Real-Time Control of Microfluidic Mixing via Machine Learning-Driven Prediction

Seonghwan Lim[1] (slim24@hawk.illinoistech.edu)

Abhinav Bhushan[2] (abhushan@illinoistech.edu)

[1] Department of Computer Science, Illinois Institute of Technology, Chicago, IL

[2] Department of Biomedical Engineering, Illinois Institute of Technology, Chicago, IL

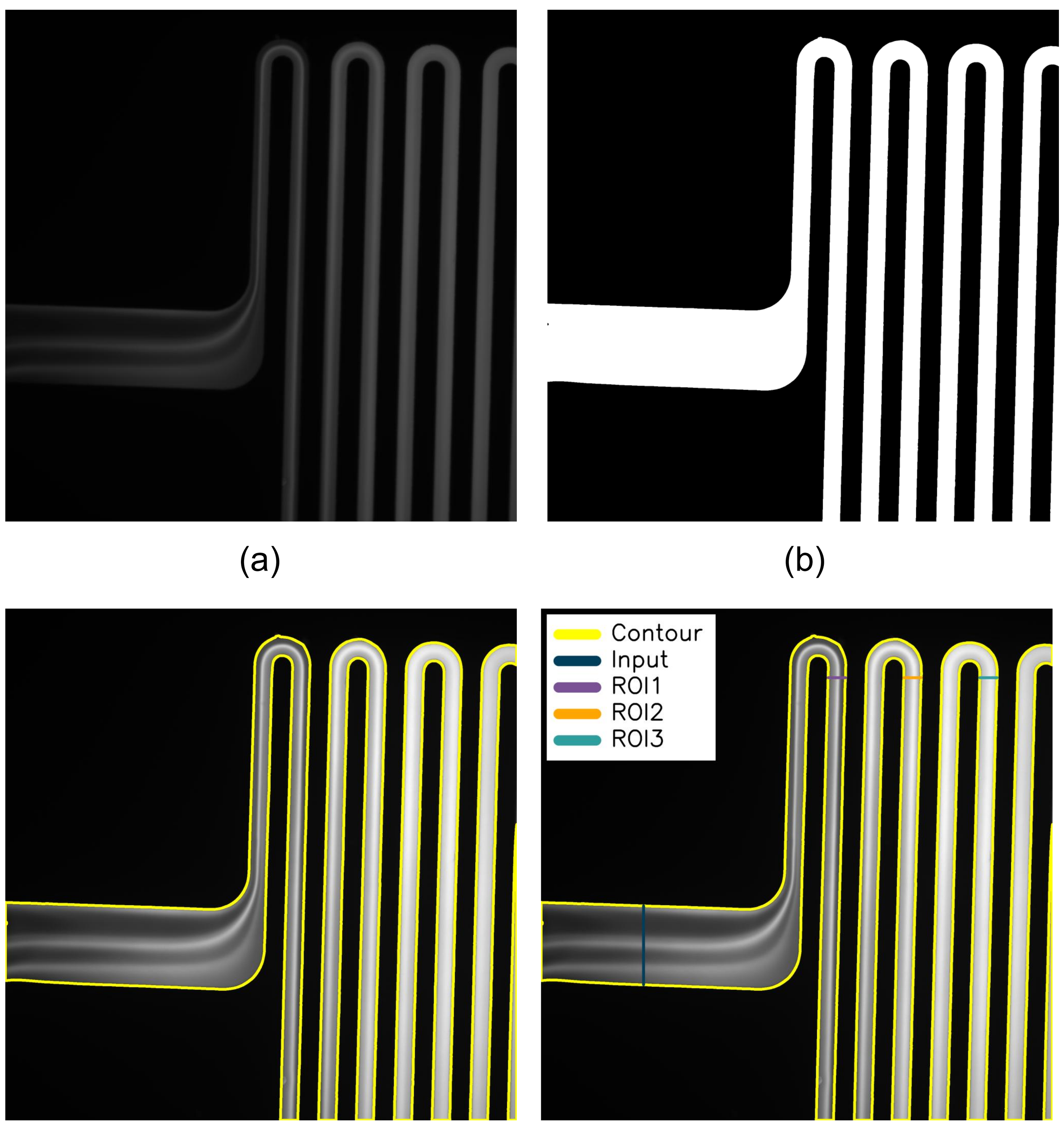
INTRODUCTION

- Microfluidic technology enables precise control of fluids for biomedical applications, such as point-of-care diagnostics and targeted drug delivery.
- Real-time monitoring of mixing rates under laminar flow is challenging due to diffusion-dominated mixing and the limitations of conventional methods (fluorescence assays, syringe pumps).
- This study introduces:
  - **Non-invasive, image-based mixing measurements**, using pixel intensity as a proxy for concentration.
  - **Machine-learning models** trained on these measurements to predict inlet inputs and enable active control of mixing rate.
- Simulation data with known inlets and experimental videos without inlet measurements demonstrate the feasibility of a data-driven control system that improves precision and responsiveness in microfluidic devices.

METHODS

ROI Extraction and Concentration Measure

- Grayscale frames of the mixing channel were processed using contour detection to define regions of interest (ROIs).
