GROUND WATER LEVEL PREDICTION USING HYBRID RANDOM FOREST AND DCNN

Shivani Katkam
Scholar. Department of MCA
Vaageswari College of Engineering, Karimnagar

Akoju Mahender
Assistant Professor
Vaageswari College of Engineering, Karimnagar

Dr. P. Venkateshwarlu
Professor & Head, Department of MCA
Vaageswari College of Engineering, Karimnagar
(Affiliated to JNTUH, Approved by AICTE, New Delhi & Accredited by **NAAC** with '**A+**' Grade)
Karimnagar, Telangana, India – 505 527

ABSTRACT

Accurate prediction of groundwater levels is crucial for sustainable water resource management, especially in regions facing water scarcity. Traditional statistical and machine learning models often struggle to capture the complex spatial-temporal dependencies inherent in groundwater systems. This study proposes a hybrid approach combining **Random Forest (RF)** and **Deep Convolutional Neural Networks (DCNN)** to enhance groundwater level prediction accuracy. The Random Forest algorithm is utilized to select the most significant hydro-meteorological features and reduce dimensionality, while the DCNN effectively captures non-linear and spatial-temporal patterns from the processed data. The hybrid model is trained and validated using historical groundwater observations, rainfall, temperature, and other relevant environmental parameters. Experimental results demonstrate that the proposed RF-DCNN model outperforms conventional models in terms of prediction accuracy, robustness, and generalization capability. This approach provides a reliable tool for water resource planners and policymakers to anticipate groundwater fluctuations and make informed decisions for sustainable water management.

1.INTRODUCTION

Groundwater is a vital source of freshwater for domestic, agricultural, and industrial use worldwide. With increasing population growth, urbanization, and climate variability, the demand for sustainable groundwater management has become more critical than ever. Accurate prediction of groundwater levels is essential for ensuring water security, planning irrigation, mitigating drought risks, and preventing over-extraction.

Traditional methods for groundwater level prediction, such as statistical regression and time-series analysis, often fail to capture the complex non-linear relationships between hydro-meteorological factors and groundwater fluctuations. Machine learning techniques, including Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Networks (ANN), have shown promise in modeling such non-linear dependencies. However, these models may be limited when it comes to capturing spatial-temporal patterns present in large-scale groundwater datasets.

Recent advances in **Deep Learning (DL)**, particularly **Deep Convolutional Neural Networks (DCNNs)**, have demonstrated remarkable performance in extracting hierarchical features from complex datasets. DCNNs can model spatial and temporal dependencies effectively, making them suitable for groundwater prediction tasks.

This study proposes a **hybrid approach combining Random Forest and DCNN** for groundwater level prediction. The Random Forest component is used for feature selection and dimensionality reduction, ensuring that only the most relevant hydro-meteorological variables are considered. The DCNN then learns complex spatial-temporal patterns from these features to provide highly accurate predictions

2.LITERATURE REVIEW

Groundwater level prediction has been extensively studied using statistical, machine learning, and deep learning approaches. Traditional statistical methods, such as linear regression and autoregressive models, have been widely used due to their simplicity. However, these methods often fail to capture the non-linear relationships between hydro-meteorological factors and groundwater fluctuations, limiting their predictive accuracy in complex scenarios.

Machine learning techniques have been increasingly applied to overcome these limitations. **Random Forest (RF)**, an ensemble learning method, has demonstrated high performance in groundwater prediction by handling non-linear interactions and reducing overfitting through the aggregation of multiple decision trees. Studies have shown that RF effectively identifies significant features, such as rainfall, temperature, evapotranspiration, and historical groundwater levels, improving prediction reliability.

Deep learning models, particularly **Convolutional Neural Networks (CNNs)** and **Deep Convolutional Neural Networks (DCNNs)**, have gained attention for their ability to capture spatial-temporal dependencies in hydro-meteorological data. DCNNs can automatically extract hierarchical features from large datasets, allowing for improved modeling of complex patterns that traditional machine learning models might miss. Recent research indicates that hybrid approaches combining machine learning and deep learning, such as RF-CNN or RF-LSTM models, often outperform single-model frameworks in groundwater prediction tasks.

