

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

Comprehensive Chemical Risk Prediction Model

Group 32 Thesis by

Shabna Ilmi – 2409661

Senuth Perera – 2409663

Kalana Kannangara – 2409090

Loganathan Thusharkanth – 2410213

Supervised by

Mr. Imesh Pathirana

Submitted in partial fulfilment of the requirements for the BSc (Hons) in
Artificial Intelligence and Data Science degree at the Robert Gordon
University.

April 2025

© The copyright for this project and all its associated products resides with
Informatics Institute of Technology

Declaration

I declare that this is our own research thesis, and this thesis does not incorporate without acknowledgement any material previously published submitted for a Degree or Diploma in any other university or institute of higher learning and to the best of our 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.

Student Name	RGU ID	IIT ID	Signature
Shabna Ilmi	2409661	20230079	
Senuth Perera	2409663	20230317	
Kalana Kannangara	2409090	20232632	
Loganathan Thusharkanth	2410213	20233168	

I have read the final thesis and it is in accordance with the approved university final thesis outline.

Signature of the supervisor:



Date: 06/04/2025

ABSTRACT

The Comprehensive Chemical Risk Prediction Model (CCRPM) is an AI-driven system designed to enhance the safety, security, and regulatory compliance of hazardous chemical management across industrial, defense, and research sectors. By leveraging machine learning, the CCRPM assesses risks at critical stages of the chemical lifecycle importation, end-user handling, recipe formulation, and future import forecasting to prevent accidents, environmental harm, and illicit use.

Inspired by Sri Lanka's National Authority for the Chemical Weapons Convention (NACWC), the system addresses gaps in monitoring explosive chemicals, from procurement to end-user distribution. Key functionalities include:

1. Importation Risk Prediction: Screening high-risk chemical imports.
2. End-User Risk Profiling: Evaluating buyers (industries, researchers) for misuse potential.
3. Recipe Risk Analysis: Preventing synthesis of dangerous compounds.
4. Predictive Import Monitoring: Forecasting trends in future chemical transactions.

The CCRPM integrates analytics and dynamic risk visualization (e.g., color-coded alerts) to empower authorities and industries in proactive decision-making. Developed as a full-stack solution (Flask backend, React frontend), it demonstrates the potential of AI to transform chemical safety protocols, aligning with global standards like the Chemical Weapons Convention (CWC). This project not only mitigates disasters but also supports Sri Lanka's defence initiatives in adopting predictive risk models.

Keywords: Chemical risk prediction, End user, Importer, Chemical recipe, future trends, Chemical weapons, AI in chemical safety, regulatory compliance, hazardous material management, machine learning for security.

ACKNOWLEDGEMENT

Enthusiasm, determination, and teamwork have always played a key role in the success of our Project. However, completing this project would not have been possible without the support and guidance of a lot of individuals.

We would like to express our gratitude towards our Module leader, Mr. Prasan Yapa for guidance and providing necessary information and resources for this project.

We are highly indebted to our supervisor, Mr. Imesh Pathirana for his guidance and supervision.

We would like to express our sincere thanks to Lecturer, Ms. Gangulie Ranawaka for her valuable guidance and support in completing our project.

We would also be thankful to our lecturer, Mr. Kanishka Hewageegana for providing feedback in completion of this project.

Besides, we would like to extend our deep appreciation to Lecturer, Mr. Prasanna Pathmanathan for sharing vast knowledge, which gives us a huge opportunity to make this project a success.

Last but not least, we would like to thank our parents and friends for always being with us and supporting us in every situation.

Finally we apologize to all others unnamed, who helped us in various ways to complete this project.

Content

Declaration	i
ABSTRACT	ii
ACKNOWLEDGEMENT	iii
Content	iv
List of Figures	ix
List of Tables.....	x
1 Chapter 1: INTRODUCTION.....	1
1.1 Chapter Overview.....	1
1.2 Problem Domain.....	1
1.3 Problem Statement	2
1.4 Research Motivation.....	2
1.5 Existing Work.....	2
1.6 Research Gap.....	4
1.7 Contribution to the body of knowledge.....	5
1.7.1 Technological Contribution	5
1.7.2 Contribution to domain	5
1.8 Research challenge	6
1.9 Research Question	6
1.10 Research Aim.....	7
1.11 Research objective	7
1.12 Project scope	8
1.12.1 In-Scope	8
1.12.2 Out-Scope.....	8
1.12.3 Feature Prototype	9
1.13 Resource Requirement.....	10
1.13.1 Hardware Requirements	10
1.13.2 Software Requirements	10
1.13.3 Data Requirements	10
1.13.4 Skill Requirements	11
1.14 Chapter Summary	11
2 Chapter 2: LITERATURE REVIEW	12
2.1 Chapter Overview.....	12
2.2 Concept Map	12

2.3	Problem Domain.....	13
2.4	Existing Work.....	13
2.4.1	Importation Risk Prediction	13
2.4.2	End-user Risk Prediction.....	15
2.4.3	Chemical Recipe Analysis and Risk Prediction.....	16
2.4.4	Future Chemical Importation and Trend Prediction	17
2.5	Technological Approach	19
2.5.1	Importation Risk Prediction	19
2.5.2	End-user Risk Prediction.....	19
2.5.3	Chemical Recipe Analysis and Risk Prediction.....	21
2.5.4	Future Chemical Importation and Trend Prediction	22
2.6	Technology/ Approach/ Algorithm Review	23
2.7	Chapter Summary.....	23
3	Chapter 3: METHODOLOGY	24
3.1	Chapter Overview.....	24
3.2	Research Methodology	24
3.3	Development Methodology	25
3.4	Project Management Methodology	26
3.4.1	Gantt Chart	26
3.4.2	Deliverables.....	26
3.5	Chapter Summary.....	27
4	Chapter 4: SOFTWARE REQUIREMENTS SPECIFICATION	28
4.1	Chapter Overview.....	28
4.2	Rich Picture	28
4.3	Stakeholder Analysis	29
4.3.1	Onion Model	29
4.3.2	Stakeholder Viewpoints	29
4.4	Selection of Requirement Elicitation Techniques	31
4.4.1	Interviews	31
4.4.2	Questionnaire	31
4.4.3	Followed Requirement Gathering Methods	32
4.5	Discussion of Results	35
4.6	Summary of Findings	35
4.7	Context Diagram	36
4.8	Use Case Diagram	37

4.8.1	Use case Descriptions	37
4.9	Functional Requirements.....	41
4.10	Non-Functional Requirements.....	42
4.11	Chapter Summary	43
5	Chapter 5: SOCIAL, LEGAL, ETHICAL AND PROFESSIONAL ISSUES ...	44
5.1	Chapter Overview.....	44
5.2	SLEP Issues and Mitigation	44
5.2.1	Social Issues	44
5.2.2	Legal Issues	44
5.2.3	Ethical Issues.....	45
5.2.4	Professional Issues	45
5.3	Chapter Summary.....	45
6	Chapter 6: SYSTEM ARCHITECTURE AND DESIGN	46
6.1	Chapter Overview.....	46
6.2	Design Goals	46
6.3	System Architecture Design.....	47
6.4	System Design	48
6.4.1	Choice of Design Paradigm	48
6.4.2	Component Diagram	49
6.4.3	Class Diagram	49
6.4.4	Sequence Diagram	50
6.4.5	UI Design	50
6.4.6	User Experience	52
6.4.7	Process Flow-Chart	53
6.5	Chapter Summary.....	53
7	Chapter 7: IMPLEMENTATION	54
7.1	Chapter Overview.....	54
7.2	Technology Selection	54
7.2.1	Technology Stack.....	54
7.2.2	Data Selection	54
7.2.3	Selection of Development Framework.....	55
7.2.4	Programming Languages	56
7.2.5	Libraries	56
7.2.6	IDE	57
7.2.7	Summary of Technology Selection.....	57

