# TERM PROJECT

# FOUNDATIONS OF MACHINE LEARNING
# EC346

**By**
**ANIRUDHA BHAT NEKKARE - 211EC268**


**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING,**
**NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL,**
**MANGALORE -575025 MAY, 2023**

# OBJECTIVE

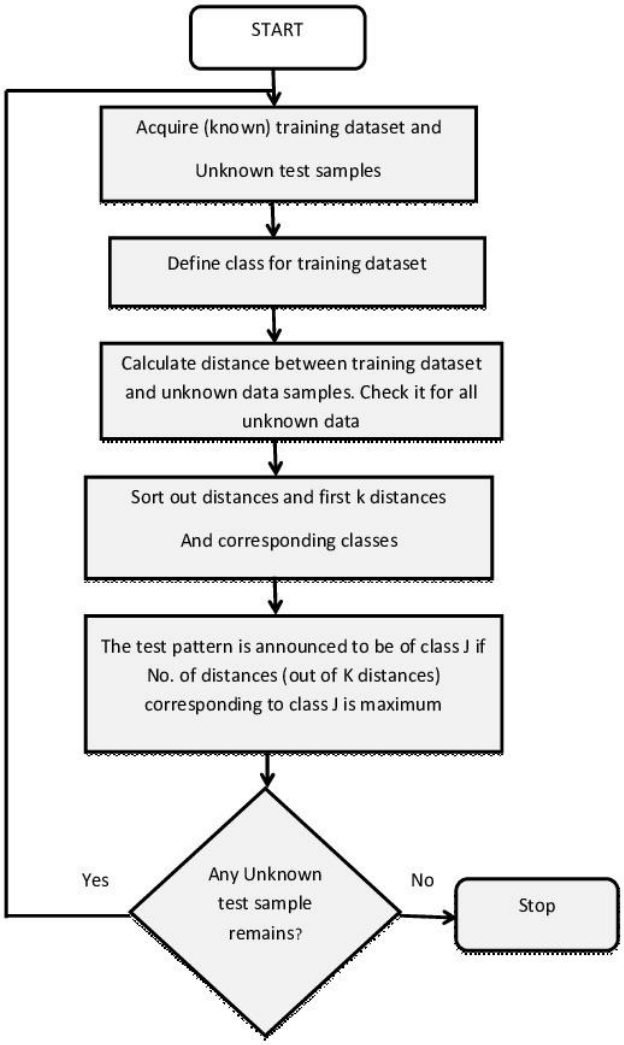
To design a machine learning project with the objective of categorizing provided images into either bleeding or non-bleeding classes, employing feature selection/extraction techniques and ensemble methods for enhanced classification accuracy.

# PROPOSED MODELS

## *1. K-Nearest Neighbour (KNN)*

- The code implements a KNN classifier for image classification, focusing on bleeding and non-bleeding samples. The dataset is preprocessed, split, and used to train a KNN model.
- Images are loaded, converted to grayscale, resized, and flattened for KNN processing and are organized into ‘Bleeding’ and ‘Non-Bleeding’ classes.
- The given data is split into 80% of training dataset and 20% of testing dataset.
- Model trained on the training set and predictions are made on the test set and accuracy is evaluated using accuracy score.

