MINOR PROJECT

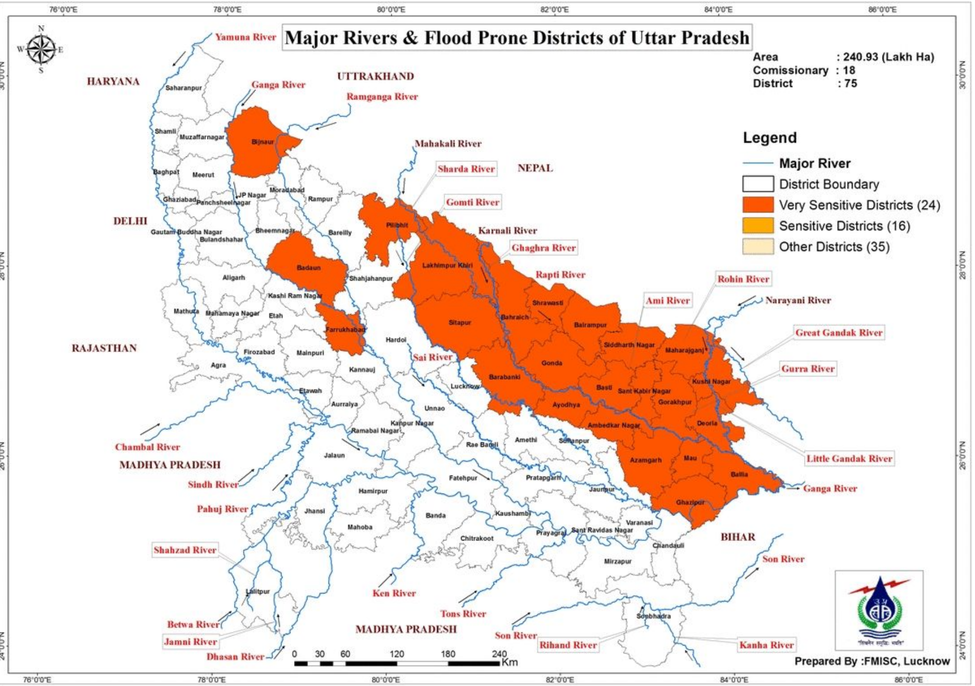
Supervised by:
Prof. Saiful Islam

Submitted by:	Faculty No.:
Anaum Khan	22COB307
Mohd. Yousuf Waseem	22COB407

FLOOD PREDICTION

Using Satellite Imagery:
River Gomti

MOTIVATION



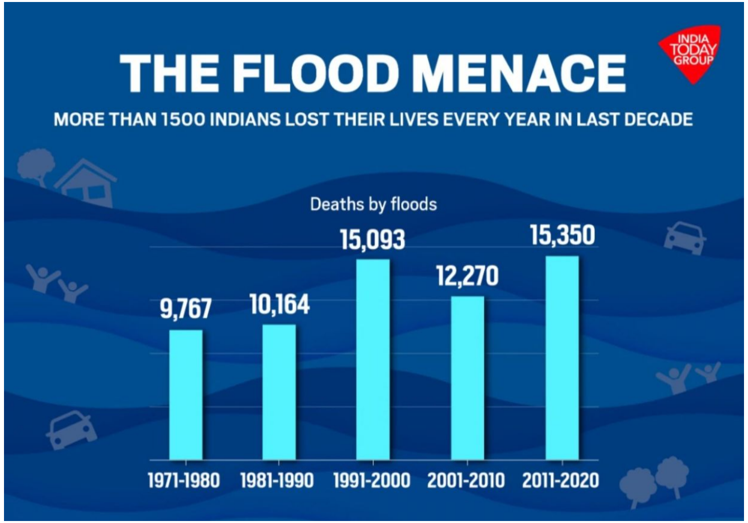
Floods are among the most devastating natural disasters, causing significant loss of life, property, and economic stability. In the context of Uttar Pradesh, the state has frequently suffered from seasonal floods due to its geographic location and river systems.

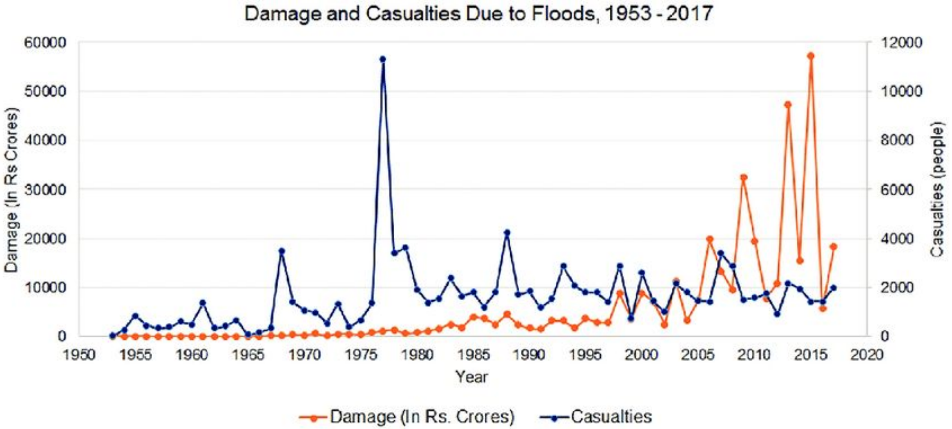
Geographical Vulnerability:

