 90 days apart are illustrated in Fig. 4. A benchmark model, *naïve* mode, is used for comparison. The naïve model uses the last measured temperature for the entire 24-h forecast. The naïve model does not made assumptions about the future state and is completely uninformed. The root mean squared errors confirm the neural network performs significantly better than the naïve model in all instances (Table 4) with an average error of 1.45°C and 6.00°C for the neural network and naïve forecast respectively.

To contextualise the performance, the neural network was compared to performance metrics from the Met Office. The 24-h predictions produced by the neural network were in 72.9% of all instances accurate to $\pm 2^{\circ}\text{C}$. By comparison, the Met Office states 92.5% of its 24-h temperature predictions are accurate to $\pm 2^{\circ}\text{C}$ while 92% of 24-h wind speed predictions are within 5 knots [41]. Note that measurements used in the weather stations were acquired with a resolution of $\pm 0.1^{\circ}\text{C}$ (Fig. 5).

A better statistical comparisons is done by looking at the probability density functions of the predicted and measured data. The 96 individual forecasts are derived from the four windows in Fig. 4. These points were used to compute a distribution function and are compared to the measured temperature distribution for

Fig. 5 Temperature probability density functions at Kew Gardens. Full temperature dataset of 52,608 samples and two predicted and measured distributions from 96 samples

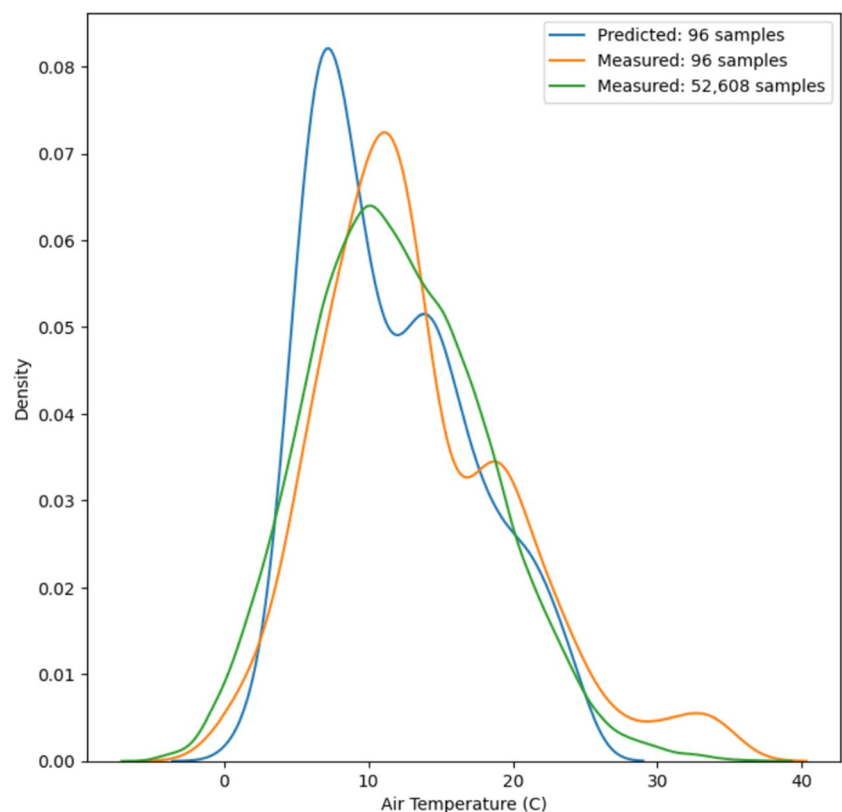
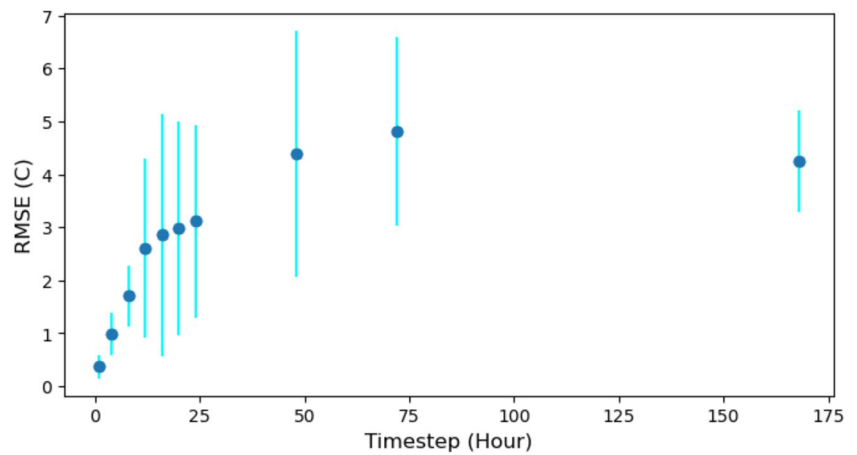