7.3	Implementation of Core Functionalities	57
7.3.1	Importer Risk Prediction	57
7.3.2	Future Trend Prediction	58
7.3.3	Recipe Generator.....	58
7.3.4	End-User Risk Prediction.....	59
7.4	Chapter Summary.....	60
8	Chapter 8: TESTING	61
8.1	Chapter Overview.....	61
8.2	Objectives and Goals of Testing	61
8.3	Testing Criteria.....	62
8.3.1	Functionality Testing Criteria	62
8.3.2	Non-Functionality Testing Criteria	62
8.4	Model Evaluation	62
8.4.1	Importer Risk Prediction	63
8.4.2	End User Risk Prediction	65
8.4.3	Chemical recipe risk analysis.....	67
8.4.4	Future Trend Prediction	68
8.5	Benchmarking	69
8.6	Functional Testing	71
8.7	Module and Integration Testing	72
8.8	Non- Functional Testing.....	73
8.8.1	Accuracy Testing	73
8.8.2	Performance Testing	73
8.8.3	Load Balancing	73
8.9	Limitations.....	74
8.10	Chapter Summary	74
9	Chapter 9: EVALUATION	75
9.1	Chapter Overview.....	75
9.2	Evaluation Methodology and Approach.....	75
9.3	Evaluation Criteria	75
9.4	Self-Evaluation	76
9.5	Selection of Evaluators.....	77
9.6	Evaluation Results	77
9.6.1	The concept of the project.....	77
9.6.2	Scope of the project.....	77

9.6.3	System design, architecture and implementation.....	78
9.6.4	Solution and prototype	78
9.7	Limitations.....	78
9.8	Evaluation on Functional Requirements	79
9.9	Evaluation on Non-Functional Requirements	80
9.10	Chapter Summary	81
10	Chapter 10: CONCLUSION	82
10.1	Chapter Overview	82
10.2	Achievements of Research Aims and Objectives	82
10.2.1	Project Aim	82
10.2.2	Completion of Objectives of the Project.....	82
10.3	Utilization of Knowledge from the Course.....	83
10.4	Use of Existing Skills	84
10.4.1	Machine Learning	84
10.4.2	Web Development - Front End	84
10.4.3	Python	84
10.5	Use of New Skills	85
10.5.1	Machine Learning	85
10.5.2	Web Development – Backend.....	85
10.6	Achievement of Learning Outcomes	85
10.6.1	Skills developed through collaborating within a team on a software development project	85
10.6.2	Analysis of the User-Centered Design Process and its Impact on Legal, Ethical, Professional, and Social Issues in Data Science Applications.....	86
10.7	Problems and Challenges Faced	86
10.8	Deviations	87
10.9	Limitations of the Development	87
10.10	Future Enhancements.....	88
10.11	Achievement of the contribution to body of knowledge	88
10.12	Individual Contribution.....	89
10.13	Chapter Summary	90
	References	i
	Appendix A - Questionnaire for Requirement Gathering	iv

List of Figures

Figure 1 - Feature Prototypes	9
Figure 2 - Concept Map	12
Figure 3 - Gantt Chart	26
Figure 4 - Rich Picture	28
Figure 5 - Onion Model	29
Figure 6 - Context Diagram	36
Figure 7 - Use Case Diagram	37
Figure 8 - System Architecture Design	47
Figure 9 - Component Diagram	49
Figure 10 - Class Diagram	49
Figure 11 - Sequence Diagram	50
Figure 12 - About us page	50
Figure 13 - Home page	50
Figure 14 - Sign in page	50
Figure 15 - Main page	50
Figure 16 - Sign up page	51
Figure 17 - Importer Risk Predict	51
Figure 18 - End user risk prediction	51
Figure 19 - Recipe Risk Analysis	52
Figure 20 - Future Trend Prediction	52
Figure 21 - Process Flow-Chart	53
Figure 22 - Technology Stack	54
Figure 23 - Importer risk classification report	64
Figure 24 - Importer risk confusion matrix	65
Figure 25 - End User classification report	66
Figure 26 - End User confusion matrix	66
Figure 27 - Chemical recipe classification report	67
Figure 28 - Chemical recipe confusion matrix	68
Figure 29 - Future trend Classification Report	68
Figure 30 - Future trend Confusion Matrix	69
Figure 31 - Survey form for Students	vi

List of Tables

Table 1 - Existing Work - Introduction Chapter	4
Table 2 - Table 2 Research Objectives	8
Table 3 - In-Scope	8
Table 4 - Out-Scope	8
Table 5 - Existing Work - Import Risk Prediction	14
Table 6 - Existing Work - End-user Risk Prediction	16
Table 7 - Existing Work - Chemical Recipe Analysis and Risk Prediction	17
Table 8 - Future Chemical Importation and Trend Prediction	18
Table 9 - Research Objectives.....	25
Table 10 – Deliverables	27
Table 11 - Stakeholder Viewpoints.....	30
Table 12 - Comparison of Interviews and Questionnaires.....	32
Table 13 - Questions and Answers – Interview	33
Table 14 - Questionnaire Findings	35
Table 15 - Summary of Findings	36
Table 16 - Use Case Description.....	40
Table 17 - Priority Level	41
Table 18 - Functional Requirements	41
Table 19 - Non-Functional Requirements.....	43
Table 20 - Design Goals.....	46
Table 21 - Data Selection.....	55
Table 22 – Libraries	56
Table 23 - Summary of Technology Selection	57
Table 24 - Importer risk model comparison.....	64
Table 25 - Importer risk confusion matrix	64
Table 26 - End User model comparison.....	65
Table 27 - End User confusion matrix	66
Table 28 - Chemical recipe model comparison.....	67
Table 29 - Chemical recipe confusion matrix	67
Table 30 - Future trend Performance Comparison.....	68
Table 31 - Future trend Confusion Matrix	69
Table 32 – Benchmarking	70
Table 33 - Functional Testing	71
Table 34 - Self Evaluation.....	76
Table 35 - Evaluation Results- 1	77
Table 36 - Evaluation Results- 2	77
Table 37 - Evaluation Results- 3	78
Table 38 - Evaluation Results- 4	78
Table 39 – Limitations	79
Table 40 - Evaluation on functional requirements	80
Table 41 - Evaluation on non-functional requirements.....	80
Table 42 - Completion of Objectives of the Project	83
Table 43 - Utilization of Knowledge from the Course	83
Table 44 - Problems and Challenges.....	87
Table 45 - Individual Contribution	90

1 Chapter 1: INTRODUCTION

1.1 Chapter Overview

This chapter discusses the Comprehensive Chemical Risk Prediction Model (CCRPM), which was inspired by the National Authority for Implementation of the Chemical Weapons Convention in Sri Lanka (NACWC). This model assists the NACWC authorities in monitoring and predicting risks associated with dual-use chemicals, which serve legitimate industrial purposes but can be repurposed for illegal activities, such as making chemical weapons or explosives. This chapter begins by identifying the key stakeholders of the system and outlining their respective roles. It then explores various requirement collection strategies, evaluating their advantages and disadvantages according to the context of the project. Furthermore, the requirement analysis stages also include the prototyping and the associated definitions. Finally, a scope description is used to describe the scope of the system by dividing the scope into categories depending on how important they are to the system.

