LEAFDETECTAI: EVALUATING ML AND DL PARADIGMS FOR SUPERIOR LEAF DISEASE IDENTIFICATION

GUDUR SUDHAKARA SANTHOSH

VIDEO PRESENTATION

MARCH 2024

WHY IS EARLY DETECTION IN PLANT HEALTH CRUCIAL?

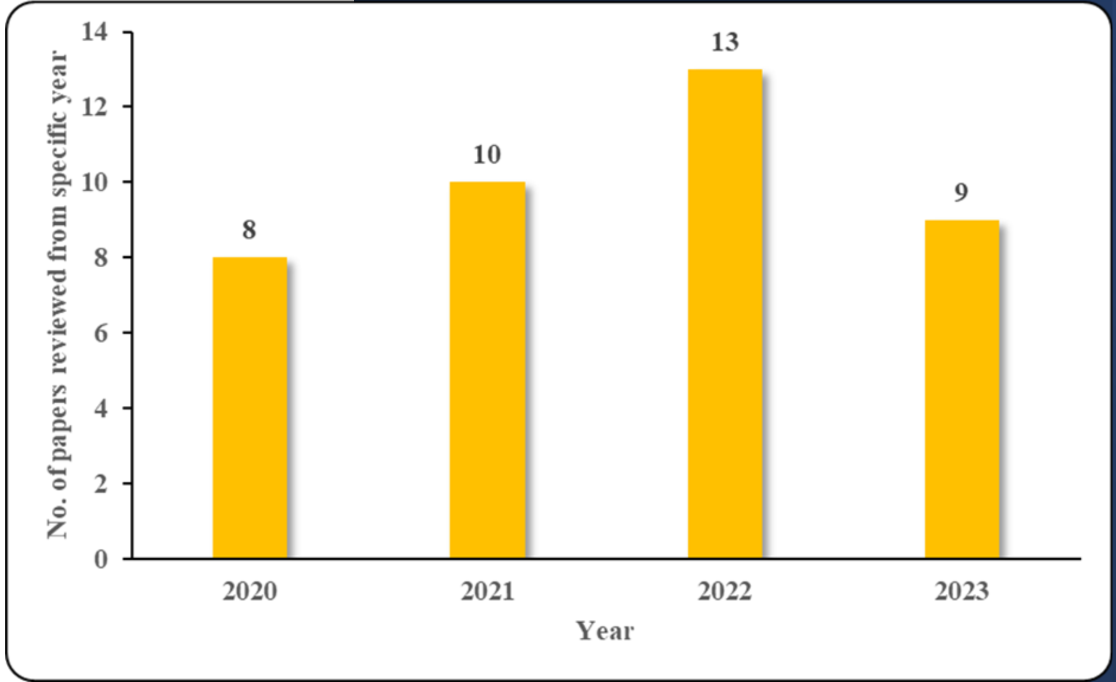
Early Detection Saves Crops

Secures Global Food Supply

AI Transforms Detection

CHALLENGES IDENTIFIED IN LITERATURE REVIEW

- **Symptom Similarity**: Difficult to differentiate.
- **Data Diversity**: Requires broader datasets.
- **Hardware Limits**: Constrains DL model training.
- **Data Quality**: Affects detection accuracy.
- **Diverse Symptoms**: Challenges robust modeling.
- **Preprocessing Hurdles**: Impacts image analysis.
- **Advanced Methods**: Need for novel algorithms



TACKLING PLANT DISEASE DETECTION CHALLENGES

Problem Statement:

- **Inadequate Traditional Methods**: Slow and error-prone.
- **Delayed Diagnoses**: Leads to yield reduction.
- **Economic & Food Security Risks**: From inaccurate detection.

Solution:

- **Leveraging AI**: Introducing ML and DL methodologies.
- **Enhanced Accuracy & Speed**: For real-time disease identification.
- **Sustainable Farming**: Contributing to global food security.

DATASET AND DISEASE FOCUS

- **Dataset:** PlantVillage
- **Dataset Link:** https://github.com/spMohanty/PlantVillage-Dataset
- **Diverse Species**: 14 plants, comprehensive coverage.
- **Multiple Diseases**: 17 fungal, 4 bacterial, 2 mold, 2 viral, 1 disease caused by mite.
- **Apple Leaf Diseases**: Focused on Black Rot, Cedar Apple Rust, and Apple Scab diseases.
- **Balanced Data**: Diseased vs. Healthy apple leaves.

EVALUATING CUTTING-EDGE MODELS IN AGRICULTURE

Machine Learning Models:

- Logistic Regression (LR)
- Linear Discriminant Analysis (LDA)
- K-Nearest Neighbors (KNN)
- Classification and Regression Trees (CART)
- Random Forest (RF)
- Gaussian Naive Bayes (NB)
- Support Vector Machine (SVM)

Deep Learning CNN Models:

- MobileNetV2
- NASNetMobile

REFINING DATA FOR PRECISION AGRICULTURE

Data Labelling and Encoding: Label encoding for ML/DL compatibility.

Dataset Split: 70% Training and 30% Testing for model rigor.

Cross-validation for ML/DL Models: Stratified K-fold for unbiased evaluation.

