Journal of Infrastructure, Policy and Development 2024, 8(10), 6807.
https://doi.org/10.24294/jipd.v8i10.6807

Article

Using ensemble learning method and binary decision tree algorithm for drought intensity level classification

Oksana Kukartseva[1,2], Vadim Tynchenko[3,4], Vladislav Kukartsev[3,5,*], Tatyana Panfilova[1,6]

[1] Center for Continuing Education, Bauman Moscow State Technical University, 105005 Moscow, Russia
[2] Laboratory of Biofuel Compositions, Siberian Federal University, 660041 Krasnoyarsk, Russia
[3] Artificial Intelligence Technology Scientific and Education Center, Bauman Moscow State Technical University, 105005 Moscow, Russia
[4] Information-Control Systems Department, Institute of Computer Science and Telecommunications, Reshetnev Siberian State University of Science and Technology, 660037 Krasnoyarsk, Russia
[5] Department of Information Economic Systems, Reshetnev Siberian State University of Science and Technology, 660037 Krasnoyarsk, Russia
[6] Department of Technological Machines and Equipment of Oil and Gas Complex, Siberian Federal University, 660041 Krasnoyarsk, Russia
*** Corresponding author:** Vladislav Kukartsev, vlad_saa_2000@mail.ru

CITATION

Kukartseva O, Tynchenko V, Kukartsev V, Panfilova T. (2024). Using ensemble learning method and binary decision tree algorithm for drought intensity level classification. Journal of Infrastructure, Policy and Development. 8(10): 6807.
https://doi.org/10.24294/jipd.v8i10.6807

ARTICLE INFO

Received: 31 May 2024
Accepted: 27 June 2024
Available online: 23 September 2024

Abstract: This study applies machine learning methods such as Decision Tree (CART) and Random Forest to classify drought intensity based on meteorological data. The goal of the study was to evaluate the effectiveness of these methods for drought classification and their use in water resource management and agriculture. The methodology involved using two machine learning models that analyzed temperature and humidity indicators, as well as wind speed indicators. The models were trained and tested on real meteorological data to assess their accuracy and identify key factors affecting predictions. Results showed that the Random Forest model achieved the highest accuracy of 94.4% when analyzing temperature and humidity indicators, while the Decision Tree (CART) achieved an accuracy of 93.2%. When analyzing wind speed indicators, the models' accuracies were 91.3% and 93.0%, respectively. Feature importance revealed that atmospheric pressure, temperature at 2 m, and wind speed are key factors influencing drought intensity. One of the study's limitations was the insufficient amount of data for high drought levels (classes 4 and 5), indicating the need for further data collection. The innovation of this study lies in the integration of various meteorological parameters to build drought classification models, achieving high prediction accuracy. Unlike previous studies, our approach demonstrates that using a wide range of meteorological data can significantly improve drought classification accuracy. Significant findings include the necessity to expand the dataset and integrate additional climatic parameters to improve models and enhance their reliability.

Keywords: sustainable growth; agricultural development; land management; soil fertility; agricultural innovation

1. Introduction

Drought is one of the most destructive natural disasters, characterized by a prolonged deficit of precipitation, leading to a lack of soil moisture. Its causes are multifaceted, encompassing natural climatic fluctuations, changes in atmospheric and oceanic circulation, and anthropogenic factors such as climate change (Mokhtar, 2021; Rhee, 2017; Tufaner, 2020). Drought affects millions of people worldwide, causing food crises and economic losses (Richman, 2016). The consequences of drought are extensive and varied, ranging from reduced crop yields and agricultural losses to ecosystem degradation and depletion of water resources.

Unlike sudden disasters such as earthquakes or hurricanes, drought develops gradually, making its prediction and management challenging (Shamshirband, 2020). This gradual development necessitates long-term monitoring and early warning systems (Bashmur, 2022; Mosavi, 2023). With the ongoing global climate change, the importance of drought monitoring and forecasting has significantly increased (Deo, 2015). Modern monitoring methods primarily include the use of satellite data and meteorological measurements (Khan, 2020), each having its own set of advantages and limitations.

Satellite data provide valuable information about soil and vegetation conditions over large areas (Kolenchukov, 2022). However, these data often suffer from limitations in resolution and accuracy (Dikshit, 2021). In contrast, meteorological measurements, such as those for precipitation and temperature, offer more precise data but are limited in their spatial coverage (Rahmati, 2020).

The integration of these diverse data sources is crucial for effective drought forecasting, which in turn helps minimize the impacts of drought, develop adaptation measures, and improve water resource management (Aghelpour, 2020). Current research is increasingly focused on enhancing drought monitoring and forecasting methods through the use of advanced technologies and innovative scientific approaches (Deo, 2017).

Machine learning plays a key role in data analysis and drought forecasting. This integration is illustrated in **Figure 1**. These algorithms can process large volumes of data, identify hidden patterns, and make accurate predictions (Dikshit, 2020; Rahmati, 2020). For example, deep learning models such as neural networks are successfully used for long-term drought forecasting, utilizing meteorological and climatic data (Kolachian, 2021; Raza, 2022).

Figure 1. Integration of machine learning methods for drought monitoring and forecasting.

The use of machine learning in drought forecasting involves developing models that take into account various climatic indices such as the Palmer Drought Severity Index (PDSI), Standardized Precipitation Index (SPI), and Soil Moisture Index (SMP) (Alkan, 2023; Liu, 2023). These models help predict the intensity and duration of droughts, which is crucial for agricultural planning and water resource management (Sharma, 2021; Vrindavanam, 2022). Successful applications of machine learning include weather forecasting, water resource management, and ecological system monitoring (Dikshit, 2021). For example, a study on the improved SPEI drought forecasting approach using the Long Short-Term Memory (LSTM) model showed that the LSTM model achieved an R^2 value of more than 0.99 for both SPEI 1 and SPEI 3, indicating very high accuracy. The model also demonstrated an AUC of 0.83 for SPEI 1 and 0.82 for SPEI 3 in ROC analysis, highlighting its effectiveness in drought category forecasting. These results show a significant improvement in forecasting capabilities compared to traditional machine learning models (Dikshit et al., 2021). Additionally, a study on the machine learning approach to flood severity classification developed a model with a 91% accuracy rate in classifying flood severity levels. The model successfully reduced false alarms by 25%, significantly enhancing the reliability of flood alerts, underscoring the potential of machine learning models in improving flood management and response systems. These methods demonstrate high accuracy and reliability, making them promising for use in various fields of science and technology (Sharma, 2021).

A study evaluating various machine learning techniques for hydrological drought forecasting in the Wadi Ouahrane basin in Algeria found that the Support Vector Machine (SVM) model outperformed other models such as Artificial Neural Networks (ANN) and Decision Trees (DT). The SVM model achieved an RMSE of 0.28, MAE of 0.19, NSE of 0.86, and an R^2 of 0.90, demonstrating high accuracy in predicting hydrological drought (Achite, 2022). Another study focused on drought prediction in the Yazd province of Iran using the Standardized Precipitation Index (SPI) and the Standardized Water-Level Index (SWI). This research highlighted the effectiveness of Random Forest (RF) and Gaussian Process Regression (GPR) models, with the RF model achieving a high R^2 of 0.85, indicating robust predictive capability for meteorological droughts (Elbeltagi, 2023). A novel deep learning model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for meteorological drought forecasting showed promising results. The model was validated with a high degree of accuracy in predicting drought events, emphasizing the advantage of integrating spatial and temporal features in the forecasting process (Dehghani, 2014).

Research and implementation of machine learning methods for classifying and forecasting drought intensity is a relevant and important task that helps mitigate the consequences of this natural disaster and improve resource management (Zhao, 2022). Despite successful examples of using machine learning in drought forecasting, further research is needed to improve model accuracy and adapt them to different climatic conditions (Yelemessov, 2023). For example, current models may not account for all important variables or their combinations, which can reduce forecast accuracy (Liu, 2023; Vrindavanam, 2022). Additionally, most studies focus on one- or a few-time scales, which may limit their applicability in other conditions (Sharma, 2021).

