Neural Network Based Hyperparameter Optimization for Random Forest Models

- DESIGN PRESENTATION -

INTRODUCTION

What are HYPERPARAMETERS in Machine Learning ?

IMPORTANCE OF HYPERPARAMETER OPTIMIZATION

- Control the training process and model complexity
- Minimize the Loss Function
- Improve model performance and Ensure reliable predictions
- Reduce overfitting and underfitting

LITERATURE REVIEW

Grid Search

Exhaustive search through all possible combinations of hyperparameters

Bayesian Optimization

Uses probabilistic models to predict hyperparameter combinations iteratively

Random Search

Samples hyperparameters randomly from a defined space

Genetic Algorithms

Evolves hyperparameters over generations using mutation and selection mechanisms

RESEARCH PROBLEM

Grid Search

- Computationally expensive
- Infeasible for large search spaces

Random Search

- More efficient than grid search but unreliable
- Struggles with complex models and large parameter spaces

Bayesian Optimization

- Need good understanding about the probabilistic model
- Struggles with large search spaces

Genetic Algorithms

- Computationally intensive
- Requires careful tuning of evolutionary parameters

OBJECTIVES

Design a neural network-based meta-learning model to

- Predict optimal hyperparameters for Random Forest models
- Minimize computational cost and Achieve higher accuracy
- Compare with traditional approaches like grid search

METHODOLOGY

- Select datasets to create the meta dataset
- Define hyperparameter grid with selected hyperparameters
- Evaluate hyperparameter combinations using traditional methods
- Create a meta-dataset with dataset characteristics and accuracy
- Design a neural network to predict accuracy for hyperparameters
- Train the neural network on the meta-dataset
- Predict hyperparameters for new datasets using the trained model

USED DATASETS

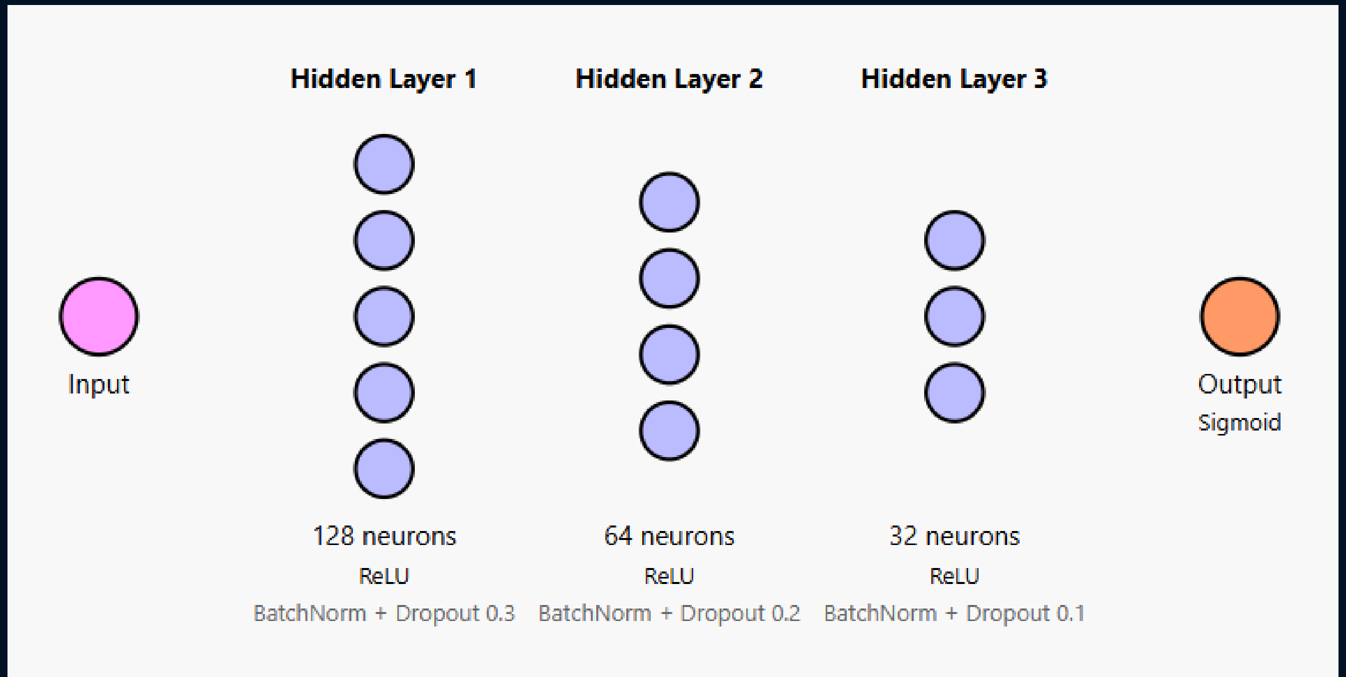
10 Binary Classification Datasets

Dataset	No. of rows	No. of features	Class imbalance ratio
Heart Disease	1025	13	1.0541
Titanic Survival	712	9	1.4722
Diabetes Diagnosis	768	8	1.8657
BankNote Authentication	1348	4	1.2098
Water Quality	7999	20	7.7708
Wine	1359	11	1.1234
Age Prediction	3524	5	1.0000
Gender Classification	3233	7	1.2296
Brain Stroke	4981	11	9.0847
AIDS Classification	5000	18	2.1666

