Single Model, Multiple Climates: Applying Modified Levenberg Marquardt Algorithm to Meteorological Time-Series Prediction

Cihat Nusret Uğurlu[1] and Mehmet Önder Efe[2]

Abstract— **Forecasting meteorological time series typically requires training separate models for distinct geographical locations due to the complexity of local climate dynamics. When a single Artificial Neural Network (ANN) is tasked with learning multiple datasets with diverse characteristics, the phenomenon known as "catastrophic forgetting" often degrades performance on previously learned tasks. In this study, the modified Levenberg-Marquardt (LM) algorithm, which utilizes a parameter masking strategy, is applied and validated for multi-site weather forecasting. Real-world hourly data spanning 2022-2024 for three cities representing different climates—Ankara (Continental), Mersin (Mediterranean), and Çanakkale (Transitional)—were utilized. A single shallow ANN architecture, equipped with randomly generated binary mask vectors specific to each city, was trained simultaneously on these datasets. The proposed method organizes the network weights into shared and task-specific subnetworks, allowing distinct climate patterns to be stored within the same model without interference. Experimental results demonstrate that the model accurately predicts temperature variations with different forecast horizons across all three cities, effectively preventing catastrophic forgetting. This study confirms that the modified LM algorithm provides robust generalization capabilities not only for theoretical regression tasks but also for complex, noisy real-world time series.**

I. INTRODUCTION

Accurate and timely prediction of meteorological events is of vital importance for numerous critical sectors, ranging from agriculture and energy management to aviation and disaster planning. In recent years, Artificial Neural Networks (ANNs) and Deep Learning techniques have emerged as robust alternatives to traditional physical models for modeling the non-linear and chaotic nature of atmospheric data [1]. The success of ANNs, particularly in time-series forecasting, has been widely demonstrated in the literature [2], [3]. However, the fact that different geographical regions and climatic zones possess unique dynamics presents a significant challenge for machine learning models.

In the conventional approach, separate models are trained for each distinct geographical location (e.g., a city with a continental climate versus one with a tropical climate). This leads to scalability issues and high computational costs. Conversely, when a single "universal" model approach is attempted for all regions, the phenomenon known in the literature as "Catastrophic Forgetting" is encountered [4], [5]. Standard learning algorithms (e.g., Backpropagation or standard Levenberg-Marquardt) tend to overwrite previously learned weights while the model attempts to adapt to a new dataset (a new climate), thereby erasing old information. This prevents a single ANN from successfully performing multiple tasks with distinct characteristics simultaneously.

To overcome this problem, a modification to the Levenberg-Marquardt (LM) optimization algorithm was proposed by Efe et al. [6]. This method employs binary masking on network parameters to facilitate the creation of specialized subnetworks for each task (or dataset) within the main network. Consequently, while preserving a single physical network structure, different datasets can be learned simultaneously through the selective updating of parameters.

In this study, the modified LM algorithm and masking strategy proposed by Efe et al. are applied to the real-world meteorological time-series prediction problem. Within the scope of this work, a single ANN model was developed to perform temperature forecasts for three cities in Turkey with distinct climate characteristics (Ankara, Mersin, Çanakkale). Unlike studies in the literature that often focus on theoretical algorithm verification, this paper demonstrates the method's performance on noisy and complex real-world data. Experimental results indicate that the proposed method can learn the distinct climate dynamics of all three cities with high precision and without interference.

The remainder of this paper is organized as follows: Section II explains the utilized modified LM algorithm and the masking structure. Section III details the dataset, preprocessing steps, and experimental setup. Section IV presents the prediction results and performance metrics, while Section V discusses the conclusion and future work.

II. METHODOLOGY

In this section, we describe the modified Levenberg-Marquardt (LM) algorithm adopted for simultaneous learning. The core optimization framework is based on the modification proposed by Efe et al. [6], which enables a single neural network to learn multiple datasets without catastrophic forgetting. Consider a dataset ensemble $D = \{d_1, d_2, ..., d_M\}$ representing M different tasks (e.g., different cities). For a standard neural network with a parameter vector $w \in \mathbb{R}^N$, the proposed method introduces a binary mask vector $m^k \in \{0, 1\}^N$ for each dataset k.

