Journal of Robotics and Control (JRC)
Volume 5, Issue 3, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i3.21716

Enhancing Long-Term Air Temperature Forecasting with Deep Learning Architectures

Denis Krivoguz [1], Alexander Ioshpa [2], Sergei Chernyi [3*], Anton Zhilenkov [4], Aleksandr Kustov [5], Anton Zinchenko [6], Pavel Podelenyuk [7], Polina Tsareva [8]
[1, 2] Oceanology Department, Southern Federal University, Rostov-on-Don, Russia
[3, 4, 5, 6, 7, 8] St. Petersburg State Marine Technical University, St. Petersburg, Russia
Email: [1] krivoguz@sfedu.ru, [2] ioshpa@sfedu.ru, [3] sergiiblack@gmail.com, [4] zhilenkovanton@gmail.com, [5] kustov@smtu.ru, [6] antel85@bk.ru, [7] podelenyk@smtu.ru, [8] tsareva@smtu.ru
*Corresponding Author

***Abstract*—Modern challenges in climate prediction necessitate the adoption of advanced deep learning architectures for enhanced precision in temperature forecasting. This study undertakes a comparative evaluation of various neural network designs, particularly focusing on Deep Recurrent Neural Networks (DRNN) and their extension with Gated Recurrent Units (DRNN-GRU), chosen for their proven efficacy in sequential data analysis and long-term dependency capture. Leveraging a comprehensive meteorological dataset, collected from 1961 to 2023, which includes atmospheric temperature, pressure, and precipitation levels, the research unfolds a nuanced understanding of the climate variability. The evaluation framework rigorously applies Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) metrics to quantify model performance. The DRNN and DRNN-GRU architectures are distinguished for their superior predictive accuracy, suggesting their high potential for real-world forecasting applications. These findings are not merely academic; they imply substantial practical implications, particularly for geographic information systems where they can enhance climate monitoring and resource management. The paper culminates with recommendations for dataset expansion and diversified analytical techniques, which are critical for refining the predictive prowess of these models. This research thereby sets a benchmark for future explorations in the field and directs towards innovative avenues to augment the scientific understanding of climate dynamics.**

***Keywords*—Deep Learning; Neural Networks; Temperature Forecasting; Meteorological Data; DRNN.**

I. Introduction

Air temperature forecasting holds significant strategic and applied importance in today's world due to several pivotal factors that directly or indirectly impact humanity and the planet's ecosystem [1]. In this context, temperature prediction becomes an indispensable tool for scientific research and practical applications.

Temperature conditions are a crucial component of climatic dynamics, and accurate forecasting of temperature variations allows for the analysis of the influence of climatic factors on the planet [2]. Studying climate changes based on long-term air temperature data provides scientists with the opportunity to devise adaptation and mitigation strategies against climatic anomalies [3]. Temperature forecasting finds direct application in sectors where temperature conditions are pivotal, such as agriculture [4]. Knowledge of future temperatures enables agricultural producers to make informed decisions regarding planting, irrigation, and harvest timings, thereby optimizing food production [5].

Social aspects also play a significant role, as accurate temperature forecasts allow adaptation to extreme temperature conditions, including heatwaves and cold snaps [6]. Effective planning and response to such events contribute to the reduction of human casualties and material losses.

In the energy and resource supply sector, temperature forecasting plays a notable role [7], [8]. For instance, managing power systems requires considering the impact of temperature on electricity consumption [9]. Analyzing temperature data is also integral to managing heating and water supply [10].

Existing air temperature forecasting methods have their inherent challenges and limitations, crucial to consider in the context of developing and refining predictive models [11], [12], [13]. A primary issue is the limited ability of traditional methods to account for the intricate and continually changing interrelations between various factors influencing temperature conditions [3]. Such methods often struggle to adapt to rapidly changing climatic conditions and account for nonlinear dynamics [14], [15]. Another significant issue is the limited spatial and temporal resolution of traditional forecasting methods. In most cases, they operate with restricted data, complicating accurate temperature prediction on fine spatial and temporal scales. This is especially critical when forecasting extreme events like frosts, heatwaves, and floods [16], [17].

Traditional methods are also susceptible to limitations in accounting for the complex interrelations between various natural processes influencing temperature conditions [3]. For instance, they might underestimate the impact of land-use change on temperature, critically important for understanding global climate changes [18], [19], [20]. Moreover, traditional methods might exert limited influence on enhancing the accuracy and reliability of forecasts, especially when analyzing large data volumes with high variability. Methods like statistical and empirical models often have a restricted ability to capture intricate nonlinear dependencies.

The challenges and limitations associated with traditional temperature forecasting methods underscore the imperative need for the development and application of more advanced and precise forecasting techniques, particularly those based on state-of-the-art deep neural network architectures. Approaches leveraging these architectures can represent a significant advancement in the domain of temperature condition forecasting.

The significance and relevance of employing deep neural networks, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), in the context of air temperature forecasting are underpinned by several fundamental factors. Foremost, deep neural networks possess the capability to automatically extract intricate hierarchical dependencies from data, becoming good tools in time series analysis [21], [22], [23]. This is especially pivotal when working with meteorological data, characterized by interrelations and nonlinear dependencies among various meteorological variables.

A key aspect underscoring the importance of deep neural networks is their aptitude to account for long-term dependencies in data. LSTM and GRU, as recurrent neural network architectures, are designed to retain and update information over extended temporal intervals, enabling them to capture long-term patterns and trends in temperature data [21], [24]. This becomes especially crucial when forecasting climatic changes and extreme weather events with long-term results.

Deep neural networks also enhance forecast accuracy through their inherent adaptability to evolving conditions. Their capacity to train on vast data volumes allows them to discern minute details in weather variations [25], [26]. Furthermore, their ability to process data of diverse natures, encompassing climate change information, land-use alterations, and other factors, renders them promising in integrating multifarious information sources for more accurate predictions.

It's noteworthy that deep neural networks can substantially mitigate the impacts of spatial and temporal data resolution constraints. Their memory and analysis capabilities across numerous time steps facilitate predictions with high sensitivity to changes on temporal and spatial scales [21]. This opens new avenues for enhancing the precision and predictive power of models.

Recent advances in deep learning have markedly improved air temperature forecasting, offering nuanced insights into the accuracy and applicability of various models. For instance, studies such as the [27] and [28] illustrate the efficacy of deep learning approaches in capturing complex climatic patterns. These studies underscore the significance of employing models like U-Net and artificial intelligence algorithms for forecasting, emphasizing their potential in improving predictive accuracy over traditional statistical methods. However, while these works have laid a solid foundation, the presented study introduces a novel comparative analysis focusing on the specific efficacy of deep recurrent neural networks (DRNN) and DRNN with Gated Recurrent Unit (DRNN-GRU) models against those incorporating attention mechanisms for air temperature forecasting. This comparative insight is particularly novel in identifying the DRNN and DRNN-GRU models as not only highly accurate but also more contextually adaptable for real-world applications in sectors like agriculture, energy management, and environmental monitoring. By directly comparing these models within the framework of air temperature forecasting, this research fills a critical gap, offering both a methodological advancement in forecasting technology and practical recommendations for its application across various sectors affected by climatic conditions. This unique contribution, set against the backdrop of existing machine learning approaches to climate forecasting, highlights the practical implications of choosing specific deep learning architecture based on their performance and applicability to real-world scenarios.

The dataset used in this study is derived from the meteorological station affiliated with the Institute of Earth Sciences at Southern Federal University, covering an extensive period from 1961 to 2023. This dataset's temporal span is particularly noteworthy, offering a rich historical record that captures a wide range of climatic variability. Such a comprehensive dataset is crucial for training deep learning models to recognize and predict complex patterns over long temporal scales, which is often a challenge with shorter or less diverse datasets. The inclusion of variables such as atmospheric air temperature, partial pressure, and atmospheric precipitation levels adds to the dataset's uniqueness, providing a multifaceted view of the climatic conditions that influence air temperature. These characteristics make the dataset an ideal candidate for exploring the efficacy of deep neural network architectures in forecasting air temperature with high precision.

