

Quantifying Regional Climate Stress in Ankara via Entropy-Weighted Random Forest Modeling

İbrahim Miraç Eligüzel and Nazmiye Eligüzel

Industrial Engineering Department

Gaziantep İslam Science and Technology University, Gaziantep, Türkiye

ibrahim.eliguzel@gibtu.edu.tr, nazmiye.eliguzel@gibtu.edu.tr

Sena Aydoğan

Industrial Engineering Department

Gazi University, Ankara, Türkiye

senaaydogan@gazi.edu.tr

Abstract

Urban centers face unprecedented challenges due to climate change and increasing environmental volatility. In this study, we developed a hybrid framework to evaluate climate-related stress in Ankara, Türkiye. To achieve this aim, a long-term monthly dataset from 1997 to 2025 is analyzed. Collected via the Meteostat library, the data were processed through the Entropy Weighting Method to objectively rank the impact of six variables, including temperature extremes and precipitation. By building a composite Climate Stress Index, we quantified monthly risk levels. After that, we applied a Random Forest regressor to capture the complex, non-linear patterns within the weather data. Results reveal that precipitation is the primary factor driving climatic stress in Ankara; also, the machine learning model performed with remarkable precision. Performance metrics show an outstanding result. In that aspect, an R^2 score of 0.9881 and an RMSE of 0.0193 are observed. These results demonstrate that the proposed hybrid approach can serve as a practical, automated tool for local authorities to monitor and manage regional climate risks more effectively.

Keywords

Climate Stress Index (CSI), Entropy-Based Methods, Random Forest Regression, Regional Climate Assessment, Ankara.

1. Introduction

Rising global surface temperatures are influencing the frequency of extreme climate events. Climate, defined as the average weather conditions for a specific location over a period (30 years, per the World Meteorological Organization), comprises average temperatures, precipitation, and sunlight in the context of climate change scenarios (Chhiddikur & Shajedur, 2023). Many aspects of the world affect it, including extreme weather events such as shifting rainfall patterns and rising temperatures, which make regions far more vulnerable (Pandit, 2023; Wanke & Tan, 2025). Due to the calibration of ecosystems and urban infrastructures to specific climatic ranges, even slight long-term changes might result in significant ecological and socioeconomic consequences (Kc et al., 2021).

These continuous disruptions in climate variables have a direct impact on the hydrological cycle and human existence. Ankara province plays a significant role as the capital of the Republic of Türkiye and is the second-most populous city (Çelik et al., 2025). Ankara is particularly susceptible to weather fluctuations due to its rapid expansion as an urban hub in a semi-arid region. Accurately monitoring and forecasting climate stress in this context requires a multifaceted strategy capable of addressing the non-linear relation of meteorological data. Conventional risk assessment techniques frequently depend on subjective expert judgments or oversimplified linear indices. Therefore, it may not adequately

reflect the complex dynamics of prolonged climatic changes. There is a pressing necessity for data-driven frameworks that integrate objective weighting systems with advanced predictive algorithms to deliver actionable insights for regional policymakers.

This study introduces a hybrid approach combining the Entropy Weight Method (EWM) with Random Forest regression to assess and predict climate stress in Ankara. The research employs an extensive longitudinal dataset covering nearly three decades (1997–2025) to create a Climate Stress Index (CSI) that objectively measures environmental risk. The Entropy approach is used to remove subjective bias by deriving weights from the intrinsic variances of six essential meteorological parameters: precipitation, average temperature, maximum temperature, minimum temperature, atmospheric pressure, and temperature range. A Random Forest model is subsequently trained to correlate these raw inputs with the CSI, yielding a very accurate predictive instrument.

The subsequent structure of the paper is outlined as follows. Section 2 addresses the literature review. Section 3 delineates the methodology and findings. Section 4 provided a discussion of the findings. Section 5 concludes the article.

