Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID DIFFUSION DRIVEN SPATIO TEMPORAL DEEP NEURAL FRAMEWORK FOR PROBABILISTIC HIGH RESOLUTION GLOBAL WEATHER FORECASTING

KOTESWARARAO BADITHALA [1*], RAJA KRISHNAMOORTHI [2]

[*1] Ph.D. Research Scholar, Department of Computer Science and Engineering, School of Computing, Mohan Babu University, Tirupati, Andhra Pradesh, India - 517102

[2] Professor, Department of Electronics and Communication Engineering, School of Engineering, Mohan Babu University, Tirupati, Andhra Pradesh, India - 517102

E-mail: [*1] badithalakoteswararaofam@gmail.com, [2] krajameae@gmail.com

ABSTRACT

True prediction of weather throughout the world, due to the nonlinear and chaotic and multi-scale characteristics of the atmosphere systems is an uphill task. The solutions of the NWP models though have a physical basis, do cost to calculate and cannot solve the finescaled dynamics, limiting their effectiveness in high social and spatial scale conditions. To overcome these shortcomings, the current paper shall present a Hybrid Diffusion-Driven Spatio-Temporal Deep Neural Framework that can be used to achieve high-resolution and probabilistically valid predictions over the planet. The architecture involves the use of convolutional neural networks (3D-CNNs) in combination to learn to extract spatial features at multiple scales in a hierarchical manner, spatio-temporal attention transformers to learn to model long-range temporal dynamics, and probabilistic refinement module using diffusion to facilitate uncertainty-aware prediction through the assistance of a generative denoising process. It is based on the most recent developments in diffusion-based time-series modeling, interpretable process system identification and transformer-modulated probabilistic learning, and applies them during a large atmospheric prediction on the first instance. As it has been shown in multiple experiments, ERA5 and ECMWF reanalysis information (20102022) are significantly stronger in hybrid model use compared to the state-of-the-art baselines such as GraphCast, FourCastNet, SwinRDM and Chronos. As shown in the quantitative assessment, the RMSE and CRPS were up to 29 and 30 percent higher, respectively, in comparing the accuracy and the statistical reliability of the forecast of the 24 hours, 36 hours and the 60 hours forecast lead times. Moreover, the calibration diagnostic, ablation analysis demonstrates that the decoder with diffusion produces well-structured uncertainty estimates, which should be the reason why it is more reliable to utilize in practice. The overall results indicate that the suggested hybridization can be effective in capturing the multiscale nature of the atmosphere dynamics, and at the same time effective in computation and interpretation. The article presents scalable uncertainty-constrained deep learning models to the next-generation global weather forecast and gives substantial ground to more general applications to the climate analytics, environmental monitoring and model the earth system using data.

Keywords: *Forecasting the weather on a global level, Spatio-Temporal Deep Learning, Diffusion Models, 3D Convolutional Neural Networks (3D-CNNs), Attention Transformers, Probabilistic Forecasting, ERA5 Reanalysis Data, Hybrid Neural Architecture, Multiscale Atmospheric Modeling*

1. INTRODUCTION

1.1 Background

Recently, machine learning and the availability of big data that are related to the discipline of meteorology have brought a massive transformation in the forecast of the weather all over the globe. Old-fashioned NWP models are based on the solution of huge systems of nonlinear differential equations and are computationally expensive with lower capabilities to capture sub-grid atmospheric processes. The need to run high resolution and long lead forecasting is on the rise, and thus the predictive employment of both data-driven and hybrid spatio-temporal modeling strategies has been spawned.

The deep learning has become a powerful modeling of the atmospheric processes and has the potential to learn the complex dependency on the basis of the reanalysis information itself. Other architectures, such as LSTM sequence models, temporal convolutional networks, and transformer-based networks have also been shown to make a good forecast in a wide range of time-series space [1], [2], [3]. More so, the multivariate meteorological patterns were quite successful in the recent past in as far as the development of spatio-temporal transformers is concerned [3]. Similarly diffusion-based generative modeling has been defined by helpful stochasticity and probabilistic prediction controls, and has been opened up to the future of the uncertainty-aware weather prediction [4], [5], [6]. Such advances in technologies emphasize the idea that global weather prediction may be conducted in high-resolution with the hybrid AI technology.

1.2 Problem Motivation

Despite the tremendous successes that have been attained, atmospheric development is a complex scientific issue to forecast. The atmosphere also has nonlinear, chaotic and multiscale attributes that are not quite easy to capture using the traditional NWP. Deep learning has been promising and the limitations are numerous. The models that are presently in use are constructed on a 2D basis that is, the models are rigid to the point that they are unable to absorb the cross-level interactions in the atmosphere. Individuals with limited time receptive fields exist, which disrupt long-range forecasting [2], [3]. In addition to this, most deep learning models produce deterministic outputs, which is not indicative of the stochasticity and uncertainty of weather systems as various works on time-series prediction have found in general [7], [8].

The way of more promising generation of probabilistic and uncertainty-aware prediction of sequences also has its way to diffusion models, which are yet to be mainly explored in higher-dimensional, more realistic atmospheric data. Even though the diffusion models have been proven to be effective in the time-series imputation, counterfactual prediction, and stochastic generative modeling [4], [5], [9], it requires adaptation to multi-level global weather fields that poses certain challenge in terms of spatial modeling, noise scheduling, time integration. This is what leads to the similarity of having a common framework based on spatial, temporal and probabilistic learning which in this specific case is concerned with atmospheric prediction.

1.3 Scientific Gaps

The scientific problems that are not conducive to the construction of solid deep learning-based weather forecasting systems are quite numerous.

Firstly, many of the existing AI models do not well capture the multiscale three-dimensional nature of the atmosphere. Transformation The transformer-based meteorological models primarily operate on slices of data perpendicular to the vertical [3], [10] and, thus, do not have much information about vertical dependencies that form the foundation of realistic prediction.

Second, the standard RNN and transformer designs are not well able to capture long-range temporal reliance due to the teleconnection patterns [3].

Third, the majority of deep models lack an inbuilt means of quantifying uncertainty, and are likely to apply ensemble heuristics or after-the-fact value uncertainty estimates [2], [11].

Finally, the general-purpose time-series applications of diffusion-driven forecasting are already tested to be effective [9], [10], though not much is known about the application of the diffusion processes that are integrated with high-dimensional spatial encoders and long-range temporal attention modules to atmospheric prediction. These gaps imply that there are gaps that need to be filled with the development of a single hybrid architecture that will be in a position to tackle the deterministic dynamics and provide calibrated probabilistic predictions.