This research explores the utility of advanced deep learning architectures, including Deep Recurrent Neural Networks (DRNN) and DRNN with Gated Recurrent Unit (DRNN-GRU), in the domain of air temperature forecasting. Deep learning, a subset of machine learning, excels in identifying complex patterns in large datasets, making it particularly suitable for climate data characterized by nonlinear relationships and high-dimensional variability. The chosen architectures, DRNN and DRNN-GRU, are renowned for their ability to process sequential data, capturing temporal dependencies that are critical in forecasting tasks.

The DRNN model, with its multiple layers of recurrent units, is designed to model time series data efficiently, making it well-suited for analyzing the temporal sequences found in meteorological data. The addition of Gated Recurrent Units (GRU) in the DRNN-GRU model enhances the network's ability to remember long-term dependencies, addressing one of the critical challenges in temperature forecasting: capturing the influence of past climatic conditions on future temperatures. This capability is pivotal for improving the accuracy of long-term forecasts, as it allows for a more nuanced understanding of climatic trends and patterns.

The deployment of these deep learning methodologies aligns closely with the study's objectives to enhance the accuracy and reliability of long-term air temperature forecasts. By harnessing the predictive power of DRNN and DRNN-GRU models, this research aims to overcome the limitations of traditional forecasting methods, which often struggle with the complex dynamics of climate systems. The anticipated outcome is not only methodological advancement in the field of climate science but also a practical tool for stakeholders in agriculture, energy, and urban planning to make informed decisions based on accurate and reliable temperature forecasts.

II. Materials and Methods

For the subsequent analysis and training of the forecasting models, a series of preprocessing steps were undertaken. The foundational data consisted of climatic observations, encompassing information on temperature, precipitation, and atmospheric pressure.

In the initial phase, the parameters of the data sequence, essential for neural network training, were delineated. The length of the step sequence (sequence_length), was determined to be 7 steps. Additionally, the number of features (num_features), was established to be 3, encompassing temperature, precipitation, and pressure. Subsequently, data normalization was executed utilizing the Min-Max Scaling method [27]-[35]. This procedure facilitated the scaling of all features to a uniform range, thereby enhancing the convergence efficiency of the neural network training process. Post-normalization, sequences and their corresponding target labels were generated. By iterating through the normalized data and forming sequences of a length of 7-time steps, the foundational data was transformed into a format appropriate to training. Each sequence encapsulated information on temperature, precipitation, and pressure over 7-time steps, while the corresponding target label represented the temperature for the ensuing time step.

The derived sequences and labels were then converted into numpy arrays for further utilization. To evaluate the efficacy of the models, the data was divided into training and testing datasets. The training dataset constituted 80% of the total data, whereas the testing dataset was employed to assess the proficiency of the trained models.

A. Architecture of the Models

The Deep Recurrent Neural Network. One of the neural network architectures created is a variation of the deep recurrent network, termed the “Deep Recurrent Neural Network” (DRNN) [36]-[45]. This architecture is created to accommodate the peculiarities of temporal data, such as climatic observations, and is designed to analyze sequences of variable lengths.

The DRNN encompasses multiple Long Short-Term Memory (LSTM) layers, each endowed with a distinct number of hidden units. This configuration empowers the model to discern intricate temporal dependencies and interrelationships among various climatic data features. Each LSTM layer relates to a Dropout layer, with a dropout rate of 0.2. This inclusion aids in mitigating the model's susceptibility to overfitting. Such an approach ensures a harmonious balance between the model's complexity and its generalization capability. In its final phase, the DRNN ends with a fully connected Dense layer, equipped with a single output unit. This is purposed for forecasting the temperature for the subsequent time step. For the model's training phase, the “Adam” optimizer was employed in conjunction with the Mean Squared Error loss function. This combination facilitates the minimization of the mean squared error between the model's temperature predictions and the actual temperature values.

The Deep Recurrent Network with Gated Recurrent Unit. The second neural network architecture presented in this research is characterized as the “Deep Recurrent Network with Gated Recurrent Unit” (DRNN-GRU). This architecture is created for the task of air temperature forecasting, leveraging temporal data amassed from a meteorological station [46]-[51].

The DRNN-GRU encompasses multiple Gated Recurrent Unit (GRU) layers, each endowed with a distinct number of hidden units. The GRU is a variant of recurrent neural networks, purposed for processing temporal sequences. They are distinguished by their capability to model long-term dependencies in data, a pivotal facet in the analysis of climatic time series. Each GRU layer relates to a Dropout layer, with a dropout rate of 0.2. This inclusion aids in mitigating the model's susceptibility to overfitting, thereby enhancing its generalization capability. In its final stage, the DRNN-GRU ends in a fully connected Dense layer with a singular output unit. As an optimizer model used “Adam” and loss function, based on MSE.

The Long Short-Term Memory with Attention Mechanism (LSTM-Attention). The third neural network architecture analyzed in this research called “Long Short-Term Memory with Attention Mechanism” (LSTM-Attention).

The model initiates with an input layer, representing temporal data sequences, considering the predefined sequence length and the number of features. After this, the architecture incorporates an LSTM layer with 64 units, which returns sequences. The LSTM is a variant of recurrent neural networks, renowned for their capability to model long-term dependencies in data. Main feature of this architecture is the attention mechanism. This mechanism furnishes the model with the capacity to focus on specific segments of the data. Such a capability can be critically pivotal in enhancing forecasts, especially in scenarios inundated with interrelated variables.

The outputs from the LSTM layer and the attention mechanism are concatenated into a unified sequence. This amalgamation empowers the model to discern both short-term and long-term dependencies within the data. The architecture ends with a fully connected Dense layer with a singular output unit, purposed for the task of temperature forecasting. The LSTM-Attention architecture also employs the "Adam" optimizer and the MSE loss function during training, facilitating the minimization of discrepancies between the forecasted and actual temperature values.

Robust Stacked LSTM (RSLSTM). The fourth neural network architecture delineated in this research is characterized as the “Robust Stacked LSTM” (RSLSTM). The model initiates with an input layer, designed to accept input data in the form of temporal sequences, considering the predefined sequence length and the number of features. After this, the architecture incorporates a primary LSTM layer with 128 units, which returns sequences. At this juncture, methodologies such as Dropout (with a coefficient of 0.2) and

Batch Normalization are employed. The Dropout helps in mitigating the risk of model overfitting, while Batch Normalization contributes to the stabilization of the training process.

Following this, the architecture integrates a secondary LSTM layer with 64 units, also augmented with Dropout (0.2) and Batch Normalization. While additional layers can be incorporated at the discretion of the researcher, in this architecture, only an output Dense layer is presented with a single unit, purposed for the task of temperature forecasting. The RSLSTM architecture also employs the “Adam” optimizer and the MSE loss function during training, facilitating the minimization of discrepancies between the forecasted and actual temperature values.

Robust Stacked LSTM with Attention Mechanism. The fifth neural network architecture presented in this research is termed Robust Stacked LSTM with Attention” (RSLSTM-Attention). The model commences with an input layer, designed to accommodate temporal data sequences, factoring in the predetermined sequence length and the number of features. After this, the architecture integrates three sequential LSTM layers: the primary LSTM layer with 128 units, which returns sequences; the secondary LSTM layer with 64 units, also returning sequences; and the tertiary LSTM layer with 32 units, which too returns sequences. These layers are pivotal in extracting temporal dependencies and key features from the input data. Following this, an attention mechanism is embedded between the last LSTM layer and the input. This mechanism endows the model with the capability to focus on the most pertinent segments of the input data, considering their respective weights [52]-[55].

