Weather Forecasting using Multilayer Perceptron Technique

Shifanaaz Abdulsab Nadaf Afrasama A. Harlapur Aleena A. Sayed Fathma Shekh, and Shashank Hegde

[1]School of Computer Science and Engineering, KLE Technological University, Hubballi 580031, India,
info@kletech.ac.in

Abstract. *The Multilayer Perceptron (MLP)* in Weather forecasting is used for regression tasks based on input features such as *pressure* values, *temperature*, *wind*. This proposed model focuses on evaluating the efficiency of MLP for accurate time series pattern predictions. This study incorporates ERA5 hourly data on pressure levels from 1940 to the present and uses a *Feedforward Neural Network(FNN)* MLP architecture. In addition, techniques such as the *Cosine annealing learning rate scheduler* and *Hyperparameter tuning* are employed to analyze temporal relationships, perform feature selection and ultimately improve model performance. Experiments conducted with MLPs demonstrate competitive accuracy with MSE 91.92, MAE 7.02 and R^2 0.9985 compared to traditional forecasting , highlighting MLPs as a valid method for meteorological applications. . . .

Keywords: Cosine annealing scheduler, Hyperparameter tuning, Multi-Layer Perceptron (MLP), Weather forecasting, Neural network depth.

1 Introduction

Weather forecasting plays a crucial role in understanding the dynamics of the atmosphere and is critical for numerous industries, including agriculture, transportation, energy, and disaster management. [4] The process of predicting the weather condition for the future is known as Weather forecasting. [1] It reduces risks, maximizes agricultural yields, allows safe transportation, and enables proper energy management. The use of real-time temperature, humidity, and pressure data using various sensors [1]. The task of weather forecasting is inherently challenging because atmospheric systems are complex and unpredictable and *Artificial Neural Networks (ANNs)* have some interesting properties that made this family of machine learning algorithms very appealing when confronted with difficult pattern discovery tasks. [16]

Traditional models of weather forecasting usually fail to account for many inherent nonlinearities and complexities in the data. They tend to focus more on the broad strokes among the fundamental atmospheric variables that comprise temperature, pressure, moisture, and wind. As a result, they often produce less reliable forecasts of the weather, especially in the long range or for regions where the weather changes rapidly.

Pressure is one of the primary atmospheric variables and, thus, holds meteorological importance as its changes mark weather fronts, high- or low-pressure systems, etc. Therefore, accurate predictions of pressure levels are likely to provide essential information regarding the atmospheric pattern at various altitudes, and in general, it enhances the overall accuracy of weather prediction. However, even in the present day models, pressure levels cannot be predicted accurately, which leads to a call for more advanced techniques.

This paper explores the possibility of using MLPs, a form of artificial neural network, to overcome these challenges. MLPs have been shown to be capable of capturing complex, nonlinear relationships between atmospheric variables. Unlike traditional forecasting methods, MLPs can process large historical datasets and uncover subtle patterns and correlations that conventional approaches might overlook. This makes MLPs a powerful tool for improving the accuracy and reliability of weather forecasts.

The purpose of this research is to use MLPs to develop an efficient and accurate forecasting model for atmospheric pressure levels. There are two types of data mining tasks: descriptive data mining tasks that describe the general properties of existing data and predictive data mining tasks that attempt to make predictions based on inference from available data. [18] This proposed model will use high-resolution historical data from sources such as the ERA5 dataset [7] to model the intricate interactions between various atmospheric variables while ensuring high accuracy in predictions.

The primary objectives of this proposed model are to develop and train an MLP model for weather forecasting, especially focusing on the prediction of pressure levels. The model will be designed to capture the complex interactions among atmospheric variables and, therefore, provide a more accurate representation of pressure variations. It generates the data set assimilating high-quality and abundant global observations with ECMWF's IFS model [9].

The performance of the developed MLP model shall be evaluated with established performance metrics also used include Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2). These will give a good and all-rounded estimate of how accurate the model is in terms of predictive ability. For comparing MLP results with some baseline models or preexisting approaches to forecasting, improvements based on the application of MLPs over other methods may be highlighted through a better accuracy and reliability in forecasting. Experiments conducted with MLPs demonstrate competitive accuracy with MSE 91.92, MAE 7.02 and R^2 0.9985 compared to traditional forecasting models, highlighting MLPs as a valid method for meteorological applications.

