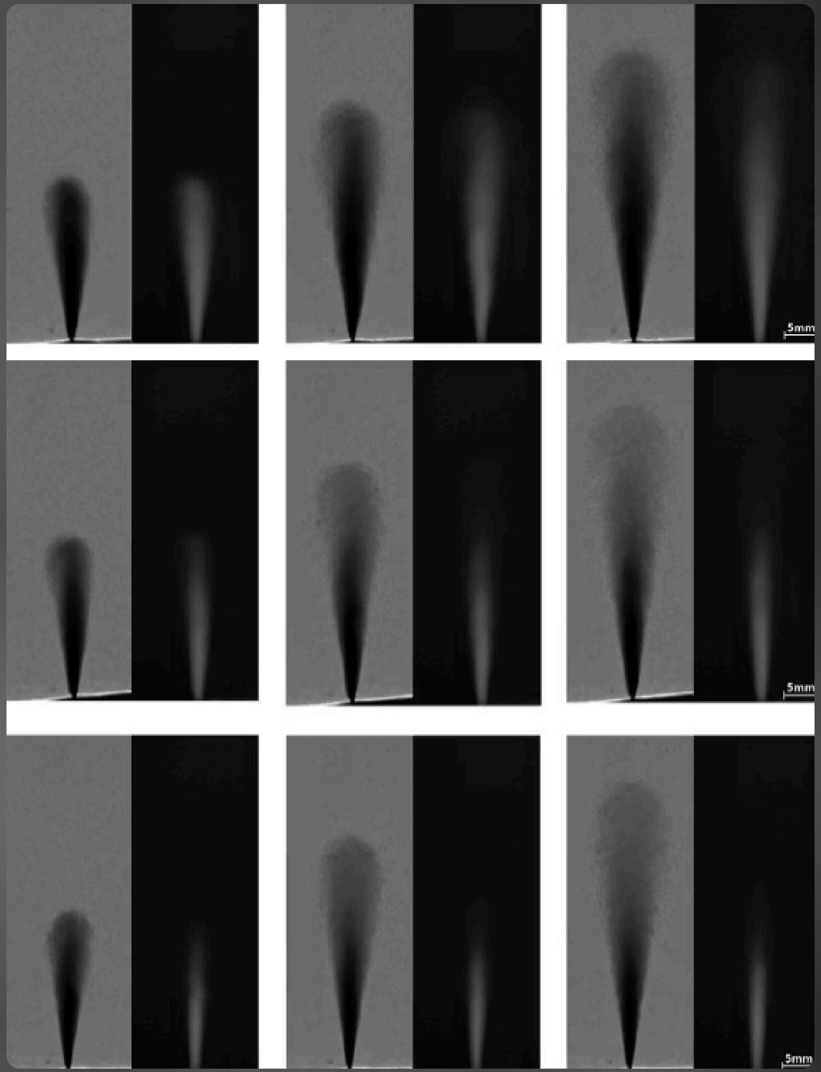
Time-Series Prediction of Spray Characteristics using Classical Machine Learning Approaches

Yuvin Raja (22BCE3942)
Faculty Guide: Prof. P. Suresh Kumar, Assistant Professor Sr. Grade 2,
Automotive Research Centre (ARC)

Review II of BCSE497J Project-I
School of Computer Science and Engineering (SCOPE)
10 September 2025



Introduction & Motivation

Critical Role of Spray Characteristics

Spray characteristics, such as angle and penetration length, are fundamental to the performance of combustion engines, agricultural systems, and various thermal applications.

Limitations of Traditional Methods

Traditional Computational Fluid Dynamics (CFD) simulations, while accurate, are often computationally expensive and time-consuming, hindering rapid iteration.

Challenges with Experimental Approaches

Experimental methods for analyzing sprays demand specialized equipment and deep expertise, limiting their scalability and real-time applicability.

The Need for ML Surrogates

There is a growing need for fast, interpretable Machine Learning surrogates capable of real-time prediction for control and optimization tasks.