- Uttar Pradesh is traversed by major rivers like the Ganga, Yamuna, Ghaghara, and Gomti. Their tributaries often overflow during monsoons, inundating vast agricultural and urban areas.
- The low-lying regions, especially around the Gomti, are particularly prone to floods, disrupting millions of lives.

Need

- <u>Localized Risk Assessment:</u>
Predicting flood risks at a granular level (e.g., city or district) allows for better planning, especially in regions like Varanasi, Jaunpur, Azamgarh and Sultanpur, which are prone to floods.
- <u>Data-Driven Decision Making:</u>
Leveraging meteorological data (e.g., rainfall, humidity, soil moisture) ensures precise predictions. Incorporating satellite imagery from Sentinel-1 and Sentinel-2 enhances model accuracy.
- <u>Integration with Disaster Management:</u>
By predicting flood probabilities, authorities can better design embankments, improve drainage systems, and prepare emergency shelters.





FLOOD PREDICTION MODEL

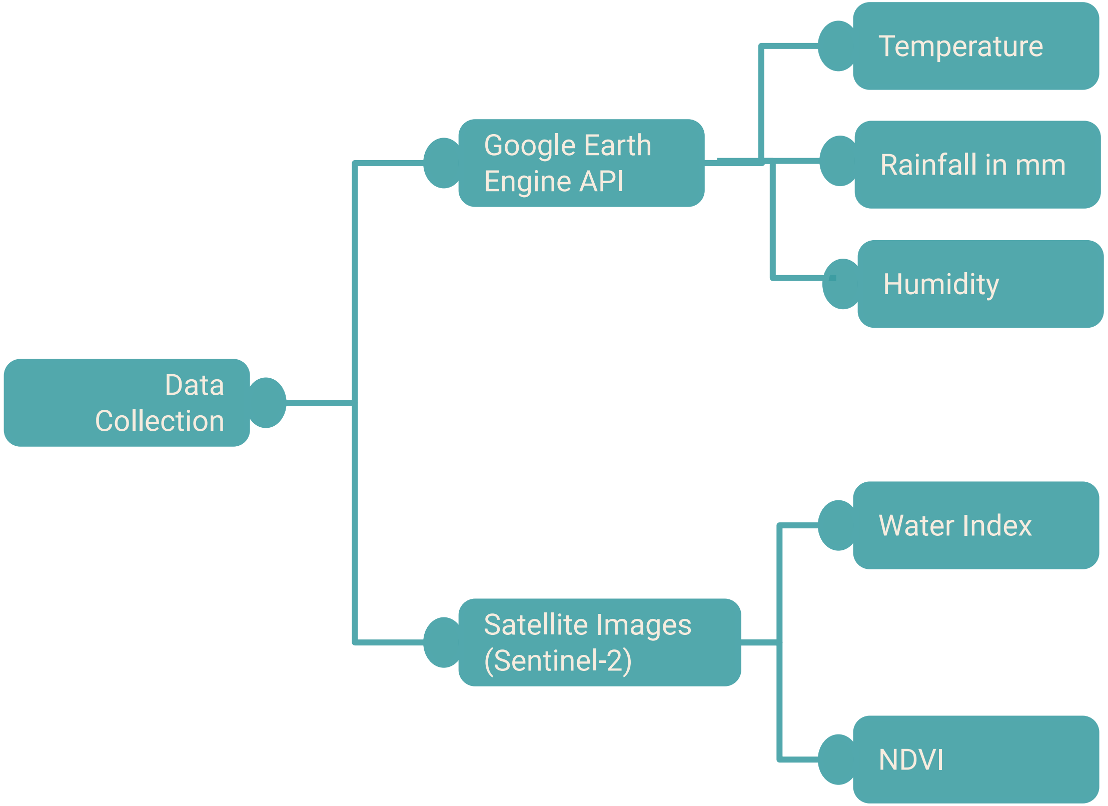
- which predicts Flood risk in binary taking satellite data and imagery as input
- and is trained on data focusing Gomti River which actually is one of the cause of severe flood in the region

