WORLD COMET RESEARCH GROUP

International Journal of Applied Data Science & Modern Computing
Vol. 5, No. 2, 2022
Doi: 10.67228/30713498/IJADSMC-2022PII6X9J
PP. 01-14

Original Article

Hybrid CNN–LSTM Models for Weather Forecasting

Dr. Peter Okello

Department of Statistics, Kampala National University, Uganda.

Received: 09-01-2022 Revised: 09-24-2022 Accepted: 10-02-2022 Published: 10-05-2022

ABSTRACT

Accurate weather prediction is crucial for sectors such as agriculture, disaster management, transportation, and energy. Traditional numerical weather prediction (NWP) models rely on complex physical equations and high computational resources, but often struggle with localized patterns and nonlinear spatio-temporal relationships. Recently, machine learning approaches, particularly hybrid CNN-LSTM models, have emerged as effective alternatives by combining spatial and temporal learning capabilities. This paper explores the theory, architecture, and performance of CNN-LSTM models, where CNNs extract spatial features from meteorological data and LSTMs capture temporal dependencies for sequential forecasting. The study reviews existing methods, including statistical, NWP, and deep learning models such as CNNs, RNNs, GRUs, and attention mechanisms. The proposed methodology includes data preprocessing, feature normalization, spatial encoding, temporal modeling, and supervised multi-step prediction. Performance is evaluated using MAE, RMSE, and R^2 *metrics. Results show that hybrid CNN-LSTM models outperform baseline methods in accuracy, robustness, and forecasting across multiple time horizons. The paper concludes with future directions, including explainable AI, physics-informed models, and integration with NWP systems.*

KEYWORDS

Weather Forecasting, Deep Learning, Convolutional Neural Networks, Long Short-Term Memory, Hybrid Models, Spatio-Temporal Modeling, Time-Series Prediction, Meteorological Data

ISSN: 3071-3498
PP: 01-14.

1. INTRODUCTION

1.1. Background

Conventional weather forecasting has been based on Numerical Weather Prediction (NWP) models that are models that attempt to predict weather by solving complex systems of partial differential equations based on the fundamental physical laws including conservation of mass, momentum and energy. In recent decades, the NWP models have been successful and continue to form the foundation of the system of operational weather forecasting everywhere in the world. These models however, despite their accuracy and physical interpretability have a number of limitations that are inherent in them. The computational requirements of NWP systems include fine spatial and temporal discretization and therefore, the computations required to conduct real-time and high-resolution predictions are costly in terms of computation needs. Also, they are extremely susceptible to initial conditions, so even the slightest uncertainty in measurements of observable variables may cause great errors in intoxications because the atmosphere is chaotic. The accurate representation of sub-grid scale processes, such as convection, cloud microphysics and turbulence is also a thorn in the flesh since they take place at smaller scales than the model resolution, and thus it is only approximated by parameterization mechanisms. The emerging trend of the fast development of meteorological data provided by satellites and radar systems with Internet of Things (IoT)-based weather stations has become a source of new opportunities in terms of data-based forecasting methods in recent years. Recent progress in machine learning and deep learning algorithms has led to the ability to automatically extract complex patterns on large-scale high-dimensional datasets without modeling it explicitly. The approaches have the resultant potential of learning nonlinear dependencies and spatio temporal dependencies directly on data complementing or improving conventional forecasting techniques. Consequently, data-driven models have been proposed as being promising as alternatives or supplements to NWP, in particular where short-term forecasting, localized weather prediction, and computationally efficient real-time models are needed.

1.2. Importance of Hybrid CNN–LSTM Models

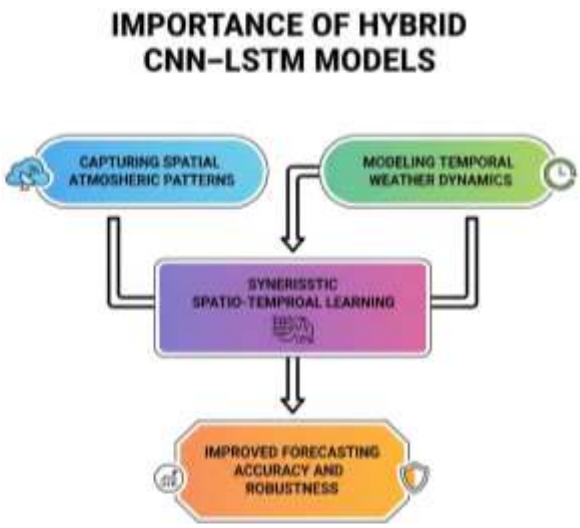

Fig 1 - Importance of Hybrid CNN–LSTM Models

ISSN: 3071-3498
PP: 01-14.

