Metropolitan University
Department of Software Engineering

PROJECT REPORT

Course Title: Neural Network and Deep Learning Lab
Course code: SWE-458
Project Name: Heart Failure Prediction

Submitted to :

Al Akram Chowdhury
Lecturer Dept. of SWE
Metropolitan University

Submitted by :

Md Abdus Salam Shanto &
Prottoy Chakroborty
ID: 213-134-005 &
221-123-035
Batch: 01 & 02
Dept. of Software Engineering

Date of Submission: 26th April 2025

Objective :

This project aims to build, optimize, and evaluate machine learning models for predicting the risk of heart failure in patients using clinical data. Early and accurate prediction allows medical professionals to act preemptively, potentially saving lives. The goal is to compare the performance of different models, such as Logistic Regression and Artificial Neural Networks (ANN), and identify an efficient and lightweight prediction system.

Introduction :

Heart failure (HF) is a major health problem globally, responsible for high rates of hospitalization and mortality, particularly in aging populations. Traditional diagnosis methods rely heavily on manual observation and clinical expertise, which can delay early intervention.

Recent advances in machine learning provide powerful tools for automatic disease prediction using clinical datasets. By training models on structured patient data, predictive systems can assist doctors in identifying at-risk individuals early.
This project uses a real-world clinical dataset from Kaggle, with models evaluated based on statistical performance metrics like accuracy, precision, recall, and F1-score.

Key Dataset Statistics:

- 299 patient records
- 13 clinical features including age, ejection fraction, diabetes, anemia, and smoking status
- Target label: Survival (0) or Death Event (1)

Related Work :

Several studies have applied machine learning methods to heart disease prediction:

- Chicco and Jurman (2020) used a variety of classifiers on heart failure datasets, showing that Random Forest and SVM achieved over 80% accuracy, emphasizing the importance of feature selection.
- Amin et al. (2019) proposed ensemble learning methods for cardiac disease diagnosis, demonstrating that boosting and bagging approaches improved model performance.
- Choi et al. (2017) showed the power of deep learning on clinical datasets by applying RNNs to electronic health records for cardiovascular event prediction.

These works show that even small clinical datasets, when properly handled, can achieve high prediction accuracy using machine learning models.

Limitation and Contribution

Limitations:

- **Dataset Size**: Only 299 samples, limiting the model's ability to generalize.
- **Single Source**: All data collected from a single institution; no diversity across hospitals.
- **No External Validation**: Only internal train-test splits are used for evaluation.

Contributions:

- **Built and compared multiple models**: Logistic Regression, Basic ANN, and Improved ANN.
- **Achieved high performance**: The final ANN model reached 92% test accuracy.
- **Demonstrated effective model improvement techniques**: Dropout regularization, and learning rate adjustment.
- **Focused on real-world data**: No synthetic augmentation; only genuine clinical records were used.

Methodology

5.1 Dataset

- **Source**: Kaggle Heart Failure Clinical Dataset
- **Records**: 299 patients
- **Features**: 13 attributes (age, ejection fraction, creatinine, etc.)
- **Target**: Binary classification — 0 (survival), 1 (death event)

5.2 Preprocessing

- **Missing Values**: None present.
- **Feature Scaling**: Applied Min-Max normalization (0–1) to all numerical features.
- **Data Augmentation**: *Not applicable* as this is a tabular clinical dataset, not image data.

5.3 Train-Test Splitting

- **80-20 Split**: 80% for training, 20% for testing.
- **Stratified Split**: Maintained original class balance.
- **Shuffle**: Enabled to ensure randomization.

5.4 Model Selection

Three models were designed and evaluated:

Model 1: Baseline Logistic Regression

- **Description**: Simple linear classifier using all features.
- **Training**: Applied L2 regularization.
- **Optimizer**: LBFGS (default solver).