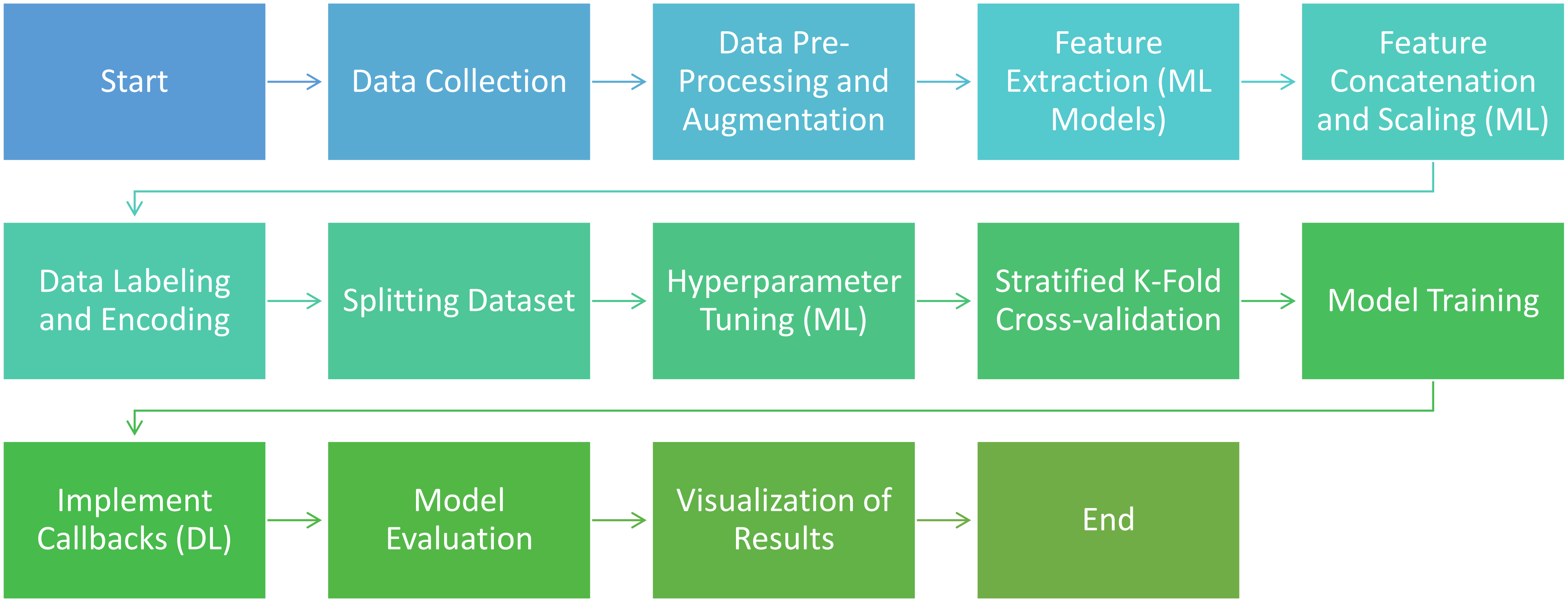
FEATURE TRANSFORMATION & NORMALIZATION IN ML

- **Resize**: Resize images to 500x500 for uniformity
- **Convert**: Convert colour spaces: RGB to BGR, HSV
- **Apply**: Apply segmentation for feature emphasis
- **Extract**: Extract crucial shape, texture, colour features
- **Utilize**: Utilize Hu Moments, Haralick, colour histograms
- **Normalize**: Normalize via MinMaxScaler for feature equity

BOOSTING DL MODELS: AUGMENTATION & TRANSFER LEARNING

- Augment data: rotations, shifts, flips
- Enhance volume and variability
- Reduce overfitting, improve generalization
- Use MobileNetV2, NASNetMobile architectures
- Adopt ImageNet pre-trained models
- Stratified K-Fold Cross Validation

