Lightweight Deep Learning for Weather Prediction and Forecasting in Africa

Kinyua Gikunda[1][0000−0001−7962−2168] and Nicolas Jouandeau[2][0000−0001−6902−4324]

[1] Dedan Kimathi University of Technology, Nyeri, Kenya
patrick.gikunda@dkut.ac.ke
[2] Université Paris 8, Vincennes - Saint-Denis
[3] n@up8.edu

Abstract. Weather forecasting in Africa is hampered by sparse meteorological data and limited computational resources. This paper addresses these challenges by proposing lightweight deep learning (DL) for weather prediction and forecasting. We integrate active learning and transfer learning methods to enhance model training efficiency and accuracy. By focusing on the informativeness and representativeness of training samples, our approach significantly reduces the need for extensive and costly labeling. After training on a source dataset, model skills are transferred to target datasets, allowing for effective weather variable predictions with minimal data. Extensive experiments on three weather datasets demonstrate that our hybrid Transfer Active Learning method achieves similar classification accuracy compared to existing methods, using only 20% of the training samples. This study highlights the potential of advanced DL techniques to improve weather forecasting in Africa, despite the constraints of data scarcity and limited computational infrastructure.

Keywords: Weather Forecasting · Deep Learning · Transfer Learning · Active Learning

1 Introduction

Weather forecasting is a critical component in managing and adapting to environmental changes, particularly in Africa [1]. The continent faces unique challenges due to its vast geographical diversity and limited availability of meteorological data. Many regions in Africa have sparse weather station networks, resulting in uneven and incomplete datasets [2]. Additionally, the computational resources required for advanced weather prediction models are often scarce, further complicating accurate forecasting efforts. These challenges necessitate innovative approaches that can leverage available data and computational resources efficiently. Deep learning (DL) combined with strategies like active learning and transfer learning offers promising solutions to enhance weather prediction and forecasting accuracy in Africa. By utilizing lightweight DL models, it is possible to achieve

weather forecasts even in data-scarce and resource-constrained environments, ultimately aiding in better decision-making and resource management across the continent.

2 Deep Learning for Weather Prediction

The non-linear behavior of meteorological data poses significant challenges for weather prediction, even with state-of-the-art numerical models [3]. This complexity has led researchers to explore emerging Artificial Intelligence (AI) approaches, which have demonstrated impressive performance in various fields [4]. Traditional parametric models, such as linear models, struggle with meteorological data due to their limited expressive power and inability to stack linear operations for more abstract representations [5]. Non-parametric learners like Gaussian kernels offer flexibility but are hindered by their reliance on local generalization and the exponential growth of input dimensionality.

Deep Learning (DL) methods address these challenges by stacking multiple feature learning layers to form deep representations, enhancing both computational and statistical efficiency. Recent advancements have improved the representation of inputs with fewer parameters, allowing for effective feature learning using both labeled and unlabeled data. Transfer Learning (TL), a process within DL, leverages learned features to apply knowledge from one domain to another related domain, improving learning efficiency and effectiveness. This makes DL particularly suitable for complex and dynamic fields like weather prediction.

Deep learning methods, especially convolutional neural network (CNN)-based time series classifiers, have proven highly effective for extracting temporal and spatial features from spatio-temporal weather data [7]. These methods offer faster and more accurate predictions and can handle large, complex datasets from weather satellites and IoT devices [8]. Unlike traditional models, DL do not require extensive feature engineering, making them more adaptable and practical for weather forecasting applications.

The flexibility and robustness of DL approaches make them well-suited for the complexities of weather data, which often exhibit non-linear and chaotic behavior. DL models, leveraging distributed and sparse representations, can capture intricate data structures that traditional parametric and non-parametric models struggle to represent effectively. This capability is crucial for processing high-dimensional meteorological datasets, where capturing subtle patterns and correlations can significantly enhance prediction accuracy.

DL's superior feature learning capabilities allow for better representation and understanding of weather patterns, leading to improved prediction accuracy and reliability [9]. These techniques reduce the need for manual data preprocessing and feature extraction, streamlining the forecasting process. Moreover, DL methods excel at learning from vast amounts of data, continually improving predictive performance as more data becomes available. Their scalability ensures that forecasting systems remain efficient and effective even as data volumes grow, making DL particularly beneficial for weather forecasting.

3 Transfer Learning and Active Learning

To address the challenge of sparse training data in time series datasets, the proposed model incorporates two primary DL techniques: Transfer Learning and Active Learning.