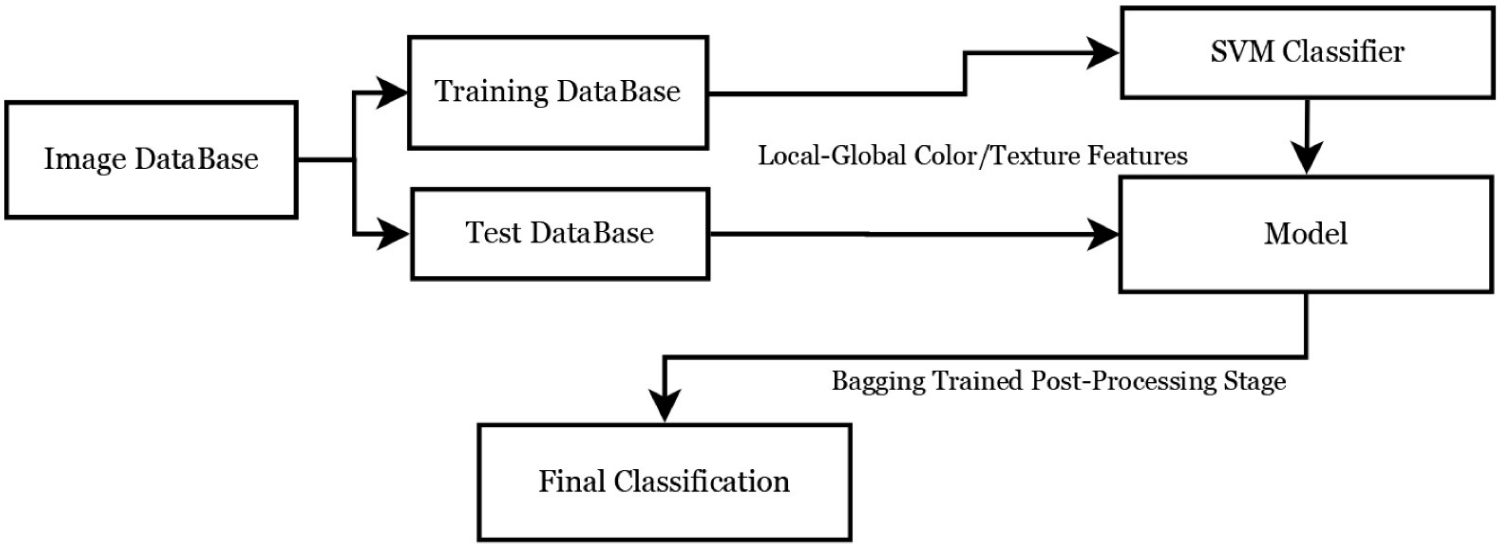
# BLOCK DIAGRAM



*2. Support Vector Machine (SVM)*

- The code uses a Support Vector Machine (SVM) classifier for image classification to focus on classification of bleeding and non-bleeding samples. Images are preprocessed, split into training and testing sets, and used to train an SVM model.
- Images loaded and preprocessed similarly to KNN, labeled as 'Bleeding' or 'NonBleeding.'
- SVM initialized with the Radial basis function or RBF kernel.
- Model is trained on the training dataset and predictions made on the test dataset and accuracy calculated using accuracy_score.
- SVM achieves reasonable accuracy, indicating its effectiveness for image classification.

