# Project Final Report

## Title:

**Hybrid Model to Predict Arrhythmia in Cancer Patients**

---

## Abstract:

This project aims to develop a hybrid machine learning model for early and accurate detection of arrhythmia in cancer patients using ECG data. Arrhythmia, a condition characterized by irregular heartbeats, poses significant health risks such as stroke or cardiac arrest. The proposed approach integrates traditional machine learning and neural network models with dimensionality reduction techniques to enhance accuracy, efficiency, and generalization. The model is trained and evaluated using a large dataset sourced from Kaggle, consisting of over 100,000 ECG signal records. Initial results demonstrate that dimensionality reduction significantly improves performance, with the K-Nearest Neighbors (KNN) algorithm combined with PCA achieving the highest accuracy.

---

## 1. Introduction

- **Background and Motivation:**
  Arrhythmia is a life-threatening condition that is especially dangerous in cancer patients due to weakened cardiovascular health. Early detection through automated means can prevent complications and improve outcomes.
- **Importance and Relevance:**
  Traditional manual ECG interpretation is time-consuming and error-prone. Machine learning provides a scalable, real-time diagnostic tool, especially beneficial in clinical environments.
- **Objectives and Research Questions:**
  - Can a hybrid ML model improve arrhythmia detection accuracy in cancer patients?
  - How do dimensionality reduction techniques influence model performance?
  - Which model architecture offers the best balance between performance and efficiency?

---

## 2. Problem Statement

- **Defined Problem:**
  Cancer patients are at increased risk for arrhythmia. Manual ECG analysis lacks scalability and speed, necessitating automated and accurate systems.

- **Challenges in Existing Solutions:**
    - High dimensionality and noise in ECG data.
    - Lack of generalization across models.
    - Performance degradation without appropriate feature engineering or dimensionality reduction.

---

## 3. Project Architecture

- **System Architecture Diagram:**
- **Components:**
    - Data Acquisition and Preprocessing
    - Dimensionality Reduction
    - Model Training (ML algorithms and Neural Networks)
    - Evaluation and Visualization
- **Data Flow and Processing Pipeline:**
  Raw ECG data → Feature Scaling → Dimensionality Reduction (PCA/KPCA/LDA) → Model Training → Evaluation → Prediction Output

---

## 4. Tools and Technologies

### Software

- **Programming Languages:** Python
- **Libraries and Frameworks:**
    - Scikit-learn (SVM, KNN, Naive Bayes, Logistic Regression, PCA)
    - NumPy, Pandas, Matplotlib
- **Development Environment:** Jupyter Notebook

### Hardware

- **CPU/GPU Requirements:** Standard CPU suffices (no GPU-intensive training involved)
- **Sensors/Edge Devices:** Not applicable (dataset-based project)

---

## 5. Methodology

- **Dataset Source:** Kaggle
- **Dataset Overview:**
    - Samples: 100,689
    - Features: 17 ECG-related and physiological metrics
    - Target: Binary classification (0 = No arrhythmia, 1 = Arrhythmia)

- **Preprocessing Techniques:**
    - Removed gender and age columns
    - Applied StandardScaler for normalization
    - Split: 40% training, 60% testing using `train_test_split(random_state=0)`
- **Dimensionality Reduction Methods:**
    - **PCA:** Linear technique capturing 95% variance
    - **Kernel PCA (RBF kernel):** For non-linear separation
    - **LDA:** Supervised method focusing on class separability
- **Model Architectures Used:**
    - **SVM:** Suitable for high-dimensional data
    - **KNN:** Performs well post dimensionality reduction
    - **Naive Bayes:** Fast, suitable for independent features
    - **Logistic Regression:** Baseline model with interpretability
    - **Random Forest:** Handles feature-rich data and prevents overfitting
    - **Neural Networks:** Captures non-linear patterns, ideal for large, complex datasets