Fig. 6 RMSE of Model A predictions against forecast length. The error bar correspond to min/max RMSE in the windows



the same period, while the entire yearly data was used to create a benchmark. The 96-sample measured temperature peak is wider than the predicted peak indicating that predictions are conservative with both curves demonstrating bimodal behaviour. Nonetheless, the predicted and measured distribution agree very well, except tails on very hot days. Outlier temperatures above 40 °C were measured that are not predicted.

Using the same network (Model A), the length of forecast was varied next to understand the deterioration of the predictions without adapting the model and parameters. Ten different forecast lengths were tested ranging from one to 168 h (seven days). The RMSE mean and standard deviation are plotted against forecast length in Fig. 6 to indicate uncertainty for increasing forecast lengths. For consistency, each prediction was run with a single epoch rather than attempting to optimise performance by identifying the most suitable number of epochs for each forecast length. The single hour prediction has the smallest mean and standard deviation, both of which increase as the forecast length increases, but become more stable after 24 h. 1–24 h predictions have a mean error less than 3 °C. Beyond 24 h, the prediction uncertainty continues to increase before rapidly converging around 4 °C. While there are many caveats to this information, the results suggest that, without further

optimisation, the model should not be used for predictions exceeding one day.

4.2 72-h temperature, relative humidity and wind velocity forecasts

The Model B setup, is shown in Table 5. The main difference with Model A is the addition of a linear layer within the hidden layer and a reduction in the dropout percentage to 10%. The hyperparameters used in the optimised model are recorded in Table 6.

As the first model, an increase in the number of epochs resulted in a reduction in the error and increase of the r-square value. However, there was no direct correlation between optimisation of these two parameters and how the 72-h forecast performed over different time periods. Therefore, once a capable architecture was identified, a similar trial-and-error approach began to optimise the hyperparameters and context length based on the RMSE from the four windows. Initially, 120 h were used for the context length but later changed to 168 h as this gave optimal performance. After upwards of twenty iterations with different conditions, the hyperparameters listed in Table 6 resulted in the best

Table 5 Architecture of Bi-LSTM model, Model B, including the number and type of layers and nodes in each layer

Layer	Type	Value	Shape	Parameters
Input	-	-	(168 × 12)	0
Hidden	Bi-LSTM	Tanh activation function	(4 × 640)	852,480
Hidden	Linear	ReLU activation function	(4 × 256)	164,096
Hidden	Dropout	0.10	(4 × 256)	0
Output	Linear	-	(4 × 12)	3,048
Total				1,019,660

Table 6 The finalised hyperparameters used to train Model B including the number of epochs and optimiser settings

Parameter	Value
Context Length	168 h
Gradient Optimisation	Adaptive Moment Estimation (ADAM)
Learning rate	0.001
Loss: model training	Mean squared error
Metrics: test data evaluation	Root mean squared error
Epochs	1
Batch size	4
Run time	187 s
Train, validate and test split	0.7, 0.15 and 0.15

