SVM-RF-CNN GRADIENT BOOSTING ENSEMBLE LEARNING FOR SKIN CANCER DETACTION

Presented By: Group No- 02
Akankshya Pattanaik (2361020015)
Asmita Bijaya (2361020030)
Abhilipsa Sahoo (2361020007)
Suchitra Das (2361020296)
4th Semester MCA-2025

Supervised By: Dr. Bichitrananda Patra
Professor, CA Department

Department of Computer Application
Institute of Technical Education and Research(ITER)
Siksha 'O' Anusandhan Deemed to be University
Bhubaneswar-30

Table of Contents

INTRODUCTION

- Early and accurate detection of skin cancer is vital, but traditional methods can be slow and subjective.
- Machine Learning models, especially CNNs, offer a faster and more reliable way to classify skin lesions, improving diagnosis accuracy and accessibility

MOTIVATION

- Skin cancer is a common and serious disease, making early detection crucial for effective treatment and survival.
- Traditional diagnostic methods are often slow, subjective, and depend on the availability of skilled professionals.
- Machine Learning offers fast, accurate analysis of skin lesion images, identifying patterns beyond human perception.
- An ML-based system can improve early diagnosis, reduce errors, and increase access to healthcare, especially in remote areas.

OBJECTIVE

- The main objective is to detect skin cancer early using machine learning and image processing techniques.
- The project uses SVM, CNN, and Random Forest models to classify skin lesions as cancerous or non-cancerous.
- An ensemble technique is applied to improve overall accuracy and reliability of the diagnosis

LITERATURE SURVEY

...

Title	Dataset Used	Classification	Feature Extraction	Accuracy (%)
"Melanoma Skin Cancer Detection using Image Processing and Machine Learning"(June 2019)	ISIC 2018 dataset	Support Vector Machine (SVM), CNN, Backpropagation Neural Network (BPNN)	Gray-Level-Co-Occurrence Matrics (GLCM) for texture features, color features, Shap, Size	87.44%
“A Melanoma Skin Cancer Detection Using Machine Learning Technique: Support Vector Machine” (2021)	ISIC Dataset	Support Vector Machine (SVM)	Asymetry, Color & texture features	86.64%
"Diagnosis of skin cancer using machine learning techniques " (2020, Elsevier)	HAM10000, ISIC datasets	Support Vector Machine (SVM), Probabilistic Nural Network (PNN), Random Forest	Gray-Level-Co-Occurrence MatricsGLCM, Generalized Low Rank Model (GLRM)	86.51%