TL allows the model to leverage pre-existing knowledge from a related source task and apply it to the target task. This technique enhances the model's ability to generalize and perform well even with limited data by re-using model skills. AL dynamically queries and selects the most informative samples to add to the training set. It uses labeled data to provide critical information about class labels or boundaries, while unlabeled data helps in understanding the base data distribution. This iterative process improves the efficiency of the learning process by focusing on the most useful data points.

Before delving into the specifics of these techniques, it is essential to define the Time Series Classification (TSC) problem.

Definition 1. *An univariate time series* $U_t = [x_1, x_2, ..., x_T]$ *is an ordered set of real values. The length of* U_t *is equal to the number of observable time-points* T*.*

Definition 2. *A multivariate time series* $M_t = U_t^1, U_t^2, ..., U_t^n$ *consist of n observations per time-point with* $U_t^i \in R^T$

Definition 3. *A dataset* $D = (X_1, Y_1), (X_2, Y_2), ..., (X_N, Y_N)$ *is a collection of pairs* (X_i, Y_i) *where* X_i *could either be Ut or Mt with* Y_i *as its corresponding label. For a dataset containing* K *classes, the label vector* Y_i *is a vector of length* K *where each element* $j \in [1, K]$ *is equal to 1 if the class of* X_i *is* j *and 0 otherwise.*

We can define Time Series Classification (TSC) as the task of mapping time-based inputs to a probability distribution over a set of labels. This can be mathematically represented by the following equation:

$$C_t = f(w * U_{t-l/2:t+l/2} + b) | \forall t \in 1, T \tag{1}$$

C denotes the convolution result on a univariate time series U_t of length T with a filter w of length l, a bias parameter b and a non-linear function f. Applying several filters on a time series will result in a multivariate time series whose dimensions are equal to the number of filters used. Using the same filter values w and b in ConvNets its possible to find the results for all time stamps $t \in [1, T]$. This is possible by using weight sharing that enables the model to learn feature detectors that are invariant across the time array

4 Deep Transfer Active Learning

During target training, the model's parameters are initialized using weights from a previous task, represented as $\Theta \leftarrow \vartheta_\theta$. After initializing the weights, a forward

pass through the model is performed using the function $f(\theta, x_i)$, which computes the output for an input x_i. The output is a vector of estimated probabilities for x_i belonging to each class. The prediction loss is then computed using a cost function, such as the negative log likelihood. Using gradient descent, the weights are updated in a backward pass to propagate the error. This iterative process of forward pass followed by backpropagation updates the model's parameters to minimize the loss on the training data. During testing, the model is evaluated on unseen data. A forward pass is performed on the new input, followed by class prediction. The predicted class corresponds to the one with the highest probability. For this, categorical cross-entropy is applied as the loss function, denoted as:

$$L(y, \hat{y}) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{2}$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability for class i. This loss function helps to measure the performance of the classification model by comparing the predicted probabilities with the actual labels.

AL is used to select instances a model is most uncertain about to improve learning efficiency. In uncertainty sampling, the model aims to identify and learn from the most informative data points. Three primary metrics used to define uncertainty are least confidence, sample margin, and entropy. To take consideration of the entire output distribution, entropy is used as a metric which is defined as:

$$f_u(x) = \arg\max_i -\sum_i P(y_i|x_i) \log P(y_i|x_i) \tag{3}$$

Here, $P(y_i|x_i)$ is the posterior probability of instance x_i belonging to class i. For binary classification, the most uncertain instances are those with nearly equal probabilities for both classes.

Besides uncertainty, considering the distribution of instances can enhance AL performance. Instance diversity helps in selecting the most representative samples, thus improving query performance and avoiding outliers.

The correlation measure assesses the pairwise similarities of instances. The informativeness of an instance is determined by its average similarity to its neighbors. For two instances x_i and x_j, the correlation measure f_c is defined as:

$$f_c(x) = \frac{1}{DU} \sum_{x_j \in DU/x_i} f_c(x_i, x_j) \tag{4}$$

The value of $f_c(x_i)$ represents the density of x_i in the unlabeled set. Higher values indicate that an instance is closely related to others, while lower values suggest outliers, which should be avoided for labeling.

To select the most informative and representative samples, a heuristic combination of correlation and uncertainty measures is employed. The most effective instance to label can be expressed as:

$$\hat{x} = \arg\max_i (f_u(x) \cdot f_c(x)) \tag{5}$$

This approach ensures that the selected samples are both uncertain and representative, enhancing the learning process.