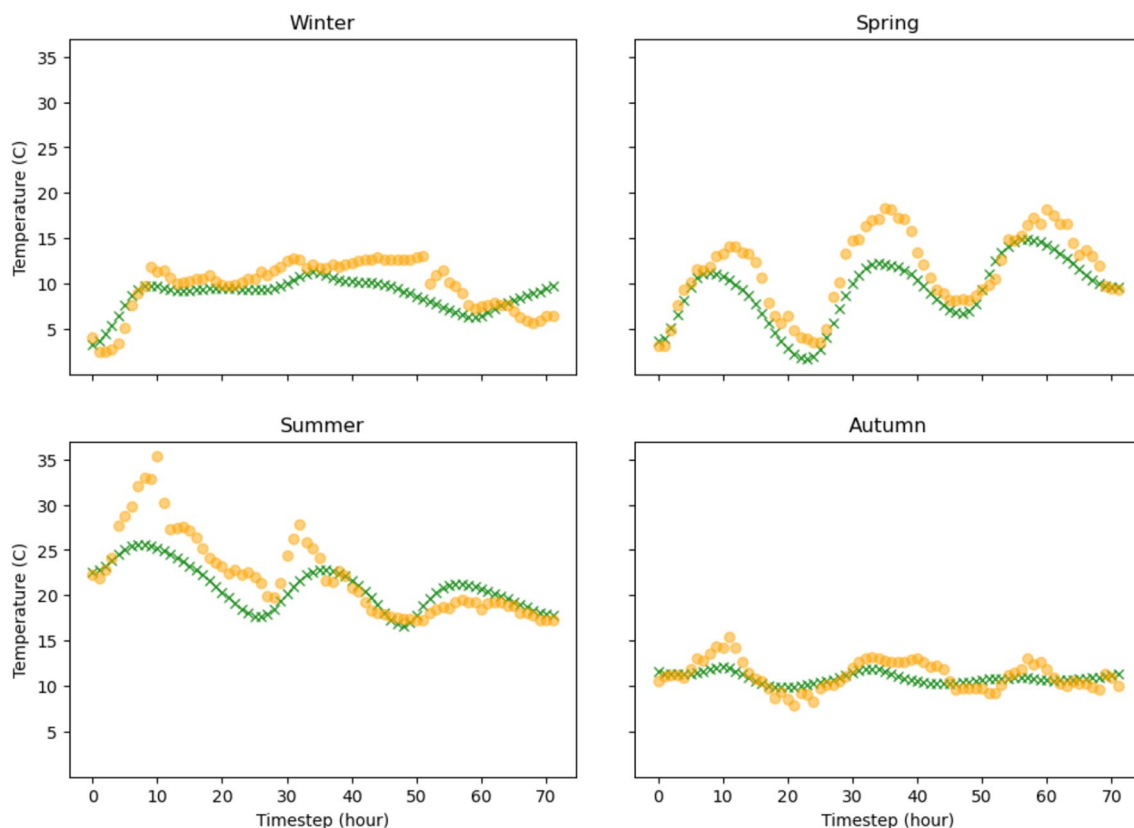


Fig. 7 72-h forecast of the air temperature at Heathrow during four days in different seasons. Symbols same as Fig. 4

performance. * Once the model was trained, it was possible to make new predictions rapidly, within 15 s. The single-step hourly prediction RMSE was 0.94°C, MAE 0.68°C and maximum error 14.94°C when calculated over the entire test dataset. While the numbers are comparable to the single-hour predictions generated in Model A, the model did not perform quite as well over three days as one day. This is to be expected as the forecast window is three times longer and the likelihood of error propagation is much higher.

