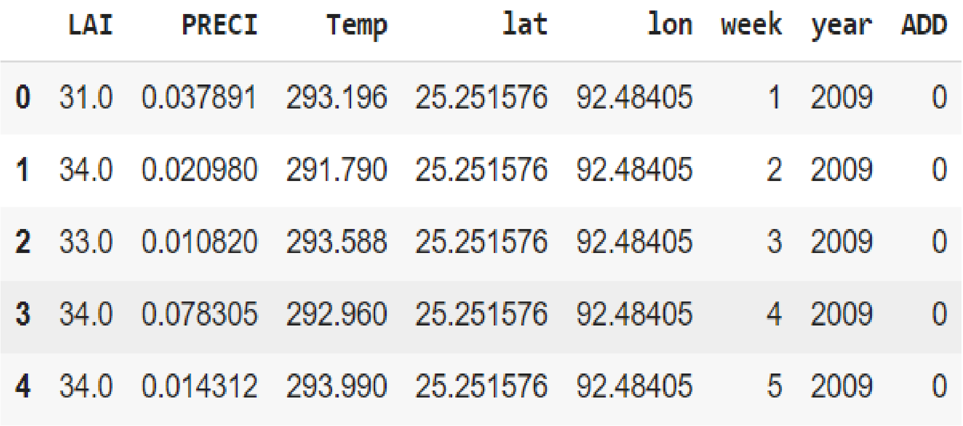
Methodology

1. **Data Collection and Understanding:**
   We Collected a comprehensive dataset with features relevant to Alzheimer's Disease prediction from the Health Ministry. The dataset consists of 516060 values for data. We understood the dataset by checking its shape, examining the first few rows, and handling missing values.

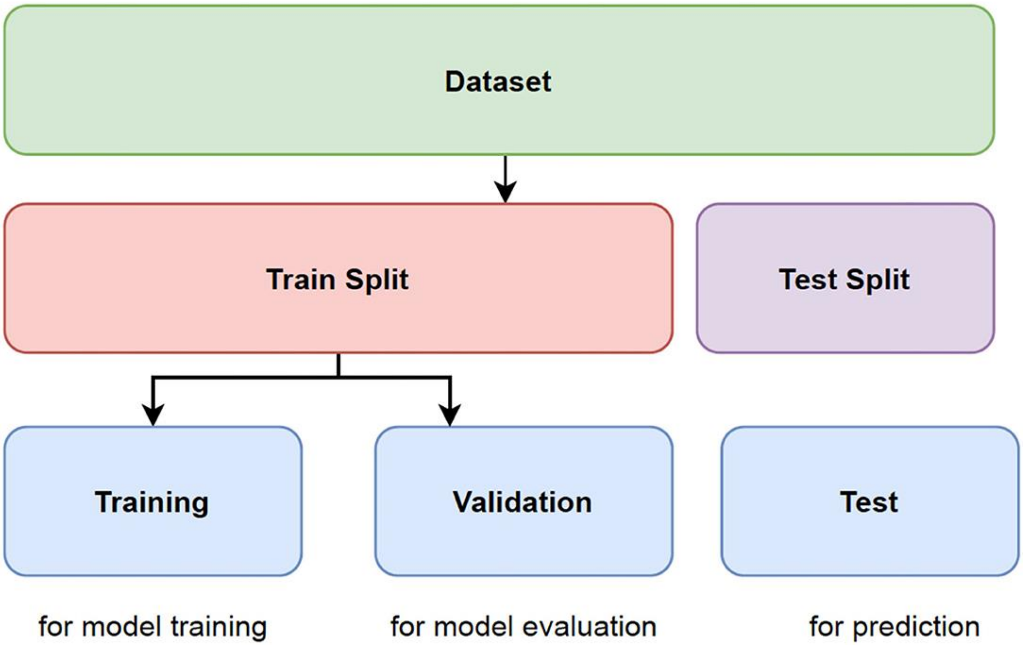
TABLE 1. Dataset description.



