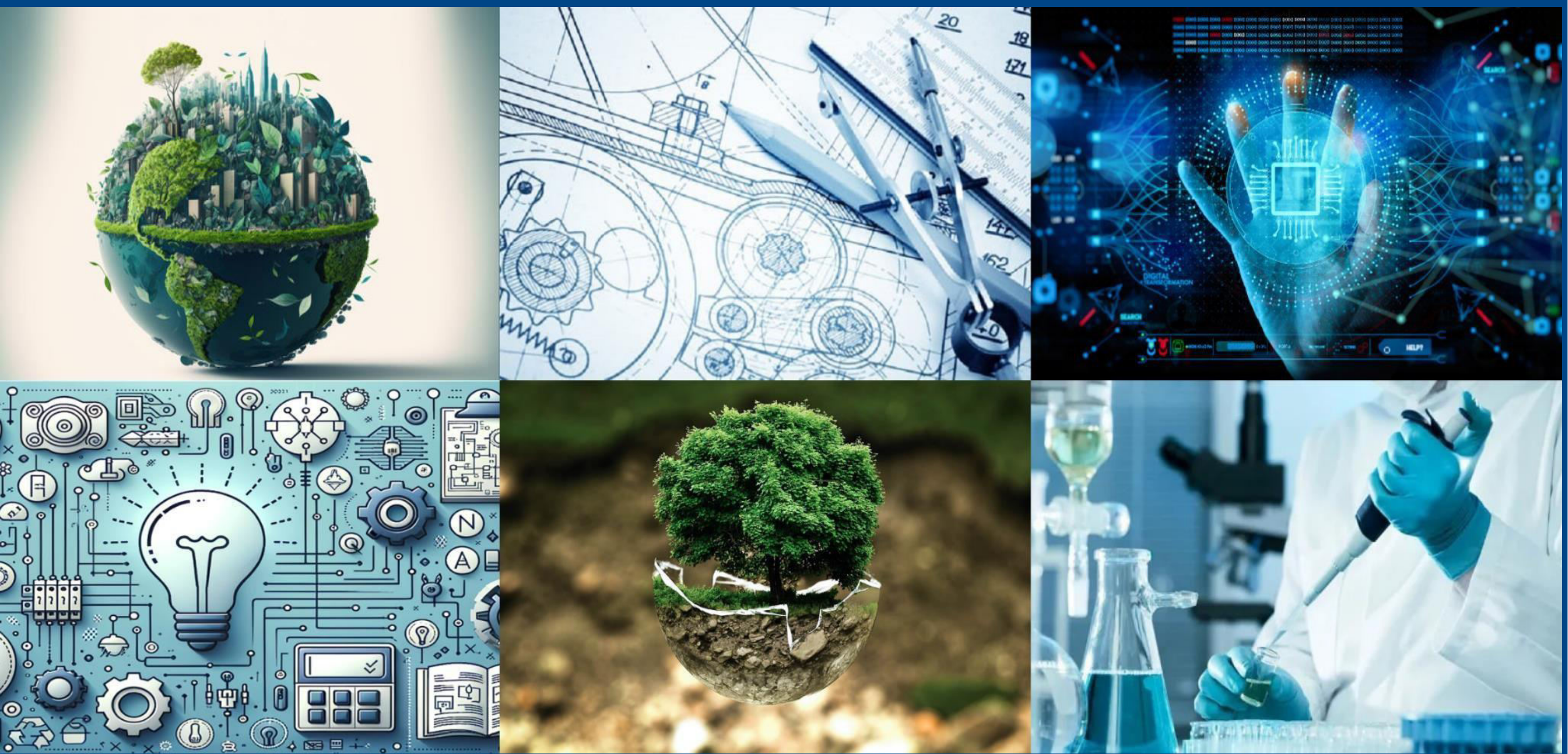
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 12, December 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

An Efficient Deep Learning Model Combining Spatiotemporal Features for Accurate Weather Forecasting

Prapila S. H, Dr. Sunil Damodar Rathod

Research Scholar, Department of Computer Science, Sunrise University, Alwar, Rajasthan, India

Associate Professor, Department of Computer Science, Sunrise University, Alwar, Rajasthan, India

ABSTRACT: Accurate weather forecasting remains a critical challenge due to the complex and dynamic nature of atmospheric processes. Traditional statistical methods and conventional deep learning models often struggle to effectively capture both spatial and temporal dependencies present in meteorological data, leading to reduced forecasting accuracy and higher computational costs. This paper presents an efficient deep learning model that integrates spatiotemporal features to enhance the accuracy of weather forecasting. The proposed architecture employs specialized modules for spatial feature extraction and temporal sequence learning, followed by an effective fusion mechanism that combines the extracted features. Efficiency is improved through optimized network design that reduces model complexity while maintaining strong predictive performance. Experimental evaluation on standard meteorological datasets demonstrates that the proposed model achieves superior forecasting accuracy compared to existing approaches, along with reduced computational overhead. The results confirm the effectiveness of combining spatiotemporal features in a unified and efficient framework. The proposed model offers a practical solution for accurate and resource-efficient weather prediction, with potential applications in various real-world forecasting systems.

