

TABLE OF CONTENTS

UNDERGRADUATE THESIS	i
APPROVAL PAGE	ii
DECLARATION PAGE	iii
PREFACE	iv
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ATTACHMENTS	xii
ABSTRACT	xiii
INTISARI	xiv
CHAPTER I INTRODUCTION	1
I.1 Background	1
I.2 Research Objectives	15
I.3 Research Benefits	16
CHAPTER II LITERATURE REVIEW AND HYPOTHESIS FORMULATION	18
II.1 Literature Review	18
II.1.1 Ionic liquids as green solvents for CO ₂ capture	18
II.1.2 Quantum chemical descriptors and molecular properties	20
II.1.3 Machine learning algorithms for property prediction	22
II.1.4 Quantitative structure-property relationship modeling	25
II.1.5 Machine learning for CO ₂ solubility prediction in ionic liquids	26
II.1.6 Comparative algorithm performance	27
II.2 Hypothesis Formulation	27
II.2.1 Algorithm performance hierarchy	27
II.2.2 Neural network performance	27
II.2.3 Descriptor importance hierarchy	28
II.2.4 Operational parameter significance	28
II.2.5 Underlying rationale	28
CHAPTER III RESEARCH METHODS	29
III.1 Materials	29
III.1.1 Experimental data sources	29
III.1.2 Ionic liquid library	29
III.1.3 Quantum chemical descriptors	31
III.1.4 Operational parameters	31
III.1.5 Target variable	31
III.2 Apparatus	32
III.2.1 Computing hardware	32
III.2.2 Software environment	33
III.2.3 Development environment configuration	34
III.3 Procedure	35
III.3.1 Data acquisition and preprocessing	35

III.3.2	Feature engineering and dataset preparation	38
III.3.3	Machine learning model development and training	39
III.3.4	Model evaluation and comparison	41
III.3.5	Statistical analysis and comparative validation	42
III.4	Schematic	43
CHAPTER IV RESULTS AND DISCUSSION		45
IV.1	Results	45
IV.1.1	Model performance metrics on hold-out test set	45
IV.1.2	Parity plots and residual plot	45
IV.1.3	Feature importance analysis	47
IV.1.4	Hypothesis testing results	48
IV.1.5	Extrapolative screening of bio-based ionic liquids: [BMIM][Oleate] and [BMIM][Laurate]	49
IV.2	Discussion	49
IV.2.1	Nonlinear machine learning methods superior performance compared to linear regression	49
IV.2.2	Pressure and temperature as dominant solubility determinants	51
IV.2.3	Mechanistic foundation and interpretation of quantum chemical descriptors	52
IV.2.4	Literature-based data acquisition strategy and methodological justification	54
IV.2.5	Ion decomposition into individual cations and anions: computational tractability	55
IV.2.6	DFT methodology selection: B3LYP/6-31G(d,p) as optimal Balance	56
IV.2.7	Feature standardization in machine learning pipeline	56
IV.2.8	Stratified data partitioning strategy for cross-validation	57
IV.2.9	Multi-algorithm comparison framework	59
IV.2.10	Complementary performance evaluation metrics	60
IV.2.11	Feature importance analysis and mechanistic understanding	61
IV.2.12	Hypothesis-driven research framework	62
IV.2.13	Performance achievement relative to published literature	62
IV.2.14	Model generalization, applicability scope, and extrapolation limitations	63
IV.2.15	Pressure and temperature as fundamental solubility controls	64
IV.2.16	Chemistry in-depth explanation	65
IV.2.17	Computational analysis of top-performing ionic liquids	84
IV.2.18	Interpretation of bio-based ionic liquids [BMIM][Oleate] and [BMIM][Laurate]	93
CHAPTER V CONCLUSION AND SUGGESTION		101
V.1	Conclusion	101

V.1.1	Comprehensive literature dataset compilation	101
V.1.2	Quantum chemical descriptor generation through DFT	101
V.1.3	Comprehensive feature matrix integration	102
V.1.4	Multi-algorithm machine learning model development	102
V.1.5	Standardized performance evaluation metrics	103
V.1.6	Optimal algorithm identification and performance benchmarking	103
V.1.7	Major research findings	104
V.1.8	Research contributions and significance	105
V.1.9	Research limitations and boundary conditions	105
V.2	Suggestion	106
REFERENCES		107

