**Project report on:-**

**ECG Heartbeat Classification: –**

**Detect abnormal heartbeats from ECG signals**

**Team: -**

1. **Shuvadip pal (22/IT/167)**
2. **Shuvojit Saha (22/IT/166)**
3. **Sirshendu Ghorai(22/IT/169)**
4. **Snigdhendu Chattopadhyay (22/IT/171)**
5. **Soham Mukherjee (22/IT/172)**

**ECG Heartbeat Classification: –**

**Detect abnormal heartbeats from ECG signals**

**1. Introduction**

* Importance of ECG in medical diagnosis.
* What is heartbeat classification?
* Real-world applications (e.g., early detection of arrhythmias, heart diseases).

**2. Literature Review**

Summarize key recent works:-

| **Authors & Year** | **Dataset Used** | **Methodology** | **Findings / Accuracy** |
| --- | --- | --- | --- |
| Acharya et al., 2017 | MIT-BIH | CNN | 97% accuracy for arrhythmia classification |
| Kiranyaz et al., 2016 | MIT-BIH | 1D Convolutional NN | Real-time classification of arrhythmias |
| Rajpurkar et al., 2017 | PhysioNet | Deep Neural Network | Comparable to cardiologist-level detection |
| Yildirim, 2018 | MIT-BIH | Wavelet + Deep CNN | 99% classification accuracy |
| Faust et al., 2016 | PTB Diagnostic | Statistical + DWT + ML | Accurate MI detection using handcrafted features |

Include:

* Traditional vs. deep learning methods
* Challenges in existing work (imbalanced data, interpretability)

**3. Problem Definition**

"To build a machine learning model capable of classifying normal and abnormal heartbeats from ECG signals, using public datasets and modern classification techniques."

**Objectives:**

* Load and preprocess ECG signal data
* Train a supervised classification model
* Evaluate performance using standard metrics
* Handle class imbalance and overfitting
* Interpret results to ensure medical reliability

**4. Dataset Information**

We will use the publicly available dataset from Kaggle:

| **File Name** | **Source Database** | **Labels** | **Classification Type** | **Records** |
| --- | --- | --- | --- | --- |
| mitbih\_train.csv | MIT-BIH Arrhythmia | 0 to 4 (5 classes) | Multiclass heartbeat classification | ~87,000 |
| mitbih\_test.csv | MIT-BIH Arrhythmia | 0 to 4 | Multiclass heartbeat classification | ~21,000 |
| ptbdb\_train.csv | PTB Diagnostic | 0 (normal), 1 (MI) | Binary classification (disease vs. healthy) | ~3,600 |

**ECG Vector Format:**

* Each row represents one heartbeat with:
  + **187** time-domain sampled values (ECG waveform)
  + **1** label (class)

**5. Methodology:-**

Following one of the two major approaches:

**A. Classical Machine Learning**

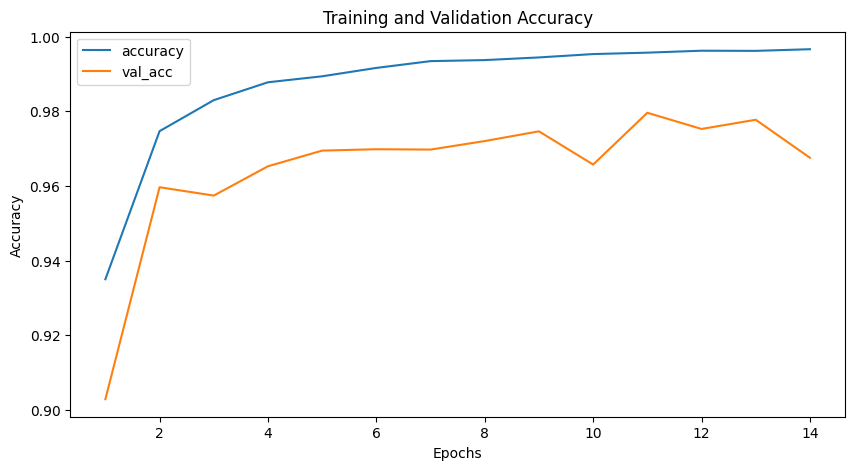
1. **Preprocessing**:
   * Normalize ECG vectors
   * Handle class imbalance (undersampling, oversampling)
2. **Feature Engineering** (optional):
   * Time-domain and frequency-domain features (e.g., DWT, FFT)
3. **Model Selection**:
   * Random Forest
   * k-Nearest Neighbors
   * SVM
   * CNN
   * XGBOOST
4. **Evaluation**:
   * Accuracy, Precision, Recall, F1-score