After this, a concatenation of the outputs from the last LSTM layer and the attention mechanism is executed, combining the information. Additionally, a Flatten layer is incorporated to transform the output into a two-dimensional format. The RLSTM-Attention architecture employs the “Adam” optimizer and the MSE loss function during its training phase, with the overarching objective of minimizing discrepancies between the forecasted and actual temperature values.

The selection of these architectures was guided by their theoretical capabilities and empirical successes in processing sequential and time-series data, which are critical in forecasting tasks such as air temperature prediction.

The choice of DRNNs was motivated by their design to handle sequential data, making them particularly apt for time-series forecasting tasks like air temperature prediction. DRNNs can learn from the temporal dependencies of data, which is crucial for understanding the dynamics of climate variables over time. Theoretically, DRNNs can maintain information in “memory” over long sequences, enabling the model to leverage past climate data effectively when making predictions. This characteristic addresses a fundamental challenge in meteorological forecasting — the need to incorporate historical climate conditions to accurately forecast future temperatures. Empirically, DRNNs have shown promising results in various time-series forecasting applications, ranging from financial market predictions to energy demand forecasting. Their ability to capture temporal dynamics and adapt to time-dependent data variance has been well-documented, providing a strong foundation for their application in air temperature forecasting.

The integration of Gated Recurrent Units (GRU) into DRNN architectures represents a targeted enhancement aimed at overcoming some of the limitations associated with traditional RNNs, particularly in handling long-term dependencies. GRUs introduced a gating mechanism that regulates the flow of information, allowing the model to better retain relevant information over longer sequences without the vanishing gradient problem that often plagues standard RNNs. The selection of DRNN-GRU was further reinforced by empirical evidence demonstrating their effectiveness in similar environmental and climatic forecasting scenarios. Studies have highlighted GRU-based models' superior performance in capturing complex, nonlinear relationships inherent in climatic data, offering significant improvements in predictive accuracy and reliability. The theoretical advantages of GRUs, coupled with their empirical success in managing sequential data with long-term dependences, underscored their suitability for the objectives of this study. By adopting DRNN-GRU, we aimed to leverage these attributes to achieve more accurate and reliable long-term temperature forecasts, contributing valuable insights into the domain of climate change forecasting.

LSTM models are renowned for their ability to capture long-term dependencies, addressing a significant limitation of traditional RNNs through a sophisticated gating mechanism. The addition of the Attention Mechanism allows the LSTM-Attention model to focus on specific parts of the input sequence that are most relevant for the prediction, enhancing the model's interpretability and performance. This selective focus is particularly useful in meteorological forecasting, where certain historical patterns may be more indicative of future conditions than others. Theoretical insights into attention mechanisms' capacity to weigh inputs dynamically, coupled with empirical evidence of their success in enhancing prediction tasks in natural language processing and sequence modeling, motivated their inclusion in our study.

The RSLSTM architecture builds on the strengths of LSTM models by stacking multiple LSTM layers, which allows for a more nuanced representation of data by capturing a broader range of dependencies across different scales. This multilayer approach is particularly beneficial for complex, nonlinear time-series data like climatic records, enabling the model to learn from the intricacies of the data more effectively.

Incorporating an Attention Mechanism into the RSLSTM architecture (RSLSTM-Attention) further refines the model's ability to prioritize crucial information within the input sequence. This architecture is designed to not only leverage the deep, stacked structure for capturing complex patterns but also to apply focused attention on elements of the sequence that significantly impact forecasting accuracy. This dual strategy of depth and focus aims to address the challenges of forecasting in the highly variable and complex domain of climate science.

The choice of these architectures was motivated by their alignment with the core challenges of long-term air temperature forecasting. Theoretical advantages, such as the ability of LSTM and GRU units to handle long-term dependencies and the attention mechanism's capability to enhance model focus and interpretability, are complemented by empirical evidence from their application in similar forecasting scenarios. Together, these architectures represent a comprehensive approach to addressing the multifaceted nature of climatic data and its prediction, underscoring our methodology's potential to advance the accuracy and reliability of air temperature forecasts.

B. *Data Description*

The dataset employed for this research was procured from a meteorological station operating under the Institute of Earth Sciences of the Southern Federal University. The data embodies temporal sequences spanning from January 1961 to 2023 (see in Fig. 1). This extensive temporal span facilitates the analysis of long-term trends and seasonal fluctuations. The observations encapsulate the following variables [40]-[50]:

- Date. A chronological marker for each observation. The date format adheres to the standard representation YYYY-MM-DD HH:MM:SS.
- Temperature. Measured in degrees Celsius, this stands as the pivotal dependent variable in presented deep learning models.
- Precipitation. Expressed in millimeters, this variable can serve as an indicator of humidity and rainfall.
- Pressure. Recorded in hectopascals (hPa), this parameter can be associated with atmospheric conditions and high/low-pressure systems.

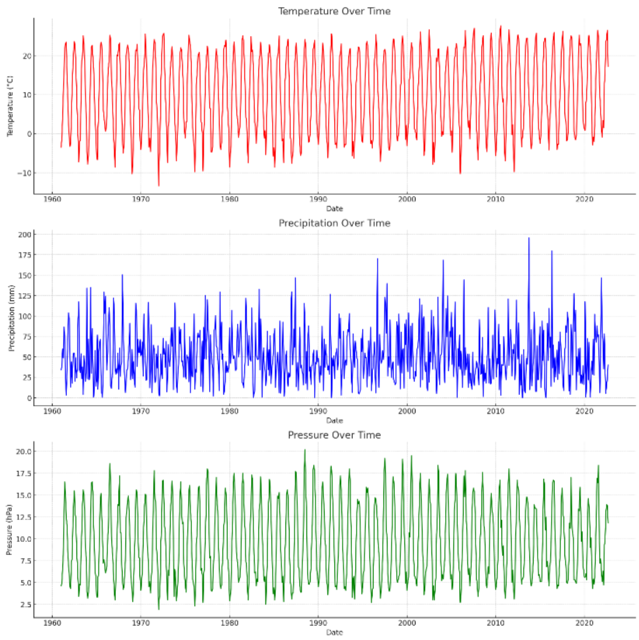

Fig. 1. Distribution of the dataset variables over time

Data collection was processed via a specialized meteorological station situated at the Institute of Earth Sciences of the Southern Federal University. The instrumentation of the meteorological station underwent myriad calibration and validation stages, ensuring the precision and reliability of measurements. These procedures, in turn, fortify the empirical foundation of the research and augment the gravitas of its conclusions [55]-[60].

Prior to integration into machine learning models, the data underwent a series of preliminary processing stages. Initially, a data audit was conducted to outliers, missing values, and other anomalies. These facets were critically imperative to ensure statistical validity and avert model result distortions. Subsequently, the data was normalized to enhance its statistical properties and ensure the convergence of learning algorithms. These measures are instituted to improve data quality and elevate the accuracy of predictive models.

C. *Data Preprocessing*

In this study, the preprocessing of the dataset is crucial for the success of the applied deep learning models. Our preprocessing strategy encompasses several critical steps: outlier detection and management, data normalization, and sequence generation, each tailored to the unique characteristics of the dataset spanning from 1961 to 2023.

Given the long-term nature of the dataset, outlier detection was paramount to ensure data integrity and model reliability. We employed an IQR (Interquartile Range) method to identify and scrutinize outliers [30]. This method calculates the IQR as the difference between the 75th and 25th percentiles and designates data points falling outside 1.5 times the IQR above the 75th percentile and below the 25th percentile as outliers. This approach was chosen due to its robustness in handling large datasets with potential non-Gaussian distributions. Upon identification, outliers were assessed with respect to historical climatic events and measurement anomalies. Outliers corresponding to known extreme weather events were retained, recognizing their importance in training models to predict under varying conditions.