- Pixel intensity was regarded as a proxy for local solute concentration.
- The mixing degree for each ROI was defined as the arithmetic mean of pixel intensities within the ROI.

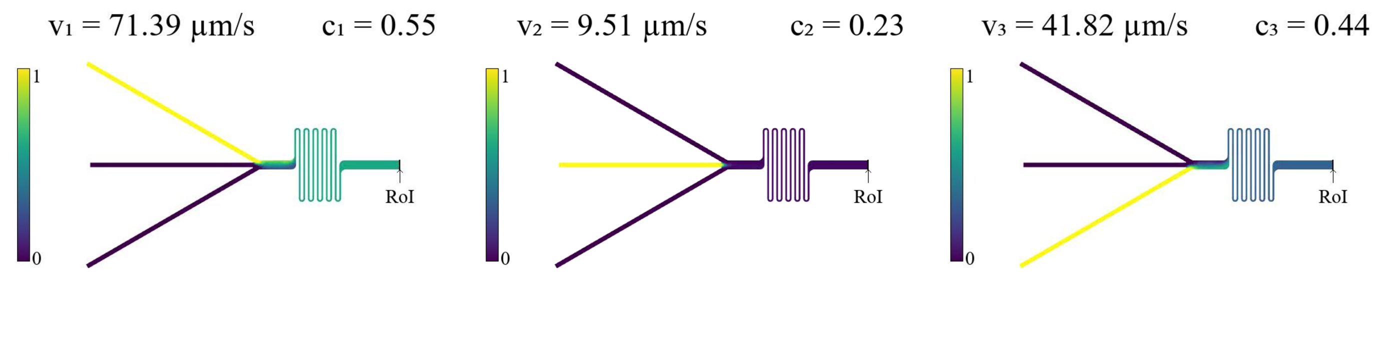


(a) (b) (c) (d)

Figure 1. (a) Grayscale image. (b) Binarized image. (c) Image with contours extracted from the binarized image. (d) Regions of interest (ROIs) for concentration measurement, with one vertical ROI and three horizontal ROIs indicated after the curved section.

Simulation Dataset (N = 100)

- Data were generated from a three-channel ANSYS Fluent simulation of laminar flow with diffusion.
- A single downstream outlet ROI was used to obtain the average concentration after mixing.
- Ground truth consisted of inlet flow rates from the three input channels.
- The task was to predict inlet velocities from the measured outlet concentration.



(a) (b) (c)

Figure 2. Representative simulation results showing fluid concentration distributions in the microfluidic channel at different inlet velocities (v_1 = 71.39 µm/s, v_2 = 9.51 µm/s, v_3 = 41.82 µm/s). Corresponding concentration values (c_1, c_2, c_3) were extracted from the designated region of interest (ROI). Color gradients indicate normalized concentration on a 0–1 scale.