The mask vector determines the active subnetwork for a specific task. If the i-th element of the mask m_i^k is 1, the corresponding weight w_i is trained for dataset k; otherwise, it remains frozen or inactive for that specific

[1]Cihat Nusret Uğurlu is with the Computer Vision and Multimedia Analysis Group, TUBITAK Space Technologies Research Institute, Ankara, 06800, Türkiye cihat.ugurlu@tubitak.gov.tr

[2]Mehmet Önder Efe is with the Department of Computer Engineering, Hacettepe University, Ankara, 06800, Türkiye onderefe@gmail.com

forward pass. The learning process involves constructing a cumulative Jacobian matrix that encompasses gradient information from all datasets simultaneously. For a single dataset k, the Jacobian J_k is computed by applying the mask to the standard gradients:

$$J_k(i,j) = m_j^k \cdot \frac{\partial e_i^k}{\partial w_j}$$

where e^k is the error vector for dataset k. These individual Jacobians are then stacked vertically to form the global Jacobian matrix J_F:

$$J_F = [J_1^T, J_2^T, ..., J_M^T]^T$$

Similarly, the error vectors are concatenated to form a global error vector E_F. The update rule for the network parameters follows the Levenberg-Marquardt approximation [7]:

$$w(t+1) = w(t) - (J_F^T J_F + \mu I)^{-1} J_F^T E_F$$

where μ is the damping parameter. By solving this system simultaneously, the algorithm finds a parameter set w that satisfies the distinct error surfaces of all datasets, effectively preventing the interference that causes catastrophic forgetting.

III. Experimental Setup

A. Data Acquisition and Preprocessing

The study utilizes real-world hourly meteorological data obtained from the Open-Meteo API [1] for three distinct locations in Turkey, selected to represent diverse climatic zones:

1) **Ankara (Continental):** Characterized by significant diurnal temperature variations.
2) **Mersin (Mediterranean):** Characterized by humid and mild weather patterns.
3) **Çanakkale (Transitional):** Characterized by high wind speeds and transitional dynamics.

The dataset spans from January 1, 2022, to December 31, 2024. Initially, a broad set of candidate meteorological variables was retrieved to capture the complex atmospheric dynamics. To identify the most relevant predictors for the target variable ($Temperature(2m)$), a cross-correlation analysis was conducted among all candidate features (Fig. 1). This analysis evaluated the linear dependencies between the target temperature and other potential inputs such as wind components, precipitation, and pressure levels. Based on the calculated correlation coefficients, a final subset of 8 robust features was selected to maximize predictive power while minimizing redundancy. The selected input vector consists of:

1) Temperature (2m)
2) Relative humidity (2m)
3) Cloud cover
4) Mean temperature for the last 6-hours
5) Cyclical time encodings: Sine and cosine components of both the Hour-of-Day and Day-of-Year (4 features).

[1]https://open-meteo.com

These cyclical features were explicitly engineered to allow the neural network to model the daily and annual seasonality inherent in meteorological data [8].

A critical step in our methodology is the normalization strategy. Since the proposed Levenberg-Marquardt algorithm utilizes the hyperbolic tangent ($\tanh$) activation function, the input space must be bounded within $[-1, 1]$. To preserve the unique climatic range of each city (e.g., preserving the difference between a maximum of $40°$C in Mersin versus $30°$C in Ankara), we employed location-specific Min-Max scalers. Each dataset d_k was scaled independently using its own statistical properties. To rigorously mitigate data leakage, the Min–Max scalers were calibrated solely using the training subsets. The resulting scaling parameters, specifically the minimum and maximum values, were subsequently fixed and consistently applied to the validation and test datasets through transformation.