1.2 Problem Domain

The misuse of dual-use chemicals—those that serve legitimate industrial purposes but can also be repurposed for harmful activities such as creating chemical weapons or explosives—presents significant challenges for governments and regulatory bodies worldwide. These chemicals are often difficult to monitor effectively due to their widespread industrial use and the complexity of tracking their movement and handling across different stages, from importation to end-use.

CCRPM addresses the need for a robust risk prediction and monitoring system to track the entire lifecycle of dual-use chemicals, from the importer to the end-user. It identifies and predicts potential risks associated with the misuse of these chemicals by analysing past importation patterns, end-user behaviour, and suspicious activity. Furthermore, it forecasts future trends in chemical importation and assesses the associated risks, enabling authorities to take proactive measures to prevent illegal activities. The model offers real-time insights and warnings, ensuring that regulatory bodies can swiftly respond to potential threats while ensuring that legitimate industrial processes are not disrupted.

1.3 Problem Statement

The Chemical Control Risk Prediction Model (CCRPM) is an AI-driven system that leverages machine learning to track dual-use chemicals, assess risks, and predict trends in importation and usage, enabling real-time alerts and early detection of suspicious activities to enhance public safety and national security.

1.4 Research Motivation

The growing complexity of global chemical supply chains, coupled with the dual-use nature of many chemicals, presents significant challenges for regulators trying to prevent their misuse in the creation of chemical weapons or explosives. Dual-use chemicals are essential to numerous industries, making it difficult to differentiate between legitimate use and potential illicit activities. Traditional monitoring systems often struggle to provide timely, actionable insights, leaving gaps in regulatory oversight and causing delayed responses to emerging threats. This research, inspired by the mission of the National Authority for the Implementation of the Chemical Weapons Convention (NACWC) of Sri Lanka, aims to address these 2 challenges by developing an AI-driven solution that improves the monitoring and regulation of dual use chemicals.

The Comprehensive Chemical Risk Prediction Model (CCRPM) will leverage machine learning (ML) to analyze large volumes of historical data, uncover hidden patterns, and predict risks in real time. By offering advanced risk predictions and future trend analyses, CCRPM will empower regulatory bodies to proactively anticipate and mitigate threats, rather than reacting to them after they have escalated. The project aims to fill gaps in current monitoring systems, ensuring both the prevention of chemical misuse and the continued support of legitimate industrial activities. This research has the potential to significantly enhance global chemical safety management efforts, contributing to the broader goal of preventing the proliferation of dangerous substances.

1.5 Existing Work

The application of AI and Machine Learning (ML) in chemical risk prediction is a relatively new and evolving field. As a result, the body of existing work specifically addressing comprehensive chemical risk prediction is limited. There are no fully developed, off-the-shelf software solutions that entirely match the needs of this problem domain. However, several research studies and systems with similar functionalities address various aspects of chemical safety, risk management, and prediction. Below is a list of notable research works

and systems that employ AI and ML techniques in related fields, providing a foundation for developing a more comprehensive chemical risk prediction model.

Citation	Technology/Algorithm	Advantages	Limitations
Na Luo et al. (2023). Fuzzy Logic and Neural Networkbased Risk Assessment Model for Import and Export Enterprises.	Fuzzy Logic, Neural Networks, ANFIS	Improves risk assessment and handles vague and uncertain data more efficiently, enhancing customs risk management.	Computational complexity can be high, especially for large datasets with many variables.
Anuar et al. (2019). Process Knowledge Management System (PKMS)	Random Forest	Emphasizes hazard control and allows updating and modification of data.	Reliant on realtime data, which might not always be available consistently.
Jeong et al. (2022). Vapor Pressure and Toxicity Prediction for Novichok Agents	Fuzzy Rules, Text Mining	Accurate prediction of vapor pressure and toxicity.	Limited to organophosphorus compounds, reducing generalization to other chemicals.

He et al. (2023). Game Mechanism and Event triggering Natural Gas Demand Model.	Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GBM), LSTM	Improved accuracy for natural gas demand forecasting.	Efficacy depends on availability and quality of external datasets
--	---	---	---

Table 1 - Existing Work - Introduction Chapter

1.6 Research Gap

Despite advancements in AI and Machine Learning (ML) for risk assessment in customs and chemical safety, a significant gap exists in developing a comprehensive predictive model specifically for monitoring dual-use chemicals. Current models often focus on isolated aspects of risk, failing to provide a holistic view of the entire lifecycle of these chemicals—from importation to end-user behaviour. There is an urgent need for an integrated system that combines various data sources and analytical techniques for dynamic, real-time risk assessment.

Moreover, existing systems typically rely on static algorithms that struggle to adapt to changing threats, highlighting the necessity for adaptive ML models capable of learning from new data and evolving circumstances. Additionally, many approaches do not adequately address the dual-use nature of certain chemicals, complicating their regulation and monitoring. Effective methods for providing real-time alerts and actionable insights for immediate intervention are lacking, along with sophisticated data fusion techniques that enhance predictive accuracy by combining diverse data types.

Finally, ethical considerations, including bias and transparency, are often overlooked in current models. Further research is required to establish frameworks that ensure the ethical use of AI in chemical risk prediction while maintaining model interpretability for regulatory stakeholders. By addressing these gaps, the Comprehensive Chemical Risk Prediction Model (CCRPM) aims to significantly enhance the capabilities of regulatory bodies in preventing the misuse of dual-use chemicals, contributing to global safety and security efforts.

1.7 Contribution to the body of knowledge

1.7.1 Technological Contribution

Our Comprehensive Chemical Risk Prediction Model (CCRPM) uses supervised machine learning algorithms, like classification and multivariate regression for the prediction of risks associated with importers, end-users, and chemical recipes. We will develop a web-based application using HTML, CSS, and JavaScript to provide an easy-to-use interface.

The backend will be implemented using the Flask REST API, which will connect to a database for secure data management. Safety and prevention from the illicit use of chemicals will be enhanced by the implementation

of ML models with the help of Python libraries, which perform pattern analysis, predict future trends, and enable compliance by performing automated checks.

1.7.2 Contribution to domain

Our system will contribute to risk prediction and management in the chemical lifecycle with machine learning at different junctures. This will check if an importer is trustworthy or not, and based on that, the amount of risk can be determined on the kind of chemicals that he imports within the country.

Once the chemicals are inputted, the system checks for various hazardous combinations. For example, if an importer has 20 chemicals, it calculates all the ways they could combine (that's like finding factorial of 20 (20!)). This helps us to determine whether substances, such as explosives or chemical weapons, can be produced.

We also track the end users and forecast whether they will misuse the chemicals. In this manner, we make certain that chemicals are used safely by authorized participants only. The system will make predictions of the import of chemicals in the future. It will check how much of which chemical may be required and flag any unusual orders that may pose a risk. The concept guarantees better safety and security since the risks are caught at an earlier stage, whether it is at the time of import, usage, or even before the order for the next one, which is the very reason this system is considered unique and effective.

1.8 Research challenge

Limited availability of research papers - No comprehensive research or projects were identified that are closely aligned with the proposed Chemical management system. Although some works have 30-40% similarity, they focus only on the specific aspects of four main functionality of our system, which makes it challenging for us to depend on existing works.