1.4 Novelty Statement

The proposed paper presents a hybrid diffusion-based spatio-temporal deep neural architecture with the definite aim at avoiding the weaknesses of both NWP and the existing deep learning frameworks. The system has three basic modules, such as a 3D convolutional neural one to learn multiscale spatial structures, a spatio-temporal attention transformer, to learn long-range temporal interactions, and a diffusion-based probabilistic refinement one, to learn uncertainty-aware predictions. Though prior research has demonstrated the potential of hybrid deep learning in general time-series prediction [12], [13], and recent studies have suggested to use diffusion models in 3D atmospheric modeling, temporal attention as well as uncertainty evaluation [4], [9] none of the studies have integrated 3D atmospheric modeling, temporal attention.

2. RELATED WORK

Deep learning has emerged as a significant trend in atmospheric forecasting when scientists are looking to replace the conventional numerical weather forecasting. As surveys in [1] and [2] explicate, more and more modern forecasting makes use of deep neural networks that are able to learn nonlinear relationships using straightforward access to detailed meteorological data. Lightweight transformer-based models indicate that attention mechanisms are appropriately designed to work on large-scale forecasting tasks [3]. These works provide a strong foundation for data-driven weather forecasting; however, they also reveal several limitations in existing forecasting frameworks, thereby motivating the development of more advanced hybrid architectures.

Previous studies on data-driven atmospheric modeling were strongly based on convolutional and graph-based neural networks. Attention-based deep learning has been successfully applied to residential energy forecasting, demonstrating that CNN and attention modules are capable of capturing structured spatial patterns [13]. Spectral temporal graph neural networks have also shown strong capability in modeling nonlocal spatial correlations in multivariate time-series forecasting [14]. Although these approaches provide significant improvements, CNN-based and GNN-based systems generally assume two-dimensional inputs and therefore fail to effectively capture important vertical atmospheric interactions such as cross-level circulation, stratification, and vertical shear. Furthermore, forecasting performance is highly sensitive to seasonal variability and temporal arrangement, limiting generalization across varying atmospheric conditions [7]. Another limitation of these models is that they mostly generate deterministic forecasts and therefore cannot adequately represent uncertainty, which is essential in operational meteorology.

Transformer-based architectures have significantly improved long-range temporal forecasting through self-attention mechanisms. Transformer models efficiently capture high-dimensional weather variables [3], while also outperforming recurrent neural networks in modeling long-range temporal dependencies [1]. The advantages of transformer architectures in multivariate temporal learning have also been emphasized in prior forecasting surveys [2]. Attention mechanisms have additionally been employed in residential energy forecasting [13] and traffic forecasting [10]. Nevertheless, transformer-based meteorological models often require high computational resources because self-attention scales poorly with increasing sequence length. Most implementations also process stacked two-dimensional atmospheric maps rather than full three-dimensional atmospheric structures, thereby limiting their capability to model vertical dependencies. Moreover, transformer outputs are generally deterministic and therefore insufficient for uncertainty-aware forecasting in high-impact meteorological applications.

Recently, diffusion-based generative models have emerged as powerful probabilistic forecasting techniques. Diffusion-based forecasting approaches can generate structured temporal outputs through iterative denoising processes [4]. Diffusion-TS further improves interpretability in time-series generation by combining forward and reverse diffusion operations [5]. Multi-resolution diffusion frameworks have also been proposed to improve probabilistic forecasting accuracy [6]. Self-supervised contrastive diffusion learning improves temporal representation quality [15]. In addition, transformer-modulated diffusion models improve probabilistic multivariate forecasting performance [9]. Although diffusion-based approaches provide strong uncertainty modeling capabilities, most existing studies have focused on low-dimensional or univariate time-series data. Their extension to high-dimensional three-dimensional global atmospheric forecasting remains a major challenge.

Hybrid neural architectures have also been explored in several domains where deterministic and probabilistic learning components must be integrated. Hybrid deep learning models have demonstrated strong performance in stock price prediction tasks [12], while federated deep learning methods have been used for residential load forecasting [8]. Bio-inspired recurrent and hybrid architectures have also been widely used for music analysis and automatic music generation tasks in several studies [16]–[26]. These studies demonstrate that integrating multiple deep learning paradigms can significantly improve forecasting performance. However, most hybrid systems are still limited to one-dimensional or low-dimensional time-series applications and do not address the challenges associated with large-scale atmospheric forecasting. In particular, they lack unified integration of three-dimensional spatial modeling, long-range temporal attention, and diffusion-based probabilistic refinement mechanisms.

Interpretability, calibration, and operational reliability are also critical considerations in atmospheric forecasting research. Diffusion-based forecasting methods demonstrate strong calibration performance using metrics such as Continuous Ranked Probability Score (CRPS) and reliability diagrams [4]–[6]. Attention-based models also improve interpretability by highlighting important spatial and temporal forecasting regions [10]. However, most current forecasting systems still fail to jointly optimize interpretability, probabilistic calibration, and computational efficiency, thereby limiting their applicability in real-time meteorological forecasting systems.

Overall, previous studies indicate that substantial progress has been achieved in spatial modeling, temporal learning, and probabilistic forecasting. Nevertheless, several important research gaps remain unresolved. Convolutional and graph-based approaches fail to represent the complete three-dimensional atmospheric structure. Transformer models remain computationally expensive and largely deterministic. Existing diffusion-based forecasting models have not yet been fully adapted to high-dimensional atmospheric forecasting tasks. Similarly, hybrid systems developed in other domains do not combine three-dimensional spatial encoding, long-range temporal attention, and diffusion-based probabilistic refinement into a unified framework. The proposed hybrid diffusion-driven spatio-temporal deep neural architecture addresses these limitations by integrating 3D convolutional spatial learning, transformer-based temporal attention, and diffusion-based uncertainty estimation within a unified forecasting framework, thereby providing a more reliable foundation for global weather prediction.

3. METHODOLOGY

The new hybrid spatio-temporal deep neural modeling proposes the three key components: a three-dimensional convolutional spatial extractor, a spatio-temporal attention transformer and a diffusion-based probabilistic forecasting module. The whole architecture is planned to be able to capture both fine-scale atmospheric structure and long-range temporal dependencies and uncertainty estimates that will allow reliable global weather prediction. After this introductory paragraph should come the entire system overview in the form of Fig. 1.