The four windows in Fig. 7 illustrate how the Bi-LSTM and linear model is highly capable of making predictions with excellent generalisability across different periods and seasons. The three day forecast resulted RMSE mean and

standard deviation 2.26°C and 0.316°C respectively, with 79.5% of the temperature forecasts are within ±3°C when making a 72-h forecast (compared to 1.45°C and 0.244°C in single day prediction) (Table 7).

Figure 8 shows predicted distribution for 72-h forecasts. Despite the qualitatively good agreement, the modelled distribution has a narrower peak with extreme high temperatures underestimated (similarly to model A), showcasing the difficulty to represent the tails of the distribution.

The model takes in all features from both locations resulting in six unique features and 12 features in total. As before, it is possible to generate a prediction for any one of the features introduced to the model in training. While the model

Table 7 Root mean squared error (RMSE), mean average error (MAE) and maximum errors for the 72-h temperature prediction (Fig. 7), values in parentheses are normalised RMSE

	RMSE [°C]		MAE [°C]		Max. Error [°C]	
	Kew G	Heathrow	Kew G	Heathrow	Kew G	Heathrow
Winter	2.22 (0.78)	1.80 (0.63)	1.79 (0.63)	1.44 (0.51)	6.25 (2.2)	4.11 (1.45)
Autumn	3.02 (1.06)	2.27 (0.80)	2.51 (0.88)	1.89 (0.67)	7.64 (2.70)	5.24 (1.85)
Summer	3.41 (1.20)	2.69 (0.95)	2.80 (0.99)	2.03 (0.72)	7.10 (2.50)	6.66 (2.35)
Spring	2.70 (0.95)	2.31 (0.82)	2.12 (0.75)	1.87 (0.66)	5.25 (1.85)	5.31 (1.87)
Mean	2.83	2.26	2.31	1.81	6.56	5.33
Std. dev	0.436	0.316	0.383	0.221	0.910	0.904

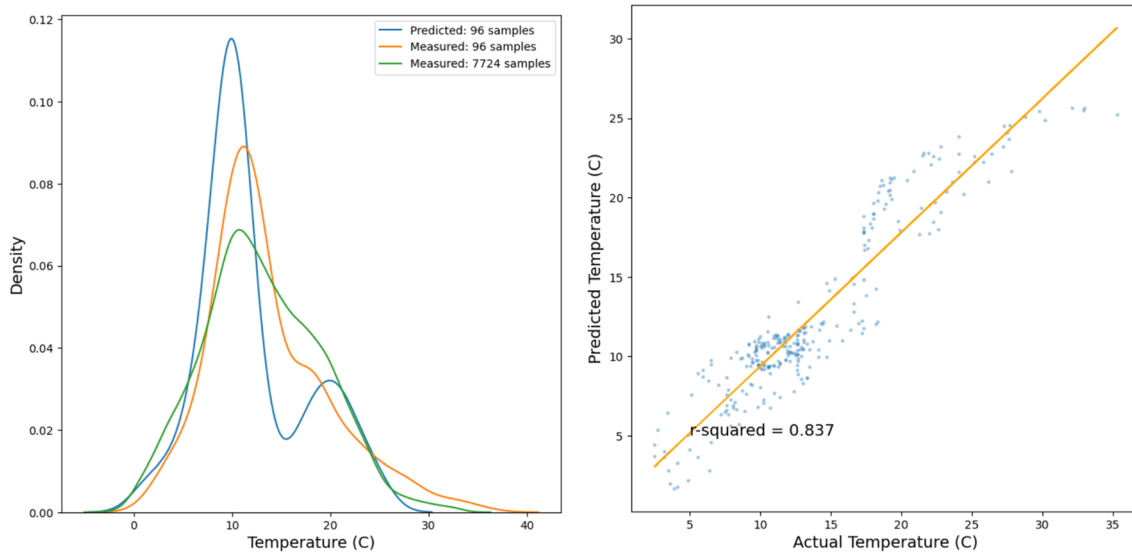


Fig. 8 Temperature probability density functions (left) and scatter plot (right) at Heathrow

does take all inputs into consideration during training and seeks to minimise the loss function with respect to all features, the performance arising from this approach does not necessarily translate into good generalisability across all timescales. When training the model, the weighted sum of

all 12 features is used when minimising the loss, assigning different levels of importance to each feature. During the training of Model B, the objective was to optimise the 72-h temperature predictions, there was no guarantee that this performance would translate into comparable performance

