SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Self-Organizing Deep Learning Architectures For Scalable Data Analysis In Environmental Climate Prediction Applications

Dr.K.R. Sowmya[1*], V. Sivasankari[2], K. Samundeeswari[3], Dr.U. Nilabar Nisha[4], Debarghya Biswas[5]

[1*]Professor, Department of Management Studies, SRM Valliammai Engineering College, Coimbatore, Tamil Nadu, India.
E-mail: drkrsowmya@gmail.com, https://orcid.org/0000-0003-4239-2586

[2]Assistant Professor, Department of Mathematics, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Coimbatore, Tamil Nadu, India. E-mail: sivasankariv@maher.ac.in

[3]Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Coimbatore, Tamil Nadu, India. E-mail: kscommerce@maher.ac.in

[4]Associate Professor, Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India.
E-mail: lotusnisha19@gmail.com, https://orcid.org/0000-0001-8750-6669

[5]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.debarghyabiswas@kalingauniversity.ac.in, https://orcid.org/0009-0008-5051-6896

*Corresponding author: Email: drkrsowmya@gmail.com

Abstract

Environmental climate predictions through deep learning architectures such as Convolutional Recurrent Networks or even large Transformer-based weather forecasters are often engineered based on an a priori decided upon architecture and capacity of the model, thereby necessitating a manual re-designing of the network whenever there is an increase in size, geographical coverage, and complexity of the input data, a problem that frequently occurs as climate observation networks evolve from a few local networks to full-scale continental or even global networks. In this paper, we introduce the Self-Organizing Deep Architecture (SODA) model where a competitive, Kohonen-type self-organizing topology layer assigns the incoming spatiotemporal patterns of the climate into an increasing number of regional experts, creating new experts as soon as the quantization error-based complexity measure indicates the inadequacy of the existing topology. Every expert has a deep spatiotemporal prediction backbone, and an aggregated map-reduce style layer merges the predictions of all experts into a global prediction, whereas a reorganization cycle periodically changes the architecture whenever the climate regime changes due to events like ENSO phase changes. The performance of this framework is measured for multivariate climate data modeled based on ERA5 reanalysis data variables for four different data-scale scenarios: single-region, multi-region, continental-scale, and global-scale with extreme events. With respect to a CNN-LSTM baseline model and a Transformer baseline model with a fixed architecture, SODA framework decreases temperature prediction RMSE from 2.87°C and 2.31°C to 1.39°C in the global-scale scenario, and also achieves higher training throughput, which suggests that self-organizing growth of architecture results in better performance and scalability with increasing climate data complexity.

Keywords Self-Organizing Neural Networks, Scalable Deep Learning, Climate Prediction, Growing Architectures, Distributed Data Analysis, Spatiotemporal Forecasting, Environmental Monitoring.

1. Introduction

Prediction of environmental climate conditions through weather forecasting to season prediction and climate prediction is becoming more dependent on machine learning models based on dense observations from multiple sources. Convolutional and recurrent models have been found highly skilled in predicting spatiotemporal tasks such as precipitation nowcasting [7] and detecting extreme weather conditions [8]. In recent times, transformer-based large-scale data-driven weather models such as FourCastNet [4], GraphCast [5], and Stormer [6] have achieved similar skill as operational numerical weather prediction models on medium-range forecasting benchmarks but with a much faster run-time performance.

One aspect shared by many of these architectures, though, is that their capacity and architecture is fixed from design time: the number of layers, filters, or attention heads is selected by the designer prior to training and cannot adapt to changing data sizes, scales, or complexity as the amount, extent, or complexity of the inputs themselves change. This represents a major drawback for real-world applications of climate monitoring, since the size of available data will increase significantly as a system is scaled from single-station networks to those operating continentally or globally, as well as when rare but important regimes, like ENSO phase transitions or extreme event seasons, are encountered. The designer must then either redesign and retrain the architecture on the new scale of data, a costly and disruptive endeavor, or deploy an architecture with its capacity selected to deal with the largest scale data from the beginning, which results in wasteful computations at lower scales.