LITERATURE SURVEY

Previous papers Authors	Models	Type of Data	Data pre-processing technique	Performance metrics
Dai et al.	LSVM	Hydrological data captured from IoT devices	Flood intensity classification and K-fold cross validation	RSME = 0.3991 $R^2 = 0.9943$
	QSVM			RSME = 0.3432 $R^2 = 0.9972$
	RF			RSME = 0.1571 $R^2 = 0.9559$
Prasad et al.	Random Forest Regressor	Rainfall level(numeric)	Used statistical methods to clean data	R2 = 1.00
Ghorpade et al.	Decision Tree	hydro data (numeric)	-	RMSE= 0.216; R2= 0.9945
Jain et al.	Deep Convolutional Neural Networks	MediaEval 2019 flood dataset-Sentinel-2 data (image)	-	F1 - 0.96 Kappa - 0.92
Moumtzidou et al.	Deep Neural Networks	Sentinel-2 data (image)	-	F-Score 62%

Tanim et al.	RF	Data From road closure reports along with SAR images for flood detection	Image pre-processing done using SNAP tool for noise removal and filtering.	Precision ~ 0.5 Accuracy ~ 0.65
	SVM			Precision ~ 0.85 Accuracy ~0.85
	MLC			Precision ~0.75 Accuracy ~0.82
	Unsupervised (CD, iso-clustering, fuzzy rules)			Precision ~ 0.8 Accuracy ~ 0.84
Razafipahatelo D. et al.	Digital Elevation Model (DEM), kernel k-means, non - linear clustering	SAR Images of flooded regions	-	Manual result = 807.63 ha Log Ratio in fs = 844.74 ha Ratio in fs = 1429.27 ha Ratio = 682 ha
Xu H. et al.	Urban flood inundation model, k-means clustering and improved entropy weight method	Information about river, rains and drainage, previous storm records and physical parameters like slope, distance from river etc,	Removal of invalid data	High risk zones overlap when compared with the references therefore making it a feasible approach (but limited by data provided).

DATA COLLECTION

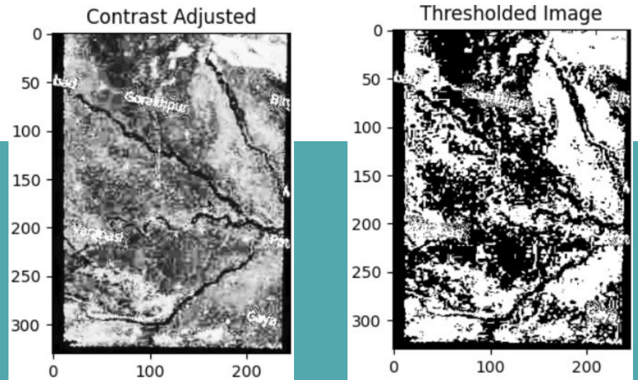



```python
ee.Authenticate()
ee.Initialize(project='imagetocsv-441315')

# Rainfall
chirps = ee.ImageCollection('UCSB-CHG/CHIRPS/DAILY') \
    .filterBounds(roi) \
    .filterDate(date_str, next_day_str) \
    .select('precipitation')

# Temperature
era5 = ee.ImageCollection('ECMWF/ERA5_LAND/HOURLY') \
    .filterBounds(roi) \
    .filterDate(date_str, next_day_str) \
    .select('temperature_2m') \

# Humidity
era5_humidity = ee.ImageCollection('ECMWF/ERA5_LAND/HOURLY') \
    .filterBounds(roi) \
    .filterDate(date_str, next_day_str) \
    .select('dewpoint_temperature_2m') \
```



DATA PREPROCESSING

TRAINING DATA

date	average_rainfall_mm	average_temperature_C	average_humidity_C	soil_moisture_0_7cm_percentage	rainfall_3day	rainfall_7day	soil_moisture_normalized	risk_index	flood_risk_binary
2011-06-01	0.000000	0.614042	0.643891	0.042398	0.000000	0.000000	0.042398	0.011005	0.0
2011-06-02	0.000000	0.664942	0.582972	0.041786	0.000000	0.000000	0.041786	0.010846	0.0
2011-06-03	0.000219	0.675961	0.540900	0.037440	0.000088	0.000048	0.037440	0.009783	0.0
2011-06-04	0.000000	0.695299	0.265529	0.026881	0.000088	0.000048	0.026881	0.007042	0.0
2011-06-05	0.000000	0.769198	0.308128	0.017098	0.000088	0.000048	0.017098	0.004503	0.0
...	...	...	...	...	...	...	...	...	...
2021-09-01	0.075771	0.348351	0.947088	0.845718	0.054864	0.224408	0.845718	0.260165	0.0
2021-09-02	0.387741	0.331301	0.932151	0.857713	0.185617	0.183096	0.857713	0.360151	1.0
2021-09-03	0.191613	0.292032	0.932824	0.877709	0.262349	0.157307	0.877709	0.422191	1.0
2021-09-04	0.003784	0.332234	0.923237	0.861254	0.233522	0.157849	0.861254	0.396562	1.0
2021-09-05	0.003043	0.384701	0.966518	0.809066	0.079466	0.157041	0.809066	0.268879	0.0

