OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Predicting Extreme Weather: A Theoretical Semi-Supervised Machine Learning Framework

Aadhya Bagree
aadhya.bagree1663@gmail.com

ABSTRACT

Weather frequency and its destructive impacts have intensified over the years, increasing demand for predictive systems that deliver accurate, relevant, and cost-effective forecasts. However, accurate prediction remains challenging due to the complexity of atmospheric dynamics and limitations in existing global models. In recent years, advancements in machine learning and artificial intelligence have offered new opportunities to improve prediction accuracy. This paper presents a meta-review of extreme weather predictions that exist, spanning from physics-based modelling to machine learning to modern AI systems, drawing key studies from 1992- 2025, and identifying existing gaps. In response, two theoretical models have been proposed– a spatiotemporal gradient boosting with Gaussian-weighted downscaling and a feedforward neural network. Both have been contextualized against a publicly available multi-city hourly dataset. Together, the proposed frameworks are intended to investigate spatiotemporal learning, resolution-aware forecasting, and rare-event detection. Continued research and validation of these frameworks can lead to more accurate and deployable extreme weather prediction systems, in the end leading to improvement in public safety and disaster preparedness.

INTRODUCTION

In 2011 alone, the United States recorded at least 14 weather disasters that led to immediate damage of 1 billion dollars (about $3.1 per person in the US), affecting the economic conditions of the country. [1] Collectively, it claimed over 1000 lives, which was nearly double the yearly average. Worldwide, an insurance company called Munich Re reported that such extreme meteorological events have increased by nearly two times in the last 20 years.[1] These figures represent a global issue that remains prevalent, where economies suffer and lives are lost. Addressing these growing threats related to weather conditions requires scientific and computational knowledge spanning from datasets to artificial intelligence.

Prediction to this day has not been advanced enough to forecast extreme weather conditions early. With a rise in global temperatures, climate change has become a pressing issue. This creates unreliability, as the weather can dramatically change at any moment. The atmosphere is a chaotic system whose sensitivity to initial conditions can lead to the probability of a correct prediction decreasing. Even though some

machine learning tools have achieved a high accuracy in predictions, up to 97%, temperatures fluctuate daily, which causes models to fail to adapt, making them appear uncalibrated. [3]

The review examines how different approaches to predicting extreme weather conditions have evolved. Starting from physics-based numerical models to machine learning and artificial intelligence, I drew from five studies ranging from 1992 to 2025. The approaches that will be discussed will be interpreted against real-time historical weather data for over 36 countries over five years. [7] This review will evaluate these approaches and find the positives along with the gaps in data availability, model interpretability, and risks associated with this problem before introducing two models as part of a theoretical framework.

Aims and Goals

The primary aim of this theoretical study is to propose frameworks through which future research could investigate local extreme-weather prediction. The study considers the use of spatiotemporal downscaling connecting broad weather observations with city-level decision-making.

Investigating how new modelling applications could address differences in predictability between dynamic-onset and thermodynamic events is also important. Through this project, we aim to assess the extent to which distributed or futuristic computing technology could improve the processing power available for real-time Earth simulations.

It should enable us to examine how the integration of different machine-learning models could mitigate model bias and spurious correlations in pattern-recognition-based AI systems and also determine the extent to which semi-supervised machine-learning systems could improve the detection of rare events in geographical regions where event labels are limited.

The study proposes to investigate whether distributed computing could increase available processing capacity, whether combining machine-learning approaches could reduce model bias and spurious correlations, and whether semi-supervised learning could support rare-event detection where labelled examples are limited. These potential advantages have not been empirically tested in this paper.

Prediction Task

This section defines the intended application of the theoretical frameworks. It specifies what the models are designed to predict and proposes a method for defining event labels. It does not report on model implementation or experimental performance.

The proposed frameworks are intended to predict whether an extreme temperature or wind event will occur in a particular city within the following six hours. Each prediction would use the previous 24 hours of observations, which would include the most recent measurements of temperature, humidity, atmospheric pressure, wind speed, and direction. To present the geographical and seasonal differences, location and time information would also be incorporated.

The 24-input window is proposed to capture the recent conditions and daily variation, while the six-hour window defines a short-term prediction task. These are initial design choices rather than empirically optimized settings, and their suitability would require future evaluation. Both models will address the same prediction task.

The target variable would be a binary indicator that demonstrates whether one qualifying extreme event occurs during the prediction period. A value of 1 would indicate that it is an extreme temperature or wind event occurring within the following six hours. A value of 0 would indicate that none of the defined events occur during that period. Therefore, temperature, humidity, pressure, and wind measurements would serve as input variables. This formulation would distinguish between event and non-event periods. However, it would not separately predict the event's type, exact timing, and/or its intensity. Three categories are proposed: unusually elevated temperature, unusually low temperature, and unusually strong wind. A viable way to label the data could be to define elevated temperature and strong wind using city- and month- specific 99th- percentile thresholds and low temperature using the corresponding 1st-percentile threshold. These are just statistical definitions of the conditions and not actual hazard warning criteria. Under this proposed rule, the prediction period would be able to receive a positive label if at least one of the hourly measurements crossed a relevant threshold. A negative label would require sufficient measurements that no qualifying event has occurred. Otherwise, its label would remain unavailable.