The implementation of new approaches, such as Decision Tree and Random Forest methods, is a promising topic in scientific research (Borodulin, 2024; Degtyareva, 2023). These methods demonstrate high accuracy and reliability in drought classification (Kolachian, 2021; Raza, 2022). Decision Tree and Random Forest methods have advantages in data interpretation and visualization, making them convenient for use in real applications (Bosikov, 2023; Martyushev, 2023). They also show high resistance to overfitting and can effectively work with large datasets (Alkan, 2023; Dikshit, 2021; Gohel, 2020).

Classifying droughts considering wind conditions will help improve the planning of infrastructure projects, such as the construction of reservoirs and irrigation systems, and the development of measures to protect ecosystems from degradation due to dry conditions (Aghelpour, 2020; Malozyomov, 2023). Additionally, the use of temperature and humidity data will allow for more accurate drought classification, which is important for planning agricultural activities, monitoring soil and vegetation conditions, and taking measures to mitigate the effects of climate change (Almikaeel, 2022). Wind speed indicators also play a significant role in the development of droughts and can be used to build effective classification models (Tynchenko, 2024).

The goal of this study is to develop models for classifying drought intensity based on machine learning methods such as Random Forest and Decision Tree (CART). Various groups of meteorological parameters, including temperature and humidity indicators, as well as wind speed indicators, were considered for this purpose. The study aims to develop effective models capable of predicting drought intensity, which can be useful for water resource management, agriculture, and infrastructure planning.

Hypotheses of the study:

1) Modeling based on meteorological parameters:
   - Models built on a wide range of meteorological parameters will be able to accurately classify drought intensity.
2) Modeling based on temperature and humidity indicators:
   - Temperature and humidity indicators are key factors influencing drought intensity, and their use in machine learning models will lead to high classification accuracy.
3) Modeling based on wind speed indicators:
   - Wind speed indicators significantly impact drought development, and their inclusion in machine learning models will improve the accuracy of drought intensity classification.

2. Materials and methods

Meteorological conditions and drought intensity data were obtained from the open dataset: US Drought Meteorological Data. This dataset was created based on information provided by drought experts in the United States and includes data on drought conditions and meteorological indicators for the years 2017 and 2018.

Although the United States already has a drought dataset created by drought experts, our study remains necessary for several key reasons. Traditional methods of drought classification, which rely heavily on expert assessments and drought indices, often fail to fully account for the complex interactions between various meteorological parameters. These conventional approaches primarily focus on straightforward metrics like precipitation and temperature, but they lack the integration of multiple interrelated factors that influence drought conditions. For example, they might not adequately consider the synergistic effects of soil moisture, evapotranspiration, and groundwater levels, which are crucial for a comprehensive understanding of drought dynamics (Durand, 2008; Murali, 2023). In contrast to these methods, we use modern machine learning algorithms that can analyze large volumes of data and identify hidden patterns, significantly improving the accuracy of drought forecasting and classification.

Moreover, the integration of various meteorological parameters, such as temperature, humidity, wind speed, and atmospheric pressure, into our models allows us to consider a broader range of factors affecting drought development. This contributes to the creation of more comprehensive and accurate models, which in turn helps improve water resource management and agricultural planning.

Thus, despite the existence of existing datasets and methods, our study provides new tools and approaches for more accurate and reliable drought forecasting, which is an important step in improving adaptation to changing climatic conditions and managing the impacts of drought.

The research presented in this article focuses on meteorological drought. Meteorological drought is defined as a period significantly below the average precipitation level, which leads to a shortage of moisture in the atmosphere and soil, and is the main cause of other types of droughts, such as hydrological and agricultural (Achite, 2022; Chu, 2018). Using meteorological data such as temperature, humidity, and wind speed, it is possible to create models for accurate forecasting and classification of meteorological droughts.

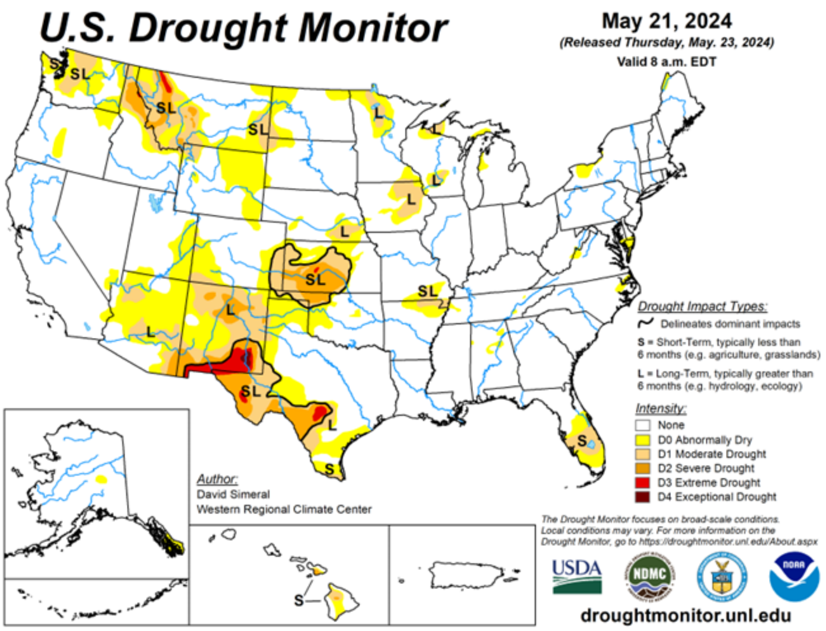

Figure 2. U.S. drought monitoring map for 21 May 2024.

Drought monitoring in the United States involves measurements of drought created manually by experts using a wide range of data. The goal of this dataset is to explore the possibility of predicting and classifying droughts using only meteorological data, which could potentially lead to the generalization of drought forecasts from the US to other regions of the world (**Figure 2**).

This dataset classifies data into six levels of drought: no drought (None) and five levels of drought, as shown below in **Table 1**.

Table 1. Drought categories and possible consequences.

Category	Description	Possible Impacts
D0	Abnormally Dry	Going into drought: • short-term dryness slowing planting, growth of crops or pastures Coming out of drought: • some lingering water deficits • pastures or crops not fully recovered
D1	Moderate Drought	• Some damage to crops. pastures • Streams, reservoirs, or wells low. some water shortages developing or imminent • voluntary water-use restrictions requested
D2	Severe Drought	• Crop or pasture losses likely water shortages common • water restrictions imposed
D3	Extreme Drought	• Major crop/pasture losses • Widespread water shortages or restrictions
D4	Exceptional Drought	• Exceptional and widespread crop/pasture losses • Shortages of water in reservoirs, streams. and wells creating water emergencies

Each entry in the dataset (a total of 543,357 records) represents a drought level at a specific moment in time in a specific US county, accompanied by data on 18 meteorological indicators over the past 90 days:

- WS10M_MIN: Minimum wind speed at 10 m (m/s)
- QV2M: Specific humidity at 2 m (g/kg)
- T2M_RANGE: Temperature range at 2 m (°C)
- WS10M: Wind speed at 10 m (m/s)
- T2M: Temperature at 2 m (°C)
- WS50M_MIN: Minimum wind speed at 50 m (m/s)
- T2M_MAX: Maximum temperature at 2 m (°C)
- WS50M: Wind speed at 50 m (m/s)
- TS: Earth skin temperature (°C)
- WS50M_RANGE: Wind speed range at 50 m (m/s)
- WS50M_MAX: Maximum wind speed at 50 m (m/s)
- WS10M_MAX: Maximum wind speed at 10 m (m/s)
- WS10M_RANGE: Wind speed range at 10 m (m/s)
- PS: Surface pressure (kPa)
- T2MDEW: Dew/Frost point at 2 m (°C)
- T2M_MIN: Minimum temperature at 2 m (°C)
- T2MWET: Wet bulb temperature at 2 m (°C)
- Output Parameter: Drought intensity classes (D0-D4)

Data preprocessing:

- Removal of missing values: missing values were removed from the dataset to avoid distorting the machine learning models.

- Data normalization: all numerical data were normalized to ensure comparability between different parameters.
- Categorical variable transformation: categorical variables, such as drought levels, were transformed into numerical format using one-hot encoding.

Some of the parameters used in the study are collected from ground-based meteorological stations, namely those that do not have the number "50" in their name. These parameters include temperature at an altitude of 2 m, humidity, and wind speed at an altitude of 10 m. The parameters, which contain the number "50" in the name, are collected from aerosols and balloons, which allows you to cover higher layers of the atmosphere. Data is also collected by county, which allows detailed analysis of the state of drought in various regions. On the U.S. website. Drought Monitor indicates that drought monitoring data is collected from a variety of stations operated by various agencies, including NOAA and the National Drought Mitigation Center (NDMC).