1073 rows × 9 columns

TESTING DATA

date	average_rainfall_mm	average_temperature_C	average_humidity_C	soil_moisture_0_7cm_percentage	rainfall_3day	rainfall_7day	soil_moisture_normalized	risk_index	flood_risk_binary
2021-09-06	0.041982	0.388748	0.947688	0.786383	0.019546	0.152861	0.786383	0.218597	0.0
2021-09-07	0.004613	0.304050	0.828671	0.769257	0.019877	0.153862	0.769257	0.214398	0.0
2021-09-08	0.000000	0.262820	0.856396	0.788709	0.018659	0.137408	0.788709	0.218544	0.0
2021-09-09	0.110472	0.266909	0.884759	0.883695	0.046086	0.077199	0.883695	0.263519	0.0
2021-09-10	0.070228	0.317109	0.913392	0.888325	0.072362	0.050840	0.888325	0.284188	0.0
...	...	...	...	...	...	...	...	...	...
2023-09-26	0.000269	0.262377	0.882766	0.789873	0.000108	0.011315	0.789873	0.205102	0.0
2023-09-27	0.000000	0.349186	0.902902	0.742994	0.000108	0.006001	0.742994	0.192934	0.0
2023-09-28	0.039490	0.365350	0.918453	0.703183	0.015922	0.008634	0.703183	0.194317	0.0
2023-09-29	0.000000	0.385848	0.906930	0.665408	0.015814	0.008634	0.665408	0.184432	0.0
2023-09-30	0.216228	0.360627	0.882672	0.634793	0.102404	0.055588	0.634793	0.240639	0.0

269 rows × 9 columns

THRESHOLDING

<u>**Thresholds for Risk Levels**</u>

The **Risk Index** values are classified into risk levels:

- Very High Risk: Index ≥ 0.8
- High Risk: 0.6 ≤ Index < 0.8
- Moderate Risk: 0.4 ≤ Index < 0.6
- Low Risk: 0.2 ≤ Index < 0.4
- No Risk: Index < 0.2

Binary Risk Conversion

- 0 (No Risk)
- 1 (Flood Risk) (Low, Moderate, High, Very High)

$$\text{Risk Index} = \alpha \times \text{Normalized Rainfall (3-day)} + \beta \times \text{Normalized Soil Moisture}$$

Normalization:

Rainfall normalized to a threshold of 300 mm.

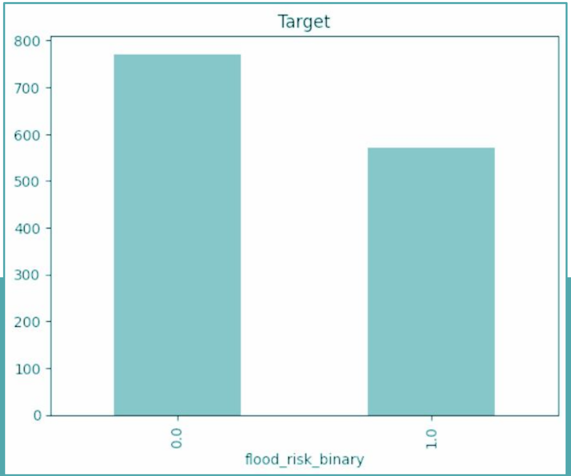
Weights:

α=0.6 (Rainfall)

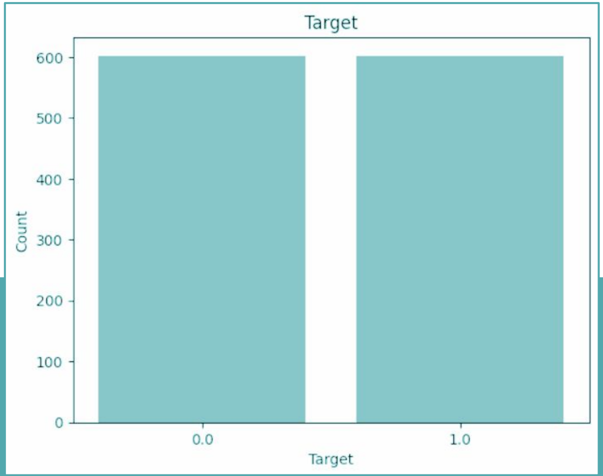
β=0.4 (Soil Moisture)