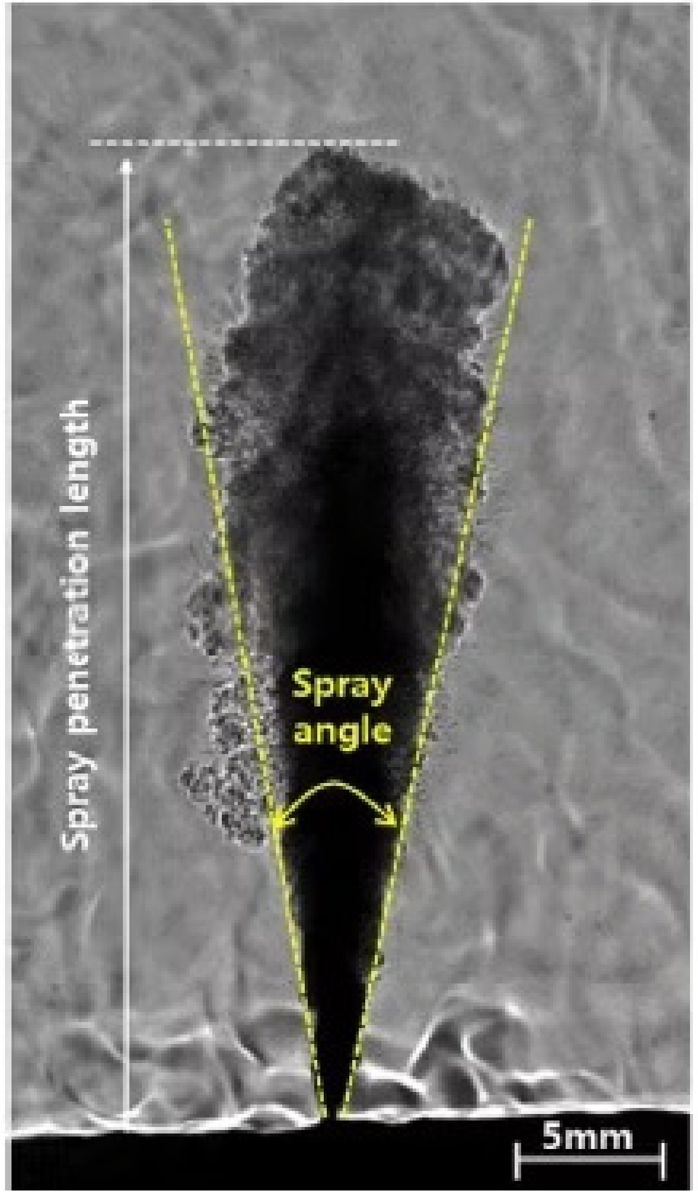


FIG. 5. Definitions of spray angle and penetration length.

Literature Review

01

Traditional Approaches

CFD simulations and optical diagnostics including shadowgraph and Mie scattering techniques for spray characterization

02

Recent Trends

Dominance of artificial neural networks and deep learning models with complex architectures requiring extensive computational resources