Normalization is vital for aligning the scales of the various meteorological measurements, thus enhancing model training efficiency. We utilized Min-Max Scaling to adjust the features into a [0, 1] range, ensuring that no single variable would dominate the model's learning process due to scale differences [31].

The generation of input sequences for the deep learning models was based on a fixed-length sliding window approach, reflecting the sequential nature of time-series data. Each sequence, set to a length of 7, was constructed to predict the temperature on the following time. This sequence length was determined empirically, balancing the need to capture relevant temporal dynamics without introducing excessive complexity that could hinder model performance.

III. RESULTS

A. *Comparison of Different Architecture Performances*

The research was conducted on an ASUS Vivobook Pro 14X laptop, equipped with an 11th Gen Intel(R) Core(TM) i7-11370H processor from the H-series and an NVIDIA GeForce RTX 3050 graphics card. The system operated under the Windows 11 Pro operating system, version 22H2.

The device boasted RAM capacity of 16.0 GB and was equipped with a 64-bit operating system with an x64-based processor. These technical specifications ensured high performance and efficiency during the execution of the air temperature forecasting study using deep neural networks.

Several pivotal libraries, specifically created for data processing and analysis in the realm of scientific research, were employed in this study. Pandas (version 2.1.1) provides efficient tools for data manipulation and analysis, facilitating ease of interaction with diverse data types, NumPy (version 1.25.2) offers functionalities for computational operations on multidimensional data arrays, TensorFlow (version 2.14.0) is a robust tool for the creation and training of deep neural networks and its versatile functionalities cater to a broad spectrum of machine learning tasks, Scikit-learn (version 1.3.1) provides an extensive range of machine learning algorithms and tools for their effective implementation, enabling the resolution of a myriad of forecasting and classification tasks, keras (version 2.14.0) serves as a high-level interface for neural network construction, scienceplots (version 2.1.0) offers a suite of stylized graphical elements optimized for scientific data representation, enhancing the visual clarity of research outcomes and Matplotlib (version 3.8.0) as an instrumental tool for data visualization and chart creation.

The consolidation of these libraries, in conjunction with the technical infrastructure, provided a robust foundation for the research, ensuring precision, efficiency, and reproducibility in the study's outcomes.

The analysis of the results from the five different experiments, which determined the corresponding values of MAE, MSE, and RMSE for each of the models, reveals insightful patterns in the predictive accuracy of the deep learning architectures employed. The results are presented in Table I.

TABLE I. RESULTS OF THE DIFFERENT DEEP LEARNING ARCHITECTURES PERFORMANCE IN THE TASK OF AIR TEMPERATURE PREDICTION

Model	MAE	MSE	RMSE
DRNN	0.07	0.006	0.08
DRNN-GRU	0.06	0.005	0.08
LSTM-Attention	0.22	0.06	0.25
RSLSTM	0.28	0.12	0.34
RSLSTM-Attention	0.27	0.11	0.33

From the data presented, it is observable that the DRNN model exhibits a low MAE value (0.07), indicating relatively minor errors in its forecasts. Additionally, the MSE and RMSE for this model remain at a considerably low level (0.006 and 0.08 respectively), signifying a high accuracy in predictions. The DRNN-GRU model also demonstrates commendable results with an MAE of 0.06 and an MSE of 0.005. The RMSE remains at 0.08. These metrics indicate a low degree of error in the forecasts for this architecture. However, the LSTM-Attention model exhibits higher values of MAE (0.22), MSE (0.06), and RMSE (0.25). These results suggest more significant errors in the forecasts, which could be attributed to the architecture's insufficient adaptation to the specific data set. The RSLSTM and RSLSTM-Attention models also display intermediate MAE values (0.28 and 0.27 respectively) and higher MSE and RMSE values (0.12 and 0.34 for RSLSTM, 0.11 and 0.33 for RSLSTM-Attention). This indicates relatively larger errors in the forecasts for these architectures.

Based on the analysis, it can be concluded that the DRNN and DRNN-GRU models achieved the best results in assessing the accuracy of air temperature forecasts, possessing lower values of MAE, MSE, and RMSE. These models can be considered the most effective in this context and may be found in tasks related to forecasting climatic changes and studying natural systems [37]-[45].

B. Comparison of the Temperature Prediction

The analysis of temperature forecasting results using different models and their juxtaposition with actual data reveals intriguing aspects and facilitates a comparative evaluation.

Based on the presented graphs (Fig. 2 and Fig. 3.) examining temperature anomalies by various neural networks, several conclusions can be drawn concerning their efficacy. The DRNN model, a relatively simple recurrent network, performs commendably throughout the entire time series. The fluctuations in the predicted anomalies range from -5 to 5 degrees. However, by 2023, these fluctuations escalate, reaching an amplitude of approximately 8 degrees.

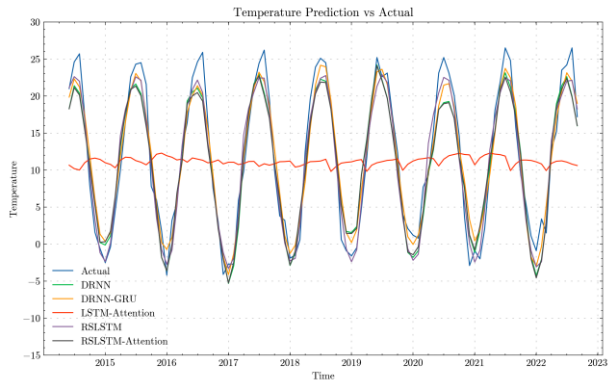

Fig. 2. Prediction of the temperature by different architectures

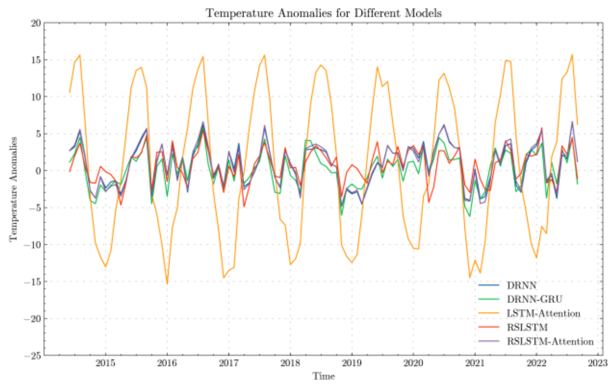

Fig. 3. Comparison of temperature anomalies from prediction of different models

The DRNN-GRU model, which integrates GRU elements into the DRNN architecture, exhibits slightly more stability. The fluctuations in predictions throughout the entire time series range from -3 to 3 degrees, but by 2023, the discrepancy between the predicted and actual data expands to

6 degrees. Despite its complexity and the incorporation of an attention mechanism, the LSTM-Attention model does not yield satisfactory results. During specific periods, especially in 2022, deviations from actual data exceed 15 degrees, representing a significant error for such tasks. Both the RSLSTM and RSLSTM-Attention models represent the highest efficacy. Specifically, the RSLSTM-Attention model displays anomaly fluctuations merely within 2-3 degrees of the actual data, attesting to its high predictive accuracy. In contrast, the RSLSTM model exhibits fluctuations around 5 degrees. In summation, based on quantitative assessment, the RSLSTM-Attention model boasts the highest accuracy, while the LSTM-Attention model yielded the least precise results.

Within our study, we employed Local Interpretable Model-agnostic Explanations (LIME) to demystify the decision-making processes of our models.