END-TO-END FLOW FOR ML AND DL MODEL EVALUATION



EVALUATION METRICS FOR MODEL PERFORMANCE

Precision: Positive predictive value

Recall: Sensitivity to condition

F1-Score: Harmonic mean of precision, recall

Accuracy: Correct predictions overall

ROC Curve: Trade-off between true positive, and false positive rates

AUC: Area under ROC, measures discriminability

Confusion Matrix: Visualizes model performance

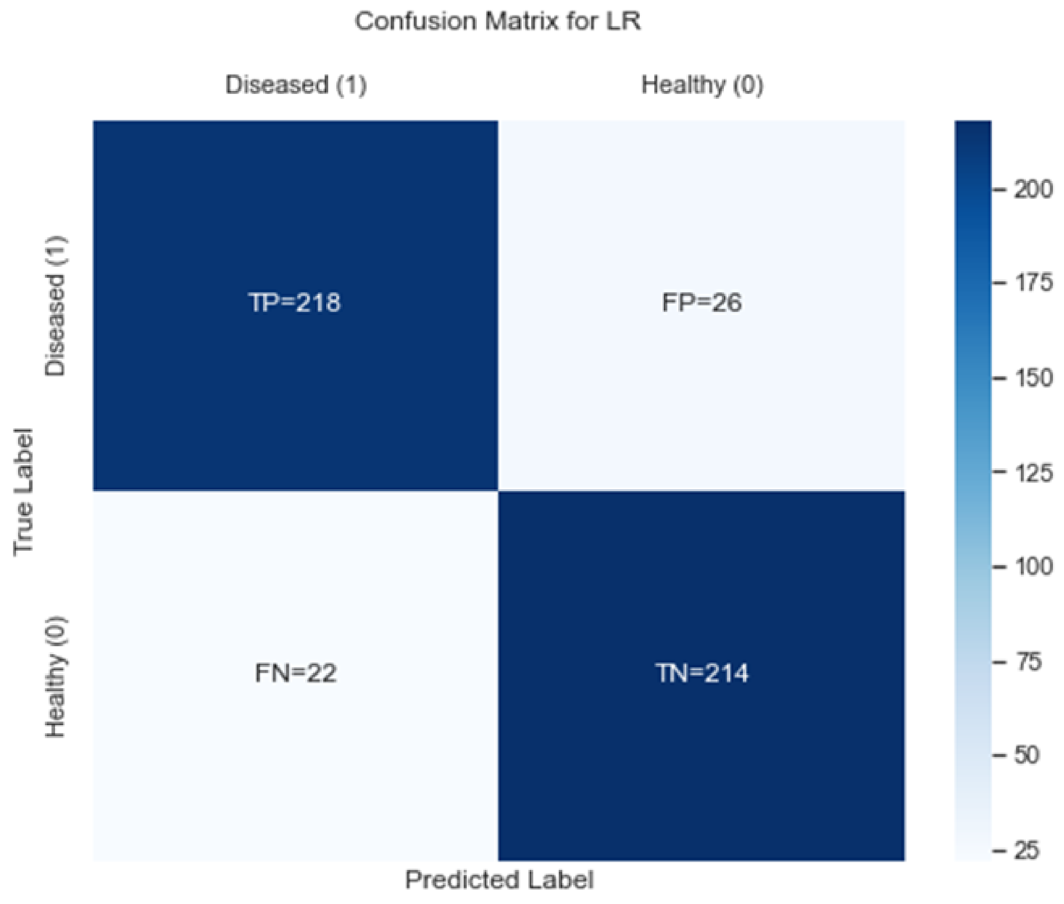
DESIGN & PERFORMANCE ANALYSIS OF LOGISTIC REGRESSION MODEL

Design Highlights:

- Initialized with scikit-learn **LogisticRegression** class, ensuring robustness and reproducibility
- **random_state** set for consistent optimization outcomes

Performance Analysis:

- **High Precision**: 91% for diseased class
- **Balanced metrics**: Precision, Recall, F1-score
- **Accuracy**: Strong at 90% overall
- **Recall**: 89% for diseased, 91% for healthy
- **F1-Score**: 90% for both classes
- Refinement needed for False Positives/Negatives
- Promising reliable classifier for future improvement



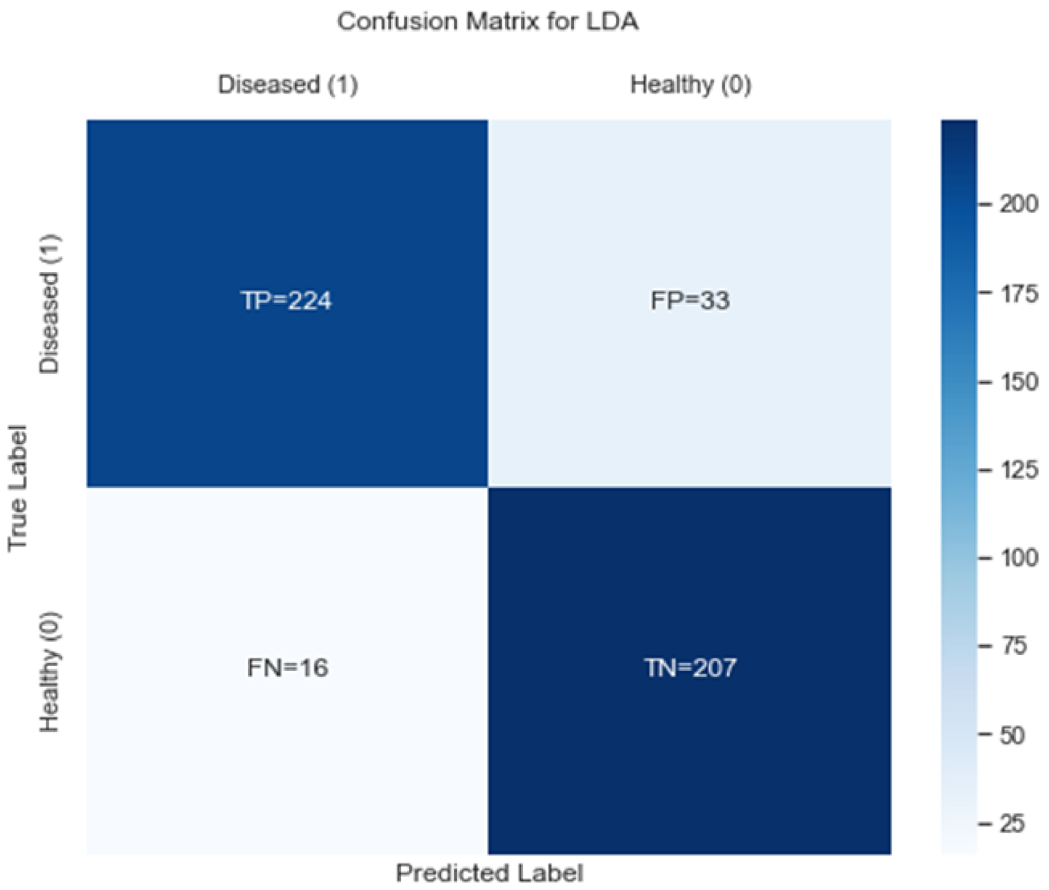
DESIGN & PERFORMANCE ANALYSIS OF LINEAR DISCRIMINANT ANALYSIS MODEL

Design Highlights:

- Employs scikit-learn **LinearDiscriminantAnalysis** for reliability
- Harnesses LDA's statistical efficiency in distinguishing classes
- Utilizes default settings to leverage inherent model strengths

Performance Analysis:

- **High Precision:** 93% for diseased class detection
- **Superior Recall:** 93% for healthy classification
- **Good Accuracy:** Overall, 89.79% performance
- **Balanced F1-Score:** 89% diseased, 90% healthy
- **Recall for Diseased:** 86%, indicates improvement area
- FP and FN: 33 and 16 cases, need refinement
- LDA proves useful, requires further tuning



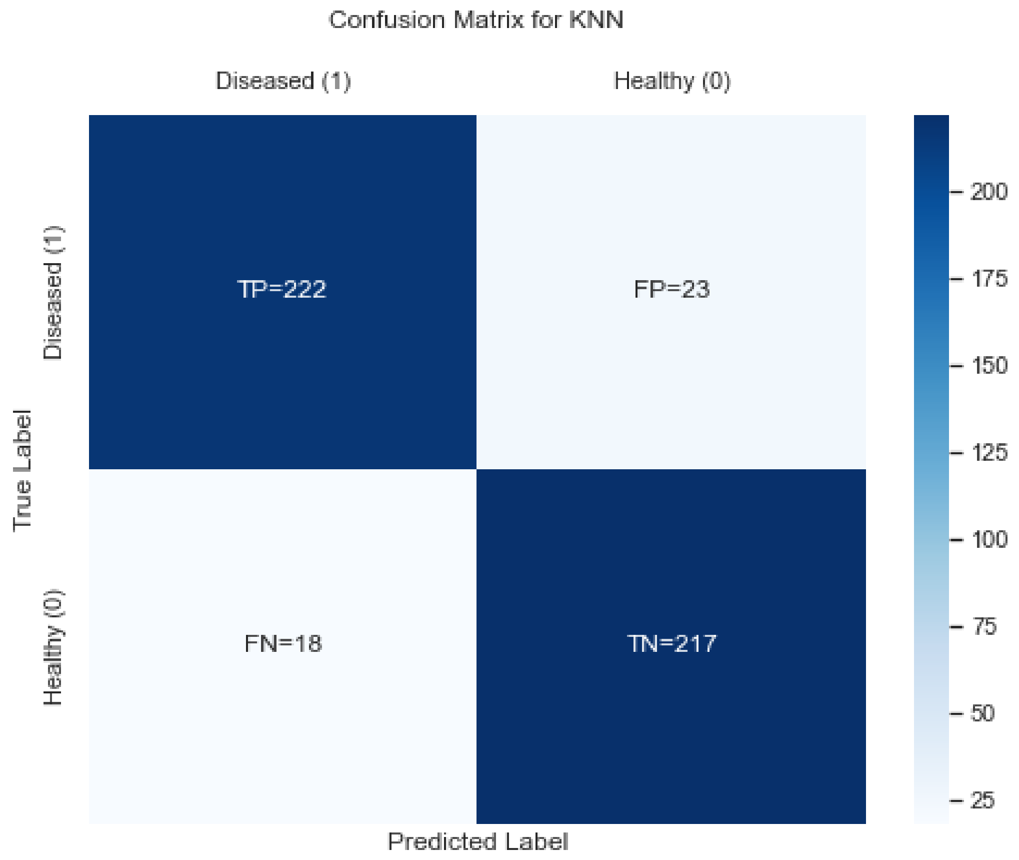
DESIGN & PERFORMANCE ANALYSIS OF K-NEAREST NEIGHBORS MODEL

Design Highlights:

- Implements K-Nearest Neighbors via scikit-learn **KNeighborsClassifier**
- Default settings used, including 5 neighbors and Euclidean distance

Performance Analysis:

- **High Precision:** 92% for diseased leaves
- **Strong Recall:** 90% diseased, 93% healthy
- **Impressive Accuracy:** Overall, 91.46%
- **F1-Score:** 91% diseased, 92% healthy
- **Low False Positives:** Only 23 cases
- **Few False Negatives:** 18 missed instances
- KNN is reliable but needs further tuning



set for consistent and reproducible results
behavior through controlled randomness in feature selection