The following machine learning algorithms were used for drought intensity classification:

- Random Forest: An ensemble learning method that builds multiple decision trees and combines their results to improve classification accuracy.
- Decision Tree (CART): A classical decision tree algorithm that uses recursive partitioning to create a classification model.
- The models were trained on a dataset split into training and testing sets in an 80/20 ratio. The main hyperparameters for each model were optimized using the Random Search method.

The following metrics were used to evaluate the quality of the models:

- Accuracy: The proportion of correctly classified examples.
- Recall: The proportion of correctly classified positive examples among all positive examples.
- F1-score: The harmonic mean between precision and recall.

In traditional methods, drought is classified based on the range of some drought indices, such as the Standardized Precipitation Index (SPI) and Standardized Runoff Index (SRI). The advantages of the drought categorization method proposed in the article are the integration of a variety of meteorological parameters such as temperature, humidity, wind speed and atmospheric pressure, which allows us to take into account a wider range of factors affecting the development of drought and provide more accurate forecasts. The use of modern machine learning algorithms such as Random Forest and Decision Tree (CART) allows you to process large amounts of data and identify hidden patterns, providing higher accuracy in drought classification compared to traditional methods that often rely on linear dependencies.

3. Results

As part of the study, a correlation analysis was conducted to identify possible linear dependencies (Kozlova, 2023) between the input meteorological data and the output parameter—drought intensity. Correlation analysis allows determining the degree of interrelation between various meteorological indicators and drought intensity, measured through D0-D4 classes. This helps to understand if there are any strong correlation links that might indicate a direct dependence. **Figure 3** presents a correlation matrix showing Pearson correlation coefficients between all meteorological indicators and the output parameter (drought intensity).

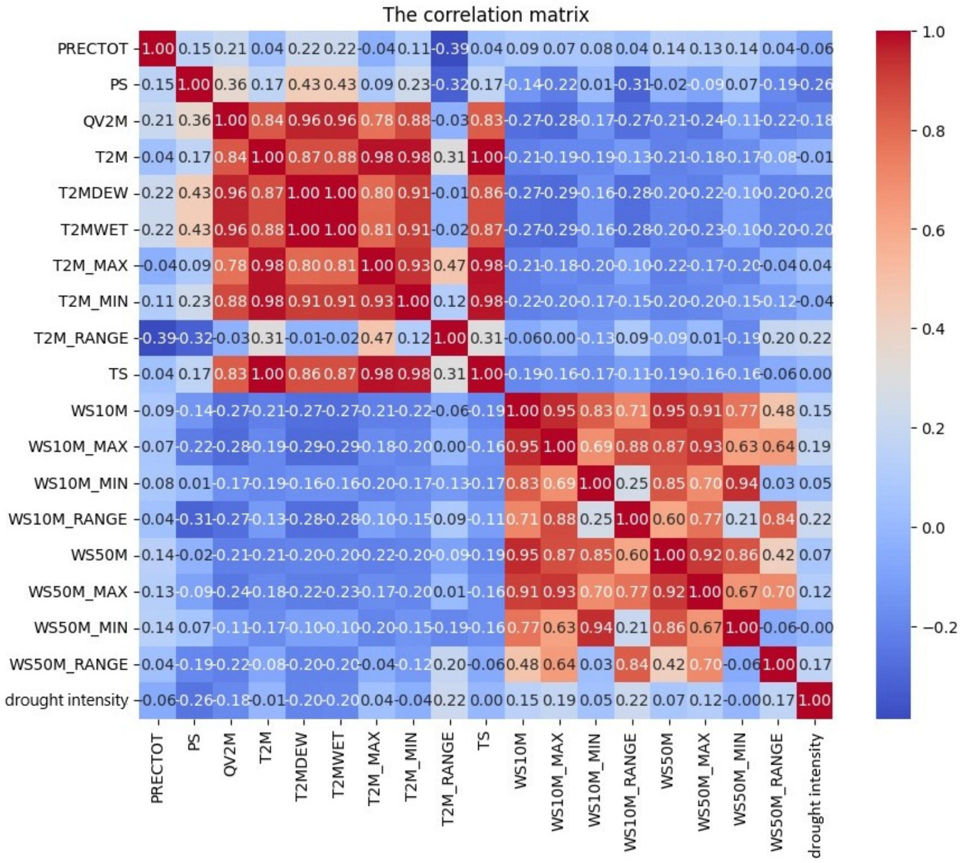

Figure 3. Correlation matrix of meteorological indicators and drought intensity.

The presented correlation matrix shows Pearson correlation coefficients between various meteorological indicators and drought intensity. The coefficients range from −1 to 1, where values close to 1 or −1 indicate strong positive or negative correlations, respectively, and values close to 0 indicate weak or no correlation.

The matrix reveals that indicators such as temperature at 2 m (T2M), minimum (T2M_MIN) and maximum (T2M_MAX) temperatures, and surface temperature (TS) have high positive correlations with each other, which is expected due to their physical relationships. However, the correlation between these parameters and drought intensity is relatively weak, with coefficients around 0.2 and below, indicating the necessity of using complex nonlinear models, such as machine learning algorithms, for more accurate drought classification.

Parameters for wind speed at 10 m (WS10M) and 50 m (WS50M) also show positive correlations with each other, but their relationship with drought intensity is similarly weak. This underscores the importance of integrating multiple meteorological indicators into machine learning models to improve forecast accuracy.

Thus, the presented data confirm the absence of strong linear dependencies between individual meteorological indicators and drought intensity, justifying the use of machine learning methods for more accurate drought classification and prediction. The source of the data is the entire dataset, providing the most comprehensive and accurate picture for analysis.

The next step towards achieving the goal was to build classification models on the entire dataset. The first algorithm used was the Decision Tree "CART". The feature importance of the model and the classification matrix with metrics are presented in **Figures 4–6**.

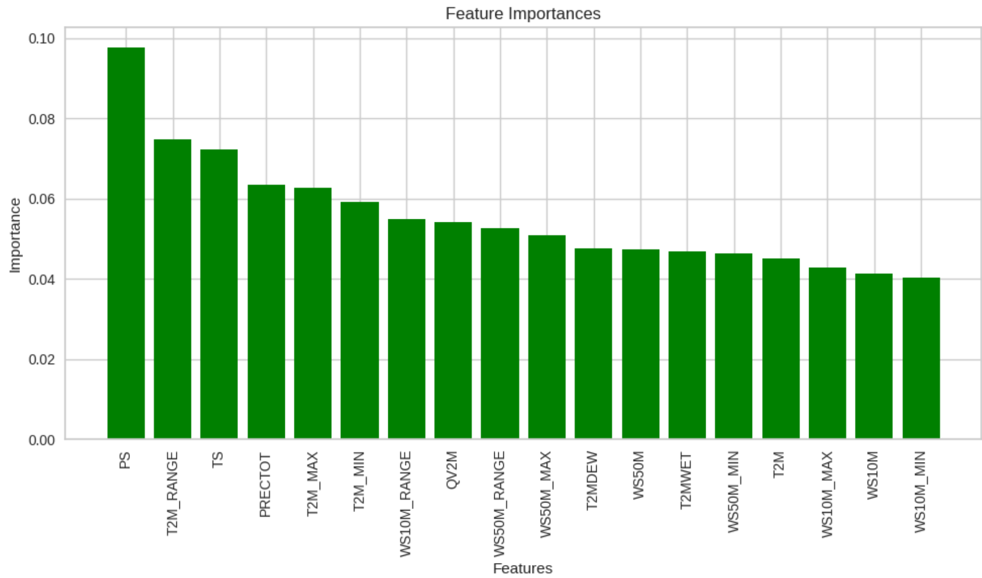

Figure 4. The importance of CART model features based on the entire dataset.