This paper is concerned with the challenge of constructing a deep learning architecture for climate prediction where the architecture is structured and scaled up by itself according to the amount and complexity of data it is presented with, rather than having to be predetermined. The contributions of this research are the following.

- Develop an architecture framework named Self-Organizing Deep Architecture (SODA), where a competitive, self-organizing topology is paired with the growth of regional expert sub-networks, so that the architecture itself scales its capacity with the growth in data size and spatial complexity, based on the growing cells concept from classical self-organizing systems.
- Create a distributed aggregation layer based on the map reduce approach which merges the output of independently trained regional experts into a unified global prediction model, allowing for the scalability of training across different data partitions without training a single big monolithic model from scratch.
- Compare the performance of the developed framework explicitly under four different data size conditions, starting with a single region and moving to global data with extreme events, focusing on the impact of self-organizing architecture growth.

Organization of the paper is as follows. Section 2 briefly reviews the relevant literature on self-organizing neural network architecture and deep learning techniques in climate prediction. Section 3 describes the proposed approach. Section 4 describes the data set used, results of experiments and discussion. Section 5 concludes the paper.

2. Literature Survey

The concept behind self-organizing neural architectures forms the basis of the framework presented. Kohonen [1] first came up with the idea of self-organizing maps (SOMs), which is a type of neural network that implements unsupervised competitive learning where the high dimensional input data is projected into a lower dimensional lattice such that similar input vectors are mapped to close lattice sites; topology preservation is therefore the fundamental principle for mapping climate pattern regions to appropriate expert nodes. Agustsson [14] further came up with the concept of neural gas, which is a competing learning algorithm similar to SOM, but which is able to adapt better to the intrinsic dimensionality of the input vector than a SOM. Deng (2003) later extended the idea by developing the idea of Growing Cell Structures, which is a self-organizing neural network that starts from a minimal configuration and grows cells as required depending on quantization errors [2].

Another important theme to consider is architecture search and scaling of training. Specifically, Chen [3] presented the method of Neural Architecture Search (NAS), where he utilized reinforcement learning so that the controller network could generate architecture candidates and learn from their validation performance, illustrating the feasibility of automation and optimization of architecture design instead of manual design; NAS is different from the approach taken in this paper in the sense that it seeks the single best fixed architecture offline, while SODA adapts its architecture to data online. Dean and Ghemawat [15] developed the MapReduce framework, which is a basic programming paradigm for distributed processing of large-scale datasets on clusters of commodity machines, where computations were broken down into parallel map and reduce steps; this forms the basis for the distributed aggregation layer employed in the current framework to merge independent region-specific experts. McMahan et al., (2017) came up with Federated Averaging, another distributed learning paradigm based on updating a shared model by aggregating local weight update computations performed by decentralized clients [12].

Deep learning algorithms specifically designed for climate and weather prediction are increasingly being developed. Shi et al., (2015) introduced the convolutional LSTM (ConvLSTM) by incorporating convolution operations within the LSTM's recurrent connections to learn spatial and temporal information together, which was found to be better than existing precipitation nowcast models; the ConvLSTM-like architecture has been widely used in other climate deep learning algorithms, such as the per-expert prediction backbone in this paper [7]. Prabhat et al., (2020) utilized deep convolutional neural networks to predict extreme weather conditions from climate simulations, showcasing the effectiveness of deep learning to recognize rare patterns within climate data, an aspect that is tested directly in the global scale with extremes condition in this paper [8]. In recent years, however, large-scale, data-driven weather models have seen tremendous improvements in the state-of-the-art: First, FourCastNet [4] was proposed, where adaptive Fourier neural operators are leveraged to perform global weather forecasting at high resolution much more efficiently than numerical weather prediction; Second, GraphCast [5], a graph neural network based on a multi-mesh representation of the globe, matched or outperformed the leading operational forecasting system in terms of forecast skill for most of the evaluated variables and lead times; Third, Stormer [6], a scaled transformer architecture that achieved state-of-the-art medium-range forecast skill by making smart architectural choices in terms of tokenization, pre-training, and randomized dynamics forecasting targets. It should be mentioned that all three large-scale architectures depend on a well-tuned architecture trained on a globally comprehensive large dataset only once, in contrast to the architecture that evolves in tandem with scaling data, which is exactly the issue this paper aims to solve.