The plot (Fig. 4) for a specific prediction made by the DRNN architecture reveals a nuanced interplay of features across the temporal sequence. Notably, "Pressure" at the initial timestep (Timestep 0) appears to be a significant driver of the model's output, indicating that early atmospheric pressure readings weigh heavily in the model's predictive synthesis.

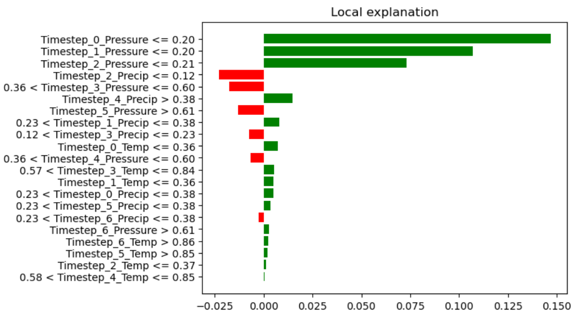

Fig. 4. Feature importance for DRNN architecture

Additionally, "'Precipitation" at Timestep 4 emerges as a substantial positive contributor, suggesting that mid-sequence precipitation levels are critically informative for the prediction at hand. Conversely, "Temperature" at Timestep 6 exhibits a strong negative influence on the prediction. This may imply a temporal relationship where temperature readings toward the end of the sequence serve as a corrective mechanism within the model's architecture, potentially overriding earlier cues to refine the prediction.

From the plot (Fig. 5) provided for the DRNN-GRU model, we observe a detailed attribution of feature importance across the sequences.

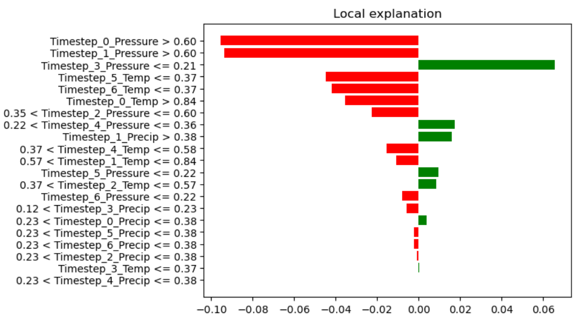

Fig. 5. Feature importance for DRNN-GRU architecture

The analysis underscores a substantial negative impact from the features "Pressure" at Timesteps 0, 1, and 6 and "Temperature" at Timesteps 2, 3, and 4. The temporal placement and influence of these features suggest that the DRNN-GRU model assigns significant predictive weight to atmospheric pressure at the start and end of the sequence and to mid-sequence temperatures, potentially alluding to a mechanistic understanding of the target phenomenon that penalizes certain conditions at key temporal junctures. Conversely, "Precipitation" at Timestep 1 and "Pressure" at Timestep 5 contribute positively, albeit to a lesser degree than the negative contributors. This hints at a model sensitivity to climatic variables that, when occurring at specific times, are deemed conducive to a rise in the predicted variable.

The gated architecture of the DRNN-GRU, which facilitates selective memory retention and deletion, likely contributes to the distinct pattern of feature importance observed. It endows the model with the ability to modulate its focus across the temporal spectrum of the data, affording each input feature a different degree of influence depending on its timestep. Our findings through LIME suggest that the DRNN-GRU model's predictions are not merely a function of the immediate input state but rather a composite reflection of the sequential data narrative. The gating mechanisms intrinsic to the GRU units appear to filter the temporal information flow, enabling the model to capture and prioritize complex, time-dependent relationships within the data.

From the plot (Fig. 6) reveals the LSTM-Attention model's interpretative mechanisms at a granular level. certain variables such as "Pressure" and "Precipitation" at Timestep 4, and "Temperature" at Timestep 1 have a positive effect on the output of the model. Notably, "Pressure" at Timestep 4 has the most substantial positive impact, suggesting that the model has learned to associate the state of atmospheric pressure at this point in the sequence with an increase in the target variable.

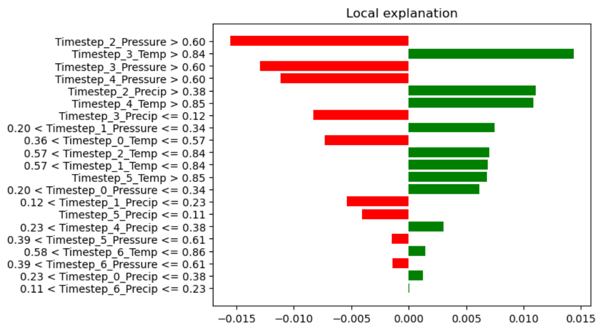

Fig. 6. Feature importance for LSTM-Attention architecture

"Pressure" at Timestep 2 and "Temp" at Timestep 3 are among the features that lead to a decrease in the predicted value. This negative attribution implies that higher pressure at Timestep 2 or higher temperatures at Timestep 3 are interpreted by the model as signals to decrease the prediction outcome. The LSTM-Attention model integrates an attention mechanism, which allows the model to assign varying degrees of importance to different parts of the input sequence. The strong positive impact of "Pressure" at Timestep 4, for instance, could be a manifestation of the attention mechanism identifying a key moment in the sequence that is crucial for prediction. In contrast to the DRNN and DRNN-GRU models, the LSTM-Attention architecture's interpretability is characterized by its ability to assign and visualize such differential weights to sequential data. This capability allows for a more nuanced understanding of temporal dependencies and the importance of sequential context in influencing the model's output.

The plot for RSLSTM model (Fig. 7) offers insights into the specific influences of sequential input features on a single prediction.

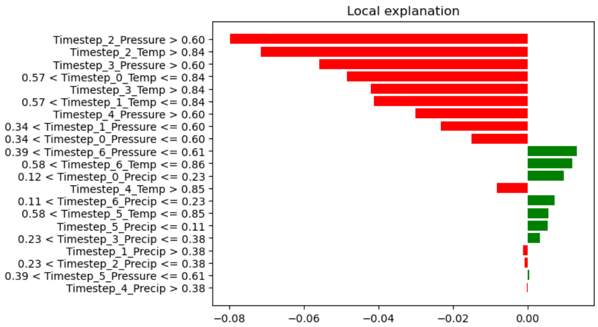

Fig. 7. Feature importance for RSLSTM architecture

Here, "Pressure" at Timesteps 2 and 3, and "Temp" at Timestep 1, appear to significantly reduce the predicted value, suggesting that the model is sensitive to higher values of these features within the early to middle parts of the sequence. Conversely, "Temp" at Timestep 6, demonstrate a positive influence, although these effects are less pronounced compared to the negative ones. This pattern of feature attribution reflects the RSLSTM's ability to leverage both the direct input data and the residuals of this data over time, likely allowing it to capture complex temporal behaviors more effectively.

In case of RSLSTM-Attention architecture (Fig. 8) we observe a predominant impact of "Pressure" at Timesteps 1 and 2, along with "Temperature" at Timestep 1, which strongly elevate the prediction value.

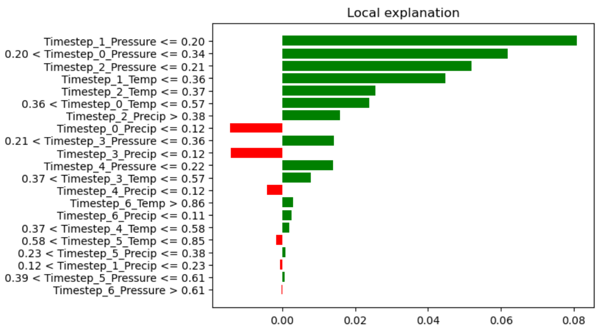

Fig. 8. Feature importance for RSLSTM-Attention architecture

This suggests the model associates early timesteps' pressure and temperature with an increase in the output variable, indicating a model sensitivity to initial sequence conditions. Conversely, "Pressure" at Timestep 3 and "Precipitation" at Timestep 6, indicate negative contributions, albeit with a lesser influence than the positive ones. The integration of an attention mechanism likely refines the model's focus, allowing it to selectively weigh these inputs and their residuals at pivotal timesteps, which aligns with the attentional dynamics that RSLSTM-Attention is designed to capture. The analysis thereby enhances interpretability by revealing how the model's attention-based processing differentially weighs temporal inputs to arrive at its forecasts.