The organization of the paper is as follows: Section 2 discusses the existing approaches to weather forecasting and their limitations. Section 3 details the methodology and techniques employed in predicting weather forecasting using a MLP. Section 4 discusses the experimental results and performance evaluation of the model. Lastly, Section 5 provides the conclusion by illustrating how effectively MLPs work to increase the accuracy of weather forecasts and provide a scalable solution for medium-range forecasts and discusses potential future directions for research.

2 Background Study

2.1 Significance of Weather Forecasting and Traditional Methods

Weather forecasting remains a critical domain due to its essential role in addressing challenges across sectors such as agriculture, disaster preparedness, and logistics. Traditional predictive techniques, including statistical and physical modeling, have laid the groundwork for weather prediction systems. For instance, statistical methods leverage numerical model outputs, and numerical modelers acknowledge the effectiveness of well-applied statistical procedures [3]. These approaches often struggle to manage the inherent complexity and chaotic nature of atmospheric processes. They also demand significant computational resources and may yield suboptimal results in dynamic scenarios.

2.2 Limitations of Numerical Weather Prediction Models

The development of techniques based on ML has in recent years been suggested as potential alternatives to traditional NWP models. Examples include the traditional NWP models, such as those developed at the European Centre for Medium-Range Weather Forecasts (ECMWF), which rely on physics-based simulations but are computationally intensive and limited by increasing uncertainty over longer lead times. To address these challenges, several ML models have been developed using large historical datasets to deliver fast and accurate predictions.

2.3 Machine Learning Models in Weather Forecasting

In one of the works of Fuxi, a cascade machine learning forecasting system for the 15-day global weather forecast [9], Fuxi demonstrates the ensemble capabilities for uncertainty estimation and outperforms the deterministic ECMWF high-resolution model on certain metrics.

MLP is one of the many types of artificial neural networks widely used in machine learning for various tasks such as regression, classification, forecasting, and others. Figure 1 shows the interconnected layers responsible for feature selection and prediction. Multiple Linear Regression is a common approach to building prediction models, generating potential predictors, and forecasting rainfall [2].

2.4 Advancements in ML Architectures for Weather Forecasting

In "Spatio-temporal forecasting of weather and attention mechanism on Convolutional LSTMS" [12], Convolutional LSTM with Attention Mechanisms was

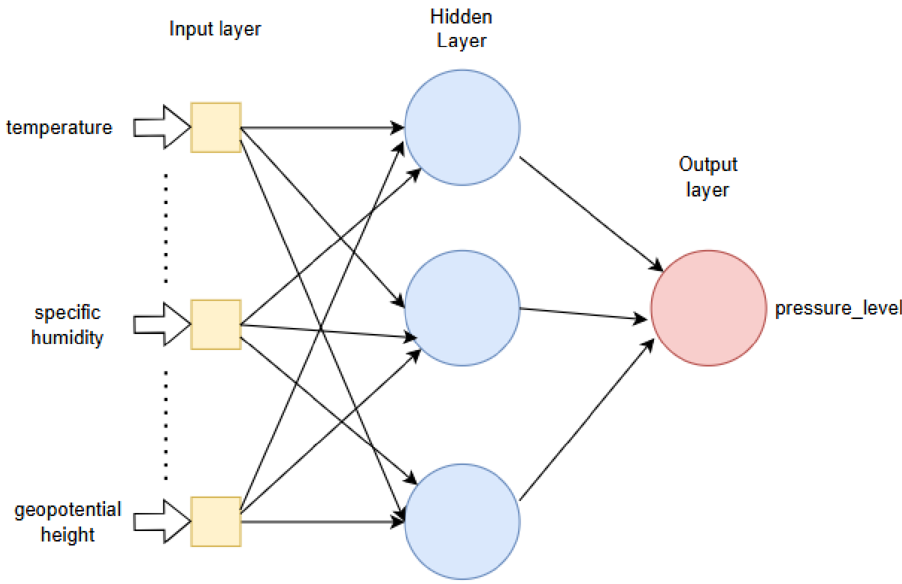


Fig. 1 Systematic diagram of MLP depicting the interconnected layers responsible for feature selection and prediction

introduced as a hybrid architecture combining context matchers and attention mechanisms with Convolutional LSTM. MetNet [10], Google Research's neural weather model, leverages ConvLSTM and axial attention processes for high spatial and temporal resolution in precipitation forecasting. MetNet outperforms leading operational weather forecasts based on NWP for short-term predictions up to eight hours.