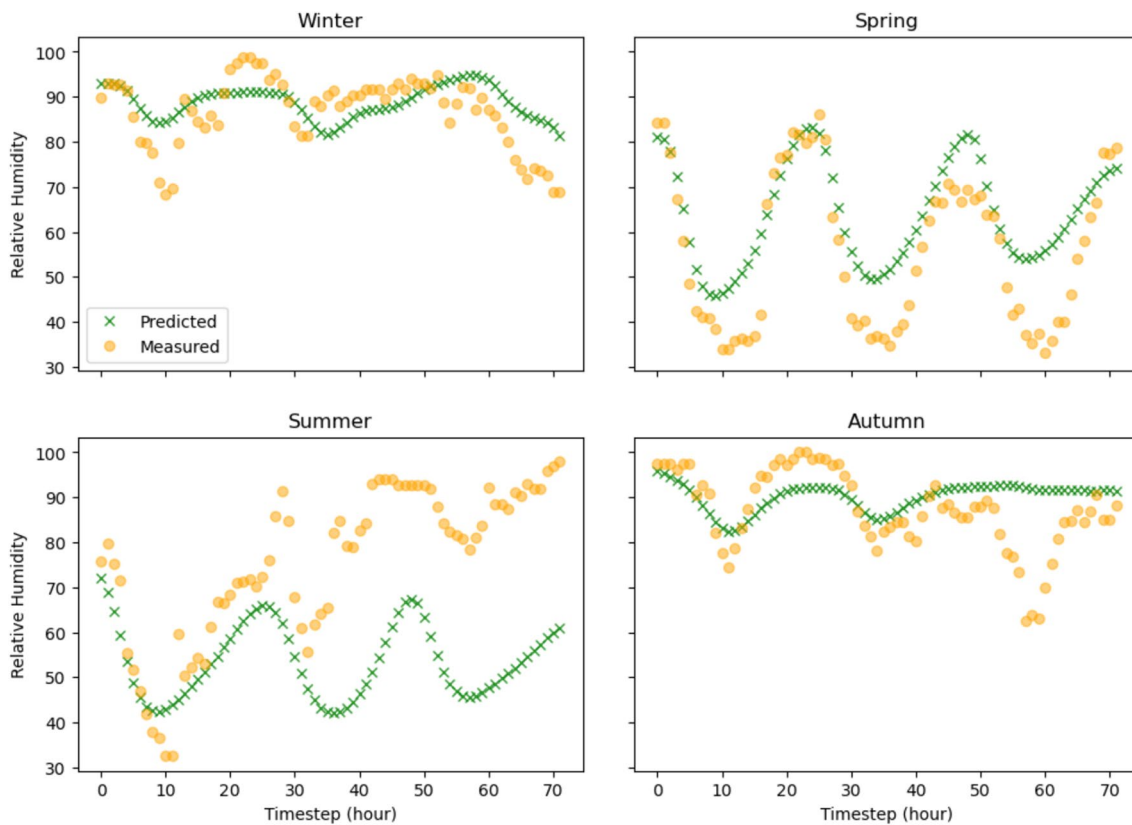


Fig. 9 72-h forecast of the relative humidity at Heathrow during four days in different seasons

Table 8 Root mean squared error (RMSE), mean average error (MAE) and maximum errors for the 72-h relative humidity prediction (Fig. 7), values in parentheses are normalised RMSE