CHALLENGES

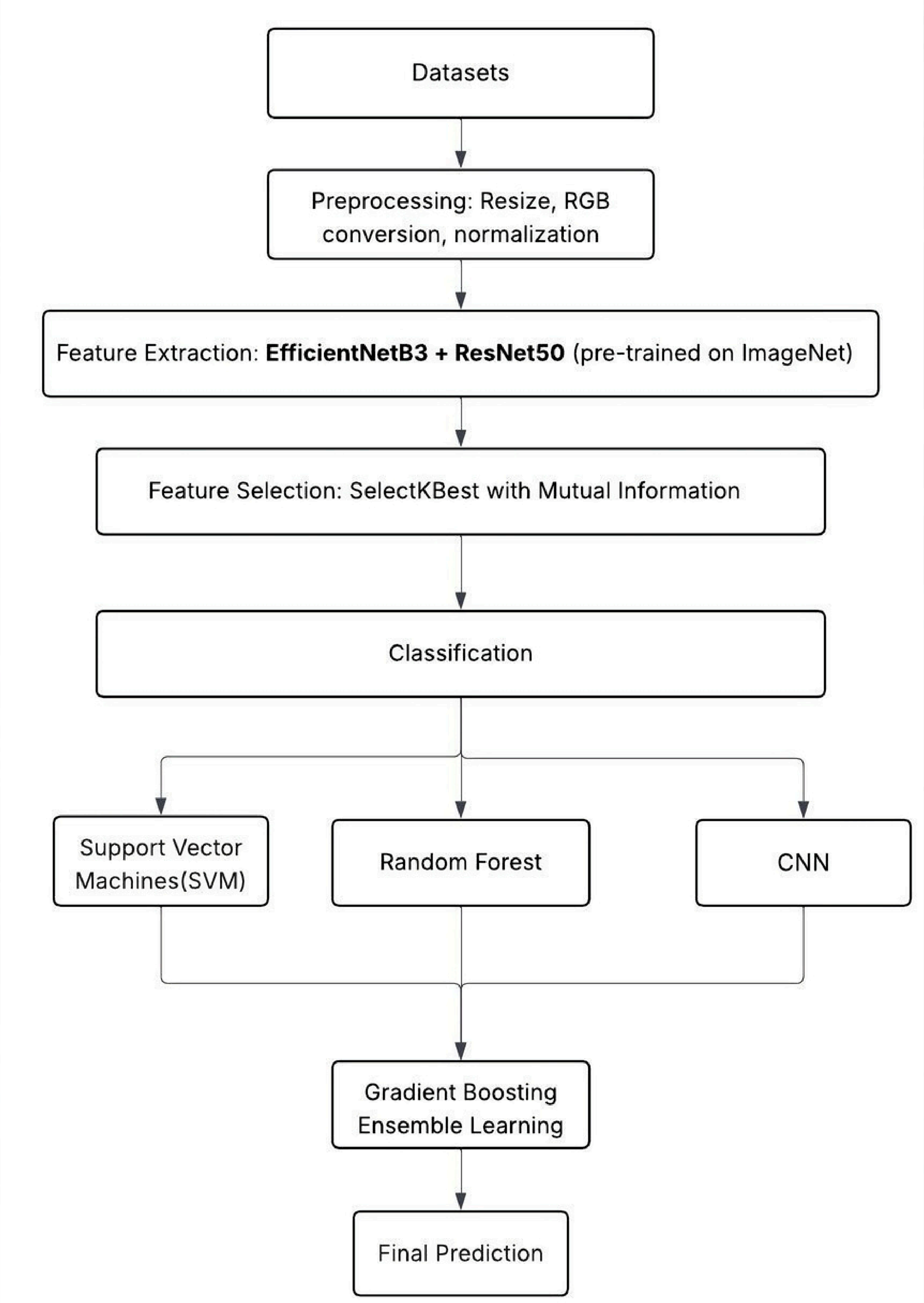
- Early detection is difficult due to similarity between benign and malignant lesions.
- Traditional diagnosis is slow and subjective, relying on expert interpretation.
- Limited access to specialists in remote or underdeveloped areas.

DATASET DESCRIPTION

- Name: Skin Cancer: Malignant vs. Benign
- Images: Approximately 3,297 images in total.
- Classes: Two main classes —
  Benign (non-cancerous)- 1,497 images
  Malignant (cancerous)- 1,800 images

(a) Benign (b) Malignant

PROPOSED MODEL



FEATURE EXTRACTION

- The project uses pre-trained EfficientNetB3 and ResNet50 models to extract important features from images.

- The main goal is to classify skin lesion images as benign or malignant. This is done by extracting deep features using CNNs and using them in machine learning models for classification.

- Output of Feature Extraction:
  EfficientNetB3 gives 1536 features per image
  ResNet50 gives 2048 features per image
  Combined total: 3584 features per image

FEATURE SELECTION

- The project uses **SelectKBest with mutual information** as the scoring method to select the top 300 most relevant features from the combined feature set.
- After extracting deep features from images, feature selection is applied to keep only the most important ones, which helps improve model performance and reduce complexity.
- The final selected feature vector has 300 features per image.

SUPPORT VECTOR MACHINE

- **Train SVM Classifier**
  Fit the Support Vector Machine with RBF kernel on tabular data.
  Enable probability=True to allow AUC/ROC analysis later.

- **Predict on Test Set**
  Use the trained SVM model to predict class labels on X_test.
  This gives a direct binary prediction: benign or malignant.

