Annotated Bibliography: Machine Learning and Deep Learning Approaches for Weather Forecasting in Aviation

Hyacinthe Chemasle

1. Narvekar, M. & Fargose, P. Daily Weather Forecasting using Artificial Neural Network. *Int. J. Comput. Appl.* 121, 9–13 (2015).

This paper aims to find a finer approach in forecasting by reviewing different machine learning approaches. Some of the ML approaches mentioned such as: Artificial Neural Network, Ensemble Neural Network, Radial Basis Function Network and Genetic algorithm. The authors experiment the use of an Artificial Neural network with a Back-propagation approach to assess forecasting ability.

The built model aims to predict Rainfall and Sky conditions. The paper makes no mention of the specific area that local data is taken from or if an agency has helped them provide the weather dataset. This is important as it helps determine if these are gridded or point based approaches at specific locations. 28 Features are used to help predict rainfall and sky conditions. Some of the features include "Temperature", "Humidity", "air pressure", "wind speed", "wind direction", "cloud coverage", and "Rainfall".

No numerical insight into the hyperparameters are provided for the NN BP approach. Authors do mention that the number of neurons in the hidden layers are to be 70%-90% of the size in the input layer. Sigmoid activation function is applied through the hidden layers and the output layers consists of a linear transfer function.

Findings of the paper noted that a back-propagated artificial neural network worked the best for predicting weather with minimal error. Performance metrics such as MSE are used to identify the networks overall discriminative ability.

The authors make no explicit mention of limitations to the review. However, the paper itself doesn't provide performance metric insight into the experiments and numerical assessment of the model accuracy and MSE. The architecture of the Neural Network and Backpropagation approach is also not provided in detail around the batch size, number of epochs and hidden layers but rather gives a guideline on what the standard practice tends to be around hyperparameters.

2. Kakar, S. A. et al. Artificial Neural Network based Weather Prediction using Back Propagation Technique. *Int. J. Adv. Comput. Sci. Appl.* 9, (2018).

This paper focuses on the specific use of an ANN by leveraging the Weather forecast standard dataset. Whereby the goal is to see if predictive accuracy is greater than a physics based NWP model.

The author utilises an Artificial Neural network with eleven features such as “MaxDew-Point”, “MaxHumidity”, “MaxPressure”, “MaxTemperature”, “MaxVisibility”, “Mean-WindSpeed”, to classify weather into four categories being “Sunny”, “Foggy”, “Thunderstorm” and “Rainy”. Two aspects are considered with the model being weather classification and weather prediction. The model classifies the weather into four types as previously mentioned and then predicts eleven weather features by utilising twenty training samples from 1997 and 2015.

Findings of the paper noted that by increasing the number of hidden layers in the ANN architecture improved predictive ability with less errors. The study didn’t provide information around activation functions used or the architecture of the neural network itself. Although the findings mention that there is less error the study itself doesn’t aid us with performance metrics like MSE or MAE and numerical insight into those accuracies. Limitations of the experiment were mentioned in the paper with only two objectives being considered for a single use case study focusing on classification and year based prediction. Future studies would look at predictions on a daily and monthly bases.

3. Araujo, G. & Andrade, F. A. A. Post-Processing Air Temperature Weather Forecast Using Artificial Neural Networks with Measurements from Meteorological Stations. *Appl. Sci.* 12, 7131 (2022).

Paper focuses on the improvement of post processing for air temperature compared to regular weather forecasts. Experiment looks at two deep learning architectures being the CNN (Convolutional Neural Network) and MLP (Multi-Layer Perceptron).

The aim of this experiment is to predict the measured air temperature which has notable relevance to airspace for aircraft movements. Data acquisition was done by utilising the MET being Norway’s national weather database. Features selected from the MET were “latitude”, “longitude”, “date”, “time” and “elevation”. It also uses the outputs of the MEPS (MetCoOp Ensemble Prediction System) with crucial data such as temperature, wind speed, gust presence etc.

The proposed Multi-Layer Perceptron model has two fully connected (hidden) layers with 64 and 16 neurons. The hyperbolic tangent (tanh) function is selected for the activation function in the hidden layers. No activation function is used for the output layer when predicting air temperature. The convolutional neural network has one 1D convolutional layer with up to 16 kernels of size 3. This is followed by a max pooling layer and two fully connected layers with 64 and 16 neurons.

Findings regarding the improvement of post processing of air temperature found the NN solution to be considered more favourable. The MLP achieved an RMSE reduction of approximately 0.5K compared to the standard MET tool used to forecast weather. The MEPS itself had the highest RMSE of 5.258 with 54.05% of error whereas the MLP had a much lower RMSE of 1.786 and 8.2% of error. The CNN achieved a RMSE of 1.79 with a 8.23% error rate. This shows that deep learning methods like the CNN or MLP perform better than standard numerical weather prediction models like the MEPS.