	RMSE [%]		MAE [%]		Max. Error [%]	
	Kew G	Heathrow	Kew G	Heathrow	Kew G	Heathrow
Winter	8.78	7.46	6.33	5.4	22.9	20.6
Autumn	9.48	8.25	7.30	6.21	21.9	19.6
Summer	28.0	29.1	22.2	23.43	58.6	61.6
Spring	11.9	11.5	8.43	8.51	36.2	33.4
Total	58.1	56.3	44.3	43.6	139.6	135.2
Average	14.5	14.0	11.1	10.9	34.9	33.8

for another feature, in this case relative humidity. The accuracy of the results in Fig. 9 are a byproduct of the process to optimise the air temperature. If the relative humidity were the focus of the optimisation, the forecast prediction would probably show considerable improvement (Table 8).

5 Conclusions and future work

This paper presented a novel, flexible, deep learning local weather forecasting. The approach is capable of rapidly predicting weather features and generating cheap, reliable short duration forecasts. The model is purely data-driven, in contrast with earlier approaches that required varying degrees of data assimilation or hybrid model. A total of two models were trained and used to predict air temperature and relative humidity. The dataset used to train the models contained six years of historical weather observations from Kew Garden and Heathrow weather observation stations in London. The objective of having multiple locations is to infer a topographical representation for the model to learn from. As the two weather observation stations are positioned 11 km apart, it is expected that they would share similar weather characteristics. Discrepancies in wind speed and humidity between the location could be explained by local land features and artificial structures. Kew Gardens is positioned near the river Thames in a built-up area while the nearest body of water to Heathrow is several kilometres away. Heathrow observation station is situated within the airport boundaries with few obstructions.

Model A is a 24-h prediction network designed to predict air temperature. This model was intended to demonstrate proof of concept and was trained with wet bulb, air and dew point temperatures. The Model A achieved its objective of establishing a baseline for further predictions. It showed that air temperature could be predicted with reasonable accuracy compared to the Met Office, predicting the air temperature within a range of 2°C in 72.9% of instances with a maximum error of 3.85°C occurring mostly in very hot days. Model B is a 72-h prediction network that attempted to predict air temperature, relative

humidity and wind speed. Despite a three-fold increase in the forecast length, the model was able to accurately predict air temperature with an RMSE of 2.26°C at Heathrow and was able to predict the temperature accurately to within $\pm 3^\circ\text{C}$ in 79.5% of instance. It was able to predict the relative humidity in the same location with an RMSE of 14%. However, Model B was optimised with respect to air temperature which impacted the accuracy.

The flexibility and speed of the model makes it attractive to short-term local forecast in locations where weather stations are present but it maybe difficult to have accurate weather predictions (due to topography, local effects, etc.) The result show that predictions up to three days have accuracy comparable to expensive numerical weather predictions. However, featured-based optimisation may be required to improve the accuracy of features such as wind speed or humidity. Future lines of research will be in this direction.

Authors contribution G.Z and S.NM contributed to the conception of the presented idea. G.Z. wrote the main text, performed the simulations and prepared all figures. S.NM revised the manuscript. All authors reviewed the manuscript.

Data availability The weather hourly data that support the findings of this study are publicly available in the NCAS British Atmospheric Data Centre, <https://catalogue.ceda.ac.uk/>

Declarations

Ethical and informed consent for data used Not Applicable.