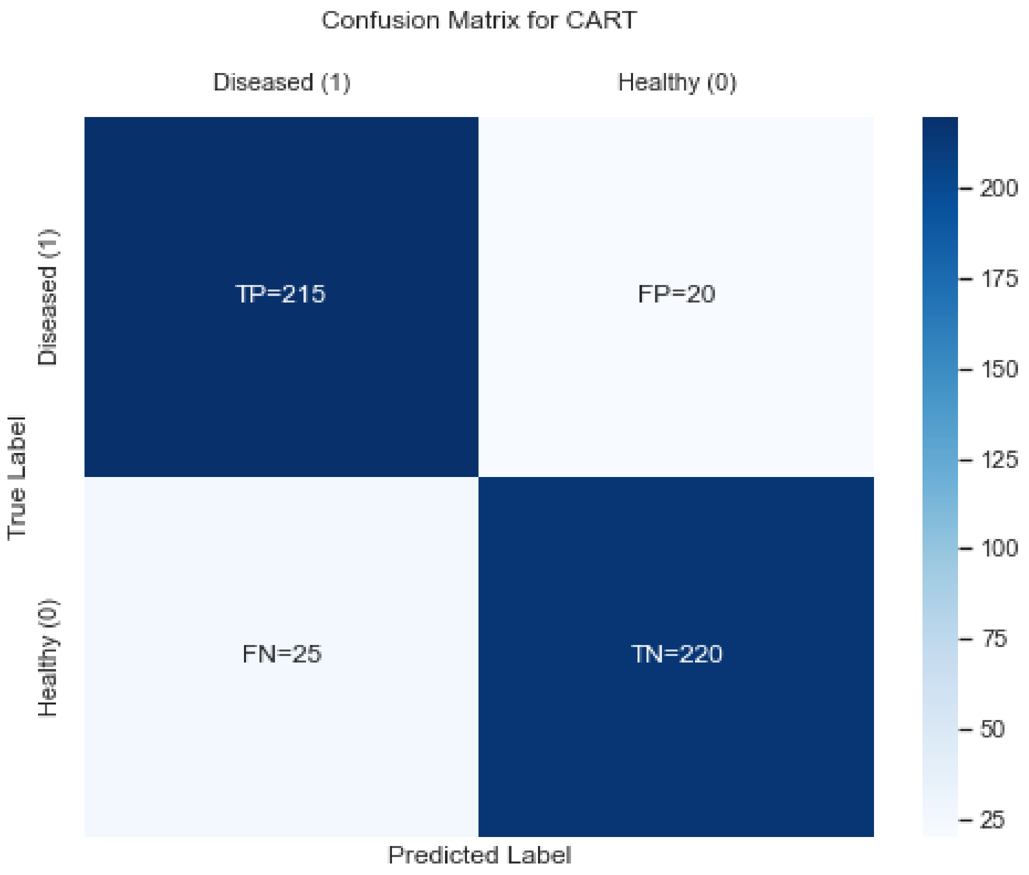
DESIGN & PERFORMANCE ANALYSIS OF CLASSIFICATION AND REGRESSION TREE MODEL

Design Highlights:

- Utilizes scikit-learn **DecisionTreeClassifier** for the CART model
- **random_state** set for consistent and reproducible results
- Deterministic behavior through controlled randomness in feature selection

Performance Analysis:

- **Balanced Classification:** Precision at 90%
- **High Recall for Diseased:** 92% effectiveness
- **Notable Accuracy:** Overall, at 90.62%
- **Equitable F1-Score:** 91% for both classes
- **Low Misclassification:** 20 False Positives
- **Reduced False Negatives:** Only 25 cases
- Effective yet requires further optimization



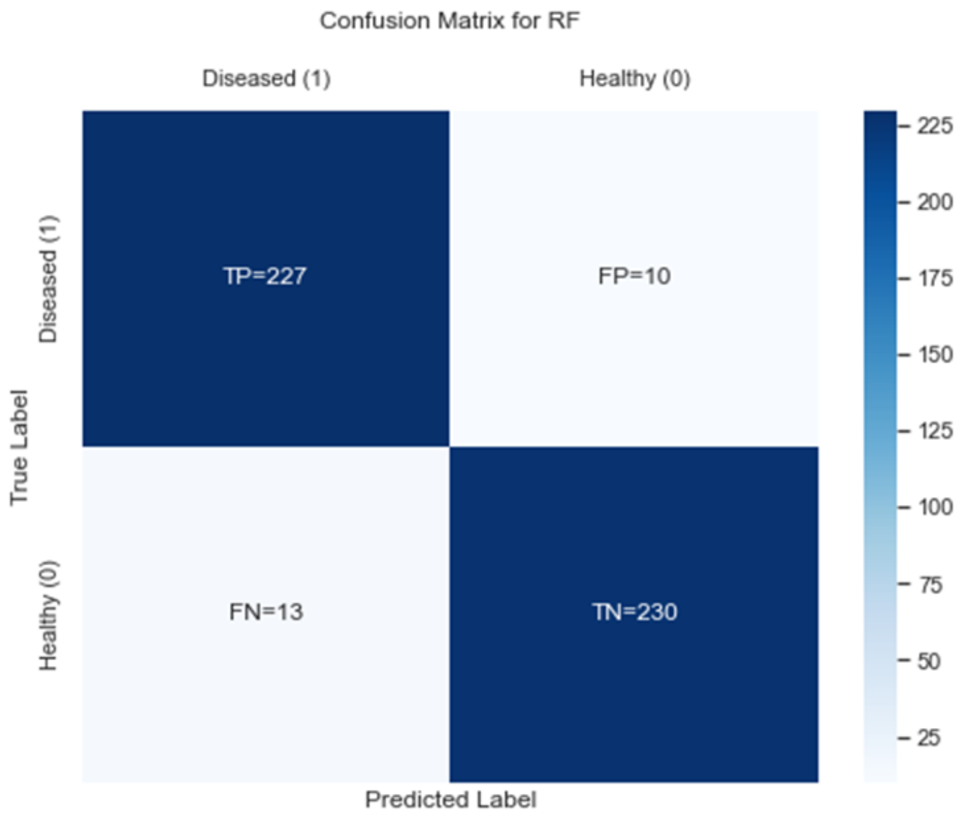
DESIGN AND PERFORMANCE ANALYSIS OF RANDOM FOREST MODEL