- **Evaluate Classification Metrics**
  Print accuracy, confusion matrix, and classification report.
  Visualize class performance using confusion matrix plot.

RANDOM FOREST

- Train Random Forest Model
  Fit RandomForestClassifier on tabular data using X_train and y_train.
  Uses 100 trees to learn patterns for classifying benign vs malignant.

- Predict on Test Data
  Generate predicted class labels using the trained random forest model.
  These predictions help assess model generalization on unseen data.

- Evaluate Classification Results
  Print accuracy, classification report, and confusion matrix.
  Visualize confusion matrix to observe class-wise performance.

CONVOLUTIONAL NEURAL NETWORK

- **Load and Prepare Data**
  Load and preprocess train/test images using EfficientNetB3's preprocessing.
  Convert labels to one-hot format and compute class weights for balance.

- **Build and Train CNN**
  Freeze base layers and build custom classifier on top of EfficientNetB3.
  Train with image augmentation, callbacks, and class weights for 25 epochs.

- **Fine-Tune EfficientNetB3**
  Unfreeze last 30 layers of EfficientNetB3 for better feature learning.
  Retrain with lower learning rate to refine weights on domain-specific images.

ENSEMBLE LEARNING

- **Get Base Model Probabilities**
  Extract predicted probabilities for malignant class (1).
  Use SVM, RF on tabular data and CNN on image data (X_test_e).
- **Create Meta-Feature Set**
  Stack the three model outputs into X_meta.
  Each row has 3 features representing the models' predicted probabilities.
- **Train Gradient Boosting Classifier**
  Fit GradientBoostingClassifier using X_meta and y_test.
  Set key hyperparameters: 200 trees, learning rate 0.05, depth 3.

RESULT AND ANALYSIS

Model	Accuracy(%)	F1-Score	Precision	Recall
SVM	74.85	0.75	0.75	0.74
RF	**81.82**	**0.82**	**0.82**	**0.82**
CNN	55.00	0.82	0.82	0.82

[Table 1: Before applying the feature extraction and selection]

CONT..

Model	Accuracy(%)	F1-Score	Precision	Recall
SVM	**87.88**	**0.88**	**0.88**	**0.88**
RF	86.52	0.87	0.87	0.87
CNN	84.09	0.84	0.84	0.84