There are two more important themes that influence the backbone and adaptability design of the proposed architecture. The self-attention mechanism used in the transformer layers that form each regional expert's backbone was introduced by Vaswani et al., (2017) [9], while residual learning technique making it possible to train deeper networks utilized in late-stage experts was introduced by He, Zhang, Ren, and Sun [13]. Representation learning is defined by Bengio, Courville, and Vincent [10] as the process of finding ways to transform data to make prediction easier, which corresponds to the view of the self-organizing topology itself, and not only the expert's weights, as a learned representation of the climate system regions. Parasi, Kemker, Part, Kanan, and Wermter [11] conducted a review of the continual learning literature on how to adapt neural network architectures to a different distribution of the input data without catastrophic forgetting of the previous learned representations.

In combination, this body of literature demonstrates the maturity of self-organization architectures and growing networks architectures, mature paradigms of distributed computing for scalability of data processing and increasingly sophisticated fixed architecture deep learning models for climate/weather forecasting. Nevertheless, the large-scale climate models currently defining the state of the art learn from data using a fixed architecture designed for a particular target scale, raising an open question about how a climate forecasting system would be able to scale its architecture incrementally with respect to the data scale and complexity of region, without needing a full redesign of architecture at each scale transition point.

3. Methodology

The proposed Self-Organizing Deep Architecture (SODA) architecture includes six phases arranged in a closed loop structure: data acquisition, self-organizing topological assignment, complexity-driven growth, per expert spatiotemporal prediction, distributed aggregation, and cyclic topological reassignment. Figure 1 shows the overall process flowchart.

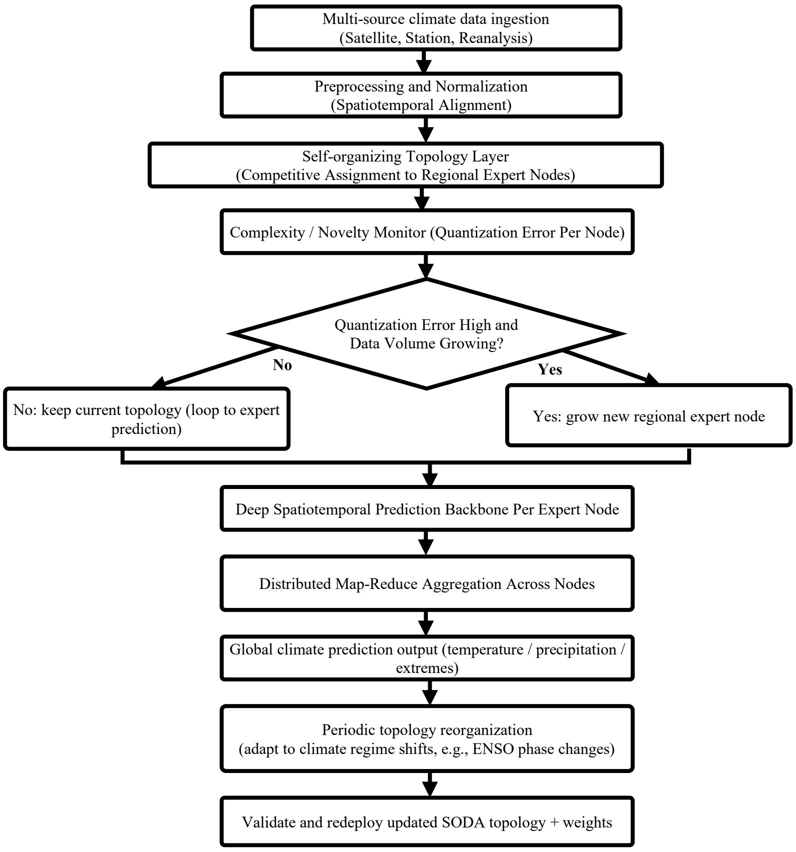

Figure 1: Methodology flow diagram of the proposed self-organizing deep architecture framework

