Name: Muhammad Asif Abbas
Roll no: BITF22M002

Comparative Study of Classical Machine Learning, Fully Connected Neural Networks, and Sequence-Based Deep Learning Models

Abstract

This report presents a comprehensive comparison of three classification approaches applied to a single dataset: feature-engineering-based classical machine learning models, fully connected deep neural networks using raw data, and sequence-based deep learning models. Performance is evaluated using accuracy and F1-score to analyze the strengths and limitations of each approach.

Introduction / Problem Statement

Classification problems can be addressed using a wide range of modeling techniques. Traditional machine learning models rely heavily on feature engineering, whereas deep learning models learn representations directly from raw data. However, fully connected neural networks often fail to capture temporal dependencies in sequential data. This project aims to compare these approaches to determine their effectiveness under different modeling paradigms.

Methodology

The project follows three approaches:

Approach-1

Feature engineering combined with classical machine learning classifiers such as Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and XGBoost.

Approach-2

Fully connected deep neural networks trained directly on raw input data.

Approach-3

Sequence-based deep learning models (RNN, LSTM, and GRU) trained on raw sequential data.

All approaches use the same dataset and evaluation metrics.

Experiments and Results

Dataset Description

Dataset Name: Epileptic Seizure Recognition Data Set
Source: UCI Machine Learning Repository / Kaggle
Data Type: Sequential time-series EEG data
Train Shape: (9,200, 178)
Test Shape: (2,300, 178)
Class Labels: 1 (Seizure), 2 (Tumor), 3 (Healthy), 4 (Eyes Closed), 5 (Eyes Open)

Data Settings

Train-Test Split: 80% training and 20% testing.

Preprocessing Techniques:

- Classical ML: Feature-engineered statistical and frequency domain features
- **DNN**: Raw flattened feature vectors (178 features per sample)
- **Sequence Models**: Raw sequential data (178 time steps × 1 feature)

Label Encoding:

- **Classical ML:** LabelEncoder (if required)
- **DNN**: LabelEncoder + One-hot encoding
- **Sequence Models**: LabelEncoder + One-hot encoding

Normalization:

- **Classical ML:** StandardScaler
- **DNN**: StandardScaler
- **Sequence Models**: StandardScaler applied per feature

Results (Comparison of All Approaches)

Approach	Model Type	Accuracy	F1-Score
Feature Engineering + ML	Best Classical Model (To be updated)	XX.XX	XX.XX
Fully Connected DNN	DNN with ReLU Activation	XX.XX	XX.XX
Sequence Model	GRU (Best Sequence Model)	0.7013	0.6928

Additional Sequence Model Results:

- LSTM: Accuracy 0.6743, F1-Score 0.6617
- RNN: Accuracy 0.4570, F1-Score 0.4209

Discussion

Feature-engineering-based models require significant domain expertise and manual effort. Fully connected neural networks reduce feature dependency but fail to explicitly model temporal relationships. Sequence-based models effectively capture temporal patterns, making them more suitable for EEG signal classification.

Among the sequence-based approaches, GRU achieved the best performance with 70.13% accuracy and 69.28% F1-score. Both GRU and LSTM outperform the basic RNN due to their gating mechanisms, which help capture long-term dependencies across the 178 time steps. This capability is crucial for identifying seizure-related temporal patterns in EEG signals.

Conclusion

This project demonstrated that model selection depends heavily on data characteristics. For sequential/time-series data like EEG signals, sequence-based deep learning models achieved the best performance by effectively capturing temporal patterns. The GRU model outperformed both LSTM and basic RNN, achieving 70.13% accuracy and 69.28% F1-score, demonstrating the effectiveness of gating mechanisms in learning long-term dependencies in brain electrical activity.