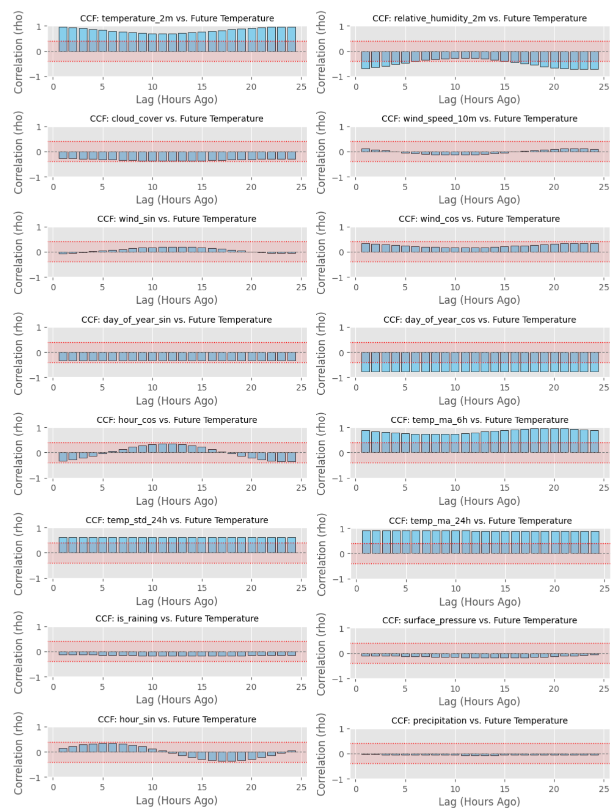

Fig. 1. Cross-correlation analysis for future temperature and different features.

B. Time-Series Formulation and Network Architecture

The forecasting problem is formulated as predicting the temperature T_{t+h} with forecast horizons of $h = 1$, $h = 6$, and $h = 24$ hours. Unlike deep learning approaches that rely on long historical sequences (large lookback windows), this study investigates the capability of the modified LM algorithm to learn dynamics using a minimalist input structure.

We set the lookback window to $L = 1$, meaning the network makes predictions based solely on the most recent observed state of the atmosphere, without an explicit buffer of past time steps. This choice was made deliberately to:

1) Minimize the number of trainable parameters (avoiding over-parameterization).
2) Test the robustness of the "simultaneous learning" capability under restricted information.

Consequently, the input dimension for the neural network is $1 \times 8 = 8$. The network architecture (Fig. 2) consists of:

- Input Layer: 8 neurons.
- Hidden Layer: 100 neurons with $\tanh$ activation.
- Output Layer: 1 linear neuron.
- Masking Threshold: 0.3 (approximately 70% active weights per task).

The total number of parameters (weights + biases) is 1001. The model was trained for 100 epochs with an initial damping parameter $\mu = 1.0$, which is adaptively updated during training. The dataset was split chronologically: the first 70% for training, the subsequent 15% for validation, and the final 15% for testing to strictly evaluate generalization on unseen future data.

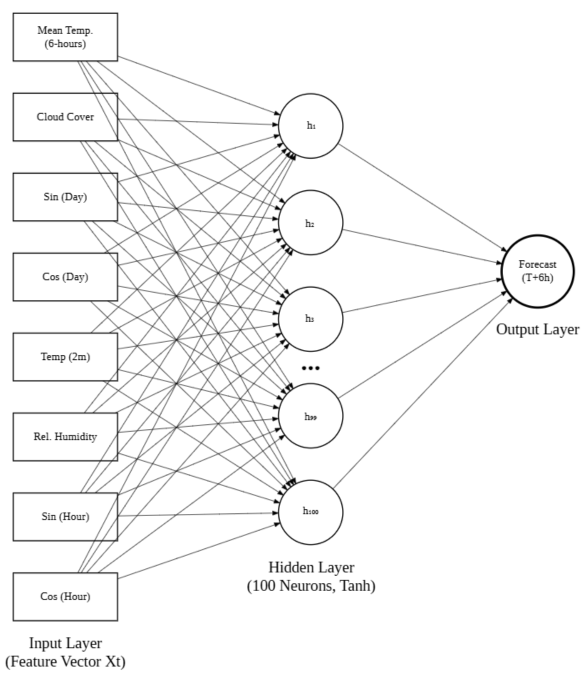

Fig. 2. The architecture of the proposed ANN.