The authors mention that there should be more work to be done in the field such as hosting data from IoT sensors that work as meteorological observers in the cloud and possibly evaluating Graph Neural Networks for future experiments.

4. Hennon, C. C., Coleman, A. & Hill, A. Short-Term Weather Forecast Skill of Artificial Neural Networks. *Weather Forecast.* 37, 1941–1951 (2022).

This paper looks at the short term predictive nature of weather forecasting using three different neural network architectures: FFB (Feed Forward Back-Propagation), GRN (Generalised Regression) and RBF (Radial Basis Function).

NNs are tested independently with real-time forecasts across 11 US cities. The authors leveraged data from a variety of US weather providers like the North American Mesoscale (NAM). No specific variables were mentioned from the real time forecasts but it is assuming that these variables from the forecasts are used to predict four categories of Maximum/Minimum temperature, wind speed and precipitation.

The Architecture of the FFB Neural network utilised the Levenberg-Marquardt (LM) method with a fixed learning rate. Each variable had 1500 training epochs. The number of hidden neurons was set between 3 and 7 for temperature, wind speed and precipitation. The GRN and RBF architectures had fixed number of hidden neurons with just one training epoch and had undergone spread tuning values to control the influence the hidden neurons had when determining output.

Findings of research outline that the NNs have a lower error rate of 25% than common guides for weather forecasting and more than a 50% reduction for forecasting wind speeds. Wind speeds specifically are important for ground movements and traffic direction in airspaces for arriving and departing aircraft. The RBF NN performed the best out of the FFB and GRN when classifying maximum wind speed, while the FFB has the lowest MAE for prediction of quantitative precipitation, while the GRN performed better than current objective aids for weather such as the HRRR and NAM in the US. Overall, for max temperature the NNs themselves achieved a 28.75 RMSE improvement compared to current NWP methods.

Experiment outlines that there were significant improvements in the short term nature of 1 day forecasts for maximum and minimum temperature. However, the research does mention that a larger amount of precipitation predictors could be added to reduce forecast variance and larger deviations when utilised for NN training.

5. Perez-Ortiz, M., Gutierrez, P. A., Tino, P., Casanova-Mateo, C. & Salcedo-Sanz, S. A mixture of experts model for predicting persistent weather patterns. arXiv (2019).

Topic chosen looks at the underlying problem of predicting low-visibility events at airports. The aim is to devise a new model to predict low visibility in airports.

The authors utilise the concept of MoE (Mixture of Experts). The model combines experts predicting previous categories by use of an autoregressive neural network where they treat the experiment as an ordinal classification problem by process of discretisation. This is done by discretising the time series into different categories.

METAR reports are used as data to predict cloud height and runway visual range from a time period of November through to February at Valladolid airport. The cloud height and runway visual range are selected for output as they are strong predictors for low visibility events. 7 Atmospheric variables are considered for training to contribute to prediction. METAR Reports contain important features such as “time, wind, runway visual range, precipitation, fog, cloud coverage.”

Overall findings suggest that ML approaches can be complemented to standard persistence models used for visibility forecasts. Concepts such as MoE in the experiment show this superior performance. When predicting Runway Visual range the STMEIC (Simultaneously Trained mixture of experts with imbalanced costs) had the lowest AMAE (average mean absolute error) of 1.67 and MMAE (Maximum Mean absolute error) of 1.89 compared to current NWP models achieving 2.89 for both AMAE and MMAE. Interestingly enough, current weather prediction methods outperformed all Mixture of Expert model categories when predicting cloud height.

Authors make a specific call to further research in testing NN models with a recurrent structure to leverage the dynamics of time series phenomena. This would be relevant for our topic since there are known weather phenomena to occur at Wellington Airport especially with wind events.

6. Akbayır, I., Yavuz, V., Demirhan, D., and İnanç, B. M. Meteorological Analysis and Prediction of Gusts at Istanbul Airport Using Machine Learning Algorithms, *EGU General Assembly 2025*, Vienna, Austria, 27 Apr–2 May 2025, EGU25-12021, 2025.

Conference paper presented around predicting the onset of wind gusts by utilising machine learning methods. Three different methods are tested being the LSTM (Long Short term Memory), XGB (Extreme Gradient Boosting), and RF (Random Forests).

Utilised METAR reports as data from Istanbul Airport between the years of 2018 and 2024. Relationships between meteorological variables such as temperature, pressure and dew point were analysed as features. Models were subsequently trained and evaluated on 1000 randomly selected subsets.

Findings of the report suggested that the three different models had high predictive success at times of high wind gust values greater than 30 knots, while prediction success was seen to be less on smaller gust values.