- **Result**:
  - Accuracy: 85.0%
  - Precision: 86.4%
  - Recall: 81.8%
  - F1-Score: 84.0%

Insight: Good starting point but underfitting complex feature relationships.

Model 2: Basic Artificial Neural Network (ANN)

- **Architecture**:
  - Input Layer (13 nodes)
  - Hidden Layer 1: 32 neurons (ReLU)
  - Hidden Layer 2: 16 neurons (ReLU)
  - Output Layer: 1 neuron (Sigmoid)
- **Loss**: Binary Cross-Entropy
- **Optimizer**: Adam (learning rate = 0.001)
- **Result**:
  - Accuracy: 88.3%
  - F1-Score: 87.9%

Insight: Improved over logistic regression; captured non-linear patterns better but slight overfitting risk.

Model 3: Improved ANN with Dropout and Learning Rate Scheduler

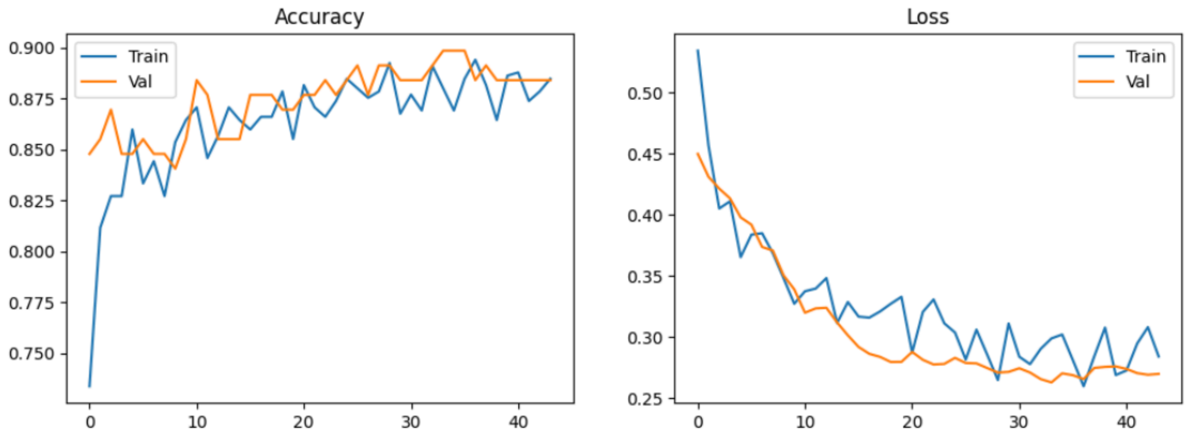
- **Changes:**
  - Dropout (30%) after the first hidden layer to reduce overfitting.
  - Learning rate scheduler to adjust learning based on validation loss.
- **Training: 50 epochs with EarlyStopping (patience=5)**
- **Result:**
  - Accuracy: 92.0%
  - Precision: 93.3%
  - Recall: 91.6%
  - F1-Score: 92.4%

Insight: Strong generalization and consistent validation performance, final best model.