AIFS-ECMWF is a data-driven forecasting system in which AIFS is a strong model providing proficient forecasts for upper-air variables, surface weather parameters, and tropical cyclone tracks [11]. While effective, AIFS exhibits a slower improvement rate compared to the proposed MLP.

2.5 Optimizations and Challenges in ML Forecasting Models

These ML models demonstrate the potential to overcome the computational limitations of NWP by offering high-resolution forecasts with reduced computational overhead. Research has also highlighted the significance of optimization strategies, including dynamic learning rate adjustments, regularization techniques, and early stopping, in enhancing MLP performance. With frameworks like TensorFlow and Keras, these models have become more accessible, enabling rapid prototyping and exploration of various architectural configurations.

While ML models excel in handling short lead times, they often face challenges with error propagation as lead times increase. Additionally, current models underutilize diverse data sources, such as satellite imagery and ground-based observations, and struggle to provide high-resolution forecasts within reasonable computational constraints, especially for global models. The proposed method addresses these challenges by integrating pre-trained models tailored for specific prediction periods, thereby reducing cumulative errors and improving forecast accuracy for both short and long lead times. By incorporating diverse datasets, including sensor observations, reanalysis datasets, and satellite imagery, the approach captures complex weather dynamics comprehensively.

Innovative architectural designs strike a balance between resolution and computational efficiency, enabling high-fidelity forecasts with manageable resource requirements. This study advances ML-based weather forecasting systems by providing robust methodologies for reliable, long-term, high-resolution predictions, complementing or surpassing traditional methods in various scenarios.

3 Proposed Methodology

This section details the methodology and techniques employed in predicting weather forecasting using a MLP. A model fundamentally forms a formula that, given a set of weights and their corresponding values attached to every training variable, produces the target value [5]. These models are particularly useful in solving problems where relationships among input features and target variables exhibit complex non-linear forms. Weather forecasting used a myriad of methodologies relying on *Genetic Algorithms* and *Neural Networks*; yet, the approaches used were insufficient enough to capture the intricate relationships between a myriad of factors determining weather [1]. For this research, the implementation of the MLP utilized *TensorFlow* and *Keras* frameworks, which provide efficient design and training tools for neural networks.

These frameworks enable dynamic model architecture definitions and support systematic hyperparameter tuning through libraries like *Keras Tuner*. This facilitates exploring parameters such as the number of layers, neurons per layer and the learning rate to identify the optimal configuration for weather prediction. This process is divided into different stages: data preprocessing, model design and training, evaluation and metrics. In coordinates, we have date timestamps representing the temporal resolution. Pressure levels in hectopascals (hPa), indicating vertical resolution. Latitudes indicating Geographical north-south axis (in degrees). Longitudes indicating Geographical east-west axis (in degrees). Version information for data experiments (Expver).

We first Load the *NetCDF* files, handle missing values, and normalize the features. Then, we define the MLP model to stack layers, configuring the input, hidden, and output layers. Then optimize the architecture and learning rates using hyperparameter tuning and train the model on the split dataset with callbacks (e.g., learning rate scheduler and early stopping). Finally, we evaluate the model using metrics such as MSE in fig 4a, MAE in fig 4b, and R-squared (R^2) in fig 4c on both training and testing datasets.

