Comparative Analysis of Shallow and Deep Learning Methods for Diabetes Prediction Using the Pima Indians Dataset

Puyang Zhao[1] Zhiyi Yue[1] Md Saifur Rahman[1]

[1]Department of Biostatistics and Data Science, The University of Texas Health Science Center at Houston

Email: Puyang.Zhao@uth.tmc.edu;Zhiyi.Yue@uth.tmc.edu;Mdsaifur.Rahman@uth.tmc.edu

UTHealth Houston
School of Public Health
Biostatistics and Data Science

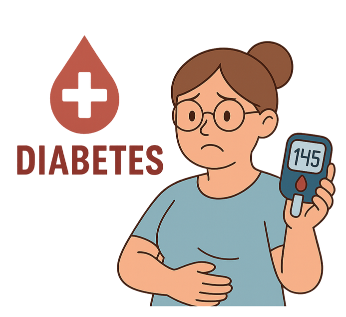
Abstract

Key Findings on Diabetes Prediction using Pima Indians Dataset:

- Compared various Shallow ML (LR, RF, SVM, etc.) and Deep Learning (MLP, DNN, CNN, LSTM) models.
- Evaluated performance using 10-fold cross-validation and an independent test set (Accuracy, AUC, etc.).
- All models provided robust predictions.
- Deep Learning models, especially CNN (AUC 0.8767), showed marginal AUC improvements over the best Shallow ML model (LR, AUC 0.8686).
- Performance differences were small, indicating competitive results from well-tuned shallow models.
- Model selection should balance predictive power, interpretability, and efficiency based on clinical needs.

Introduction

Diabetes is a significant and growing public health issue in the United States [1]. Early detection through accurate predictive models is crucial for timely intervention and effective disease management. This study presents a comparative analysis of shallow machine learning (ML) and deep learning (DL) techniques for diabetes risk prediction using the Pima Indians Diabetes Dataset.



Key Contributions:

- Develop predictive models to enhance early diabetes detection.
- Integrate and validate both traditional shallow ML models and advanced DL architectures.
- Promote reproducibility by providing a comprehensive account of methods and parameters.

Dataset & Preprocessing

Dataset: Pima Indians Diabetes Dataset [2] (768 records).
Features: Pregnancies, Glucose, Blood Pressure, Skin Thickness, BMI, Diabetes Pedigree Function, Age.
Outcome: Binary (1 for diabetic, 0 for non-diabetic).

Preprocessing Highlights:

- Clinically implausible zero values (e.g., in Glucose, BMI) treated as missing.
- Missing values imputed using k-Nearest Neighbors (kNN, $k = 5$).
- Features standardized using Z-score normalization: $x_{\text{scaled}} = \frac{x-\mu_{\text{train}}}{\sigma_{\text{train}}}$.
- Data split (stratified): Training (70%, 429 samples), Validation (15%, 108 samples), Test (15%, 231 samples).