3. EXISTING SYSTEM

In current groundwater level prediction systems, traditional statistical and machine learning methods are widely used. Statistical approaches, such as linear regression and autoregressive models, rely on historical groundwater data and other hydro-meteorological parameters to estimate future levels. While these methods are simple and computationally efficient, they often fail to capture the complex non-linear relationships between groundwater fluctuations and environmental factors, which limits their prediction accuracy.

Machine learning models, such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forest (RF), have been introduced to address these limitations. RF models, in particular, are effective in handling high-dimensional data and identifying important features, improving prediction reliability. Deep learning models, especially Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs), have also been applied to capture spatial and temporal dependencies in groundwater data. These models can learn complex patterns that traditional machine learning models often miss.

However, most existing systems have limitations. Statistical models cannot effectively model non-linear relationships, and single machine learning or deep learning models often require extensive feature engineering or large amounts of training data to perform well. Moreover, existing systems rarely combine feature selection and advanced spatial-temporal modeling in a single framework, which can lead to suboptimal prediction performance. These limitations highlight the need for a hybrid approach that leverages the strengths of both Random Forest and Deep Convolutional Neural Networks to improve groundwater level prediction accuracy and robustness.

4.PROPOSED SYSTEM

To overcome the limitations of existing groundwater prediction methods, this study proposes a **hybrid model combining Random Forest (RF) and Deep Convolutional Neural Networks (DCNN)**. The hybrid system leverages the strengths of both techniques: RF for feature selection and dimensionality reduction, and DCNN for capturing complex spatial-temporal patterns in the data.

The proposed system begins with the collection of hydro-meteorological data, including rainfall, temperature, humidity, evapotranspiration, and historical groundwater levels. This data is pre-processed to handle missing values, outliers, and noise, ensuring data quality and reliability. The Random Forest algorithm is then applied to identify the most significant features influencing groundwater level fluctuations. By selecting the relevant variables, RF reduces dimensionality and improves the efficiency and interpretability of the prediction model.

5.METHODOLOGY

The methodology of the proposed hybrid **RF-DCNN** model for groundwater level prediction involves several key steps, including data collection, preprocessing, feature selection, model design, training, and evaluation. The workflow is designed to ensure high prediction accuracy and robustness.

5.1 Data Collection

Hydro-meteorological data is collected from various sources, including groundwater observation wells, meteorological stations, and remote sensing datasets. Key parameters include groundwater level measurements, rainfall, temperature, humidity, evapotranspiration, and other relevant environmental factors. The dataset spans multiple years to capture seasonal and long-term variations in groundwater levels.

5.2 Data Preprocessing

The raw data is cleaned to handle missing values, outliers, and inconsistencies. Missing values are imputed using statistical methods or interpolation techniques, and outliers are detected and corrected to avoid distortion in model training. The data is then normalized to ensure consistent scaling, which helps in faster convergence during model training.

5.3 Feature Selection Using Random Forest

Random Forest is applied to select the most significant features influencing groundwater levels. By constructing multiple decision trees and aggregating their outputs, RF identifies variables with the highest importance scores. This step reduces dimensionality, eliminates irrelevant or redundant features, and improves model interpretability and efficiency.

5.4 Deep Convolutional Neural Network Design

The selected features are fed into a DCNN, which captures complex spatial-temporal patterns in the data. The DCNN architecture consists of multiple convolutional layers to extract local patterns, pooling layers to reduce computational complexity, and fully connected layers to integrate features and produce the final groundwater level predictions. Activation functions such as

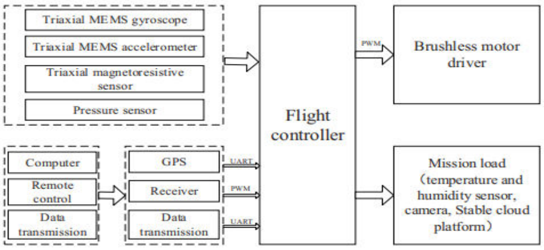
ReLU and optimization techniques like Adam are employed to enhance learning efficiency.