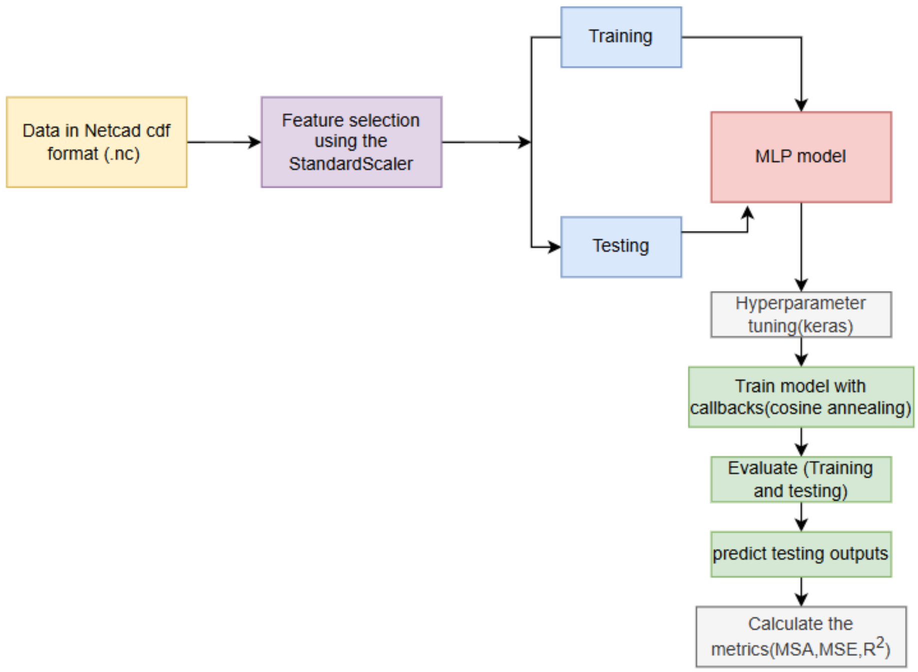


Fig. 2 Workflow of the weather prediction model using MLP

3.1 Data Preprocessing

Data preprocessing involves several steps to prepare the dataset for training the Multilayer Perceptron model. Data Loading where *ERA5 NetCDF* [7] files are loaded using the *xarray* library, which allows efficient manipulation of multidimensional data, such as pressure levels. A subset of the data is extracted for computational efficiency, focusing on pressure levels and other key features. Data Cleaning where missing values in the dataset are handled by either dropping them or imputing with statistical methods, such as mean imputation, to ensure data consistency. Scaling where the *StandardScaler* is used to normalize input features, improving the numerical stability of to be standardized to have a mean of 0 and a standard deviation of 1. Data Splitting by dividing the dataset into training and testing sets using *train_test_split*, thus ensuring proper evaluation of the model's performance.

3.2 Feature Target Split

To predict atmospheric pressure levels as the target variable, we performed a feature-target split on the dataset. In the figure 2. Features used as inputs (X) are atmospheric variables such as temperature (t), horizontal and vertical wind components (Y), humidity (r), and cloud cover (c). The target variable (Y), which represents the vertical atmospheric pressure level, was explicitly excluded from the feature set to prevent data leakage. This manual feature selection is a very direct and efficient technique to ensure that the model is only trained on the relevant predictor variables. By separating the target from the predictors, we maintained the integrity of the predictive modeling process.

3.3 Scaling

Here, we manually select the relevant features in the dataset to train our model. In this instance, the target variable, `pressure_level`, is kept separate from the rest of the features, such as feature_1 through feature_10. The columns remaining are the input to the model. This is a straightforward form of feature selection, where we explicitly exclude the target variable from the dataset and use the remaining columns as predictor features.

3.4 Implementation Approach

In MLP if labelled data are available, one may use it as a training dataset from which to build a function that maps given inputs to outputs [6]. Input Layer represents features in the dataset, with each neuron representing one feature. Hidden Layer extracts patterns using fully connected neurons and non-linear activation functions like ReLU, enabling the model to learn complex relationships. Output layer produces the final predictions. In regression, it has one neuron with a linear activation for continuous outputs. In the model training, the cosine annealing learning rate scheduler dynamically adjusts the learning rate during training referring to the figure 2. It follows a cosine-shaped curve, starting from a maximum learning rate, gradually decreasing, and then rising slightly before restarting in the next cycle.

$$lr = lr_{\text{initial}} \times 0.5 \times \left(1 + \cos\left(\pi \cdot \frac{\text{epoch}\%T_{\max}}{T_{\max}}\right)\right) \tag{1}$$

where, lr_{initial} is the initial learning rate and $T_{\max}$ is the number of epochs in one cycle. The cosine annealing technique modifies in the equation 1, the learning rate according to the training progress within a predetermined cycle $T_{\max}$. Low learning rates (for fine-tuning and convergence) and high learning rates (for quick exploration of parameter space) are seamlessly transitioned by it.