These definitions make the theoretical framework compact and precise for future testing. They do not establish that either model can predict the defined events accurately or outperform existing approaches.

LITERATURE REVIEW

With increasing development of artificial intelligence and machine learning in the coming years, models are being used to predict extreme weather conditions. This has been developing over time with better machine learning, computer vision, large language models, and pipeline structures. Current literature available focuses on various different models such as backpropagation through time used for modelling dependencies in climate data, GCMs enhanced with machine learning for climate simulation, simple energy models for foundations of atmospheric modelling, and environmental intelligence with pipelined structure to help in the prediction of weather events that may occur.

Predicting and Managing Weather Events

Weather forecasting in the very beginning relied on observation networks and global climate models to predict atmospheric patterns. However, in recent years, these models, operating at a spatial resolution level, have become obsolete for practical and local-based decision-making, such as a flash flood that

could occur within a drainage basin. To combat these problems, the study has suggested using the "Environmental Intelligence" which combines both data from weather radars and satellites with numerical modelling and scientific assessment. This framework uses a model called nested modeling, where a high-resolution climate model fits the global level, allowing broader atmospheric context while the local models fit the local items. Statistical downscaling applies mathematical relationships between large-scale and smaller-scale conditions to help forecast weather to a better degree. However, a significant drawback to this methodology is that it is not engineered in a way to fit extreme weather conditions like intense snow and rain. Additionally, the source notes that there is limited available data for detecting changes in phenomena such as tornadoes and cyclones.[1]

Weather Prediction Using Futuristic Technologies: From Industry Perspective

Accurate prediction of weather data to this day remains a task due to the increased atmospheric vulnerability. However, in recent years, the use of futuristic technology, which includes artificial intelligence, machine learning, and high-level weather models, has opened a stream to improve prediction for researchers. This study by Kumari and Muthulakshmi analyses the role of futuristic technology in weather prediction. The authors analyze a range of futuristic technologies, such as Artificial Intelligence (AI) and Machine Learning (ML) for pattern recognition in satellite images, Quantum Computing for more complex simulations, Digital Twin technology for virtual Earth modeling, and the secure data-sharing option provided by Blockchain. The study highlights that the use of these technologies brings accuracy of prediction to 97%, especially in rainfall and temperature. It concluded that with the use of such technology combined with distributed paradigms such as Edge and Fog computing, this technology can revolutionize the industry from agriculture to disaster response. However, the research paper highlights that data is often limited in rural areas, which restricts the ability of the model to predict data accurately. Additionally, computational power is still inefficient in handling massive amounts of data and can lead to processing overheads. [3]

Explainable Machine Learning for Extreme Weather Event Prediction

A study by Tupayachi et al. (2026) proposed an explainable machine learning framework for predicting extreme weather events using a multi-tower data fusion strategy. The research compared Gradient Boosting models such as XGBoost and LightGBM with Deep Learning models including LSTM-FCN and Convolutional Neural Networks (CNNs). The study also integrated physics-based meteorological features and used SHAP (SHapley Additive exPlanations) analysis to improve interpretability of the prediction models. The results showed that gradient boosting models performed better for structured tabular meteorological data, while deep learning models struggled with rapidly changing high-wind events due to their stochastic and turbulent behavior. The study concluded that although physics-based features improved interpretability, the addition of high-dimensional features sometimes introduced noise that reduced the effectiveness of sequence-based deep learning architectures. However, the research does not effectively handle turbulent weather patterns in complex terrains and lacks efficient hybrid models that can reduce noise sensitivity while maintaining accurate temporal predictions.[4]

Introduction to Climate Modeling

A study by Schneider (1992) discussed diverse types of climate models, ranging from simple energy balance models to complex three-dimensional General Circulation Models (GCMs) based on physical laws such as thermodynamics. The study also explained empirical statistical methods that use historical climate data to identify weather patterns. The research highlighted that GCMs are computationally expensive and cannot accurately capture small-scale events like clouds or storms because of limited grid resolution. To overcome this, statistical approximations called parameterization methods are used, which introduce uncertainty in climate predictions, especially in cloud-feedback analysis. However, the study does not clearly address how small-scale atmospheric processes behave under climate change and faces limitations in providing accurate local weather predictions due to computational constraints and coarse model resolution. **[2]**

Prediction of Weather Forecasting by Using Machine Learning

A study by Sri Lakshmi et al. (2021) proposed the use of Back Propagation Neural Networks (BPNN) for predicting temperature and rainfall patterns. The research also explored Hidden Markov Models (HMM) for identifying weather patterns and k-means clustering for grouping and analyzing weather data. The study found that BPNN reduced prediction error rates for short-term forecasting, but accuracy decreased significantly for long-range predictions due to atmospheric turbulence and changing environmental conditions. The research also highlighted that errors in initial atmospheric measurements affected prediction reliability. However, the study lacks an effective mechanism for handling real-time dynamic weather updates, shows limited performance for long-term forecasting, and does not effectively integrate clustering and sequential learning techniques for improving prediction stability across diverse geographical regions. Additionally, the paper does not address scalability issues when handling large multi-dimensional weather datasets or methods for reducing noise caused by incomplete atmospheric observations. [5]