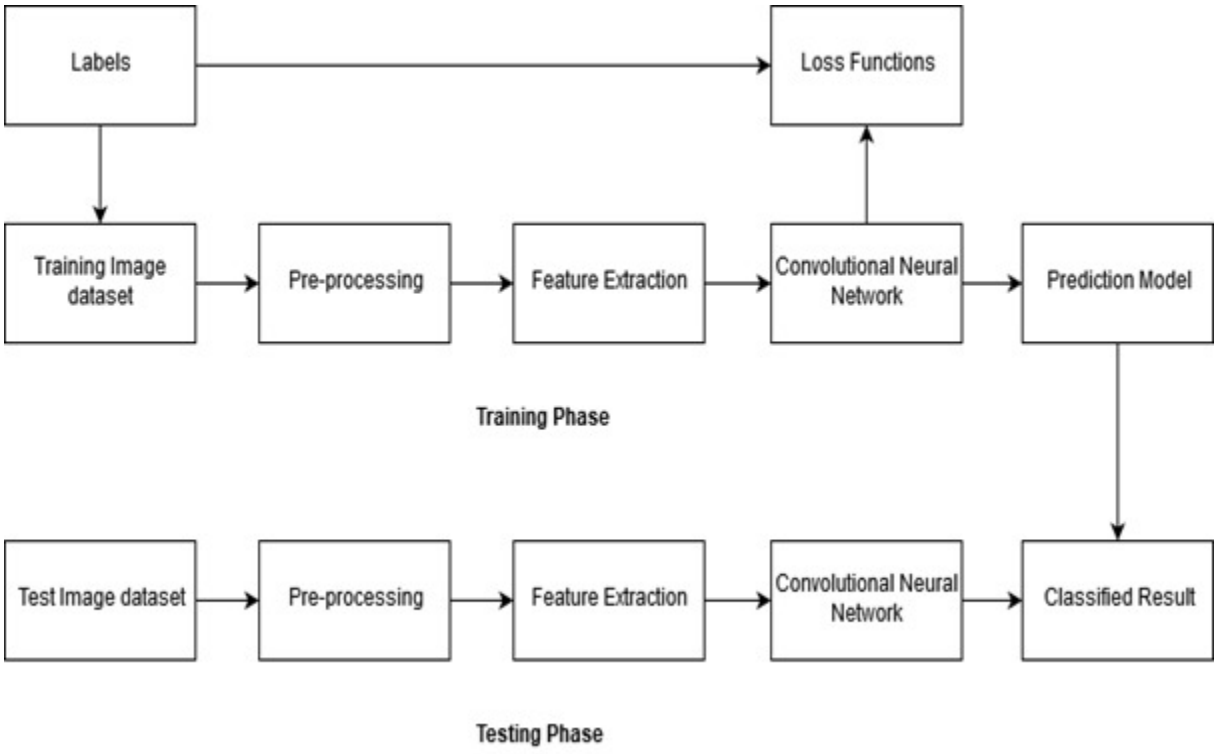
# BLOCK DIAGRAM



*3. Convolutional Neural Network (CNN)*

- Sequential CNN model constructed with convolutional, pooling, and dense layers.
- The model is compiled with Adam optimizer and categorical cross-entropy loss.
- Image data is augmented for diverse training samples using ‘ImageDataGenerator’.
- Model trained on augmented training data and validated on the test set.

# BLOCK DIAGRAM

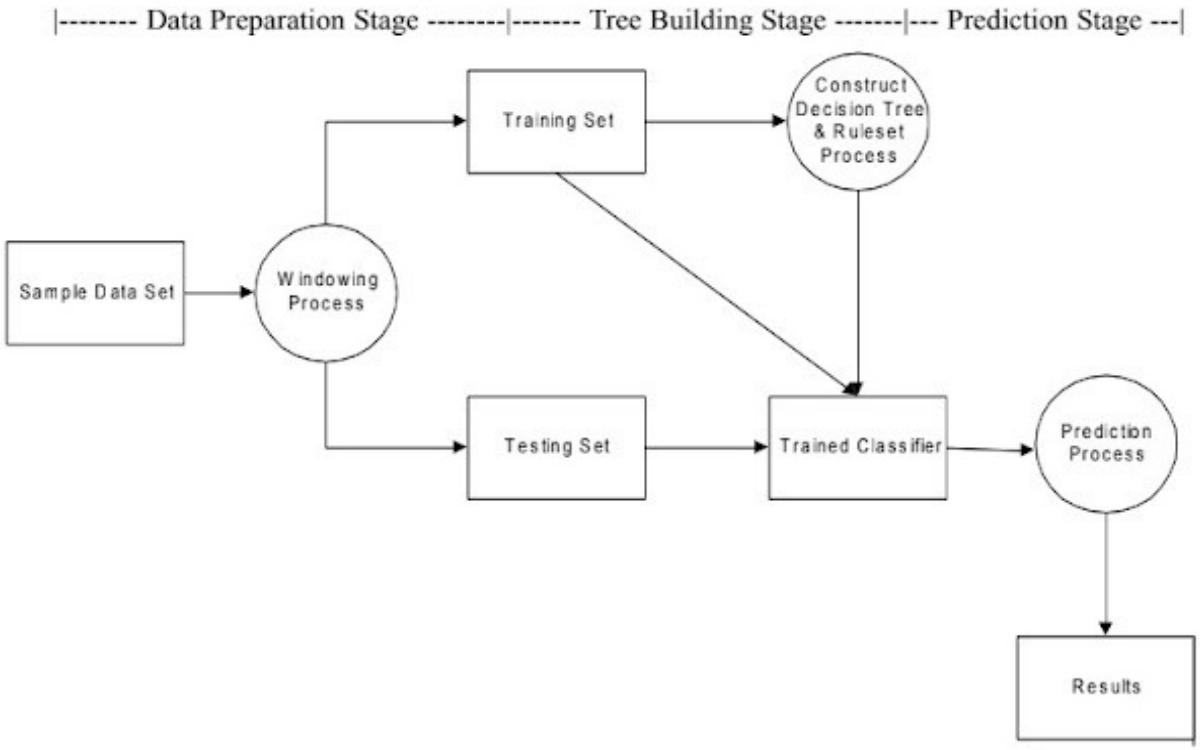


*4. Decision Trees (DT)*

- The model is trained on the training set, learning patterns and relationships within the data.
- The Decision Tree classifier is initialized with a specified random state for reproducibility.
- Predictions made on the test set, and accuracy calculated using accuracy_score.