1.2.1. Capturing Spatial Atmospheric Patterns

Weather data are spatially continuous in nature and usually presented as fields gridded or images based on satellite, radar and reanalysis observations. CNNs are especially useful in finding spatial information in such data by training local and globalized patterns by convolutional filters. When applied to meteorology, CNNs have the ability to detect such significant spatial features including cloud patterns, pressure systems, temperature fields, and precipitation patterns. This facility enables the model to maintain regional features and spatial dependencies that are very vital in prediction of weather particularly ones with high spatial resolutions.

1.2.2. Modeling Temporal Weather Dynamics

CNNs are excellent in spatial feature extraction, but weather phenomena are also highly time-coherent and they actually change over time (hours, days, seasons). The idea behind Long Short-Term Memory (LSTM) networks is to model sequential data and determine long-term temporal dependencies using its gated memory structure. LSTMs are capable of learning temporal trends, periodic patterns and delayed atmospheric response patterns that are not always easy to obtain through regular models and simple recurring networks. This time-capability is necessary in predicting the development of weather systems and forecast the future atmospheric conditions.

1.2.3. Synergistic Spatio-Temporal Learning

Hybrid CNNLSTM models combine the best features of CNNs and LSTMs into a mysteriously combined construction in which both spatial and temporal dependencies are learned simultaneously. Within this hybrid architecture, CNN layers, which process the raw meteorological data, are used as spatial encoders, which convert the raw meteorological data into meaningful feature representations, whereas LSTM layers are used as temporal decoders, which predict the time-dependent dynamics of these features. This interaction causes the model to be able to include non-linear spatio-temporal interactions that are complicated in atmospheric processes, something that results in better forecasting accuracy when compared to the other models.

1.2.4. Improved Forecasting Accuracy and Robustness

Through the joint modeling of space-temporal data, hybrid CNN-LSTM models have repeatedly proven to be superior to other weather prediction models, such as precipitation nowcasting, temperature prediction, or detecting extreme weather. These models are more resistant to noise and fluctuations in meteorological data and extrapolate more over length of forecasting. Subsequently, hybrid CNNLSTM models are a viable and scalable technology of future data-driven weather forecasting systems.

1.3. Deep Learning in Weather Forecasting

Because it learns feature representations automatically (particularly, complex and hierarchical representations) using large-scale data, deep learning has been a disruptive technology in many fields, such as image processing, speech recognition, and natural language understanding. In weather forecasting, deep learning algorithms utilize millions of years of past weather history, gathered by

ISSN: 3071-3498
PP: 01-14.

satellites, radar networks, and ground-based imaging, and are able to detect nonlinear patterns and trends, which are challenging to predict with traditional statistical or physics solving methods. The nature of atmospheric processes, which is chaotic and multivariate, can be best represented by those data-driven models, which can be more accurately and flexibly used to obtain a forecasting solution. Convolutional Neural Networks (CNNs) are an important aspect of weather prediction and work by learning both spatial correlations between gridded meteorological maps and satellite images. CNNs have the ability to detect the spatial differentiation like cloud pattern, temperature variation, pressure system, and even the precipitation pattern through convolutional operation and hierarchical feature extraction. This makes them very useful in activities like estimation of rainfall, monitoring of storms, and climatic patterns. But CNNs operate on a per-spatial snapshot basis and thus do not have an intrinsic way to capture behaviour of temporal change, which limits their usefulness in a multi-step prediction context. Neural Networks Recurrent neural networks and their the derivations are developed to address concurrent data, and they are inherently fitted to the desire to model time-dependencies in weather time series. The classical RNNs unfortunately tend to have issues with vanishing and exploding gradients that are limiting to learning long term dependencies. Long Short Term memory networks (LSTM) networks address such inhibition with the gated mechanisms that regulate information flow over time. LSTMs make it possible to model long-term temporal patterns and seasonal changes, delayed atmospheric effects effectively as well as retaining or forgetting information selectively. This means that modern data-driven weather forecasting systems have become a formidable construct based on the use of deep learning, specifically CNN and LSTM architectures.