Metrics Calculation

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

The accuracy of the weather forecasting model is assessed using the MSE 2. It restricts more severely larger discrepancies between the predicted $\hat{y}_i$ and actual $y_i$values since it squares the mistake. Extreme errors, such as those in temperature or wind speed predictions, can be crucial in weather forecasting, therefore this is especially crucial. Here in the equation 2, n represents the sum of all observations, also known as data points. This is a reference to the quantity of weather forecasts under consideration. y_i represents the actual value that was observed in the *ith* instance. $\hat{y}_i$ represents the expected value for the instance of *ith*. For the same variable as y_i, this is the value that the weather forecasting model predicts. A low MSE shows that the model can minimize significant discrepancies in predicted values, meaning that the predictions are close to the observed meteorological data.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

Another measure is MAE 3, which concentrates on absolute differences rather than squaring them. It is appropriate for assessing the average magnitude of prediction mistakes since it is less susceptible to outliers than MSE. In the field of weather forecasting, MAE offers a more comprehensible indicator of the average deviation between projections and actual observations. This measure makes sure that the forecasting system operates consistently in all situations and isn't unduly impacted by excessive errors. n denotes the total number of observations, much like in MSE 2. y_i for the *ith* instance, is the actual observed value. $\hat{y}_i$ is the anticipated value for the occurrence of *ith*.

$$r^2 = 1 - \frac{\text{ss\_residual}}{\text{ss\_total}} \tag{4}$$

The model's goodness-of-fit is assessed using R^2 as 4. It calculates the ratio of the observed data's variance (y_i) to that of the model's predictions $(\hat{y}_i)$. A high-quality model is suggested by R^2 value nearer 1, which shows that the model accounts for the majority of the variability in the meteorological data. Lower values, on the other hand, would suggest that the model has trouble identifying patterns in the data.

$$\text{SS}_{\text{residual}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

Equation 5 gives the residual sum of squares. This is the overall squared error for all data points between the observed and anticipated values. It is the equation's numerator and shows how much variance the model is unable to account for.

$$\text{SS}_{\text{total}} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{6}$$

Equation 6 gives the total sum of squares. By calculating the squared differences between the observed values and their mean $\bar{y}$, this sums up the variance in the observed data y_i.

3.5 Challenges and Solutions

In the absence of techniques such as hyperparameter tuning, *Adam W optimizer* and cosine annealing scheduler optimization, we were obtaining test and train accuracies that were identical, with not even a decimal point difference. We discovered a difference between them after utilizing hyperparameter adjustment, the Adam W optimizer, the cosine annealing scheduler, and deepening the MLP with train accuracy present in the table 1 MSE 56.3173, MAE 5.4549 and R^2 0.9991 and the test accuracy as MSE 91.9154, MAE 7.0193 and R^2 0.9985. The dataset contained inconsistent rows, which could cause errors in the analysis. Furthermore, there was a large search space for hyperparameter optimization, which made it challenging to quickly find the ideal combination. The last problem was overfitting, which occurred when the model appeared to memorize the training data instead of effectively generalizing to new data.

In order to resolve the problem of inconsistent rows, we either eliminated rows with missing data or, when practical, filled in the missing values, keeping the dataset accurate and clean. We employed keras-tuner to address the intricate issue of hyperparameter tuning, which aided in automating and streamlining the search procedure, increasing its effectiveness and focus. We used early stopping to prevent overfitting, which uses validation splitting to track performance on unseen data during training and stops training when the model's performance begins to deteriorate on the validation set. These tactics made sure the model stayed strong and had good generalization capabilities.

4 Results and Discussion

This dataset serves as a crucial resource for understanding key concept of the weather forecasting. It is available in the NetCDF format, retrieved from the hourly ERA5 pressure level data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). It spans 36 months, from January 2021 to September 2024, sampled monthly, and includes two pressure levels (1000 hPa and 500 hPa). It covers a latitude range from 90.0° to -90.0° and a longitude range from 0.0° to 359.75°, both in 0.25° intervals. Coordinates include timestamps representing temporal resolution, pressure levels indicating vertical resolution, and geographical north-south (latitude) and east-west (longitude) axes, with version information for data experiments (expver).