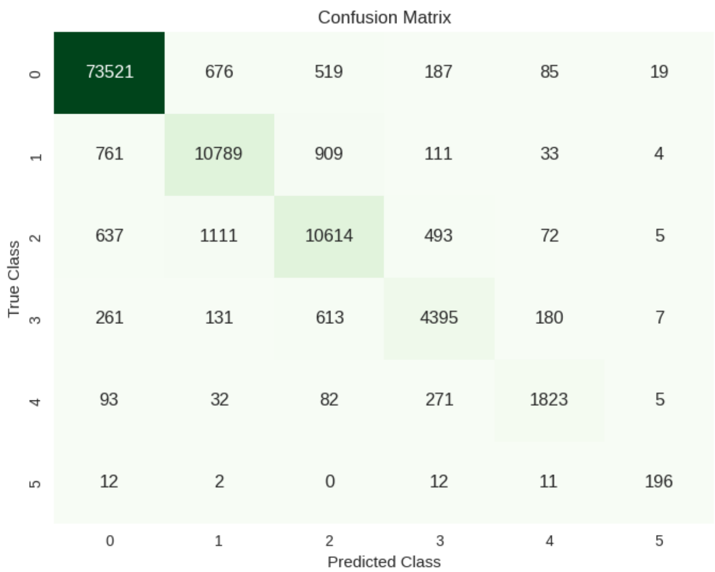

Figure 5. The error matrix of the CART model based on the entire dataset.

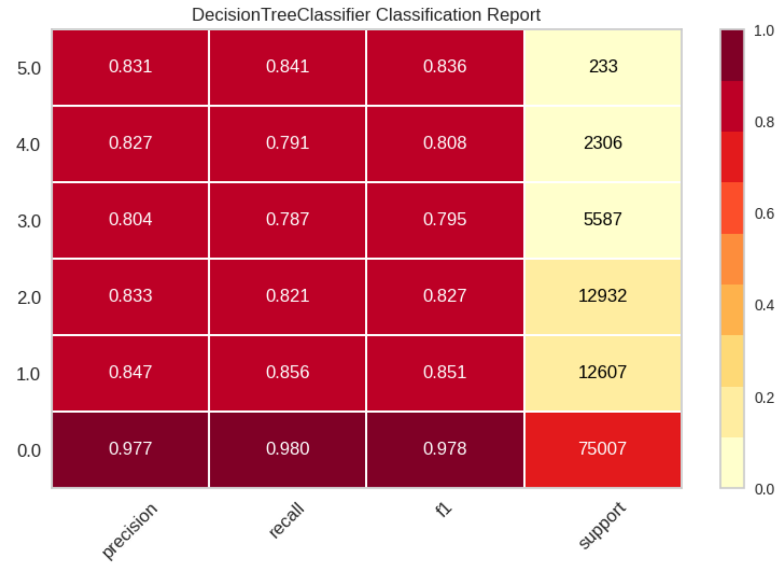

Figure 6. CART model classification report based on the entire dataset.

The classification matrix showed the distribution of true and predicted drought intensity classes. Values on the diagonal of the matrix represent the number of correctly classified examples for each class, while values off the diagonal indicate the number of classification errors. Based on the classification matrix, it was evident that the model achieved high classification accuracy for most classes.

Feature importance was visualized, allowing the identification of key meteorological parameters influencing the model's predictions. The most important features in the model were surface pressure (PS), temperature range at 2 m (T2M_RANGE), and earth skin temperature (TS). This helps to better understand which factors contribute most to predicting drought intensity.

A classification report was also presented, showing the precision, recall, and F1-score metrics for each class, as well as the number of examples (support) in each class. These metrics help evaluate the model's performance across different levels of drought intensity. The Decision Tree model achieved an overall accuracy of 93.2%, indicating high efficiency in classifying drought intensity based on meteorological data.

However, it is important to note the limitations related to the amount of data for classes 4 and 5. These classes are represented by fewer examples compared to other drought levels, which can affect the model's prediction accuracy for these classes. The small amount of data in these classes makes it difficult to train the model and may lead to reduced classification accuracy for severe drought levels.

Another method used for model building was "Random Forest". The entire dataset was used for training and testing the model. The feature importance of the model and the classification matrix are presented in **Figures 7** and **8**.

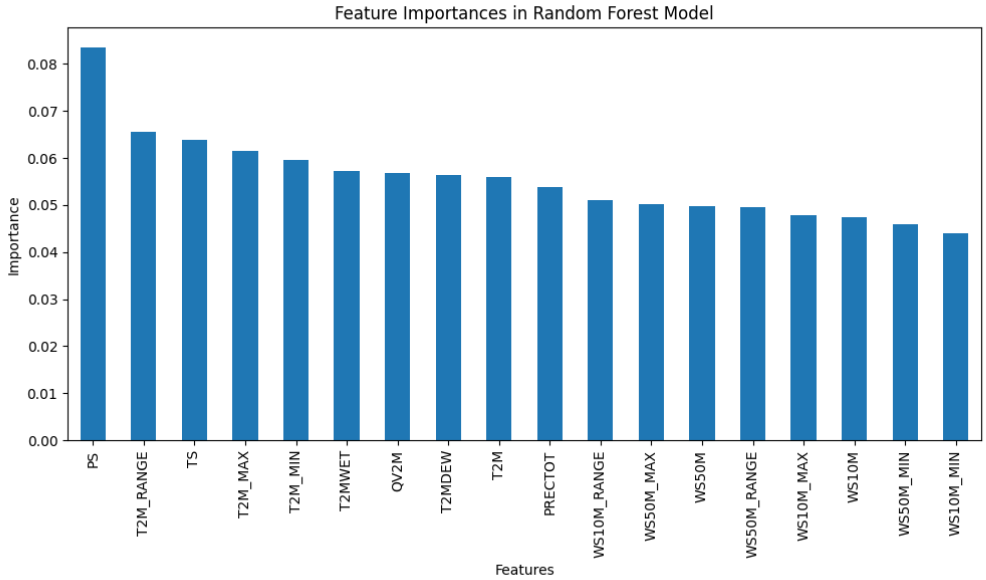

Figure 7. The importance of features of the Random Forest model based on the entire dataset.

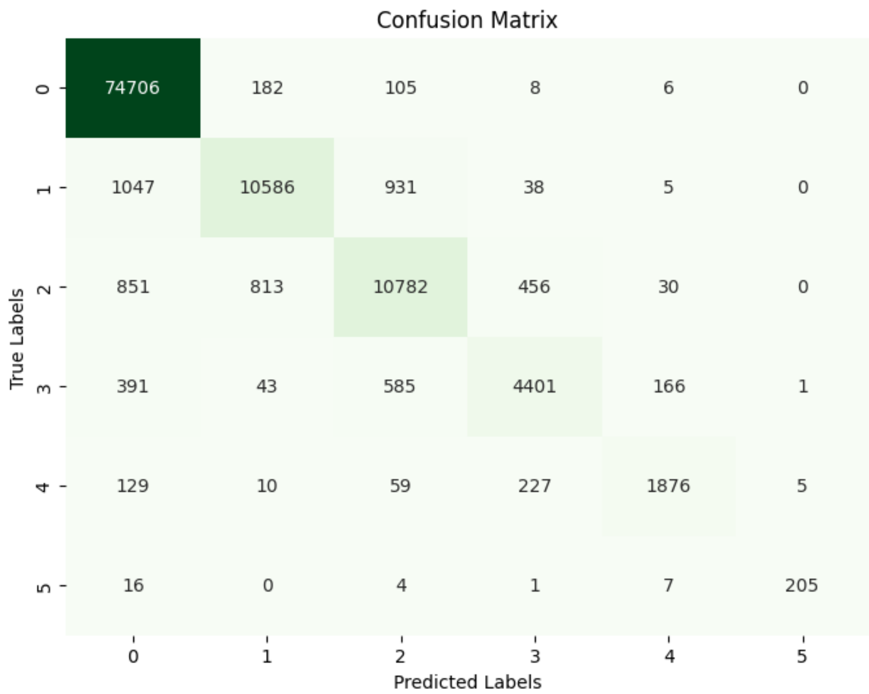

Figure 8. The error matrix of the Random Forest model based on the entire dataset.

The Random Forest model for classifying drought intensity based on meteorological data achieved an overall accuracy of 94.4%, demonstrating high efficiency. The classification matrix showed that the model successfully distinguishes most drought classes, however, there are difficulties in classifying classes 4 and 5 due to the limited amount of data for these levels. The importance of the features revealed key meteorological parameters such as atmospheric pressure, a temperature range at an altitude of 2 m and the temperature of the earth's surface, which have a significant impact on the predictions of the model. The limited amount of data for high levels of drought highlights the need to collect additional data to improve model accuracy for these classes.