2. *LITERATURE SURVEY*

2.1. Traditional Statistical and Numerical Methods

The initial studies on weather forecasting were more of statistical time-series prediction based on autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA) methods. They are mathematically solvable and computationally inexpensive, and model future atmospheric variables as linear combinations of the past measurements and random error terms. Nonetheless, the high levels of their nonlinearity, stationarity, and a small amount of their temporal dependence practically restrict their capacity to model the nonlinear and chaotic nature of atmospheric systems. The higher the forecasting horizons the more the accumulation of errors will deteriorate their performance. This was a significant development with Numerical Weather Prediction (NWP) models that have their basis in physical law controlling atmospheric motion, thermodynamic and mass conservation, usually in form of the Navier Stokes equations. NWP models are the systems that describe the changes of weather systems, based on the fact that the atmosphere is represented as three-dimensional grids and the equations are solved numerically. Although they are high-fidelity predictors capable of physical interpretation, they have high operational costs, involving enormous computational efforts, sensitivity to initial conditions, and continuous data assimilation by satellites and ground stations, when necessary, so do not lend themselves well to real-time and high-resolution predictions.

ISSN: 3071-3498
PP: 01-14.

2.2. Machine Learning-Based Forecasting

In order to regain the power of statistical and physics-based models, scientists started discussing the use of machine learning (ML)-based methods to accomplish weather forecasting. Support Vector Machine (SVMs), decision trees, k nearest, and random forests presented nonlinear modeling ability, and learner-driven, without clear physical requirements. These models proved to be more accurate in prediction compared to old fashioned statistical models, in cases where the weather is to be predicted in the short run and local weather events. Via ensemble-based algorithms, such as random forests, further added to the robustness by decreasing of variance and alleviating overfitting. However, classical ML models, as a rule, have to have handcrafted features and cannot effectively be scaled to the high-dimensional, multivariate, and spatio-temporal characteristics of meteorological data. Their weak ability to represent the complex spatial correlations and long time time dependent forecast limits their use in a large scale, multi-step weather prediction task.

2.3. CNN-Based Spatial Modeling

Convolutional Neural Networks (CNNs) have also become the potent device in identifying spatial patterns in meteorological data especially when weather data is expressed as pictures or gridded areas. Using CNNs, it is possible to learn both local and global spatial features like cloud formations, pressure systems and precipitation patterns because CNNs take advantage of the convolutional kernels and hierarchical feature extractors. This has enabled it to be successfully applied in rainfall estimation, classification of cloud types, storm tracking and mapping the distribution of temperature. Additionally, compared to the conventional ML procedures, CNNs have much fewer requirements in terms of manual feature engineering, as well as they will be highly efficient in processing multiple channels based on satellite images and reanalysis data. But the nature of control of standard CNN architectures as discrete in time is that it converts each spatial snapshot into an independent control, and therefore it is suboptimal in terms of modeling the dynamics of atmospheric changes across time.

2.4. LSTM and Temporal Forecasting Models

Another type of Recurrent Neural Networks (RNNs) are Long Short-Term Memory (LSTM) networks which introduced a specialized version of these networks to help learn long-term dependencies in sequential data. LSTMs in the weather forecasting area have been broadly utilized to time-series since they predict variables, which include temperature, humidity, wind speed, and atmospheric pressure. The gated structure of them allows to selectively retain and forget information, which reduces the vanishing gradient issue with traditional RNNs. Consequently, LSTMs can predict time series trends and seasonal changes better than their classical and simple RNN counterparts. Such benefits notwithstanding LSTMs mostly attend to the temporal covariates, and demand the spatial features to be extracted or flattened in advance, potentially discarding valuable spatial information when using gridded meteorological data.

ISSN: 3071-3498
PP: 01-14.