Novel concept with limited precedent - This idea was proposed by the National Authority for implementation of the chemical weapons convention of Sri Lanka, but they have not collected the necessary data or implemented a similar system it has remained a significant challenge for us while developing models-the data we need is unavailable.

Data collection difficulties - There is no existing dataset for chemical importation, combinations, and misuse tracking, therefore we need to collect and create our own datasets, which will require extensive time and effort. As a result, generating accurate predictions and performing reliable risk assessments is challenging.

Selecting an appropriate machine learning model - It is challenging to identify the right ML models for each functionality. The lack of previous work to guide us makes it harder to determine whether to use models like regression, classification, or clustering for risk prediction.

1.9 Research Question

RO1: Which supervised machine learning algorithm is most effective for predicting the risk level of chemical importers?

RO2: How can AI models detect potential misuse by analyzing combinations of imported chemicals?

RO3: What methods can be used to design a secure and scalable platform for tracking import activities and chemical risks?

RO4: How can predictive analytics be used to forecast future chemical import demands and assess associated risks

1.10 Research Aim

The research aims to develop a comprehensive, AI-driven risk prediction model that leverages machine learning techniques to monitor dual-use chemicals throughout their lifecycle. This model will predict potential risks associated with importers, end-users, and hazardous chemical combinations, providing real-time alerts and trend analyses. By filling critical gaps in existing systems, the project aims to empower regulatory authorities with proactive tools to prevent the misuse of chemicals, ultimately contributing to safer and more secure industrial and public environments.

1.11 Research objective

Research objective	Explanation	Learning out coming
Problem Identification	RO1: Using the importers profile and their past import pattern identifies high-risk chemical importer. RO2: Predicting hazard chemical combination, which can be used to manufacture chemical weapons and explosives RO3: tracking activities of end-user and calculating the associated risk of misuse RO4: Predict the future chemical import needed and assess associated risk	LO1
Data gathering and analysis	<ul style="list-style-type: none"> • Chemical import records • Past chemical import pattern • End-user records 	LO1, LO3
Research design	A Quasi-Experimental Design will be employed to examine the relationship between chemical import patterns, potential misuse, and risk prediction. (Maciejewski, M. L. 2018).	LO4, LO3
Implementation	Web platform for tracking chemical imports, risk analysis, hazardous chemical combination, end-user monitoring and future import pattern along with ML model and other technical aspects.	LO2, LO3, LO4

Testing and evaluation	Surveys and questionnaires will be used to gather feedback from supervisors to check system performance and usability.	LO2, LO4
------------------------	--	----------

Table 2 - Table 2 Research Objectives

1.12 Project scope

1.12.1 In-Scope

NO	Description
01	Identifying risk factors based on importer history, types of chemicals, and compliance records.
02	Identifying safety concerns, usage trends, or improper handling and Predicting risks associated with different end-user groups or usage patterns.
03	Allowing users to input chemical combinations and Analyzing the safety and risk associated with various chemical combinations.
04	Predicting future import trends of specific chemicals and Analyzing potential risks based on these imports and their potential impact.

Table 3 - In-Scope

1.12.2 Out-Scope

NO	Description
01	Implementing a chatbot to assist with user queries, report generation, or data retrieval will be considered in future expansions.
02	The System can be customize for different countries' legal frameworks or standards

Table 4 - Out-Scope

1.12.3 Feature Prototype

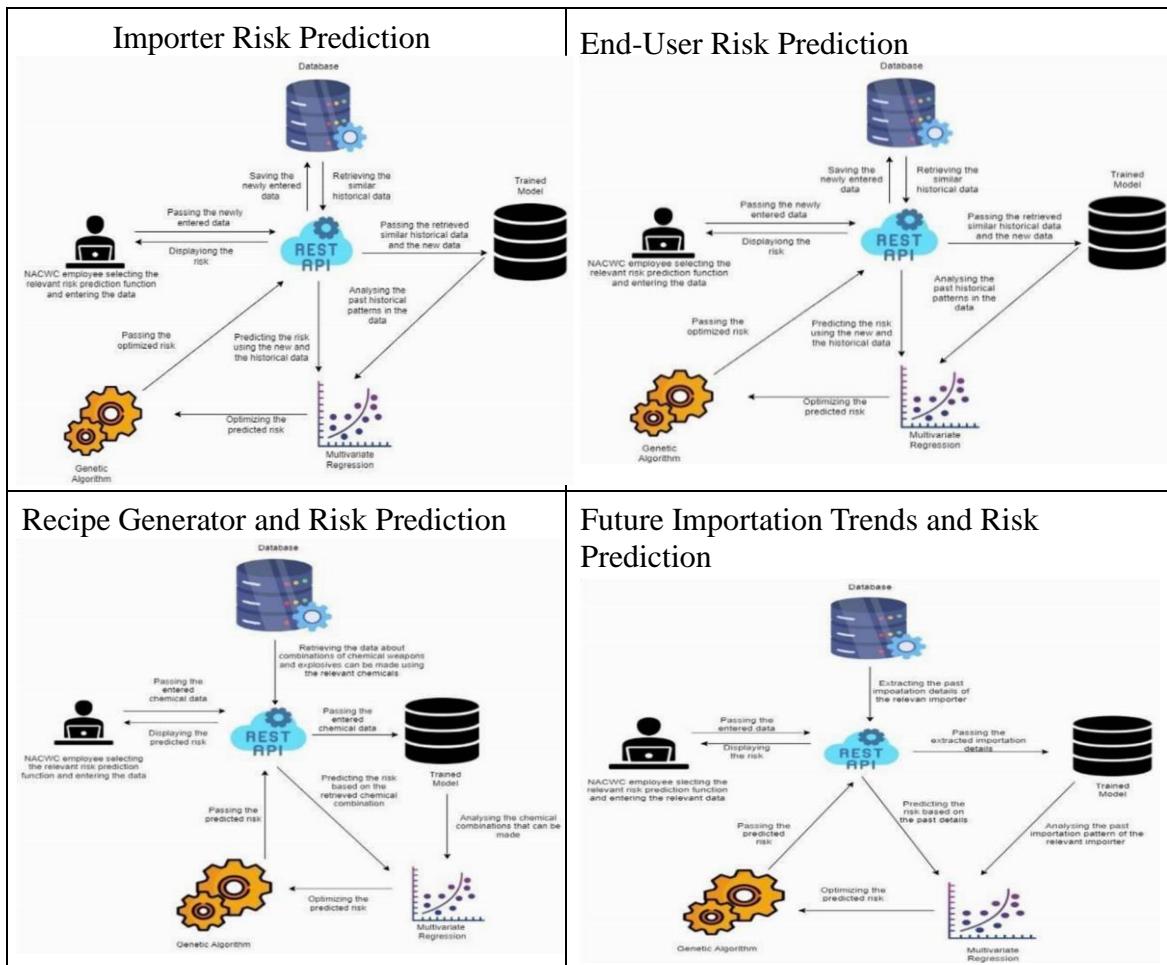


Figure 1 - Feature Prototypes

The Comprehensive Chemical Risk Prediction Model (CCRPM) consists of four independent functionalities designed to assess chemical risks effectively. These include Importer Risk Prediction, End-User Risk Prediction, Recipe Generator and Risk Prediction, and Future Importation Trends and Risk Prediction, each leveraging machine learning and data-driven techniques for real-time risk assessment.

