Integrating AI-ML And GIS For Flood Assessment And Resilience Planning

Krithika A [1], Deeptha G [2], Prof. Mrs. Anitha Sarafin X [3], Prof. Mrs. Jeyanthi A [4]
[1, 2] Dept of Artificial Intelligence and Data Science
[3]Dept of Computer Science and Engineering
[4]Dept of Artificial Intelligence and Data Science
[1, 2, 3, 4] Misrimal Navajee Munoth Jain Engineering College, Chennai, Tamilnadu – 600097

Abstract- *This paper introduces a pioneering initiative integrating Artificial Intelligence and Machine Learning (AI-ML) with Geographic Information Systems (GIS) to revolutionize flood management. The escalating frequency and intensity of flooding worldwide demand innovative solutions for effective flood assessment and resilience planning. Traditional flood assessment and planning approaches often fall short in providing timely and accurate insights to inform decision-making, necessitating innovative solutions that leverage advanced technologies. By harnessing AI-ML algorithms and GIS capabilities, the project aims to provide accurate flood risk assessment and inform proactive resilience strategies. Data sources include Landsat imagery from USGS, Chennai shape files from QGIS, flood inventory maps and rainfall data from Bhuvan and IMD (Indian Meteorological Department).Furthermore, the project endeavors to deploy AI-powered decision support tools to facilitate timely response and adaptive management. Ultimately, this project aspires to contribute to the development of safer and more sustainable communities in the face of escalating flood risks. The project achieves an accuracy of 95.7%, aiming to contribute to safer and more sustainable communities amidst rising flood risks.*

Keywords- Geographic Information System (GIS), Artificial Intelligence (AI), Machine Learning (ML), Disaster Prediction, Chennai, Flood.

I. INTRODUCTION

Flooding represents one of the most significant natural disasters affecting communities worldwide, posing immense challenges to infrastructure, livelihoods, and public safety. Particularly in regions like South India, characterized by low-lying coastal areas and susceptible to extreme weather events, the impact of flooding is pronounced. Chennai, a prominent coastal metropolis in South India, has experienced recurrent flooding episodes over the years, resulting in substantial damage to property, infrastructure, and human health. The failure of major rivers and drainage systems has exacerbated the vulnerability of Chennai and underscored the urgent need for proactive flood management and resilience planning.

Addressing the complexities of flood assessment and mitigation requires innovative approaches that leverage advancements in technology and data analysis. Machine Learning (ML) techniques have emerged as valuable tools for assessing the susceptibility of regions to natural hazards, including flooding. By harnessing ML algorithms and Geographic Information Systems (GIS), researchers can develop predictive models to identify flood-prone areas and assess their vulnerability comprehensively. These models rely on various factors such as rainfall patterns, terrain characteristics, land use patterns, and historical flood data to generate accurate assessments and inform proactive resilience strategies.

This paper introduces a case study focusing on the application of the Random Forest method to generate an urban flood susceptibility map of Chennai Corporation. The study aims to address the technical challenges associated with flood susceptibility assessment and provide valuable insights into the risk distribution of hazard-inducing environmental factors. By systematically analyzing input parameters and environmental constraints, the study seeks to develop a robust methodology for assessing flood risk and identifying urban flood susceptibility zones. Through the integration of AI-ML techniques with GIS capabilities, this research endeavors to contribute to the development of effective flood management strategies and enhance the resilience of communities facing escalating flood risks. By providing a systematic framework for flood susceptibility assessment, this study aims to empower decision-makers and urban planners with valuable tools to mitigate the impact of flooding and create safer, more sustainable environments for communities at risk.

II. EXISTING SYSTEM

The existing approach to flood prediction and management often relies on conventional methods that have several limitations. Typically, these methods involve the use of historical data and simplistic statistical models to forecast flood events. While these approaches may provide some level of insight into potential flood risks, they often lack the precision and timeliness required for effective disaster preparedness. Moreover, the integration of various geospatial data sources and the utilization of advanced technologies like artificial intelligence (AI) and machine learning (ML) is limited.