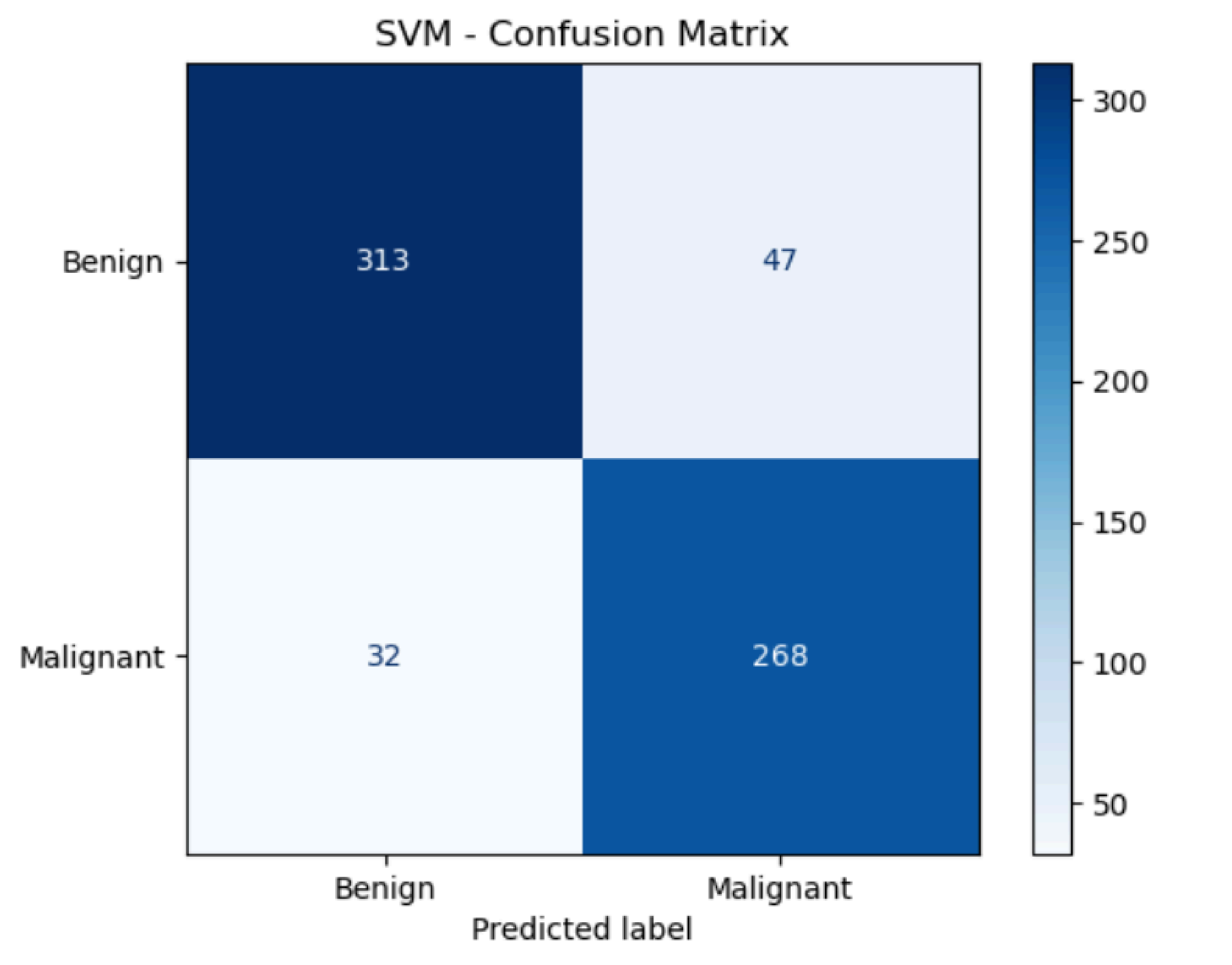
[Table 2: After applying the feature extraction and selection]

CONT..

Model	Accuracy(%)	F1-Score	Precision	Recall
Ensemble Learning	95.91	0.96	0.96	0.96