date	cloud_cov	average_ra	average_te	average_h	soil_moist	rainfall_3d	rainfall_7d	soil_moist	risk_index	flood_risk_	flood_risk_binary
24-06-2011		0.507035	303.6087	296.3388	30.42482	3.027129	50.49947	0.304248	0.127754	No Risk	0
25-06-2011		5.856679	302.4905	297.599	29.5303	8.55873	56.15094	0.295303	0.135239	No Risk	0
26-06-2011		8.058068	301.9205	298.4207	34.69054	14.42178	57.37796	0.346905	0.167606	No Risk	0
27-06-2011		7.791306	301.1606	298.1537	38.898	21.70605	27.36856	0.38898	0.199004	No Risk	0
28-06-2011		11.12623	300.4248	297.9844	40.01378	26.97561	35.85942	0.400138	0.214006	Low Risk	1
29-06-2011		58.62102	300.11	297.8438	40.51132	77.53856	94.15536	0.405113	0.317122	Low Risk	1
30-06-2011		40.11999	299.1873	297.8947	42.53527	109.8672	132.0803	0.425353	0.389876	Low Risk	1
01-07-2011		23.51566	301.2184	297.8665	41.53758	122.2567	155.089	0.415376	0.410664	Moderate	1
02-07-2011		8.048651	303.727	297.3646	38.49671	71.6843	157.2809	0.384967	0.297355	Low Risk	1
03-07-2011		6.366578	304.1261	297.4365	35.73937	37.93089	155.5894	0.357394	0.218819	Low Risk	1
04-07-2011		7.125435	302.8474	298.6193	36.47667	21.54066	154.9236	0.364767	0.188988	No Risk	0
05-07-2011		16.3824	301.5504	298.0147	40.1718	29.87441	160.1797	0.401718	0.220436	Low Risk	1
06-07-2011		0.478003	302.4403	297.722	38.05554	23.98584	102.0367	0.380555	0.200194	Low Risk	1
07-07-2011		0.770416	302.3586	297.7447	36.97037	17.63082	62.68714	0.369704	0.183143	No Risk	0