2. Literature Review

Recent studies have increasingly emphasized integrating objective Multi-Criteria Decision-Making (MCDM) methodologies with ensemble learning algorithms to enhance the precision of climate-related risk assessments. Several studies have used EWM with the Random Forest algorithm in climate research. Wu et al. (2025) proposed a novel methodology for evaluating subway flood risks by comprehensively analyzing the hazard, exposure, vulnerability, and adaptive capacity of subway systems within the framework of urban flooding. The EWM was used to determine the weights for these indicators. A case study focused on the Beijing subway system. A multi-index system was created that includes three hazard indicators, eight exposure indicators, five vulnerability indicators, and seven coping capacity indicators. We used Pearson's correlation analysis and Random Forest regression, along with the Shapley Additive exPlanations approach, to identify the association between each assessment indicator and the flood risk level and quantify how much each indicator contributed to the flood risk. Ding et al. (2026) conducted a thorough assessment of county-level heat vulnerability in Henan Province using an entropy-weighted, machine-learning-enhanced index methodology. They employed the entropy method to assign various weights to different socio-environmental factors. After that, they employed a Random Forest regressor to examine how excessive heat affects public health. Li and Shi (2025) expanded the use of the entropy method to historical climate reconstruction. They employed the entropy method and Natural Language Processing (NLP) to change old weather records into standard grades. Some of the signs in that system were how severe the weather was, how it affected farming, the economy, and society, and how many people died. The entropy method and expert opinion were used to assign each indicator a scientific, objective weight. They employed a Random Forest model to check their index. Their research underscores the importance of objective weighting in assigning scientific values to qualitative variables, such as weather intensity and societal impact, thereby creating a reliable basis for long-term climate analysis.

General climate risk frameworks provide a comprehensive picture of environmental instability. At the same time, a specific literature stream focuses on rainfall-related problems, such as drought and waterlogging, using EWM and Random Forests together. Bezdan et al. (2025) suggested an impartial approach to assessing waterlogging risk, particularly in predominantly agricultural rural areas. The risk of waterlogging was assessed by examining the susceptibility and hazard posed by essential environmental, anthropogenic, and climatic factors. The EWM was used to determine the weights of the factors that affect the likelihood of waterlogging, ensuring that the overall evaluation results were reasonable. By comparing the generated waterlogging risk map with actual and known waterlogged areas using Sentinel-2 data and Random Forest classification, it was confirmed. Their research emphasizes that objective weighting is essential for appropriately depicting the spatial complexities of water-related disasters, providing a more reliable basis for water management strategies than subjective approaches. Zhu et al. (2025) investigated the resilience of terrestrial ecosystems to drought stress within the Yellow River Basin. Their method uses EWM to assign different weights to driving factors, such as precipitation, temperature, and soil moisture, to determine how drought-resistant each factor is. Han et al. (2024) developed a comprehensive evaluation methodology to assess the adaptation of maize varieties across diverse locations. Using the EWM, they combined nine agronomic characteristics to develop a Variety Adaptability Comprehensive Evaluation Index. After that, the index was used as the target variable in a machine learning pipeline that used Random Forest to predict yield stability.

The effectiveness of hybrid MCDM-machine learning models depends on the quality and relevance of the input meteorological features. Zhang et al. (2026) proposed an advanced framework for short-term forecasting that employs meteorological feature selection and hybrid optimization clustering. Their study showed that meteorological factors, including wind speed, temperature, and air pressure, are the most important external factors in complex forecasting situations. The study improved the precision of nonlinear prediction models by using optimization techniques to identify critical meteorological factors. This highlights an essential analogy to the current study: just as feature selection identifies the most relevant inputs for power prediction, the EWM identifies the most critical factors influencing climate stress.