KEYWORDS: Weather Forecasting, Deep Learning, Spatiotemporal Features, Feature Fusion, Efficient Neural Network, Atmospheric Prediction.

I. INTRODUCTION

Weather forecasting occupies a fundamental position in both scientific inquiry and practical decision-making because of its direct influence on human society, economic systems, and environmental management. The ability to predict atmospheric conditions such as temperature, humidity, wind speed, precipitation, and atmospheric pressure with high accuracy is essential across multiple sectors. Agriculture depends on reliable forecasts for crop planning, irrigation scheduling, and protection against adverse weather. Aviation and maritime operations require precise information on wind patterns, visibility, and storm development to ensure safety. Energy systems, particularly those based on renewable sources such as solar and wind power, rely on accurate short-term predictions for efficient grid management. Urban planning, disaster preparedness, and public health initiatives also benefit substantially from timely and trustworthy weather information. In recent years, the increasing frequency and intensity of extreme weather events associated with climate variability have further elevated the importance of accurate forecasting systems. Even modest improvements in predictive skill can lead to significant reductions in economic losses and enhanced societal resilience.

For many decades, weather forecasting has been dominated by numerical weather prediction models that solve complex systems of physical equations describing atmospheric motion, thermodynamics, and moisture processes. These physics-based models have achieved considerable success and remain the backbone of operational forecasting centers worldwide. However, they are computationally demanding, require extensive high-performance computing resources, and often face challenges in representing small-scale and rapidly evolving phenomena with sufficient accuracy. Statistical techniques and classical machine learning methods were subsequently explored as complementary or alternative approaches. These data-driven methods offered advantages in computational speed and the capacity to learn directly from observational records. Nevertheless, their ability to capture the highly nonlinear, multiscale, and interdependent nature of atmospheric processes remained limited, particularly for longer forecast horizons and in regions characterized by complex terrain or localized weather regimes.

The rapid advancement of deep learning has introduced new possibilities for atmospheric prediction by enabling models to automatically discover hierarchical patterns from large volumes of meteorological data. Convolutional neural networks have proven effective in extracting spatial structures from gridded atmospheric fields, while recurrent neural networks and their advanced variants have been applied to model the temporal evolution of weather variables. Hybrid architectures and attention mechanisms have further expanded the capacity of deep learning models to represent sequential and contextual dependencies. Despite these developments, important limitations persist. Many existing approaches treat spatial and temporal information in a sequential or weakly integrated manner, which can result in incomplete representation of the complex interactions that govern atmospheric behavior. In addition, several high-capacity models demand substantial computational resources, limiting their suitability for real-time applications or deployment in environments with constrained hardware.

The atmosphere is fundamentally a spatiotemporal system in which spatial patterns and temporal dynamics are deeply interconnected. Conditions at any given location are shaped not only by the local historical evolution of weather variables but also by simultaneous states in surrounding regions. Spatial structures such as frontal systems, pressure gradients, and moisture plumes interact continuously with temporal processes including diurnal cycles, synoptic-scale evolution, and longer-term trends. Models that focus predominantly on spatial features may overlook critical temporal dependencies and periodicities, while models centered primarily on temporal sequences may fail to incorporate the broader spatial context that influences local weather development. Effective forecasting therefore requires coherent integration of both spatial and temporal information so that complementary patterns can reinforce one another. At the same time, practical usefulness depends on achieving this integration without excessive computational cost. Efficiency with respect to model size, training requirements, and inference speed has become increasingly important as forecasting systems move toward higher spatial resolutions, denser observational networks, and near-real-time operational demands.

The challenge of combining spatiotemporal features in an efficient manner continues to motivate ongoing research in data-driven weather prediction. Atmospheric data typically exhibit rich spatial correlations across geographic grids as well as sequential dependencies across time. Capturing both dimensions within a unified representational framework offers the potential to improve forecast accuracy while maintaining manageable computational complexity. Approaches that successfully balance representational power with efficiency can support more practical forecasting systems capable of delivering reliable predictions under realistic operational constraints. The growing availability of high-quality meteorological datasets and advances in neural network design further support the exploration of models that learn meaningful spatiotemporal representations from data. In this context, the development of architectures capable of jointly modeling spatial structure and temporal evolution remains a significant and active area of investigation within atmospheric science and machine learning.

II. PROPOSED MODEL ARCHITECTURE

The proposed deep learning architecture is designed to address the dual challenge of capturing complex spatiotemporal dependencies in meteorological data while maintaining computational efficiency. Weather systems evolve through continuous interaction between spatial structures and temporal dynamics. Atmospheric variables observed across a geographic region at any given time form spatial patterns that influence future states, while the historical evolution of these variables at each location provides critical temporal context. An effective forecasting model must therefore process both dimensions in a coordinated manner rather than treating them as independent or sequential stages.