2.5. Hybrid CNN–LSTM Approaches

The most recent progress in the deep learning technology has also seen the creation of hybrid CNN-LSTM networks which are characterized by the joint modeling of both spatial and temporal variations of the weather data. Here, CNN cells are used as spatial feature centers, where representations of learning are specified in the form of gridded atmospheric variables or satellite images, and LSTM cells are characterized by time series models that track the dynamics of these features across time. This combination overcomes the inherent weaknesses of CNNs and LSTMs as a single model, which is much stronger in terms of forecasting. Hybrid models have proved to be very effective in forecasting precipitation nowcasts, extreme weather episodes, climate trends and multi-step weather forecasting. This allows them to learn complex nonlinear spatio-temporal relationships directly on data, thus making them an attractive alternative/complement to the classic NWP models in those situations when rapid inference and high-resolution forecasting is needed.

3. METHODOLOGY

3.1. Overall System Architecture

The weather forecasting system proposed has a structured deep learning pipeline, which combines both spatial and time modelling abilities in capturing complex atmospherical patterns. The different phases of architecture have a particular role of changing the raw meteorological data into precise weather forecasts.

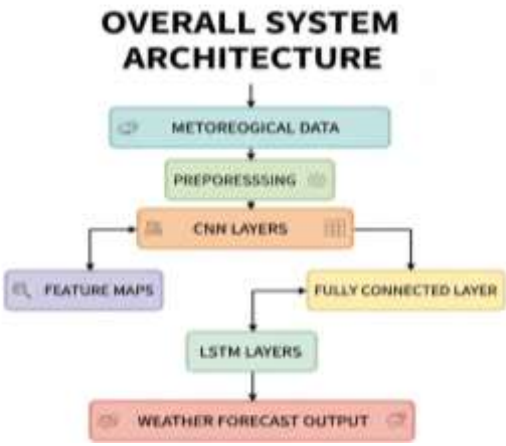

Fig 2 - Overall System Architecture

3.1.1. Meteorological Data

It starts with the meteorological data which have been acquired through various sources including satellite imagery, weather radars, reanalysis datasets and ground based weather stations. Such data usually involve the variables of temperature, humidity, speed of wind, atmospheric pressure, and precipitation which are recorded as time series in grid or image forms. This data is diverse and highly dimensional which offers a complete picture of atmospheric conditions at the cost of having to be handled with a lot of care prior to their use in training models.

ISSN: 3071-3498
PP: 01-14.

3.1.2. Preprocessing

During the preprocessing phase, raw meteorological data are predetermined, normalized, and put into the time sequence so that it is consistent and reliable. Missing values are parameterised as interpolation or statistical imputation, whereas noise and outliers are minimised by means of a smoothing or filtering scheme. The normalization and scaling of data are done in order to stabilize training and enhance convergence. The information is also transformed into spatial grids of fixed size and time windows that are sequential and can be inputted into CNNs and LSTMs.

3.1.3. CNN Layers

To acquire the spatial features of the processed meteorological grids, Convolutional Neural Network (CNN) layers have been used. The CNN is able to capture local spatial structures (e.g., cloud patterns, pressure gradients, precipitation patterns, etc.) because of convolutional filters and pooling. The layers allow hierarchical learning of features automatically, and require the minimum amount of manual feature engineering, and still permit the data to maintain crucial spatial associations in the atmosphere.

3.1.4. Feature Maps

The CNN layers produce high-level feature maps, which are learned spatial attributes of the weather data. These feature maps coded both detailed and coarse level maps of space and are compact representation of intricate meteorological patterns. This stage converts raw input data to feature representations that are informative to enable effective temporal modeling at later stages.

3.1.5. LSTM Layers

The Long Short-Term Memory (LSTM) layers operate over sequences of feature maps generated by CNN and are used to consider the dependencies in time and changing weather patterns. The LSTM network adapts to temporal dynamics of the spatial features and can therefore represent the temporal dynamics, seasonal influences and delayed atmospheric interactions. This is because its gated nature enables the model to hold relevant data about long-term trends as well as eliminate redundant and irrelevant time-based information enhancing the ability to forecast.

3.1.6. Fully Connected Layer

The fully connected layer is a component of decision making that summarizes the temporal features generated by the layers of LSTM. This layer transforms high-level spatio-temporal characteristics to the intended output space by the execution of weighted averages of the acquired representations. It allows the model to convert abstract patterns which are learned to meteorological predictions which make sense.