Gaps in Literature

There are many gaps in the literature that can be identified. Global climate models operate at a spatial scale that is too coarse for practical local decision-making, such as predicting a flash flood in a drainage basin. Current models available for weather prediction show a decline in the accuracy of dynamic, rapid-onset events in comparison to thermodynamic events. Processing overheads and power remain insufficient to run a high volume of data on multi-resolution and dimensional Earth simulations. Many AI-driven models focus on pattern recognition rather than the underlying causes and structure, which may lead to model biases. Limited availability of region-specific and geographically tagged datasets for rare and extreme weather events reduces the model's ability to capture local climatic patterns, terrain-based variations, and temporal weather dependencies accurately.

METHODS

Analysis of Dataset

The dataset employed in this study is the Historical Hourly Weather Data from Kaggle containing hourly meteorological observations. [7] This data is recorded in over 36 cities spanning from North America to Europe and the Middle East over five years from 2012 to 2017. The dataset contains two distinct types of information. The first is numerical variables, which include temperature, humidity, atmospheric pressure, wind speed, and wind direction. This is recorded in each city at every hourly timestamp. Another category includes weather description labels such as "sky is clear" or "light rain." In a future implementation, complete records could be divided into labelled and unlabeled subsets for evaluating the proposed semi-supervised framework. The usefulness of this dataset for this purpose would depend on the number of reliable target labels produced by the proposed labelling procedure.

The numerical time series and the weather-description field should not be confused with the target labels proposed in this paper. The existing weather descriptions are auxiliary categorical annotations; they are not the binary "extreme event/non-event" target defined in the Prediction Task section. In a future implementation, the binary target could be assigned to a labelled subset of suitable records using the proposed percentile rule. Other complete meteorological observations could then be used without exposing their target labels during the unsupervised or pretraining stage of a semi-supervised model.

The dataset may also contain gaps where one or more numerical measurements are not recorded for a particular city and hour. These are missing data cases. Missing measurements are different from unlabeled observations: an unlabeled observation has the meteorological inputs needed by the model, but its binary target label is not supplied to the learning stage. A record with missing required inputs should first be handled by preprocessing, for example by exclusion or a separately justified imputation method, and should not automatically be counted as unlabeled training data.

Throughout this paper, missing data and unlabeled data refer to different situations. Missing data means that one or more required meteorological measurements, such as temperature, pressure, humidity, or wind speed, are absent for a city and time. An unlabeled observation, by contrast, has the required meteorological measurements available but does not have the binary extreme-event target label assigned for the learning stage. Existing weather-description entries such as "sky is clear" or "light rain" are descriptive fields in the Kaggle dataset and are different from the binary extreme/non-extreme target labels defined for the proposed prediction task. Records with missing feature measurements would require preprocessing or exclusion before they could be used as unlabeled training observations.

To establish a thorough understanding of how the dataset can be used to model, a series of graphs are presented below. They help reveal the trends, structure, and distribution of the key meteorological variables in both proposed models. Moreover, this data helps surface important characteristics of weather and reveals the severe imbalance between normal and extreme weather conditions, which is discussed further below.