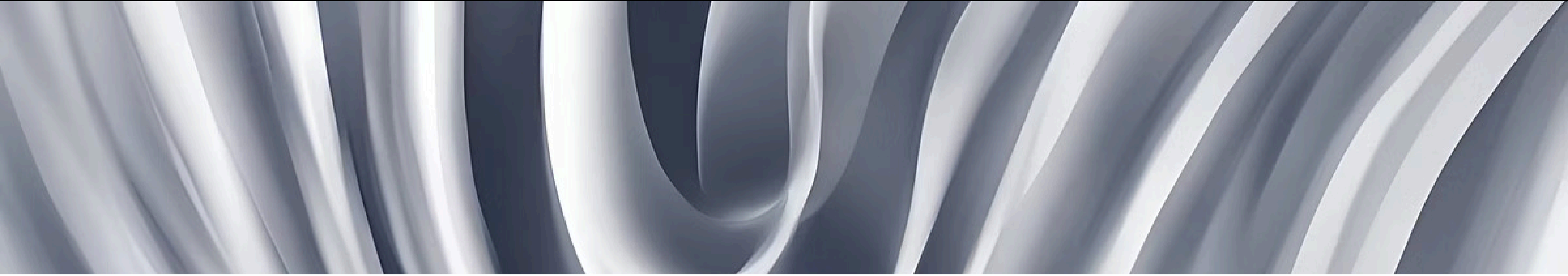
03

Advanced Methods

GA-BP neural networks, LSTM models, and optimized machine learning frameworks for enhanced prediction accuracy

Key References

- Hwang et al. (2022): Gasoline spray ML prediction
- Park (2022): Flash boiling spray characteristics
- Wu & Nadimi (2024): GA-BP neural networks for diesel sprays
- Khan & Masood (2024): Direct injection gasoline fuel sprays



Research Gaps Identified

Limited Classical ML Evaluation

Systematic assessment of traditional machine learning models remains underexplored in spray prediction applications

Black Box Problem

Overemphasis on complex models with poor interpretability limits understanding of physical spray phenomena

Computational Barriers

High resource requirements prevent real-time deployment in resource-constrained industrial environments

Benchmarking Deficiency

Absence of standardized performance benchmarks for classical regressors in small-data experimental scenarios

Project Objectives

1 Model Evaluation Framework

Systematically evaluate six classical ML models: Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Regression, and K-Nearest Neighbors

2 Multi-Output Prediction

Predict four critical spray characteristics: spray angle and penetration length via shadowgraph and Mie scattering measurement techniques

3 Temporal Data Processing

Utilize comprehensive time-series dataset with 726 samples and six input features while maintaining temporal integrity through chronological splitting

4 Performance Benchmarking

Compare classical models against artificial neural network implementation to establish competitive performance baselines

Hardware & Software Specifications

Hardware Configuration

Intel Core i7-10700K (8 cores, 3.8 GHz), 16GB DDR4-3200 RAM, NVIDIA GTX 1650Ti (4GB), 1TB NVMe SSD storage

Development Environment

Python 3.12 with Anaconda distribution, scikit-learn (0.24.2), pandas, numpy, matplotlib, seaborn libraries