IV. Discussion

A. *Study Results and Their Strategic Implications*

The interpretation of the results presented in this study extends beyond mere statistical analysis; it also holds strategic significance for various applied domains. Specifically, the DRNN and DRNN-GRU models, which showcased the best results based on metrics MAE (0.07 and 0.06 respectively), MSE (0.006 and 0.005), and RMSE (0.08 for both), are of particular interest for tasks related to climate change forecasting and the study of natural systems. These models can be integrated into Geographic Information Systems (GIS) for real-time resource monitoring and management, potentially influencing policies in sustainable development and environmental safety.

Furthermore, given the high accuracy of these models, they can be employed in agriculture to optimize water resource management systems and regulate microclimates in both open and enclosed areas. In the medical realm, these models can be utilized to assess the impact of climatic conditions on the spread of infectious diseases. At the macroeconomic level, precise climate condition forecasts can underpin strategic planning in sectors like energy, logistics, and insurance.

Models with high error values, such as the LSTM-Attention with MAE 0.22, MSE 0.06, and RMSE 0.25, or the RSLSTM and RSLSTM-Attention with MAE 0.28 and 0.27 respectively, as well as MSE 0.12 and 0.11 and RMSE 0.34 and 0.33, offer valuable insights for the further refinement of machine learning algorithms. They highlight the need for additional research into the influence of various factors on predictive accuracy, including data structure, loss function characteristics, and optimization strategies. These models can be enhanced for more specialized applications where the consideration of additional parameters or conditions is required.

B. *Research Limitations and Opportunities of their Overcoming*

During this study, certain limitations were identified that need to be considered when interpreting the results and planning future research endeavors. Firstly, although the DRNN and DRNN-GRU models showcased exemplary results based on the MAE, MSE, and RMSE metrics, these models were trained and tested solely on data obtained from a single meteorological station. This constrains the generalizability of the results and their applicability across diverse geographical and climatic conditions. To overcome this limitation, it is recommended to expand the dataset by incorporating data from multiple meteorological stations distributed across various climatic zones.

The second limitation pertains to the choice of evaluation metrics. While the use of MAE, MSE, and RMSE is a standard approach, they do not always fully encapsulate the nuances of forecasting tasks in the context of climate change. For instance, these metrics do not account for potential nonlinear dependencies and data asymmetries. For a more comprehensive analysis, the application of additional metrics and evaluation methods, such as cross-validation or bootstrap analysis, might be beneficial.

The third limitation is related to the network architectures employed. While the study encompasses diverse architectures like DRNN, DRNN-GRU, and LSTM with attention mechanisms, it does not explore other approaches like convolutional neural networks or GANs that might also be effective for forecasting tasks. To address this limitation, the research can be expanded to include these and other contemporary architectures.

Lastly, the study focuses exclusively on temperature forecasting, neglecting other crucial climatic parameters such as humidity, wind speed, or pollutant concentration. Incorporating these parameters could significantly influence the accuracy and reliability of forecasts, especially in the context of intricate ecosystems. In summary, these limitations highlight avenues for future research and offer opportunities for both methodological enhancements and the precision improvement of predictive models.

C. Future Research Directions

Future research in the realm of temperature forecasting using deep learning could progress significantly by building upon the interpretability insights gained from current model architectures. One promising direction is the exploration of hybrid models that capitalize on the complementary strengths of diverse neural network designs. Envisioning a model that combines the nuanced spatial pattern recognition capabilities of convolutional neural networks with the intricate temporal dependencies captured by recurrent networks might yield a tool of enhanced predictive power and deeper climatic insight. Additionally, the field stands to benefit from an investigation into advanced attention mechanisms. Beyond the LSTM-Attention and RSLSTM-Attention models, delving into more sophisticated variants, such as multi-headed attention, could provide a more granular understanding of which aspects of the data are most informative and when. This could lead to the development of attention mechanisms that adapt not just to different features within the data but also to different temporal scales, recognizing patterns that evolve over both short and extended periods.

Another promising research trajectory involves augmenting deep learning models with domain-specific knowledge. Integrating principles from atmospheric science into the structure and training of neural networks might guide the learning process, resulting in models that not only perform well empirically but also align more closely with theoretical expectations and physical laws. Moreover, the pursuit of novel attention mechanisms that can dynamically adjust to changing atmospheric conditions could offer breakthroughs in how models handle abrupt climatic shifts or rare events. Such mechanisms would ideally provide the model with the capacity to discern and respond to the onset of significant weather phenomena, potentially improving forecasting accuracy and timeliness.

In light of the interpretability analyses conducted, future research should also emphasize the importance of model trustworthiness and user comprehension. Ensuring that predictive models remain interpretable to human experts is crucial, as it underpins the models' practical applicability and the confidence with which their output can be utilized in decision-making processes.

Altogether, these future research directions aim not just to enhance predictive accuracy but also to ensure that advancements in machine learning resonate with and are informed by the rich tapestry of knowledge in atmospheric science, leading to robust, reliable, and interpretable forecasting tools for meteorology.

V. Conclusion

Within the scope of the conducted research, a comparative analysis of the effectiveness of various deep neural network architectures for temperature forecasting based on meteorological station data was carried out. The results showcase significant disparities in forecast accuracy among the models examined. Specifically, the DRNN and DRNN-GRU models exhibited the highest accuracy according to the MAE, MSE, and RMSE metrics, while architectures with attention mechanisms (LSTM-Attention, RSLSTM-Attention) displayed a relatively high level of errors. These findings can serve as a foundation for the further advancement and optimization of machine learning methods in predicting climatic parameters.

Based on the results obtained, several practical recommendations can be made. Firstly, the DRNN and DRNN-GRU models, which demonstrated high precision, can be recommended for integration into Geographic Information Systems (GIS) with the aim of monitoring and managing climatic resources. They may also find application in agriculture for optimizing water resource management and in medical research for analyzing the influence of climatic factors on disease spread.

Secondly, there is a need for further research into architecture with attention mechanisms. Despite their relative inefficiency in the current study, these models possess the potential to process complex nonlinear dependencies and can be optimized for specific forecasting tasks.

Additionally, the necessity to expand the initial dataset by incorporating data from multiple meteorological stations can be highlighted. This would enhance the generalizability of the results, rendering them more universal.

Funding

The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation as part of the World-class Research Center program: Advanced Digital Technologies (contract No. 075-15-2022-312 dated 20 April 2022).

References

[1] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An Overview of CMIP5 and the Experiment Design," *Bulletin of the American Meteorological Society*, vol. 93, no. 4, pp. 485–498, 2012, doi: 10.1175/BAMS-D-11-00094.1.

[2] C. Schultz, "Climate dynamics: Why does climate vary?," *Eos, Transactions American Geophysical Union*, vol. 92, no. 34, pp. 285–286, 2011, doi: 10.1029/2011EO340006.

[3] K. Abbass, M. Z. Qasim, H. Song, M. Murshed, H. Mahmood, and I. Younis, "A review of the global climate change impacts, adaptation, and sustainable mitigation measures," *Environ. Sci. Pollut. Res.*, vol. 29, no. 28, pp. 42539–42559, 2022, doi: 10.1007/s11356-022-19718-6.

[4] R. G. Newell, B. C. Prest, and S. E. Sexton, "The GDP-Temperature relationship: Implications for climate change damages," *Journal of Environmental Economics and Management*, vol. 108, p. 102445, 2021, doi: 10.1016/j.jeem.2021.102445.