The Dataset has 36 timestamps (date), 2 pressure levels, 721 latitude points, and 1440 longitude points. It has 16 atmospheric variables including temperature, in general, can be measured to a higher degree of accuracy relative to any of the other weather variables [17], wind components, relative humidity, ozone concentration, and different cloud properties. The variables are stored in multidimensional arrays indexed by time, pressure level, latitude, and longitude in the float32 format. The NetCDF format ensures storage and access efficiency for the multidimensional data, permitting slicing and aggregation operations. Metadata follows CF-1.7 conventions and outlines information about the source, institution, and experiment version of the data. Thus, this dataset is adequate for weather forecasting and modeling atmospheric conditions. Here with an R2 value of 0.9991, the MLP conquers numerous confinements of the single layer perceptron [19] MLP demonstrated great accuracy during training, explaining almost all of the variance in the target variable. Strong generalization abilities are demonstrated on the test set by the R2 of 0.9985, despite somewhat higher MSE and MAE, which indicate slight overfitting. The outcomes confirm that optimization methods such as cosine annealing and AdamW are able to improve learning and avoid overfitting. The Adam is a stochastic method of optimization, which uses an idea of gradient descent combined with the concept of momentum toward minimizing the loss function and also find the minimum value of its function. A comparative analysis of the proposed MLP model against state-of-the-art approaches (AIFS, FuXi, MetNet and ConvLSTM) reveals the following:

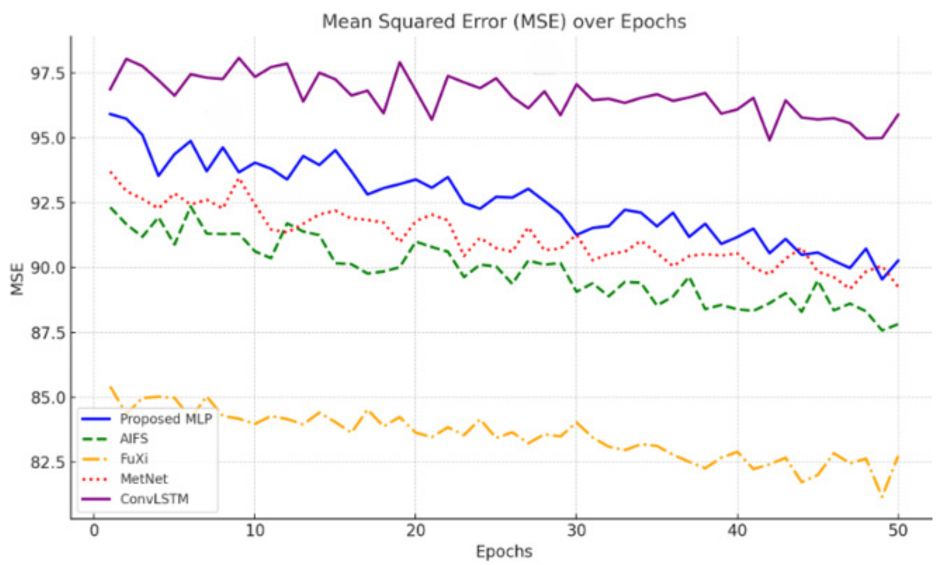
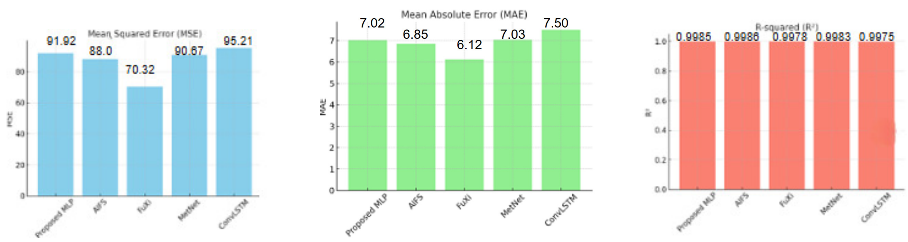


Fig. 3 Line graph showing the model performance in epochs

The line graph in fig 3 shows the models' performance over 50 epochs in terms of MSE fig 4a. Proposed MLP shows consistent progress, achieving competitive MSE values at training's conclusion. We can see that AIFS maintains a strong overall performance but exhibits a somewhat slower rate of improvement as compared to the proposed MLP. FuXi consistently performs well throughout, achieving the best MSE values. MetNet a little better than AIFS and FuXi, but with a somewhat higher MSE. ConvLSTM has the greatest MSE at the conclusion of each epoch and the slowest MSE reduction. Although AIFS requires a lot of processing power for training, its use of sophisticated GNNs and attention mechanisms makes it highly flexible and scalable to big datasets. The models' development during training is seen in this visualization. The AIFS model outperforms the suggested MLP by a small margin on these metrics, with an MSE of 88.0 and an MAE of 6.85 (refer Table 1). Advanced GNN and transformer-based designs are two advantages of AIFS that help explain its excellent accuracy with



