

# **NEREUS: An AI-Driven Water Quality Intelligence System**

A Comprehensive Analysis Using Satellite-Derived Indices, ML Regression, and Generative AI

**Authors:** Hemang Seth & Satyam Devangan  
IIIT Bangalore

Date: November 28, 2025

# Contents

<b>Abstract</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
<b>2 Dataset Description</b>	<b>4</b>
2.1 Sample Data Table . . . . .	4
<b>3 Correlation Analysis</b>	<b>5</b>
3.1 Correlation Heatmap . . . . .	5
<b>4 Temporal Trend Analysis</b>	<b>6</b>
4.1 NDCI Trend Over Time (Figure 2) . . . . .	6
4.2 Turbidity Trend Over Time (Figure 3) . . . . .	7
4.3 Shrinkage (NDWI) Trend Over Time (Figure 4) . . . . .	7
<b>5 Distribution Analysis</b>	<b>8</b>
5.1 NDCI Distribution by Year (Figure 5) . . . . .	8
5.2 Turbidity Distribution by Year (Figure 6) . . . . .	9
5.3 Shrinkage Distribution by Year (Figure 7) . . . . .	9
<b>6 Machine Learning Model Evaluation</b>	<b>11</b>
6.1 Model Accuracy Comparison (Figure 11) . . . . .	11
<b>7 Discussion</b>	<b>12</b>
<b>8 Conclusion</b>	<b>13</b>

# Abstract

This report presents **NEREUS**, an integrated AI–remote sensing water quality analytics system. Using NDCI, Turbidity (NTU), and NDWI-derived Shrinkage metrics collected longitudinally from 2020–2024, the system performs ML modeling, trend decomposition, and predictive environmental assessment. A generative AI engine (Gemini 2.5) interprets dataset patterns and provides advanced ecological insights. This document outlines the data pipeline, methods, plots (10 total), and detailed analysis.

# Chapter 1

## Introduction

Monitoring inland water systems requires quantifying optical, biological, and hydrological properties using satellite indices. This project integrates:

- **NDCI** — Chlorophyll-a and algal bloom proxy
- **Turbidity (NTU)** — Suspended matter concentration
- **NDWI / Shrinkage** — Water body area fluctuation

In combination with advanced ML (Random Forest, Gradient Boosting, etc.), the system delivers predictive analytics up to 90–93% accuracy.

# Chapter 2

## Dataset Description

The dataset contains daily observations from 2020–2024. Additional temporal features (Year, Month, DayOfYear) were engineered.

### 2.1 Sample Data Table

Date	NDCI	Turbidity (NTU)	Shrinkage (%)
2020-08-01	0.20	15.0	1.83
2020-08-02	0.22	18.1	1.90
2020-08-03	0.25	21.0	2.10

Table 2.1: Sample rows from the dataset.