During real-time inference, the model functions as a conditional multi-task architecture. The city identifier is not derived from the raw meteorological inputs; instead, it is explicitly provided as a metadata label (e.g., $Task\ ID = 0, 1, or\ 2$). This identifier serves as a routing signal, directing the system to employ the corresponding pre-calibrated local Min–Max scaler for input normalization and to enforce the city-specific binary parameter mask prior to executing the forward pass.

IV. Results

In this section, we present the quantitative performance of the proposed Single-Model Multi-Site forecasting framework. The model was evaluated on the test sets of Ankara, Mersin, and Çanakkale, which were strictly excluded from the training process. To rigorously assess the generalization capability and the limits of the proposed shallow architecture (with $Lookback = 1$), three separate instances of the model were trained specifically for short-term (1-hour), medium-term (6-hour), and long-term (24-hour) forecast horizons. This approach allows us to observe how the modified LM algorithm performs under varying degrees of predictive difficulty while maintaining the same minimalist topology.

A. Quantitative Analysis

Table I summarizes the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) for all three cities across different horizons.

TABLE I
Performance Comparison of Proposed Model Against Baselines (Best:**Bold**, Second:Underlined)

Horizon	City	Model	MSE	RMSE (°C)	MAE (°C)	R^2
		Naive	1.7137	1.3091	1.0178	0.9826
		ANN	0.6144	0.7838	0.5715	0.9937
	Ankara	Ind. LM	**0.4244**	**0.6515**	**0.4297**	**0.9957**
		LSTM	0.4681	0.6842	0.4599	0.9952
		Proposed	0.4293	0.6552	0.4347	0.9956
		Naive	1.3987	1.1827	0.8334	0.9784
		ANN	0.5617	0.7495	0.5416	0.9913
1-Hour	Mersin	Ind. LM	**0.3189**	**0.5647**	**0.4089**	**0.9951**
		LSTM	0.3993	0.6319	0.4515	0.9938
		Proposed	0.3201	0.5658	0.4106	0.9950
		Naive	1.2712	1.1275	0.8481	0.9828
		ANN	0.4617	0.6795	0.4971	0.9938
	Çanakkale	Ind. LM	**0.3025**	**0.5500**	**0.3813**	**0.9959**
		LSTM	0.3469	0.5889	0.4131	0.9953
		Proposed	0.3073	0.5543	0.3893	0.9958
		Naive	34.3786	5.8633	4.8600	0.6492
		ANN	3.2571	1.8047	1.3814	0.9668
	Ankara	Ind. LM	**2.6137**	**1.6167**	**1.1992**	**0.9733**
		LSTM	3.0251	1.7393	1.3211	0.9691
		Proposed	2.6510	1.6282	1.2049	0.9729
		Naive	24.2538	4.9248	3.9511	0.6238
		ANN	2.6236	1.6198	1.2688	0.9593
6-Hour	Mersin	Ind. LM	**1.8455**	**1.3585**	**1.0207**	**0.9714**
		LSTM	2.1970	1.4822	1.1500	0.9659
		Proposed	1.9258	1.3877	1.0448	0.9701
		Naive	26.0285	5.1018	4.1333	0.6465
		ANN	3.7593	1.9389	1.4239	0.9489
	Çanakkale	Ind. LM	**2.7096**	**1.6461**	**1.1963**	**0.9632**
		LSTM	3.4626	1.8608	1.3550	0.9530
		Proposed	2.7185	1.6488	1.1967	0.9631
		Naive	6.0923	2.4683	1.7465	0.9378
		ANN	5.5504	2.3559	**1.7310**	0.9433
	Ankara	Ind. LM	5.8464	2.4179	1.8126	0.9403
		LSTM	**5.5172**	**2.3489**	1.7377	**0.9437**
		Proposed	5.9372	2.4366	1.8174	0.9394
		Naive	3.3534	1.8312	1.2472	0.9479
		ANN	**2.8666**	**1.6931**	**1.2175**	**0.9555**
24-Hour	Mersin	Ind. LM	4.0110	2.0027	1.4392	0.9377
		LSTM	2.9767	1.7253	1.2866	0.9538
		Proposed	3.4886	1.8678	1.3584	0.9458
		Naive	7.3663	2.7141	1.9086	0.8996
		ANN	6.1662	2.4832	1.7918	0.9160
	Çanakkale	Ind. LM	6.9796	2.6419	1.8916	0.9049
		LSTM	**5.8375**	**2.4161**	**1.7477**	**0.9205**
		Proposed	6.7613	2.6003	1.8529	0.9079