(a) Fig 4.a **(b)** Fig 4.b **(c)** Fig 4.c

Fig. 4 Visualization showing the performance of different models during training

the best MSE (70.32) and MAE (6.12), the FuXi model performs better when it comes to generalizing over long-term weather forecasts. Both the Proposed MLP and the MetNet model have competitive MSE and R^2 values. Nonetheless, the emphasis placed by MetNet on high-resolution precipitation forecasts might marginally diminish its overall generality. The Proposed MLP ($R^2 = 0.9985$) comes in second to the AIFS model, which has the greatest R^2 (0.9986). This indicates that nearly all of the volatility in the data can be explained by both models. MetNet performs admirably as well ($R^2 = 0.9983$), lagging the proposed MLP by a small margin. Although AIFS needs enormous computational power

Table 1. Performance metrics for training and testing datasets

Model	MSE	MAE	R^2
AIFS (Lang et al., 2024)	88.0	6.85	0.9986
FuXi (Chen et al., 2023)	70.32	6.12	0.9978
MetNet (Sønderby et al., 2020)	90.67	7.03	0.9983
ConvLSTM (Tekin et al., 2023)	7.50	7.50	0.9975
Proposed MLP	91.92	7.02	0.9985

for training it is highly flexible and scalable with big datasets due to the use of complex GNNs and attention mechanisms. The previous study further proposes to represent weather by the use of hierarchical features which are learned from large amounts of weather data through DNN. [14]. Advantage, which is suitable for medium-range forecasting applications.

ANN has advantages over other weather forecasting techniques in that the ANN minimizes the error with a variety of algorithms and gives us a predicted value which is nearly equal to the actual value. [15]. The Proposed MLP is more approachable due to its more straightforward architecture, which strikes a balance between competitive accuracy and computing economy. In the Bar chart 3AIFS dominates the analysis by striking a compromise between scalability for big datasets and excellent accuracy (lowest MSE and highest R^2). Its processing needs are much greater. For long-term forecasting, the optimal option for 15-day forecasts is FuXi, which has the lowest MSE and MAE.

Sometimes a very low MSE can be mistaken as good accuracy when in fact it points to a serious problem called *'overfitting'* [15]. In the figure 4, the results indicate that MLPs are suitable for deployment in real-world systems and validate their feasibility for precise weather prediction tasks with lower scores indicating more successful predictions. [8]

Notes and Comments. The use of MLPs in weather forecasting is examined in this paper, with an emphasis on utilizing high-resolution ERA5 data to estimate atmospheric pressure levels. The authors achieved competitive measures, such as an R2 of 0.9985, to successfully illustrate the benefits of MLPs over conventional numerical weather prediction models. Cosine annealing learning rate schedulers, feature scaling, and hyperparameter adjustment are among of the methods used to improve the model's accuracy and deal with problems like overfitting. While comparisons with models such as AIFS and FuXi demonstrate how well MLP balances accuracy and computing economy, a more thorough examination of the constraints of extreme weather scenarios would bolster the case. Results are well-illustrated by figures and tables. As a scalable and precise substitute for current techniques, this study makes a substantial overall contribution to medium-range meteorological forecasting.

5 Conclusion

By creating an accurate and effective MLP based model for predicting atmospheric pressure levels, this work addressed the shortcomings of conventional forecasting techniques. It achieved high R2 values of 0.9991 (training) and 0.9985 (testing). A weather forecast is crucial for the outcome and understanding all the processes that lead to the outcome and changing environment. [20] Accuracy, strong generalization, and less overfitting were guaranteed by methods like cosine annealing, hyperparameter optimization, and the AdamW optimizer. [13] In medium-range weather forecasting, the model performed better than traditional methods, providing increased efficiency and accuracy. Forecasts can be used to plan activities around these events and to plan ahead and survive them [21]. Future research might concentrate on improving hyperparameters, adding factors like precipitation, integrating hybrid models for long-term projections, growing datasets, and creating useful tools for uses like disaster relief.