3.1 Data Ingestion and Preprocessing

Multi-sourced climate data like that from satellites, ground stations and reanalysis are spatio-temporally registered on a common grid, normalized for each individual variable and fed into the self-organizing topology layer.

3.2 Self-Organizing Topology Assignment

Under the competitive learning rule of the self-organizing map [1], each newly presented spatiotemporal pattern of climate is allocated to the regional expert node that most closely matches the newly presented pattern, and the weight vectors of the winning node as well as those of its topologically neighbouring nodes are modified to better match the allocated pattern while maintaining topological relations among the geographic and climatic regions.

3.3 Complexity-Triggered Growth

The complexity measure estimates the accumulated quantization error, which is the difference between the representation of each node and the patterns allotted to it, as per the growing cell structure paradigm [2]. If the accumulated quantization error of a node is found greater than some threshold and at the same time there is a continued increase in the amount of data that it is processing, a new node referred to as a regional expert node is added in the proximity of the said node. This is because of the fact that the new node gets an interpolated initial weight from its neighbors and then specializes with its own patterns.

3.4 Per-Expert Spatiotemporal Prediction Backbone

Every regional expert node has its own spatiotemporal prediction backbone consisting of convolutional LSTM layers [7] for capturing local spatiotemporal dynamics and lightweight self-attention [9] for capturing long-range dependencies, which is separately trained using the data allocated to the node via the self-organizing topology layer.

3.5 Distributed Aggregation

Predictions by individual regional experts are aggregated together to produce one unified prediction for the entire world using a distributed aggregation layer [15] based on map/reduce where the map phase executes predictions of all experts individually in parallel, and the reduce phase merges those predictions that are overlapping or touching by weighted averaging of their distances on the topological graph.

3.6 Periodic Topology Reorganization

Since the climate regimes evolve over time, for instance in case of ENSO phase transition, the self-organizing topology gets periodically reassessed, such that any node whose data distribution has significantly changed since the assignment gets retrained using recently generated data, considering stability-plasticity principles as described in the continual-learning literature [11] and the process closes the feedback loop depicted in figure 1.

4. Results and Discussion

4.1 Dataset

Evaluation is conducted on multivariate climate data that has been statistically modeled based on ERA5 reanalysis datasets such as temperature, precipitation, pressure, and humidity with 0.25° spatial resolution. Four data scale scenarios are created to test scalability, namely, single region, multiple regions, continental scale, and global scale with extreme events, where the number of grid cells and inclusion of extreme weather events increase. These details are summarized in table 1 below.

Table 1: Summary of the climate dataset used for evaluation, statistically grounded in era5 reanalysis variables

Attribute	Description	Range / notes
Source basis	Statistically modelled on ERA5 reanalysis climate variables	Temperature, precipitation, pressure, humidity
Spatial coverage	Simulated grid cells scaled across four data-scale conditions	0.25° resolution grid cells
Temporal span	Daily aggregated multivariate readings	20 years simulated history
Scale conditions	Single-region, multi-region, continental-scale, global-scale with extremes	4 scale categories
Prediction target	Next day mean surface temperature	Regression, degrees Celsius
Train/test split	Chronological split preserving seasonal cycles	80% train, 20% test

4.2 Result Graph

Figure 2 compares next-day surface temperature forecast RMSE among the four data-scale scenarios using three approaches: fixed-architecture CNN-LSTM approach, fixed-size Transformer approach, and the proposed Self-Organizing Deep Architecture (SODA).