Research demonstrates a significant methodological advancement in combining EWM with the Random Forest algorithm. This combination works especially well because the EWM objectively weights environmental factors based on facts, and the Random Forest algorithm is good at handling the nonlinearity in weather data. These hybrid models have been employed for many risks, including urban flooding, heat vulnerability, and agricultural adaptability. Nonetheless, a significant shortcoming remains in their use for long-term, localized climate assessment in semi-arid urban environments like Ankara. Most current research focuses on short-term predictions or specific agricultural outcomes, underscoring the need for a comprehensive diagnostic tool that integrates decades of climate data. This study addresses the deficiency by developing a tailored CSI for Ankara, incorporating objective multi-criteria weighting with sophisticated machine learning to improve regional economic and environmental planning.

3. Methodology

The proposed method employs a hybrid framework that integrates MCDM with ensemble learning. The core concept of this hybrid framework is the collaboration between the EWM and the Random Forest algorithm. EWM offers a mathematically precise method to mitigate subjective bias by assigning weights based on the intrinsic variance of meteorological data. At the same time, Random Forest provides the computational flexibility to correlate these multidimensional inputs into a consolidated risk metric. Employing nearly 30 years of longitudinal data (1997–2025), the methodology ensures that the resulting CSI is both historically grounded and predictive. The procedure is divided into four principal stages:

3.1 Data Collection

This study employs a collection of historical meteorological observations for Ankara, Türkiye, covering the period from 1997 to 2025. To ensure seasonal representation and cycle accuracy, the data were organized at the monthly resolution, yielding 347 observations. Each observation consists of six variables: average temperature (monthly mean temperature), minimum and maximum temperature, precipitation (total monthly rainfall in millimeters), atmospheric pressure (mean sea-level pressure), and temperature range (the difference between monthly maximum and minimum temperatures). The data were obtained using the Meteostat library, which consolidates observations from national and international meteorological stations (e.g., NOAA, DWD, and Environment Canada) (Meteostat, 2026). For data quality, a complete-case approach was applied, any month for which the source did not provide a full set of the six variables was excluded rather than imputed, ensuring that the entropy weighting and the CSI computation relied solely on fully observed records. After this step, the final dataset contained 347 complete monthly observations with no missing values in any field.

3.2 Data Pre-processing and Normalization

Min-Max Normalization was employed to standardize all criteria on a comparable scale. In the drought analysis, precipitation was treated as a 'benefit' criterion and inverted to correspond with the 'cost' characteristics of the temperature and pressure variables. This indicates that a score of 1 indicates the highest climate stress, while 0 denotes the lowest.

3.3 Objective Weighting via the Entropy Weight Method

The Entropy Weight Method (EWM) was introduced by Shannon (1948) to ascertain objective weights. EWM offers a substantial advantage over subjective weighting models by removing the impact of human variables on indicator weights, hence augmenting the objectivity of the overall evaluation outcomes (Eligüzel & Eligüzel, 2024; Y. Zhu et al., 2020). The decision matrix is normalized to eliminate the effects of varying measurement units. This study defines precipitation as a maximization criterion (benefit), whereas temperature variables, pressure, and temperature range are characterized as minimization criteria (cost) to illustrate climate stress. The normalization is executed via Equation (1):

$$e_{ij} = \begin{cases} \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} & \text{for maximization criteria} \\ \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})} & \text{for minimization criteria} \end{cases} \quad (1)$$

Entropy values for the criteria are calculated using Eq. 2.

$$E_j = \frac{\sum_{i=1}^m e_{ij} \ln(e_{ij})}{\ln(m)} \quad (2)$$

m represents the total number of monthly observations (347) from 1997 to 2025.

The weight of each criterion is calculated as Eq. 3.

$$w_j = \frac{1 - E_j}{\sum_{j=1}^n (1 - E_j)} \quad (3)$$

The EWM was used to objectively assess the importance of each criterion (see Table 1).

Table 1. Weights of Criteria

Average temperature	Minimum temperature	Maximum temperature	Precipitation	Pressure	Temperature range
0.155744	0.141829	0.165903	0.304304	0.056703	0.175517

The Entropy-based weighting determined Precipitation (0.3043) as the primary factor influencing climatic stress in Ankara, followed by Temperature Range (0.1755) and Maximum Temperature (0.1659). These findings align with the semi-arid climatic traits of Central Anatolia, where water scarcity and significant temperature variations constitute the principal environmental risks.