Conducting experiments on building models on a group of meteorological parameters

The models were divided based on the aim and hypotheses of the study. The primary goal of our research was to develop models for classifying drought intensity using various groups of meteorological parameters, including temperature and humidity indicators, as well as wind speed metrics. In the introduction of the article (see the Introduction section), the research hypotheses are described, according to which models built on different groups of parameters should provide high accuracy in drought classification. This division allows for a detailed study of the influence of each group of parameters on classification accuracy and helps to identify which factors are most significant.

To conduct the experiment on building models based on various classification methods, the following groups of factors were selected:

1) Temperature and Humidity Indicators:
    - T2M: Temperature at 2 m (°C)—the main indicator of air temperature near the ground surface.
    - QV2M: Specific Humidity at 2 m (g/kg)—the amount of water vapor in the air at a height of 2 m.
    - T2M_MAX: Maximum temperature at 2 m (°C)—the highest temperature recorded over a specific period.
    - T2M_MIN: Minimum temperature at 2 m (°C)—the lowest temperature recorded over a specific period.
    - T2M_RANGE: Temperature range at 2 m (°C)—the difference between the maximum and minimum temperatures.
    - T2MDEW: Dew point at 2 m (°C)—the temperature at which the air at a height of 2 m becomes saturated and condensation begins.
    - T2MWET: Wet Bulb temperature at 2 m (°C)—the temperature measured using a wet bulb thermometer, indicating cooling due to evaporation.
2) Wind Speed Indicators:
    - WS10M_RANGE: Wind speed range at 10 m (m/s)—the difference between the maximum and minimum wind speeds at a height of 10 m.
    - WS50M_RANGE: Wind speed range at 50 m (m/s)—the difference between the maximum and minimum wind speeds at a height of 50 m.
    - WS10M_MAX: Maximum wind speed at 10 m (m/s)—the highest wind speed recorded at a height of 10 m.
    - WS50M_MAX: Maximum wind speed at 50 m (m/s)—the highest wind speed recorded at a height of 50 m.

Each of the above parameters was selected to analyze their influence on drought intensity. Temperature and humidity indicators provide data on the temperature and humidity conditions that can affect drought development. Wind speed indicators help understand the impact of wind conditions on regional dryness. Atmospheric pressure and specific humidity play important roles in shaping weather conditions that influence drought levels.

Figures 9 and **10** show the correlation matrices for the two classification methods, allowing a visual assessment of the relationships between the selected parameters and drought intensity.

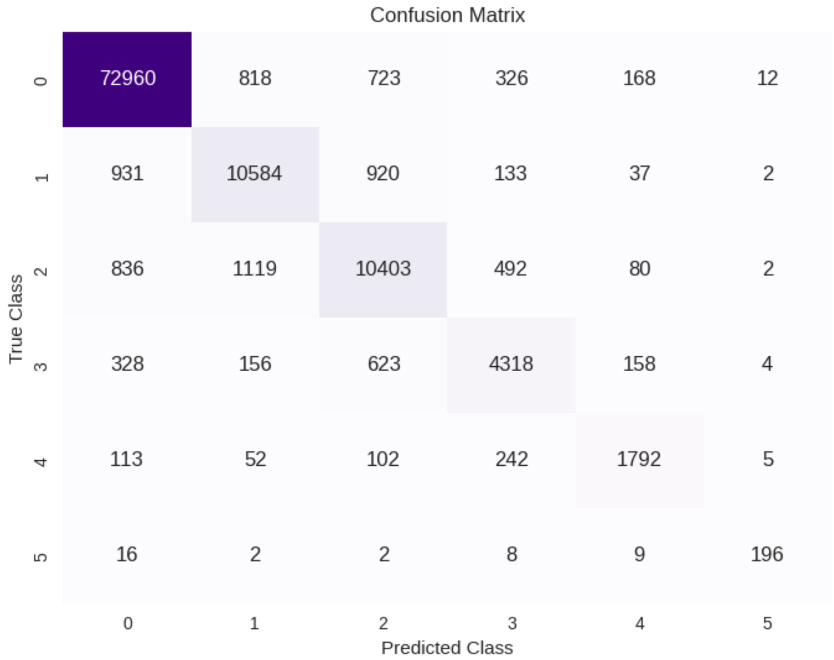

Figure 9. Error matrix of the CART model of the first experiment.

Figure 10. The error matrix of the Random Forest model of the first experiment.

During the experiment, drought intensity classification models were built using two methods: Decision Tree (CART) and Random Forest. The first model (CART) achieved an accuracy of 92.2%, while the second model (Random Forest) reached 93.4%.

The classification matrices for both methods are presented in the figures above. These matrices show that both models successfully distinguish most drought classes;

however, there are difficulties in classifying classes 4 and 5 due to the limited amount of data for these levels.

The feature importance and classification report for the models were also visualized, showing the key parameters influencing the models' predictions, as well as the precision, recall, and F1-score metrics for each class. The visualizations of feature importance and classification reports are presented in **Figures 11** and **12**.

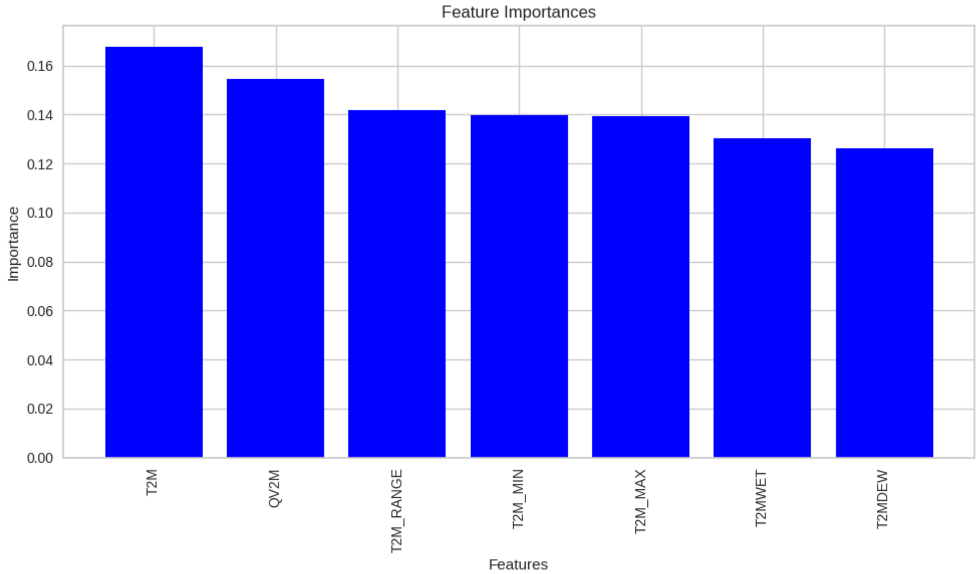

Figure 11. The importance of the features of both models of the first experiment.

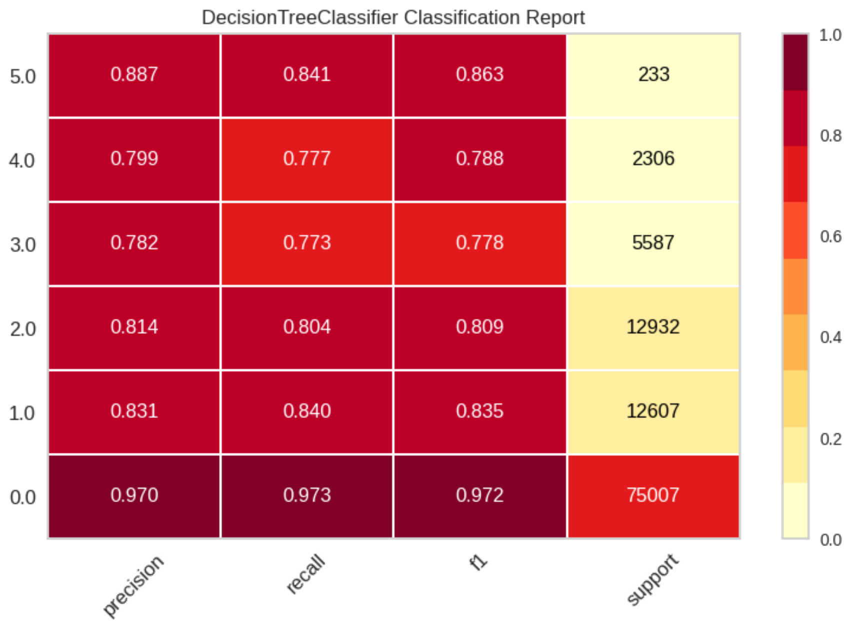

Figure 12. CART model classification report of the first experiment.