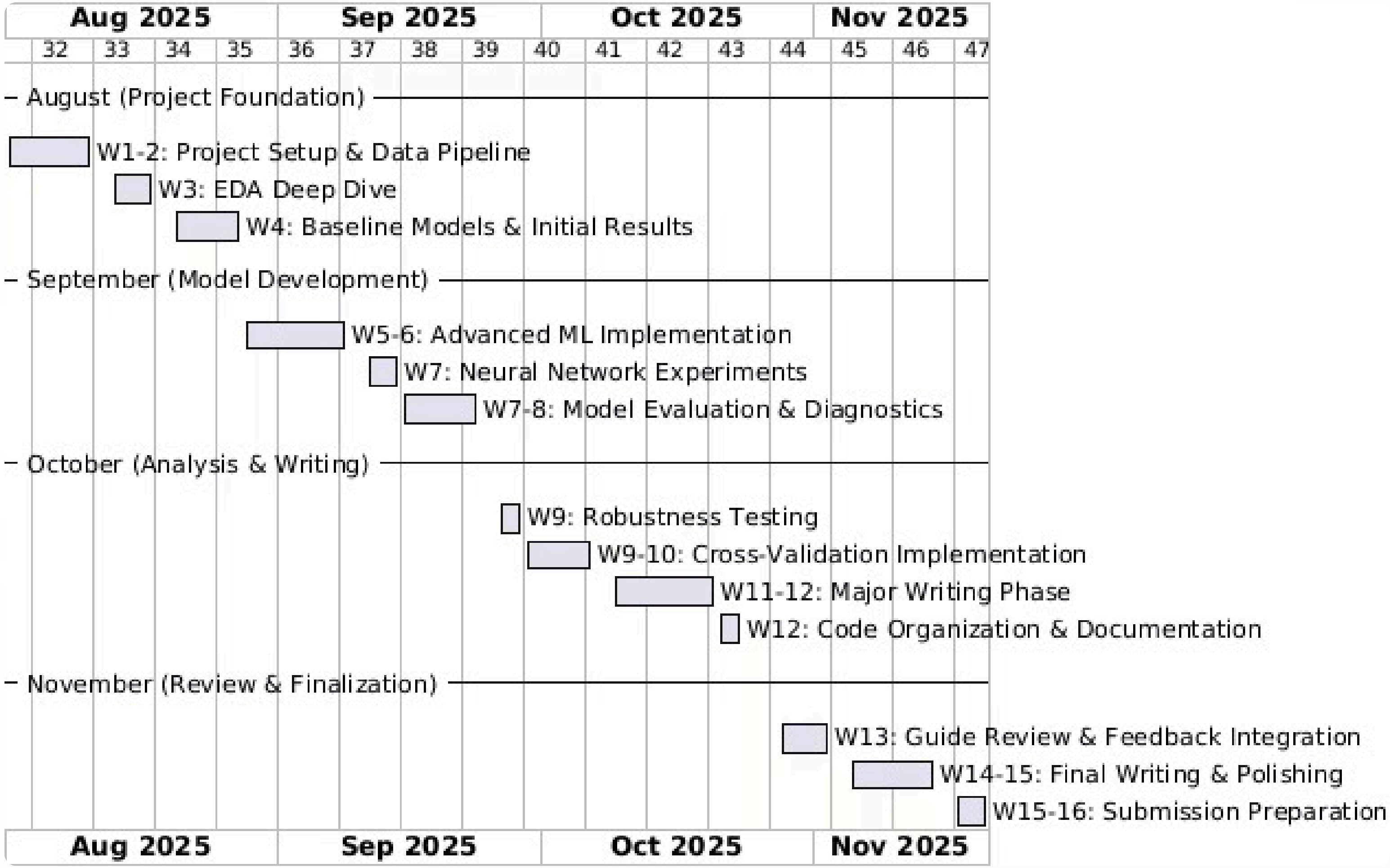
Operating System

Windows 10 Home (64-bit) providing stable platform for computational workflows

Neural Network Implementation

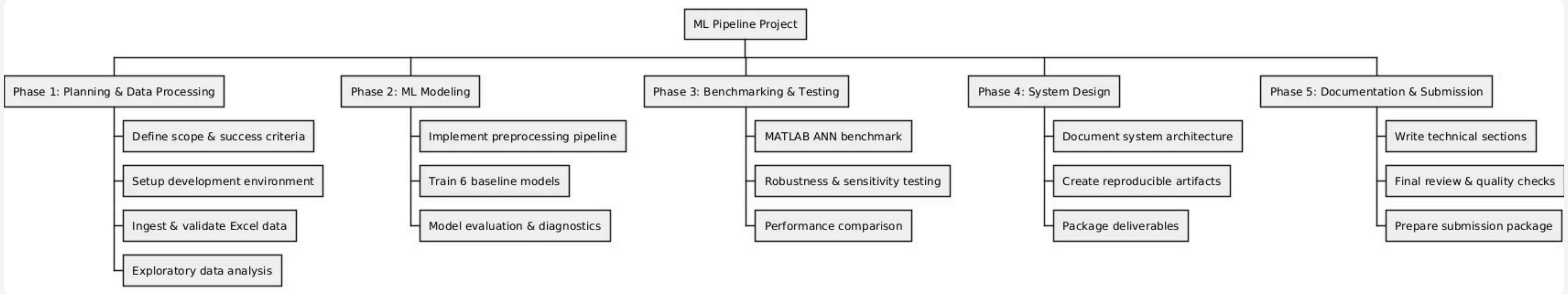
MATLAB R2025a Deep Learning Toolbox for ANN baseline comparison