3.4 Calculation of CSI

CSI for each observation was computed in Eq. 4 as a linear weighted summation of the normalized criteria.

$$CSI_i = \sum_{j=1}^n w_j \cdot x'_{ij} \quad (4)$$

3.5 Random Forest Regression

A Random Forest regressor was employed to describe the non-linear relationship between raw meteorological inputs and the computed CSI. This approach, unlike conventional time-series models, treats each month as an independent data point and employs cross-sectional analysis. The model was trained with 200 decision trees, using an 80%/20% split for training and testing data. The Random Forest model achieved an R-squared score of 0.9881, indicating that it explains nearly 99% of the variance in climatic stress. The Root Mean Square Error (RMSE) is roughly 0.0193, indicating a remarkably low value for an index between 0 and 1, hence demonstrating good prediction accuracy. It should be noted that this high accuracy is expected by construction: the target CSI is computed as a weighted linear combination of the same six normalized meteorological variables that serve as model inputs. The task therefore consists of recovering a deterministic, closed-form index from its own constituent features rather than forecasting an independent external outcome, and even an ordinary linear model reproduces the index almost perfectly. The Random Forest step should thus be interpreted as a validation that a non-parametric ensemble can faithfully reconstruct the multi-criteria index, and as a means of ranking the relative discriminative importance of each variable (Figure 1), rather than as evidence of predictive skill on unseen climate events. The significance of the input variables was assessed to identify the fundamental factors influencing the Random Forest model's decision-making process. The EWM provided the foundational mathematical basis for the index. At the same time, the feature importance analysis (see Figure 1) identifies the meteorological factors with the greatest predictive capacity for correlating current climate states with stress levels.

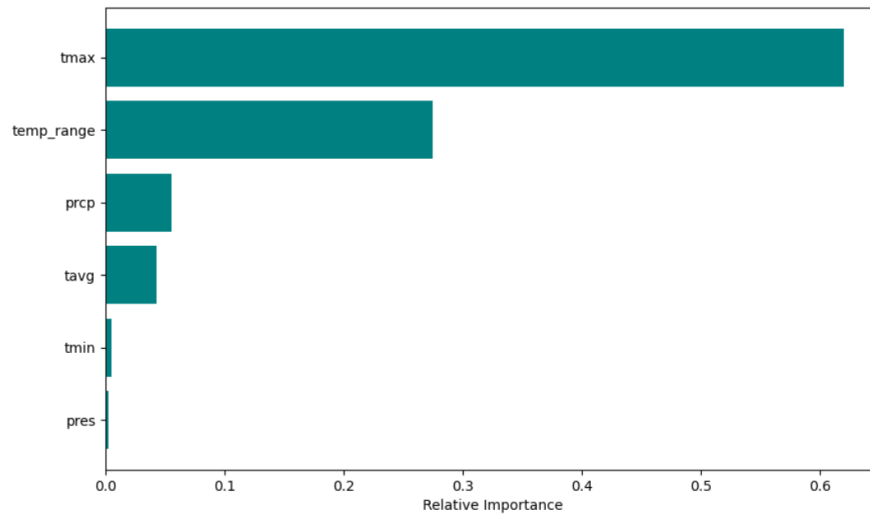


Figure 1. The Importance of Variables in the Decision-Making Mechanism

Figure 1 demonstrates that Maximum temperature and Temperature range are the most prominent predictors in the model. This indicates that climate stress in Ankara is characterized predominantly by extreme heat events and significant thermal variability. In a continental climate like Ankara's, these variables serve as key markers of atmospheric stability and thermal stress, often tipping the overall index. Notably, while Precipitation had the highest weight in the entropy phase, its relative significance in the Random Forest model is ranked somewhat lower than that of heat extremes. This indicates that although precipitation is an essential structural element of the CSI, the model predominantly depends on temperature variations to distinguish minor changes in monthly risk profiles. Atmospheric pressure and minimum temperature have the least relative significance. This affirms that these variables, although providing essential background, do not significantly influence the variance in drought stress levels relative to the predominant heat- and moisture-related factors. Figure 2 depicts the correlation between the Actual CSI, obtained from the Entropy-weighted meteorological variables, and the Predicted CSI produced by the Random Forest regressor.