[5] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Enhancing smart farming through the applications of Agriculture 4.0 technologies," *International Journal of Intelligent Networks*, vol. 3, pp. 150–164, 2022, doi: 10.1016/j.ijin.2022.09.004.

[6] K. Arbuthnott, S. Hajat, C. Heaviside, and S. Vardoulakis, "Changes in population susceptibility to heat and cold over time: assessing adaptation to climate change," *Environmental Health*, vol. 15, no. 1, 2016, doi: 10.1186/s12940-016-0102-7.

[7] D. Cawthorne, A. R. de Queiroz, H. Eshraghi, A. Sankarasubramanian, and J. F. DeCarolis, "The Role of Temperature Variability on Seasonal Electricity Demand in the Southern US," *Frontiers in Sustainable Cities*, vol. 3, 2021.

[8] T. P. Agyekum, P. Antwi-Agyei, and A. J. Dougill, "The contribution of weather forecast information to agriculture, water, and energy sectors in East and West Africa: A systematic review," *Frontiers in Environmental Science*, vol. 10, 2022.

[9] S. Silva, I. Soares, and C. Pinho, "Climate change impacts on electricity demand: The case of a Southern European country," *Utilities Policy*, vol. 67, p. 101115, 2020, doi: 10.1016/j.jup.2020.101115.

[10] S. Halilovic, L. Odersky, and T. Hamacher, "Integration of groundwater heat pumps into energy system optimization models," *Energy*, vol. 238, p. 121607, 2022, doi: 10.1016/j.energy.2021.121607.

[11] P. Joe *et al*., "Predicting the Weather: A Partnership of Observation Scientists and Forecasters," in *Towards the "Perfect" Weather Warning: Bridging Disciplinary Gaps through Partnership and Communication*, pp. 201–254, 2022, doi: 10.1007/978-3-030-98989-7_7.

[12] D. Fister, J. Pérez-Aracil, C. Peláez-Rodríguez, J. Del Ser, and S. Salcedo-Sanz, "Accurate long-term air temperature prediction with Machine Learning models and data reduction techniques," *Applied Soft Computing*, vol. 136, p. 110118, 2023, doi: 10.1016/j.asoc.2023.110118.

[13] J. Jun and H. K. Kim, "Informer-Based Temperature Prediction Using Observed and Numerical Weather Prediction Data," *Sensors*, vol. 23, no. 16, 2023, doi: 10.3390/s23167047.

[14] A. G. Bhave, D. Conway, S. Dessai, and D. A. Stainforth, "Barriers and opportunities for robust decision making approaches to support climate change adaptation in the developing world," *Climate Risk Management*, vol. 14, pp. 1–10, 2016, doi: 10.1016/j.crm.2016.09.004.

[15] G. Fedele, C. I. Donatti, C. A. Harvey, L. Hannah, and D. G. Hole, "Transformative adaptation to climate change for sustainable social-ecological systems," *Environmental Science & Policy*, vol. 101, pp. 116–125, 2019, doi: 10.1016/j.envsci.2019.07.001.

[16] A. Rutgersson *et al.*, "Natural hazards and extreme events in the Baltic Sea region," *Earth System Dynamics*, vol. 13, no. 1, pp. 251–301, 2022, doi: 10.5194/esd-13-251-2022.

[17] C. Raymond *et al.*, "Understanding and managing connected extreme events," *Nat. Clim. Chang.*, vol. 10, no. 7, 2020, doi: 10.1038/s41558-020-0790-4.

[18] S. Jia, C. Yang, M. Wang, and P. Failler, "Heterogeneous Impact of Land-Use on Climate Change: Study From a Spatial Perspective," *Frontiers in Environmental Science*, vol. 10, 2022.

[19] J. Pongratz, C. Schwingshackl, S. Bultan, W. Obermeier, F. Havermann, and S. Guo, "Land Use Effects on Climate: Current State, Recent Progress, and Emerging Topics," *Curr. Clim. Change Rep.*, vol. 7, no. 4, pp. 99–120, 2021, doi: 10.1007/s40641-021-00178-y.

[20] N. M. Mahowald, D. S. Ward, S. C. Doney, P. G. Hess, and J. T. Randerson, "Are the impacts of land use on warming underestimated in climate policy?," *Environ. Res. Lett.*, vol. 12, no. 9, p. 094016, 2017, doi: 10.1088/1748-9326/aa836d.

[21] S.-I. Ao and H. Fayek, "Continual Deep Learning for Time Series Modeling," *Sensors*, vol. 23, no. 16, 2023, doi: 10.3390/s23167167.

[22] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13521–13617, 2023, doi: 10.1007/s10462-023-10466-8.

[23] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *Sn Comput. Sci.*, vol. 2, no. 6, p. 420, 2021, doi: 10.1007/s42979-021-00815-1.

[24] I. A. Araya, C. Valle, and H. Allende, "A Multi-Scale Model based on the Long Short-Term Memory for day ahead hourly wind speed forecasting," *Pattern Recognition Letters*, vol. 136, pp. 333–340, 2020, doi: 10.1016/j.patrec.2019.10.011.

[25] J. A. Weyn, D. R. Durran, and R. Caruana, "Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 9, p. e2020MS002109, 2020, doi: 10.1029/2020MS002109.

[26] M. Safia, R. Abbas, and M. Aslani, "Classification of Weather Conditions Based on Supervised Learning for Swedish Cities," *Atmosphere*, vol. 14, no. 7, 2023, doi: 10.3390/atmos14071174.

[27] T. R. Andersson *et al.*, "Seasonal Arctic sea ice forecasting with probabilistic deep learning," *Nat. Commun.*, vol. 12, no. 1, p. 5124, 2021, doi: 10.1038/s41467-021-25257-4.

[28] A. D. Kenne, M. Toure, L. Logamou Seknewna, and H. L. Ketsemen, "Subseasonal Prediction of Summer Temperature in West Africa Using Artificial Intelligence: A Case Study of Senegal," *International Journal of Intelligent Systems*, vol. 2024, p. e8869267, 2024, doi: 10.1155/2024/8869267.

[29] D. Krivoguz, "Methodology of physiography zoning using machine learning: A case study of the Black Sea," *Russian Journal of Earth Sciences*, vol. 20, no. 2, 2020, doi: 10.2205/2020ES000707.

[30] A. Smiti, "A critical overview of outlier detection methods," *Computer Science Review*, vol. 38, p. 100306, 2020, doi: 10.1016/j.cosrev.2020.100306.

[31] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 10173–10196, 2023, doi: 10.1109/TPAMI.2023.3250241.

[32] N. Abid *et al*., "Burnt Forest Estimation from Sentinel-2 Imagery of Australia using Unsupervised Deep Learning," *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 01-08, 2021, doi: 10.1109/DICTA52665.2021.9647174.

[33] Y. Yuan, L. Mou, and X. Lu, "Scene Recognition by Manifold Regularized Deep Learning Architecture," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2222-2233, Oct. 2015, doi: 10.1109/TNNLS.2014.2359471

[34] A. Srivastava, S. Neog, and K. Medhi, "An Efficient Deep Learning Architecture for Internet of Medical Things," *2023 4th International Conference on Computing and Communication Systems (I3CS)*, pp. 1-6, 2023, doi: 10.1109/I3CS58314.2023.10127240.

[35] D. J. Richter and R. A. Calix, "Using Double Deep Q-Learning to learn Attitude Control of Fixed-Wing Aircraft," *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 646-651, 2022, doi: 10.1109/SITIS57111.2022.00102.