Design:

- Adopts RF's ensemble learning for complex pattern recognition
- Utilizes GridSearchCV for optimal hyperparameter tuning

Performance Analysis:

- **High Precision/Recall**: 95% & 96% respectively
- **Exceptional Accuracy**: Overall, at 95.21%
- **F1-Score**: Robust at 95% for both classes
- **Very Low Misclassifications**: FP=10, FN=13
- **Reliable Class Discrimination**: High consistency
- **RF Complexity**: Consider computational needs
- **Interpretability**: Less transparent decision process



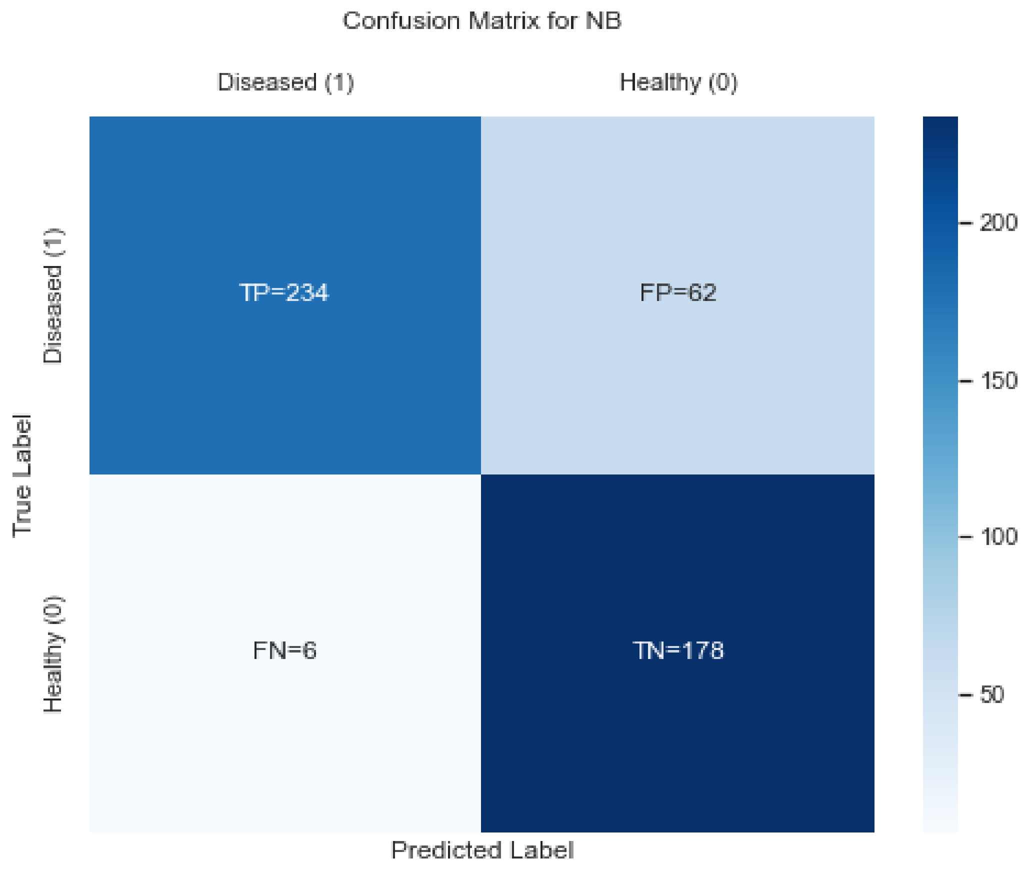
DESIGN AND PERFORMANCE ANALYSIS OF GAUSSIAN NAÏVE BAYES MODEL

Design:

- Adopts scikit-learn **GaussianNB** for its simplicity and efficiency
- Default parameterization leverages data-driven parameter estimation
- The probabilistic approach suits biological data's inherent variability

Performance Analysis:

- **Strong Detection:** High true positive rate
- **Effective Specificity:** Solid true negative count
- **Notable Accuracy:** 85.83% overall performance
- **False Positives:** Number at 62, indicates over-diagnosis
- **Precision/Recall Gap:** High precision, lower recall
- **F1-Score:** 84% diseased, 87% healthy
- **Naive Bayes:** Powerful, yet needs refinement