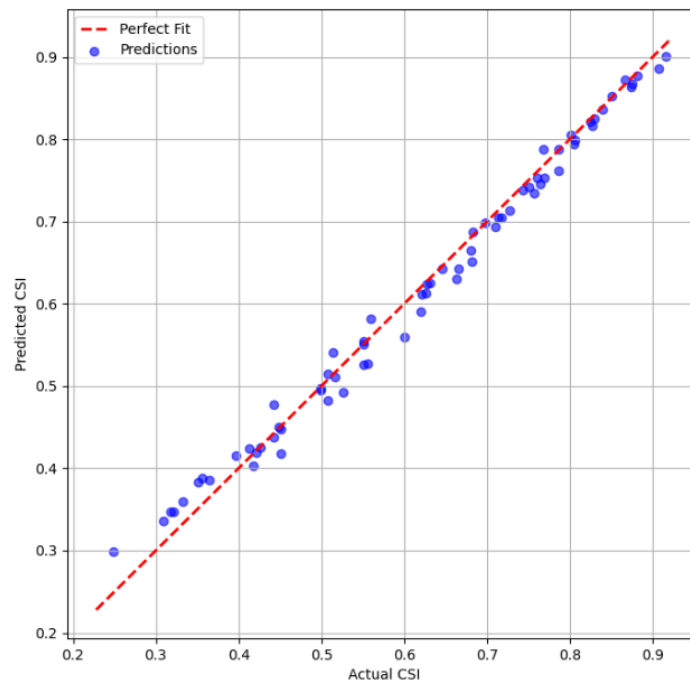


Figure 2. Random Forest Point Prediction Performance

The most notable aspect of Figure 2 is the nearly flawless alignment of the data points (blue circles) along the 45° identity line (red dashed line). This linear consistency demonstrates that the model has effectively encapsulated the fundamental mathematical principles of the multi-criteria index. The limited dispersion from the diagonal indicates that the Random Forest algorithm effectively captures the overall trend while remaining extremely responsive to the unique differences in each monthly data point.

4. Results and Discussion

The Random Forest model demonstrated high predictive accuracy, achieving an R^2 score of 0.9881 and an RMSE of 0.0193. These metrics indicate that the model accounts for approximately 99% of the variance observed in the CSI. Furthermore, the scatter plot analysis (Figure 2) reveals a near-perfect alignment of the data points along the identity line. This shows that the ensemble learner can successfully decode the complex, non-linear mathematical reasoning embedded in the EWM-weighted index. An intriguing finding from this investigation is the observed shift in variable significance between the weighting and prediction stages. The EWM analysis identified Precipitation (0.3043) as the most critical structural component, primarily because of its considerable variability in semi-arid regions. Conversely, the Random Forest model identified Maximum Temperature and Temperature Range as the most influential predictors. This divergence suggests that, although water scarcity (precipitation) is the primary driver of climatic stress in Ankara, thermal fluctuations serve as the principal indicator used by the machine learning model to differentiate among varying degrees of environmental pressure. As global temperatures rise rapidly, these extreme heat events are emerging as the primary driver of instability, affecting diverse areas, including public health and urban energy demand. Our research indicates that Ankara's vulnerability is increasingly defined by heat-related events rather than solely by water scarcity. Consequently, regional policymakers should shift their priorities from conventional water management strategies to include microclimatic thermal mitigation and adaptive cooling interventions. Furthermore, the model's robust performance on cross-sectional data suggests that even a single month of climate data can effectively inform a reliable risk profile. This offers a transparent, objective framework for regional assessment, thereby facilitating prompt action in Central Anatolia, where environmental challenges are intensifying.