Limitations to the report lack the specific methodology used and performance metrics across the different machine learning methods. Specifically, the comparison between deep learning architectures like the LSTM could be compared against other methods in terms of predictive success around wind gusts.

7. Nagimov et al. Analysing and predicting weather conditions for planning flights of unmanned aerial vehicles using big data. arXiv (2025).

Researchers propose the effective use of deep learning algorithms like LSTMs and RNNs in building a predictive weather model to aid UAV flight planning. Various meteorological sources are used across the study like IoT sensors, global climate repositories, and real time weather databases. Experiment has 6 features used to guide the Neural networks being "Temperature", "Wind speed", "Wind direction", "Humidity", "Atmospheric pressure" and "Precipitation levels". The networks are tasked to predict Temperature and Wind speed variations.

Findings reported that the LSTM architecture achieved the best performance of an accuracy of 91% and a RMSE of 2.31 for prediction of temperature whereas other architectures like the CNN achieved 89% accuracy with an RMSE of 2.74. A key limitation outlined in the research is one applicable to the current topic where deep learning models may limit real-time applications such as aircraft movements requiring additional optimisation.

8. Patriarca, R., Simone, F. & Di Gravio, G. Supporting weather forecasting performance management at aerodromes through anomaly detection and hierarchical clustering. *Expert Systems with Applications* 213, Part C (2023).

Researchers look at a different way to utilise machine learning to detect forecasting anomalies in historic data by leveraging an anomaly detection algorithm and hierarchical clustering.

METARs and TAFs were the primary source of data used to train the unsupervised models consisting of 500,000 METARs and 50,000 TAFs of historic data from 40 aerodromes worldwide. These consist of variables such as "Wind", "Visibility", "Runway Visual Range", "Rain", "Fog", "Cloud coverage", "Temperature", "Pressure".

Findings suggested that there were significant differences between TAF predictions and observations made by the METAR by use of an error propensity metric. When utilising testing of their anomaly detector they found 19 anomalous time periods out of the 365 days identified for the two airports.

Limitations outlined by the authors mention that there is limited data for their application and only utilise a single location focus. Although the authors didn't use deep learning methods in their study the research carried is still of value for topic. The study successfully created the groundwork by demonstrating ML approaches to predict forecast errors. This creates a strong foundation for necessary future work to be carried out on deep learning methods which the authors mentioned for better performance.

9. Choi, S. & Kim, Y. J. Artificial neural network models for airport capacity prediction. *Journal of Air Transport Management* 97 (2021).

Researchers in this article propose different Deep Learning architectures such as MLP, RNN and LSTM to predict arrival/departure capacity of airports. The relevance to the topic being notable as the paper utilises past airport weather data to help aid prediction.

These models are trained with historic data on hourly capacity and weather observations from Atlanta International Airport (ATL) during the periods of 2013 to 2017. NOAA's ISD (Integrated Surface Database) was utilised to include variables such as "Wind", "Speed" "Wind Direction", "Temperature", "Cloud Height", "Sea Level Pressure", "Visibility", "Rainfall" and "Snow Depth". These are variables that can be considered strong predictors for arrival/departure capacity.

The models are trained with the defined variables. Pre-processing steps are used to convert categorical variables to one hot vectors. The MLP is trained by stochastic gradient descent whereas the RNN and LSTM were trained by Adam algorithm. Each NN architecture had hyper parameters of 300 epochs with a batch size of 200.

Findings suggested that the LSTM model was the best performing with the lowest RMSEs across predicted capacity for departures and arrivals. The MLP, RNN and LSTM achieved Test RMSEs between 1.7 and 3 for departure and 1.84–4.27 for arrivals.

Limitations to the experiment include deterministic weather being used and as a result that there needs to be research for integrating the uncertainty of weather forecasts to predict future capacity.

10. Schultz, M., Reitmann, S. & Alam, S. Predictive classification and understanding of weather impact on airport performance through machine learning. *Transportation Research Part C: Emerging technologies* 131 (2021).

Paper looks at quantifying the impacts that severe weather events have locally on the efficiency of airport operations by use of a machine learning approach. Local Weather data and airport performance data are used in conjunction to train recurrent and convolutional neural networks.

METARs are used in conjunction with Terminal Aerodrome forecasts (TAFs) in the form of tabular data from London Gatwick airport. Features from the METAR tables are used to help aid the Neural Networks in predicting the level of airport efficiency. Features such as "date of month", "time", "wind direction", "wind speed", "visibility", "precipitation", "clouding", "air temperature" and "pressure" are all used to contribute to overall prediction. Permutation importance is also applied as a heuristic approach to aid with computational efficiency and extract feature importance.