Importer Risk Prediction analyzes historical chemical import data to determine potential risks, while End-User Risk Prediction evaluates purchase patterns to assess the likelihood of misuse. Recipe Generator and Risk Prediction identifies hazardous chemical combinations, predicting whether they could be used to create explosives or toxic substances. Lastly, Future Importation Trends and Risk Prediction forecasts potential risks based on past import trends, allowing regulatory bodies to take preventive measures. Each function operates autonomously through a Flask REST API, using multivariate regression and Genetic Algorithms to optimize risk.

1.13 Resource Requirement

1.13.1 Hardware Requirements

- Intel Core i7 7th generation processor or higher - To get more powerful and high performance.
- 16GB RAM or high - To train heavy algorithms
- Storage (minimum 256GB SSD / 1TB HDD) - To store a large amount of data

1.13.2 Software Requirements

- Windows Operating System - To handle huge computational functionalities
- Python - The primary language used to create the proposed system is Python
- PyCharm Enterprise - Used for developing proprietary and commercial software.
- MongoDB - To manage databases and servers
- HTML - To handle huge computational functionalities
- CSS - The primary language used to create the proposed system is Python
- JavaScript - Used for developing proprietary and commercial software.
- MS Word and Google Docs - For writing documents
- Flask - To develop dynamic and interactive websites
- TensorFlow - For pre-processing and training the model

1.13.3 Data Requirements

- Stockholders – NACWC
- Website - <https://www.volza.com/p/general-chemical/>
- Website - <https://www.seair.co.in/hs-codes/chapter-28-inorganicchemicalscompounds-precious-metals.aspx>
- Generative AI tools – Chatgpt, Gemini

1.13.4 Skill Requirements

- Searching for information
- Time management
- Problem solving
- Report writing
- Critical thinking
- Planning and scheduling

1.14 Chapter Summary

This chapter outlines the key contributions of stakeholders in the Comprehensive Chemical Risk Prediction Model (CCRPM) project. The system aims to improve chemical risk monitoring, prevent misuse, and enhance regulatory measures. Stakeholders, including regulatory authorities, benefit from better risk assessment, early detection of threats, and improved compliance management.

2 Chapter 2: LITERATURE REVIEW

2.1 Chapter Overview

Effective chemical management is crucial for safety, security, and regulatory compliance across industries. Some chemicals pose health and environmental risks, while others, like cyanides, have dual-use properties, making them susceptible to misuse. The Comprehensive Chemical Risk Prediction Model (CCRPM) is an AI-driven system designed to assess and mitigate risks throughout the chemical lifecycle by automating analysis and improving decision-making. Using Random Forest, a hybrid model combining LSTM and Gradient Boost, the system predicts risks in four key areas: importation, end-user applications, hazardous formulations, and future imports. Inspired by the National Authority for Implementation of the Chemical Weapons Convention (NACWC) in Sri Lanka, under the Ministry of Defence, CCRPM enhances monitoring from importation to end-use, reducing risks of accidents, environmental damage, and illicit activities. This chapter reviews existing literature on chemical risk management, AI-driven predictive models, and regulatory frameworks, identifying research gaps and demonstrating CCRPM's role in advancing safety and compliance.

2.2 Concept Map

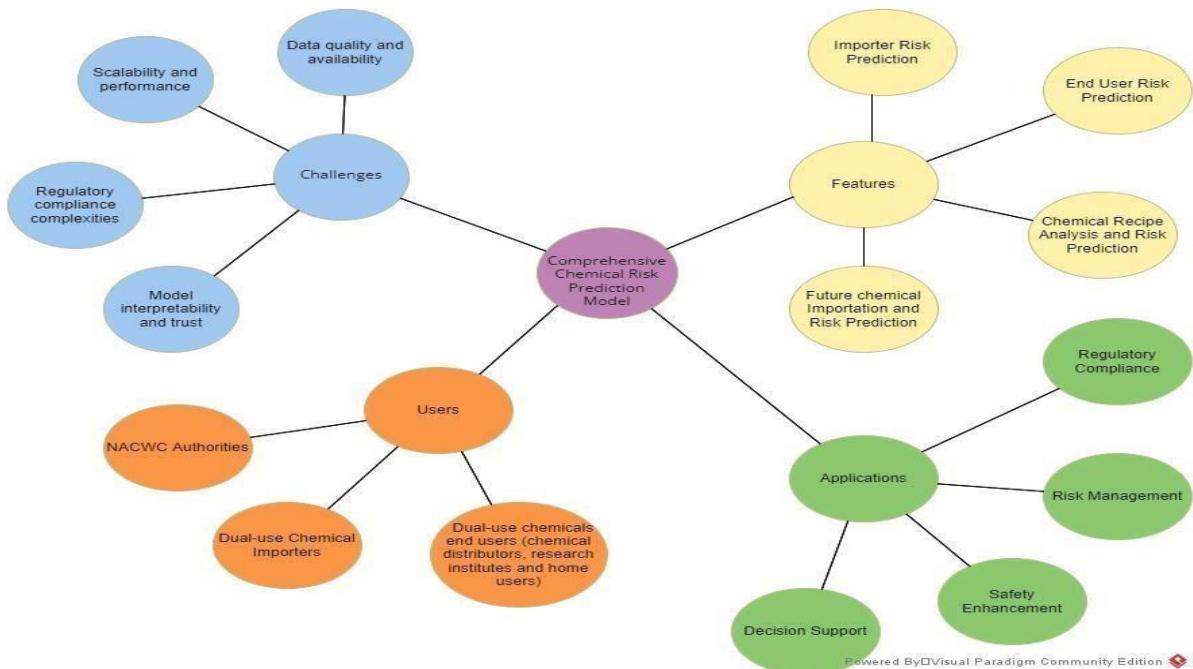


Figure 2 - Concept Map

2.3 Problem Domain

The misuse of dual-use chemicals—those that serve legitimate industrial purposes but can also be repurposed for harmful activities such as creating chemical weapons or explosives—presents significant challenges for governments and regulatory bodies worldwide. These chemicals are often difficult to monitor effectively due to their widespread industrial use and the complexity of tracking their movement and handling across different stages, from importation to end-use.

CCRPM addresses the need for a robust risk prediction and monitoring system to track the entire lifecycle of dual-use chemicals, from the importer to the end-user. It identifies and predicts potential risks associated with the misuse of these chemicals by analysing past importation patterns, end-user behaviour, and suspicious activity. Furthermore, it forecasts future trends in chemical importation and assesses the associated risks, enabling authorities to take proactive measures to prevent illegal activities. The model offers real-time insights and warnings, ensuring that regulatory bodies can swiftly respond to potential threats while ensuring that legitimate industrial processes are not disrupted.