1) **Short-Term Accuracy (1-Hour):** At the $h = 1$ forecast horizon, the results clearly demonstrate the effectiveness of the masking mechanism. The proposed model significantly outperforms both the naive baseline and the unmasked neural network, achieving performance nearly equivalent to the independent LM models across all three cities ($R^2 > 0.995$). These findings indicate that binary parameter masking effectively isolates task-specific decision boundaries within a shared weight matrix, thereby mitigating the multi-task interference commonly observed in conventional unmasked neural networks.
2) **Medium-Term Robustness (6-Hour):** As the forecast horizon extends to $h = 6$, the physical inertia of weather systems diminishes, as reflected in the complete failure of the Naive persistence baseline. In this setting, the ablation study becomes particularly informative. The unmasked ANN struggles to model the diverging dynamics of three distinct climates, resulting in higher error rates. In contrast, the proposed model maintains strong predictive performance, closely matching the upper-bound results of the Independent LMs. These results support the claim that the mask-induced subnetworks remain stable and effectively capture distinct regional meteorological patterns without cross-contamination.
3) **Long-Term Limits (24-Hour):** At the $h = 24$ horizon, a clear shift in performance emerges: the LSTM and standard ANN surpass the LM-based models, while the proposed model shows a noticeable decline in accuracy. Notably, the Independent LM exhibits the same degradation, indicating that the limitation arises not from the masking algorithm but from the strict $Lookback = 1$ architectural constraint. A shallow model relying on a single prior time step cannot capture a full 24-hour diurnal cycle. The superior performance of the LSTM at this horizon supports this explanation, demonstrating that although the masking mechanism remains effective, long-term forecasting requires recurrent architectures or extended lookback windows.

B. Temporal Forecast Analysis

To visually assess the model's tracking capability, we analyzed the time-series predictions for short, medium, and long-term horizons across all three cities. The comparative plots are presented in Figures 3, 4, and 5.

1) **Short-Term Precision (Fig. 3):** For the 1-hour forecast horizon, the model exhibits near-perfect synchronization with the ground truth. As seen in Fig. 3, the predicted curves (dashed lines) almost completely overlap with the real values (solid lines) for all three cities. The model successfully captures high-frequency fluctuations and sudden temperature drops, validating that the shallow ANN effectively maps the current atmospheric state to the immediate future state without losing information.

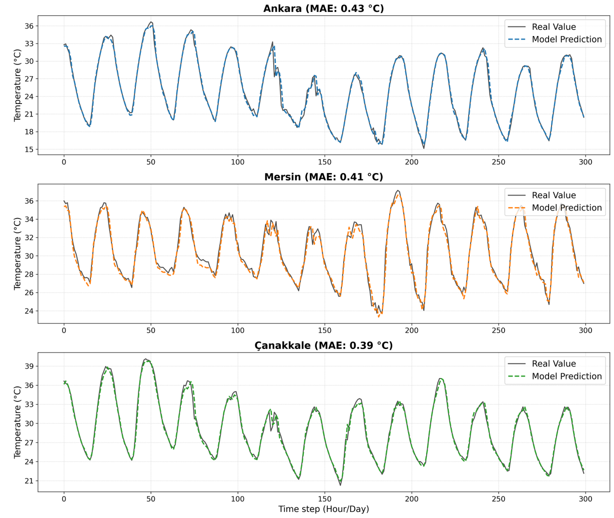

Fig. 3. Temporal temperature predictions for 1-hour forecast horizon.