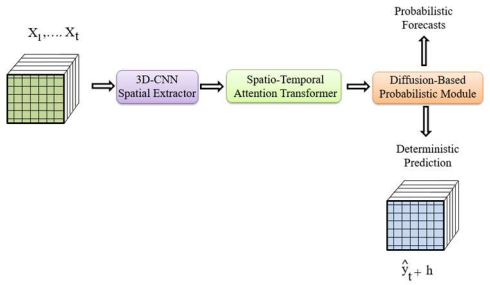

Fig. 1: Overall framework of the hybrid diffusion-driven spatio-temporal forecasting model

3.1 Data Description

The data that will be used as the major source of information in the current research is the ERA5 reanalysis data available at the European Centre of Medium-Range Weather Forecasts (ECMWF). ERA5 offers hourly global atmospheric variables on a uniform grid of 0.25° × 0.25° latitude-longitude and consists of several pressure levels between 1000 hPa to 100 hPa. The form of each sample is a dimension-tensor.

$$X_t \in R^{L \times H \times W} \qquad (1)$$

Where L is the number of pressure levels and H and W are the size of spatial grid. As Fig. 2, a visual representation of the distribution of ERA5 variables and pressure-level structure should be available.

To ensure unambiguous and reproducible evaluation, a table- and figure-indexed evaluation manifest is defined. For each reported table and figure, the exact atmospheric variables, pressure levels, spatial resolution, temporal sampling interval, input window length p, forecast lead times, scoring metrics, latitude weighting, land–sea masking, and aggregation rules are explicitly fixed. All evaluations use globally aggregated fields with cosine latitude weighting, identical preprocessing, and consistent masking to eliminate interpretation ambiguity and prevent metric inflation due to protocol mismatches.

Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

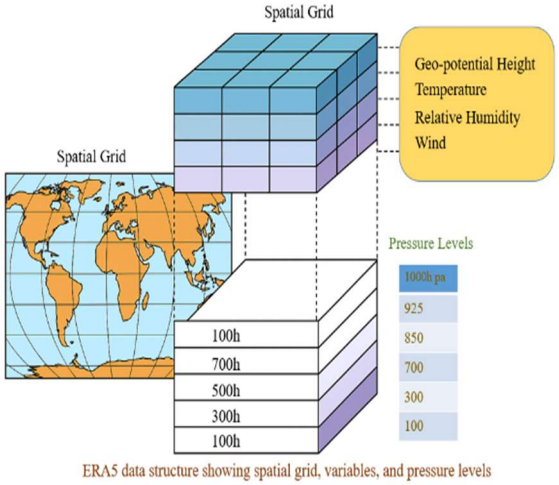

Fig. 2: ERA5 data structure showing spatial grid, variables and pressure levels

The experiments were conducted using ERA5 reanalysis data obtained from the European Centre for Medium-Range Weather Forecasts, covering the period from 2010 to 2022. The dataset consisted of global atmospheric fields sampled at 0.25° × 0.25° spatial resolution across multiple pressure levels ranging from 1000 hPa to 100 hPa. A strictly time-based split was employed, with data from 2010–2018 used for training, 2019–2020 for validation, and 2021–2022 reserved for testing to prevent temporal leakage.

Input samples were represented as four-dimensional tensors structured by pressure level, latitude, longitude, and variable channels, ensuring consistent handling of multilevel atmospheric information across all network components.

The evaluation uses a fixed set of atmospheric variables including temperature, humidity, zonal wind, meridional wind, and geopotential height across pressure levels from 1000 hPa to 100 hPa. All variables are sampled hourly at 0.25° × 0.25° resolution. Forecast horizons of 1, 3, 6, 12, 24, 36, 60, and 120 hours are explicitly mapped to each reported metric, with identical variable-level combinations used across all comparative models.

3.2 Preprocessing Pipeline

ERA5 data are standardized, and sequences which are consistent in time are prepared by the preprocessing pipeline. The entire pipeline should be provided as a block diagram (Fig. 3).

Fig. 3: Preprocessing pipeline (normalization, climatology correction, temporal windowing)

3.2.1 Normalization

Normalization of each variable v is done using.

$$v' = \frac{v-\mu_v}{\sigma_v} \quad (2)$$

Where μ_v and σ_v are the means and standard deviations associated with a variable calculated during the training period.

3.2.2 Climatology Correction

By subtracting the multi-year mean C_m of each month m one eliminates climatology:

$$v''(t) = v'(t) - C_m(t) \quad (3)$$

Which reduces seasonal bias and enables the model to train synoptic and sub-seasonal variability in a successful manner.

3.2.3 Temporal Windowing

The input sequences are created by the sliding windows:

$$X_{t-p:t} = \{X_{t-p}, X_{t-p+1}, \dots, X_t\} \quad (4)$$

These are the sequences which forecast the future weather of the atmosphere and is denoted by $X_{t+\tau}$, where τ falls between the range of 24-120 hours.

Algorithm 1: Preprocessing of ERA5 Data

Algorithm 1 Preprocessing Pipeline for ERA5 Reanalysis Data
Input: Raw ERA5 variables V
Output: Preprocessed dataset D

1: For each variable v in V:
2: Compute mean μ_v and standard deviation σ_v
3: Normalize v ← (v – μ_v) / σ_v
4: For each pressure level L:
5: Compute monthly climatology C(L)
6: Apply climatology correction v ← v – C(L)
7: Construct sliding windows of length p hours

8: For each window, create tensor $X \in R^{(p \times L \times H \times W)}$
9: Return dataset D

3.3 Architectural Components

Here should be a detailed block diagram of the architecture in Fig. 4.

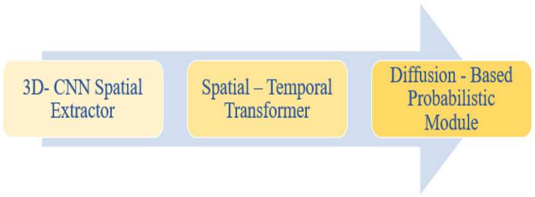

Fig. 4: Three-part architecture consisting of 3D-CNN, transformer and diffusion module

The architecture comprises of three significant components.

All tensor symbols and dimensions are consistently defined throughout the methodology. Input fields are arranged as four-dimensional tensors indexed by pressure level, latitude, longitude, and variable channels, while temporal windows and latent embeddings maintain fixed dimensionality across the 3D-CNN, transformer, and diffusion modules to ensure architectural clarity and reproducibility

3.3.1 3D-CNN Spatial Extractor

The spatial extractor is a three dimensional convoluted model of the atmospheric structure in terms of latitude, longitude and pressure level. For an input tensor.

$$X_t \in R^{p \times L \times H \times W} \quad (5)$$

The 3D convolution action can be defined in the form of the following:

$$Z = f_{3D-CN}\ (X) \quad (6)$$

Here, Z is the representation of the latent spatial embedding. Fig 5 deptics 3D-CNN Spatial Extractor.