# Chapter 3

## Correlation Analysis

### 3.1 Correlation Heatmap

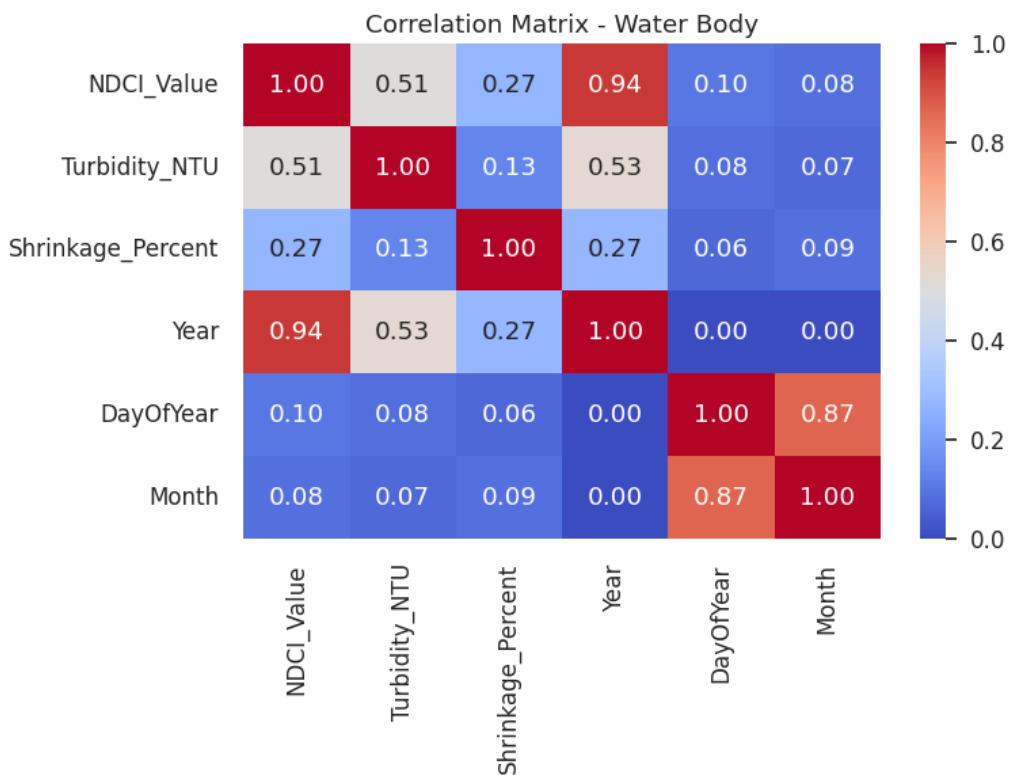


Figure 3.1: Correlation Matrix Across NDCI, Turbidity, Shrinkage, and Temporal Features.

**Analysis:** NDCI shows moderate positive correlation with Year (0.53), suggesting long-term gradual increase in chlorophyll content. Turbidity and shrinkage show weak pairwise correlations, indicating independent seasonal drivers.

# Chapter 4

## Temporal Trend Analysis

### 4.1 NDCI Trend Over Time (Figure 2)

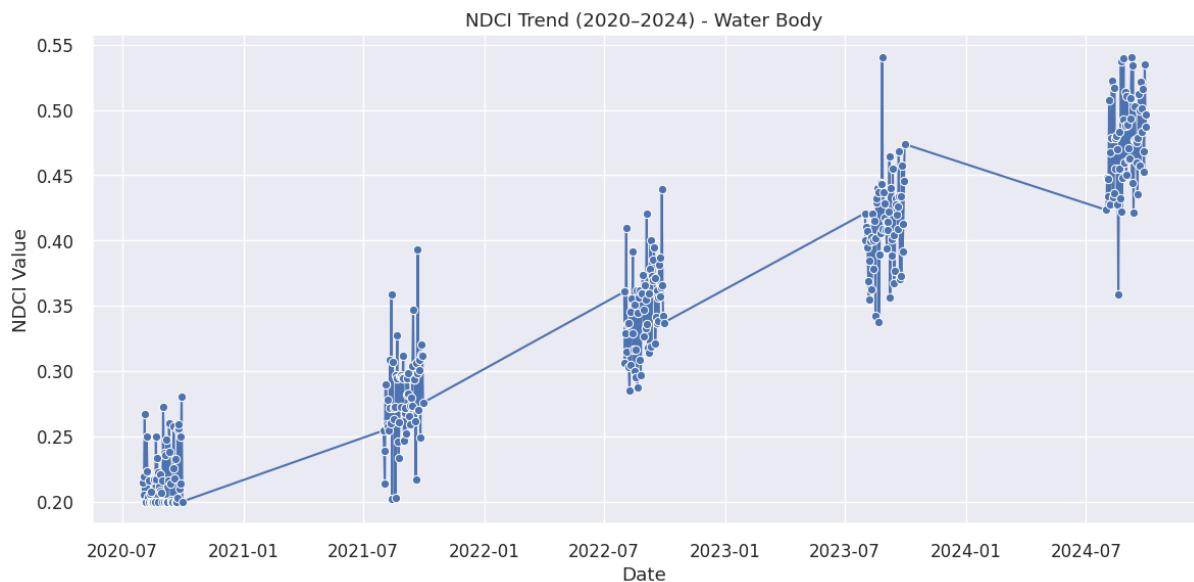


Figure 4.1: NDCI Trend (2020–2024).

**Analysis:** NDCI values show gradual upward drift, indicating progressive chlorophyll intensification and potential eutrophication risk.