Overall, the results show that the Decision Tree (CART) and Random Forest models effectively classify drought intensity, with Random Forest demonstrating slightly better performance.

We move on to the second experiment, which used wind speed indicators. In this experiment, the following factors were considered: wind speed range at 10 m

(WS10M_RANGE), wind speed range at 50 m (WS50M_RANGE), maximum wind speed at 10 m (WS10M_MAX), and maximum wind speed at 50 m (WS50M_MAX). **Figures 13** and **14** present the classification matrices for the Decision Tree (CART) and Random Forest models. **Figures 15** and **16** show the feature importance for the second experiment and the classification report.

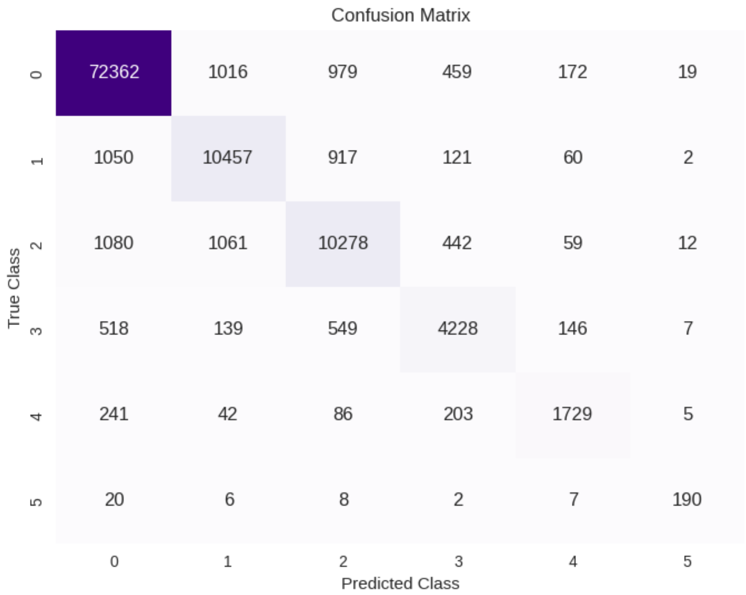

Figure 13. The error matrix of the CART model of the second experiment.

Figure 14. The error matrix of the Random Forest model of the second experiment.

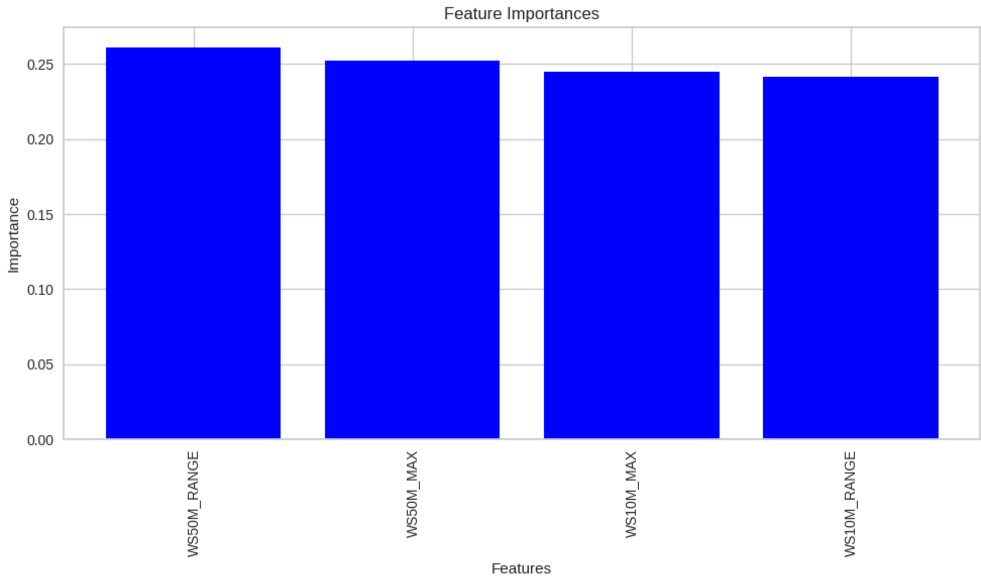

Figure 15. The importance of the features of both models of the second experiment.

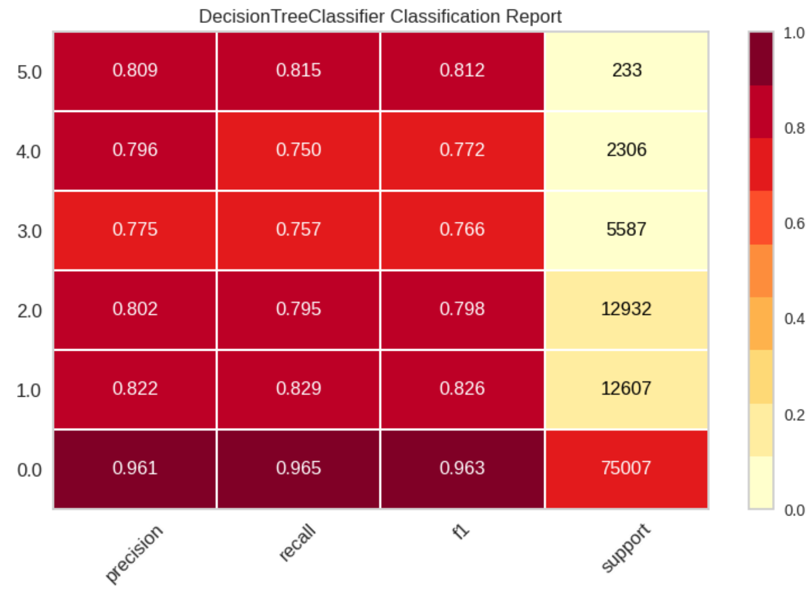

Figure 16. CART model classification report of the second experiment.

The Decision Tree (CART) model showed an accuracy of 91.3%. The classification matrix for this model shows that it distinguishes most drought classes well, but has some classification errors, especially for classes 4 and 5. Feature importance revealed that the wind speed range and maximum wind speed at heights of 10 and 50 m have the greatest influence on the model's predictions.

The Random Forest model demonstrated higher accuracy—93%. The classification matrix for this model shows higher prediction accuracy compared to the Decision Tree. The visualization of feature importance also confirms the significance of the wind speed range and maximum wind speed at various heights.

4. Discussion

Hybrid ARIMA-ANN models have shown superior accuracy in drought forecasting compared to individual models. In the Kansabati River basin, these hybrids were effective in providing accurate predictions. Hybrid models like ANFIS also enhance performance by combining linear and non-linear approaches, addressing the limitations of individual models (Fung, 2019).

Wavelet-based models such as the Wavelet Extreme Learning Machine (W-ELM) have proven effective in capturing the temporal characteristics of droughts. These models provided accurate drought forecasts in Eastern Australia, demonstrating their suitability for regions with complex climatic patterns. The hybrid wavelet-ANN model also showed high efficiency in forecasting long-term droughts (Deo, 2017; Gorgij et al., 2016).

The results of the conducted experiments showed that machine learning models such as Decision Tree (CART) and Random Forest can effectively classify drought intensity based on various meteorological parameters. On the entire dataset, the Decision Tree and Random Forest models achieved accuracies of 92.2% and 93.4%, respectively, confirming their high performance in complex factor analysis. In the first experiment, which analyzed temperature and humidity indicators, the Decision Tree and Random Forest models showed improved results with accuracies of 93.2% and 94.4%, respectively. This indicates the significance of temperature and humidity parameters in predicting drought intensity. In the second experiment, which used wind speed indicators, the models demonstrated accuracies of 91.3% for Decision Tree and 93.0% for Random Forest, also confirming the importance of wind conditions in the context of drought.