2.4 Existing Work

2.4.1 Importation Risk Prediction

Research	Author	Year	Dataset	Model Used	Metric/Outcome
Fuzzy Logic and Neural Networkbased Risk Assessment Model for Import and Export Enterprises	Na Luo, Hua Yu, Zeqing You, Yao Li, Tunan Zhou, Yuwei Jiao, Nan Han, Chenxu Liu, Zihan Jiang, Shaojie Qiao	2023	Customs risk assessment data for import and export enterprises	Fuzzy logic, Neural networks, ANFIS	Improved risk assessment, better handling of vague and uncertain data, enhanced efficiency of customs risk management

Customs Fraud Detection	Vanhoeveld et al.	2020	9,624,124 records of Belgian Customs administration	EasyEnsemble (EE), Support Vector Machines (SVM)	Improved fraud detection, advanced feature selection for customs fraud cases
Customs Risk Assessment Based on Data Mining	Song et al.	2019	121,506 samples of goods and 333 foods from China Customs food department	Fuzzy rules, Text mining	Enhanced accuracy in customs risk assessment, combination of text mining with fuzzy rule reasoning
Developing a Model for Managing the Risk Assessment of Import Declarations in Customs Based on Data Analysis Techniques	Hassan Ali Khojasteh Aliabadi, Saeed DaeiKarimzadeh, Majid Iranpour Mobarakeh, Farsad Zamani Boroujeni	2022	575,006 import declaration from all Iranian customs (2019-2020)	PCA, LDA, Fast ICA, Random Forest, Decision Trees, Logistic Regression, k-Nearest Neighbors, Gaussian Naïve Bayes, Multilayer Perceptron, Fuzzy pattern, Fuzzy Pattern Tree Top Down, Gradient Boosting, Ada Boost, Bagging, Quadratic Discriminant Analysis, Multimodal Evolutionary	Improved risk prediction with Random Forest showing 69% accuracy; Better resource allocation in customs risk management

Table 5 - Existing Work - Import Risk Prediction

2.4.2 End-user Risk Prediction

Research	Author	Year	Dataset	Model Used	Metric/Outcome
Best Practices in Chemical Management for Textile Manufacturing	Doug Cahn Robert Clifford	2014	Data provided by the InterAmerican Development Bank	Deep Learning	Textile manufacturing factories in Haiti are increasing chemical usage and need to identify and control the associated hazards.
Chemical substances management system at the University of Tokyo	<u>Yoshiko Tsuji</u> <u>Kenichi Tonokura</u> <u>Rumiko Hayashi</u>	2016	University Data	Deep Learning	The University of Tokyo has a comprehensive chemical management system that tracks chemical usage from purchase to disposal.
The Chemical Management System (CMS): A Useful Tool for Inventory Management	Maurice K. Payne Andrew W. Nelson Walter R. Humphrey Christine M. Straut	2020	Text Data	Logistic Regression	The Chemical Management System (CMS) is a free software tool that provides effective inventory management and chemical security.
Process Knowledge Management System	N I P Anuar H A Aziz R Ahmad	2019	data was collected based on regulatory requirements, Real Plant Data and User-Entered Data	Random Forest	PKMS allows the end user to store, review, modify and update the data. Most importantly, PKMS emphasizes the hazard related to the process chemical, technology and equipment within company in systematic way and vital to control the process hazard

Enhancing Chemical Inventory Management in Laboratory through a Mobile-Based QR Code Tag	M A M Shukran M S Ishak M N Abdullah	2017	Real time Data	Logistic Regression	A fully functional QR Code inventory system that was developed in this research has proven that the information that was retrieved from the database at the time of scanning is the status record of the inventory such as chemical quantity or volume, date purchased, disposal scheduled, and the vendor
--	--	------	----------------	---------------------	--

Table 6 - Existing Work - End-user Risk Prediction

2.4.3 Chemical Recipe Analysis and Risk Prediction

Research	Author	Year	Dataset	Model Used	Metric/Outcome
Vapor Pressure and Toxicity Prediction for Novichok Agents	Keunhong Jeong et al.	2022	Organophosphorus compounds (265 descriptors)	Machine Learning (Classification)	Accurate predictions of vapor pressure and toxicity with reasonable error
Hazardous Chemicals Detection and Classification Through Millimeter Wave and Machine Learning	Lorena C. Ilagan, Elmer P. Dadios	2024	Millimeter wave data for hazardous chemicals	Decision Tree, KNN, SVM, Random Forest	Enhanced chemical detection accuracy through machine learning
Interpreting Chemical Detection Alarms with	Patrick C. Riley, Samir V.	2022	Ion Mobility Spectrometry (IMS) and	Random Forest, LSTM	Reduced false alarms in chemical detection

Live Analysis of ML Algorithms	Deshpande, et al.		multi-gas detector data		
The EUSENSE System for Chemical Hazards Detection, Identification, and Monitoring	Małgorzata Gawlik-Kobylińska et al.	2021	Heterogeneous sensor node data	Machine Learning (Modelling Algorithms)	Improved chemical reconnaissance and reduced false alarm rate
Machine Learning-based Approach for Efficient Prediction of Toxicity of Glasses	Ahmet Murat Erturan et al.	2023	Toxic and nontoxic gases (144 compounds)	Bayesian Network, Random Forest, J48, K-Nearest Neighbor (KNN), Simple Logistic	89.65% accuracy with Bayes Net for external validation

Table 7 - Existing Work - Chemical Recipe Analysis and Risk Prediction

2.4.4 Future Chemical Importation and Trend Prediction

Research	Author	Year	Dataset	Model	Metric
Game mechanism and event-triggering based Natural Gas Demand Prediction (GMET-NGDP) model for Chemical and Fertilizer Industry	Zhiming He Wei Xiao Zizi Li Fuping Wang Yujing Chen Tianxiang Yang	25-27 August 2023	1. Gas Consumption Data 2. Production days 3. Annual average price 4. Natural gas index data 5. Weather data 6. Commodity 7. exchange data 8. City air quality index.	Support vector Machine (SVM) Random Forest (RF) Gradient Boosting Machine (GBM) Long ShortTerm memory (LSTM) network	--

Demand Forecasting in Supply Chain Using Uni-Regression Deep Approximate Forecasting Model	<u>Emad Aldahmani</u> <u>Kolawole Iyiola</u> <u>Ahmad Alzubi</u>	2024	Demand Forecasting Dataset M5 Forecasting Dataset Daily Demand Forecasting Orders Dataset	Recurrent BiLSTM (Bidirectional Long Short-Term Memory) Network NARX (Nonlinear Autoregressive with Exogenous Inputs) Model	--
Water Chemical Oxygen Demand Detection System Based on LASSO Algorithm	<ul style="list-style-type: none"> • <u>Ning Wang</u> • <u>Yazhen Wang</u> • <u>Yang Yu</u> • <u>Zhongxing Pan</u> • <u>Rui Sun</u> • <u>Yuanyuan Kong</u> • <u>Chunfang Zhang</u> 	28-29 December 2022	MOOC Learner Dataset Engagement Dataset Cohort Dataset Publicly Available Datasets	<ol style="list-style-type: none"> 1. Logistic Regression 2. Support Vector 3. Machines (SVM) 4. Decision Trees 5. Random Forest 6. K-Means Clustering 7. Multi-Layer Perceptron (MLP) 8. Convolutional Neural Networks (CNN) 	90%
Forecasting Model of Fishery Import and Export Trade Data Using Deep Learning Method	<u>Mingyu Sun</u> <u>Huanhai Yang</u>	25-27 August 2023	China Fishery Statistical Yearbook (2008-2021) Import and export trade data from major fishing areas (Shandong, Liaoning, Zhejiang, Fujian, Shanghai, Jiangsu)	ARIMA (AutoRegressive Integrated Moving Average) LSTM (Long Short-Term Memory)	98%

Table 8 - Future Chemical Importation and Trend Prediction

2.5 Technological Approach

2.5.1 Importation Risk Prediction

The Importer Risk Prediction system is developed using the Agile Development Methodology, ensuring iterative improvements and flexibility in feature enhancement. This approach allows continuous feedback and adaptation to evolving requirements (Hassan Ali Khojasteh Aliabadi, 2022). Data exchange is managed by a centralized server accessible by administrators. The server handles data using the HTTP protocol and REST API, facilitating efficient communication between different system components (Song et al., 2019).