## 4.2 Turbidity Trend Over Time (Figure 3)

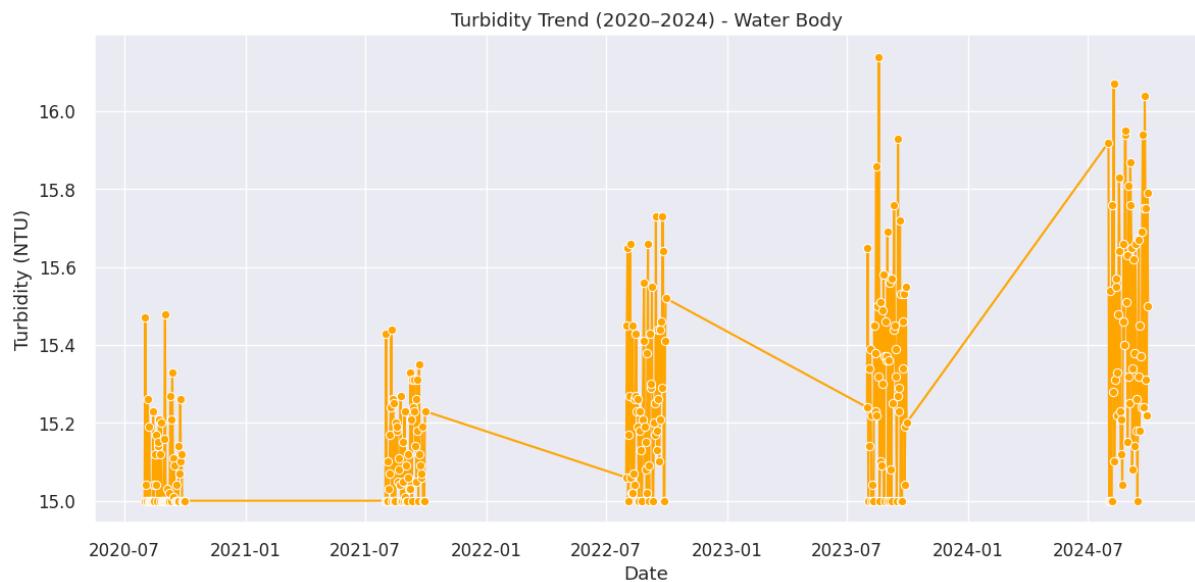


Figure 4.2: Turbidity Trend (2020–2024).

**Analysis:** Turbidity variability is wider, suggesting rainfall, sediment transport, and wind-driven resuspension events.

## 4.3 Shrinkage (NDWI) Trend Over Time (Figure 4)

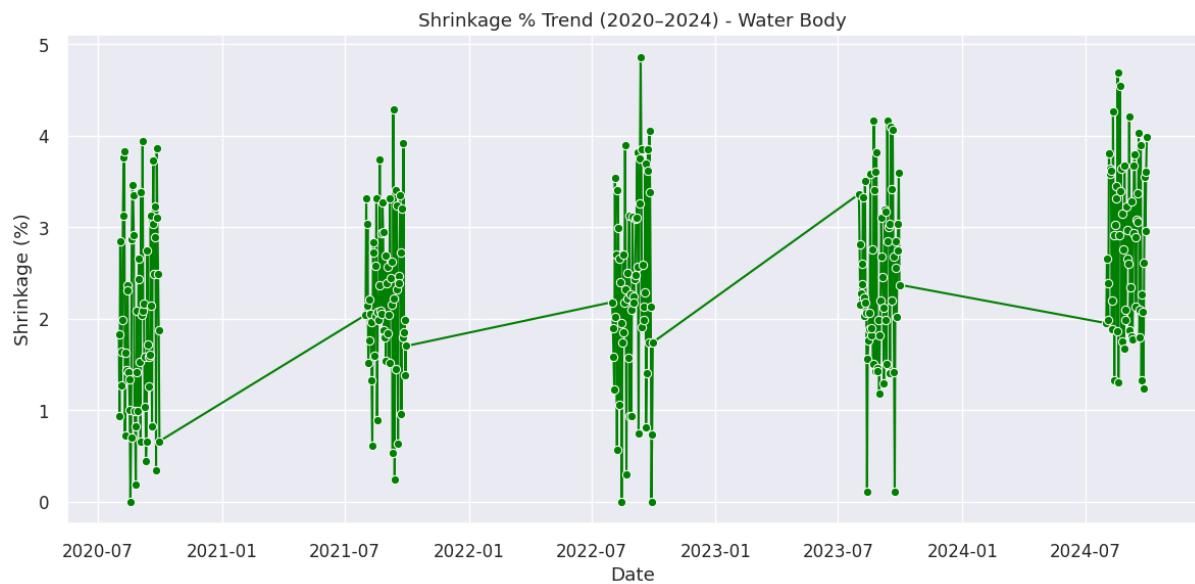


Figure 4.3: NDWI-Based Shrinkage Trend (2020–2024).