However, despite the high accuracy rates, the models faced certain limitations, especially in classifying the rarer classes 4 and 5. The small amount of data for these classes makes it difficult to accurately predict drought intensity, highlighting the need for additional data collection to improve the models. These results emphasize the importance of data diversity and volume in enhancing the accuracy and reliability of forecasts, which is a critical factor for effective water resource management and infrastructure planning in the context of climate change.

Our models were divided based on the aim and hypotheses of the research. The primary goal was to develop models for classifying drought intensity using various groups of meteorological parameters, including temperature and humidity indicators, as well as wind speed metrics. In the introduction of the article, the research hypotheses are described, according to which models built on different groups of parameters should provide high accuracy in drought classification. This division allowed us to study the influence of each group of parameters on classification accuracy in detail and identify the most significant factors.

Details of each stage are presented in **Table 2**.

Table 2. Accuracy of Decision Tree (CART) and Random Forest models for different experiments.

Experiment	Decision Tree (CART)	Random Forest
Entire dataset	92.2%	93.4%
First experiment	93.2%	94.4%
Second experiment	91.3%	93.0%

In summary, our results are consistent with previous studies showing that the integration of several meteorological indicators can significantly improve the accuracy of the model. Moreover, our results highlight the effectiveness of hybrid approaches in improving forecasting efficiency, as has been shown in other studies using hybrid ARIMA-ANN and wavelet models. Thus, our study complements the body of evidence supporting the use of advanced machine learning techniques for drought forecasting, offering new insights into the integration of various meteorological data to more accurately classify drought.

5. Conclusion

In this study, machine learning methods such as Decision Tree (CART) and Random Forest were applied to classify drought intensity levels based on various meteorological data. The main objective of the study was to assess the effectiveness of these methods in classification tasks and their applicability for water resource management and agriculture.

The study showed that the Decision Tree (CART) model on the entire dataset demonstrated an accuracy of 92.2%, while Random Forest achieved 93.4%. In the first experiment, which used temperature and humidity indicators, the accuracies were 93.2% for Decision Tree and 94.4% for Random Forest. In the second experiment, which considered wind speed indicators, the accuracies were 91.3% and 93.0% respectively. These results confirm the high effectiveness of machine learning models in drought classification tasks.

The feature importance visualized during the analysis showed that key meteorological parameters such as atmospheric pressure, temperature at 2 m, and wind speed play a significant role in classifying drought intensity. This underscores the necessity of including these parameters in future models to improve their accuracy. However, the study also revealed a limitation related to the insufficient amount of data for high drought levels (classes 4 and 5), making it difficult to accurately predict these classes. This indicates the need for additional data collection to improve the models and enhance their reliability.

For future research, it is recommended to expand the dataset to increase the accuracy of the models and their ability to predict rare events such as extreme droughts. Including additional meteorological and climatic parameters, such as soil moisture data, solar radiation, and evaporation, can also contribute to improving model accuracy. Developing combined models that integrate machine learning methods with physical drought models can provide more accurate and reliable forecasts.

Supporting research and development in machine learning and climatology is important for creating more accurate drought classification and forecasting models. Developing and implementing drought adaptation strategies based on forecast data may include improving water resource management, developing sustainable agricultural practices, and increasing preparedness for dry periods.

Thus, the use of machine learning methods such as Decision Tree and Random Forest has demonstrated high effectiveness in classifying drought intensity. To achieve more accurate and reliable forecasts, it is necessary to continue research and expand the dataset, which will improve water resource and agricultural management under climate change conditions.

To conclude our study, we present an illustration demonstrating how the Random Forest machine learning model helps in classifying drought intensity. **Figure 17** shows the data analysis process, where the model processes the input meteorological data (temperature, humidity, wind speed and atmospheric pressure) and outputs a map with a classification of drought levels from D0 (no drought) to D4 (extreme drought).

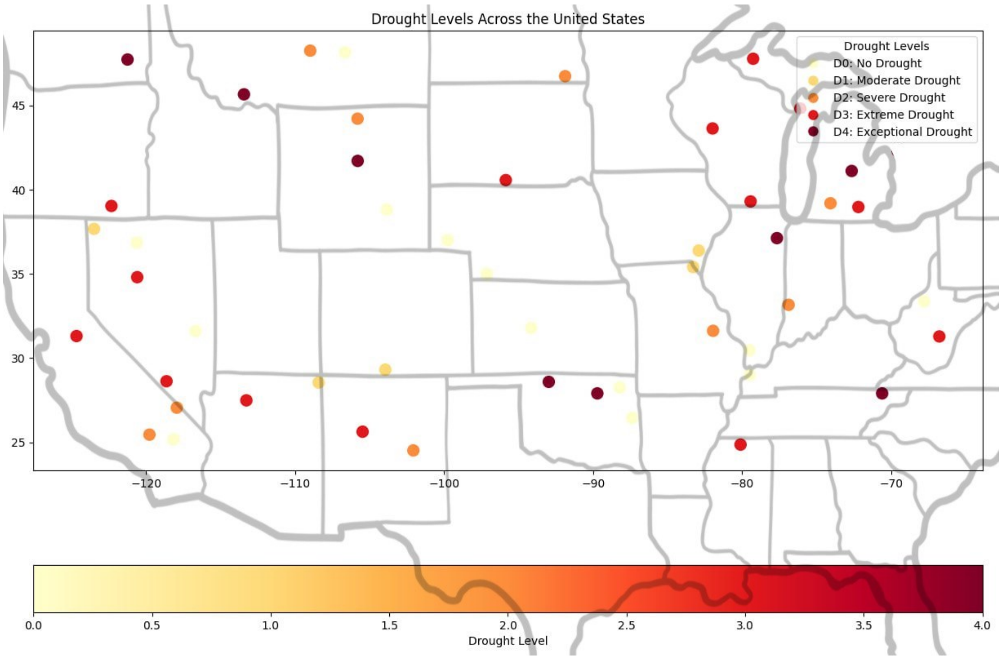

Figure 17. An illustration that shows how the machine learning model helps in drought classification.

This visualization clearly shows how the use of modern machine learning algorithms can significantly improve the accuracy of drought forecasts, which is an important step in water resource management and agricultural planning in the face of climate change.

Author contributions: Conceptualization, VK and VT; methodology, OK; software, VT; validation, VK and VT; formal analysis, OK; investigation, TP; resources, VT; data curation, VT; writing—original draft preparation, VK; writing—review and editing, VT; visualization, TP; supervision, VK; project administration, VK; funding acquisition, OK. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

Aghelpour, P., Mohammadi, B., Biazar, S. M., et al. (2020). A Theoretical Approach for Forecasting Different Types of Drought Simultaneously, Using Entropy Theory and Machine-Learning Methods. ISPRS International Journal of Geo-Information, 9(12), 701. https://doi.org/10.3390/ijgi9120701

Achite, M., Jehanzaib, M., Elshaboury, N., et al. (2022). Evaluation of Machine Learning Techniques for Hydrological Drought Modeling: A Case Study of the Wadi Ouahrane Basin in Algeria. Water, 14(3), 431. https://doi.org/10.3390/w14030431

Aghelpour, P., Mohammadi, B., Biazar, S. M., et al. (2020). A Theoretical Approach for Forecasting Different Types of Drought Simultaneously, Using Entropy Theory and Machine-Learning Methods. ISPRS International Journal of Geo-Information, 9(12), 701. https://doi.org/10.3390/ijgi9120701

Alkan, A. (2023). Drought Forecasting using Palmer Drought Severity Index with Wavelet Transform Technique and Machine Learning Methods. International Journal of Research Publication and Reviews, 04(01), 2177–2185. https://doi.org/10.55248/gengpi.2023.4158

Almikaeel, W., Čubanová, L., & Šoltész, A. (2022). Hydrological Drought Forecasting Using Machine Learning—Gidra River Case Study. Water, 14(3), 387. https://doi.org/10.3390/w14030387

Bashmur, K. A., Kolenchukov, O. A., Bukhtoyarov, V. V., et al. (2022). Biofuel Technologies and Petroleum Industry: Synergy of Sustainable Development for the Eastern Siberian Arctic. Sustainability, 14(20), 13083. https://doi.org/10.3390/su142013083