2) **Medium-Term Trend Following (Fig. 4):** In the 6-hour horizon (Fig. 4), the model continues to follow the diurnal cycles reliably. While a slight smoothing effect is observed at extreme peaks (local maxima) and troughs (local minima), the general phase alignment remains accurate. The model correctly anticipates the rising and cooling trends of the day, proving its operational viability for medium-term planning.

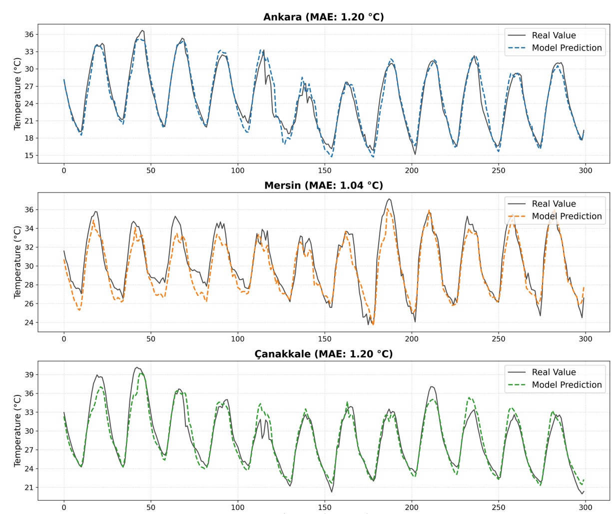

Fig. 4. Temporal temperature predictions for 6-hour forecast horizon.

3) **Long-Term Persistence and Phase Shift (Fig. 5):** For the 24-hour horizon, a distinct behavioral change is observed. As shown in Fig. 5, while the model correctly reproduces the amplitude and periodicity of the temperature cycles, a noticeable phase shift (time lag) is present. The predictions appear to trail the actual values. This phenomenon is a direct consequence of the minimalist design choice ($Lookback = 1$). Lacking a sequence of historical data, the network cannot infer

the rate of change (first derivative) or momentum of the weather front. Consequently, to minimize the error loss, the model converges towards a "persistence" behavior, heavily relying on the strong autocorrelation between t and $t + 24$. It effectively predicts that "tomorrow will be similar to today," which results in the observed lag when weather patterns shift dynamically. This limitation highlights the trade-off between architectural simplicity and long-term predictive responsiveness.

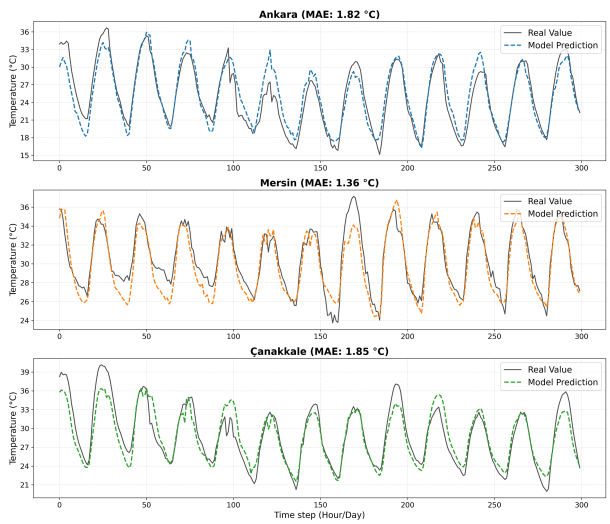

Fig. 5. Temporal temperature predictions for 24-hour forecast horizon.

C. Simultaneous Learning Verification

The consistent performance across all three cities confirms the efficacy of the modified LM algorithm. In a standard "universal" model without masking, the error gradients from the volatile weather of Çanakkale would likely disrupt the learned weights for the stable weather of Mersin (catastrophic forgetting). However, our results show that Mersin maintains low error rates independent of the other tasks, validating that the masking strategy successfully isolates task-specific knowledge within the shared network architecture.

One of the primary contributions of this study is the ability of a single neural network to learn disparate climatic behaviors without interference. To verify this capability, we conducted a "probing analysis" on the trained model (specifically the 24-hour forecast instance).We generated a synthetic input grid where normalized Feature 1 ($Temperature$) and Feature 2 ($RelativeHumidity$) varied across the range $[-1, 1]$, while all other features were held constant at their mean values. This synthetic data was then fed into the network three separate times, applying the specific binary mask (m^k) corresponding to each city.