Experimental Dataset (n = 47)

- Data consisted of two-channel microfluidic mixing videos.
- The initial ROI mixing rates were used as surrogate inputs in place of directly measured inlet velocities.
- One vertical inlet ROI at the mixing front (v1) and three downstream ROIs (h1–h3) were extracted.
- The task was to predict the initial mixing concentration (v1) from the downstream ROIs.

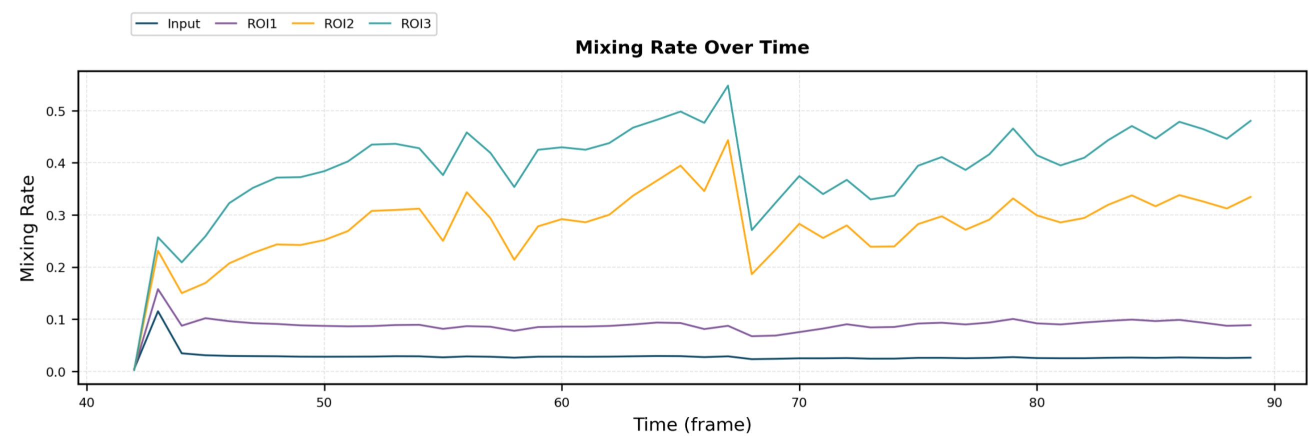


Figure 2. Mixing rate over time from experimental video frames, showing normalized mean pixel intensity for the inlet (Input) and downstream ROIs (ROI1–ROI3).

Machine Learning Models

- Models included a mean baseline predictor, Ridge regression as a linear regularized approach, and Random Forest and Gradient Boosting as ensemble approaches.
- These were chosen to compare simple linear models with non-linear ensemble methods under different data conditions.

Evaluation Metrics

- Both datasets were divided into training and test sets for model evaluation.
- Performance was assessed using mean absolute error (MAE), mean relative error (MRE), and root mean squared error (RMSE).