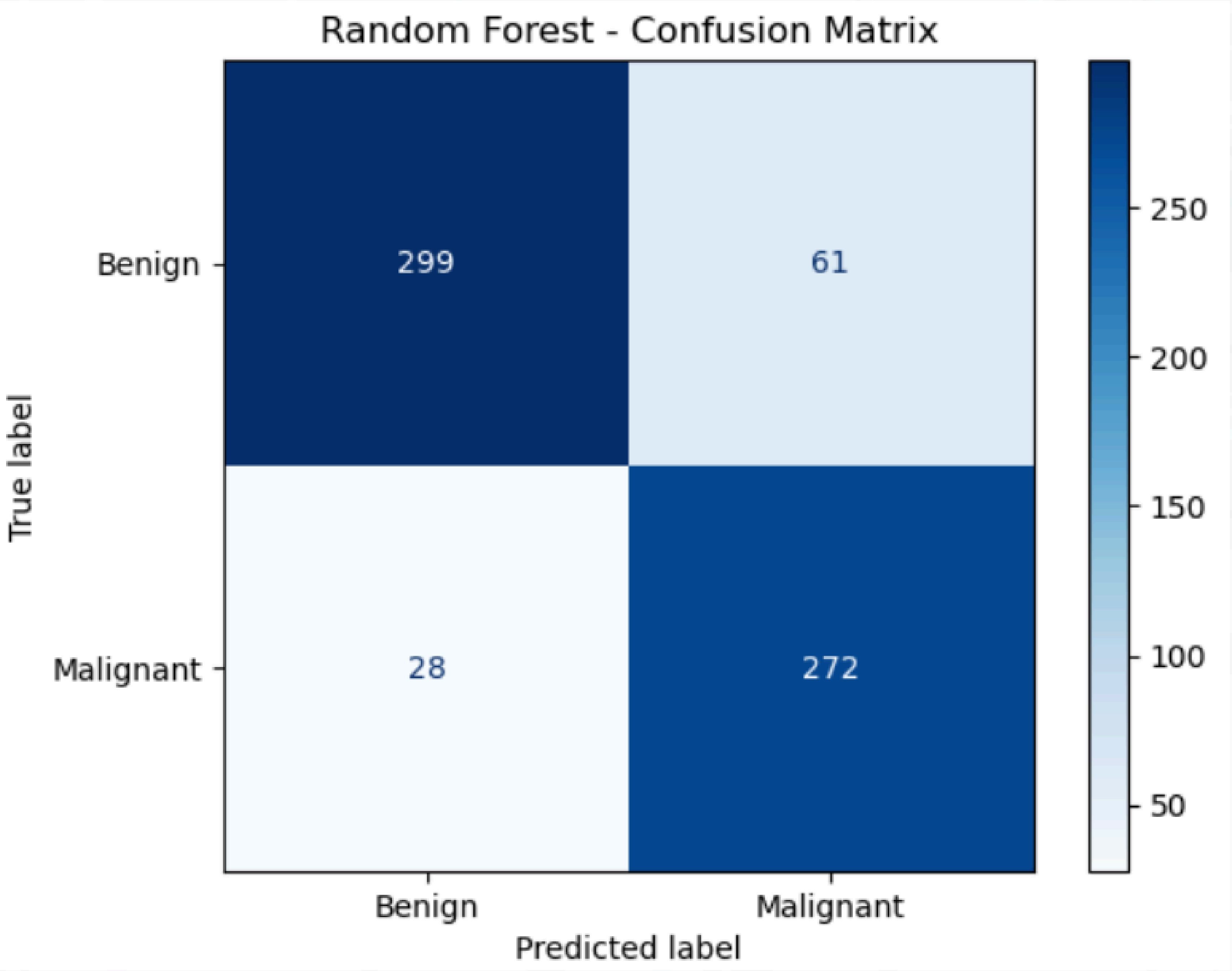
[Table 3: After applying Ensemble Learning]

CONT..