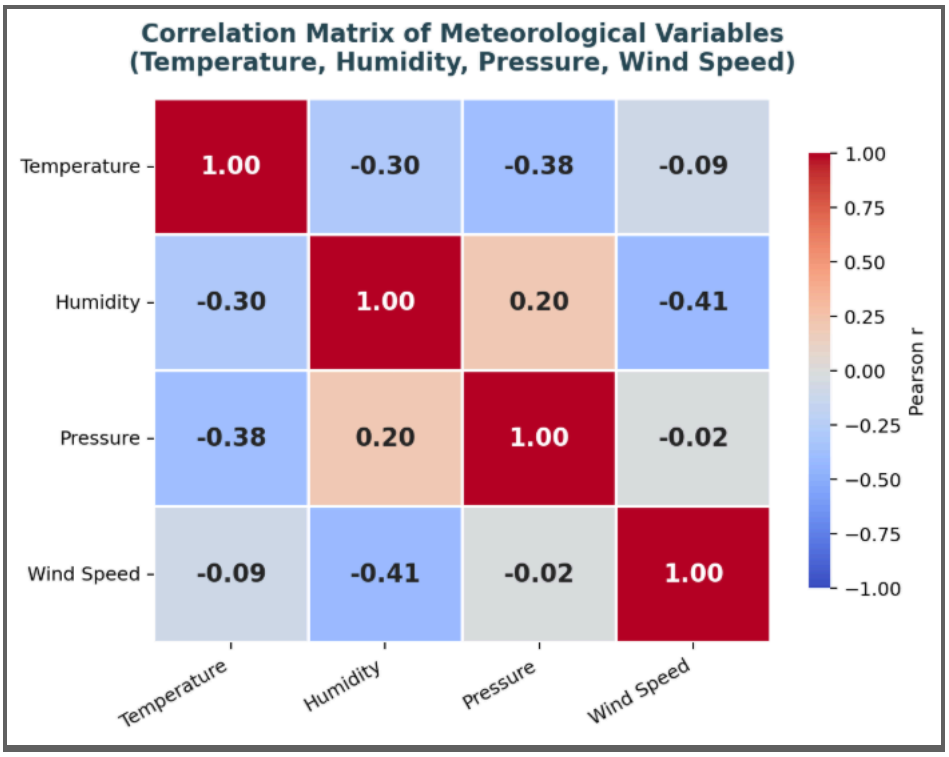


Figure 1: Correlation Matrix of Meteorological Variables

Figure 1 reveals a correlation matrix including the Pearson correlation coefficient to determine the relationships between many variables such as temperature, humidity, pressure, and wind speed. The most notable relationships are the moderate negative correlations between temperature and pressure and between humidity and wind speed. Temperature and humidity also have a moderately negative relationship reflecting the inverse link between heat and moisture retention. These correlations provide a descriptive overview of relationships among the variables or demonstrate non-linear relationships, requiring modelling and evaluation.

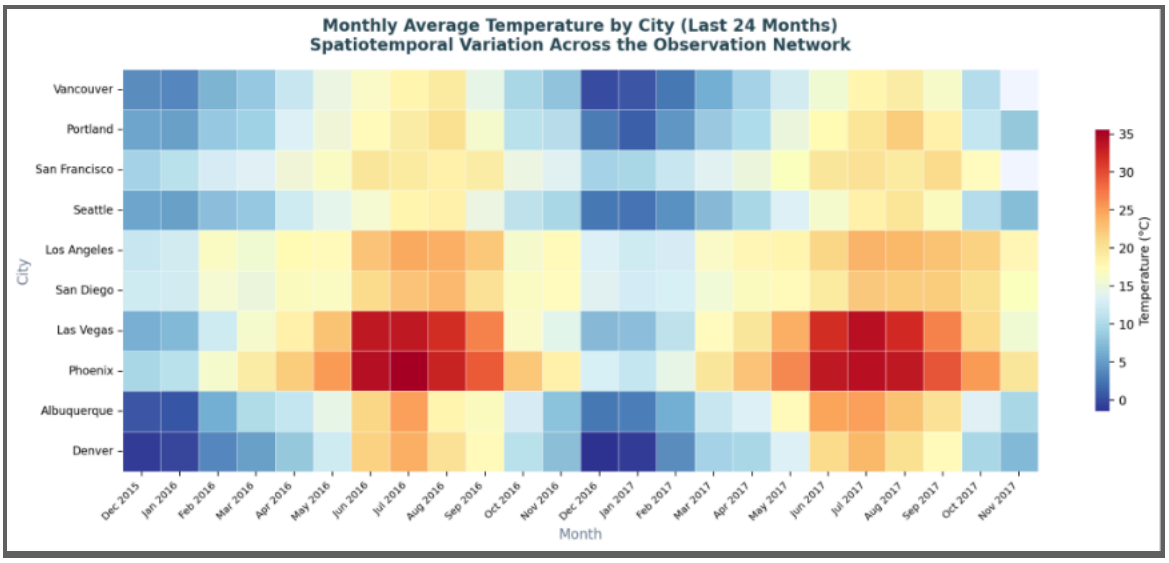


Figure 2: Average Temperature Monthly by City.

Figure 2 shows the monthly average temperature across 10 cities over 24 months. Cities that are south, such as Las Vegas and Phoenix, stay hot throughout the year, while cities like Denver and Vancouver, which are north, show seasonal swings with cold winters. Moreover, coastal cities like San Francisco remain stable throughout the year. These geographical and seasonal variations motivate the proposed investigation of spatiotemporal information. It does not establish that a spatiotemporal model would produce better predictions.