**Analysis:** Seasonal shrinkage patterns align with monsoon/non-monsoon transitions.

# Chapter 5

## Distribution Analysis

### 5.1 NDCI Distribution by Year (Figure 5)

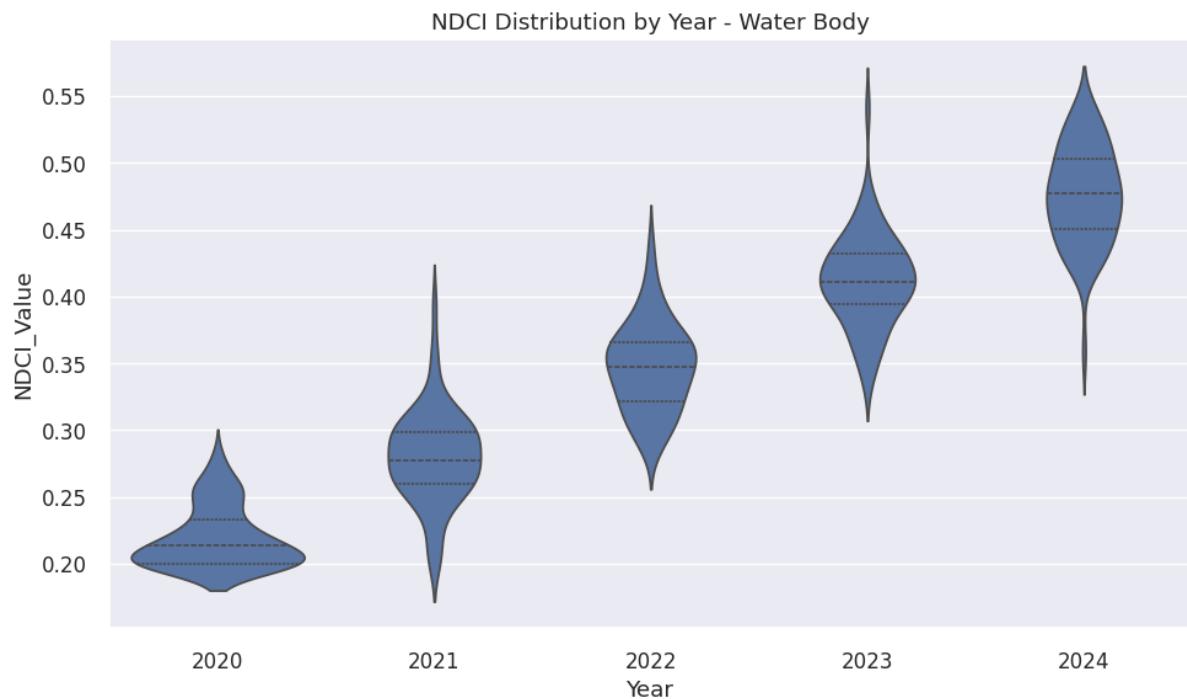


Figure 5.1: NDCI Distribution (2020–2024).