5 Results

Three datasets were used in the experiments namely: a) RAUS[4] dataset contains daily weather observations from various Australian weather stations for a period of 10 years, b) KenCentralMet (Kenya Meteorological Department[5] privately acquired daily weather observations covering Central Kenya for a period of 3 years from 2012-2014) and c) MeteoNet[6] a meteorological dataset developed and made available by the French national meteorological service. For each of the dataset, less than 20% of the labeled samples was used as the initial training set. We present comparison of the proposed DTAL method, as detailed in the previous section, against: i) Random selection of data samples to query, iii) QUIRE method inspired by the margin based active learning from the minimax viewpoint with emphasize on selecting unlabeled instances that are both informative and representative [10], iv) DFAL method that selects unlabeled samples with the smallest perturbation. The distance between a sample and its smallest adversarial example better approximates the original distance to the decision boundary [11], v) Core-Set non-uncertainty based AL method [12].

	RAUS			**KenCentralMet**			**MeteoNet**		
	$\mathbb{P}$	$\mathbb{R}$	$\mathbb{A}$	$\mathbb{P}$	$\mathbb{R}$	$\mathbb{A}$	$\mathbb{P}$	$\mathbb{R}$	$\mathbb{A}$
Random	81	80	79	64	67	62	89	85	91
DTAL	80	**85**	**85**	**68**	64	67	**91**	90	**93**
QUIRE	**89**	84	81	67	**68**	67	87	88	86
DFAL	83	82	80	60	62	64	**91**	88	**93**
Core-Set	79	83	84	65	65	**68**	90	**91**	91

Table 1. Experimental results with Precision $\mathbb{P}$, Recall $\mathbb{R}$ and Accuracy $\mathbb{A}$.

Table 1 shows that DTAL generally outperforms a bit other methods (except QUIRE that is better with RAUS), especially in terms of precision and recall, demonstrating the effectiveness of the proposed hybrid strategy in selecting the most valuable training samples from the distribution. However, performance varies depending on the dataset, highlighting the importance of dataset characteristics in the efficacy of active learning methods and demonstrates that results can be equivalent even with less samples.

6 Conclusion

This paper demonstrates the efficacy of lightweight deep learning, integrating active and transfer learning, for weather prediction in Africa. Our hybrid Transfer

[4] https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package
[5] https://meteo.go.ke/
[6] https://www.kaggle.com/datasets/katerpillar/meteonet

Active Learning method significantly enhances forecasting accuracy with minimal data, using only small portion of the training samples compared to existing methods. Despite challenges of data scarcity and limited computational resources, our approach shows promise in providing good weather forecasts essential for effective decision-making and resource management in Africa. Future work will focus on refining these techniques and validating their practical benefits in real-world applications.

References

1. Cooper, P. J., Dimes, J., Rao, K. P. C., Shapiro, B., Shiferaw, B., & Twomlow, S. (2008). Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: an essential first step in adapting to future climate change?. Agriculture, ecosystems & environment, 126(1-2), 24-35.
2. Radeny, M., Desalegn, A., Mubiru, D., Kyazze, F., Mahoo, H., Recha, J., ... & Solomon, D. (2019). Indigenous knowledge for seasonal weather and climate forecasting across East Africa. Climatic Change, 156, 509-526.
3. Benavides Cesar, L., Amaro e Silva, R., Manso Callejo, M. Á., & Cira, C. I. (2022). Review on spatio-temporal solar forecasting methods driven by in situ measurements or their combination with satellite and numerical weather prediction (NWP) estimates. Energies, 15(12), 4341.
4. Das, M., & Ghosh, S. K. (2018). Data-driven approaches for meteorological time series prediction: a comparative study of the state-of-the-art computational intelligence techniques. Pattern Recognition Letters, 105, 155-164.
5. Cohen, N., Sharir, O., & Shashua, A. (2016, June). On the expressive power of deep learning: A tensor analysis. In Conference on learning theory (pp. 698-728). PMLR.
6. Langer, M., He, Z., Rahayu, W., & Xue, Y. (2020). Distributed training of deep learning models: A taxonomic perspective. IEEE Transactions on Parallel and Distributed Systems, 31(12), 2802-2818.
7. Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: a survey. Big Data, 9(1), 3-21.
8. Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z. (2023). Machine learning methods in weather and climate applications: A survey. Applied Sciences, 13(21), 12019.
9. Huang, G., Wang, Y., Ham, Y. G., Mu, B., Tao, W., & Xie, C. (2024). Toward a learnable climate model in the artificial intelligence era. Advances in Atmospheric Sciences, 1-8.
10. Huang, S. J., Jin, R., & Zhou, Z. H. (2010). Active learning by querying informative and representative examples. Advances in neural information processing systems.
11. Ducoffe, M., & Precioso, F. (2018). Adversarial active learning for deep networks: a margin based approach. arXiv preprint arXiv:1802.09841.
12. Sener, O., & Savarese, S. (2017). Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489.