Competing Interests The authors have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Lynch P (2008) The origins of computer weather prediction and climate modeling. *J Comput Phys* 227(7):3431–3444
- Scher S, Messori G (2018) Predicting weather forecast uncertainty with machine learning. *Q J R Meteorol Soc* 144(717):2830–2841
- Rasp S, Dueben PD, Scher S, Weyn JA, Mouatadid S, Thuerey N (2020) WeatherBench: a benchmark data set for data-driven weather forecasting. *J Adv Model Earth Syst* 12:e2020MS002203. <https://doi.org/10.1029/2020MS002203>
- Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. *Nature* 525(7567):47–55
- Rihan FA, Collier CG, Roulstone I (2005) Four-dimensional variational data assimilation for doppler radar wind data. *J Comput Appl Math* 176(1):15–34. <https://doi.org/10.1016/j.cam.2004.07.003>
- Brunton SL, Noack BR, Koumoutsakos P (2020) Machine learning for fluid mechanics. *Annu Rev Fluid Mech* 52(1):477–508
- Vinuesa R, Brunton SL (2022) Enhancing computational fluid dynamics with machine learning. *Nat Comput Sci* 2(6):358–366
- Sarghini F, de Felice G, Santini S (2003) Neural networks based subgrid scale modeling in large eddy simulations. *Comput Fluids* 32(1):97–108
- Prat A, Sautory T, Navarro-Martinez S (2020) A priori sub-grid modelling using artificial neural networks. *Int J Comput Fluid Dyna* 34:6:397–417. <https://doi.org/10.1080/10618562.2020.1789116>
- Milano M, Koumoutsakos P (2002) Neural Network Modeling for Near Wall Turbulent Flow. *J Comput Phys* 182(1):1–26
- Schizas C, Michaelides S, Pattichis C, Livesay R (1991) in 1991 Second International Conference on Artificial Neural Networks, pp. 112–114
- Ochiai K, Suzuki H, Shinozawa K, Fujii M, Sonehara N (1995) in Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 2, pp. 1182–1187
- Hewage P, Trovati M, Pereira E, Behera A (2021) Deep learning-based effective fine-grained weather forecasting model. *Pattern Anal Appl* 24(1):343–366
- Ma X, Fang C, Ji J (2020) Prediction of outdoor air temperature and humidity using xgboost. *IOP Conf Ser Earth Environ Sci* 427(1):012–013
- Slingo J, Palmer T (2011) Uncertainty in weather and climate prediction. *Phil Trans R Soc A Math Phys Eng Sci* 369(1956):4751–4767
- Frnda J, Durica M, Rozhon J et al (2022) ECMWF short-term prediction accuracy improvement by deep learning. *Sci Rep* 12:7898. <https://doi.org/10.1038/s41598-022-11936-9>
- Shao B, Song D, Bian G, Zhao Y (2021) Wind speed forecast based on the lstm neural network optimized by the firework algorithm. *Adv Mater Sci Eng* 2021:1–13
- Dong X, Deng S, Wang D (2022) A short-term power load forecasting method based on k-means and SVM. *J Ambient Intell Humaniz Comput* 13(11):5253–5267. <https://doi.org/10.1007/s12652-021-03444-x>
- Fasil OK, Rajesh R (2022) Epileptic seizure classification using shifting sample difference of EEG signals. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-022-03737-9>
- Gupta V (2023) Wavelet transform and vector machines as emerging tools for computational medicine. *J Ambient Intell Humaniz Comput* 14(4):4595–4605. <https://doi.org/10.1007/s12652-023-04582-0>
- Wang P, Fan E, Wang P (2021) Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recogn Lett* 141:61–67
- Yang M, Wang J (2022) Adaptability of Financial Time Series Prediction Based on BiLSTM. *Procedia Comput Sci* 199:18–25
- Murali P, Revathy R, Balamurali S, Tayade AS (2020) Integration of RNN with GARCH refined by whale optimization algorithm for yield forecasting: a hybrid machine learning approach. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-020-01922-2>
- Nayak JR, Shaw B, Sahu BK (2022) A fuzzy adaptive symbiotic organism search based hybrid wavelet transform-extreme learning machine model for load forecasting of power system: a case study. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-022-04355-1>