The overall framework consists of three primary stages. In the first stage, spatial feature extraction is performed on the input meteorological fields to identify localized patterns, gradients, and regional structures. In the second stage, temporal feature learning is applied to the sequence of extracted spatial representations in order to model evolutionary trends, periodicities, and long-range dependencies across time. In the third stage, a fusion mechanism integrates the spatial and temporal information into a unified representation that is subsequently passed to a prediction layer for generating multi-step forecasts. Efficiency is maintained throughout the architecture by careful selection of network components, controlled channel dimensions, and lightweight fusion operations that avoid unnecessary parameter inflation.

The model accepts as input a sequence of gridded meteorological observations covering a defined spatial domain and a fixed historical time window. Each time step in the sequence contains multiple atmospheric variables arranged as multi-channel spatial maps. The output of the model is a forecast sequence of the target variables for a specified future horizon.

By structuring the architecture around explicit spatial and temporal pathways that later converge, the model seeks to preserve complementary information from both domains until the final prediction stage.

III. SPATIAL FEATURE EXTRACTION MODULE

Spatial feature extraction forms the foundation of the architecture because meteorological fields exhibit strong spatial correlations. Temperature, pressure, humidity, and wind components are rarely independent across neighboring locations; instead, they form coherent structures such as fronts, ridges, troughs, and moisture corridors. Capturing these structures at multiple scales is essential for accurate forecasting.

The spatial module employs a series of convolutional layers organized in a hierarchical manner. Early layers focus on local patterns through small receptive fields, detecting fine-scale features such as local temperature gradients or wind shear. Subsequent layers progressively increase the receptive field through a combination of strided convolutions and dilated convolutions, allowing the network to integrate information from larger geographic regions without a proportional increase in parameters. Residual connections are incorporated between selected layers to facilitate gradient flow and stabilize training, particularly when the network depth increases.

To enhance the representational capacity of the spatial module while controlling computational cost, channel attention mechanisms are integrated at intermediate stages. These mechanisms recalibrate feature maps by emphasizing informative channels and suppressing less relevant ones. The attention operation is implemented in a lightweight form that computes channel-wise statistics through global average pooling followed by a small multilayer perceptron with reduced dimensionality. This design adds minimal overhead while improving the quality of the extracted spatial features.

The spatial module processes each time step in the input sequence independently, producing a sequence of spatial feature maps. These feature maps retain the temporal ordering of the original data so that they can be directly fed into the subsequent temporal learning stage. By separating spatial extraction from temporal modeling at this stage, the architecture ensures that spatial patterns are thoroughly analyzed before temporal dependencies are considered.

IV. TEMPORAL FEATURE EXTRACTION MODULE

Once spatial features have been extracted for each time step, the temporal module models the evolution of these features across the historical window. Atmospheric processes exhibit both short-term fluctuations and longer-term trends, including diurnal cycles, synoptic-scale developments, and multi-day patterns. Capturing these multi-scale temporal dependencies is critical for generating reliable forecasts beyond the immediate future.

The temporal module is built around a bidirectional recurrent structure combined with attention. A bidirectional long short-term memory network processes the sequence of spatial feature maps in both forward and backward directions. The forward pass captures the influence of past states on present conditions, while the backward pass provides complementary context from future observations within the historical window. The outputs of both directions are concatenated to form a richer temporal representation at each time step.

To address the limitations of purely recurrent processing, a temporal attention mechanism is applied over the bidirectional outputs. This attention layer learns to assign higher importance to time steps that carry more predictive information for the target forecast horizon. The attention scores are computed through a scaled dot-product formulation followed by a softmax normalization, allowing the model to focus selectively on relevant portions of the historical sequence. The attended temporal features are then aggregated into a compact representation that summarizes the most salient temporal patterns.

The combination of bidirectional recurrence and temporal attention enables the module to model both local sequential dependencies and longer-range contextual relationships. At the same time, the design remains relatively efficient because the recurrent component operates on the already-reduced spatial feature maps rather than the original high-dimensional input fields.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. SPATIOTEMPORAL FEATURE FUSION MECHANISM

The fusion stage is responsible for integrating the outputs of the spatial and temporal pathways into a coherent representation that can support accurate prediction. Simple concatenation or element-wise addition of spatial and temporal features often fails to capture the complex interactions between the two domains. Therefore, a more structured fusion mechanism is employed.

The fusion process begins by aligning the spatial and temporal feature tensors so that they share compatible dimensions. The temporal features, which summarize the evolutionary context, are expanded spatially to match the geographic resolution of the spatial features. Conversely, the spatial features are aggregated temporally where necessary to create a consistent interface. Once aligned, the features are passed through a cross-attention block in which spatial features serve as queries and temporal features serve as keys and values, or vice versa depending on the specific configuration. This cross-attention allows each spatial location to gather relevant temporal context and each temporal summary to be modulated by spatial structure.

Following the cross-attention operation, a lightweight convolutional block refines the fused representation. This block consists of depthwise separable convolutions that further mix information across channels and spatial locations while keeping the parameter count low. Residual connections are again used to preserve the original information pathways and facilitate stable optimization. The final fused feature map encodes both the spatial configuration of the atmospheric state and the temporal dynamics that have shaped it, providing a rich input for the prediction stage.