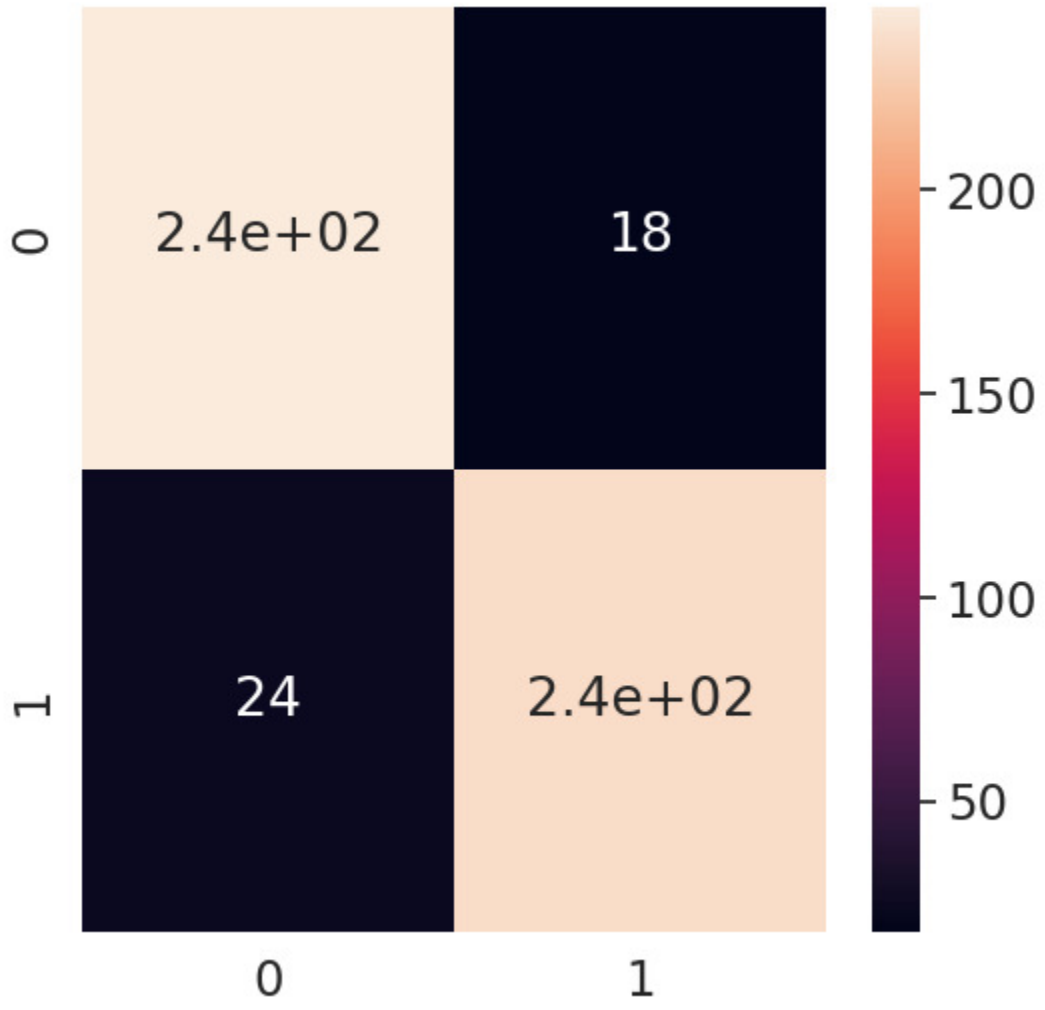
Consider implementing ensemble methods, such as Random Forests, to harness the strengths of multiple decision trees for improved robustness.