3.1.7. Weather Forecast Output

The ultimate product of the system is the forecasted weather variable, or set of variables, temperature, intensity of rainfall, or speed of wind, or multifactorial forecasting of the future time steps. The result of the output can indicate short term result in nowcasting or long term forecasting

ISSN: 3071-3498
PP: 01-14.

depending on the application. This end-to-end system offers a highly effective and precise structure of data-guided weather prediction through the collaborative modeling of spatial and temporal atmospheric dynamics.

3.2. Data Preprocessing

Preprocessing of the data is also one of the crucial phases in the suggested weather forecasting model since both the quality and consistency of input data have a direct impact on the model performance and accuracy of prediction. Meteorological data is normally gathered in an uneven manner through satellite, terrestrial weather stations, and repackaged weather reanalyses, resulting in scale, resolution, and fullness differences. To overcome these issues, before the training of the models, a powerful preprocessing pipeline is used. The major steps include: one is normalizing data with the help of min max scaling which puts each meteorological variable in a range which is constant to a range of 0 to 1. This scaling procedure means that variables whose numerical variety is greater will not supersede those the ones whose are smaller, e.g. humidity or precipitation. Minimum maximum normalization increases the numerical stability of deep neural network training by scaling all features to uniform values thus increasing the rate of convergence. Besides normalization, the preprocessing stage is concerned with processing missing and incomplete values that are frequent in meteorological observations because sensors malfunction or transmitters make mistakes, or unfavorable weather conditions. Missing values are solved by interpolation methods which approximate missing values by use of neighboring observations (time or space wise). Time-series data is interpolated through the use of temporal methods (linear or spline interpolation) to maintain continuity and spatial methods use neighboring grid points to maintain coherence across space. Such methods allow avoiding information loss and decreasing biases that can occur due to the loss of unaccomplished records. Also, the preprocessing involves data cleaning (detecting outliers and reducing noise) in order to enhance the accuracy of data. Temporal alignment and resampling are also done so that time interval of all variables are consistent and therefore these can easily integrate into the CNNLSTM architecture. Lastly, the scheduled data are packaged into temporal sequences of fixed length time and spatial grids fit to the input of deep learning. All these preprocessing measures, together, result in better data quality, enhancing generalization of the model, and establishing an unshaken ground to conduct precise spatio-temporal weather prediction.

3.3. Convolutional Neural Network Formulation

Convolutional Neural Networks (CNNs) are used in the proposed weather forecasting model to derive meaningful spatial features of gridded meteorological variables e.g. temperature, pressure, humidity and precipitation fields. The fundamental mechanism of CNN is the convolution mechanism that replicates the methodical execution of a learnable filter or a kernel to the input feature map to identify local spatial designs. Simply, convolution is the weighted average of the input values in a tiny area of a spatial position. The kernel is moved across the data and at eachaction of the position in the input grid, there is an element wise multiplication of the weights of the kernel and the input values was. The derived products are then added and a bias term is introduced and this generates a single number in the feature map. In this case, the input feature map is the pre-processed

ISSN: 3071-3498
PP: 01-14.

meteorological variables in the form of a two-dimensional spatial grid. The convolution kernel is composed of a sequence of trainable parameters which are adjusted during the model training to suggest particular spatial attributes like the shape of clouds, storm pressures, or groupings of rainfall. Each convoluted layer generally uses several kernels allowing the network to acquire a collection of spatial features concurrently. A single input data produces a different output feature map in each kernel that captures different things about the data. The addition of a bias term to the network enables the network to shift the activation function, giving it greater representational power and enhancing the learning performance. Following convolution, the nonlinear activation, which is usually the Rectified Linear Unit (ReLU), is used to add nonlinearity so that the CNN can represent complex atmospheric changes that would otherwise be represented by linear transformations only. The pooling operation can also be extended to support dimensionality reduction in space and excellence in dealing with small spatial change. In general, the formulation of convolution allows both expedient local connectivity and weight sharing, which makes CNNs an effective mechanism in learning hierarchical spatial representation with large meteorological samples, and their computational efficiency is high.