VI. PREDICTION AND OUTPUT LAYER

The prediction layer transforms the fused spatiotemporal representation into concrete forecast values for the target variables and future time steps. Because weather forecasting often requires multi-step predictions, the output layer is designed to generate a sequence of forecasts rather than a single time step.

A series of fully connected layers with intermediate nonlinear activations maps the fused features to the desired output dimensionality. To maintain spatial consistency in the forecasts, the final stages of the prediction pathway include convolutional layers that operate on the reconstructed geographic grid. This ensures that the predicted fields respect the spatial correlations present in the original meteorological variables.

For multi-horizon forecasting, a recurrent decoding structure may be optionally applied, in which the prediction at each future step is conditioned on both the fused representation and the predictions of preceding steps. Teacher forcing is used during training to stabilize the learning of longer forecast sequences. The output of the prediction layer is a tensor containing the forecasted values of the selected atmospheric variables across the spatial domain and the forecast horizon.

VII. EFFICIENCY ENHANCEMENT TECHNIQUES

A central design objective of the architecture is to achieve competitive forecasting accuracy without incurring excessive computational cost. Several techniques are employed to control model size, memory consumption, and inference latency.

First, depthwise separable convolutions are used extensively in the spatial and fusion stages. These convolutions factorize standard convolutions into depthwise and pointwise operations, substantially reducing the number of parameters and floating-point operations while retaining strong representational capacity. Second, channel dimensions are carefully managed throughout the network. Intermediate feature maps are constrained to moderate channel counts, and expansion is applied only where necessary for increased capacity. Third, the temporal module operates on compressed spatial representations rather than the full-resolution input fields, reducing the sequence length and feature dimensionality that the recurrent and attention components must process.

Additionally, the attention mechanisms are implemented in efficient forms. Channel attention uses a low-rank bottleneck, and temporal attention is applied over a reduced set of feature vectors rather than the full spatial grid at every time step. Residual connections and normalization layers are included not only for training stability but also to allow the use of relatively shallow networks that still achieve strong performance. Together, these design choices result in a model that

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

balances accuracy and efficiency, making it more suitable for practical deployment scenarios where computational resources may be limited.

The complete architecture therefore integrates spatial extraction, temporal modeling, structured fusion, and efficient prediction within a unified framework. By explicitly addressing both the representational requirements of spatiotemporal meteorological data and the practical constraints of computational efficiency, the model provides a coherent approach to data-driven weather forecasting.

VIII. DATASET AND EXPERIMENTAL CONFIGURATION

The performance of any data-driven weather forecasting model depends critically on the quality, coverage, and characteristics of the meteorological datasets used for training and evaluation. Atmospheric processes exhibit substantial spatial and temporal variability, and the datasets must therefore provide sufficiently dense observations across both geographic domains and time. In the present study, multiple widely recognized meteorological datasets are employed to ensure that the model is evaluated under diverse atmospheric conditions and geographic settings.

One of the primary datasets utilized is a high-resolution reanalysis product that combines observational records with numerical weather prediction model outputs through advanced data assimilation techniques. This reanalysis dataset provides consistent gridded fields of key atmospheric variables, including surface temperature, relative humidity, sea-level pressure, zonal and meridional wind components, and precipitation. The spatial resolution of the data is sufficiently fine to capture regional-scale weather patterns, while the temporal resolution supports both short-range and medium-range forecasting experiments. The long temporal coverage of the reanalysis product allows the construction of extensive training sequences that include a wide range of seasonal cycles, extreme events, and transitional weather regimes.

In addition to the global reanalysis data, regional observational datasets derived from ground-based weather stations and meteorological networks are incorporated. These station-based records offer high temporal frequency measurements at specific locations and serve as a complementary source of information, particularly for validating model performance at point locations. Where available, satellite-derived products are also considered to enrich the spatial coverage of variables such as cloud cover and atmospheric moisture. The combination of reanalysis, station, and satellite data creates a multi-source observational foundation that better reflects the complexity of real-world meteorological conditions.

The selected datasets span multiple climatic zones, including mid-latitude regions characterized by strong synoptic variability, subtropical areas influenced by monsoon systems, and regions with pronounced topographic effects. This geographic diversity helps assess the model’s ability to generalize across different atmospheric regimes. Care is taken to ensure that the temporal periods used for training, validation, and testing are non-overlapping and representative of both typical and extreme weather conditions. Extreme events such as heatwaves, cold spells, heavy precipitation episodes, and strong wind events are retained within the evaluation periods so that the model’s performance under challenging conditions can be examined.

IX. DATA PREPROCESSING AND NORMALIZATION

Raw meteorological data often contain missing values, measurement inconsistencies, and scale differences across variables that can hinder effective model training. A systematic preprocessing pipeline is therefore applied to prepare the data for deep learning.