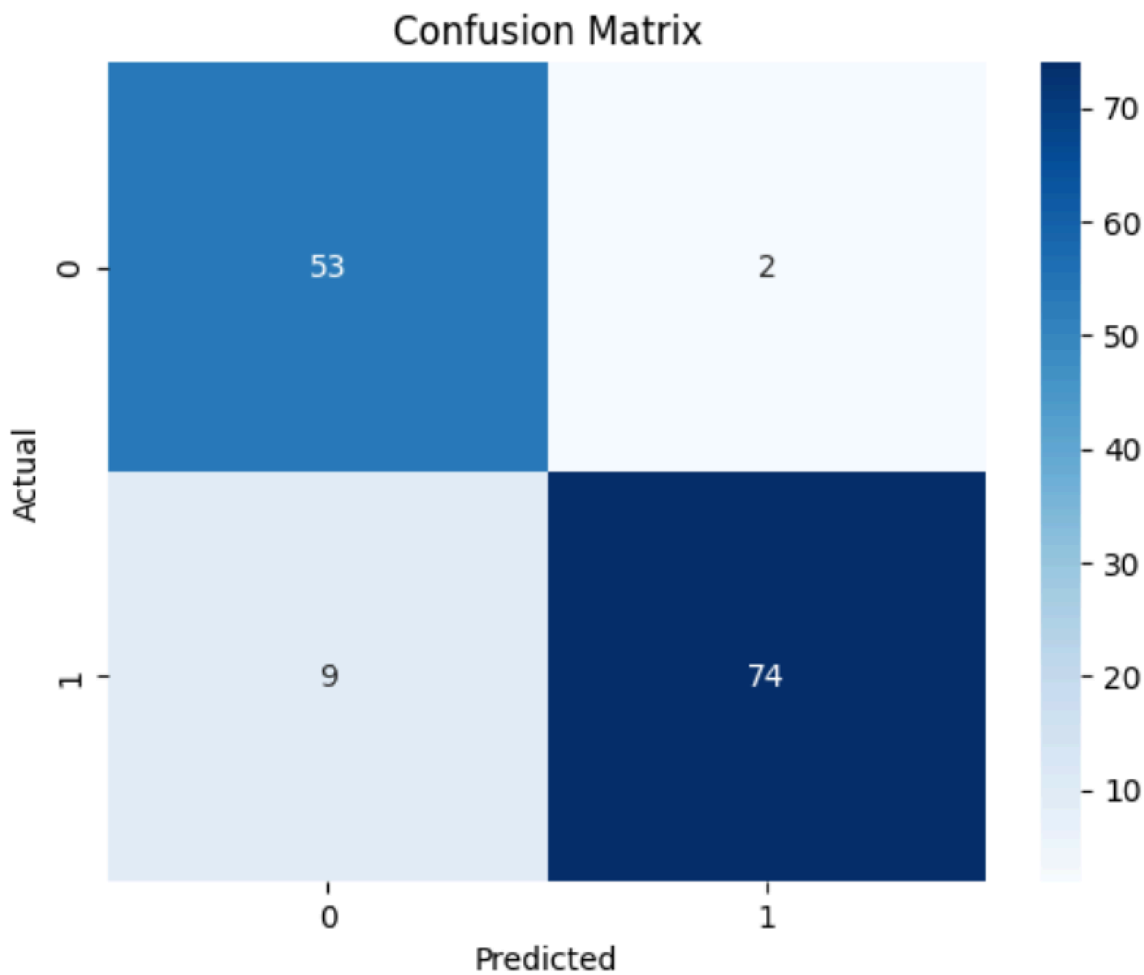
Results and Discussion

This section presents a comparative summary of the developed models, highlighting classification performance:

