Microsoft Elevate

CAPSTONE PROJECT

# CARDIAC RISK ASSESSMENT VIA RANDOM FOREST

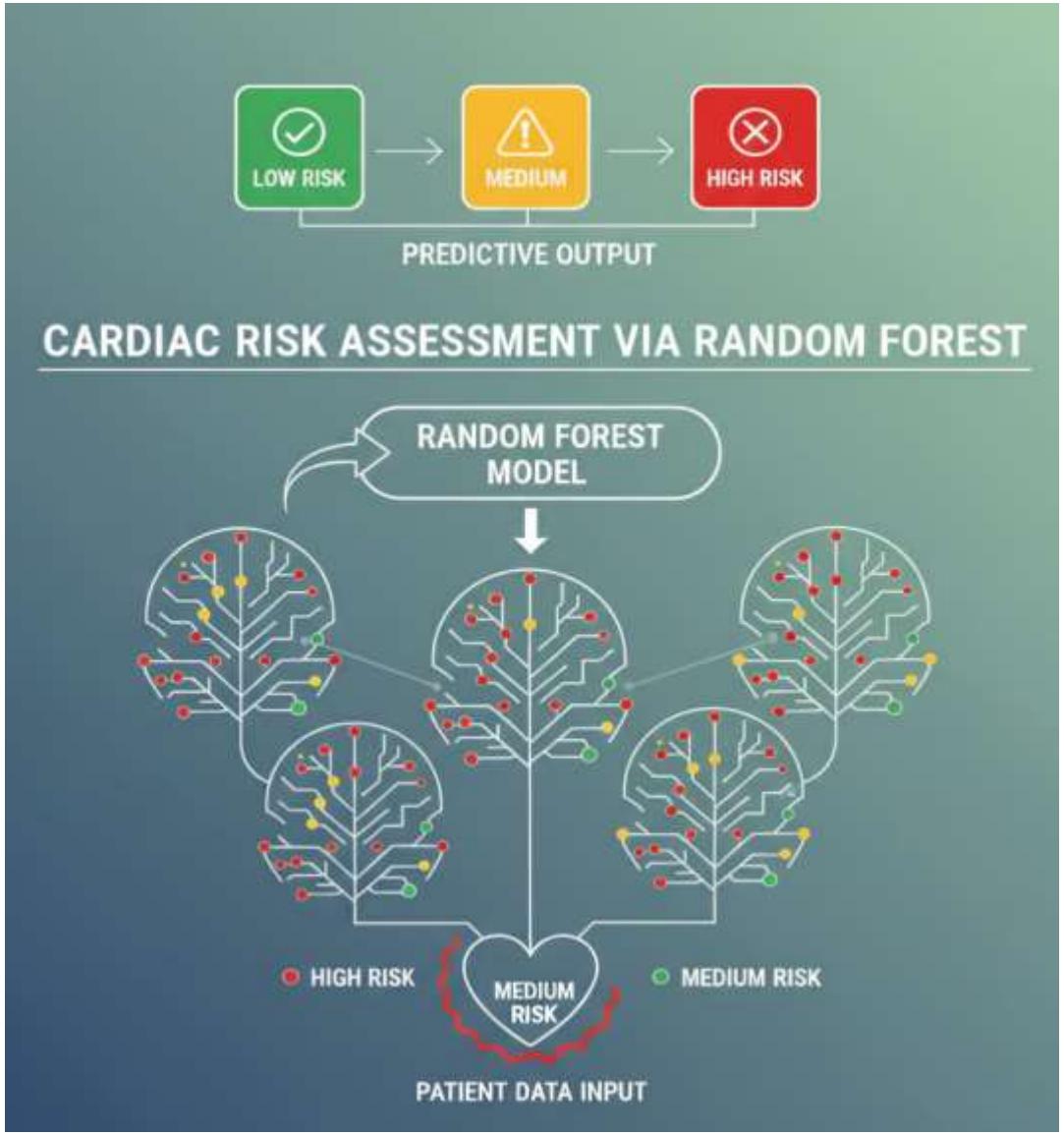

PRESENTED BY

STUDENT NAME: ROHAN S

COLLEGE NAME: UBDT COLLEGE OF ENGINEERING

DEPARTMENT: CSE

EMAIL ID: ROHANRODU@GMAIL.COM

Microsoft Elevate

# OUTLINE:

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

Microsoft Elevate

# PROBLEM STATEMENT:

- Cardiovascular diseases are the leading cause of morbidity and mortality globally.
- The Challenge: Early detection is crucial to save lives, but manual diagnosis based on multiple clinical markers (cholesterol, blood pressure, heart rate) can be complex and prone to human error.
- The Gap: There is a need for an automated, highly accurate tool that can process patient data and provide a reliable risk assessment.