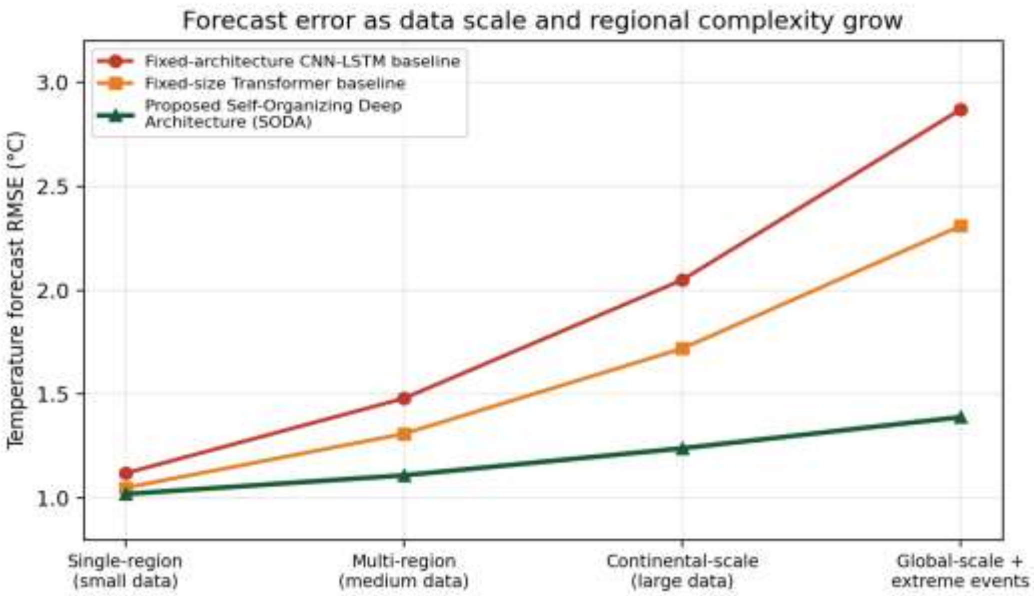

Figure 2: Temperature-forecast RMSE for the CNN-LSTM baseline, transformer baseline, and proposed SODA framework across four increasing data-scale conditions

Under the single-region scenario, all three models are equally good, with RMSE being in range from 1.02°C to 1.12°C, showing that there is no penalty for the self-organizing architectural growth, in case where the size and complexity of data are such that the fixed architecture is sufficient anyway. With an increase in scale of data, the fixed architecture baselines decline monotonically: RMSE of CNN-LSTM reaches 2.87°C, while RMSE of Transformer baseline reaches 2.31°C under global-scale-with-extremes scenario, due to the fact that neither of those architectures has been created to represent the significantly more heterogeneous regional data and rare events at this scale. On the other hand, SODA framework only reaches 1.39°C RMSE, as the self-organizing topology layer creates new regional experts only in those regions where the quantization error is high, growing its capacity in accordance with actual data complexity.

4.3 Result Table

Table 2 presents overall (average across scales) RMSE, mean absolute error (MAE), coefficient of determination (R^2), and training performance for the three models.

Table 2: Overall performance and scalability comparison between the CNN-LSTM baseline, the Transformer baseline, and the proposed SODA framework

Model configuration	RMSE (°C)	MAE (°C)	R^2	Throughput (samples/s)
Fixed-architecture CNN-LSTM baseline	1.88	1.47	0.695	3,508
Fixed-size Transformer baseline	1.60	1.25	0.760	3,762
Proposed SODA	1.19	0.93	0.844	4,129