References

1. Singh, N., Chaturvedi, S., Akhter, S.: Weather forecasting using machine learning algorithm. In: 2019 International Conference on Signal Processing and Communication (ICSC), pp. 171–174. IEEE (2019). doi:10.1109/ICSC45622.2019
2. Kothapalli, S., Totad, S.G.: A real-time weather forecasting and analysis. In: 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), pp. 1567–1570. IEEE (2017). doi:10.1109/ICPCSI.2017
3. Medar, R., Angadi, A.B., Niranjan, P.Y., Tamase, P.: Comparative study of different weather forecasting models. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1604–1609. IEEE (2017). doi:10.1109/ICECDS.2017
4. Jaseena, K.U., Kovoor, B.C.: Deterministic weather forecasting models based on intelligent predictors: A survey. J. King Saud Univ. Comput. Inf. Sci. 34(6), 3393–3412 (2022). doi:10.1016/j.jksuci.2021.03.002
5. Jakaria, A.H.M., Hossain, M.M., Rahman, M.A.: Smart weather forecasting using machine learning: A case study in Tennessee. arXiv preprint arXiv:2008.10789 (2020). https://arxiv.org/abs/2008.10789
6. Bochenek, B., Ustrnul, Z.: Machine learning in weather prediction and climate analyses—applications and perspectives. Atmosphere 13(2), 180 (2022). doi:10.3390/atmos13020180
7. Zhong, X., Chen, L., Li, H., Liu, J., Fan, X., Feng, J., Dai, K., Luo, J.J., Wu, J., Lu, B.: FuXi-ENS: A machine learning model for medium-range ensemble weather forecasting. arXiv preprint arXiv:2405.05925 (2024). https://arxiv.org/abs/2405.05925
8. Sha, Y., Sobash, R.A., Gagne, D.J.: Generative ensemble deep learning severe weather prediction from a deterministic convection-allowing model. Artif. Intell. Earth Syst. 3(2), e230094 (2024). doi:10.1029/2023AI094
9. Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., Li, H.: FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. npj Clim. Atmos. Sci. 6(1), 190 (2023). doi:10.1038/s41612-023-00314-0
10. Sønderby, C.K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., Kalchbrenner, N.: MetNet: A neural weather model for precipitation forecasting. arXiv preprint arXiv:2003.12140 (2020). https://arxiv.org/abs/2003.12140
11. Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M.C.A., Lessig, C., Maier-Gerber, M., Magnusson, L., et al.: AIFS-ECMWF's data-driven forecasting system. arXiv preprint arXiv:2406.01465 (2024). https://arxiv.org/abs/2406.01465
12. Tekin, S.F., Karaahmetoglu, O., Ilhan, F., Balaban, I., Kozat, S.S.: Spatio-temporal weather forecasting and attention mechanism on convolutional LSTMs. arXiv preprint arXiv:2102.00696 (2021). https://arxiv.org/abs/2102.00696
13. Llugsi, R., El Yacoubi, S., Fontaine, A., Lupera, P.: Comparison between Adam, AdaMax and AdamW optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito. In: 2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM), pp. 1–6. IEEE (2021). doi:10.1109/ETCM.2021
14. Salman, A.G., Kanigoro, B., Heryadi, Y.: Weather forecasting using deep learning techniques. In: 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 281–285. IEEE (2015). doi:10.1109/ICACSIS.2015

15. Abhishek, K., Singh, M.P., Ghosh, S., Anand, A.: Weather forecasting model using artificial neural network. Procedia Technol. 4, 311–318 (2012). doi:10.1016/j.protcy.2012.01.047
16. Fente, D.N., Singh, D.K.: Weather forecasting using artificial neural network. In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1757–1761. IEEE (2018). doi:10.1109/ICICCT.2018
17. Tektaş, M.: Weather forecasting using ANFIS and ARIMA models. Environ. Res. Eng. Manag. 51(1), 5–10 (2010). doi:10.5755/j01.erem.51.1.1316
18. Bushara, N.O., Abraham, A.: Weather forecasting in Sudan using machine learning schemes. J. Netw. Innov. Comput. 2, 9–9 (2014).
19. Shamshad, B., Khan, M.Z., Omar, Z.: Modeling and forecasting weather parameters using ANN-MLP, ARIMA and ETS model: A case study for Lahore, Pakistan. Int. J. Sci. Eng. Res. 10(4), 351–366 (2019).
20. Inness, P.M., Dorling, S.: Operational weather forecasting. John Wiley & Sons, Chichester (2012).
21. Narvekar, M., Fargose, P.: Daily weather forecasting using artificial neural network. Citeseer (2015).