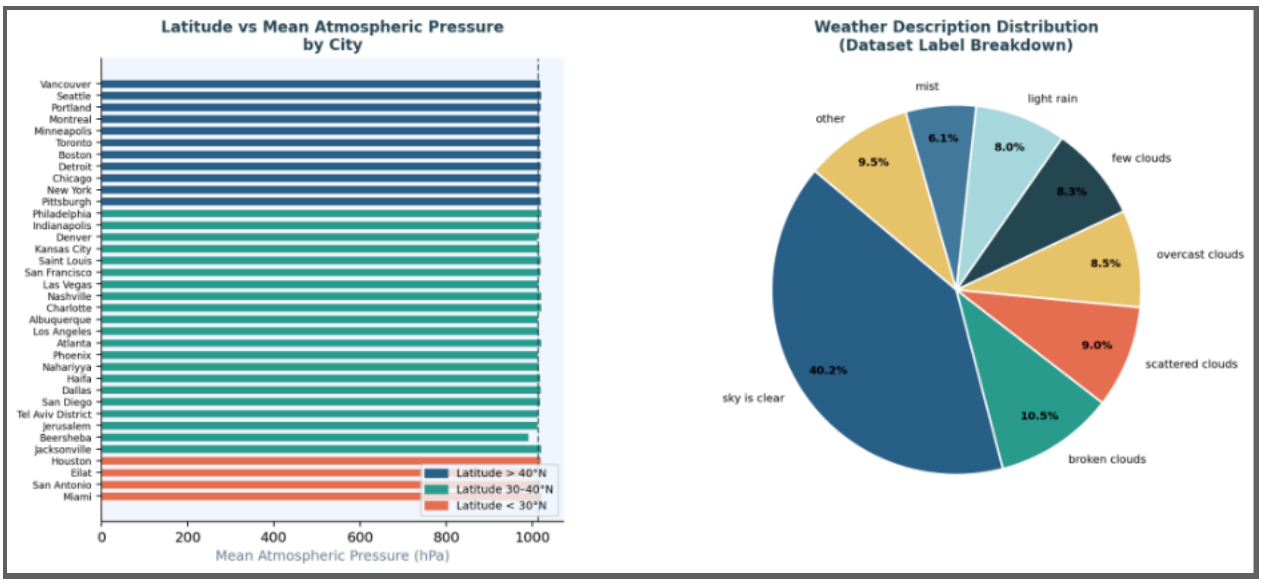


Figure 3: Latitude vs Mean Atmospheric Pressure by City & Weather Description Distribution

Figure 3 represents two views of the dataset structure. The bar chart shows mean atmospheric pressure per city, which is grouped by latitude. The graph shows that pressure alone is insufficient for prediction, as the data collected are clustered closely across all locations. The other view is a pie chart which shows that

40.2% of all observations are towards normal conditions while extreme events form a small minority, but they do not make use of the proposed binary extreme/non-extreme target. Therefore, the degree of target class imbalance can only be calculated after the proposed labelling procedure.

Introduction of Semi-Supervised Model

A fully supervised model would learn only from observations whose target labels are supplied. Because labelled extreme event examples can be limited, the proposed semi-supervised framework would use complete observations without supplied target labels during an initial learning stage, followed by a smaller labelled subset.

The hypothesis may help directly address the scarcity of data available and the different challenges identified throughout the literature review. Whether this improves extreme event detection would require comparison with fully supervised baselines. [8]

Deep learning has become increasingly applied to extreme weather forecasting, showing future promise in identifying and forecasting a range of atmospheric conditions [6]. However, existing deep learning approaches have worked in fully supervised learning or in large-scale climate data analysis; this leaves a gap in the application of semi-supervised learning, and the models proposed could be helpful but cannot be confirmed without any implementation.

For both the models described below, we describe how they will be trained using semi-supervised learning.

Model 1 — Spatiotemporal Gradient Boosting

Gradient boosting is a learning method that constructs a sequence of simple prediction models, building predictions iteratively based on errors and correcting them based on previous iterations. The final prediction is a combined result of these successive corrections. Thus, it can be effective on tabular data as well. [9]

However, in general, a gradient boosting model treats all training data with the same weight and relevance to give a prediction. When using weather data, this can negatively impact the prediction made by the model, particularly if observations closer to the target city and time are more accurate for the prediction. To address this, it uses a Gaussian weighting function, which will be applied across space and time. The Gaussian weighting function assigns each object a weight based on the geographic and temporal separation so that data from cities that are closest to the target city and from recent times contribute to a higher weight. Using the latitude and longitude available in datasets, the model will use that to set the spatial weights, and the timestamps available will give the temporal weights. [12] The weights are then applied to the gradient boosting system and ensure that the model incorporates spatial and temporal information when making a new prediction.

The significance of this model lies in the gap it addresses within the literature reviewed. Approaches that have already been analyzed include simple gradient boosting, and others have shown that without geographical data taken into account, forecast accuracy staggers.[1] [3] Therefore, to combat these problems, a spatiotemporal weighting allows for a network of distributed points. Existing global models cannot be adapted to local decision-making, and existing machine learning models often treat spatial relationships either uniformly or through discrete cut-offs. Instead of treating it like this, the proposed model offers a technique called smooth fade, where, like a flashlight, it shines brightest on the data from the city and the most recent hour. The light slowly fades as it looks at cities further away from the target or hours further in the past. The hypothesis is that this could make city-level predictions more locally relevant, although an improvement in predictive performance would require testing.

Analytical Flow of the Model

The model works in two phases to find and predict extreme weather events.