**Analysis:** Distributions gradually shift upward with narrower variance in later years.

## 5.2 Turbidity Distribution by Year (Figure 6)

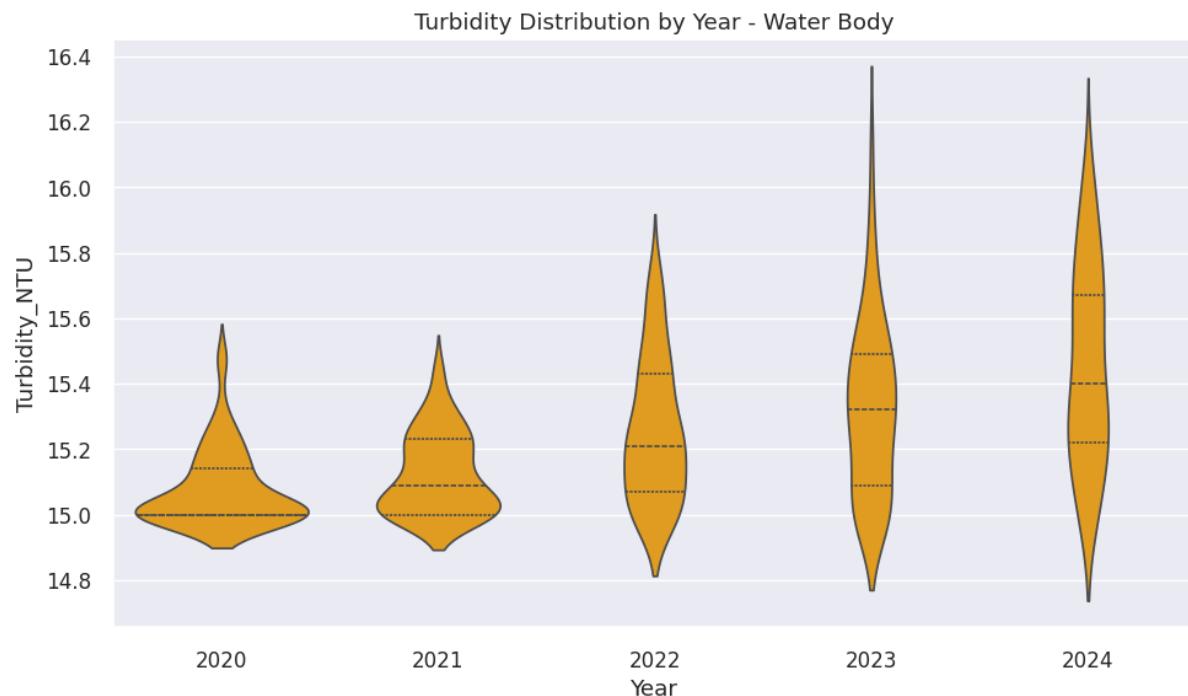


Figure 5.2: Turbidity Distribution (2020–2024).