APPLIED SMOTE



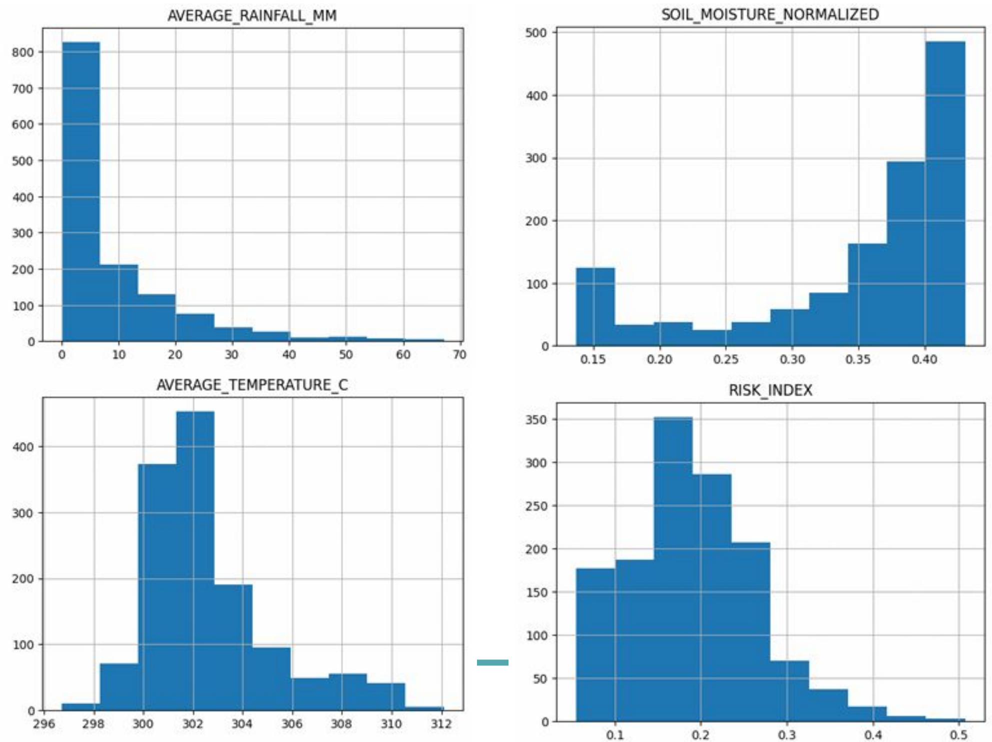
Before SMOTE



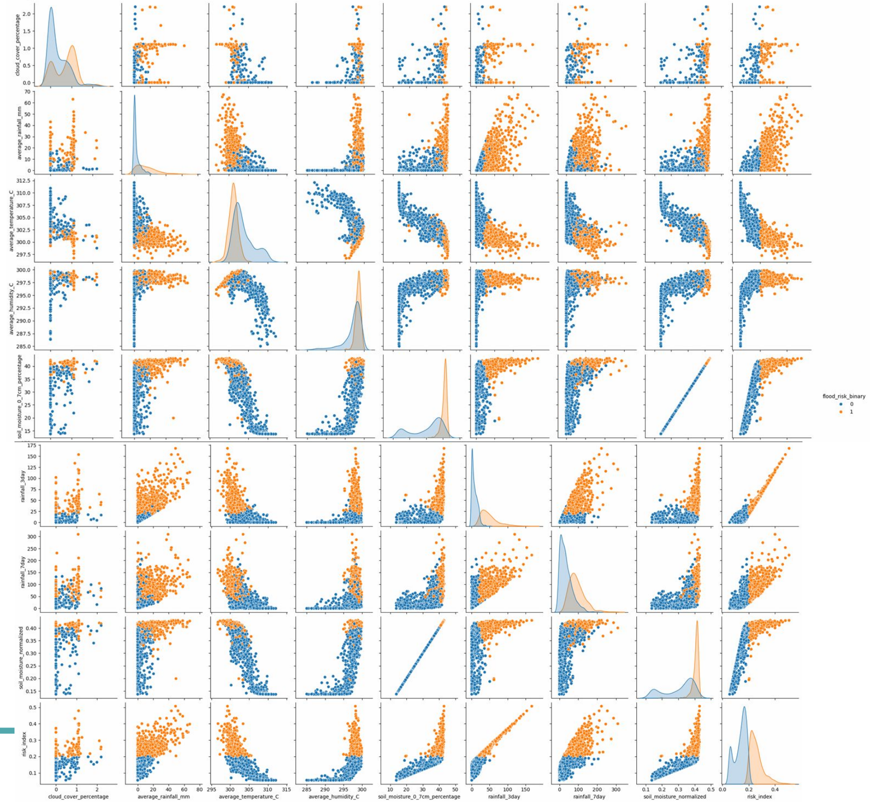
After SMOTE

Applied SMOTE to get balanced data

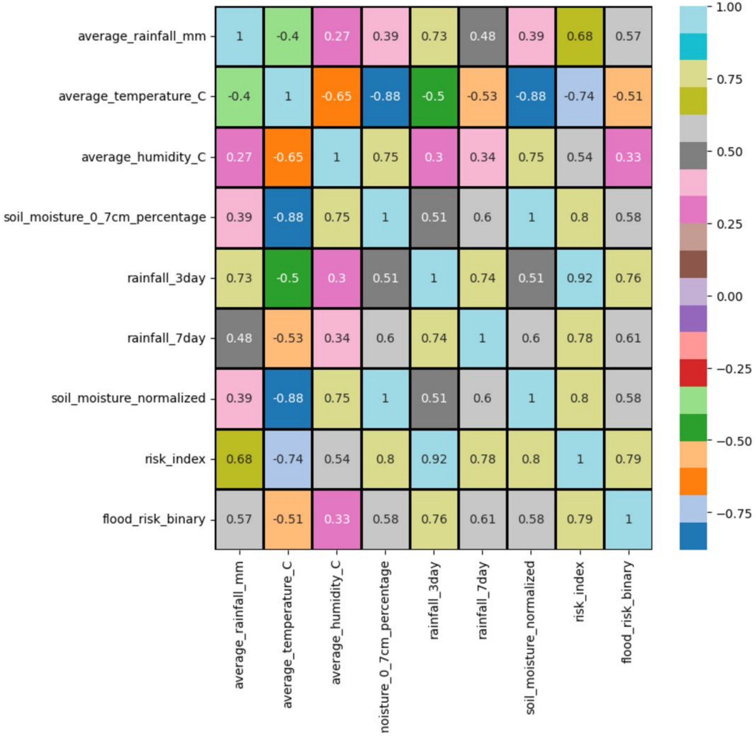
DATA EXPLORATION

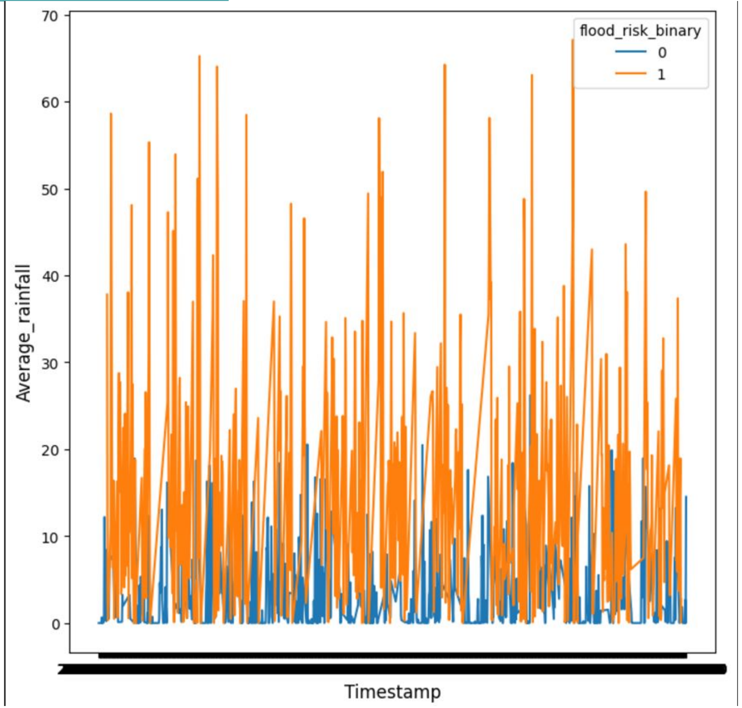


<seaborn.axisgrid.PairGrid at 0x788ab818ee60>
<Figure size 800x800 with 0 Axes>

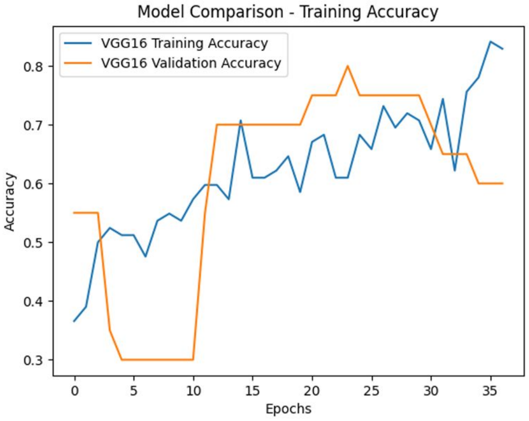


DATA EXPLORATION