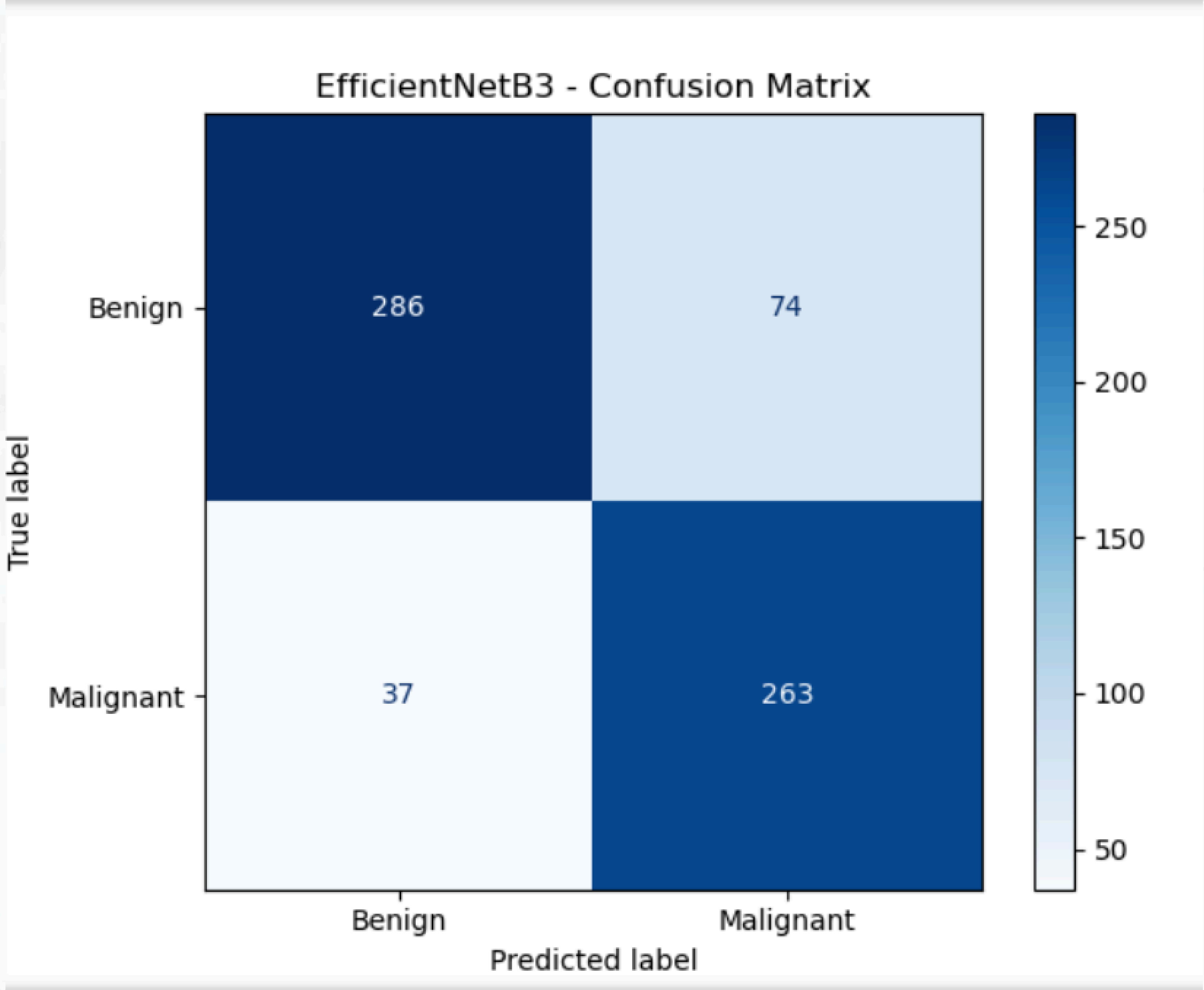


$$\begin{aligned} precision &= \frac{TP}{TP+FP} \\ recall &= \frac{TP}{TP+FN} \\ F1 &= \frac{2 \times precision \times recall}{precision + recall} \\ accuracy &= \frac{TP+TN}{TP+FN+TN+FP} \end{aligned}$$

CONT..

