

LB/TH/15/2026
TH6172

**ROBUST REGRESSION AS A BENCHMARK FOR
REGRESSION BASED FEDERATED LEARNING**

Karunarithna J.H.S.P

219441K

Degree of Master of Science

Department of Electronic and Telecommunication
Faculty of Engineering

University of Moratuwa
Sri Lanka

April 2025

ROBUST REGRESSION AS A BENCHMARK FOR REGRESSION BASED FEDERATED LEARNING

Karunarathna J.H.S.P

219441K

Dissertation submitted in partial fulfillment of the requirements for the
degree
Degree of Master of Science

Department of Electronic and Telecommunication
Faculty of Engineering

University of Moratuwa
Sri Lanka

April 2025

DECLARATION

I declare that this is my own work and this Dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 2025.07.02

The supervisor should certify the Dissertation with the following declaration.

The above candidate has carried out research for the Degree of Master of Science Dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr.Upeka Premaratne

Signature of the Supervisor:

Date: 2025.07.23

ACKNOWLEDGEMENT

This MSc was carried out from 2021 to 2024 in the Department of Electronic and telecommunication Engineering University of Moratuwa Sri Lanka. I express my heartfelt gratitude to everyone who supported and guided me throughout this MSc journey. In particular, I extend my deepest gratitude to my supervisor, Dr. Upeka Premaratne and course coordinator, Dr. Kasun Hemachandra for their unwavering support, insightful advice, and invaluable guidance. Their expertise, dedication and encouragement have been instrumental in the successful completion of this research.

Further, I am grateful to the University of Moratuwa and the Department of Electronic and Telecommunication Engineering for providing facilities, resources and a supportive academic environment to carry out this research. Also, I would like to thank all the owners of the materials that have been referred to in this research work. Finally, I want to express my gratitude to my wife, family, friends, and colleagues for the support provided to complete this research.

ABSTRACT

This research evaluates the effectiveness of robust regression techniques, specifically Theil-Sen and Repeated Median Regression (RMR) as benchmarks for regression-based Federated Learning (FL) across a range of simulated sensor configurations. FL is a privacy-preserving paradigm that enables collaborative model training across distributed nodes without centralizing the raw data. Standard regression techniques such as linear regression, which is not robust against outliers and heterogeneity are commonly employed in real world FL environments. In contrast, RMR demonstrates the highest resilience and lowest error rates, with Theil-Sen also showing strong performance as a more computationally efficient alternative. RMR achieves the overall highest performance gains, with the best efficiency improvement with respect to the linear federated regression. The findings of the study support the use of robust regression as a dependable alternative in distributed, real-world sensor networks.

Keywords: Federated Learning, Repeated Median Regression, Linear Regression, Theil-Sen Regression

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Acknowledgement	ii
Abstract	iii
Table of Contents	iv
List of Figures	vii
List of Abbreviations	vii
1 Introduction	1
1.1 Research Problem	1
1.2 Scope and Objectives of the Research	2
1.3 Contribution	3
1.4 Organization of the thesis	4
2 Literature review	6
2.1 Introduction to Federated Learning and Robust Regression	6
2.2 Elimination of Statical bias	7
2.2.1 Sources of Statistical Bias in Federated Learning	7
2.2.2 Methods for Eliminating Statistical Bias in FL	8
2.2.3 Personalized Federated Learning	8
2.2.4 Fair Aggregation Algorithms	8
2.2.5 Differential Privacy and Bias Correction	8
2.2.6 Conclusion- Elimination of Bias Federated Learning	9
2.3 Regression	9
2.3.1 Introduction	9
2.3.2 Sources of Statistical Bias in Regression Models	9
2.3.3 Methods for Eliminating Statistical Bias in Regression	10
2.3.4 Regularization Techniques	10
2.3.5 Robust Regression Methods	10
2.3.6 Instrumental Variable Approach	10

2.3.7	Bias Correction Techniques	11
2.3.8	Fairness-Aware Regression	11
2.4	Robust regression	12
2.4.1	Fundamentals of Robust Regression	12
2.4.2	Integrating Robust Regression into Federated Learning	13
2.4.3	Methodological Advances and Practical Considerations	13
3	Methodology	15
3.1	Introduction	15
3.2	Data Generation and Noise Model	15
3.2.1	Gaussian Noise	16
3.2.2	Exponential Noise	16
3.2.3	Non-Linear Quadratic Noise	16
3.3	Data Generation	16
3.3.1	Experimental Design	17
3.3.2	Sensor Parameter Scenarios	17
3.4	Implementation of Robust Regression Algorithms	19
3.4.1	Least Squares Linear Regression	19
3.4.2	Theil-Sen Robust Regression	19
3.4.3	Repeated Median Regression	20
3.5	Federated Regression	21
3.5.1	Federated Regression Formulation	21
3.5.2	Federated Regression Techniques	22
3.6	Evaluation Metrics	25
3.6.1	Mean Absolute Error	25
3.6.2	Relative Federated Error	25
3.6.3	Efficiency Metric for Federated Regression Model Evaluation	26
4	Experimental Result	28
4.1	Results and Discussion	28
4.2	Initial Evaluation Under Uniform Sensor Parameters	28
4.2.1	Performance of Regression Techniques	30
4.2.2	Effect of Sensor Configuration Distributions	32

4.2.3	Impact of Sensor Quantity and Convergence Trends	36
4.2.4	Selecting the Best Method	36
5	Conclusion and Future Work	37
5.1	Conclusion	37
5.2	Future Work	38
	References	40

LIST OF FIGURES

Figure	Description	Page
Figure 2.1	Federated Model Architecture	7
Figure 4.1	Linear regression federated model for N=10	29
Figure 4.2	Theil-sen regression federated model for N=10	29
Figure 4.3	Repeated mean regression federated model for N=10	29
Figure 4.4	Regression models federated model for N=10	30
Figure 4.5	Relative federated error (MSE) of uniform distribution method 1	31
Figure 4.6	Relative federated error (MAE) of uniform distribution method 1	31
Figure 4.7	Relative federated error (MAE) of exponential distribution	31
Figure 4.8	Relative federated error (MSE) of exponential distribution	32
Figure 4.9	Relative federated error (MAE) of uniform distribution method 2	33
Figure 4.10	Relative federated error (MSE) of uniform distribution method 2	33
Figure 4.11	Relative federated error (MAE) of Gaussian distribution	33
Figure 4.12	Relative federated error (MSE) of Gaussian distribution	34
Figure 4.13	Efficiency comparison for exponential method across sample sizes	34
Figure 4.14	Efficiency of federated regression models	35
Figure 4.15	Efficiency improvement with respect to Theil-Sen method	36

LIST OF ABBREVIATIONS

Abbreviation	Description
FL	Federated Learning
IID	Independent and Identically Distributed
MAE	Mean Absolute Error
MSE	Mean Square Error
OLS	Ordinary Least Squares
RFE	Relative Federated Error
RMR	Repeated Median Regression