Gantt Chart



Aug 2025
Sep 2025
Oct 2025
Nov 2025
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
August (Project Foundation)
W1-2: Project Setup & Data Pipeline
W3: EDA Deep Dive
W4: Baseline Models & Initial Results
September (Model Development)
W5-6: Advanced ML Implementation
W7: Neural Network Experiments
W7-8: Model Evaluation & Diagnostics
October (Analysis & Writing)
W9: Robustness Testing
W9-10: Cross-Validation Implementation
W11-12: Major Writing Phase
W12: Code Organization & Documentation
November (Review & Finalization)
W13: Guide Review & Feedback Integration
W14-15: Final Writing & Polishing
W15-16: Submission Preparation
Aug 2025
Sep 2025
Oct 2025
Nov 2025

Work Breakdown Structure



Phase 1

Data Preparation (data collection, cleaning, temporal ordering)

Phase 3

Evaluation & Diagnostics (compute metrics, residual analysis)

1

2

3

4

Phase 2

Model Development (implement 6 ML models, configure pipelines)

Phase 4

Analysis & Interpretation (feature importance, literature positioning)

Requirement Analysis (SRS)

Functional Requirements

- Accept 6 input features for spray prediction
- Predict 4 continuous spray outputs
- Multi-output regression support
- Temporal integrity preservation

Performance Requirements

- R^2 score $\geqslant$ 0.95 target accuracy
- Inference time $<$ 0.5 seconds
- Interpretability through feature importance
- Computational efficiency optimization

System Constraints

- Limited dataset: 726 samples
- No cross-validation implementation
- Prevention of data leakage
- Time-aware processing requirements

System Design Overview

Data Preprocessing

Handles data loading, cleaning, and ensures temporal ordering to prevent leakage.

Feature Engineering

Standardizes inputs and applies multi-output wrappers for simultaneous prediction.

Model Training

Implements 6 classical ML models + ANN baseline with time-aware train/test splits.