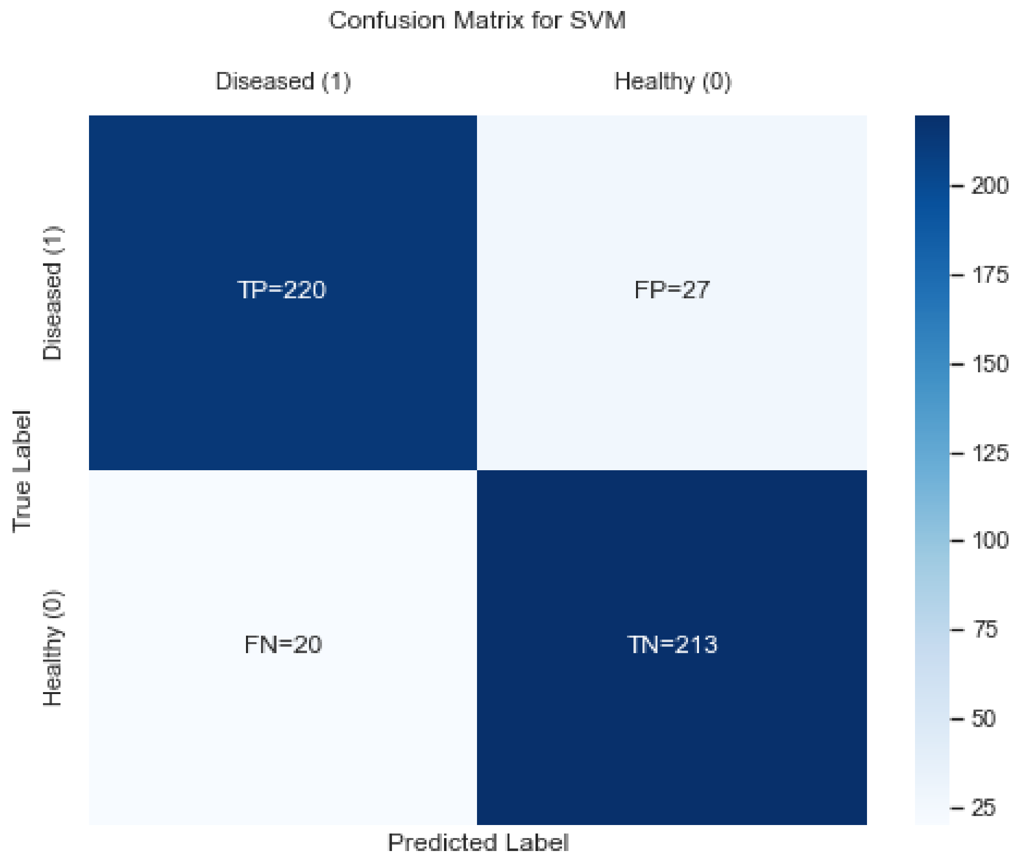
DESIGN AND PERFORMANCE ANALYSIS OF SUPPORT VECTOR MACHINE MODEL

Design:

- Employs scikit-learn **SVC** class, tailored for classification
- Probability estimation enabled for prediction confidence
- **random_state** ensures reproducible and consistent training

Performance Analysis:

- **Precision**: 91% diseased, 89% healthy
- **True Positives**: 220, reliable disease detection
- **Accuracy**: Solid at 90.21%
- **True Negatives**: 213, accurate healthy classification
- **Manageable False Positives/Negatives**: 27/20
- **F1-Score**: Equally strong at 90%
- **SVM**: Robust yet requires careful tuning



COMPARATIVE ANALYSIS OF ML MODELS FOR APPLE LEAF DISEASE DETECTION

Model	Accuracy	Precision (Diseased)	Recall (Diseased)	Key Points
LR	90.00%	0.91	0.89	Reliable in positive identification, slight recall disparity
LDA	89.79%	0.93	0.86	High precision, improvement needed in diseased recall
KNN	91.46%	0.92	0.90	Effective identification with low false positives
CART	90.62%	0.90	0.92	Balanced classification, minor misclassification errors
RF	95.21%	0.95	0.96	Superior accuracy, excellent class differentiation
NB	85.83%	0.97	0.74	High true positive rate, high false positive rate
SVM	90.21%	0.91	0.89	Balanced classification, scope for misclassification reduction