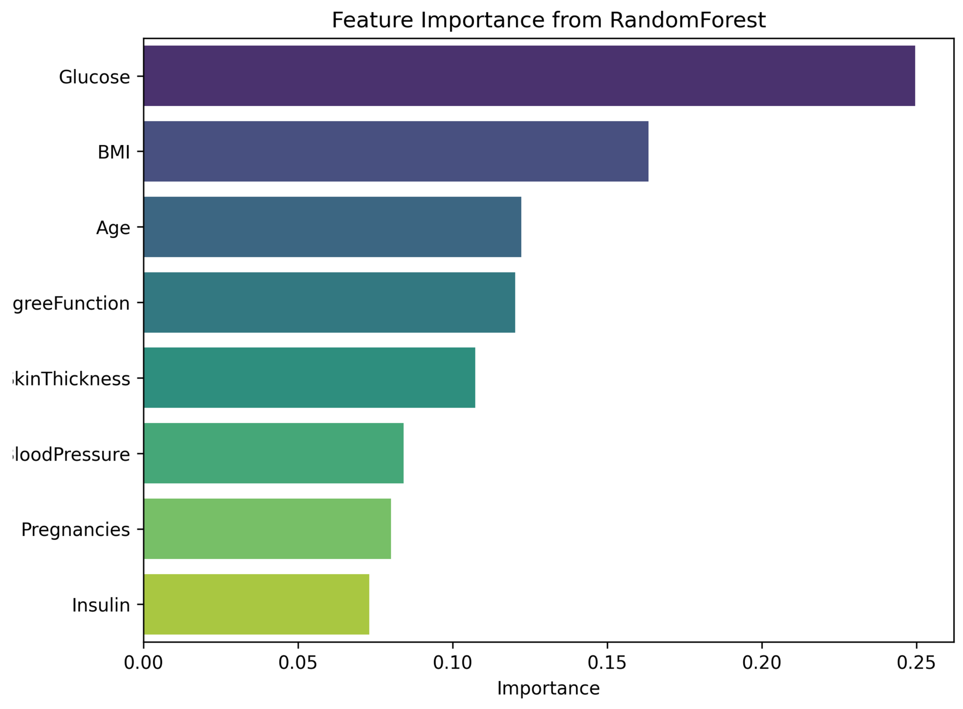


Figure 1. Feature importance scores (Random Forest).

Methodology: Shallow Machine Learning

Five traditional models were implemented and evaluated:

- Logistic Regression (LR)
- Random Forest (RF)
- Gradient Boosting Machine (GBM)
- Support Vector Machine (SVM) with RBF kernel
- k-Nearest Neighbors (kNN)

Key hyperparameters were tuned (e.g., LR: max_iter=1000; RF/GBM: n_estimators=100; kNN: n_neighbors=5).

Methodology: Deep Learning Architectures

A range of DL models were developed using Keras/TensorFlow:

- Multilayer Perceptrons (MLP): Basic, with Dropout (20-50%), with Batch Normalization.
- Deep MLP: Deeper architecture (e.g., Dense(32)-Dense(16)).
- Deep Neural Network (DNN): e.g., Dense(64)-Dropout(0.2)-Dense(32)-Dropout(0.2).
- Convolutional Neural Network (CNN): e.g., Conv1D(32)-MaxPool-Conv1D(64)-Dense(16).
- LSTM and LSTM with Attention.

All DL models used Adam optimizer (lr=0.001), categorical cross-entropy loss, and early stopping based on validation loss.

Results: CNN Model Training

The CNN model, which achieved the highest AUC, demonstrated stable training.

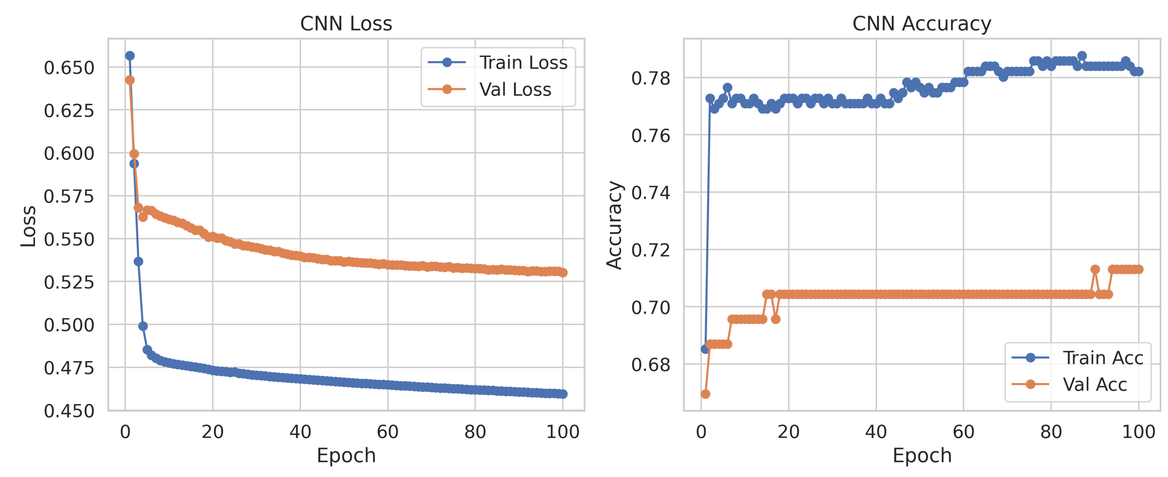


Figure 2. Training and validation loss/accuracy for the CNN model. Early stopping mitigated overfitting.