5. Conclusion

This study successfully developed and verified a hybrid analytical framework for measuring and forecasting climatic stress in Ankara, Türkiye. This research synthesizes nearly three decades of meteorological data (1997–2025) into an objective CSI, offering a statistically rigorous diagnostic analysis for regional environmental management. The Random Forest model achieved an impressive R^2 score of 0.9881 and a low RMSE of 0.0193, thereby demonstrating the capacity of ensemble learning algorithms to model complex multi-criteria weighting processes with considerable precision.

This study is nonetheless subject to several assumptions and limitations. The CSI is formulated as a linear, additive aggregation of six normalized meteorological indicators, and the entropy weights are derived from the statistical dispersion of the specific 1997–2025 Ankara dataset, consequently, both the weights and the resulting index are context-dependent and may not transfer directly to other regions or periods. The framework adopts a cross-sectional perspective that treats each month as an independent observation and therefore does not explicitly model temporal trends, seasonality, or autocorrelation. Moreover, the CSI is an internally constructed relative measure that has not been validated against external impact data such as agricultural losses, water-supply deficits, or officially declared drought events, and the analysis relies on a single semi-arid station. These constraints should be considered when interpreting the results and motivate the directions outlined below.

Future research endeavors will focus on extending this approach to a multi-provincial scale, encompassing all 81 provinces of Türkiye, to produce a comprehensive national climate risk map. Furthermore, incorporating socio-economic indicators, such as industrial power consumption and regional water demand, into the hybrid model could enhance its effectiveness as a holistic sustainable development instrument.

References

- Bezdan, A., Bezdan, J., Markovi, M., Baumgartel, A., and Salvai, A., An objective methodology for waterlogging risk assessment based on the entropy weighting method and machine learning, *Catena*, vol. 249, p. 108618, 2025. <https://doi.org/10.1016/j.catena.2024.108618>
- Çelik, M. F., Pınarlık, M., and Selek, Z., Seasonal and annual time period precipitation trend analysis, case study of Ankara province, *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 40, no. 2, pp.