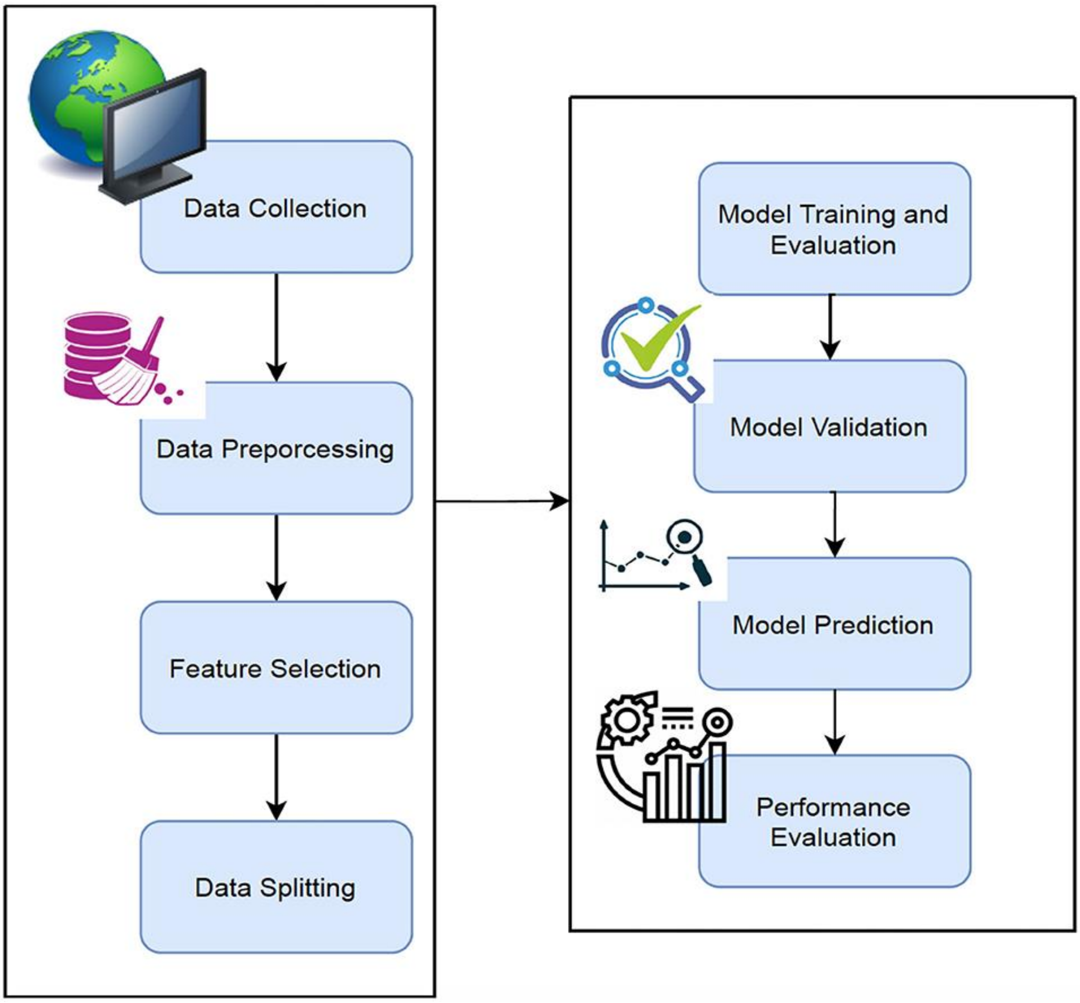
	LAI	PRECI	Temp	lat	lon	week	year	ADD
0	31.0	0.037891	293.196	25.251576	92.48405	1	2009	0
1	34.0	0.020980	291.790	25.251576	92.48405	2	2009	0
2	33.0	0.010820	293.588	25.251576	92.48405	3	2009	0
3	34.0	0.078305	292.960	25.251576	92.48405	4	2009	0
4	34.0	0.014312	293.990	25.251576	92.48405	5	2009	0

2. **Data Preprocessing:**
   We dropped or impute missing values. Then we splitted the dataset into features (X) and target variable (y). We performed a train-test split for model evaluation.
   We splitted the dataset in 80-20 ratio for the train and the test dataset. Then We subdivided the train set into training and validation dataset.