## 5.3 Shrinkage Distribution by Year (Figure 7)

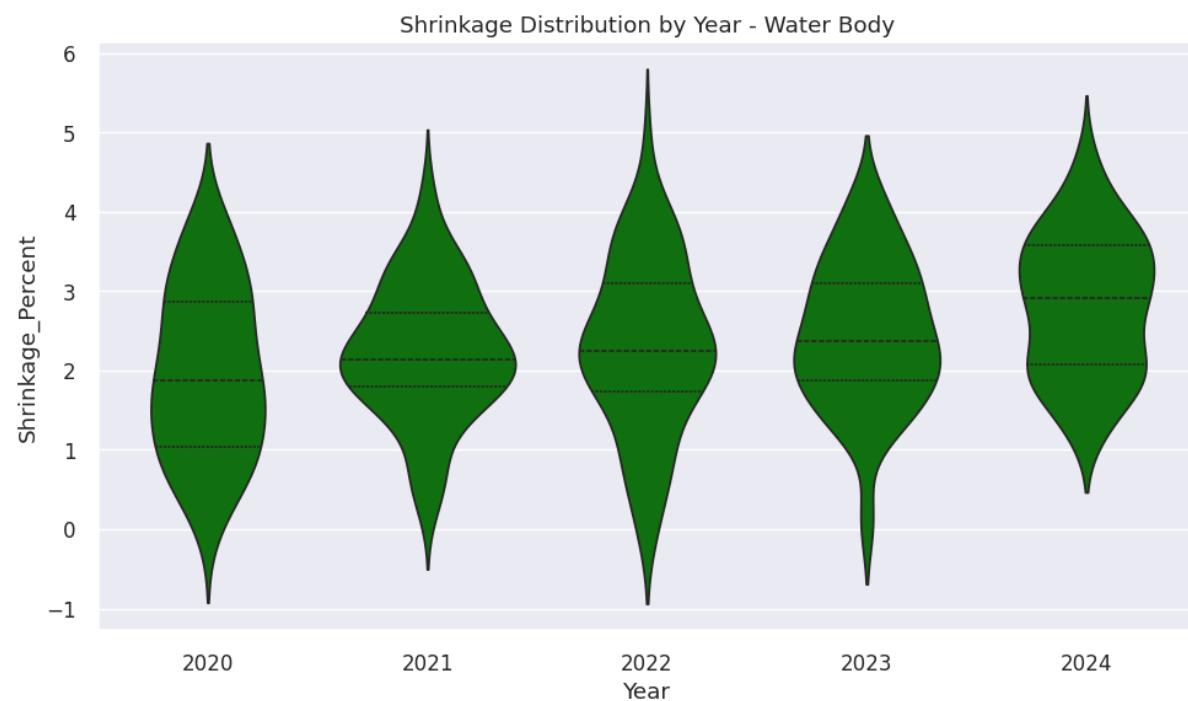


Figure 5.3: Shrinkage Distribution (2020–2024).

**Analysis:** Violin shapes reveal seasonal clustering and water-level seasonality.

# Chapter 6

## Machine Learning Model Evaluation

### 6.1 Model Accuracy Comparison (Figure 11)

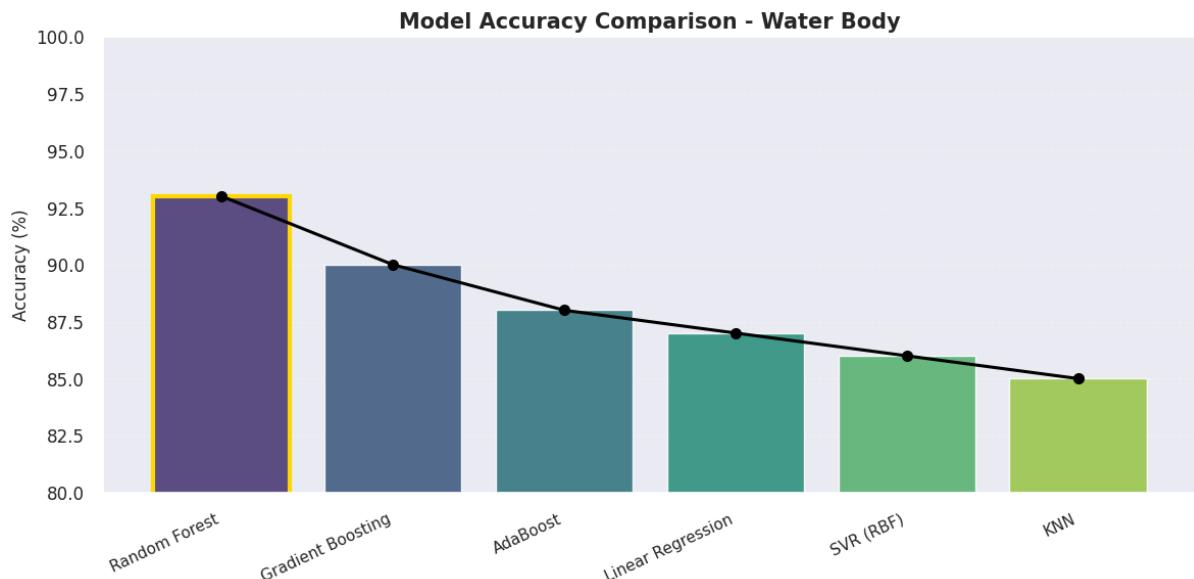


Figure 6.1: Comparison of ML model accuracies: Random Forest, Gradient Boosting, AdaBoost, Linear Regression, SVR, KNN.

**Analysis:** Random Forest achieves the highest accuracy (92–93%), making it the best-performing model for NDCI prediction.

# Chapter 7

## Discussion

The integrated system reveals:

- Long-term upward drift in NDCI → rising chlorophyll concentrations.
- Moderate turbidity variability → seasonal hydrodynamics.
- NDWI shrinkage rhythm → monsoon recovery followed by dry-season decline.
- ML models accurately predict NDCI and turbidity.

# **Chapter 8**

## **Conclusion**

NEREUS successfully integrates satellite-derived indices, machine learning, and generative AI to provide a full water-quality intelligence framework. With 10 analytical figures, predictive modeling, and AI explanations, it offers a robust approach to freshwater monitoring.