In the first, unsupervised stage, there are many unlabeled numerical observations such as temperature, humidity, pressure, and wind speed, which would be used to learn the spatiotemporal normal rhythm using Gaussian weights. This stage requires no labels and can be executed properly. The theoretical purpose is that it processes each condition through the weighting function and assigns higher relevance to data points that are closer to the target city and time step. This would allow the model to learn not just how normal conditions work in isolation but how they are related to detecting more extreme conditions. To do this, it must learn the basics of what the normal conditions are. Learning from the unlabeled data, the model may be able to build a baseline that could not be established with the number of labelled records. [12]

In the second stage, or the supervised stage, a smaller set of labelled weather data would be introduced to teach the gradient boosting systems to map these patterns onto specific outcomes, particularly the difference between extreme and normal weather conditions. Gradient boosting achieves this through a series of sequential data models. These act as a team of specialists. The first specialist makes a guess, then the second specialist corrects the mistakes the first specialist made. The third then corrects any mistakes made by the second specialist, and so on through the chain. The combined result of these corrections is proposed to be more accurate than any single model could produce. However, it is important to note that these are the expected advantages of the proposed model since it has not been experimentally tested yet.[13]

Extreme weather occurrence, however, occurs rarely, constituting only a small percentage of all observations across the globe. A standard model would be trained on imbalanced data that will predict normal conditions because its reward is statistically greater than a single extreme event. This is where the proposed model uses “class weights,” which are assigned to the model so that they are penalized more for failing to detect a single extreme weather event. This is telling the model that missing one day of a hurricane comes with a greater cost than not predicting 50 sunny days. The hypothesis is that this could

increase the model's sensitivity to rare events. Its effects on recall, precision, and false alarm rates would need to be evaluated experimentally.

Model 2: Feedforward Neural Network

This model employs a feedforward neural network in comparison to Model 1, which is a tree-based approach. As in the preceding model, the underlying mechanism is described before presenting the pipeline of the model. A feedforward neural network is a computational model that is made of successive layers of interlinked nodes. Through these nodes, information flows in a single direction from the input layer to the output layer. Each connection stores a numerical weight, and the interconnected nodes combine their weighted inputs and improve the model by applying a transformation before allowing the results to pass forward. During training, these weights are adjusted so that the network output matches the correct values. Due to this process of adjustment through many layers, the network learns complex details and relationships between the input layer and meteorological variables and makes a predicted outcome. In model 1, the use of spatial weighting allowed us to understand relationships better; however, here, it learns the relationship from a full set of input data. Whether this produces useful representations or accurate predictions would require empirical evaluation. [10][11]. Using a supervised network would be limited to learning from the small fraction of observations that carry useful, labelled data. The semi-supervised method is hypothesized to address this limitation by allowing the model to learn from a large body of unlabeled data before specializing on scarce labelled data. [8]

The gap this model addresses is significant. Previously, neural networks and related architectures have been applied to weather data and climate tasks, and a semi-supervised method combining labelled and unlabeled data has shown to be a step in correctly identifying extreme weather events. However, existing work focuses on large-scale fully supervised settings. This does not allow the application of a straightforward semi-supervised feedforward network to tabular city data considered here. Through this, the model will be able to assess the gap by adapting a standard neural network into a semi-supervised format suited to tabular data structures. [14][15]

Analytical Flow of the Model

In the first unsupervised phase, a large pool of unlabeled numerical observations, including temperature, humidity, atmospheric pressure, and wind speed, is used from the dataset to pretrain the early layers of the feedforward neural network. The network can reconstruct its own inputs with training. This means it learns to compress the input features, in this case the meteorological data, into a compact internal representation and then rebuilds it as accurately as possible. By this process, the model can understand what normal atmospheric conditions look like in a range of cities and time steps in the dataset, learning the structure and irregularities in weather without being told what any specific observation means. By building a general foundation from the abundant unlabeled data, the network can arrive at the second phase with a basic understanding of weather conditions, against which extreme conditions can be contrasted. [8]

In the second supervised phase, the smaller set of labelled data is introduced to fine-tune the model [8] from the first stage and compact the later layers of the network for the prediction task. The network may be able to differentiate between extreme and normal based on representations developed during the pre-training. Each connection between nodes carries a weight, and during this phase, these weights are changed to match the known labels. The depth of the layered structure is what would allow the network to capture the complex, nonlinear relationships between the input variables and the output of the weather outcomes. [10][11]

Unlike Model 1, this model does not apply any spatial weighting between cities, meaning it does not explicitly account for geographical distance when making predictions. This would allow Model 2 to become a meaningful point to compare Model 1 with. By holding the semi-supervised learning strategy constant and varying the presence of spatiotemporal weighting, the contribution of the Gaussian weighting system can be more isolated and evaluated.

How do the proposed models bridge the objectives?

In the case of Model 1, the following objectives highlight the novelty in bridging the existing gaps hypothetically.

- Referring to research objective 1, i.e., downscaling techniques bridging global and local decision-making, here, rather than relying on a global model to find patterns, it is hypothesized to sharpen predictions at the level of individual cities by drawing preferentially on the most spatially and temporally relevant observations.

- Referring to research objective 2, i.e., predictability between dynamic and thermodynamic events. Model 1's temporal weighting could be adjusted depending on whether the event is fast-changing, like wind, or if it is slow-changing, like temperature, which would help the model handle both types differently.