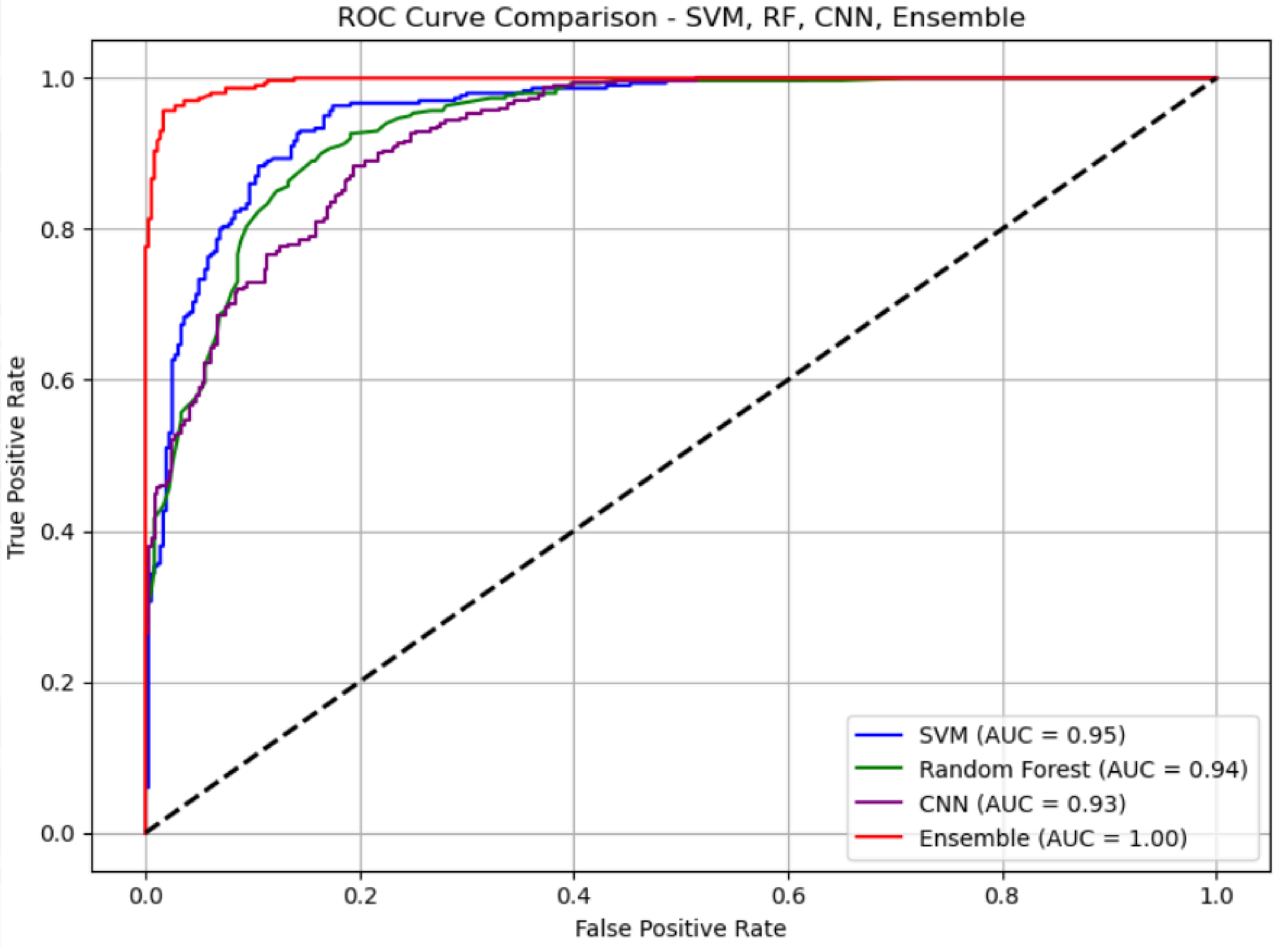


Random Forest - Confusion Matrix
True label
Predicted label
Benign
Malignant
299
61
28
272
250
200
150
100
50



EfficientNetB3 - Confusion Matrix
True label
Predicted label
Benign
Malignant
286
74
37
263
250
200
150
100
50

CONT..


ROC Curve Comparison - SVM, RF, CNN, Ensemble
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
SVM (AUC = 0.95)
Random Forest (AUC = 0.94)
CNN (AUC = 0.93)
Ensemble (AUC = 1.00)

CONCLUSION

- The project successfully uses machine learning techniques for early skin cancer detection.
- SVM, Random Forest, and CNN models were applied to classify skin lesions as malignant or benign.
- Ensemble techniques were used improve overall accuracy.
- The system supports faster, more accurate diagnosis and can help doctors make better decisions, potentially saving lives.

REFERANCE

- Vidya, M., & Karki, M. V. (2020, July). Skin cancer detection using machine learning techniques. In 2020 IEEE international conference on electronics, computing and communication technologies (CONECCT) (pp. 1-5). IEEE.
- Vijayalakshmi, M. M. (2019). Melanoma skin cancer detection using image processing and machine learning. International Journal of Trend in Scientific Research and Development (IJTSRD), 3(4), 780-784.
- Tembhurne, Jitendra V., Nachiketa Hebbar, Hemprasad Y. Patil, and Tausif Diwan. "Skin cancer detection using ensemble of machine learning and deep learning techniques." Multimedia Tools and Applications 82, no. 18 (2023): 27501-27524.
- https://www.kaggle.com

Thank you!