LIST OF TABLES

Table III.1	Descriptors used	31
Table III.2	Computing hardware details	32
Table III.3	Software environment details	34
Table III.4	Sample input list of ionic liquid components	36
Table III.5	Processed output showing generated quantum descriptors	37
Table IV.1	Performance metrics of four machine learning algorithms on independent hold-out test set	45
Table IV.2	Feature importance rankings for ridge regression	47
Table IV.3	Feature importance rankings for random forest	47
Table IV.4	Feature importance rankings for gradient boosting	48
Table IV.5	Feature importance rankings for neural network	48
Table IV.6	Summary of hypothesis testing results	48
Table IV.7	Quantitative performance comparison of bio-based ionic liquids	49
Table IV.8	Comparison of quantum chemistry methods for descriptor generation	56
Table IV.9	Dataset partitioning strategy: two-step hierarchical splitting process	58
Table IV.10	Performance comparison with published literature	62
Table IV.11	Cation categories and characteristics	65
Table IV.12	Anion categories used in this study	66
Table IV.13	Top-performing IL combinations from this study	79
Table IV.14	Design rules applied to cations/anions in this study	82

LIST OF FIGURES

Figure I.1	Global CO ₂ emissions trend (1960-2024) and atmospheric CO ₂ concentration rise with annotations showing 37 Gt/year in 2023 and 420 ppm current level	1
Figure I.2	Global greenhouse gas emissions in 2023 (United Nations Environment Programme, 2024)	5
Figure I.3	Country-specific CO ₂ emission in the ASEAN region in 2019 (Rehman et al., 2024)	6
Figure I.4	Source-wise CO ₂ emissions from the electricity and heat sector in ASEAN countries (Rehman et al., 2024)	6
Figure II.1	Flowsheet of CO ₂ capture with ionic liquids (Zhang et al., 2012)	19
Figure II.2	Machine learning workflow overview (Kittichotsatsawat et al., 2022)	23
Figure III.1	Schematic diagram of the procedures	44
Figure IV.1	Parity plots (predicted vs. actual CO ₂ solubility) for (a) Ridge Regression; (b) Random Forest; (c) Gradient Boosting; and (d) Neural Network	45
Figure IV.2	Residual plots (residuals vs. predicted CO ₂ solubility) for (a) Ridge Regression; (b) Random Forest; (c) Gradient Boosting; and (d) Neural Network	46
Figure IV.3	Top 10 most important quantum descriptors for (a) Ridge regression; (b) Random forest; (c) Gradient boosting; and (d) Neural network	47
Figure IV.4	Predicted CO ₂ solubility isotherms at 298.15, 313.15, and 333.15 K for (a) [BMIM][Oleate]; and (b) [BMIM][Laurate]	49
Figure IV.5	HOMO isosurface of the [BMIM] ⁺ cation	67
Figure IV.6	LUMO isosurface of the [BMIM] ⁺ cation	68
Figure IV.7	Electrostatic potential map of [BMIM][Cl] ion pair	69
Figure IV.8	HOMO isosurface of [Tf ₂ N] ⁻	71
Figure IV.9	Pyridinium cation core in [PMPY][Tf ₂ N]	72
Figure IV.10	Best-performing IL combination - [BMIM][Tf ₂ N] (maximum capacity)	73
Figure IV.11	Industrial-scale IL combination – [BMIM][Cl] (cost-effective alternative)	74
Figure IV.12	Thermally stable IL combination – [PMPY][Tf ₂ N] (high-temperature operation)	75
Figure IV.13	NBO interaction diagram for [BMIM][Cl]	84
Figure IV.14	Simulated IR spectra of [BMIM][Cl], [BMIM][Tf ₂ N], and [PMPY][Tf ₂ N]	86
Figure IV.15	NCI plot for [BMIM][Tf ₂ N]–CO ₂ complex	88
Figure IV.16	RDF comparison	89
Figure IV.17	Box-and-whisker plot of cation solubility distribution	92
Figure IV.18	2D structural formula of [BMIM][Oleate]	94
Figure IV.19	2D structural formula of [BMIM][Laurate]	95
Figure IV.20	electron density isosurface for oleate anion	96

Figure IV.21 Electron density isosurface for laurate anion	97
Figure IV.22 HOMO isosurface of oleate anion	98
Figure IV.23 HOMO isosurface of laurate anion	99

LIST OF ATTACHMENTS

Attachment 1 List of Ionic Liquids	122
Attachment 2 Sample Computational Output	136
Attachment 3 Machine Learning Hyperparameters	139
Attachment 4 Source Code	141