MobileNetV2 Architecture for Disease Detection

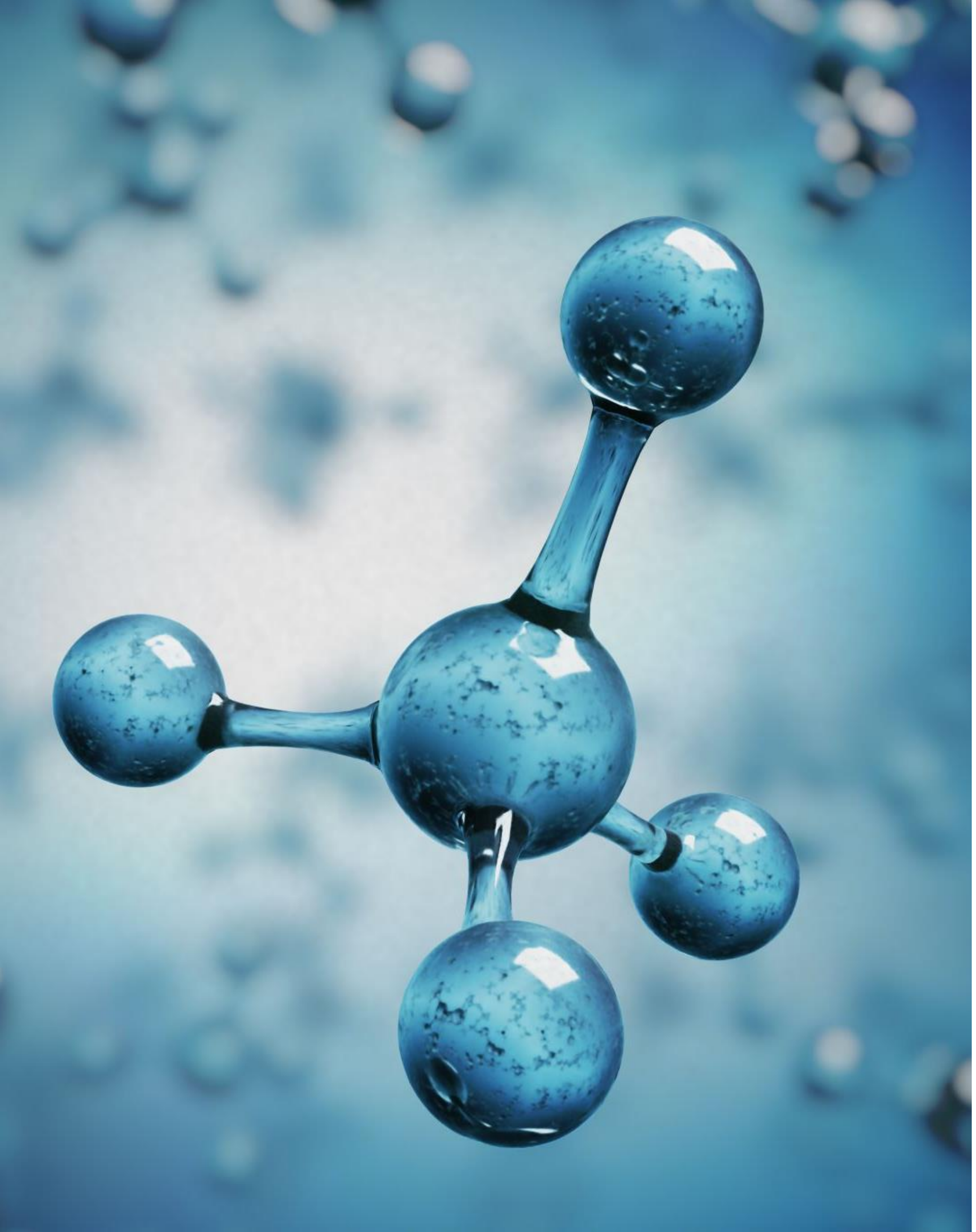
- Built on efficient MobileNetV2 architecture
- Non-trainable base, fine-tuned for disease detection
- Global Average Pooling, Dense, and Dropout layers
- Optimized with Adam, EarlyStopping, ReduceLROnPlateau
- High validation accuracy, precision, recall
- Evaluated using precision-recall and ROC-AUC curves
- Ideal for real-time, resource-constrained environments

NASNetMobile Architecture in Disease Detection

- Utilizes advanced NASNetMobile architecture
- Integrates Global Pooling, Dense, Dropout layers
- Tailored and optimized for precise disease identification
- Employs Adam optimizer with adaptive learning
- Exhibits high validation accuracy and metrics
- Analyzed through precision-recall, ROC-AUC curves
- Suitable for real-time analysis in diverse conditions

COMPARATIVE PERFORMANCE OF DL MODELS FOR APPLE LEAF DISEASE DETECTION

Feature	NASNetMobile	MobileNetV2	Remarks
Design	Advanced CNN	Advanced CNN	Both models use sophisticated CNN designs for high-dimensional data processing.
Accuracy	Overall accuracy ranges from 92% to 94%	Often nearing or achieving 98%	MobileNetV2 demonstrates superior accuracy, making it more reliable in distinguishing healthy from diseased plants.
Precision (Healthy)	Up to 0.96	Consistently >= 0.98	MobileNetV2 shows higher precision, indicating fewer false positives in healthy plant detection.
Precision (Diseased)	Up to 0.92	Consistently >= 0.98	MobileNetV2's precision for diseased plants suggests high reliability in disease detection.
Recall (Healthy)	Up to 0.91	Consistently >= 0.98	MobileNetV2 has a higher recall for healthy plants, ensuring fewer false negatives.
Recall (Diseased)	Up to 0.96	Consistently >= 0.98	MobileNetV2 demonstrates higher recall for diseased plants, crucial for early disease detection.
Diagnostic Tools	Confusion matrices, ROC-AUC, Precision-Recall curves	Confusion matrices, ROC-AUC, Precision-Recall curves	Both models utilize comprehensive diagnostic tools to assess performance.
Preferred Model		MobileNetV2	MobileNetV2 is preferred for its consistently superior performance across key metrics.



BEST MODEL FOR APPLE LEAF DISEASE DETECTION

- **RF**: Top ML model, 95.21% accuracy
- **MobileNetV2**: Best DL model, ~98% metrics
- **Precision & Recall**: DL model surpasses ML
- **RF vs. MobileNetV2**: DL offers slight edge
- Deep architecture aids complex feature extraction
- MobileNetV2 balances efficiency, performance
- MobileNetV2 recommended for practical use

ADVANCING APPLE LEAF DISEASE DETECTION: LIMITATIONS AND FUTURE DIRECTIONS

Expand	Expand dataset diversity for model robustness
Invest in	Invest in high-quality data annotation
Optimize	Optimize RF model through hyperparameter tuning
Develop	Develop hybrid ML and DL models
Improve	Improve computational efficiency for DL
Pilot	Pilot real-world deployment and user-centric design
Encourage	Encourage continuous learning and open-source collaboration

ACKNOWLEDGEMENTS

- Deep gratitude to supervisor **Mr. Jignesh Patel**
- Immense thanks to **Prof. Dr. Manoj J** and **UpGrad**
- Heartfelt appreciation to parents and family
- Tribute to all contributors of my academic journey
- Collective efforts recognized, eternally thankful