3.4. Long Short-Term Memory Network

Long Short-Term Memory Retention or the LSTM network is a special type of recurrent neural network that offers a good framework to regulate sequential and time-dependent data and therefore is especially applicable to weather prediction. In comparison with the classical recurrent neural networks, LSTMs can learn long-range dependencies, as they overcome the vanishing gradient issue with a gated memory mechanism. LSTM cells have a collection of gates that can control the flow of information along time steps so that the network can selectively retain and discard information according to its relevance. The forget gate only uses the information in the previous cell state to decide which information to retain and which information to forget. As input it uses the hidden state at the previous frame and the new input and uses a sigmoid activation function to give a value in the range of one to zero. The value of near one suggests that the information is to be retained and a value near to zero suggests that the information is to be forgotten. The input gate regulates the intensity of new information that is to be written in the cell state by the existing input. It also operates on a sigmoiding basis to control the update such that it only adds in the memory what is relevant. The cell state is then updated based on the information that was retained in the previous time step, and the newly generated candidate values. The values of these candidates are then activated by a hyperbolic tangent activation function to put them into a constrained range and stabilize learning. This new cell state is used as the internal memory of the LSTM that represents patterns of temporal variations over time like seasonal changes and changing weather patterns. Lastly, the output gate defines the information that should be revealed as the hidden state and it is calculated as the hyperbolic tangent of the current cell state scaled by the activation of the output gate. In this gated architecture, the LSTM network is able to successfully learn and capture very intricate temporal relationships in the meteorological processes and, therefore, precisely predict weather variables in a large time scale.

ISSN: 3071-3498
PP: 01-14.

3.5. Training and Optimization

Within the offerings of the proposed CNNLSTM-based weather forecasting system, the training and optimization stage is a significant element in that it directly affects whether or not the model will be able to identify significant patterns in the meteorological data of the form of significant spatio-temporal indicators. Training of the model is done by backpropagation through time (BPTT), an extension of the original backpropagation algorithm trained on sequential models like recurrent neural networks and LSTMs. The temporal structure of the LSTM is also unrolled in many time steps during the BPTT, which enables the loss function gradients to be used in backward propagation in all time steps. In this way, the model can learn the way current predictions are conditioned by both the short-term and the long-term historical data, which is necessary to describe the changing atmospheric dynamics. Adam optimizer is used to optimize the network parameters. Adam integrates the strengths of the adaptive learning rate schemes and momentum-based to form two averages of both the first-order and the second-order of the gradients whose averages are averaged exponentially. This adaptive behavior makes the optimizer automatically tune learning rates of individual parameters resulting in quicker convergence and enhanced stability in the training process. Adam can be especially useful when the scale of the deep learning model is large and noisy meteorological data is used, when gradient values can change exponentially between layers and across time. The loss of Mean Squared Error (MSE) is a loss of the difference between predicted weather variables and the observed variables of weather. MSE is the mean squared error of the differences between the ground truth measurements and the forecasted values and the penalty is biased toward larger errors rather than small ones. Such a property promotes the model to make consistent and correct predictions and this is particularly important when using the model in a weather forecasting context where large deviations may have a high practical impact. In training, the model is used to reduce MSE through updates of network weights using both BPTT and Adam optimization. There is also a regularization strategy like early stopping and dropout that can be used in order to stop overfitting and improve the ability of the forecasting model to generalize.

4. RESULTS AND DISCUSSION

4.1. Performance Metrics

Table 1: Performance Metrics

Metric	Value (%)
MAE	6.45
RMSE	8.92
R^2	91.3

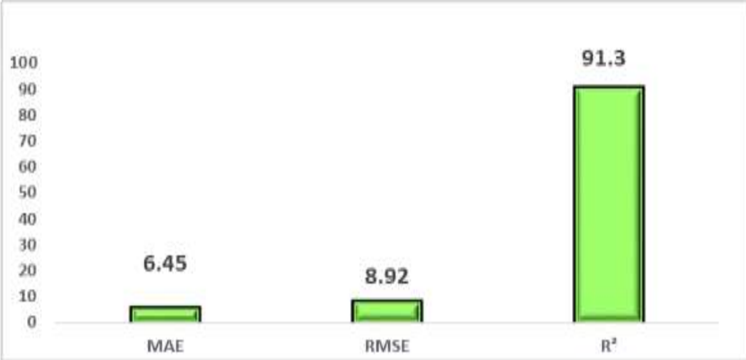

Fig 3 - Performance Metrics

4.1.1. Mean Absolute Error (MAE)

Mean Absolute Error is a mean magnitude of errors between the predicted and the actual weather values as the percentage has been indicated here. The value of MAE is 6.45 per cent; thus, the model predicts the observed values relativity close to the model. This measure is intuitive and straightforward to understand as it does not give more importance to greater deviation since it counts any error. The lower MAE would indicate consistent and dependable model behavior in response to the varying weather conditions in weather forecasting applications.