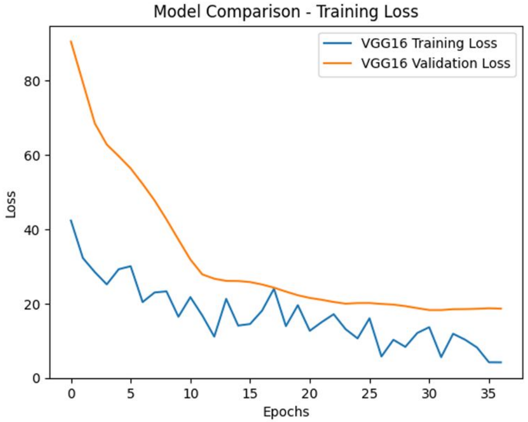




MODEL COMPARISON: VGG16



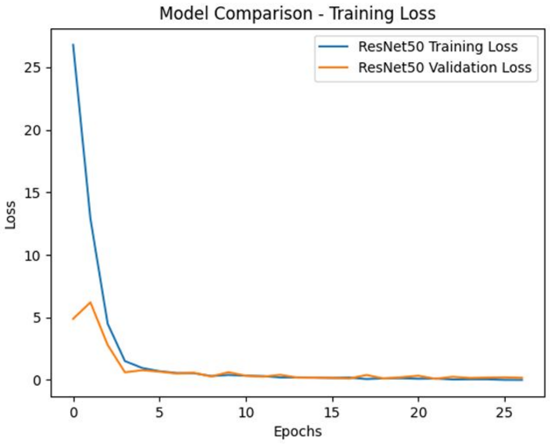
Vgg-Training Accuracy Comparison.



Vgg-Training Loss Comparison.

Class	Precision	Recall	F1-Score	Support
Accuracy			0.60	
macro avg	0.62	0.63	0.60	20
weighted avg	0.66	0.60	0.61	20

MODEL COMPARISON: ResNet50

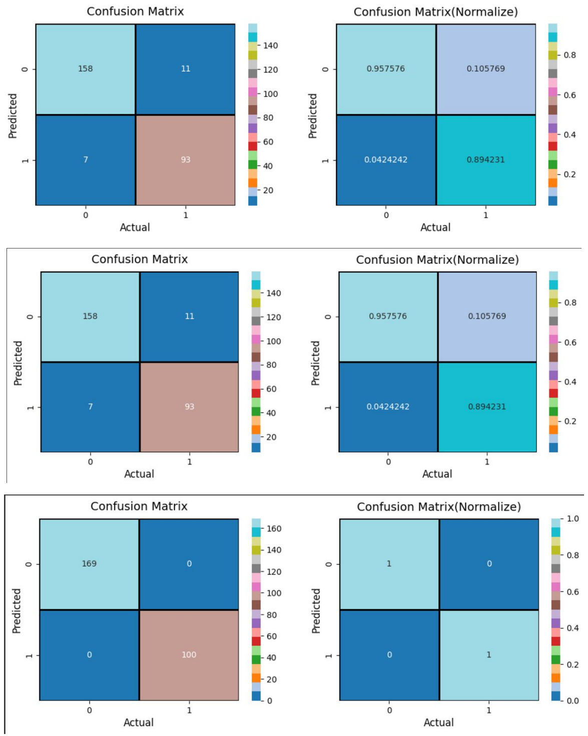


ResNet-Training Loss Comparison.

