An Analysis and Early Warning of Toxic Gas Outbursts in Small-Scale Gem Mining: Model Evaluation and GIS-Based Risk Mapping in Sri Lanka

W M U Supun Wanasundara
(Reg. No.: MS24018746)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2025

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dasuni Nawinna

Approved for MSc. Research Project:

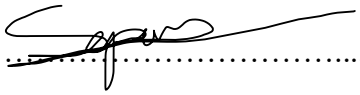
MSc in IT Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: 

W M U S WANASUNDARA

Date: 2026/01/03

ABSTRACT

An Analysis and Early Warning of Toxic Gas Outbursts in Small-Scale Gem Mining: Model Evaluation and GIS-Based Risk Mapping in Sri Lanka

W. M. U. Supun Wanasundara

MSc. in Information Technology

Supervisor: Prof. Dasuni Nawinna

December 2025

Artisanal and small-scale gem mining in Sri Lanka, particularly within the Pelmadulla region, is highly susceptible to toxic gas accumulation due to inadequate ventilation and the absence of systematic early-warning mechanisms. This research aimed to develop a predictive–GIS integrated framework for the detection and spatial mapping of toxic gas hazards in small-scale mining environments. Utilizing globally available gas sensor datasets (UCI, IEEE, Mendeley) and localized geological and spatial data, multiple predictive algorithms—Random Forest (RF), XGBoost, LSTM, GRU, TCN, and IBWO-TCN—were trained and evaluated using precision, recall, F1-score, AUROC, and lead-time metrics. The Random Forest model exhibited the highest predictive performance (F1 = 0.93, AUROC = 0.97) and was subsequently integrated with GIS-based hazard mapping for the Pelmadulla study area. The spatial analysis indicated that approximately one-third of mining sites fall within high or very high-risk zones. The findings highlight the potential of integrating predictive analytics and geospatial modeling to establish a low-cost, data-driven early-warning system, thereby enhancing occupational safety and supporting sustainable mining practices in Sri Lanka.

Keywords: Toxic gas detection, Predictive modeling, GIS-based risk mapping, Random Forest, Early-warning system, Small-scale gem mining, Sri Lanka.

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to my supervisor, Prof. Dasuni Nawinna, for her invaluable guidance, encouragement, and continuous support throughout this research. Her expert insights, constructive feedback, and patience have been instrumental in shaping the direction and quality of this study.

I am deeply thankful to my family members and relatives who provided their generous assistance during the data collection process at the gem mining sites. Their practical support and cooperation made it possible to gather essential field information under challenging conditions.

My heartfelt appreciation also goes to all individuals, colleagues, and organizations who extended their help, both directly and indirectly, during the course of this research. Their contributions, whether through technical advice, moral support, or resource facilitation, were vital to the successful completion of this work.

Finally, I extend my gratitude to the Sri Lanka Institute of Information Technology (SLIIT) for providing the academic environment, facilities, and resources necessary to carry out this study.

TABLE OF CONTENTS

DECLARATION ii
ABSTRACT iii
ACKNOWLEDGEMENT iv
TABLE OF CONTENTS v
List of Figures xi
List of Tables xii
List of Equations xiii
Chapter 1 Introduction 1
1.1 Background of the Study 1
1.2 Problem Statement 2
1.3 Research Aim and Objectives 4
1.3.1 Research Aim 4
1.3.2 Specific Objectives 4
1.4 Research Questions 4
1.4.1 Main Research Question 5
1.4.2 Sub-Questions 5
1.5 Scope of the Study 6
1.5.1 Geographical Scope 6
1.5.2 Data Scope 6
1.5.3 Methodological Scope 6
1.5.4 Temporal Scope 7
1.6 Significance of the Study 7
1.6.1 Academic Significance 8
1.6.2 Practical Significance 8
1.6.3 Socio-Economic and Policy Importance 8
1.7 Conceptual Framework 9
1.7.1 Data Inputs and Preprocessing Layer 9
1.7.2 Layer of Predictive Modeling and Evaluation 10
1.7.3 GIS-Based GIS-Based Integration and Early-Warnings Output Layer. 11
1.7.4 Conceptual Model Flow 12
1.7.5 Visual Representation (Conceptual Framework Diagram) 13
1.8 Organization of the Thesis 14
1.9 Summary of Chapter 15
Chapter 2 Literature Review 16
2.1 Introduction 16
2.2 Gas Detection in Mining 17