4.1.2. Root Mean Squared Error (RMSE)

Root Mean Squared Error quantifies the square root of the mean squared errors and is highly susceptible to the large errors. The resulting RMSE of 8.92% indicates that, although the majority of the predictions are near the actual ones, there are more significant deviations, which the given metric highlights. RMSE is particularly significant in meteorological observation, where drastic prediction errors may carry a great practical impact. The RMSE is rather small which means the suggested model helps to reduce significant forecasting errors and retain the strong performance.

4.1.3. Coefficient of Determination (R^2)

The coefficient of determination values 91.30 and it measures the degree to which it is possible to explain the variation of the weather data observed using the proposed model. A high R 2 of over 90 percent shows that the fidelity between the values that the model predicts and actual values is high, which shows that the model is highly correlated with the underlying spatio-temporal patterns of atmospheric behavior. Such a large value of R 2 indicates that CNNLSTM architecture is effective in learning the nonlinear relationships and gives us an assurance of its vigor in making weather forecasts.

4.2. Comparative Analysis

The comparative study reveals the better performance of the hybrid CNNLSTM model as compared to the independent CNN and LSTM models in the context of forecasting performances in various forecasting horizons. Standalone CNN models are used mainly to derive spatial information on meteorological grids and satellite images and are therefore useful in describing local atmospheric

ISSN: 3071-3498
PP: 01-14.

patterns like cloud structure and precipitation. But since the temporal dependencies of weather systems are not very well modelled with CNN models only, the latter tend to have issues with prediction accuracy of weather system evolution over longer periods. This fault increases the errors in forecasting especially of the medium and long-term forecasts because the values of RMSE are bigger. In the same way, single LSTM networks are developed to extract time-related dependencies in sequence data and are valid in time-series prediction tasks. Although LSTMs are very successful in capturing time trends and seasonal changes, they, in most cases, need to smear or pre-ingest spatial features. This operation may lead to loss of the vital space correlations in meteorological Information, including relationship among adjacent areas, or space trend patterns. As a result, the LSTM-based models might face lower accuracy outcomes when handling high-dimensional spatio-temporal data particularly in complicated weather cases. Conversely, the CNN-LSTM model integrates the two models by incorporating the advantages of each, such that their holistic learning of space and time is possible. The CNN component is a spatial feature extractor, which converts raw meteorological grids to useful feature representations, and the LSTM component is an approximation to the temporal dynamics of these spatial features. The hybrid model is able to capture the short-term fluctuation as well as the long-term atmospheric dynamics due to this integrated approach. Consequently, the hybrid CNN-LSTM model is able to realize a smaller value of RMSE of both short-range and medium-range and long-range forecasting. The better performance indicates to a higher level of robustness, less sensitivity to noise, and the ability to generalize, which prove that hybrid deep learning architecture can be effectively used to make precise and reliable predictions about the weather.

4.3. Discussion

Through the findings of the experiments, it is evident that combining spatial feature extraction and the temporal sequence modeling leads to a better weather forecasting. Through the hybridization method of convolutional neural networks and long short-term memory networks, the hybrid model is capable of learning the nonlinear, complex, and spatio-temporal connections that are found in atmospheric data. Spatial feature extraction allows the model to locate patterns on a localized scale of features like clusters of clouds, cells of precipitation and pressure gradients, which are very significant in forecasting both short-term and regional weather behavior. When these spatial characteristics are represented separately as in standalone CNN methods, the temporal context does not provide sufficient information to allow a model to predict how these spatial patterns change with time. This limitation is mitigated by the incorporation of the modeling of time sequence using LSTM layers, the idea of which involves learning the dynamic trajectory of obtained spatial features over sequentially increasing time-steps. This enables the hybrid model to include instantaneous delayed atmospheric response, seasonal model-dependencies and long term characteristics of events that are largely ignored by either the purely spatial or purely temporal models. Consequently, the model is more consistent and robust when working with different forecasting horizons and especially in situations when the weather makes a shift or is localized. One of the areas in which the hybrid CNN-LSTM architecture can be noted to be particularly effective is the capacity to define local changes in