RESULTS & DISCUSSION

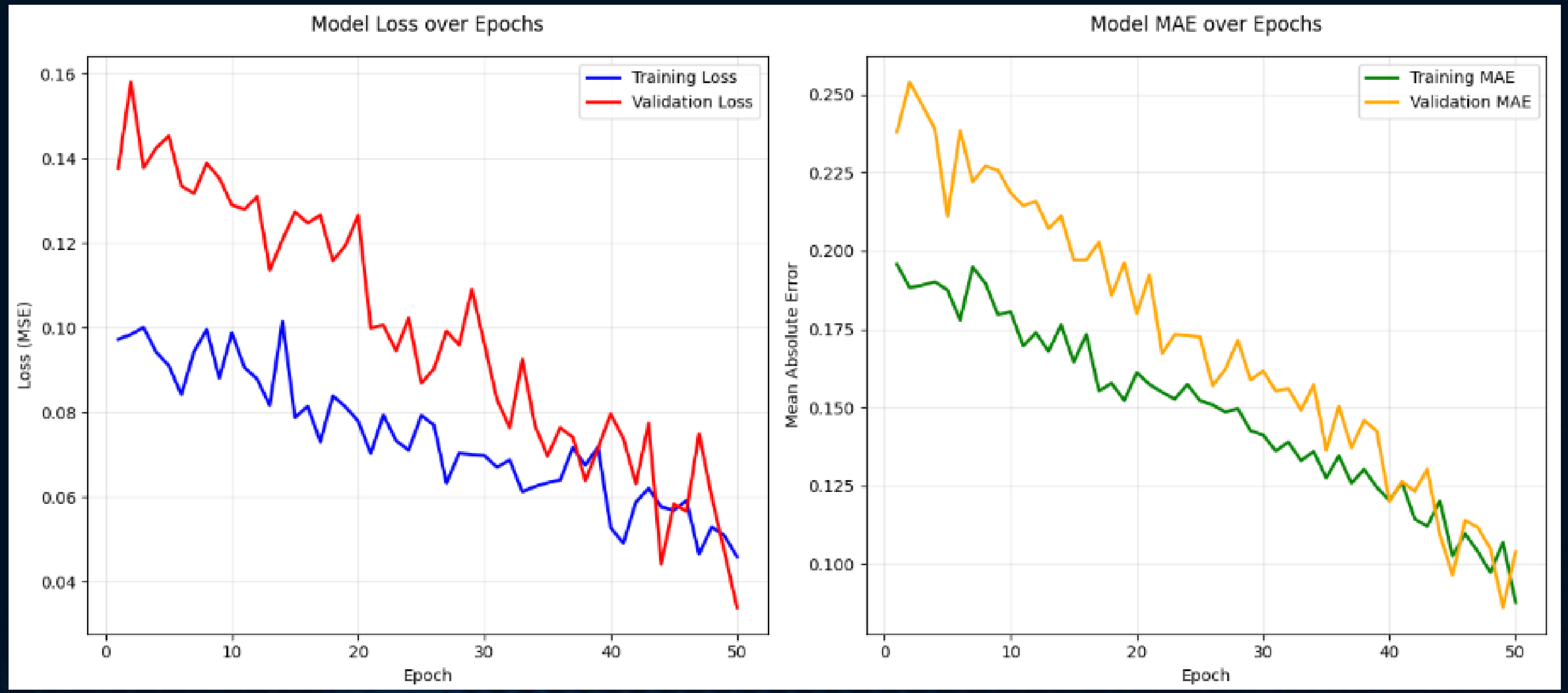
Considered Hyperparameters

- n_estimators
- max_depth
- max_features
- min_samples_split



Dataset Characteristics

- number of samples
- number of features
- class imbalance ratio
- number of categorical features
- number of numerical features



RESULTS & DISCUSSION Ctd.

Dataset	Grid Search		Bayesian Optimization		NN based approach	
	Accuracy	Time(s)	Accuracy	Time(s)	Accuracy	Time(s)
Liver disease	0.9749	87.21	1.0000	6.91	0.9875	13.26
Diabetes classification	0.9695	260.98	0.9993	22.46	0.9878	13.43
Patient treatment	0.9456	149.61	1.0000	15.35	0.9924	12.19
Loan prediction	0.9613	54.21	1.0000	5.52	0.9688	10.82

RESULTS & DISCUSSION Ctd.

Strengths:

- Balances accuracy and computational efficiency.
- Generalizability: Pre-trained model can be applied across datasets.
- Practical use for real-time predictions with minimal computational cost.

Limitations:

- Marginally slower than Bayesian Optimization in some cases.
- Relatively lower accuracy than Bayesian Optimization.
- Requires an initial meta-dataset and training phase.

CONCLUSION

- **Proposed a Neural Network-based meta-learning method for hyperparameter optimization for Random Forest models**
- **Demonstrated competitive accuracy and efficiency across datasets**
- **The method offers a practical alternative for scenarios where traditional methods are computationally prohibitive**
- **Future Work:**
  - Expand the meta-dataset with more diverse datasets
  - Improve the model architecture to achieve higher accuracy and efficiency