2.2.1 The nature of Toxic Gases and their impact 17
2.2.2 Gas Detection Technologies 18
2.2.3 Gas Detection Literature Review 20
2.3 Predictive Modeling Approaches 20
2.3.1 Conventional Statistical Methods 21
2.3.2 Machine Learning (ML) Models 21
2.3.3 Deep Learning (DL) Models 22
2.3.4 Summary of Predictive Model Comparison 23
2.3.5 Relevance to Current Study 24
2.4 Early-Warning Systems in Mining 24
2.4.1 Idea of Early-Warning Systems 24
2.4.2 Development of Early- warning System in mining 25
2.4.3 Global Best Practices 25
2.4.4 Early-Warning System Evaluation Metrics 26
2.4.5 Mining safety gaps in the mining sector of Sri Lanka 26
2.5 GIS-Based Hazard Mapping 27
2.5.1 GIS as a tool in Hazard Assessment 27
2.5.2 GIS Techniques of Risk Mapping 27
2.5.3 Global Case Studies 28
2.5.4 GIS in Sri Lanka 29
2.5.5 GIS Hazard Mapping Workflow 29
2.6 Gem Mining in Sri Lanka: Context and Challenges 30
2.6.1 There are historical and economic significance 30
2.6.2 There were two operational forms of Pelmadulla Mines listed as 30
2.6.3 Regulatory Landscape and Safety 31
2.6.4 Data and Research Gaps 31
2.6.5 Summary 32
Chapter 3 Methodology 33
3.1 Introduction 33
3.2 Research Design 34
3.2.1 Research Type and Approach 34
3.2.2 Research Framework 35
3.2.3 Research Philosophy 36
3.2.4 Research Process Overview 36
3.2.5 Research Instruments and Tools 37
3.2.6 Ethical Considerations 37
3.2.7 Reliability and Validity 37

3.3 Study Area: Pelmadulla, Rathnapura 38
3.3.1 Geographic Location and the Physical Features 38
3.3.2 Socio-Economic Context 38
3.3.3 Geological and Environmental Conditions 39
3.3.4 Relevance of Pelmadulla as Study Area 39
3.4 Data Collection 40
3.4.1 Overview 40
3.4.2 Global Gas Sensor Datasets 40
3.4.3 Local Spatial and Contextual Data 41
3.4.4 Data Integration and Storage 42
3.5 Data Preprocessing 43
3.5.1 Data Cleaning 43
3.5.2 Data Transformation and Synchronization 44
3.5.3 Feature Engineering 45
3.5.4 Event Labeling 45
3.5.5 Dimensionality Reduction 46
3.5.6 Dataset Partitioning 46
3.6 Model Development 47
3.6.1 Overview 47
3.6.2 Machine Learning Models 48
3.6.3 Deep Learning Models 50
3.6.4 Model Optimization – IBWO-TCN 51
3.6.5 Model Training and Model Execution Environment 52
3.7 Model Evaluation 52
3.7.1 Purpose of Model Evaluation 52
3.7.2 Evaluation Metrics 53
3.7.3 Validation Strategies 54
3.7.4 Handling Class Imbalance 55
3.7.5 Statistical Significance Testing 55
3.7.6 Visualization of Model Performance 55
3.7.7 Model Selection 56
3.7.8 Relevance to Pelmadulla 56
3.8 GIS-Based Risk Mapping 56
3.8.1 Purpose of GIS Integration 57
3.8.2 Spatial Data Sources 57
3.8.3 Integration of Predictive Model Outputs 58
3.8.4 Spatial Interpolation Procedures 59

3.8.5 Risk Classification 59
3.8.6 Validation of GIS Outputs 60
3.8.7 Application of GIS Maps 60
3.8.8 Visualization of an example (to go in Chapter 4) 61
3.9 Conceptual Workflow 61
3.9.1 Rationale 61
3.9.2 Workflow Stages 61
3.9.3 Conceptual Framework Diagram Description 62
3.9.4 Implementation Platform 63
3.9.5 Summary of Methodology 63
Chapter 4 Results And Findings 64
4.1 Introduction 64
4.1.1 Overview of the Analytical Process 64
4.1.2 Visual Overview 65
4.1.3 Model Comparison Overview 65
4.1.4 Model Performance Comparison 66
4.1.5 Receiver Operating Characteristic (ROC) Analysis 67
4.1.6 PrecisionRecall Curve Analysis 68
4.1.7 Confusion Matrix Analysis 69
4.2 Statistical Validation and Reliability 69
4.2.1 Cross-Validation Consistency 70
4.2.2 Statistical Significance Tests 70
4.2.3 Importance of Features Analysis 71
4.3 GIS-Based Risk Mapping 72
4.3.1 Purpose and Overview 72
4.3.2 Spatial Data Integration 72
4.3.3 Surface Generation and Interpolation 72
4.3.4 The classification of Hazard Zones is given in 73
4.3.5 GIS Hazard Map of Pelmadulla 73
4.3.6 Hazard Probability Distribution 74
4.3.7 Kriging Variance Map (Uncertainty Analysis) 75
4.3.8 Geologic Relationship with Geological Features 75
4.3.9 Multi-Layer Visualization 76
4.3.10 Spatial Validation 77
4.4 Conceptual Diagram 77
4.4.1 Purpose 78
4.4.2 Framework Overview 78