Fig. 6 illustrates the resulting decision surfaces. The visual evidence confirms that the model has constructed distinct decision manifolds for each city, despite sharing the same underlying weight matrix.

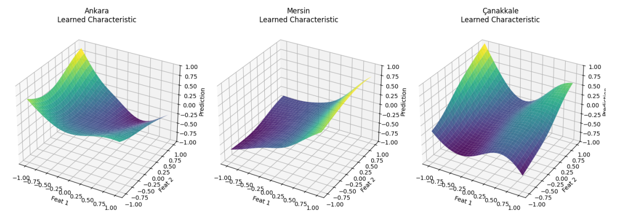

Fig. 6. Visualization of the learned decision surfaces for the 24-hour forecast model. The axes represent normalized Feature 1 and Feature 2, while the Z-axis represents the model's prediction.

- **Ankara (Left):** The surface exhibits a specific curvature reflecting the sharp temperature-humidity response typical of a continental climate.
- **Mersin (Middle):** The decision surface is noticeably flatter and has a different gradient orientation. This aligns with the physical reality of a Mediterranean coastal city, where high humidity buffers temperature fluctuations, leading to a more suppressed response curve compared to Ankara.
- **Çanakkale (Right):** The model generates a highly non-linear, saddle-like geometry, distinct from the other two.

If the "catastrophic forgetting" phenomenon had occurred, these three plots would look identical or represent an averaged compromise between the climates. The distinct geometries prove that the modified LM algorithm has successfully allocated task-specific subnetworks, allowing the single model to specialize in the unique atmospheric physics of each location simultaneously.

D. Regression Analysis

The scatter plot (Fig. 7) shows that the predictions for all three cities tightly cluster around the ideal $y = x$ line, further validating the high R^2 scores presented in Table I.

V. Conclusion

In this study, we addressed the challenge of "catastrophic forgetting" in multi-site meteorological forecasting by applying the modified Levenberg-Marquardt algorithm equipped with a parameter masking strategy. We developed a single, shallow Artificial Neural Network to simultaneously predict temperatures for three distinct climatic regions in Turkey: Ankara, Mersin, and Çanakkale.

The experimental results lead to three key conclusions:

1) **Elimination of Catastrophic Forgetting:** The proposed method successfully trained a single model on multiple distinct datasets without interference. The model adapted to the specific climatic characteristic of each city (e.g., lower variance in Mersin vs. higher variance in Ankara).
2) **High Accuracy with Minimalist Input:** Even with a restricted input structure ($Lookback = 1$), the model achieved high precision. For the 6-hour horizon, it maintained a Mean Absolute Error (MAE) of approximately 1°C and an $R^2 > 0.96$, demonstrating the algorithmic efficiency of the modified LM approach.