III. LITERATURE SURVEY

Bhoktear Mahbub Khan et al., 2023 [1] provided a comprehensive comparative analysis of machine learning models for flood prediction in India, revealing the stacked generalization model's superior performance with an accuracy of 93.3% and a low standard deviation of 0.098. Building on this research, our study aims to extend flood prediction capabilities using deep learning techniques, leveraging their capacity to capture complex patterns in data. Through this endeavor, we seek to enhance the accuracy and reliability of flood forecasting systems, contributing to the broader objective of building disaster-resilient communities and infrastructure.

Khalifa M. Al-Kindi et al., 2023 [2] investigated the utility of advanced machine learning models, namely XGBoost (XGB) and Random Forest (RF), in assessing flood susceptibility within the arid riverbeds of Wilayat As-Suwayq, Oman. By focusing on key topographical and environmental variables such as curvature, elevation, and stream power index, the study elucidated the pivotal factors influencing flood susceptibility in hydrologically vulnerable regions. Through the integration of geographical information systems (GIS) with deep learning techniques and rich remote-sensing data, the research underscored the significance of spatial analysis and data fusion in enhancing flood susceptibility mapping accuracy. This comprehensive approach not only advances flood dynamics in arid environments but also offers valuable insights for proactive mitigation strategies aimed at minimizing the impact of flooding events. By elucidating the intricate relationships between terrain features and flood susceptibility, the study contributes to the development of more effective risk assessment frameworks, ultimately bolstering resilience in regions susceptible to hydrological hazards.

K.H.V. Durga Rao et al., 2022 [3] they present a pioneering flood risk mapping approach tailored to resource-constrained environments. Their methodology utilizes tree-based machine learning models to evaluate flood susceptibility, leveraging Geomorphic Flood Descriptors (GFDs) derived from Digital Terrain Models (DTMs). Additionally, the study incorporates an efficient Data Envelopment Analysis (DEA)-based technique to map socio-economic vulnerability. By integrating flood susceptibility and socio-economic vulnerability assessments, the research generates a comprehensive GIS-based flood risk map at the finest administrative level. This novel approach not only addresses the limitations posed by resource constraints but also facilitates a holistic understanding of flood risk by considering both geomorphic and socio-economic dimensions. The resulting flood risk map provides valuable insights for decision-makers and stakeholders, enabling targeted interventions and risk mitigation strategies to enhance community resilience in flood-prone areas.

Anjeel Upreti, 2022 [4] This paper discussed the most popular supervised ML models (classification and regression) in G.I.S. and remote sensing. The classification (Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Decision Trees (DT)) and regression models (Random Forest (RF), Support Vector Machine (SVM) performance were studied. It emphasized the superiority of ML Models over traditional classifiers due to their higher accuracy in complex datasets. It serves as a valuable resource for those considering ML integration into G.I.S. and remote sensing projects.

Sankaranarayanan, S et al., 2019 [5] conducted a study that employs Deep Neural Network (DNN) for this purpose, comparing its efficacy with SVM, KNN, and Naïve Bayes, revealing DNN's superior accuracy in monsoon prediction, in addition to past efforts. The past efforts were mostly IoT and ML based for flood prediction which haven't explored temperature and rainfall intensity. By focusing on key meteorological factors, it offers a promising avenue for more precise and timely flood forecasting in flood-prone regions like India.

IV. PROPOSED SYSTEM

This project proposes a GIS platform merging elevation, land use, population density, and flood susceptibility data for comprehensive spatial understanding. Utilizing AI and ML, particularly Random Forest algorithms, for advanced flood prediction. Through sophisticated data analysis, it enables precise flood susceptibility mapping and dynamic risk assessments, facilitating informed decision-making and efficient resource allocation during flood emergencies.