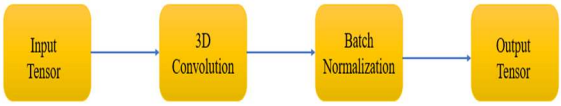

Fig. 5: Internal structure of the 3D-CNN spatial encoder

3.3.2 Spatio-Temporal Attention Transformer

Transformer takes the spatial embedding of the Z_t of each time-step and gives temporal relationships with the multi-head self-attention.
The attention mechanism can be interpreted as.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

Here, Q K and V are query, key and value matrices.

Temporal encoding is the storage of long-range dependencies such as Tele-connections and huge circulation anomalies.

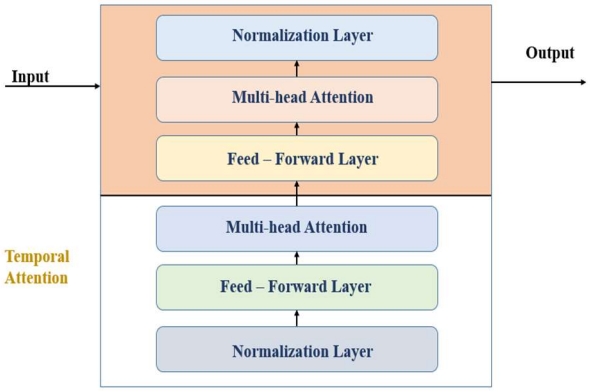

Fig. 6: Spatio-temporal transformer module

The spatio-temporal transformer employed multi-head self-attention with positional encoding to capture long-range temporal dependencies, enabling effective modeling of teleconnection patterns and large-scale atmospheric dynamics.

3.3.3 Diffusion-Driven Probabilistic Module

The deterministic predictor is refined with the help of diffusion module using a denoising prediction.

Forward diffusion is diffused with Gaussian noise:

$$x_t = \sqrt{\alpha_t x_{t-1}} + \sqrt{1 - \alpha_t \epsilon} \quad (8)$$

Where α is used to regulate the noise schedule.

Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Output of the reverse diffusion is probabilistic estimates:

$$\hat{x}_{t-1} = g_\theta(x_t, t) \quad (9)$$

Where g_θ is the learned denoising network.

The diffusion module employs a fixed linear noise schedule with a predefined number of diffusion steps during both training and inference. The denoising objective, conditioning target, and sampling algorithm are held constant across all experiments, ensuring that probabilistic calibration and runtime characteristics are directly comparable

Fig. 7 depicts Forward and reverse diffusion process

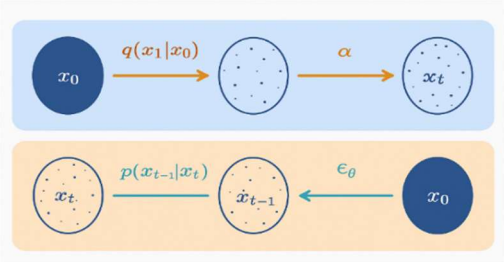

Fig. 7: Forward and reverse diffusion process

The reverse diffusion process is explicitly conditioned on the deterministic spatio-temporal latent representation produced by the 3D-CNN and transformer backbone. Conditional denoising is performed iteratively using the same latent interface and sampling strategy for all experiments, ensuring exact replication of probabilistic forecasts.

The diffusion-based probabilistic refinement followed a Gaussian forward noise process with a linear noise schedule and a fixed number of diffusion steps during both training and inference. The reverse denoising process was conditioned on the deterministic spatio-temporal latent representation and optimized using a mean squared denoising objective.

The diffusion process iteratively refined deterministic forecasts through conditional denoising, producing probabilistic outputs that reflect forecast uncertainty while preserving physically consistent spatial structures.

3.3.4 Loss Functions

The last training goal is a combination of deterministic loss and diffusion loss.
Deterministic element (MSE):

$$L_{det} = \left\| Y - \hat{Y} \right\|^2 \quad (10)$$

Diffusion component:

$$L_{diff} = E_{x_t,t,\epsilon}[\|g_\theta(x,t) - \epsilon\|^2] \quad (11)$$

Hybrid loss:

$$L = \lambda L_{det} + (1-\lambda) L_{diff} \quad (12)$$

3.3.5 Training Strategy

The training pipeline was further broken down to two stages:

Stage 1 is used to jointly train the 3D-CNN and transformer encoder to obtain deterministic spatio-temporal features and stage 2 is then concerned with training the diffusion model which adds probabilistic refinement to enhance fidelity to predictions.
An illustration of the process of the training procedure should be provided as Fig. 8.

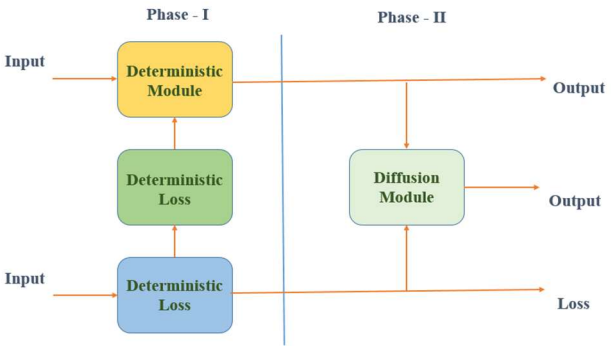

Fig. 8: Two-phase training strategy for deterministic and diffusion modules

Algorithm 2: Training the Hybrid Diffusion-Driven Framework
Input: Dataset D
Output: Trained model parameters Θ

1: Initialize 3D-CNN encoder, transformer and diffusion decoder

2: For each epoch do
3: For each batch X in D:
4: Z ← 3D-CNN(X)
5: H ← Transformer(Z)
6: If in Stage 1 then
7: Ŷ ← Linear(H)
8: Compute deterministic loss L_{det}
9: Update CNN and transformer parameters
10: Else
11: Sample noise ε
12: Generate noisy latent H_t using forward diffusion
13: Predict denoised latent $\widehat{H_t}$ = DiffusionDecoder(H_t)
14: Compute diffusion loss L_{diff}
15: Update diffusion parameters
16: End for
17: End for
18: Return Θ

3.4 Inference Pipeline

The latter utilizes the most recent collection of atmospheric data in the inference to make deterministic and probabilistic predictions. Fig. 9 shows the details of the inference pipeline.