For server-side data management, the system uses the Python-based framework. The framework is used to build and maintain the REST API, ensuring secure and scalable access to importer risk-related data (Vanhoeveld Jellis, 2020). The REST API is developed using the Framework, which simplifies API creation and ensures robust authentication mechanisms. This approach accelerates development and enhances maintainability (Hassan Ali Khojasteh Aliabadi, 2022).

Risk prediction is performed using machine learning models implemented in ScikitLearn and TensorFlow. Data preprocessing, risk factor identification, and classification (e.g., Random Forest, Gradient Boosting, and Neural Networks) are executed through Python scripts integrated with the Django backend (Na Luo, 2023).

2.5.2 End-user Risk Prediction

To develop an end-user risk prediction system, the first step involves data collection and preprocessing. Historical data on chemical importation patterns, usage, and compliance records are gathered, including details such as purchase frequency, volume, chemical types, and supplier information. Data cleaning techniques, such as imputation for missing values and normalization for scaling numerical features, are applied to ensure accuracy (M A M Shukran, 2017). Categorical encoding is also used to convert non-numeric data into a format suitable for machine learning models. For example, (M A M Shukran, 2017) demonstrated the effectiveness of real-time data collection and normalization in

their QR Code-based chemical inventory system, which tracks chemical quantities and disposal schedules.

Next, feature engineering is performed to identify and create meaningful risk factors. Key features include the frequency and volume of chemical purchases, the hazard level of chemicals, supplier risk, and compliance history. Additional features, such as temporal trends and geospatial data, can further enhance the model's predictive power. (Maurice K. Payne, 2020) emphasized the importance of systematic inventory management in identifying risk factors, which can be adapted for end-user risk prediction. This step ensures that the model captures underlying patterns in the data, enabling accurate risk classification.

For model selection, machine learning algorithms such as logistic regression, random forest, and gradient boosting are considered. Logistic regression is suitable for binary classification tasks, while random forest and gradient boosting models can handle nonlinear relationships and provide insights into feature importance (N I P Anuar, 2019). Deep learning models, such as LSTMs, can also be explored for complex timeseries data, as demonstrated by (Doug Cahn, 2014) in their work on chemical usage patterns in textile manufacturing. The model's performance is evaluated using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to ensure robustness.

Finally, the system is deployed as a web application or API, allowing stakeholders to input data and view risk predictions in real-time. Continuous monitoring and feedback loops are implemented to improve the model's accuracy over time. This approach aligns with best practices in chemical management, as highlighted by (Yoshiko Tsuji, 2016) and (Maurice K. Payne, 2020), who emphasized the importance of systematic tracking and risk assessment. By integrating these technologies, the end-user risk prediction system can provide actionable insights to mitigate risks and enhance safety in chemical management.

2.5.3 Chemical Recipe Analysis and Risk Prediction

The first step involves collecting and preprocessing chemical data, including properties such as toxicity, vapor pressure, and reactivity. Data from heterogeneous sources, such as sensor nodes or chemical databases, are cleaned and normalized to ensure consistency (Małgorzata Gawlik-Kobylińska, 2021). Techniques like imputation for missing values and scaling for numerical features are applied. For example, (Keunhong Jeong, 2022) used 265 descriptors for organophosphorus compounds to predict vapor pressure and toxicity, highlighting the importance of robust preprocessing.

Next, feature engineering is performed to identify chemical combinations and their associated risks. Key features include chemical properties, interaction effects, and environmental factors. Advanced techniques, such as molecular descriptor extraction and interaction modelling, can be used to capture complex relationships between chemicals. (Lorena C. Ilagan, 2024) demonstrated the effectiveness of machine learning in enhancing chemical detection accuracy, which can be adapted to identify hazardous combinations. This step ensures that the model captures the underlying patterns in chemical interactions.

For risk prediction, machine learning models such as Random Forest, Bayesian Networks, and LSTM are considered. Random Forest is effective for handling nonlinear relationships and providing feature importance, while Bayesian Networks offer probabilistic reasoning for risk classification (Ahmet Murat Erturan, 2023). LSTM models, as used by (Patrick C. Riley, 2022), are suitable for sequential data analysis, such as time-dependent chemical interactions. The model is trained on preprocessed data and evaluated using metrics like accuracy, precision, recall, and F1score.

The final step involves classifying risk levels of chemical combinations (e.g., low, medium, high) based on model predictions. The system is deployed as a web application or API, allowing users to input chemical recipes and receive risk assessments in real-time. Continuous monitoring and feedback loops are implemented to improve accuracy, as emphasized by (Małgorzata Gawlik-Kobylińska, 2021) in their EU-SENSE system for chemical hazard detection.

2.5.4 Future Chemical Importation and Trend Prediction

The first step involves collecting and preprocessing historical importation data, including details such as import volumes, chemical types, supplier information, and external factors like economic indicators, weather data, and commodity prices. Data cleaning techniques, such as handling missing values and normalizing numerical features, are applied to ensure consistency. For example, (Zhiming He, 2023) used diverse datasets, including gas consumption and weather data, to predict natural gas demand, highlighting the importance of robust preprocessing.

Next, trends and patterns in chemical importation are identified using exploratory data analysis (EDA) and feature engineering. Key features include temporal trends (e.g., seasonality, annual growth), chemical categories, and external factors like economic or environmental conditions. Advanced techniques, such as time-series decomposition and clustering, can be used to uncover hidden patterns. (Mingyu Sun, 2023) demonstrated the effectiveness of trend analysis in their fishery import and export forecasting model, which can be adapted for chemical importation. For predicting future importation trends, machine learning models such as LSTM, ARIMA, and Random Forest are considered. LSTM networks are particularly effective for time-series data, as shown by (Zhiming He, 2023) and (Mingyu Sun, 2023). ARIMA models are suitable for capturing linear trends and seasonality, while Random Forest can handle non-linear relationships and provide feature importance. The model is trained on historical data and evaluated using metrics.

The final step involves predicting associated risks based on future importation trends. Risk factors, such as supply chain disruptions, regulatory changes, and market volatility, are incorporated into the model. The system is deployed as a web application or API, allowing stakeholders to input data and receive predictions in real-time. Continuous monitoring and feedback loops are implemented to improve accuracy, as emphasized by (Emad Aldahmani, 2024) in their supply chain demand forecasting model.

2.6 Technology/ Approach/ Algorithm Review

The project leverages a comprehensive technology stack to ensure successful execution:

- Python - Serving as our primary programming language for machine learning development due to its versatility and extensive libraries.
- PyCharm IDE - Our chosen development environment that enhances Python programming with intelligent code completion, debugging tools, and efficient project management capabilities.
- TensorFlow Framework - The open-source machine learning platform that enables implementation of our proposed predictive models and algorithms.
- Scikit-learn Library - Provides streamlined tools for data analysis and modeling, complementing our machine learning workflow.
- React.js - The JavaScript library powering our dynamic and responsive user interface components.
- Flask Microframework - Our lightweight Python web framework that simplifies backend development and API creation.
- Git Version Control - The essential system for tracking code changes, facilitating team collaboration, and maintaining project versions.
- GitHub Platform - Hosts our Git repositories while providing additional collaborative features for distributed development.