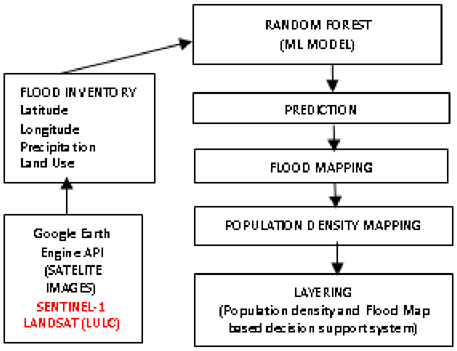

Fig 1: System Architecture

As illustrated in Figure 1, the architecture diagram depicts a flood prediction and mapping system using machine learning. It starts with a "Flood Inventory" containing data like latitude, longitude, precipitation, and land use. This data feeds into a "Random Forest (ML Model)," which processes the information and outputs to "Flood Mapping" and "Population Density Mapping." The model also integrates satellite data from "Google Earth Engine API (SENTINEL-1)" and "LANDSAT (LULC)." The final output is a decision support system that combines flood maps with population density, providing valuable insights for disaster management and planning. This system exemplifies the use of AI and geospatial data in addressing environmental challenges.

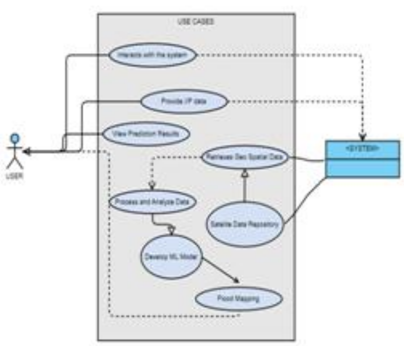

Fig 2: Use Case Diagram

As illustrated in Figure 2, the use case diagram depicts a system where a "USER" interacts with multiple functionalities: inputting IP data, viewing predictive results, retrieving and processing geospatial data, developing machine learning models, and conducting flood mapping. An external entity, "SYSTEM" provides additional geospatial data. This visual tool outlines the system's workflow, from data input to analysis and result visualization, highlighting the user's central role in navigating through the system's comprehensive capabilities.

V. IMPLEMENTATION

A. DATA ACQUISITION

The Data Acquisition module focuses on gathering relevant geospatial data necessary for flood prediction and management. This includes data such as satellite imagery, elevation data, land cover data, historical flood records, and socio-economic data. The quality and comprehensiveness of the acquired data significantly impact the accuracy and effectiveness of the flood prediction models and decision-making processes.

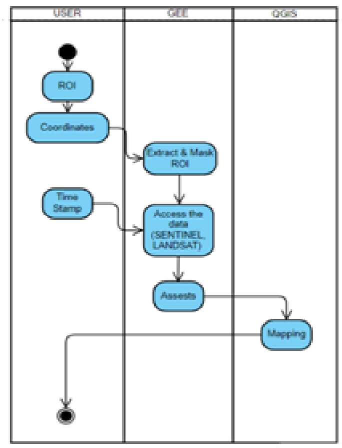

Fig 3: Module A Activity Diagram

As illustrated in Figure 3, the user defines a Region of Interest (ROI) and coordinates, which are processed in Google Earth Engine (GEE) to extract and mask the ROI. Data from satellites like SENTINEL and LANDSAT is accessed to create assets. Quantum GIS (QGIS) is then used for mapping. This workflow is essential for remote sensing and GIS applications, enabling users to analyze and visualize geographic information effectively.The steps followed are given below,

1. Identifying Data Sources:

Determine the sources from which the required geospatial data can be obtained. This may include public repositories, government agencies, research institutions, and commercial providers.

2. Accessing Data:

Retrieve the identified data from the selected sources. This may involve accessing data APIs, downloading datasets from online repositories, or requesting data directly from relevant organizations.

3. Data Preprocessing:

Clean, preprocess, and format the acquired data to ensure its compatibility and usability for subsequent analysis and modeling tasks. This may include data cleaning, filtering, transformation, and geo referencing.