# BLOCK DIAGRAM

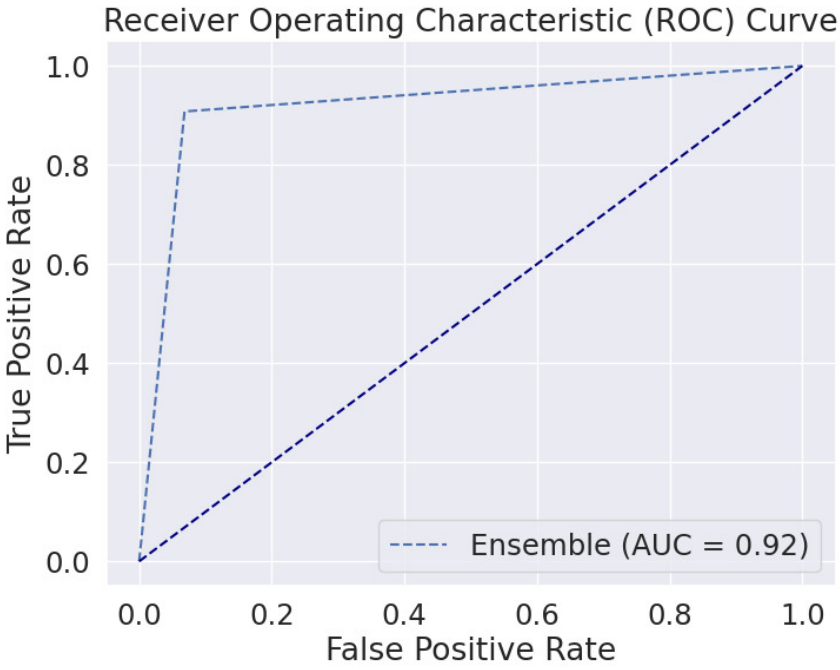


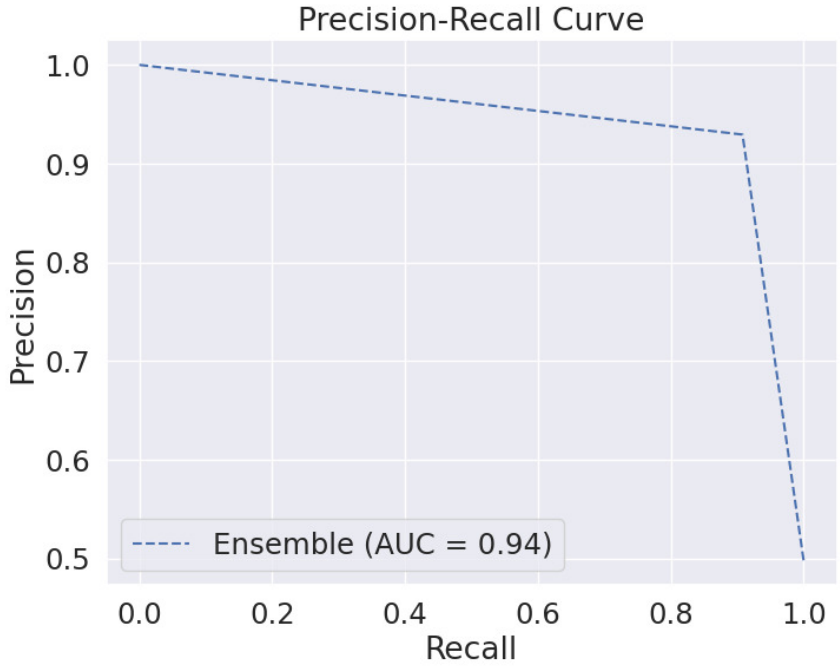
## Ensemble Method

- Individual models (KNN, SVM, CNN, Decision Tree) are trained on the training set.
- Predictions are combined using simple averaging.
- Ensemble performance is evaluated using accuracy, precision, recall, and F1 score.
- Confusion matrix is visualized.



- ROC curve and precision-recall curve are plotted for additional performance insights.



## EVALUATION METRICS FOR VALIDATION DATASET

- KNN Accuracy: 0.92
- SVM Accuracy: 0.94
- CNN Accuracy: 0.90
- DT Accuracy: 0.90
- Ensemble Accuracy: 0.92
- Ensemble Precision: 0.93
- Ensemble Recall: 0.91
- Ensemble F1 Score: 0.92
- Ensemble Confusion Matrix:
  [[245  18]
   [ 24 237]]