---

## 6. Result

- **Best Performing Model:**
  **K-Nearest Neighbors (KNN)** combined with **PCA**
- **Performance Metrics:**
    - Accuracy: Highest among tested models(96.6%)
    - Improved runtime and efficiency due to dimensionality reduction
    - PCA enhanced distance metric reliability for KNN
- **Practical Applications:**
    - Can be integrated into clinical decision support systems
    - Assists physicians with rapid diagnosis and treatment planning
- **Energy Efficiency & Computational Metrics:**
    - Reduced computational complexity through PCA/KPCA/LDA
    - Lower memory and time requirements compared to raw high-dimensional input models

---

## 7. Discussion

- **Obstacles and Mitigation:**
    - **High-dimensional noise:** Mitigated through PCA/KPCA/LDA
    - **Model overfitting:** Addressed with regularization and reduced feature space
    - **Algorithm bias/sensitivity:** Cross-validated with multiple models
- **Challenges and Risks:**
    - Generalization to real-time or clinical data
    - Limited interpretability of neural networks
    - Dataset bias or lack of patient diversity

## 8. References


1. Cheung, Christopher C., Andrew D. Krahn, and Jason G. Andrade. "The emerging role of wearable technologies in detection of arrhythmia." *Canadian Journal of Cardiology* 34.8 (2018): 1083-1087.
2. Cheung CC, Krahn AD, Andrade JG. The emerging role of wearable technologies in detection of arrhythmia. Canadian Journal of Cardiology. 2018 Aug 1;34(8):1083-7.
3. Yao Q, Wang R, Fan X, Liu J, Li Y. Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network. Information Fusion. 2020 Jan 1;53:174-82.
4. Yao, Q., Wang, R., Fan, X., Liu, J. and Li, Y., 2020. Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network. *Information Fusion*, *53*, pp.174-182.
5. Gao J, Zhang H, Lu P, Wang Z. An effective LSTM recurrent network to detect arrhythmia on imbalanced ECG dataset. Journal of healthcare engineering. 2019 Oct 13;2019.
6. Dinakarrao, S.M.P., Jantsch, A. and Shafique, M., 2019. Computer-aided arrhythmia diagnosis with bio-signal processing: A survey of trends and techniques. *ACM Computing Surveys (CSUR)*, *52*(2), pp.1-37.
7. Beauchaine, T.P., Bell, Z., Knapton, E., McDonough-Caplan, H., Shader, T. and Zisner, A., 2019. Respiratory sinus arrhythmia reactivity across empirically based structural dimensions of psychopathology: A meta-analysis. *Psychophysiology*, *56*(5), p.e13329.
8. Wang G, Chen G, Huang X, Hu J, Yu X. Deep Learning-Based Electrocardiograph in Evaluating Radiofrequency Ablation for Rapid Arrhythmia. Computational and Mathematical Methods in Medicine. 2022 Mar 23;2022.
9. Ribeiro HD, Arnold A, Howard JP, Shun-Shin MJ, Zhang Y, Francis DP, Lim PB, Whinnett Z, Zolgharni M. ECG-based real-time arrhythmia monitoring using quantized deep neural networks: A feasibility study. Computers in Biology and Medicine. 2022 Apr 1;143:105249.
10. Izci, E., Ozdemir, M.A., Degirmenci, M. and Akan, A., 2019, October. Cardiac arrhythmia detection from 2d ecg images by using deep learning technique. In *2019 Medical Technologies Congress (TIPTEKNO)* (pp. 1-4). IEEE.
11. Muhammed, L.A., 2021. Arrhythmia Classification Using One Dimensional Conventional Neural Network. *International Journal of Advances in Soft Computing & Its Applications*, *13*(3).
12. Xiong, Z., Nash, M.P., Cheng, E., Fedorov, V.V., Stiles, M.K. and Zhao, J., 2018. ECG signal
classification for the detection of cardiac arrhythmias using a convolutional recurrent neuralnetwork. *Physiological measurement*, *39*(9), p.094006.