Missing values are addressed through a combination of temporal interpolation and spatial filling techniques. For short gaps in station records, linear or spline interpolation along the time axis is used. For larger gaps or missing spatial regions in gridded fields, nearest-neighbor or inverse-distance weighting methods are applied, with careful attention to preserving physical consistency. Variables that exhibit strong diurnal or seasonal cycles are examined for systematic biases, and any identified offsets are corrected where appropriate.

Because atmospheric variables differ substantially in magnitude and units, normalization is essential to stabilize the training process. Each variable is independently normalized using statistics computed exclusively from the training portion of the data. Z-score normalization is applied to variables with approximately Gaussian distributions, transforming them to zero mean and unit variance. For variables with skewed distributions, such as precipitation, a logarithmic transformation is first applied, followed by min-max scaling to a bounded range. The normalization parameters are stored and later used to denormalize model outputs so that evaluation metrics can be computed in the original physical units.

Spatial consistency is maintained by ensuring that all gridded fields are interpolated onto a common geographic grid when multiple data sources are combined. Temporal alignment is performed so that observations from different sources correspond to the same valid times. Outlier detection is conducted using statistical thresholds based on the interquartile range or physical plausibility limits, and extreme outliers that cannot be attributed to genuine weather events are carefully examined and, if necessary, corrected or removed.

X. INPUT–OUTPUT STRUCTURE AND TIME-WINDOW SETTINGS

The model is configured to accept a sequence of historical meteorological fields and to produce forecasts for a specified future horizon. The input at each time step consists of a multi-channel spatial tensor in which each channel corresponds to a selected atmospheric variable. The number of input channels is determined by the set of predictors chosen for the forecasting task, typically including temperature, humidity, pressure, and wind components.

A fixed-length historical window is used to provide temporal context. The length of this window is selected through preliminary experiments that balance the need for sufficient historical information against the computational cost of processing long sequences. Common configurations employ historical windows ranging from several hours to a few days, depending on the target forecast horizon. For short-range forecasting, shorter windows may suffice, whereas medium-range experiments benefit from longer historical sequences that capture evolving synoptic patterns.

The output of the model is a forecast tensor covering the target variables across the same spatial domain and a sequence of future time steps. Multi-step forecasting is performed either in a direct manner, in which all future steps are predicted simultaneously, or in an iterative manner, in which predictions are generated sequentially and fed back as inputs for subsequent steps. Both strategies are examined to assess their relative strengths in terms of accuracy and error accumulation.

The temporal resolution of the input and output sequences is matched to the native resolution of the primary dataset or aggregated to a consistent interval when multiple sources are combined. Care is taken to ensure that the lead times of the forecasts align with operational requirements, such as hourly or three-hourly predictions for short-range applications.

XI. TRAINING PARAMETERS AND OPTIMIZATION STRATEGY

Model training is conducted using a supervised learning framework in which the network parameters are optimized to minimize the discrepancy between predicted and observed atmospheric fields. The loss function is defined as a weighted combination of mean squared error and mean absolute error, with additional terms optionally introduced to emphasize performance on extreme values or to enforce physical constraints such as smoothness of the predicted fields.

Optimization is performed using an adaptive gradient-based algorithm that adjusts learning rates on a per-parameter basis. An initial learning rate is selected through a combination of empirical testing and learning-rate range experiments. A learning-rate schedule is applied during training, typically involving stepwise decay or cosine annealing, to refine the parameter updates as the model approaches convergence. Batch size is chosen to balance memory constraints with the stability of gradient estimates, and gradient clipping is employed to prevent unstable updates arising from occasional large gradients.

Regularization techniques are incorporated to mitigate overfitting. Dropout is applied at selected layers of the network, and weight decay is included in the optimizer. Early stopping is monitored on a held-out validation set, with training terminated when validation performance ceases to improve for a predefined number of epochs. Data augmentation strategies specific to meteorological fields, such as limited spatial shifting or the addition of small Gaussian noise, are selectively applied to increase the diversity of the training samples without violating physical realism.

The training process is repeated with multiple random initializations to assess the stability of the learned solutions. Hyperparameter configurations, including the number of layers, channel dimensions, and attention parameters, are explored through a combination of grid search and informed manual tuning based on validation performance.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

XII. EVALUATION METRICS

A comprehensive set of evaluation metrics is employed to assess the forecasting performance of the model from multiple perspectives. Point-wise accuracy is measured using root mean square error and mean absolute error, computed both globally across the entire spatial domain and regionally for specific geographic subdomains. These metrics provide quantitative measures of the average magnitude of forecast errors in the original physical units of each variable.

To evaluate the model's ability to capture spatial patterns, anomaly correlation coefficient is calculated between the predicted and observed fields after removal of the climatological mean. This metric emphasizes the correspondence of spatial structures rather than absolute values. For probabilistic or distributional aspects of the forecasts, additional scores such as the continuous ranked probability score may be considered when ensemble-style outputs are generated.