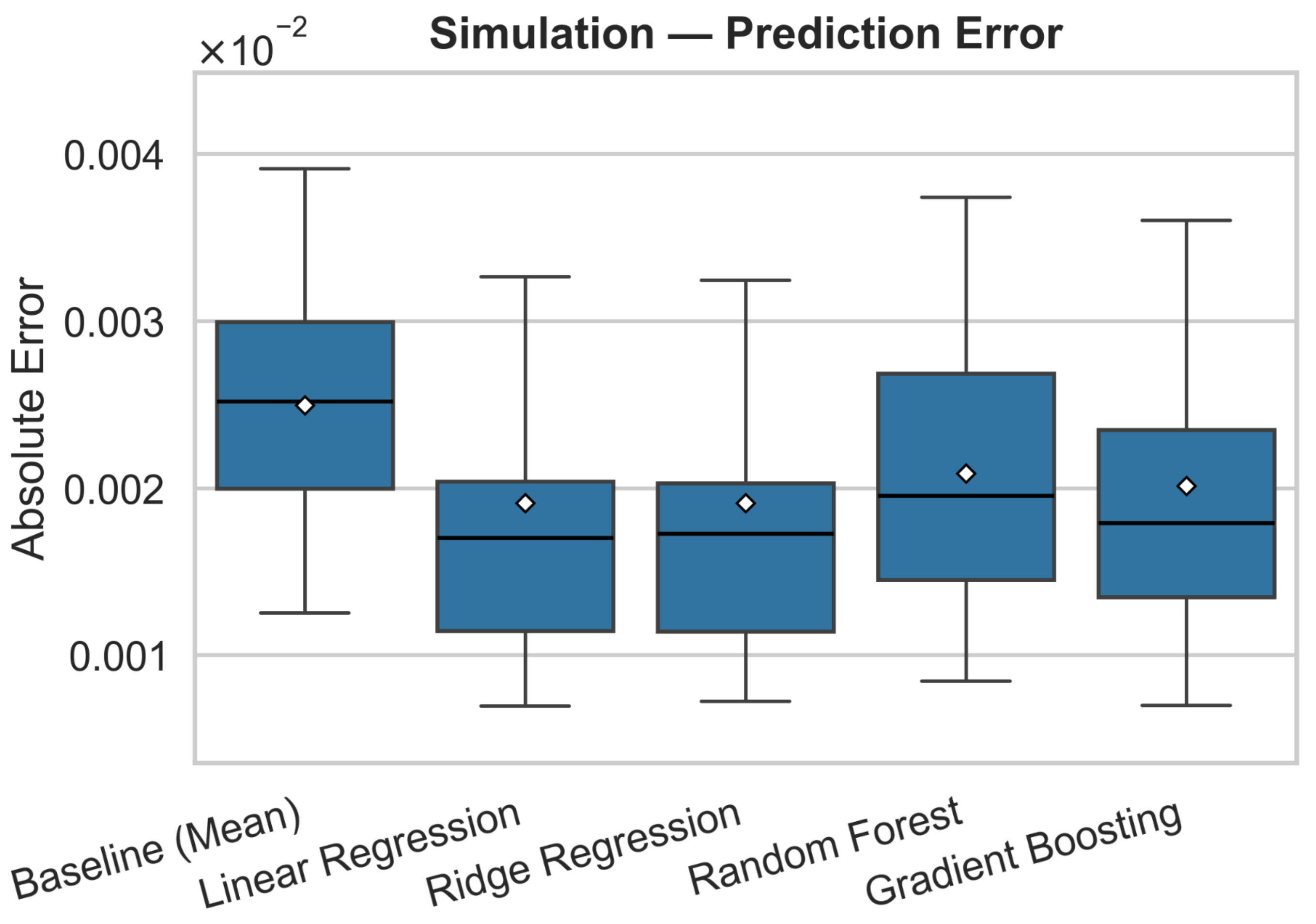
RESULTS

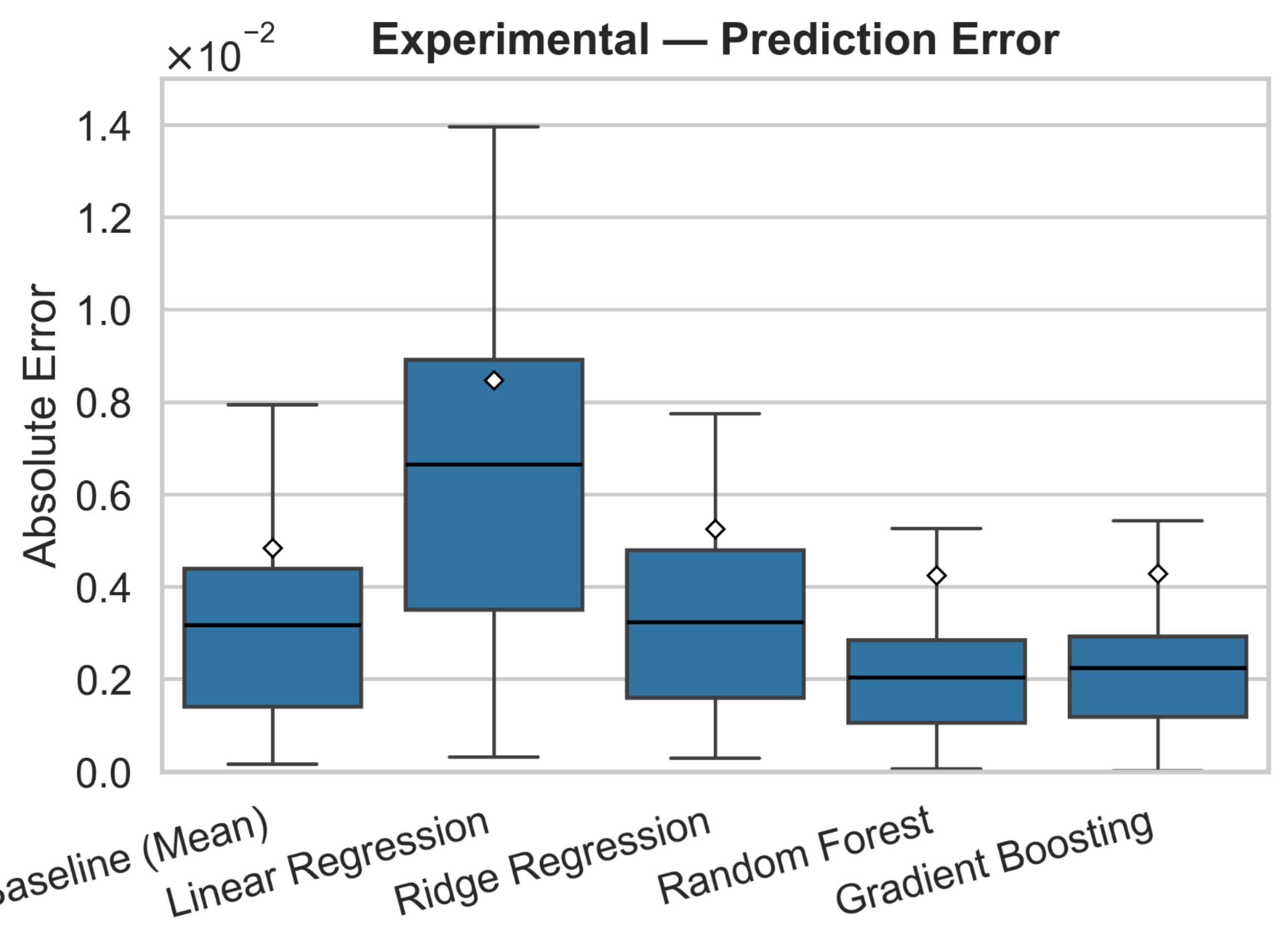
Simulation Results (N = 100)

Model	MAE $(\times 10^{-5})$	MRE	RMSE $(\times 10^{-5})$
Baseline (Mean)	2.50	1.59	2.88
Linear Regression	**1.91**	**0.60**	**2.32**
Ridge Regression	**1.91**	0.61	**2.32**
Random Forest	2.09	0.95	2.48
Gradient Boosting	2.01	0.70	2.50

Experimental Results (N = 47)

Model	MAE $(\times 10^{3})$	MRE	RMSE $(\times 10^{3})$
Baseline (Mean)	4.77	0.13	8.35
Linear Regression	8.37	0.27	12.39
Ridge Regression	5.16	0.14	8.58
Random Forest	**4.12**	**0.10**	**7.81**
Gradient Boosting	4.16	**0.10**	7.78





DISCUSSION

- **Simulation:** The linear regression model achieved the lowest MAE and RMSE, reflecting the near-linear relationship between inlet flows and outlet concentration under controlled conditions.
- **Experiment:** Since individual inlet flow rates were not available, early-stage mixing signals were used as surrogate inputs. Under these conditions, Random Forest achieved the best performance.
- **Task difference:** Simulation predicts inlet flow rates directly, while Experiment predicts early mixing dynamics without explicit inlet inputs.
- **Metrics:** R^2 was excluded for experiments due to instability under temporal variation; MAE and MRE were more reliable for evaluation.
- **Insight:** The linear models are suitable for structured simulations with explicit inputs. Ensemble models (e.g., Random Forest) were better suited for the experimental dataset, where individual inlet velocities or flow rates were unavailable and only indirect mixing signals could be used as inputs.

ILLINOIS TECH