- Weyn JA, Durran DR, Caruana R (2019) Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data. *Adv Model Earth Syst* 11(8):2680–2693
- Roy DS (2020) Forecasting the air temperature at a weather station using deep neural networks. *Procedia Comput Sci* 178:38–46
- Ravuri S, Lenc K, Willson M, Kangin D, Lam R, Mirowski P, Fitzsimons M, Athanassiadou M, Kashem S, Madge S, Prudden R, Mandhane A, Clark A, Brock A, Simonyan K, Hadsell R, Robinson N, Clancy E, Arribas A, Mohamed S (2021) Skilful precipitation nowcasting using deep generative models of radar. *Nature* 597(7878):672–677
- Casper, Espeholt L, Heek J, Dehghani M, Oliver A, Salimans T, Agrawal S, Hickey J, Kalchbrenner N (2020) Metnet: A neural weather model for precipitation forecasting. arXiv pre-print server. <https://arxiv.org/abs/2003.12140>. Accessed Mar 2023
- Espeholt L, Agrawal S, Sønderby C, Kumar M, Heek J, Bromberg C, Gazen C, Carver R, Andrychowicz M, Hickey J, Bell A, Kalchbrenner N (2022) Deep learning for twelve hour precipitation forecasts. *Nat Commun* 13(1):5145. <https://doi.org/10.1038/s41467-022-32483-x>
- Fu Q, Niu D, Zang Z, Huang J, Diao L (2019) in 2019 Chinese Control Conference (CCC), pp. 3771–3775
- Son Y, Zhang X, Yoon Y, Cho J, Choi S (2022) LSTM-GAN based cloud movement prediction in satellite images for PV forecast. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-022-04333-7>
- Danandeh Mehr A, Rikhtehgar Ghiasi A, Yaseen ZM et al (2023) A novel intelligent deep learning predictive model for meteorological drought forecasting. *J Ambient Intell Human Comput* 14:10441–10455. <https://doi.org/10.1007/s12652-022-03701-7>
- Sengar S, Liu X (2020) Ensemble approach for short term load forecasting in wind energy system using hybrid algorithm. *J Ambient Intell Humaniz Comput* 11(11):5297–5314. <https://doi.org/10.1007/s12652-020-01866-7>
- Singh U, Rizwan M (2022) SCADA system dataset exploration and machine learning based forecast for wind turbines. *Results Eng* 16:100,640
- Mujeeb S, Alghamdi TA, Ullah S, Fatima A, Javaid N, Saba T (2019) Exploiting deep learning for wind power forecasting based on big data analytics. *Appl Sci* 9:4417. <https://doi.org/10.3390/app9204417>
- Torres JM, Aguilar RM, Zuñiga-Meneses KV (2018) Deep learning to predict the generation of a wind farm. *J Renew Sustain Energy* 10(1):013305
- Gers JS, Felix A, Cummins F (2000) Learning to forget: Continual prediction with lstm. *Neural Comput* 12(10):2451–2471
- Met-Office (2006) UK daily temperature data, part of the met office integrated data archive system (midas).ncas british atmospheric data centre, date of citation. https://catalogue.ceda.ac.uk/uuid/1bb479d3b1e38c339adb9c82c15579d8?_ga=2.185710890.2071699676.1677353726-899430677.1677353726. Accessed Mar 2023

39. Wang JQ, Du Y, Wang J (2020) Lstm based long-term energy consumption prediction with periodicity. *Energy* 197(117):197
40. Kreuzer D, Munz M, Schlüter S (2020) Short-term temperature forecasts using a convolutional neural network - an application to different weather stations in germany. *Mach Learn Appl* 2:100,007
41. Met-Office (2022) How accurate are our public forecasts? <https://www.metoffice.gov.uk/about-us/what/accuracy-and-trust/how-accurate-are-our-public-forecasts> . Accessed 30/08/2022

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Gabriel Zenkner received his MSc at Imperial College London in Advanced Mechanical Engineering in 2022. His current research interests include data science, data engineering and prediction of physical and social phenomena using machine learning techniques.

Salvador Navarro-Martinez is a reader at Imperial College London in the Department of Mechanical Engineering. His research interests lie in the modelling of turbulent flows using stochastic and machine learning techniques.