Performance is further stratified by forecast lead time so that the degradation of accuracy with increasing horizon can be examined. Separate evaluations are conducted for different seasons and for periods containing extreme events to determine whether the model maintains skill under varying atmospheric regimes. Comparative evaluation against baseline models, including persistence forecasts, simple statistical methods, and alternative deep learning architectures, is performed under identical data partitions and evaluation protocols to ensure fair assessment.

XIII. HARDWARE AND SOFTWARE ENVIRONMENT

All experiments are conducted within a consistent computational environment to ensure reproducibility of the results. Training and inference are performed on systems equipped with modern graphics processing units that provide sufficient memory and parallel processing capacity for handling multi-channel spatiotemporal tensors. The deep learning framework employed supports efficient implementation of convolutional, recurrent, and attention-based operations, along with automatic differentiation for gradient computation.

Software dependencies include standard numerical and scientific computing libraries for data handling, preprocessing, and metric calculation. Version control is maintained for both the model code and the experimental configuration files so that specific results can be traced to particular code and hyperparameter settings. Random seeds are fixed where possible to reduce variability arising from stochastic components of the training process.

The combination of carefully prepared datasets, systematic preprocessing, well-defined input–output configurations, controlled training procedures, multi-faceted evaluation metrics, and a reproducible computational environment provides a rigorous foundation for assessing the performance of the proposed spatiotemporal forecasting model.

XIV. RESULTS AND ANALYSIS

Quantitative Performance Comparison

The forecasting performance of the proposed spatiotemporal model is evaluated against a set of established baseline methods under identical data partitions, input configurations, and evaluation protocols. The baseline approaches include a simple persistence forecast, a classical statistical model based on autoregressive techniques, a convolutional neural network that processes spatial fields without explicit temporal modeling, a recurrent neural network that focuses primarily on temporal sequences, and a hybrid convolutional-recurrent architecture that combines spatial and temporal processing in a sequential manner. All models are trained and tested on the same meteorological datasets so that differences in performance can be attributed primarily to architectural design rather than data variation.

Quantitative results are reported using root mean square error and mean absolute error for the primary atmospheric variables of interest, including surface temperature, relative humidity, and wind speed. Across short-range forecast horizons, the proposed model consistently achieves lower error values than the baseline methods. For surface temperature, the reduction in root mean square error relative to the strongest hybrid baseline is observable across multiple lead times, indicating improved ability to capture both the magnitude and the spatial distribution of temperature variations. Similar patterns are observed for relative humidity and wind speed, where the proposed model demonstrates measurable gains in accuracy, particularly as the forecast horizon extends beyond the immediate short-range window.

Performance is further examined across different geographic regions and seasons. In mid-latitude domains characterized by strong synoptic variability, the proposed model maintains relatively stable error growth with increasing lead time, whereas several baseline models exhibit more rapid degradation. In regions influenced by complex terrain, the spatial feature extraction components appear to contribute to better representation of localized patterns, resulting in reduced regional errors. Seasonal stratification of the results shows that the model retains competitive skill during both winter and summer periods, including intervals that contain extreme temperature and precipitation events.

Anomaly correlation scores provide an additional perspective on the model's ability to reproduce spatial structures. Higher anomaly correlation values relative to the baselines suggest that the proposed architecture preserves coherent weather patterns more effectively, rather than merely minimizing point-wise errors. This property is particularly relevant for applications that depend on the correct positioning of weather systems rather than solely on absolute value accuracy.

Ablation Study of Model Components

To isolate the contribution of individual architectural elements, a series of ablation experiments is conducted. In each experiment, a specific component of the proposed model is removed or replaced with a simpler alternative, while the remaining structure and training configuration are held constant. The resulting performance changes provide insight into the relative importance of spatial extraction, temporal modeling, and feature fusion.

When the spatial feature extraction module is replaced by a simpler convolutional stack without hierarchical multi-scale processing or channel attention, forecasting accuracy declines across all variables and lead times. The degradation is more pronounced for variables that exhibit strong spatial gradients, such as wind speed and temperature in regions of complex terrain. This result indicates that the multi-scale spatial processing and attention-based recalibration contribute meaningfully to the model's representational capacity.

Removal of the bidirectional temporal modeling component, replaced by a unidirectional recurrent structure or a purely feed-forward temporal aggregator, leads to a noticeable increase in error, particularly at longer forecast horizons. The bidirectional processing and temporal attention appear to improve the model's ability to incorporate both past evolution and contextual information from the historical window, thereby supporting more stable multi-step predictions.

The fusion mechanism is examined by substituting the cross-attention-based integration with simple concatenation followed by a fully connected layer. The simplified fusion produces higher errors and lower anomaly correlation scores, suggesting that explicit interaction between spatial and temporal features through attention yields richer combined representations than passive combination methods. Collectively, the ablation results support the design choice of maintaining distinct spatial and temporal pathways that interact through a structured fusion stage.

Visualization of Forecasting Results

Visual examination of predicted and observed atmospheric fields provides qualitative insight into the model's behavior. Spatial maps of forecasted temperature and wind fields at selected lead times are compared with corresponding observed analyses. In many cases, the proposed model reproduces the large-scale structure of weather systems with reasonable fidelity, including the positioning of temperature gradients and the orientation of wind patterns.