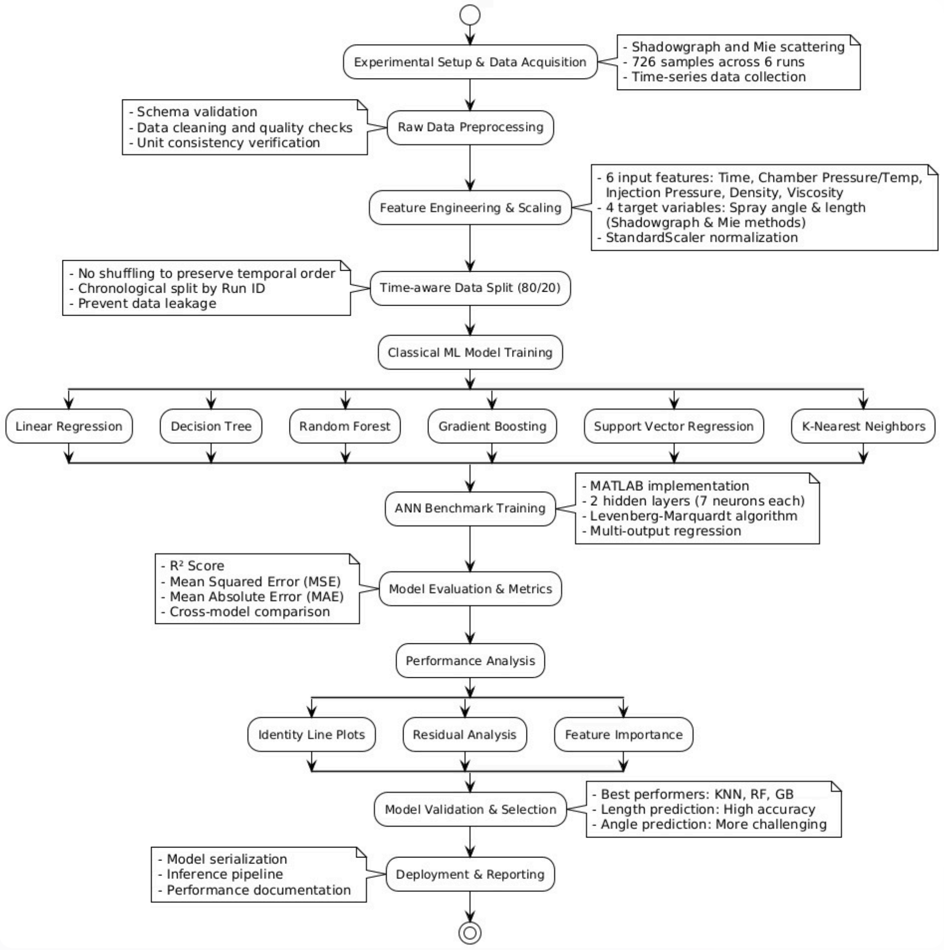
Performance Evaluation

Computes metrics (MSE, MAE, R^2) with visualization capabilities.

Model Interpretation

Provides feature importance analysis for model explainability.

Workflow Model



Implementation Results & Performance

726

Total Samples

581 training / 145 testing with temporal ordering preserved

6

ML Models

Plus ANN baseline successfully trained and evaluated

60%

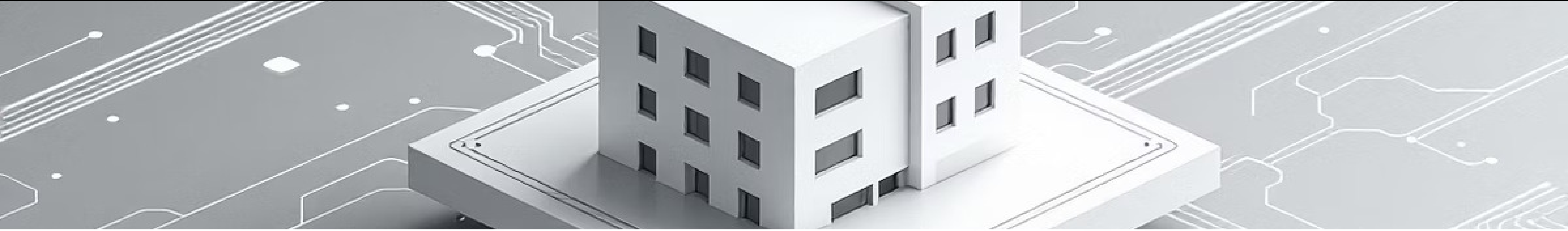
Project Complete

Core modules implemented with visualization in progress

Training Performance (seconds)

- Linear Regression: 0.023s
- K-Nearest Neighbors: 0.012s
- Decision Tree: 0.041s
- Support Vector Regression: 0.089s
- Random Forest: 0.156s
- Gradient Boosting: 0.234s
- ANN Baseline: 2.45s

Model	R2 Score (higher, the better)
Linear Regression	0.843
Decision Tree	0.948
Random Forest	0.962
Gradient Boosting	0.955
SVR	0.916
KNN	0.972
ANN Benchmark	0.998



Key Achievements & Future Directions

Project Achievements

Functional ML pipeline established with competitive classical model performance demonstrating $R^2 > 0.85$ for penetration length prediction while maintaining superior interpretability

Current Status (60% Complete)

Core implementation finished. Remaining work includes advanced visualization modules, hyperparameter optimization, and real-time deployment interface development

Future Research Directions

Extended datasets for improved generalization, cross-validation implementation, uncertainty estimation methods, and industrial deployment optimization

Key Finding: Classical machine learning models achieve competitive accuracy compared to complex neural networks while providing superior interpretability and significantly reduced computational requirements for spray characteristic prediction

Thank You

Questions & Answers