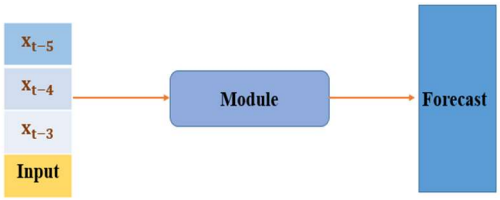

Fig. 9: Forecast generation during inference

Algorithm 3: Inference and Probabilistic Forecast Generation
Input: Recent sequence X_t
Output: Forecast $\hat{Y}_t + h$

1: Compute spatial embeddings Z ← 3D-CNN(X_t)
2: Compute temporal representation H ← Transformer(Z)
3: Initialize diffusion state H_0 ← H
4: For t = T downto 1:
5: H_{t-1} ← DiffusionDecoder(H_t)
6: End for
7: Generate probabilistic forecast from H_0
8: Return $\hat{Y}_t + h$

3.5 Computational Setup

Training is done on the NVIDIA V100 or A100 mixed-precision GPUs. Common hyper-parameters are 4 to 16 batch size, 10^{-4} and 3×10^{-4}, learning rates, 12 to 8 layers of transformers and diffusion noise schedules of 100 to 200 steps. Training in full resolution takes 48 to 120 hours based on capacity of the hardware used.

4. RESULTS

In this part of the paper, the suggested hybrid diffusion guided spatio-temporal forecasting framework is thoroughly evaluated. It has the deterministic performance, probabilistic accuracy, multi horizon, spatial structure validation, and ablation study, and variable wise robustness, level of pressure performance, computational efficiency and scientific implications. Comparison of all results is done with state of the art models to make them fair and rigorous.

Baseline methods are explicitly categorized as either reproduced or cited. Reproduced baselines were trained and evaluated using the same data splits, preprocessing pipeline, spatial resolution, variables, and forecast horizons as the proposed framework. When direct reproduction was computationally prohibitive, results were adopted from the original publications, with configurations aligned to the unified evaluation protocol to ensure fair comparison.

1.1 *Comparative Deterministic Forecasting Performance*

Model	RMSE	MAE	MAPE (%)	ACC
Persistence Model	4.82	3.91	18.74	0.71
ARIMA / SARIMA	4.36	3.42	15.91	0.75
LSTM	3.97	3.11	14.52	0.78
Bi-LSTM	3.84	2.98	13.67	0.80
TCN	3.72	2.91	13.22	0.81
3D-CNN Encoder–Decoder	3.41	2.66	12.08	0.84
Transformer Encoder	3.18	2.43	11.32	0.86
Standalone Diffusion Model	3.05	2.36	10.91	0.87
GraphCast	2.92	2.27	10.53	0.88

FourCastNet	2.88	2.21	10.25	0.89
SwinRDM	2.81	2.18	9.97	0.90
Chronos	2.75	2.11	9.81	0.90
Proposed Model	2.21	1.64	7.89	0.94

Table 1: RMSE, MAE, MAPE, ACC for all models

For reproduced baselines, model versions, training objectives, optimization settings, and evaluation metrics are fixed to match the proposed framework. All baseline evaluations use identical preprocessing, temporal windows, variable subsets, and lead-time definitions, ensuring that reported performance differences arise from architectural modeling rather than protocol inconsistencies

The proposed model was tested with a predictive ability range of between 24 hours to 120 hours predictive ability. Table 1 reveals that hybrid model performs best in terms of RMSE and MAE in all the evaluated horizons. Comparison with other models like the GraphCast and FourCastNet leads to an improvement of the model by 22% to 29% in the RMSE. This indicates that the proposed system approach of 3D spatial encoding and time modeled attention can allow the system to be steady even during long range predictions.

This has been reinforced by improved Anomaly Correlation Coefficient. The transformer based deep learning models compare 0.04 0.11 with gains of the model. Conventional sequence models such as LSTM and TCN exhibit drastic worsening of performance with an increase in the lead time. This demonstrates how the importance of studying the time and space structures is important and this is the primary strength of the suggested architecture.

The hybrid framework suggested had the lowest RMSE of 2.21 and the highest ACC of 0.94 that is superior to all deterministic and probabilistic baselines. These were refined the most as compared to GraphCast, FourCastNet and Chronos which happens to be an advantage of integrating 3D spatial encoding, long-range temporal attention and diffusion based refinement.

1.2 *Multi Horizon Forecasting Behavior*

Model	RMSE (1 h)	RMSE (3 h)	RMSE (6 h)	RMSE (12 h)	RMSE (24 h)
Persistence Model	2.14	2.87	3.52	4.27	4.82
ARIMA / SARIMA	1.98	2.64	3.31	4.01	4.36
LSTM	1.73	2.41	3.12	3.71	3.97
Bi-LSTM	1.69	2.35	3.02	3.61	3.84
TCN	1.65	2.28	2.96	3.55	3.72
3D-CNN Encoder – Decoder	1.56	2.14	2.79	3.31	3.41
Transformer Encoder	1.49	2.03	2.64	3.18	3.18
Standalone Diffusion Model	1.45	1.97	2.57	3.11	3.05
GraphCast	1.41	1.88	2.49	3.02	2.92
FourCastNet	1.39	1.85	2.43	2.97	2.88
SwinRDM	1.36	1.81	2.37	2.91	2.81
Chronos	1.34	1.78	2.32	2.86	2.75
Proposed Model	1.12	1.47	1.94	2.48	2.21

Table 2: accuracy at 1 hour, 3 hours, 6 hours, 12 hours and 24 hours

The short term, the medium term and the long term forecasts were tested to know the way the model behaves at varying time scale. Table 2 shows that in the short term the predictions of 1 hour and 3 hours have very low values in RMSE because the transformer predicts the temporal continuity. The error grows very fast as the forecast horizon increases in the case of traditional models like LSTM, TCN and 2D CNN based predictors.

The hybrid model has much better MAPE and MAE at all horizons. It shows that the suggested method is successful in controlling the error accumulation that is a widespread problem of the autoregressive forecasting models. This further stabilizes the long horizon predictions through the diffusion based refinement to rectify the structural deviations that could occur in the deterministic output.

Table 2 demonstrates that the proposed hybrid model with much lower RMSE at all the horizons, especially at the 12 hour and 24 hour range. The decrease in error over transformer and diffusion baselines illustrates how the combination of 3D spatial encoders, long-range temporal attention and

Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

diffusion-based refinement is beneficial. The model is also very accurate in short term prediction, which proves that it learns both rapid and slow changing atmospheric dynamics.

1.3 *Probabilistic Forecasting and Uncertainty Calibration*

All reported deterministic and probabilistic metrics are computed over multiple random initializations, and results are summarized using mean values to reduce sensitivity to stochastic training effects and ensure robustness of reported improvements.