Three models are tested in this experiment being an LSTM, CNN, and a comprising of CNN-LSTM Model. The hyperparameters of the LSTM consist of 200 hidden layers, a batch size of 40, 50 epochs and Adamax optimiser with a SoftMax activation function for output. The CNN has 64 filters, 3 kernels and 10 hidden layers. Results from the experiments showed that the LSTM particularly struggled to correctly classify arrivals only achieving 48.9% accuracy. CNN also struggled with only 55% accuracy. Departures were the highest performing in terms of accuracy by the CNN with 96.9% and the LSTM with 82.1%.

Findings suggest that the paradigms of the CNN and LSTM show promising results and could be further researched to provide weather-related decisions for future airport operations.

Authors make a specific mention for future research in response to METAR components being quantified due to the complex nature of weather variables interacting with one another. This would be an interesting factor to consider for our METAR reports pertaining to Wellington Airport.

11. Alves, D., Mendonça, F., Mostafa, S. S. & Morgado-Dias, F. Low Tropospheric Wind Forecasts in Aviation: The Potential of Deep Learning for Terminal Aerodrome Forecast Bulletins. *Pure Appl. Geophys.* 181, 2265–2276 (2024).

Paper looks at the effective use of Deep Learning approaches to make accurate wind predictions regarding speed and direction in low tropospheric wind forecasts. Five deep learning models are evaluated being the LSTM, vRNN (Vanilla Recurrent Neural Network), 1dCNN, CNN–LSTM, and Gated Recurrent unit (GRU).

METAR reports are used as the primary dataset from GCLP airport covering a timeframe from 2018 to 2022. The authors carry out data preprocessing steps such as segmenting the data chronically and extracting features of importance such as "Day", "Hour" "Wind Direction" and "Wind Speed". These features are strong indicators toward helping set up a model that performs well when forecasting wind.

The general architecture chosen for all of the five models consisted of 3 layers where there was an input layer, one hidden layer and one output layer. Adam optimiser was used in conjunction with the MSE as the primary loss function. Each of the five models also had the same hyperbolic tangent activation apart from the GRU that utilises a sigmoid activation for recurrent step.

Findings reported that the LSTM was the model out of the five DL approaches to achieve the highest precision. This was notable for extended forecasting periods having achieved an MAE (Mean Absolute Error) of 1.23 m/s for wind speed and a cMAE (circular Mean Absolute error) of 15.80 degrees. The 1 Dimensional CNN approach achieved the best forecasting for wind direction over short intervals whereas the Vanilla Recurrent Neural Network fell behind in performance. These were findings that were able to align with the acceptable threshold for TAF predictions which is the current use of aerodrome forecasting.

The authors make a call for the testing of these models on diverse geographical contexts especially airports that have severe wind phenomena which makes it specifically relevant to an airport like Wellington that records many severe wind weather events during the year.

12. Huang, L., Zhang, Y., Yin, Y., Zhang, S. & Zhang, Y. Efficient Real-Time Aircraft ETA Prediction via Feature Tokenization Transformer (2025).

This paper considers the Feature Tokenization-based Transformer model as a Deep Learning approach to efficiently predict estimated arrival time of aircraft. The model inputs a range of data such as aircraft latitude, longitude, ground speed and most importantly the weather context. The experiment is carried at WSSS (Singapore Changi International Airport).

The transformer architecture proposed mentions the use of a self-attention mechanism. This is done by employing feature tokenization to create ETA predictions from aircraft that are automatically projected into latent spaces and fed back through into the transformer network. The main appeal is the parallelism that reduces training time and has greater computational efficiency compared to other models like the LSTM. The output regarding the final ETA prediction is done by use of a SoftMax function.

Numerous data sources are used to help aid in prediction such as Aircraft Track Data, Meteorological Data and Flight plan data. The Meteorological data is of importance since this is relevant to our topic of using METAR reports to aid in prediction. The study uses specific features of wind direction, wind speed, visibility, sky coverage and altitude from the METAR to be used in the transformer model.

Findings compared past methods of predicting ETA for aircraft such as the use of XG-Boost where the FFT model achieved an MAE (Mean Absolute Error) of 104.6 seconds compared to an MAE of 112.2 seconds from XGBoost. This shows a considerable 7.1% improvement in predictive accuracy. The training time from the FFT was also drastically reduced by 39% compared to XGBoost. The arrival directions in these predictions were also incredibly accurate with the R squared value metric being at 0.99 for east, north and south direction and 0.94 for west directions.

No specific limitations were outlined but the relevance to the overall topic could be of consideration since the estimated arrival time could have some correlation with the weather events observed at an airport. This is why METAR reports were used as it was seen to be a significant contributor to an aircraft's ETA and could also be applied for the observation of weather events at Wellington Airport.