**B. Deep Learning (Recommended)**

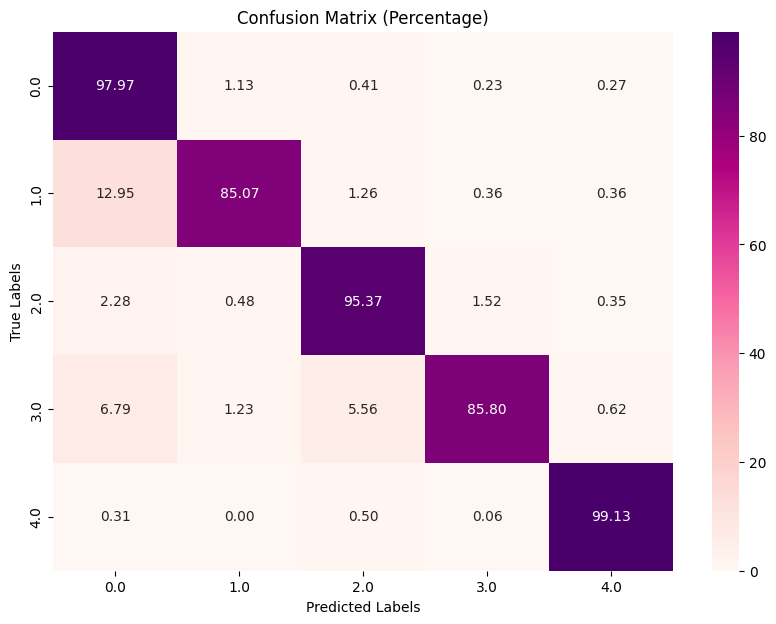
1. **Input Shape**: 1D ECG signal vector (shape: 187 x 1)
2. **Model Architecture**:
   * **1D CNN layers**: for capturing spatial patterns
   * **Dropout / BatchNorm**: for regularization
   * **Dense layers + Softmax/Sigmoid**
3. **Loss Functions**:
   * Binary Cross-Entropy (for PTBDB)
   * Categorical Cross-Entropy (for MIT-BIH)
4. **Handling Class Imbalance**:
   * Use **weighted loss function**
   * **SMOTE** or oversampling of minority classes
5. **Training & Evaluation**:
   * Train/Validation split
   * Use callbacks: EarlyStopping, ReduceLROnPlateau
   * Evaluate using metrics + ROC-AUC curve
6. **Evaluation Metrics:-**

| **Metric** | **Use** |
| --- | --- |
| Accuracy | General performance |
| Precision | Important for minimizing false positives |
| Recall | Important for detecting all abnormal cases |
| F1-score | Balance between precision & recall |

1. Results







Classification Report of CNN model–

Classification Report:

precision recall f1-score support

0.0 0.99 0.98 0.99 18118

1.0 0.69 0.85 0.76 556

2.0 0.93 0.95 0.94 1448

3.0 0.68 0.86 0.76 162

4.0 0.97 0.99 0.98 1608

accuracy 0.97 21892

macro avg 0.85 0.93 0.89 21892

weighted avg 0.98 0.97 0.98 21892

**Classification report of MLP model –**

Classification Report:

precision recall f1-score support

0.0 0.99 0.96 0.98 18118

1.0 0.53 0.79 0.63 556

2.0 0.83 0.96 0.89 1448

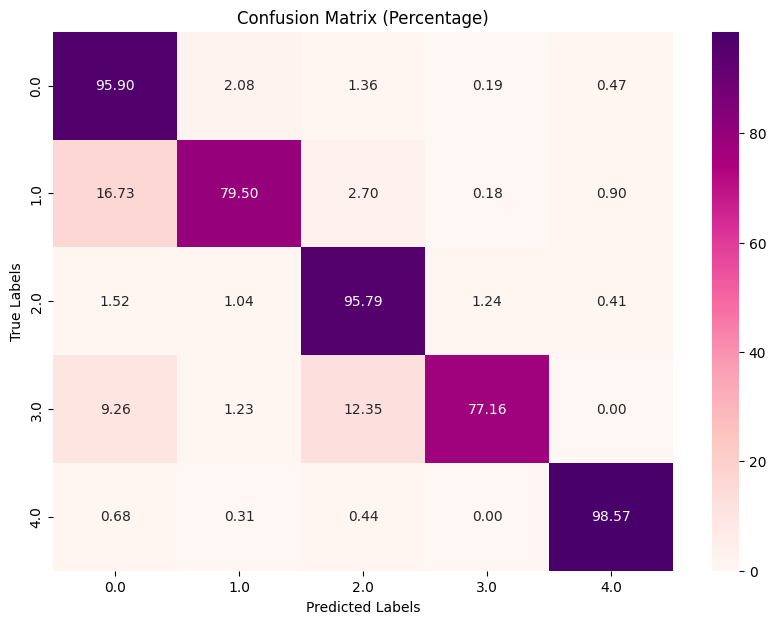
3.0 0.70 0.77 0.73 162

4.0 0.94 0.99 0.96 1608

accuracy 0.96 21892

macro avg 0.80 0.89 0.84 21892

weighted avg 0.96 0.96 0.96 21892



**8. Challenges & Future Scope**

**Challenges:**

* **Class imbalance**: Normal beats dominate the dataset
* **Real-world variability**: Different patient conditions and noise
* **Model interpretability**: Important in clinical settings

**Future Work:**

* Use **transfer learning** with pre-trained models
* Integrate with **real-time ECG monitoring devices**
* Deploy on **mobile or edge devices** for on-the-go diagnosis
* Apply **explainable AI (XAI)** techniques for transparency

**References**

1. Acharya, U. R., et al. “A deep convolutional neural network model to classify heartbeats.” Computers in Biology and Medicine (2017).
2. Shayan Fazeli. *Heartbeat Dataset*, Kaggle. https://www.kaggle.com/datasets/shayanfazeli/heartbeat
3. Goldberger AL, et al. *PhysioNet: MIT-BIH Arrhythmia Database*. https://www.physionet.org
4. Kiranyaz, S., et al. “Real-time patient-specific ECG classification by 1-D convolutional neural networks.” IEEE T-BME (2016).