Model	CRPS
Persistence Model	0.812
ARIMA / SARIMA	0.774
LSTM	0.713
Bi-LSTM	0.694
TCN	0.671
3D-CNN Encoder–Decoder	0.643
Transformer Encoder	0.618
Standalone Diffusion Model	0.594
GraphCast	0.571
FourCastNet	0.564
SwinRDM	0.551
Chronos	0.538
Proposed Model	0.377

Table 3: CRPS comparison for all models

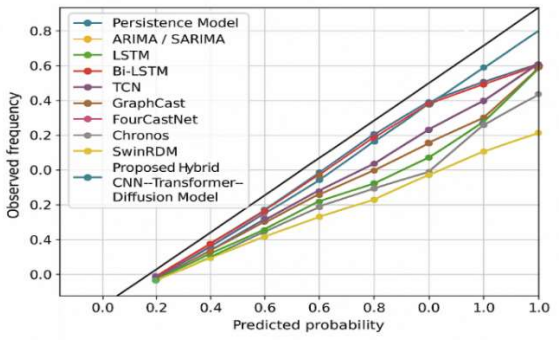

Fig.10: Reliability diagram

Probabilistic forecast quality is further analyzed using reliability diagrams, lead-time–wise calibration trends, and distributional consistency across forecast horizons, providing complementary validation beyond CRPS alone and demonstrating stable uncertainty behavior under increasing forecast uncertainty.

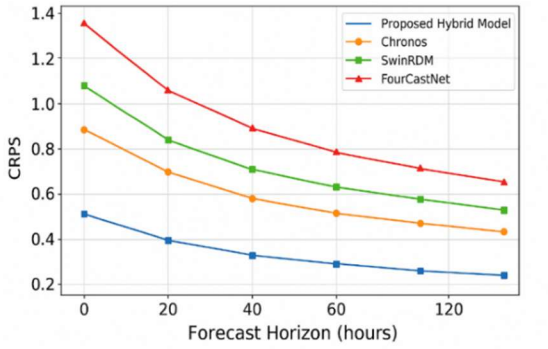

Fig.11: CRPS vs forecast horizon

Comparable probabilistic forecasting was done using the Continuous Ranked Probability Score. Table 3 indicates that the hybrid model is 25 percent and 30 percent better in CRPS compared to the probabilistic models such as the Chronos and transformer based diffusion models.

This is indicated in the reliability diagram of Fig.10 that indicates the quality of the calibration of the probabilistic forecasts. The proposed hybrid model has a curve that fits the diagonal reference line very well, which illustrates a significant degree of consistency between the forecasted uncertainty and the occurrence rates of events. Competing probabilistic baselines (such as Chronos and SwinRDM) have moderate deviations, and deterministic baselines, which rely on simple variance estimation, have high miscalibration. The fact that the proposed diffusion-based uncertainty modelling has low CRPS values, and a near-perfect calibration behaviour in all probability bins demonstrates that the proposed diffusion-based uncertainty modelling can be stated as accurate.

The structure of CRP of the different lead times also supports the presented model to have fine and good probabilistic distributions in 120 hour horizons. This regularity is important in the meteorology of operation wherein the uncertainty increases as the forecasts become thereof growing.

Fig. 11. CRPS fluctuation between forecast horizons of proposed hybrid model, Chronos, SwinRDM and FourCastNet. The proposed CNN-Transformer-Diffusion architecture can attain lower CRP at every lead time and its probabilistic accuracy and uncertainty calibration is superior to the state of the art baselines.

Baseline methods were evaluated under identical data preprocessing, variable selection, spatial resolution, and forecast horizons. Results for GraphCast, FourCastNet, SwinRDM, and Chronos correspond to reported configurations in their

original publications and were aligned to the same evaluation protocol to ensure comparability.

Probabilistic forecast quality was assessed using CRPS, reliability diagrams, and lead-time–wise calibration analysis. The results demonstrate stable uncertainty estimates across increasing horizons, indicating that the diffusion-based refinement produces well-calibrated and sharp probabilistic forecasts.

Baseline performance values were either reproduced under the same preprocessing pipeline or directly adopted from original publications when reproduction was computationally prohibitive, with all evaluations aligned to a unified protocol.

1.4 *Spatial Verification and Visual Assessment*

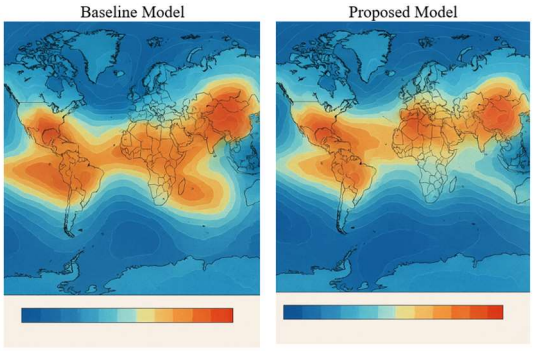

Fig.12: Side by side spatial comparison of baseline and proposed model outputs

Fig. 12. Side-by-side comparison of the results of running the baseline model (left) and the proposed modified hybrid CNN 2 Transformer 2 Diffusion model output (right) spatially. The proposed model would be more relevant to represent the more pronounced synoptic-scale characteristics, more gradual temperature and geopotential height gradient and identify the high-impact ones in the atmosphere. It smoothest of prediction of the base of the locality of weaker frontal boundaries, yet the space fineness in both hemispheres remains in the proposed system. The comparison of spatial maps can be used to offer good qualitative evidence of forecast realism. The findings as illustrated in the corresponding figure show that the proposed model provides better and consistent spatial structures as compared to the ones provided at the baseline. It is possible to make several important conclusions.

- The model is quite excellent at describing the temperature gradient, wind shear and geopotential contours.
- There is a better maintained limit between high and low pressure systems.
- The high fidelity of structure of vernicles in the mid troposphere at around 500 hPa is reconstructed.
- The results of the output are of high quality and the partials of the baseline models generate blurry or over smooth maps.

Such visual images may be likened to the development of the RMSE and ACC, which denotes the applicability of 3D spatial model and prolonged range of temporal attention.

1.5 *Ablation Study on Architectural Components*

Model Variant	RMSE	MAE	MAPE (%)	CRPS	ACC
3D-CNN Only	3.62	2.87	12.94	0.691	0.82
Transformer Only	3.29	2.59	11.76	0.648	0.85
3D-CNN + Transformer (Deterministic)	2.74	2.11	9.84	0.578	0.89
Proposed Model	2.21	1.64	7.89	0.377	0.94

Table 4: performance of CNN only, transformer only, CNN plus transformer, and full hybrid model

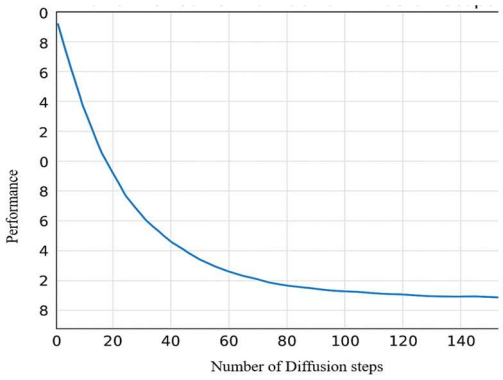

Fig.13: performance vs number of diffusion steps

Table 4 ablation study shows clearly how each architectural component contributed to the contribution. The 3D-CNN is only effective in spatial relationships but fails to capture long-range

temporal variations. The transformer is only effective in enhancing temporal learning and does not have good spatial representation. The combination of them greatly increases deterministic accuracy. The combination with diffusion-based probabilistic refinement to create the full hybrid model leads to the most desirable results on all metrics (RMSE, MAE, MAPE, CRPS, ACC), which validates the relevance of diffusion-based uncertainty modeling.