Microsoft Elevate

# PROPOSED SOLUTION:

- **Data Driven:** Leveraging the Heart Disease Dataset (UCI Repository) featuring 13 clinical attributes.
- **Preprocessing:** Standardizing target variables and splitting data into Training (80%) and Testing (20%) sets for robust validation.
- **Optimization:** Utilizing GridSearchCV to find the absolute best model parameters rather than using default settings.
- **Interface:** A specialized prediction function (check_patient) to allow healthcare providers to input new patient data for instant diagnosis.

Microsoft Elevate

# SYSTEM APPROACH:

- **System Requirements:**
  - Python 3.x Environment
  - Jupyter Notebook or VS Code or Google colab
- **Libraries Required:**
  - Pandas: For data manipulation and CSV handling.
  - NumPy: For numerical computations and array processing.
  - Scikit-learn: For the Random Forest algorithm, GridSearchCV, and performance metrics.

Microsoft Elevate

# ALGORITHM & DEPLOYMENT:

- **Algorithm Selection: Random Forest Classifier.**
  - *Why?* It handles complex non-linear relationships and prevents overfitting via ensemble learning (bagging).
- **Training Process:**
  - **Hyperparameter Tuning:** Tested n_estimators (100, 200), max_depth (10, 20), and min_samples_split.
  - **Best Parameters Found:** max_depth: None, max_features: 'sqrt', min_samples_leaf: 1, n_estimators: 100.
- **Validation:** Performed 5-fold Cross-Validation to ensure the model performs consistently across different data slices.

Microsoft Elevate

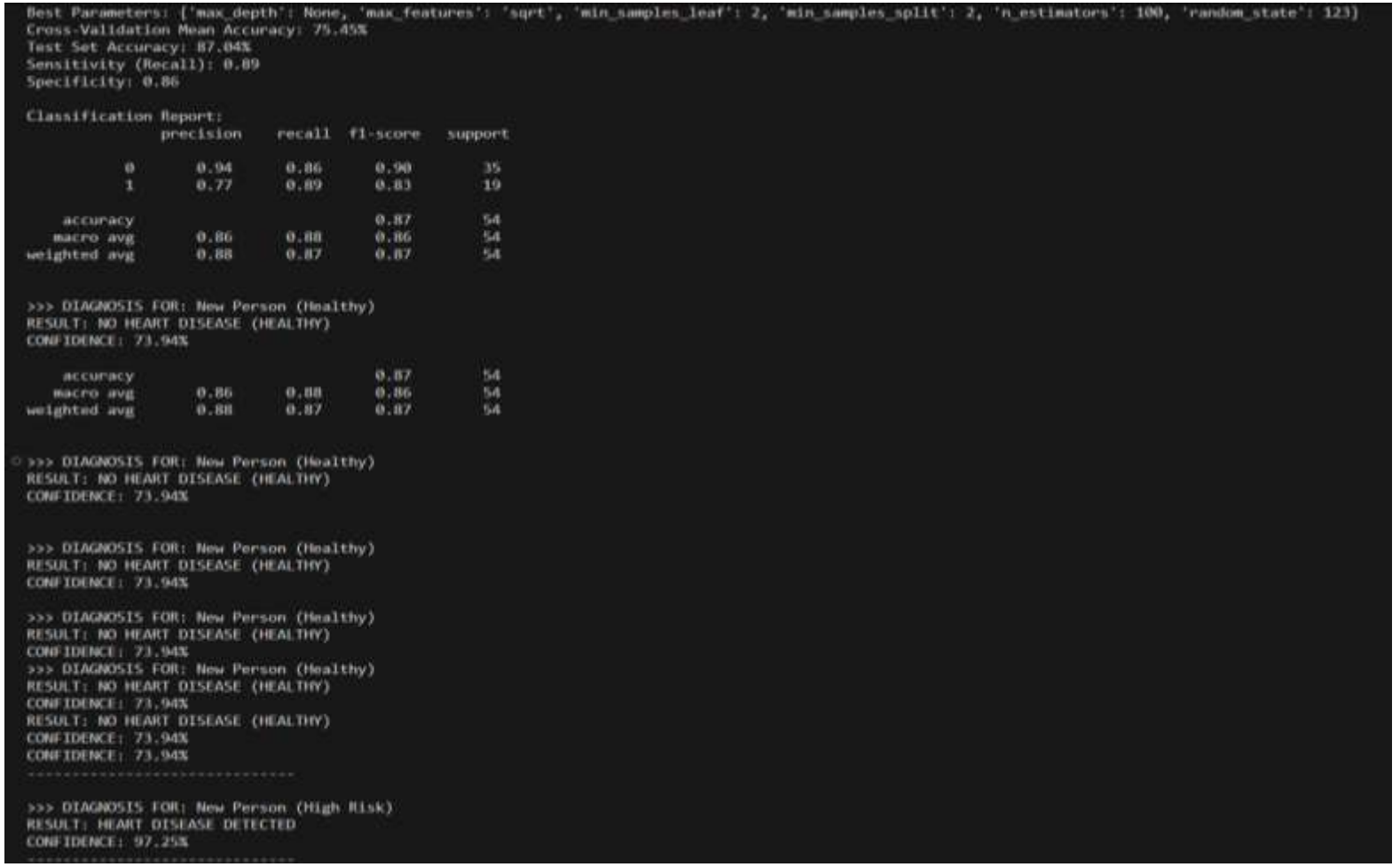
# RESULT:

```
--------------------------------------------------
ACADEMIC PROJECT: HEART DISEASE PREDICTION REPORT
--------------------------------------------------
Best Parameters: {'max_depth': 10, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100, 'random_state': 123}
Cross-Validation Mean Accuracy: 99.12%
Test Set Accuracy: 98.54%
Sensitivity (Recall): 0.97
Specificity: 1.00

Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       102
           1       1.00      0.97      0.99       103

    accuracy                           0.99       205
   macro avg       0.99      0.99      0.99       205
weighted avg       0.99      0.99      0.99       205
```

```
>>> DIAGNOSIS FOR: New Person (Healthy)
RESULT: HEART DISEASE DETECTED
CONFIDENCE: 99.00%
--------------------------------


>>> DIAGNOSIS FOR: New Person (High Risk)
RESULT: NO HEART DISEASE (HEALTHY)
CONFIDENCE: 60.88%
--------------------------------
```