Stable training with minimal overfitting was observed. Early convergence and consistent validation loss confirm the CNN model's strong generalization performance.

Results: Shallow vs. Deep Learning

Shallow ML (Test Set): Logistic Regression achieved the highest AUC and overall balanced performance. Tree-based methods also performed competitively but showed slightly lower precision.

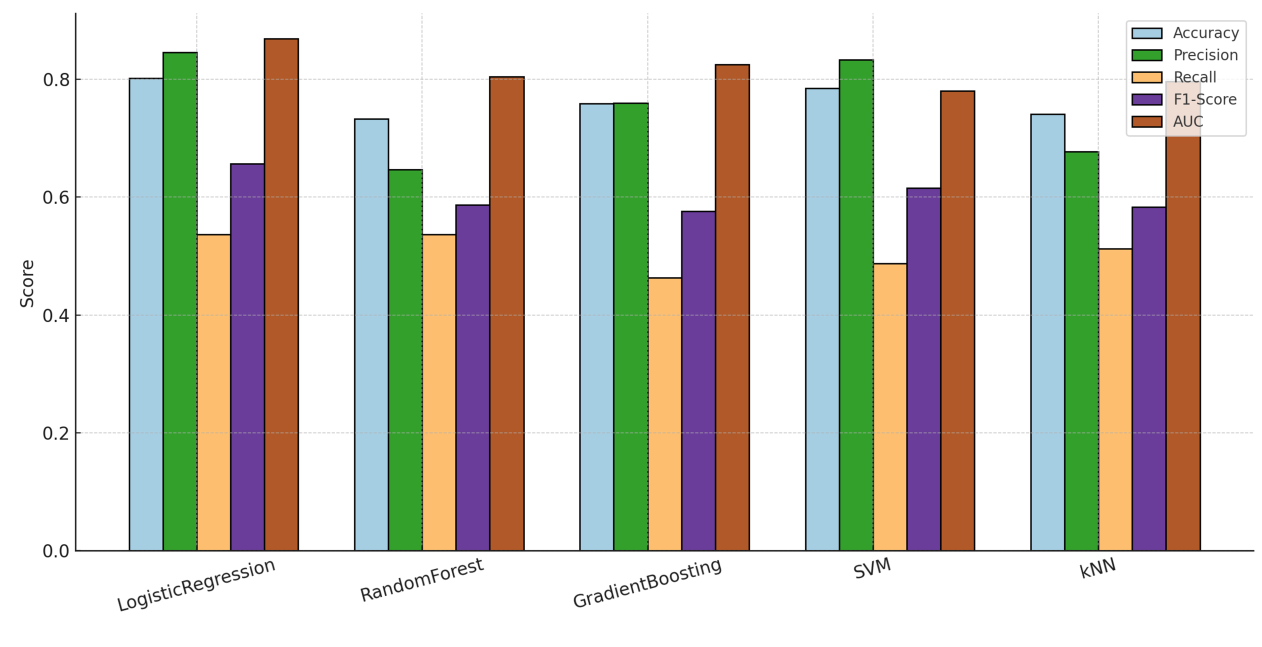


Figure 3. Shallow ML Performance.

Deep Learning (Test Set): CNN achieved the highest AUC; basic MLP had highest accuracy.