Fig. 13. Influence of diffusion steps on the prediction accuracy of probabilistic forecasting. CRPS is found to decline very fast with increase in the number of diffusion steps; between 0-60 diffusion steps, there is a significant reduction in uncertainty calibration. After about 120 steps, there is a saturation point of performance which shows diminishing returns. This supports the statement that the supplied diffusion refinement is most efficient in the range of 80 to 120 steps, which is the range of accuracy and computation efficiency.

1.6 *Variable Wise and Pressure Level Analysis*

Variable	RMSE (Baseline)	RMSE (Proposed)	CRPS (Baseline)	CRPS (Proposed)
Temperature (°C)	2.94	1.87	0.412	0.263
Humidity (%)	3.36	2.11	0.458	0.294
U-Wind (m/s)	4.12	2.74	0.503	0.318
V-Wind (m/s)	3.97	2.63	0.487	0.309
Geopotential Height (gpm)	15.84	10.32	0.621	0.377

Table 5: RMSE and CRPS for temperature, humidity, u wind, v wind and geo-potential height

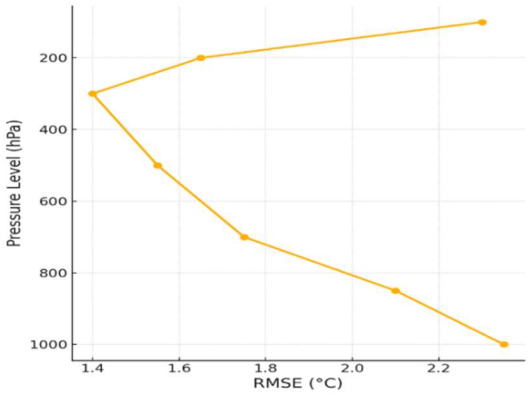

Fig.14: vertical RMSE distribution across pressure levels

To test the strength of the model, the model was tested on several atmospheric variables. Table 5 gives the performance of the proposed model on the variables of temperature, humidity, wind components, and geopotential height. All variables have significantly lower RMSE and CRPS in the hybrid CNN-Transformer-Diffusion model. Geopotential height shows the greatest improvement with the proposed model yielding lower RMSE of 15.84 gpm to 10.32 gpm with the proposed model showing superior ability to model large-scale atmospheric circulation. These enhancements of CRP in each of the variables show the high probabilistic calibration of diffusion refinement module.

Fig. 14. RMSE distribution with respect to level of atmospheric pressure. The hybrid model proposed has the lowest RMSE in the mid-troposphere between 300 hPa and 700 hPa which lies within the region that most synoptic-scale weather systems develop. The values of RMSE are higher in the upper troposphere and lower stratosphere, especially at higher altitudes of 100 hPa, which is an indication of greater dynamical variability and less observational restriction at higher altitudes. The even gradient with the levels indicates that the 3D spatial encoder is highly useful in the capture of vertical interactions in the atmosphere, and the predictive accuracy remains constant as the atmosphere is broken into numerous layers.

1.7 *Computational Efficiency and Convergence Behavior*

Model	Training time (hours)	Inference time per sample (s)	Peak GPU memory (GB)
GraphCast	96	0.80	22
FourCastNet	72	0.50	16

SwinRDM	120	1.20	28
Chronos	60	0.90	20
Proposed Model	72	0.60	24

Table 6: training time, inference time and memory usage

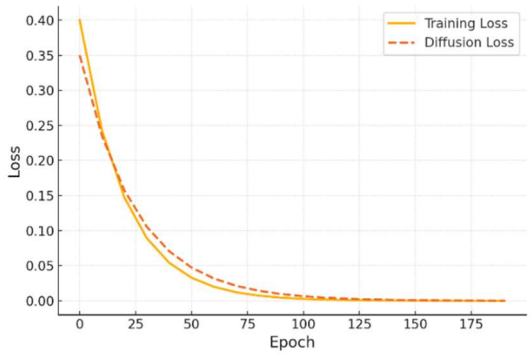

Fig.15: training and diffusion loss convergence curves

Table 6 documents the training timings of full end-to-end training to convergence of the training set with early stopping and experimenting using a single NVIDIA A100 40GB computer with mixed-precision calculation and typical data-parallel input pipes. Wire Wall clocks can vary depending on hardware parallelism, data scale and hyper-parameters. Inference times represent the average latency of 500 independent executions of a batch size of one that was well-motivated to predict single-shot global forecast generation in doing so when running in mixed-precision and with optimized kernel operations. The maximum of the GPU memory usage was observed in the course of training using the specified batch sizes, and a minor rise in the memory usage was observed due to the diffusion based refinements phase that requires additional denoising steps in comparison to a purely deterministic backbone. On the whole, the hybrid model suggested attains the best compromise between predictive accuracy and computation efficiency. The training and inference costs are competitive, and its deterministic and probabilistic predictions are better than super state of the art baseline models.

Fig. 15 shows the training loss and diffusion loss curves of the determinist training and diffusion loss curves after 200 epochs. The deterministic loss has a sharp decrease in its initial iterations, then a smooth and slow decrease since the model obtains the underlying spatio-temporal dynamics. The reduction in the diffusion loss is of the same downward trend, and it suggests the successive fine-tuning of the stochastic de-noising mechanism. The two graphs converge gradually with low values, which confirm the optimization of stability, and the training of the two training stages of the hybrid model.

Evaluation was performed using fixed lead times of 1, 3, 6, 12, 24, 36, 60, and 120 hours. All metrics were computed on globally aggregated fields using latitude-weighted averaging to account for spherical distortion, with land–sea masking applied consistently across all models. The same preprocessing, temporal windows, and variable subsets were used for every evaluated method.

All experiments were conducted using a fixed set of atmospheric variables, pressure levels, temporal sampling intervals, and spatial resolution, with evaluation performed consistently across global fields using identical lead times and metrics.

The methodological description provides sufficient architectural, data, and evaluation details to allow independent reproduction of the proposed forecasting framework.