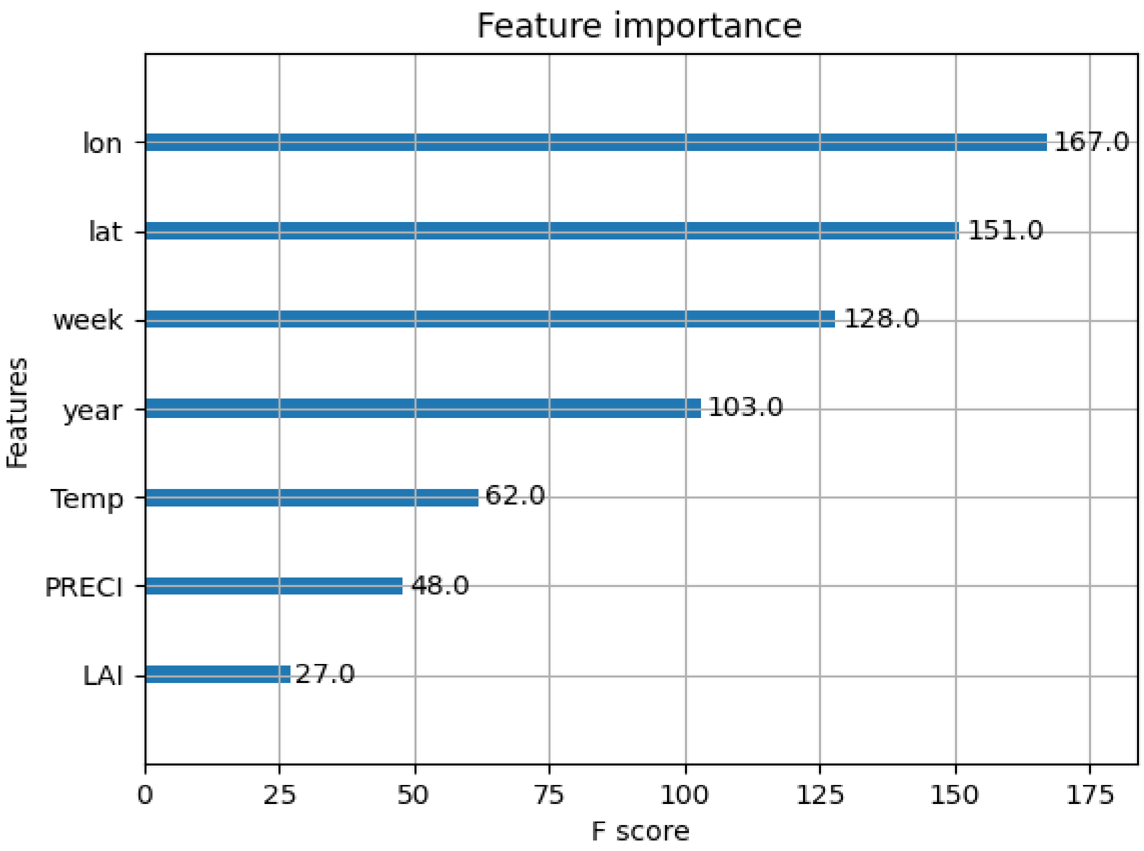
FIGURE 1. Representation of data splitting.



3. **Addressing Class Imbalance:**
   We investigated and handled class imbalance using techniques like SMOTE and random under sampling. We considered the imbalance ratio in the model training.



4. **Xgboost Model Training:**
   We set up the Xgboost model with initial hyperparameters.
   Then we trained the model on the training set and evaluated its performance on the test set. We monitored key metrics such as F1 score, precision, recall, accuracy, and log loss.



5. **Model Evaluation:**
   We assess the model's performance using metrics like F1 score, accuracy, precision, recall, and log loss.
   Then we visualized the confusion matrix to understand true positives, true negatives, false positives, and false negatives.

6. **Hyperparameter Tuning (Optuna):**
   We used Optuna for hyperparameter tuning to find the optimal combination of hyperparameters.
   We defined the objective function to maximize the F1 score. Then we conducted multiple trials to find the best hyperparameters.

7. **Final Model Evaluation:**
   We created the final Xgboost model with the best hyperparameters obtained from Optuna. We trained the final model on the training set. We evaluated the final model on the test set using key metrics.