- 1279–1298, 2025. <https://doi.org/10.17341/gazimmfd.1334926>
- Chhiddikur, M. and Shajedur, R., Climate change and risk scenario in Bangladesh, *Asia-Pacific Journal of Regional Science*, vol. 7, no. 2, pp. 381–404, 2023. <https://doi.org/10.1007/s41685-022-00252-9>
- Ding, D., Chen, Y., Wang, H., Liu, H., Li, S., Xing, T., Yang, W., Xue, M., Yu, Z., and Gao, Z., Assessing and managing heat vulnerability in Henan Province, China: An entropy-weighted and random forest framework, *Environmental and Sustainability Indicators*, vol. 30, p. 101169, 2026. <https://doi.org/10.1016/j.indic.2026.101169>
- Eligüz el, İ. and Eligüz el, N., Evaluation of Sustainable Transportation in 25 European Countries Using GRA and Entropy MABAC, *Journal of Transportation and Logistics*, vol. 9, no. 2, pp. 245–259, 2024. <https://doi.org/10.26650/jtl.2024.1437521>
- Han, Y., Wang, K., Zhang, Q., Yang, F., and Pan, S., Developing a comprehensive evaluation model of variety adaptability based on machine learning method, *Field Crops Research*, vol. 306, p. 109203, 2024. <https://doi.org/10.1016/j.fcr.2023.109203>
- Kc, B., Anthony, J. M. S., Cassandra, W. K., and Gaither, J., Multi-hazard climate risk projections for the United States, *Natural Hazards*, vol. 105, no. 2, pp. 1963–1976, 2021. <https://doi.org/10.1007/s11069-020-04385-y>
- Li, T. and Shi, T., High-resolution climate reconstruction from historical Chinese weather records using optimized natural language processing, *Scientific Reports*, vol. 15, no. 44447, pp. 1–17, 2025.
- Meteostat, Meteostat: Open weather and climate data for developers, *Meteostat Web Platform*, <https://meteostat.net/>, 2026.
- Pandit, J., A comprehensive review of climate change’s imprint on ecosystems, *Journal of Water and Climate Change*, vol. 14, no. 11, pp. 4273–4284, 2023. <https://doi.org/10.2166/wcc.2023.476>
- Shannon, C. E., A Mathematical Theory of Communication, *The Bell System Technical Journal*, vol. XXVII, no. 3, pp. 379–423, 1948. [https://doi.org/10.1016/s0016-0032\(23\)90506-5](https://doi.org/10.1016/s0016-0032(23)90506-5)
- Wanke, P. and Tan, Y., Climate risk assessment at Chinese provinces: A novel Criteria and Context Decomposition for Trade-Off Weighting (CCDTW) approach, *Journal of Environmental Management*, vol. 392, p. 126734, 2025. <https://doi.org/10.1016/j.jenvman.2025.126734>
- Wu, W., Zhu, Z., Zheng, X., Liu, S., Zhang, D., and Peng, D., A flood risk assessment framework for urban subway systems: A case study in Beijing, China, *Sustainable Cities and Society*, vol. 131, p. 106681, 2025. <https://doi.org/10.1016/j.scs.2025.106681>
- Zhang, Z., Ge, Q., Jia, C., Li, L., and Lei, J., Short-term wind power prediction based on hybrid optimization clustering, *International Journal of Systems Science*, pp. 1–15, 2026. <https://doi.org/10.1080/00207721.2026.2618571>
- Zhu, X., Huang, S., Singh, V. P., Huang, Q., and Zhang, H., Terrestrial ecosystem resilience to drought stress and driving mechanisms thereof in the Yellow River Basin, China, *Journal of Hydrology*, vol. 649, p. 132480, 2025. <https://doi.org/10.1016/j.jhydrol.2024.132480>
- Zhu, Y., Tian, D., and Yan, F., Effectiveness of Entropy Weight Method in Decision-Making, *Mathematical Problems in Engineering*, vol. 2020, pp. 1–5, 2020. <https://doi.org/10.1155/2020/3564835>

Biographies

Assist. Prof. İbrahim Miraç ELİGÜZEL is a faculty member at Gaziantep Islam Science and Technology University. He obtained his Ph.D. and Master’s degree in Industrial Engineering from Gaziantep University. Dr. Eligüz el has made significant contributions to his profession through his publications and research, with a particular emphasis on machine learning, mathematical modeling, and metaheuristic algorithms. He is widely recognized for his expertise in supply chain networks, meta-heuristics, and humanitarian supply chain management. Dr. Eligüz el remains committed to advancing knowledge and fostering academic excellence among his students and colleagues.

Assoc. Prof. Nazmiye ELİGÜZEL is a faculty member at Gaziantep Islam Science and Technology University. She holds both a Ph.D. and a Master’s degree in Industrial Engineering from Gaziantep University. With a specialized focus on machine learning and natural language processing, she has made substantial contributions to her field through numerous publications and innovative research projects. Dr. Eligüz el is widely recognized for her expertise in machine learning, disaster management, and natural language processing, and she remains dedicated to advancing knowledge and nurturing academic excellence in her students and peers.

Assoc. Prof. Sena AYDOĞAN is a faculty member at Gazi University. She holds both a Ph.D. and an M.Sc. degree in Industrial Engineering from Gazi University. With a specialized focus on linguistic summarization, machine learning, and natural language processing, she has made substantial contributions to her field through numerous

publications. Dr. Aydoğan is widely recognized for her expertise in linguistic summarization, fuzzy sets and systems, and machine learning applications.