Tools and Technologies:

- Google Earth Engine
- Remote sensing data repositories (e.g., Landsat, Sentinel)
- Open data portals (e.g., USGS Earth Explorer, Copernicus Open Access Hub)
- Python libraries for data manipulation (e.g., pandas, numpy)

B. MODEL BUILDING

The Model Building module is responsible for developing machine learning models to predict flood occurrences based on the acquired geospatial data. These models utilize historical flood data, environmental variables, and socio-economic indicators to identify patterns and relationships that can be used for accurate prediction of future flood events. The processes involved are given below:

1. Feature Selection and Engineering:

Identify relevant features from the acquired data that are likely to influence flood occurrences. This may involve analyzing correlation matrices, conducting exploratory data analysis, and domain knowledge expertise.

2. Model Selection:

Choose appropriate machine learning algorithms for flood prediction based on the nature of the data and the problem at hand. Commonly used algorithms include Random Forest, Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and Convolutional Neural Networks (CNN) for remote sensing data.

3. Model Training and Evaluation:

Train the selected machine learning models using historical flood data and evaluate their performance using appropriate metrics such as accuracy, precision, recall, and F1-score. This process may also involve cross-validation to assess the generalization capability of the models.

```
# Train the model
rf_classifier.fit(X_train, y_train)

# Make predictions
y_pred = rf_classifier.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