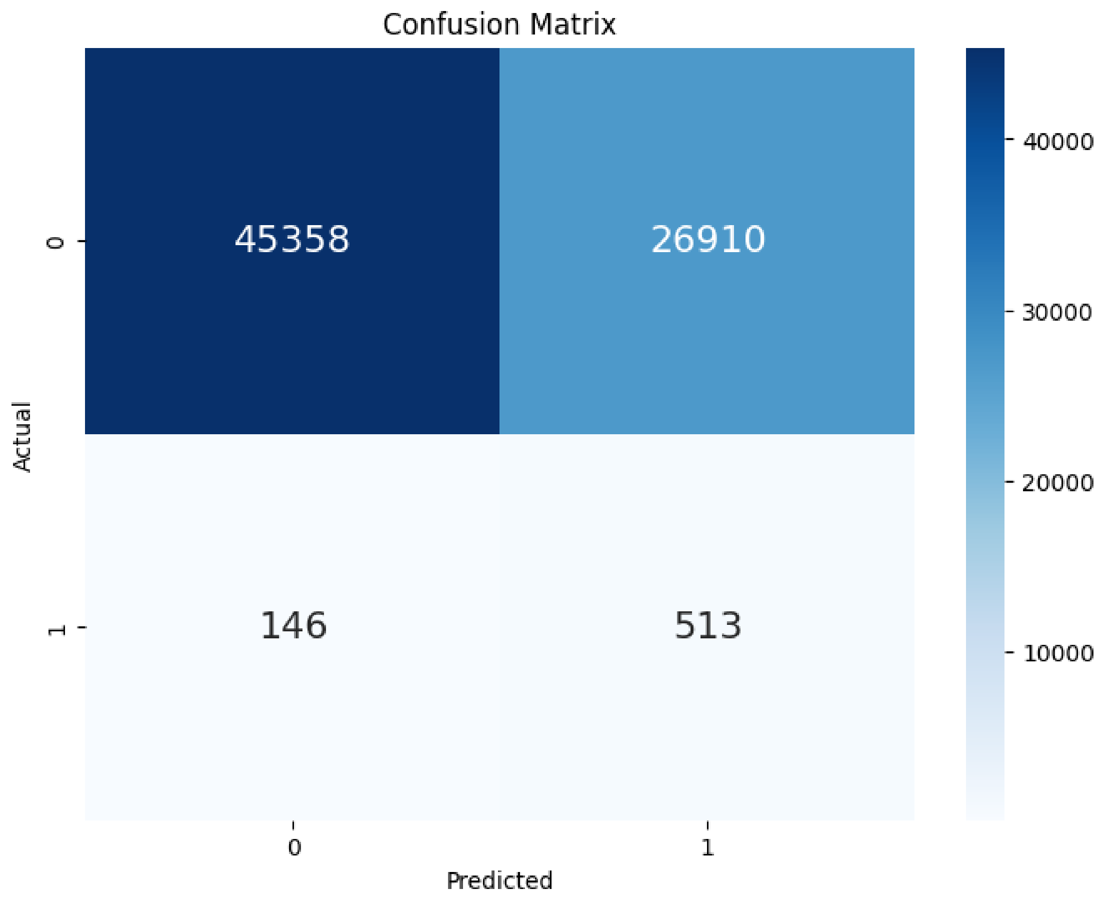
Observations

Model Training Performance:

The Xgboost model demonstrated commendable performance during the initial training phase, achieving high accuracy, precision, recall, and F1 score on the training set.

Testing Set Performance:

The model's performance on the test set was consistent with the training results, indicating generalization capability. The F1 score and accuracy metrics remained high, suggesting the model's effectiveness in predicting Alzheimer's Disease.



Optuna Hyperparameter Tuning:

The Optuna hyperparameter tuning process effectively identified optimal hyperparameters that further improved the model's performance on the validation set. This demonstrates the importance of fine-tuning to achieve the best possible predictive capability.