ResNet-Training Accuracy Comparison.

Class	Precision	Recall	F1-score	Support
Accuracy			0.75	
Macro avg	0.75	0.68	0.69	20
Weighted avg	0.75	0.75	0.73	20

ML ALGORITHMS



- **Support Vector Classifier**

Precision: 89.0 %
Recall: 93.0 %
F1 Score: 91.0 %

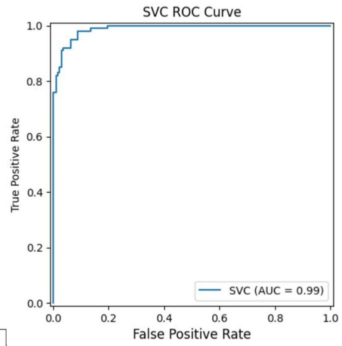
- **Random Forest Classifier**

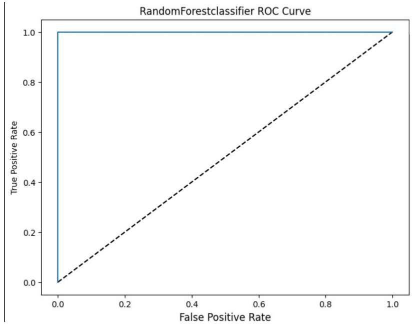
Precision: 89.0 %
Recall: 93.0 %
F1 Score: 91.0 %

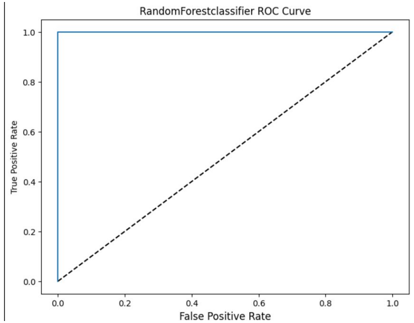
- **XGBoost Classifier**

Precision: 100.0 %
Recall: 100.0 %
F1 Score: 100.0 %

INFERENCE







Support Vector Classifier:

- Works well with linearly separable data and in high-dimensional spaces.
- Struggles with noisy data and large datasets.

Random Forest Classifier:

- Reduces overfitting, handles high-dimensional data, and doesn’t require normalization.
- High computational cost, less interpretable, and slow training.

XGBoost Classifier:

- Prevents overfitting and excels with large datasets having few features.
- Sensitive to outliers and performs poorly on unstructured, sparse data.

FUTURE WORK

1. **Data Enhancement:**
   - Use higher-resolution spatial and temporal datasets for localized predictions.
2. **Model Improvement:**
   - Implement ensemble methods combining SVC, RFC, XGBoost, and neural networks for robust predictions.
3. **Deployment and Real-Time Applications:**
   - Develop interactive dashboards for visualization and decision-making.
4. **Validation and Scalability:**
   - Collaborate with stakeholders for early warning systems and disaster management.

REFERENCES

- Keerthika, T., Ashwini, A., Amirta Varsni, R., Angela Princy, A., & Dinesh Kumar, S. (2020). Feature Extraction from Satellite Images Using Deep Learning. *International Journal of Emerging Technology and Innovative Engineering*, 6(3), 174. ISSN: 2394–6598.
- Nouioua, T. (n.d.). Flood Detection in Satellite Images using Deep Learning. *University of Tebessa, Algeria*. Department of Mathematics and Computer Sciences.
- Flood Management Information System Centre
- Akhyar, A., Cho, C., Zulkifley, M. A., Hyun, S., Lee, J., Son, Y., Song, T., & Hong, B.-W. (n.d.). Deep artificial intelligence applications for natural disaster management systems: A methodological review.
- Gautam, P. K., Chandra, S., & Henry, P. K. (2024). Estimation of flood inundation in river basins of Uttar Pradesh using Sentinel 1A-SAR data on Sentinel Application Platform (SNAP). *Arabian Journal of Geosciences, 17*(107). https://doi.org/10.1007/s12517-024-11910-x

THANK YOU

All we wish for is that people stay safe and that floods are predicted and mitigated before they can pose any harm.

~Anaum Khan
~Mohd. Yousuf Waseem