[36] G. Zhong, K. Zhang, H. Wei, Y. Zheng, and J. Dong, "Marginal Deep Architecture: Stacking Feature Learning Modules to Build Deep Learning Models," in *IEEE Access*, vol. 7, pp. 30220-30233, 2019, doi: 10.1109/ACCESS.2019.2902631.

[37] A. C. Salian, S. Vaze, P. Singh, G. N. Shaikh, S. Chapaneri, and D. Jayaswal, "Skin Lesion Classification using Deep Learning Architectures," *2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA)*, pp. 168-173, 2020, doi: 10.1109/CSCITA47329.2020.9137810.

[38] M. Alqahtani and L. Macheeri Ramaswamy, "A Layer Decomposition Approach to Inference Time Prediction of Deep Learning Architectures," *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 855-859, 2022, doi: 10.1109/ICMLA55696.2022.00141.

[39] M. Rothmann and M. Porrmann, "A Survey of Domain-Specific Architectures for Reinforcement Learning," in *IEEE Access*, vol. 10, pp. 13753-13767, 2022, doi: 10.1109/ACCESS.2022.3146518.

[40] N. Prabhu, S. K. T, and J. G, "A Comprehensive Study of Web Phishing Detection Using Machine and Deep Learning Architecture," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-14, 2023, doi: 10.1109/ICCCNT56998.2023.10306583.

[41] F. Shuang, W. Junzheng, S. Wei, and C. Guangrong, "The research of loading model of eddy current dynamometer based on DRNN with double hidden layers," *The 27th Chinese Control and Decision Conference (2015 CCDC)*, pp. 2485-2490, 2015, doi: 10.1109/CCDC.2015.7162339.

[42] Z. -h. Cen, J. -l. Wei, R. Jiang, and X. Liu, "Real-time fault diagnosis of satellite attitude control system based on sliding-window wavelet and DRNN," *2010 Chinese Control and Decision Conference*, pp. 1218-1222, 2010, doi: 10.1109/CCDC.2010.5498162.

[43] S. Reddy, S. Goel, and R. Nijhawan, "Real-time Face Mask Detection Using Machine Learning/ Deep Feature-Based Classifiers For Face Mask Recognition," *2021 IEEE Bombay Section Signature Conference (IBSSC)*, pp. 1-6, 2021, doi: 10.1109/IBSSC53889.2021.9673170.

[44] C. Tang, X. Zheng, X. Yu, C. Chen, and W. Zhu, "Design and Research of Intelligent Quantitative Investment Model Based on PLR-IRF and DRNN Algorithm," *2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp. 1187-1191, 2018, doi: 10.1109/ITOEC.2018.8740741.

[45] B. Li, X. Fan, C. Jiang, and G. Jiang, "Decoupling control of thickness and tension based on DRNN-PID in cold-rolling," *Proceeding of the 11th World Congress on Intelligent Control and Automation*, pp. 1180-1184, 2014, doi: 10.1109/WCICA.2014.7052886.

[46] L. Xingjie, Z. Yanqing, M. Zengqiang, F. Xiaowei, and W. Junhua, "Direct multi-step prediction of wind speed based on chaos analysis and DRNN," *2009 International Conference on Sustainable Power Generation and Supply*, pp. 1-5, 2009, doi: 10.1109/SUPERGEN.2009.5348114.

[47] Z. Zhou and S. Shi, "Research on discrete PR controller with LCL filter based on DRNN for photovoltaic grid-connected inverter," *2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, pp. 2084-2088, 2015, doi: 10.1109/DRPT.2015.7432584.

[48] K. S. Zaman, M. B. I. Reaz, S. H. Md Ali, A. A. A. Bakar, and M. E. H. Chowdhury, "Custom Hardware Architectures for Deep Learning on Portable Devices: A Review," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6068-6088, Nov. 2022, doi: 10.1109/TNNLS.2021.3082304.

[49] H. C. Kaskavalci and S. Gören, "A Deep Learning Based Distributed Smart Surveillance Architecture using Edge and Cloud Computing," *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, pp. 1-6, 2019, doi: 10.1109/Deep-ML.2019.00009.

[50] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of Deep Learning and Reinforcement Learning to Biological Data," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063-2079, June 2018, doi: 10.1109/TNNLS.2018.2790388.

[51] Y. Xie, K. Wang, and Y. Zhang, "Coordination control system study based on DRNN for marine power system," *Proceedings of the 29th Chinese Control Conference*, pp. 292-295, 2010.

[52] Y.-H. Xie, K.-Q. Wang, and L.-K. Zhu, "Intelligent control system design based on DRNN for wood drying kiln," *2010 2nd International Conference on Industrial and Information Systems*, pp. 514-517, 2010, doi: 10.1109/INDUSIS.2010.5565767.

[53] Z. -q. Wei, Y. -k. Wang, X. -f. Qin, and J. Zheng, "Single Neuron PID Model Reference Adaptive Control for Filling Machine Based on DRNN Neural Network On-Line Identification," *2010 International Conference on Electrical and Control Engineering*, pp. 5599-5602, 2010, doi: 10.1109/iCECE.2010.1360.

[54] H. P. S. D. Weerathunga and A. T. P. Silva, "DRNN-ARIMA Approach to Short-term Trend Forecasting in Forex Market," *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pp. 287-293, 2018, doi: 10.1109/ICTER.2018.8615580.

[55] S. Smyl, G. Dudek, and P. Pelka, "ES-dRNN with Dynamic Attention for Short-Term Load Forecasting," *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, 2022, doi: 10.1109/IJCNN55064.2022.9889791.

[56] Y. Jin and C. Su, "Adaptive Model Predictive Control Using Diagonal Recurrent Neural Network," *2008 Fourth International Conference on Natural Computation*, pp. 276-280, 2008, doi: 10.1109/ICNC.2008.575.

[57] S. Smyl, G. Dudek and P. Pełka, "ES-dRNN: A Hybrid Exponential Smoothing and Dilated Recurrent Neural Network Model for Short-Term Load Forecasting," in *IEEE Transactions on Neural Networks and Learning Systems*, 2023, doi: 10.1109/TNNLS.2023.3259149.

[58] S. Banerjee and S. Chakraborty, "Deepsub: A Novel Subset Selection Framework for Training Deep Learning Architectures," *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1615-1619, 2019, doi: 10.1109/ICIP.2019.8803096.

[59] R. A. Borgalli and S. Surve, "Deep Learning Framework for Facial Emotion Recognition using CNN Architectures," *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, pp. 1777-1784, 2022, doi: 10.1109/ICEARS53579.2022.9751735.

[60] R. G. Tiwari, A. Misra, and N. Ujjwal, "Image Embedding and Classification using Pre-Trained Deep Learning Architectures," *2022 8th International Conference on Signal Processing and Communication (ICSC)*, pp. 125-130, 2022, doi: 10.1109/ICSC56524.2022.10009560.

[61] V. Shankar and S. Chang, "Performance of Caffe on QCT Deep Learning Reference Architecture — A Preliminary Case Study," *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, pp. 35-39, 2017, doi: 10.1109/CSCloud.2017.49.

[62] S. -M. Chiu *et al*., "Development of Lightweight RBF-DRNN and Automated Framework for CNC Tool-Wear Prediction," in *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-11, 2022, doi: 10.1109/TIM.2022.3164063

[63] H. Duan, H. Xie, and Y. Lu, "Transformer On-line Monitoring and Fault Diagnosis System Based on DRNN and PAS," *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 1-4, 2019, doi: 10.1109/ICEIEC.2019.8784584.

[64] C. Ni and T. Zhao, "Decoupling Control for a Class of Dynamic Nonlinear Coupling Systems via ADRC based on DRNN," *2020 Chinese Control And Decision Conference (CCDC)*, pp. 5374-5379, 2020, doi: 10.1109/CCDC49329.2020.9164477.