Time-series plots at representative station locations illustrate the model's ability to track diurnal cycles and synoptic-scale changes. The predicted temperature curves generally follow the observed evolution, with smaller phase and amplitude errors than those produced by simpler baseline models. For humidity and wind speed, the model captures major transitions associated with frontal passages, although residual discrepancies remain in the precise timing and intensity of rapid changes.

Error maps, constructed by subtracting predicted fields from observed fields, reveal the spatial distribution of residual errors. In many instances, larger errors are concentrated near regions of sharp gradients or complex topography, which is consistent with the inherent difficulty of forecasting in such areas. Nevertheless, the spatial extent and magnitude of these error regions are generally smaller for the proposed model than for the comparative baselines. These visual diagnostics complement the quantitative metrics and help identify conditions under which the model performs well or encounters greater difficulty.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Computational Efficiency Analysis

In addition to predictive accuracy, computational efficiency is assessed in terms of model size, training time, and inference latency. The number of trainable parameters in the proposed architecture is compared with that of the hybrid baseline models. Through the use of depthwise separable convolutions, controlled channel dimensions, and efficient attention implementations, the proposed model maintains a moderate parameter count that is lower than several deeper hybrid architectures while still delivering superior or comparable accuracy.

Training time per epoch is measured under consistent hardware conditions. The proposed model requires less wall-clock time per epoch than more elaborate spatiotemporal networks that employ heavier attention mechanisms or higher-dimensional feature maps. Inference latency, measured as the time required to generate a complete forecast sequence for a given input window, remains within a range suitable for near-real-time applications. Memory consumption during both training and inference is also monitored and found to be manageable on the graphics processing units employed in the experiments.

These efficiency characteristics indicate that the architectural choices aimed at reducing computational overhead do not come at the expense of forecasting skill. The balance between accuracy and efficiency supports the practical potential of the model in operational or resource-constrained settings.

Error Analysis and Model Behavior

A more detailed examination of forecast errors reveals patterns related to lead time, variable type, and atmospheric regime. Error growth with increasing forecast horizon follows a gradual rather than abrupt trajectory for the proposed model, suggesting that the spatiotemporal representations help constrain the accumulation of prediction errors. Temperature forecasts generally exhibit lower relative errors than wind speed forecasts, consistent with the smoother spatial and temporal characteristics of temperature fields compared with the more variable nature of wind.

Residual analysis indicates that the model tends to under-predict the intensity of certain extreme events, a common challenge in data-driven forecasting systems trained predominantly on typical conditions. Nevertheless, the magnitude of under-prediction is reduced relative to several baselines, particularly when the historical window contains precursor signals of the developing extreme. In regions with sparse observational coverage, errors tend to be larger, highlighting the dependence of data-driven methods on the density and quality of training observations.

Overall, the combination of quantitative metrics, ablation experiments, visual diagnostics, efficiency measurements, and residual analysis provides a multi-faceted assessment of the proposed model. The results indicate that the integration of spatial and temporal feature learning through a structured and efficient architecture yields consistent improvements in forecasting performance across a range of conditions, while maintaining computational characteristics that support practical applicability.

Key Findings and Contributions

The experimental results indicate that the explicit combination of spatial and temporal feature learning within a unified framework yields consistent improvements in forecasting accuracy relative to methods that emphasize only one of the two dimensions or that combine them in a less structured manner. Reductions in root mean square error and mean absolute error are observed across primary atmospheric variables and multiple forecast horizons. Anomaly correlation scores further suggest improved preservation of spatial weather patterns. Ablation studies confirm that each major component of the architecture—multi-scale spatial extraction, bidirectional temporal modeling, and cross-attention-based fusion—contributes measurably to overall performance.

In addition to predictive skill, the architecture demonstrates favorable computational characteristics. Parameter counts, training times, and inference latencies remain moderate compared with more elaborate hybrid models, indicating that accuracy gains need not come at the cost of excessive resource demands. The balance between representational capacity and efficiency constitutes a central finding of the study and supports the practical relevance of the design choices.

Practical Implications

The ability to generate accurate forecasts with manageable computational requirements has direct implications for operational meteorology and related applications. Resource-efficient models are better suited to deployment in regional forecasting centers, mobile platforms, or environments where high-performance computing infrastructure is limited.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Improved short- and medium-range predictions of temperature, humidity, and wind can support decision-making in agriculture, renewable energy management, aviation, and disaster preparedness. The modular structure of the architecture also allows individual components to be adapted or extended for specific forecasting tasks or geographic domains.

Future Research Directions
Several avenues for further investigation emerge from the present work. One direction involves the integration of physical knowledge into the learning process, for example through physics-informed loss terms or hybrid architectures that combine data-driven components with simplified dynamical cores. Another promising area is the extension of the model to probabilistic forecasting, enabling the generation of ensemble-style outputs that quantify prediction uncertainty. Adaptation of the architecture to higher-resolution datasets and to longer forecast horizons would test its scalability and robustness. Domain adaptation techniques could be explored to improve performance in data-sparse regions or under changing climatic conditions. Finally, further optimization of the efficiency mechanisms, including neural architecture search and model compression methods, may yield even more compact yet accurate forecasting systems suitable for real-time operational use.