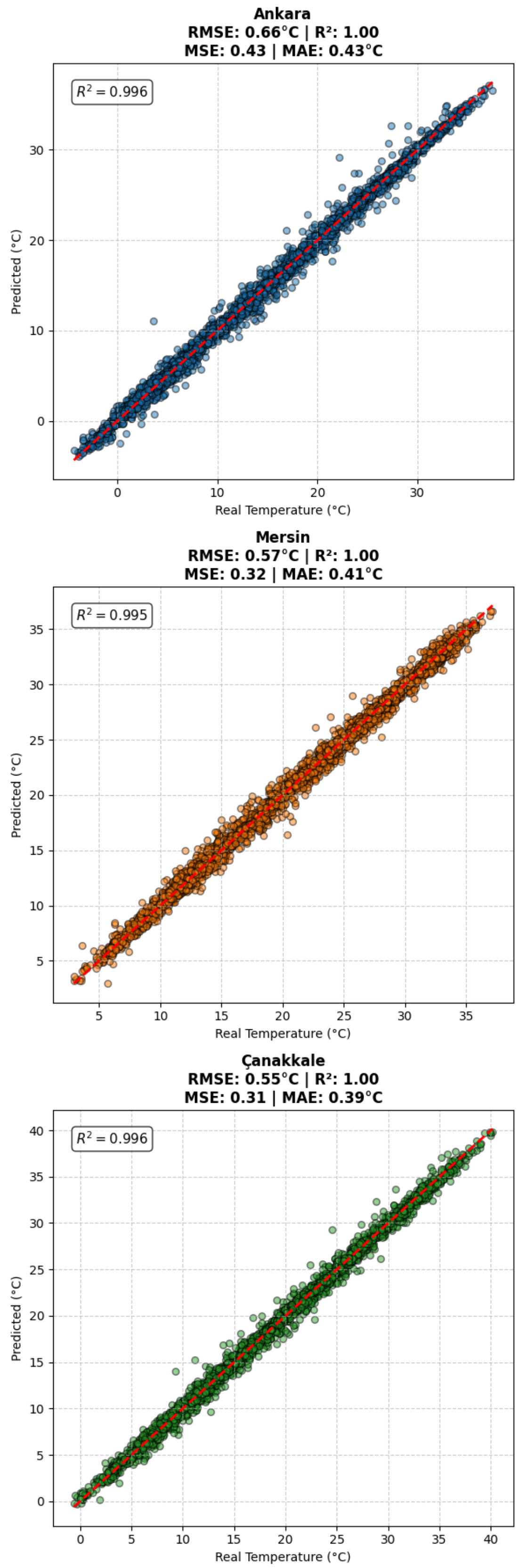

Fig. 7. Real vs. predicted temperatures for each city (1-hour forecast model).

3) **Scalability:** The ability to handle multiple climates within a single parameter set (1001 parameters) offers a significant advantage over ensemble approaches that require training and storing separate models for each location.

Future work will focus on further enhancing the predictive capability of the proposed shallow ANN architecture without increasing computational complexity. First, we aim to relax the minimalist input constraint ($Lookback = 1$) and investigate the impact of wider temporal windows (e.g., $L = 12$ or $L = 24$). By feeding the network with historical sequences, we anticipate significantly reducing the error rates for longer forecast horizons (24h+), as the model will be able to capture atmospheric trends and momentum explicitly. Second, instead of strict binary masking, we plan to explore "soft masking" techniques, where mask values can vary continuously between 0 and 1. This would enable partial knowledge transfer between cities with similar climatic traits, potentially improving generalization. Furthermore, although the current architecture effectively supports three concurrent tasks, expanding the number of locations may surpass the representational capacity of the 100-neuron hidden layer. Therefore, evaluating the scalability of the proposed masked architecture across a broader range of climatic zones (e.g., $M = 5, 10, 20$ cities) represents an important avenue for future work. Finally, we intend to enrich the input vector with engineered features such as the first and second derivatives of pressure and temperature, allowing the shallow network to perceive rates of change without requiring deep recurrent layers.

References

[1] H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine learning methods for weather forecasting: A survey," *Atmosphere*, vol. 16, no. 1, 2025. [Online]. Available: https://www.mdpi.com/2073-4433/16/1/82

[2] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207097000447

[3] A. Tealab, "Time series forecasting using artificial neural networks methodologies: A systematic review," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 334–340, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2314728817300715

[4] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, p. 3521–3526, Mar. 2017. [Online]. Available: http://dx.doi.org/10.1073/pnas.1611835114

[5] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends in Cognitive Sciences*, vol. 3, no. 4, pp. 128–135, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661399012942

[6] M. O. Efe, B. Kürkçü, C. Kasnakoğlu, Z. Mohamed, and Z. Liu, "A modified levenberg marquardt algorithm for simultaneous learning of multiple datasets," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 71, no. 4, pp. 2379–2383, Apr. 2024.

[7] M. Hagan and M. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[8] S. Khazem and H. Kanso, "Cyclical temporal encoding and hybrid deep ensembles for multistep energy forecasting," 2025. [Online]. Available: https://arxiv.org/abs/2512.03656