Borodulin, A., Gladkov, A., Gantimurov, A., et al. (2024). Using machine learning algorithms to solve data classification problems using multi-attribute dataset. BIO Web of Conferences, 84, 02001. https://doi.org/10.1051/bioconf/20248402001

Bosikov, I. I., Martyushev, N. V., Klyuev, R. V., et al. (2023). Modeling and Complex Analysis of the Topology Parameters of Ventilation Networks When Ensuring Fire Safety While Developing Coal and Gas Deposits. Fire, 6(3), 95. https://doi.org/10.3390/fire6030095

C.M, A. M., Chowdary, V. M., Kesarwani, M., et al. (2022). Integrated drought monitoring and assessment using multi-sensor and multi-temporal earth observation datasets: a case study of two agriculture-dominated states of India. Environmental Monitoring and Assessment, 195(1). https://doi.org/10.1007/s10661-022-10550-6

Chu, H. J. (2018). Drought Detection of Regional Nonparametric Standardized Groundwater Index. Water Resources Management, 32(9), 3119–3134. https://doi.org/10.1007/s11269-018-1979-4

Degtyareva, K., Ageev, D. A., & Kukartsev, V. V. (2023). Finding patterns in employee attrition rates using self-organizing Kohonen maps and decision trees. 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). https://doi.org/10.1109/icses60034.2023.10465548

Dehghani, M., Saghafian, B., Nasiri Saleh, F., et al. (2013). Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. International Journal of Climatology, 34(4), 1169–1180. https://doi.org/10.1002/joc.3754

Deo, R. C., & Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153, 512–525. https://doi.org/10.1016/j.atmosres.2014.10.016

Deo, R. C., Tiwari, M. K., Adamowski, J. F., et al. (2016). Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stochastic Environmental Research and Risk Assessment, 31(5), 1211–1240. https://doi.org/10.1007/s00477-016-1265-z

Dikshit, A., Pradhan, B., & Huete, A. (2021). An improved SPEI drought forecasting approach using the long short-term memory neural network. Journal of Environmental Management, 283, 111979. https://doi.org/10.1016/j.jenvman.2021.111979

Durand, M., Molotch, N. P., & Margulis, S. A. (2008). Merging complementary remote sensing datasets in the context of snow water equivalent reconstruction. Remote Sensing of Environment, 112(3), 1212–1225. https://doi.org/10.1016/j.rse.2007.08.010

Elbeltagi, A., Pande, C. B., Kumar, M., et al. (2023). Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. Environmental Science and Pollution Research, 30(15), 43183–43202. https://doi.org/10.1007/s11356-023-25221-3

Fung, K. F., Huang, Y. F., & Koo, C. H. (2019). Coupling fuzzy–SVR and boosting–SVR models with wavelet decomposition for meteorological drought prediction. Environmental Earth Sciences, 78(24). https://doi.org/10.1007/s12665-019-8700-7

Gohel, H. A., Upadhyay, H., Lagos, L., et al. (2020). Predictive maintenance architecture development for nuclear infrastructure using machine learning. Nuclear Engineering and Technology, 52(7), 1436–1442. https://doi.org/10.1016/j.net.2019.12.029

Khan, N., Sachindra, D. A., Shahid, S., et al. (2020). Prediction of droughts over Pakistan using machine learning algorithms. Advances in Water Resources, 139, 103562. https://doi.org/10.1016/j.advwatres.2020.103562

Kolachian, R., & Saghafian, B. (2021). Hydrological drought class early warning using support vector machines and rough sets. Environmental Earth Sciences, 80(11). https://doi.org/10.1007/s12665-021-09536-3

Kolenchukov, O. A., Bashmur, K. A., Bukhtoyarov, V. V., et al. (2022). Experimental Study of Oil Non-Condensable Gas Pyrolysis in a Stirred-Tank Reactor for Catalysis of Hydrogen and Hydrogen-Containing Mixtures Production. Energies, 15(22), 8346. https://doi.org/10.3390/en15228346

Kozlova, A., Kukartsev, V., Melnikov, V., et al. (2023). Finding dependencies in the corporate environment using data mining. E3S Web of Conferences, 431, 05032. https://doi.org/10.1051/e3sconf/202343105032

Liu, J., Jiang, W., Han, H., et al. (2023). Drought Level Prediction Based on Meteorological Data and Deep Learning. In: Proceedings of the 2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). https://doi.org/10.1109/secon58729.2023.10287478

Malozyomov, B. V., Martyushev, N. V., Kukartsev, V. A., et al. (2023). Study of Supercapacitors Built in the Start-Up System of the Main Diesel Locomotive. Energies, 16(9), 3909. https://doi.org/10.3390/en16093909

Martyushev, N. V., Bublik, D. A., Kukartsev, V. V., et al. (2023). Provision of Rational Parameters for the Turning Mode of Small-Sized Parts Made of the 29 NK Alloy and Beryllium Bronze for Subsequent Thermal Pulse Deburring. Materials, 16(9), 3490. https://doi.org/10.3390/ma16093490

Mokhtar, A., Jalali, M., He, H., et al. (2021). Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms. IEEE Access, 9, 65503–65523. https://doi.org/10.1109/access.2021.3074305

Mosavi, A., & Ardabili, S. (2023). Machine Learning for Drought Prediction; Review, Bibliometric Analysis, and Models Evaluation. In: Proceedings of the 2023 IEEE 27th International Conference on Intelligent Engineering Systems (INES). https://doi.org/10.1109/ines59282.2023.10297771

Rahmati, O., Falah, F., Dayal, K. S., et al. (2020). Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. Science of The Total Environment, 699, 134230. https://doi.org/10.1016/j.scitotenv.2019.134230

Raza, M. A., Almazah, M. M. A., Ali, Z., et al. (2022). Application of Extreme Learning Machine Algorithm for Drought Forecasting. Complexity, 2022, 1–28. https://doi.org/10.1155/2022/4998200

Rhee, J., & Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. Agricultural and Forest Meteorology, 237–238, 105–122. https://doi.org/10.1016/j.agrformet.2017.02.011

Richman, M. B., Leslie, L. M., & Segele, Z. T. (2016). Classifying Drought in Ethiopia Using Machine Learning. Procedia Computer Science, 95, 229–236. https://doi.org/10.1016/j.procs.2016.09.319

Shah, H., Rane, V., Nainani, J., et al. (2017). Drought Prediction and Management using Big Data Analytics. International Journal of Computer Applications, 162(4), 27–30. https://doi.org/10.5120/ijca2017913276

Shamshirband, S., Hashemi, S., Salimi, H., et al. (2020). Predicting Standardized Streamflow index for hydrological drought using machine learning models. Engineering Applications of Computational Fluid Mechanics, 14(1), 339–350. https://doi.org/10.1080/19942060.2020.1715844

Sharma, P., Kar, B., Wang, J., et al. (2021). A machine learning approach to flood severity classification and alerting. In: Proceedings of the 4th ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities. https://doi.org/10.1145/3486626.3493432

Tufaner, F., & Özbeyaz, A. (2020). Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms. Environmental Monitoring and Assessment, 192(9). https://doi.org/10.1007/s10661-020-08539-0

Tynchenko, Y., Kukartsev, V., Gladkov, A., et al. (2024). Assessment of technical water quality in mining based on machine learning methods. Sustainable Development of Mountain Territories, 16(1), 56–69. https://doi.org/10.21177/1998-4502-2024-16-1-56-69

Vrindavanam, J., Babu, T., Gandiboina, H., et al. (2022). A Comparative Analysis of Machine Learning Algorithms for Agricultural Drought Forecasting. In: Proceedings of the 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). https://doi.org/10.1109/icict55121.2022.10064511

Yelemessov, K., Isametova, M., Saydinbayeva, N., et al. (2023). Mathematical and computer modeling of gantry crane load-beam system oscillation. Sustainable Development of Mountain Territories, 15(2), 450–461. https://doi.org/10.21177/1998-4502-2023-15-2-450-461

Zhao, Y., Zhang, J., Bai, Y., et al. (2022). Drought Monitoring and Performance Evaluation Based on Machine Learning Fusion of Multi-Source Remote Sensing Drought Factors. Remote Sensing, 14(24), 6398. https://doi.org/10.3390/rs14246398