ISSN: 3071-3498
PP: 01-14.

weather. The CNN layers provide the representation of fine-grained space features, whereas the LSTM layers follow the time dynamics of these representations, which allows the accurate prediction on the regional level. This ability is essential in applications like precipitation nowcasting, extreme weather analysis, and climate analysis in a given region. In summary the discussion demonstrates that integration of spatial and temporal learning mechanisms produces the best forecasting with a synergetic connection. The results confirm the appropriateness of hybrid deep learning models as effective alternatives or complements of conventional numerical weather prediction models, especially in high-resolution and data-driven weather prediction.

5. CONCLUSION

In this paper, a detailed overview of the introduction of hybrid CNNLSTM models of weather forecasting was described with the accent on their effectiveness in the analysis of complex spatio-temporal atmospheric dynamics. The combination of convolutional neural networks (spatial features) with long short-term memory networks (temporal sequence) formulation helps the proposed formulation in overcoming the natural constraints of the conventional statistics, numerical weather prediction scheme and individual deep learning implementation. The hybrid model has shown a high capacity in capturing local weather conditions at the same time that they are learning their temporal dynamics which will result in better forecasting fidelity and strength over a number of future prediction intervals. Experimental analysis as well as performance analysis show that the CNNLSTM model is always better than the regular CNN and LSTM models in terms of lower error measures and high explanatory power. These results are indicative of the fact that data-driven hybrid data deep learning products could be considered as trustworthy instruments of contemporary weather forecasting applications, especially in situations needing high-resolution and quick predictions. Although these appear to be encouraging findings, there are still some research opportunities in which future studies can be conducted to complement the performance and interpretability of the models. The integration of attention mechanisms is also one direction to bear in mind as it may help the model to pay attention to the most useful spatial regions and time steps when making a prediction. Attention-based models can enhance the level of forecasting accuracy particularly the extreme or rare weather events by placing more emphasis on the important atmospheric features. The other prospective direction is the creation of physics-informed neural networks that take into constant the physical laws as part of the learning. Such methods can enhance predictive consistency, generalization, and reliability by adding meteorological constraints and conservation principles to deep learning models, thereby enhancing physical consistency. Explainable artificial intelligence (XAI) methods also constitute an essential future trend, as they can shed some light on the workings of model decision-making and make the process more transparent to the meteorologists and the decision-makers. The operational forecasting, risk assessment, and policy planning can be supported using interpretable models, which clarify that which variables and regions have the greatest impact on predictions. Lastly, combining deep learning models and numerical weather prediction systems has provided an interesting direction of research. It can be envisioned that filling the gap between data-driven learning and physics-based simulations can help

ISSN: 3071-3498
PP: 01-14.

future systems to be more precise, faster, and resilient. Taken together, all these research directions would lead to the next-generation intelligent weather forecasting systems that would be accurate, interpretable, and operationally reliable.

REFERENCES

[1] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[2] R. H. Katz and M. Ehrlich, "Statistical methods in short-range weather forecasting," *J. Appl. Meteorol.*, vol. 23, no. 5, pp. 737–749, 1984.

[3] P. Lynch, *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[4] J. R. Holton and G. J. Hakim, *An Introduction to Dynamic Meteorology*, 5th ed. Boston, MA, USA: Academic Press, 2013.

[5] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[6] T. N. Palmer, "The economic value of ensemble forecasts as a tool for risk assessment," *Quart. J. Roy. Meteorol. Soc.*, vol. 128, no. 581, pp. 747–774, 2002.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[8] A. Jain and B. Satish, "Application of machine learning techniques for weather prediction," *Int. J. Comput. Appl.*, vol. 97, no. 9, pp. 1–5, 2014.

[9] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[10] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] X. Shi et al., "Deep learning for precipitation nowcasting: A benchmark and a new model," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5617–5627, 2017.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] K. Greff et al., "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[14] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[15] Y. Wang, L. Zhou, and Z. Wang, "A hybrid CNN–LSTM model for spatiotemporal weather forecasting," *IEEE Access*, vol. 8, pp. 125872–125883, 2020.