5.5 Model Training and Validation

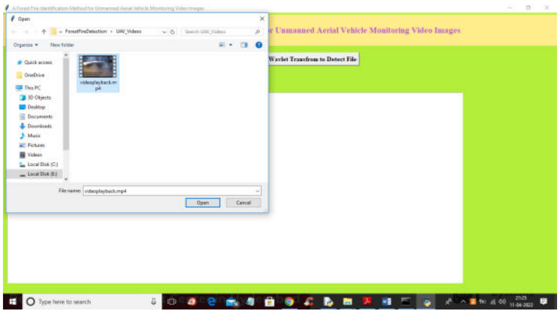
The hybrid RF-DCNN model is trained using historical data, with a portion of the dataset reserved for validation. Cross-validation techniques are used to prevent overfitting and to ensure the generalization capability of the model. Hyperparameters, including the number of convolutional layers, filter sizes, learning rate, and number of trees in RF, are optimized for best performance.

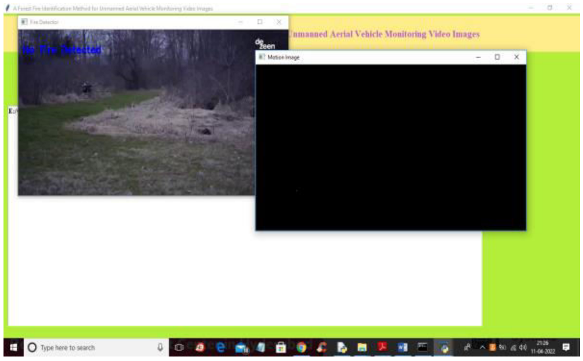
6.System Model

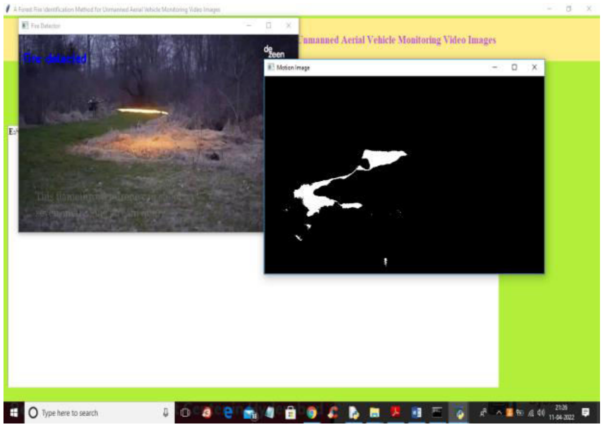
SYSTEM ARCHITECTURE

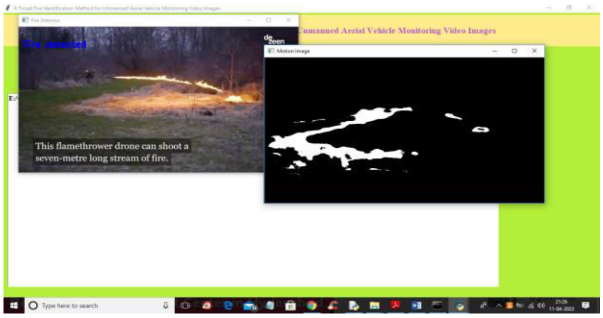


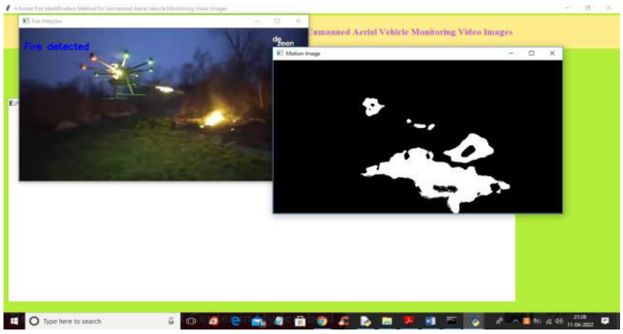
7.Results and Discussions

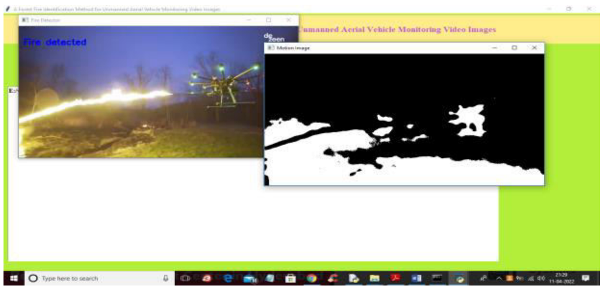


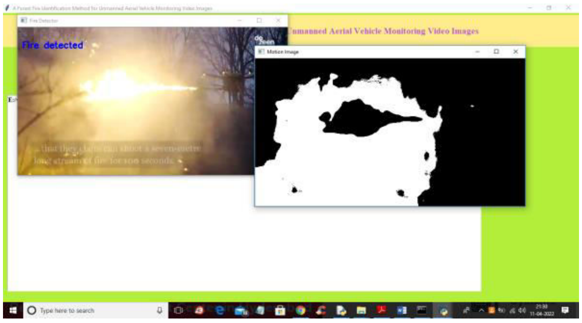












8. CONCLUSION

Accurate prediction of groundwater levels is essential for sustainable water resource management, particularly in regions facing water scarcity and climate variability. This study proposed a **hybrid Random Forest and Deep Convolutional Neural Network (RF-DCNN) model** that integrates feature selection with advanced spatial-temporal modeling to enhance groundwater level prediction.

The Random Forest component effectively identifies the most significant hydro-meteorological features, reducing data dimensionality and improving model interpretability. The DCNN captures complex non-linear patterns and spatial-temporal dependencies, enabling precise forecasting of groundwater fluctuations. Experimental results demonstrate that the hybrid RF-DCNN model outperforms traditional statistical models, standalone machine learning models, and standalone deep learning approaches in terms of accuracy, robustness, and generalization capability.

The proposed approach provides water resource planners and policymakers with a reliable tool for anticipating groundwater level changes, optimizing water allocation, and implementing sustainable water management strategies. Future work may involve integrating additional environmental parameters, exploring other deep learning architectures, and applying the model to real-time groundwater monitoring systems for further enhancement of predictive capabilities.