- Referring to research objective 4, i.e., mitigating model biases. Since Gradient boosting corrects its own mistakes step by step, it will be able to pick up false patterns in the data.

In the case of Model 2, the following objectives highlight the novelty in bridging the existing gaps hypothetically:

- Referring to research objective 3**,** i.e., processing power of real-time Earth simulation. Model 2, once trained, is likely to generate predictions through a forward pass, which is amazingly fast and has low processing power, making it suitable for real-time use.

- Referring to research objective 4, i.e., mitigating model biases. Model 2's pre-training on the unlabeled data would allow one to build a strong understanding before seeing labels, reducing the risk of learning false patterns from a biased labelled dataset.

- Referring to research objective 5, i.e., detection of weather events in geographical locations with less data available. As the model learned from unlabeled data before applying the same methodology to the labelled data, it could make meaningful predictions where such data may be scarce or missing.

DISCUSSION

The following table presents theoretically expected characteristics based on model architecture and prior literature. It does not report experimentally observed performance on the selected dataset.

Aspect	**Model 1 - Spatiotemporal Gradient Boosting**	**Model 2 - Feedforward Neural Network**
Core Approach	Will use an ensemble of decision trees that gradually correct errors made by previous trees.	Will use several layers of interconnected neurons to learn patterns from the data.
How it learns	Will learn through sequential boosting. Every new tree focuses on correcting mistakes from earlier iterations.	Will learn by adjusting weights between neurons using backpropagation and optimization algorithms.
Geographical information	Will directly make use of latitude and longitude through a Gaussian weighting mechanism. Nearby locations will be given more importance.	Location can be provided as an input feature; however, the model may not automatically prioritize nearby observations.
Temporal Information	Recent observations will be assigned greater weights, making them more influential in predictions.	Time will be treated as another variable unless more temporal processing is introduced.
Interpretability	Hypothesized to be easier to interpret and explain. Feature importance can also be examined.	It is more difficult to interpret because decision making is distributed across hidden layers.

Performance with limited labelled data	Expected to perform well even when only a small amount of labelled data is available.	It usually benefits from larger labelled datasets for effective learning.
Learning complex relationships	May be able to capture non-linear relationships through multiple tree splits and boosting stages.	May be effective at identifying extraordinarily complex and non-linear interactions between weather variables.
Computational Cost	May require lower training time and computational coordination.	May require more processing power and take longer to train.
Training Stability	Expected to be stable and less sensitive to parameter selection.	Performance may vary significantly depending on architecture, activation functions, and hyperparameters.
Handling Class Imbalance	Class weights can be assigned easily so that rare extreme events receive greater attention.	Class weights can also be used, but severe imbalance may still affect performance.
Noise and Outliers	May be robust to noisy weather observations and outliers.	May become sensitive to noise and may overfit if not regularized properly.
Feature Engineering	Requires more carefully designed input features.	Hidden layers can automatically learn useful feature interactions.
Weather data suitability	Particularly suitable for structured tabular weather datasets.	Could potentially be applicable to weather data too, although boosting methods often perform competitively on tabular datasets.
Prediction Behavior	May be able to produce stable and consistent forecasts.	May identify subtle relationships that may be missed by tree-based methods, although predictions may vary more.
Scalability	Intended to work efficiently on small and medium-sized datasets.	Could potentially be evaluated for exceptionally large datasets and high-dimensional inputs.

Overfitting Risk	May contain lower risk due to boosting regularization techniques.	May have higher risk if the network becomes too complex for the available data.
Semi-supervised learning stage	Use unlabeled data first to understand normal spatiotemporal weather behavior and then fine-tune on labelled events.	Learns underlying weather patterns from unlabeled data before associating them with labelled extreme weather classes.
Main Strength	Explicitly accounts for spatial and temporal proximity, which could make local forecasts more relevant.	Theoretical capacity for discovering hidden patterns and deeper relationships in atmospheric data is important.
Main Limitation	May not capture extremely deep feature interactions that neural networks can learn.	May require larger datasets, more tuning effort, and may lack built-in spatial awareness.
Best suited for	Regional or city-level forecasting where interpretability and geographical relevance will be important.	Will be useful in large-scale forecasting applications where learning extraordinarily complex patterns is the primary objective.

LIMITATIONS AND FUTURE WORK

The main limitation of this study is that both the proposed models remain theoretical. They have not been implemented, trained, or validated on the dataset selected. Therefore, no conclusion can be drawn about their accuracy, computational efficiency, and ability to detect rare events. The proposed 24-hour input window, six-hour prediction horizon, percentile thresholds, and model architecture are initial design choices rather than experimentally optimized settings.

The Historical Hourly Weather Dataset also imposed limitations. It only contains observations from 36 cities from 2012 to 2017 and may not accurately represent the rural regions, complex terrain, or the present-day climatic conditions. The proposed labelling defines statistically unusual temperatures and wind conditions rather than operational hazards. They do not include other extreme events such as floods, heavy rainfall, cyclones, or snowfall.