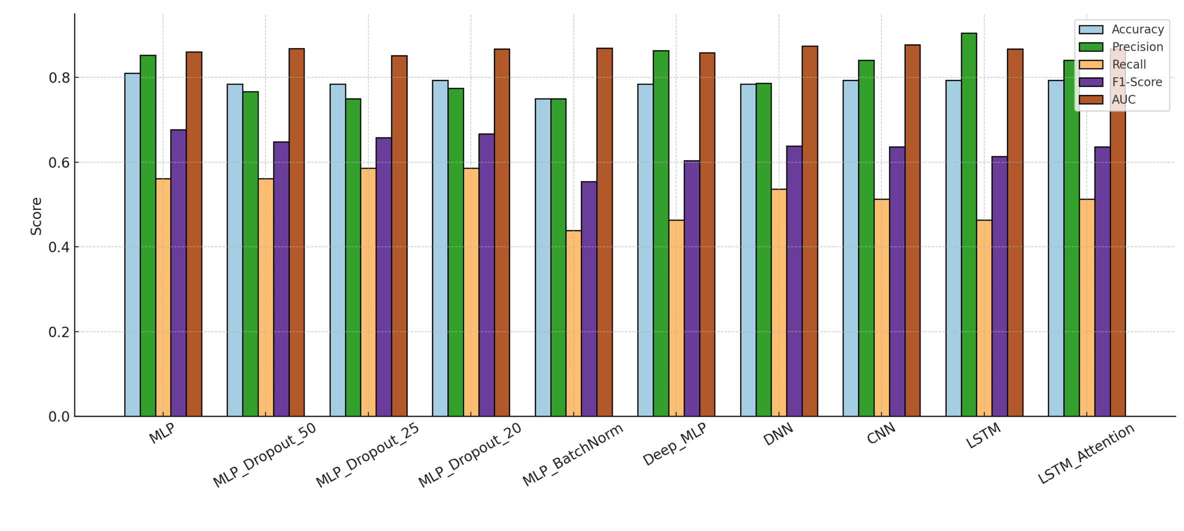


Figure 4. Deep Learning Performance.

Model Evaluation

- Rigorous evaluation via 10-fold cross-validation on the training set.
- Final performance assessed on an independent test set.
- Metrics: Accuracy, Precision, Recall, F1-Score, Area Under the Curve (AUC).

Comparative Analysis Summary

- Both shallow ML and DL models yielded robust predictions.
- DL models, particularly CNN (AUC 0.8767), achieved marginal AUC improvements over the best shallow model, Logistic Regression (AUC 0.8686).
- The basic MLP showed the highest accuracy (0.8103).
- The performance gap is narrow, highlighting the competitiveness of well-tuned shallow models.
- Shallow models offer better interpretability and computational efficiency.
- Deep models excel at capturing complex non-linear interactions.

Conclusion and Future Work

Shallow and deep learning methods are effective for diabetes prediction. While DL models like CNN can offer a slight performance edge in capturing complex data patterns, simpler models like Logistic Regression remain highly competitive and interpretable. Model selection should balance predictive power with clinical utility, interpretability, and resource constraints.

Future Work:

- Develop hybrid models combining shallow model transparency with DL representation power.
- Expand datasets to include more diverse populations for improved generalizability.
- Incorporate explainability techniques (e.g., SHAP values, attention mechanisms) to enhance clinical trust and insight.
- Further refine feature engineering and hyperparameter optimization strategies.

References

[1] Centers for Disease Control and Prevention. National diabetes fact sheet: General information and national estimates on diabetes in the united states. https://www.cdc.gov, 2021. Accessed: 2023-11-30.

[2] Leslie O Schulz, Peter H Bennett, Eric Ravussin, Judith R Kidd, Kenneth K Kidd, Julian Esparza, and Mauro E Valencia. Effects of traditional and western environments on prevalence of type 2 diabetes in pima indians in mexico and the us. *Diabetes care*, 29(8):1866–1871, 2006.

Authors: Zhiyi Yue(Left), Puyang Zhao and Md Saifur Rahman(Right)