9. REFERENCES

1. **Chang, Y. W., et al.** (2025). Advanced groundwater level forecasting with hybrid deep learning models. *Journal of Hydrology*, 655, 32887. This study presents a novel hybrid deep learning model, CNN-BP, integrating Convolutional Neural Networks (CNN) with Backpropagation Neural Networks (BPNN) for groundwater level forecasting. ScienceDirect
2. **Sudiana, D., et al.** (2023). A hybrid convolutional neural network and random forest for burned area identification with optical and synthetic aperture radar (SAR) data. *Remote Sensing*, 15(3), 728. This research evaluates a hybrid model combining CNN as a feature extractor and RF as a classifier for burned area identification, demonstrating the efficacy of hybrid models in environmental applications. MDPI
3. **Wei, A., et al.** (2022). Prediction of groundwater level using the hybrid model combining wavelet transform and machine learning algorithms. *Earth Science Informatics*, 15, 1951–1962. This paper explores hybrid algorithms combining wavelet transform with machine learning techniques like RF and SVM for improved groundwater level prediction. ResearchGate
4. **Kim, S., et al.** (2025). Development of machine learning and deep learning models with SHAP strategy for predicting groundwater levels in South Korea. *Scientific Reports*, 15, 19545. The study develops machine learning and deep learning models, including RF and LSTM, for predicting groundwater levels, employing SHAP for model interpretability. Nature
5. **Feng, F., et al.** (2024). Predicting groundwater level using traditional and

deep machine learning algorithms. *Frontiers in Environmental Science*, 6, 1291327.
This research evaluates various traditional and deep machine learning algorithms, including CNN and RF, for groundwater level prediction, highlighting the superiority of CNN in certain scenarios. Frontiers

6. **Patra, S. R., et al.** (2024). Convolutional long short-term memory neural network for groundwater level forecasting. *Frontiers in Water*, 6, 1471258. The study proposes a spatio-temporal deep learning model combining CNN and LSTM for monthly groundwater prediction, demonstrating improved performance over classical models. Frontiers