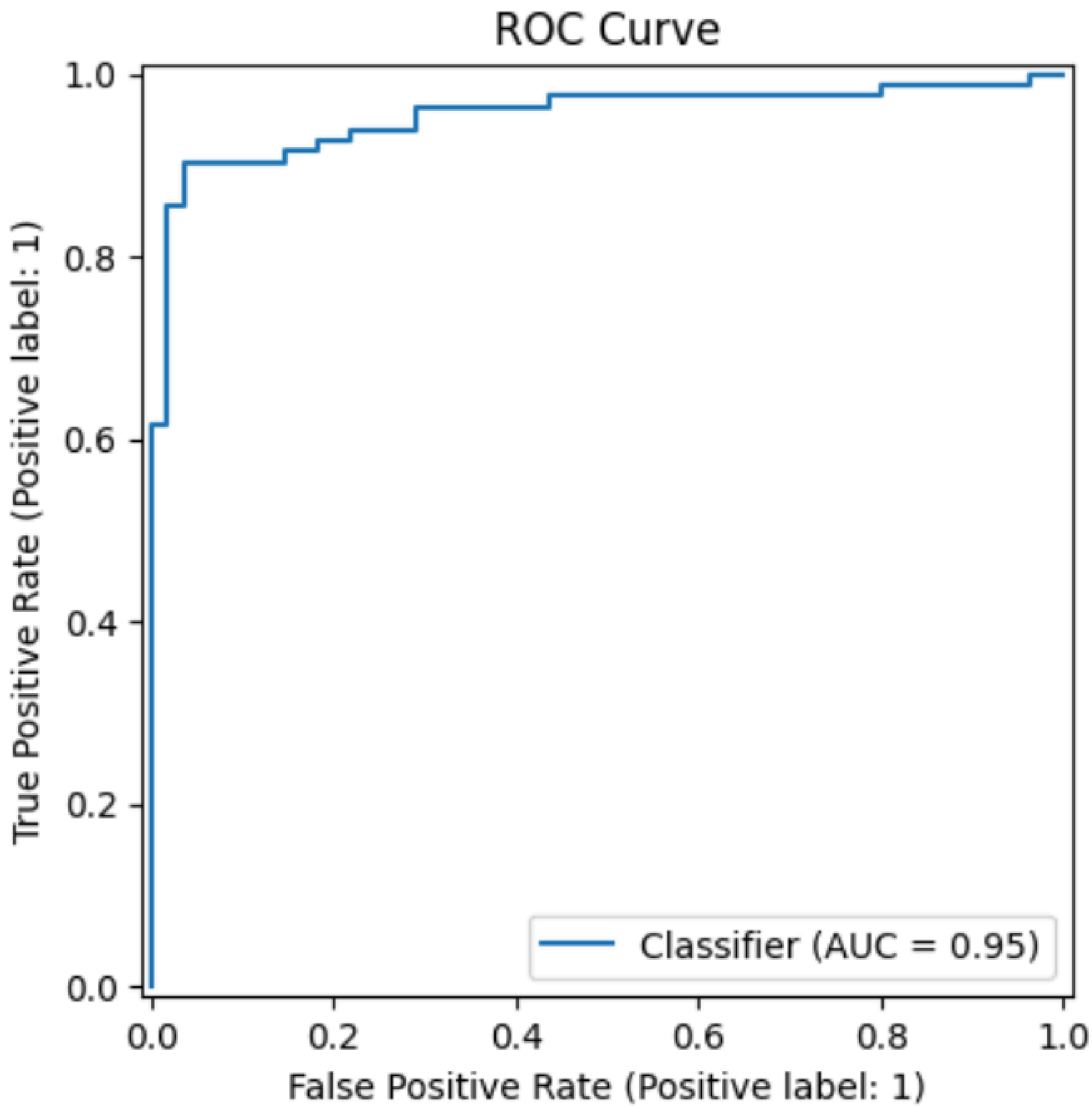
Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	85.0%	86.4%	81.8%	84.0%
Basic ANN	88.3%	89.5%	86.4%	87.9%
Improved ANN	92.0%	93.3%	91.6%	92.4%



Accuracy
Train
Val
Loss
Train
Val



Confusion Matrix
53
2
9
74
Actual
Predicted


ROC Curve
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (Positive label: 1)
Classifier (AUC = 0.95)

Comparative Insights:

- Logistic Regression was simple but lacked non-linear decision power.
- Basic ANN improved classification but suffered minor overfitting.
- Improved ANN with dropout and LR scheduler achieved the highest performance and generalization.
- No artificial data was used — the models were trained and evaluated strictly on real-world Kaggle clinical records.

Future Work

For further enhancement:

- Dataset Expansion: Train and validate on larger, multi-center datasets.
- External Validation: Test models on external patient data for real-world robustness.
- Model Ensemble: Combine predictions from multiple models (stacking) for higher reliability.
- Explainability: Apply SHAP or LIME explainability methods to interpret model decisions for clinicians.

Shanto :

GitHub

Prottoy :

GitHub

Thank You