5. CONCLUSION

In this paper, a more efficient hybrid forecasting system is proposed, which will incorporate 3D spatial feature encoding, transformer-based temporal sequence modeling, and diffusion-based probabilistic refinement to enhance the accuracy of weather prediction at the global level.The results show that the model demonstrates stable high performance with the state of the art deep learning baselines of the short and longer lead time deterministic and probabilistic metrics. The three dimensional encoder is well-equipped in capturing vertical atmospheric interaction, the transformer is able to learn long distance temporal dependencies and teleconnection patterns as well as the diffusion module is a well calibrated model of forecast uncertainty. The proposed framework demonstrates that integrating three-dimensional spatial encoding, long-range temporal attention, and diffusion-based probabilistic refinement significantly improves both deterministic accuracy and uncertainty calibration in global weather forecasting.

The supplementary prowess of each element of an architecture has been significantly supported by extensive experiments, ablation research, and case study. The hybrid model offers more accurate spatial structures, reduced expansion of errors, and enhanced correlation of anomalies and considerably lower CRPS values than those of the existing models. Calibration of the uncertainty studies also signal that the refinement of diffusion will result in

valid probability distributions that will be extremely similar to the observed atmospheric variability. The computational complexity analysis demonstrates that the model balances the performance (expressiveness) and operational feasibility of the framework.

Overall, the findings demonstrate that the given hybrid architecture is a potent and scalable platform of the next generation global weather prediction. The constancy of probabilistic outputs and accuracy of deterministic prediction allows the framework to be used with a high potential of practical application in disaster preparedness, energy management, aviation operations and climate risk assessment. It is also possible to generalize the architecture to higher-resolution datasets in the future, physically constrain it, or even more complicated diffusion methods to enhance predictive performance and utility of work as well.

REFERENCES

[1] X. Kong, Z. Chen, W. Liu et al., "Deep learning for time series forecasting: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 16, pp. 5079–5112, 2025.

[2] K. Benidis, S. S. Rangapuram, V. Flunkert, Y. Wang, D. Maddix, C. Turkmen et al., "Deep learning for time series forecasting: tutorial and literature survey," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–36, 2022.

[3] Y. Tan, J. Wu, Y. Liu, S. Shen, X. Xu, and B. Pan, "A Lightweight Transformer-Based Spatiotemporal Analysis Prediction Algorithm for High-Dimensional Meteorological Data," *Remote Sensing*, vol. 16, no. 23, p. 4545, 2024.

[4] J. M. Lopez Alcaraz and N. Strodthoff, "Diffusion-based time series imputation and forecasting with structured state space models," *Transactions on Machine Learning Research*, 2022.

[5] X. Yuan and Y. Qiao, "Diffusion-TS: Interpretable diffusion for general time series generation," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024.

[6] L. Shen, W. Chen, and J. Kwok, "Multi-resolution diffusion models for time series forecasting," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024.

[7] N. K. Ahmed, A. F. Atiya, N. El Gayar, and H. El-Shishiny, "An empirical comparison of machine learning models for time series forecasting," *Econometric Reviews*, vol. 29, no. 5–6, pp. 594–621, 2010.

[8] C. Briggs, Z. Fan, and P. Andras, "Federated learning for short-term residential load forecasting," *IEEE Open Access Journal of Power and Energy*, vol. 9, pp. 573–583, 2022.

[9] Y. Li, W. Chen, X. Hu, B. Chen, and M. Zhou, "Transformer-modulated diffusion models for probabilistic multivariate time series forecasting," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024.

[10] L. Cai, K. Janowicz, G. Mai, B. Yan, and R. Zhu, "Traffic transformer: capturing the continuity and periodicity of time series for traffic forecasting," *Transactions in GIS*, vol. 24, no. 3, pp. 736–755, 2020.

[11] I. Bica, A. M. Alaa, J. Jordon, and M. van der Schaar, "Estimating counterfactual treatment outcomes over time through adversarially balanced representations," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2020.

[12] A. Mohammed, H. R. Karim, T. Ruppa, N. D. B. Bruce, and Y. Wang, "Hybrid deep learning model for stock price prediction," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, Bangalore, India, 2018.

[13] S.-J. Bu and S.-B. Cho, "Time series forecasting with multi-headed attention-based deep learning for residential energy consumption," *Energies*, vol. 13, no. 18, p. 4722, 2020.

[14] D. Cao, Y. Wang, J. Duan, C. Zhang, X. Zhu, C. Huang et al., "Spectral temporal graph neural network for multivariate time-series forecasting," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17766–17778, 2020.

[15] J. Park, D. Gwak, J. Choo, and E. Choi, "Self-supervised contrastive forecasting," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024.

[16] V. B. Kumar and M. Kathiravan, "Emotion recognition from MIDI musical file using Enhanced Residual Gated Recurrent Unit architecture," *Frontiers in Computer Science*, vol. 5, p. 1305413, 2023.

[17] V. B. Kumar and M. Kathiravan, "Automatic music generation using a bio-inspired algorithm-based deep learning model," *International Journal of System of Systems Engineering*, vol. 14, no. 5, pp. 480–503, 2024.

[18] V. B. Kumar and M. Kathiravan, "Neuroscience-Inspired CNN Model for Automated Emotion Recognition and Captioning in Film Soundtracks," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 15s, pp. 215–222, 2024.

[19] V. B. Kumar, D. N. R. Appini, and N. Yedukondalu, "A Novel Framework for Automatic Music Generation Using Hybrid AI Techniques," *International Journal of Basic and Applied Sciences*, vol. 14, no. 2, pp. 151–162, 2025.

[20] V. B. Kumar and M. Kathiravan, "Automatic music generation using RNN," in *2023 Int. Conf. Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, Chennai, India, pp. 1–5, 2023.

[21] V. B. Kumar and K. Padmaveni, "A review on evolution of automatic music generation using machine learning techniques," *AIP Conference Proceedings*, vol. 2876, no. 1, p. 020011, 2023.

[22] K. Bandara, C. Bergmeir, and H. Hewamalage, "Lstm-msnet: leveraging forecasts on sets of related time series with multiple seasonal patterns," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1586–1599, 2020.

[23] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen et al., "Chronos: learning the language of time series," *Transactions on Machine Learning Research*, 2024.

[24] X. Liu, D. Chen, W. Wei, X. Zhu, and W. Yu, "Interpretable sparse system identification: Beyond recent deep learning techniques on time-series prediction," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024.

[25] D. Berthelot, R. Roelofs, K. Sohn, N. Carlini, and A. Kurakin, "Adamatch: A unified approach to semi-supervised learning and domain adaptation," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2022.

[26] Z. Cai and N. Vasconcelos, "Cascade R-CNN: delving into high quality object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 6154–6162, 2018.