4.4.3 The potential of future implementation is..... 79
4.5 Summary of Findings..... 79
4.5.1 Model Findings..... 79
4.5.2 GIS Findings..... 80
4.5.3 Resolution of Frames Findings..... 80
4.5.4 Policy and Practice Implications..... 80
4.5.5 Chapter Summary..... 81
Chapter 5 Discussion..... 82
5.1 Introduction to the Discussion..... 82
5.2 Interpretation of Predictive Model Findings..... 82
5.2.1 Overview of Model Outcomes..... 83
5.2.2 Why Random Forest Outperformed Other Models..... 83
5.2.3 Model Comparison with Literature..... 84
5.2.4 Insights from Feature Importance..... 85
5.2.5 Implications for Real-Time Implementation..... 86
5.3 Discussion on GIS and Spatial Hazard Insights..... 86
5.3.1 Role of GIS in Understanding Risk Distribution..... 86
5.3.2 Relationship Between Topography and Gas Accumulation..... 87
5.3.3 Geological Correlation Analysis..... 88
5.3.4 Validation with Historical and Expert Data..... 88
5.3.5 Uncertainty and Improvement Areas..... 88
5.4 Integration of Predictive and Spatial Insights..... 89
5.4.1 Overview of Integration..... 89
5.4.2 Complementarity of Predictive and Spatial Models..... 89
5.4.3 Theoretical Contribution of Integration..... 90
5.4.4 Operational Value of Integration..... 91
5.4.5 Integration Validation..... 91
5.5 Policy, Technical, and Socio-Economic Implications..... 92
5.5.1 Policy Implications..... 92
5.5.2 Technical Implications..... 92
5.5.3 Socio-Economic Implications..... 93
5.5.4 Academic Implications..... 94
5.6 Limitations and Future Work..... 94
5.6.1 Key Limitations..... 94
5.6.2 Recommendations for Future Research..... 95
5.6.3 Final Statement of Research Contribution..... 96
Chapter 6 Conclusion and Recommendations..... 98

6.1 Conclusion 98
6.1.1 Overview 98
6.1.2 Achievement of Research Objectives 98
6.1.3 Theoretical and Methodological Contributions 99
6.1.4 Empirical Findings 100
6.1.5 Broader Significance 100
6.2 6.2 Recommendations 101
6.2.1 To Government and Regulatory Bodies (NGJA, GSMB, NBRO) 101
6.2.2 For Mining Operators and Cooperative Societies 102
6.2.3 For Researchers and Academic Institutions 103
6.2.4 For Development Agencies and Donors 103
6.3 Future Prospects and Final Remarks 104
6.3.1 Future System Implementation 104
6.3.2 Long-Term Vision 104
6.3.3 Reflection on Knowledge Advancement 105
6.3.4 Concluding Statement 105
Bibliography 107
Appendix 110
Appendix 1: Cloud Run Functions and Cloud Storages 110
Appendix 2: Realtime Database – Firebase 111
Appendix 3: IOT Code Snip 112
Appendix 4: Backend Module (ML) 113
Appendix 5: Google Colab 113

List of Figures

Figure 3.1: Methodological Framework..... 35
Figure 3.2: Data Preprocessing..... 43
Figure 3.3: Research Methodology for early warning model for toxic gas outburst..... 47
Figure 3.4: Symbols and Meanings..... 49
Figure 3.5 : E.W.F for toxic gas outburst in small-scale gem mining operations in Pelmadulla..... 51
Figure 3.6: Terms and Meaning of IDW..... 59
Figure 4.1: Model Performance Comparison..... 66
Figure 4.2: Best three models (RF, XGBoost, IBWO-TCN) with the ROC curves..... 67
Figure 4.3: PrecisionRecall Curve Analysis..... 68
Figure 4.4: Confusion Matrix Analysis..... 69
Figure 4.5: Cross-Validation Consistency..... 70
Figure 4.6: Importance of Features Analysis..... 71
Figure 4.7: GIS Hazard Map..... 73
Figure 4.8: Model Comparison Overview..... 74
Figure 4.9: Kriging Variance Map..... 75
Figure 4.10: Multi-Layer Visualization..... 76
Figure 4.11: Conceptual framework for early warning..... 77
Figure 6.1: Cloud Run Functions..... 110
Figure 6.2: Cloud Run Function Overview..... 110
Figure 6.3: Cloud Run Function Code..... 110
Figure 6.4: Cloud Storage (sensor data)..... 111
Figure 6.5: Modified Realtime Database – Firebase..... 111
Figure 6.6: Code for IOT Part..... 112
Figure 6.7: Backend Program for Machine Learning (ML)..... 113
Figure 6.8: Data Loading section code Snip - Colab..... 113
Figure 6.9: Feature creating for baseline Models..... 114
Figure 6.10: Setup Sequences for Deep Learning Models..... 114

List of Tables

Table 2.1: Gas Detection 20
Table 2.2 : Reductive Models Comparisons 23
Table 3.1: Local Spatial and Contextual Data 41
Table 3.2: Feature Engineering 45
Table 3.3: Event Labeling 45
Table 3.4: Evaluation Metrics 53
Table 3.5: Spatial Data Sources 57
Table 4.1: Model Comparison Overview 65
Table 5.1: Model Comparison with Literature 84
Table 5.2: Complementarity of Predictive and Spatial Models 89

List of Equations

Equation 3.1: Logistic Regression probability equation 48
Equation 3.2: general form of a regularized loss function 49
Equation 3.3: Inverse Distance Weighted (IDW) Interpolation 59