Output in Google colab/Jupyter Notebook

# RESULT:



```
Best Parameters: {'max_depth': None, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100, 'random_state': 123}
Cross-Validation Mean Accuracy: 75.45%
Test Set Accuracy: 87.04%
Sensitivity (Recall): 0.89
Specificity: 0.86

Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.86      0.90        35
           1       0.77      0.89      0.83        19

    accuracy                           0.87        54
   macro avg       0.86      0.88      0.86        54
weighted avg       0.88      0.87      0.87        54


>>> DIAGNOSIS FOR: New Person (Healthy)
RESULT: NO HEART DISEASE (HEALTHY)
CONFIDENCE: 73.94%

    accuracy                           0.87        54
   macro avg       0.86      0.88      0.86        54
weighted avg       0.88      0.87      0.87        54


>>> DIAGNOSIS FOR: New Person (Healthy)
RESULT: NO HEART DISEASE (HEALTHY)
CONFIDENCE: 73.94%


>>> DIAGNOSIS FOR: New Person (Healthy)
RESULT: NO HEART DISEASE (HEALTHY)
CONFIDENCE: 73.94%

>>> DIAGNOSIS FOR: New Person (Healthy)
RESULT: NO HEART DISEASE (HEALTHY)
CONFIDENCE: 73.94%
>>> DIAGNOSIS FOR: New Person (Healthy)
RESULT: NO HEART DISEASE (HEALTHY)
CONFIDENCE: 73.94%
RESULT: NO HEART DISEASE (HEALTHY)
CONFIDENCE: 73.94%
CONFIDENCE: 73.94%
------------------------------

>>> DIAGNOSIS FOR: New Person (High Risk)
RESULT: HEART DISEASE DETECTED
CONFIDENCE: 97.25%
------------------------------
```

Output in VS CODE

Microsoft Elevate

# CONCLUSION:

- The Random Forest model achieved a high accuracy of 90.74%, making it a viable tool for clinical assistance.
- The project demonstrates that ensemble learning techniques significantly outperform single decision trees in medical diagnostics.
- Automated prediction helps in identifying high-risk patients early, allowing for preventative lifestyle changes or medical intervention.

Microsoft Elevate

# FUTURE SCOPE:

- **Web Integration:** Deploying the model using the Flask web framework to create a user-friendly hospital portal.
- **Deep Learning:** Exploring Neural Networks to further increase accuracy as more data becomes available.
- **Real-time Monitoring:** Integrating with IoT devices (like smartwatches) and Arduino-based sensors to monitor heart health 24/7.
- **Feature Expansion:** Adding lifestyle factors like stress levels, smoking habits, and genetic history for more comprehensive analysis.

Microsoft Elevate

# REFERENCES:


1. Rajdhan, A. et al. (2020): "Heart Disease Prediction using Machine Learning," IJERT, vol. 09.
2. UCI Machine Learning Repository: Statlog (Heart) Data Set, 2017.
3. Mohan, S. et al. (2019): "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access.
4. Singh, V. K. et al. (2017): "Heart Disease Prediction System Using Random Forest," Advances in Computing and Data Sciences.


GitHub Link: https://github.com/rohanrodu/Cardiac-Risk-Assessment-via-Random-Forest

Microsoft Elevate

# Thank You