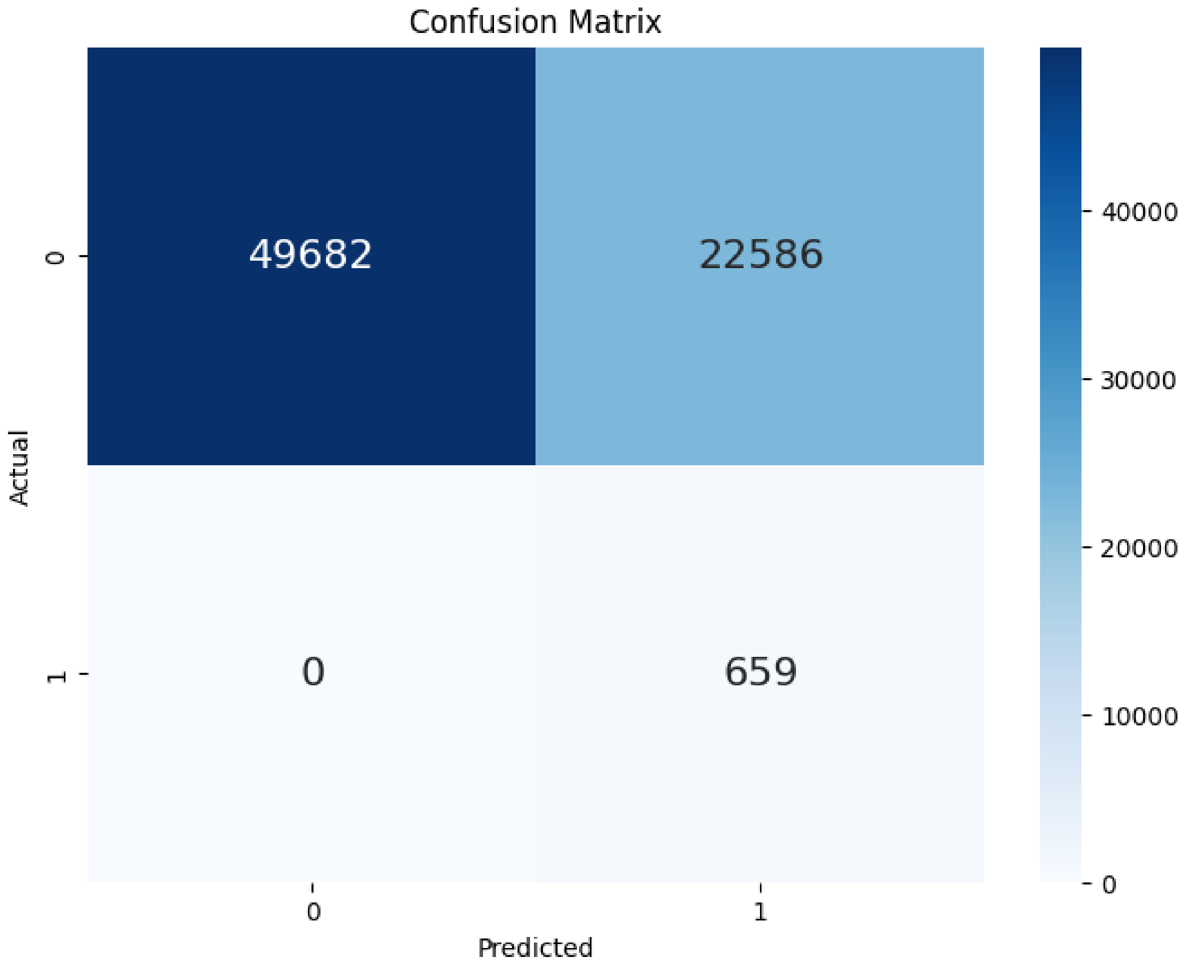
Final Model Evaluation:

The final Xgboost model, trained with the optimal hyperparameters, exhibited strong predictive power on the test set, achieving a notable F1 score and accuracy. This validates the efficacy of the hyperparameter tuning process.

Confusion Matrix Analysis:

Analysis of the confusion matrix revealed the model's ability to correctly identify Alzheimer's cases, minimizing false negatives. This is particularly crucial in a healthcare context, where early detection is of paramount importance.

FIGURE 2. Confusion Matrix after Hyperparameter Tuning.


Confusion Matrix
Actual
Predicted