```
Accuracy: 0.9566160520607375
```

Fig 4: Accuracy

The Figure.4, shows the resultant accuracy of the random forest classifier. This accuracy is used for evaluation of the model.

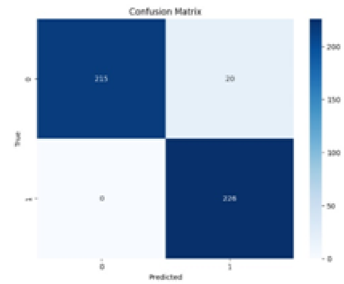

Fig 5: Confusion Matrix

The Figure.5, shows the visualization of performance of classification model.

Tools and Technologies:

- Scikit-learn
- TensorFlow
- Keras
- PyTorch

C. GIS INTEGRATION & DECISION RESCUE SYSTEM

The GIS Integration & Decision Rescue System module focuses on integrating geographic information system (GIS) functionalities into the flood prediction system and providing decision support tools for emergency response and rescue operations during flood events.

Visualization on Maps:

Visualize the results of the spatial analysis and flood prediction models on interactive maps. This provides stakeholders with a spatial context for understanding flood risks and making informed decisions.

Real-time Monitoring:

Develop systems for real-time monitoring of flood events and related factors such as rainfall intensity, river discharge, and water level changes. This enables timely response and mitigation measures during flood emergencies.

Decision Support:

Provide decision support tools and systems for emergency responders, policymakers, and other stakeholders involved in flood management. This may include risk assessment tools, evacuation planning systems, and resource allocation optimization algorithms.

Tools and Technologies:

- Geographic Information Systems (GIS) software (e.g., ArcGIS, QGIS)
- Web mapping libraries (e.g., Leaflet, OpenLayers)
- Real-time data acquisition systems (e.g., IoT sensors, weather stations)
- Decision support systems (e.g., optimization algorithms, risk assessment models)

By implementing these modules, the flood prediction and management system can effectively acquire, analyze, visualize, and utilize geospatial data for accurate prediction of flood occurrences and informed decision-making during flood events.

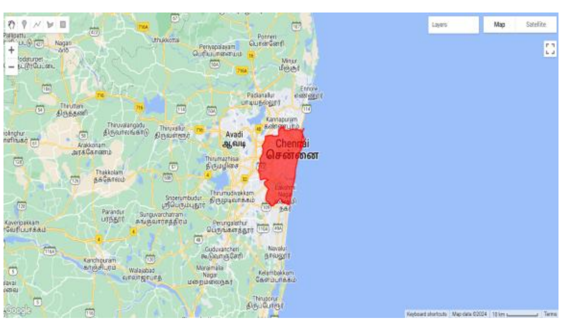

Fig 6: Region of Interest

As illustrated in Figure.6, the image showcases the Chennai region in India, delineated for geospatial analysis in QGIS. The highlighted area in red represents the extracted region of interest (ROI) using Chennai region locators.

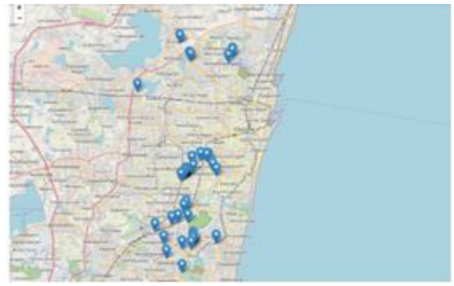

Fig 7: Resultant Flood Mapping

As illustrated in Figure.7, the Random Forest model predicts the flood region based upon the rainfall and past flood data which are acquired from IMD as shape file and converted into CSV that is used as input for prediction.

VI. CONCLUSION & FUTURE ENHANCEMENT

A. CONCLUSION

This project marks a significant endeavour in leveraging geospatial analysis techniques to tackle flood monitoring and prediction challenges in Chennai. By integrating Python libraries such as Google Earth Engine, scikit-learn, pandas, matplotlib, seaborn, and Folium, a robust framework was devised for satellite imagery analysis, flood occurrence assessment, land cover change detection, and flood prediction using machine learning. Through Landsat satellite imagery and Google Earth Engine, flood-prone areas in Chennai were successfully identified. The integration of simulated and actual flood occurrence data provided crucial insights into the spatial distribution and frequency of flood events, aiding in understanding vulnerability. Analysis of Landsat imagery enabled the quantification of land cover changes, essential for identifying regions at heightened risk of flooding. This information, coupled with population density measurements, facilitates informed decision-making and proactive flood management strategies. Leveraging the Random Forest classifier, the predictive model achieved commendable accuracy, with an impressive 95.7% success rate in forecasting flood occurrences. This comprehensive approach holds promise for enhancing flood resilience and preparedness in Chennai and similar regions.

B. FUTURE ENHANCEMENT

1. Future research could focus on enriching predictive models with additional datasets, enhancing the spatial resolution of satellite imagery.
2. Incorporating water resources data such as dams, rivers and lakes as a parameter in the model through datasets or APIs.
3. Advanced models like Deep learning and Ensemble methods can be used to improve the performance.

4. Integrating real-time data streams for more timely and accurate flood forecasts.

REFERENCES

[1] Bhoktear Mahbub Khan, Tanvir Rahman, Miah Mohammad Asif Syeed, Maisha Farzana, Ishadie Namir, Ipshita Ishrar, Meherin Hossain Nushra. "Flood Prediction using Ensemble Machine Learning Model" - IEEE 2023

[2] Khalifa M. Al Kindi, Zahra Alabri. "Investigating the Role of the Key Conditioning Factors in Flood Susceptibility Mapping Through Machine Learning Approaches" -Springer 2023.

[3] Prakhar Deroliya, Mousumi Ghosh, Mohit P. Mohanty, Subimal Ghosh, K.H.V. Durga Rao, Subhankar Karmakar. "A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions" - Elsevier 2022.

[4] Anjeel Upreti. "Machine learning application in G.I.S. and remote sensing; An overview"- Research Gate 2022.

[5] Sankaranarayanan, S., Prabhakar, M., Satish, S., Jain, P., Ramprasad, A., & Krishnan, A. (2019). Flood prediction based on weather parameters using deep learning. Journal of Water and Climate Change 2019.

[6] Yosoon Choi. **"**GeoAI: Integration of Artificial Intelligence, Machine Learning, and Deep Learning with GIS**" -** MDPI 2023.

[7] Song, H., Yang, W., Dai, S., Yuan, H. J. M. B., & Engineering. Multi-source remote sensing image classification based on 2-channel densely connected convolutional networks – NIH 2020

[8] Alem, A., & Kumar, Transfer Learning Models for Land Cover and Land Use Classification in Remote Sensing Image - Research Gate 2022.

[9] Avand, M., & Moradi, Using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability- NASA 2021

[10] Alnaim, A., Sun, Z., & Tong, Evaluating Machine Learning and Remote Sensing in Monitoring NO2 Emission of Power Plants - MDPI 2022