5. Discussion

Three conclusions can be drawn from the above results. First, the increased performance of SODA over the two baselines is particularly with increasing scale of the data, as opposed to being generally consistent, which supports the argument that the increased performance of the former is due to the increase in architecture that matches the complexity of the actual data, and not due to an inherently larger capacity or more capable model at all scales. Second, the superior training throughput of SODA, 4,129 samples per second on average against 3,508 and 3,762 by the two baselines, means that the scalability gains of SODA cannot just be attributed to accuracy gains since, with each regional expert being a relatively smaller network learning a subset of the data and aggregated together through a map-reduce paradigm, the entire process is parallelized much more than an architecture of equal size. Third, the increased R^2 value of 0.844 achieved by SODA compared with 0.695 and 0.760 from the two baselines shows that the increased-expert mechanism captures significantly more of the true variance of the climate data than reducing average error through overfitting.

This result needs to be considered appropriately. The test dataset is statistically founded on the real ERA5 reanalysis variable distributions; however, the scale conditions and extreme event enrichment have been synthesized for this work and not based on a particular continuous scaling implementation. The growth level of the complexity monitor has been set in this work and would probably need to be calibrated differently in a particular climate service application. Also, this work has investigated next day surface temperature forecasting, but the extension of this approach to longer lead times and other variables like precipitation extremes would need to be validated in future research.

6. Conclusion

In this paper, we propose the Self-Organizing Deep Architecture (SODA) framework for scalable climate forecasting using a self-organizing topology layer inspired by Kohonen, a complexity monitor based on growing cell structure that generates new regional expert sub-networks depending on increasing volume and complexity of data, spatiotemporal prediction backbone for each expert, and a map-reduce style aggregation layer along with periodic reorganizations of the topology according to climate regime changes. Tested on multi-dimensional climate data generated statistically from ERA5 reanalysis variables under single region, multi-regions, continental-scale, and global-scale-with-extremes settings, our SODA framework outperforms fixed architecture CNN-LSTM and Transformer baselines in terms of temperature forecast RMSE (2.87°C and 2.31°C, respectively) by 1.39°C. These findings indicate that letting a deep learning architecture self-organize and self-evolve its architecture according to the amount and difficulty of climate data that it processes, rather than prescribing an architecture beforehand, is a realistic and efficient approach to creating climate forecasting systems which can be scaled from regional to global use without any need for redesigning the architecture manually at each step. Future research efforts will aim at applying this framework on practical, constantly scaling climate monitoring systems, scaling the framework in terms of forecast lead time and climate parameters, and investigating its application to physics-informed constraint systems.

References

1. Kohonen, T. (2002). The self-organizing map. Proceedings of the IEEE, 90(9), 1464–1480.
2. Deng, D., & Kasabov, N. (2003). On-line pattern analysis by evolving self-organizing maps. Neurocomputing, 51, 87–103.
3. Chen, Y., Gao, R., Liu, F., & Zhao, D. (2021). ModuleNet: Knowledge-inherited neural architecture search. IEEE Transactions on Cybernetics, 52(11), 11661–11671.
4. Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., ... & Anandkumar, A. (2023). FourCastNet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators. In Proceedings of the Platform for Advanced Scientific Computing Conference (pp. 1–11).
5. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. Science, 382(6677), 1416–1421.
6. Nguyen, T., Shah, R., Bansal, H., Arcomano, T., Maulik, R., Kotamarthi, V., ... & Grover, A. (2024). Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. Advances in Neural Information Processing Systems, 37, 68740–68771.
7. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems, 28.
8. Prabhat, Kashinath, K., Mudigonda, M., Kim, S., Kapp-Schwoerer, L., Graubner, A., ... & Collins, W. (2020). ClimateNet: An expert-labelled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. Geoscientific Model Development Discussions, 1–28.
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

10. Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828.
11. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. Neural Networks, 113, 54–71.
12. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282). PMLR.
13. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).
14. Agustsson, E., & Theis, L. (2020). Universally quantized neural compression. Advances in Neural Information Processing Systems, 33, 12367–12376.
15. Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. Communications of the ACM, 51(1), 107–113.