So, the systematic combination of spatiotemporal feature learning within an efficiency-conscious deep learning framework offers a viable pathway toward improved data-driven weather forecasting. The results and analyses presented in this study provide evidence of the value of such an approach while also highlighting opportunities for continued refinement and broader application.

XV. CONCLUSION

Accurate weather forecasting continues to present significant scientific and practical challenges because of the complex, nonlinear, and multiscale nature of atmospheric processes. Spatial patterns and temporal dynamics interact continuously, and effective prediction systems must account for both dimensions in a coherent manner. At the same time, the growing demand for timely forecasts in operational settings places increasing importance on computational efficiency. The work presented in this study addresses these dual requirements through the design of a deep learning architecture that systematically combines spatiotemporal features while controlling model complexity.

The architecture is organized around distinct yet complementary pathways for spatial feature extraction and temporal sequence modeling. Spatial processing employs hierarchical convolutional operations enhanced by lightweight attention mechanisms to capture multi-scale atmospheric structures. Temporal modeling utilizes bidirectional recurrent processing combined with attention to represent evolutionary trends and contextual dependencies across historical windows. A structured fusion stage integrates the outputs of these pathways, allowing spatial and temporal information to interact before the final prediction layer generates multi-step forecasts. Efficiency is maintained through the use of depthwise separable convolutions, controlled feature dimensions, and streamlined attention implementations.

The model is evaluated using meteorological datasets that provide extensive spatial and temporal coverage, supported by systematic preprocessing, consistent input–output configurations, and a multi-metric evaluation protocol. Comparative assessment against a range of baseline methods, together with ablation experiments, visual diagnostics, and efficiency measurements, provides a comprehensive examination of the architecture’s behavior and performance characteristics.

REFERENCES

1. Bi, K., et al. (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619(7970), 533–538.
2. Chen, Y., et al. (2024). Coupling physical factors for precipitation forecast in China with graph neural network. Geophysical Research Letters, 51.
3. Hersbach, H., et al. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
5. Karlbauer, M., et al. (2024). Advancing parsimonious deep learning weather prediction using the HEALPix mesh. Journal of Advances in Modeling Earth Systems, 16.
6. Lam, R., et al. (2023). Learning skillful medium-range global weather forecasting. Science, 382(6677), 1416–1421.

7. Olivetti, L., & Messori, G. (2024). Advances and prospects of deep learning for medium-range extreme weather forecasting. Geoscientific Model Development, 17, 2347–2365.
8. Rasp, S., et al. (2020). WeatherBench: A benchmark data set for data-driven weather forecasting. Journal of Advances in Modeling Earth Systems, 12(11).
9. Rasp, S., & Thuerey, N. (2021). Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: A new model for WeatherBench. Journal of Advances in Modeling Earth Systems, 13(2).
10. Shi, X., et al. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Advances in Neural Information Processing Systems (Vol. 28, pp. 802–810).
11. Shi, X., et al. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. In Advances in Neural Information Processing Systems (Vol. 30).
12. Tekin, S. F., Fazla, A., & Kozat, S. S. (2023). Numerical weather forecasting using convolutional-LSTM with attention and context matcher mechanisms. arXiv preprint.
13. Vo, T. (2022). An integrated dual attention with convolutional LSTM for short-term temperature forecasting. Cybernetics and Systems, 55(2), 511–533.
14. Wang, Y., et al. (2017). PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In Advances in Neural Information Processing Systems (Vol. 30).
15. Wang, Y., et al. (2019). Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9154–9162).
16. Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. Journal of Advances in Modeling Earth Systems, 12(9).
17. Wu, R., et al. (2024). Spatiotemporal multivariate weather prediction network based on CNN-Transformer. Sensors, 24(23), 7837.
18. Zhang, J. L., Huang, X. M., & Sun, Y. Z. (2024). Multiscale spatiotemporal meteorological drought prediction: A deep learning approach. Advances in Climate Change Research, 15(2), 211–221.
19. Zheng, et al. (2024). TISE-LSTM: A LSTM model for precipitation nowcasting with temporal interactions and spatial extract blocks. Neurocomputing, 590, 127700.
20. Chattopadhyay, A., et al. (2022). Towards physics-inspired data-driven weather forecasting: Integrating data assimilation with a deep spatial-transformer-based U-NET in a case study with ERA5. Geoscientific Model Development, 15, 2221–2237.
21. Chen, Y., et al. (2024). Global-local feature learning via dynamic spatial-temporal graph neural network in meteorological prediction. IEEE Transactions on Knowledge and Data Engineering, 36(11), 6280–6292.
22. Li, B., & Qian, Y. (2024). Weather prediction using CNN-LSTM for time series analysis: A case study on Delhi temperature data. arXiv preprint.

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com