Further choices would need to be specified before implementation. These include the Gaussian spatial and temporal bandwidths, model hyperparameters, neural network architecture, and proportion of labelled observations. For Model 1, the development of the exact procedure through which unlabeled observations

would contribute to gradient-boosting training would be required. These could include pseudo-labelling or a separate representation learning stage. These choices can influence the behavior of both models.

Future work on this topic should focus on the implementation and validation of both proposed models. Extensions to datasets with more extreme events and more diverse weather datasets should be included. Integration of physics-informed constraints allows for a more promising future for the prediction of extreme weather conditions. Records should be divided chronologically for training, validation, and test sets rather than randomly because random division could allow observations from neighboring time periods to land in both the training and testing datasets.

The proposed model should be compared with several baselines. A city-and-month-specific baseline would predict according to the historical frequency of the extreme events, while a tenacity baseline would use the most recent weather conditions. Applying logistic regression would allow a simple statistical baseline. Model 1 should be compared with a standard gradient boosting, allowing for the contribution of spatial and temporal weighting to be examined. Model 2 should be compared with an otherwise identical supervised feedforward network.

Without the proposed pretraining stage. This would help assess whether either proposed component provides an advantage beyond simpler approaches.

The goal is not only to improve prediction and accuracy but to build systems and models that are accurate, efficient, require less processing power, and are accessible to people most vulnerable to extreme weather conditions.

CONCLUSION

Climate change affects the frequency and severity of extreme weather and requires a framework that is accurate, deployable, and equitable worldwide. This paper has reviewed the evolution of approaches to extreme weather prediction, from physics-based numerical modelling through early machine learning to modern AI systems. Through surveying the literature, there were multiple gaps that were identified in the already existing models.

To combat these gaps, two theoretical models have been proposed. The first proposed model is the spatiotemporal gradient boosting with a Gaussian-weighted downscaling model, which addresses the resolution and prediction challenges by using a spatiotemporal function. The semi-supervised feedforward neural network addresses rare event detection by building a solid foundation from unlabeled meteorological observations before specializing on labelled data. It is important to note that the above model is a purely theoretical vision and has not been empirically tested and validated.

ACKNOWLEDGMENTS

I would like to thank Daattavya Tav for his mentorship and guidance throughout this research. I would also like to thank Janani Hariharakhrishanan for her help with constructive feedback and careful reading. Sincere thanks are extended to my family for their continued support.

REFERENCES

1. Lubchenco, Jane, and Thomas R. Karl. "Predicting and managing extreme weather events." Physics today, 65.3 (2012): 31-37.
2. Trenberth, Kevin E., ed. Climate system modeling. Cambridge University Press, 1992.
3. Kumari, Shabnam, and P. Muthulakshmi. "Weather Prediction Using Futuristic Technologies: From Industry's Perspective." Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing. 2023.
4. Tupayachi, Jose, et al. "Explainable Machine Learning for Extreme Weather Event Prediction at the Oak Ridge Reservation." Available at SSRN 6409258 (2026).
5. Lakshmi, N. Sri, et al. "Prediction of weather forecasting by using machine learning." International Journal of Innovative Research in Computer Science & Technology (IJIRCST) 9 (2021).
6. Fang, Wei, et al. "Survey on the application of deep learning in extreme weather prediction." Atmosphere 12.6 (2021): 661.
7. Beniaguev, David. "Historical Hourly Weather Data 2012-2017." *Kaggle.com*, 2017, www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data/code.
8. Racah, Evan, et al. "Semi-supervised detection of extreme weather events in large climate datasets." (2016).
9. Safarov, Ruslan Z., et al. "Solving of classification problem in spatial analysis applying the technology of gradient boosting catboost." Folia Geographica 62.1 (2020): 112.
10. Haykin, Simon. Neural networks and learning machines, 3/E. Pearson education india, 2009.
11. Sharma, Pankaj, et al. "feedforward neural network: A Review." International Journal of Advanced Research in Engineering and Applied Sciences (IJAREAS) 2.10 (2013): 25-34.
12. "Spatiotemporal Analysis | Columbia Public Health." www.publichealth.columbia.edu/research/population-health-methods/spatiotemporal-analysis.
13. Tang, Rongnian, et al. "Numerical forecast correction of temperature and wind using a single-station single-time spatial LightGBM method." Sensors 22.1 (2021): 193.
14. Chattopadhyay, Ashesh, Pedram Hassanzadeh, and Saba Pasha. "Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data." Scientific reports 10.1 (2020): 1317.
15. Vitanza, Eleonora, Giovanna Maria Dimitri, and Chiara Mocenni. "A multi-modal machine learning approach to detect extreme rainfall events in Sicily." Scientific Reports 13.1 (2023): 6196.

ABOUT THE AUTHOR

Aadhya Bagree is a high school senior at Modern High School International, Kolkata. She takes an interest in statistical analysis and data science, hoping to build and explore models that could be helpful for society. She plans to study a STEM-related field in college, with a specific focus on data analysis and prediction.