2.7 Chapter Summary

Accurately predicting risks in the chemical supply chain is inherently challenging because of the dynamic and multifaceted nature of chemical data. By employing a comprehensive set of methodologies, the CCRPM enhances risk assessment accuracy and operational efficiency. The systematic integration of diverse machine learning models not only improves reliability over traditional manual assessments but also streamlines the decision-making process for regulators. Through this approach, the research advances the overall effectiveness of chemical risk detection and supports safer, more compliant chemical management practices.

3 Chapter 3: METHODOLOGY

3.1 Chapter Overview

The proposed Comprehensive Chemical Risk Prediction Model (CCRPM) integrates multiple machine learning techniques to analyze and predict risks associated with chemical handling throughout its lifecycle. Currently, chemical risk assessments are largely manual, time-consuming, and prone to inaccuracies due to the complexity of tracking dual-use chemicals. This project aims to automate the evaluation process by classifying risk levels from historical import data, end-user behavior, hazardous chemical combinations, and forecasting future trends—ultimately empowering regulatory bodies with actionable insights for proactive risk mitigation.

3.2 Research Methodology

Research Methodology	Description
Research Philosophy	The research adopts positivism as the philosophy since the CMS relies on objective, observable facts such as importer data, end-user chemical handling, and recipe safety to predict risks. The role of the researcher is limited to collecting and analyzing data to derive risk scores, and the results are factual rather than subjective.
Research Approach	The approach follows an inductive method, starting with the collection and analysis of real-world data (importer records, chemical usage data, recipe combinations). The system then uses this data to build risk prediction models, theorizing safety and compliance risks based on patterns observed in the data. The inductive approach allows for new patterns and relationships to emerge as the system learns from the data.
Research Strategy	A Quantitative Analysis strategy is used in the CMS. The system gathers numerical data related to chemical imports, handling incidents, and compliance records. Data mining, statistical techniques, and machine learning models are then applied to predict risks associated with chemical imports and usage.

Research Choice	A mono-method research approach is used, where the system predominantly focuses on quantitative data analysis. All data (import trends, chemical reactions, and user handling records) are numerically driven, and predictive analytics are applied using machine learning algorithms and statistical techniques.
Time zone	A cross-sectional time zone approach is used, as the system analyzes data at specific points in time (e.g., analyzing current importer data to predict risks or chemical recipe safety based on real-time input). Each analysis is performed on a specific dataset relevant to the time of import or usage.

Table 9 - Research Objectives

3.3 Development Methodology

The Chemical Risk Management System follows an Agile methodology with Scrum, ensuring flexibility through iterative sprints. Each sprint focuses on key aspects like risk prediction, data preprocessing, and the web-based portal. The project applies Object-Oriented Analysis and Design (OOAD) principles, keeping components modular for scalability, allowing future enhancements like chatbot integration or realtime tracking.

For model evaluation, metrics like accuracy, precision-recall, and F1-score are used, though techniques may vary across modules. Benchmarking is challenging since no existing system offers similar functionalities. Future improvements include continuous model retraining with real-world data and potential integration of government regulations and live import data, though sensitive data requires careful approval. This approach ensures adaptability, security, and ongoing enhancement.

3.4 Project Management Methodology

3.4.1 Gantt Chart

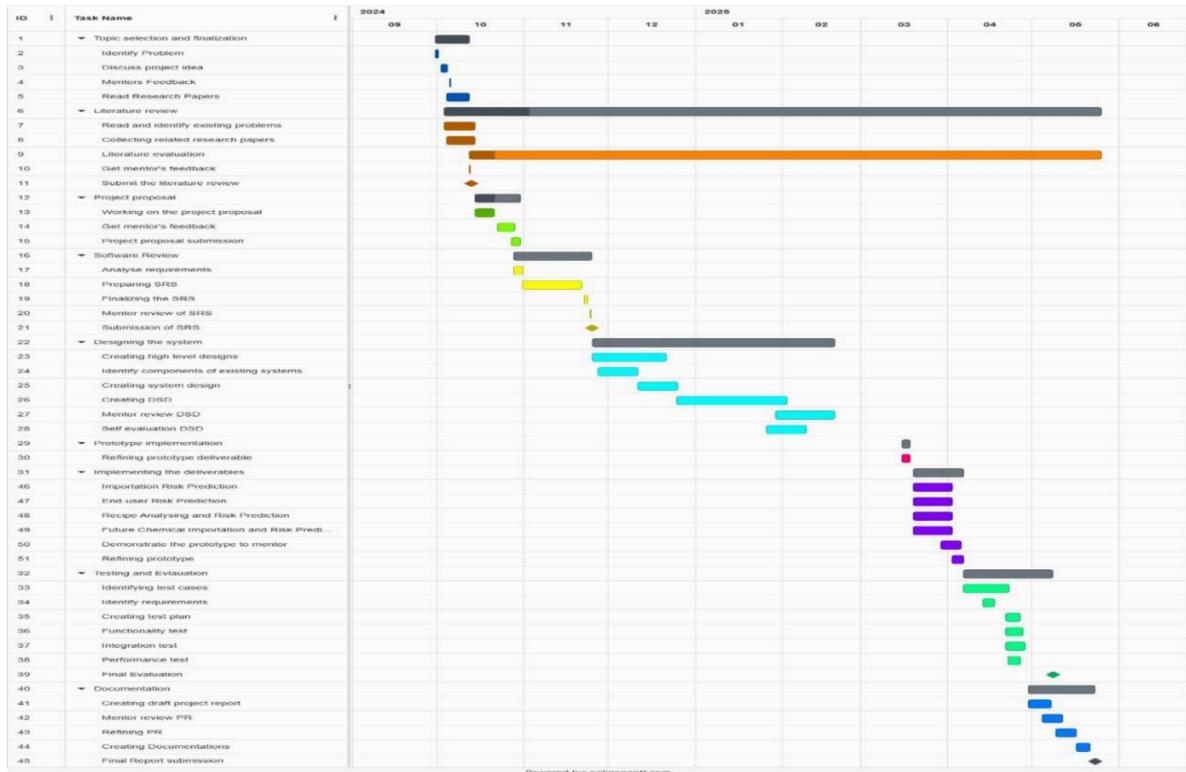


Figure 3 - Gantt Chart

3.4.2 Deliverables

Deliverable	Date
Semester 01	
Submission of literature review	Week 3
Submission of the Project Proposal to the supervisor	Week 4
Submission of the Project Proposal (Final PP)	Week 5
Submission of the SRS to the supervisor	Week 8
Submission of the SRS (Final SRS)	Week 9

Semester 02	
Prototype Implementation	Week 14
Testing and Evaluation	Week 19
Documentation and final report submission	Week 23

Table 10 – Deliverables

3.5 Chapter Summary

This chapter outlines the key contributions of stakeholders in the Comprehensive Chemical Risk Prediction Model (CCRPM) project. The system aims to improve chemical risk monitoring, prevent misuse, and enhance regulatory measures. Stakeholders, including regulatory authorities, benefit from better risk assessment, early detection of threats, and improved compliance management. The research relies on online sources and academic papers to establish a solid foundation for chemical risk prediction using AI-driven techniques.

4 Chapter 4: SOFTWARE REQUIREMENTS SPECIFICATION

4.1 Chapter Overview

This chapter outlines the requirements gathering process for the Comprehensive Chemical Risk Prediction Model (CCRPM). It identifies stakeholders, their roles, and responsibilities, and details the methodologies used to elicit requirements effectively. By leveraging these techniques, the chapter defines the functional and non-functional requirements of the CCRPM, ensuring that the system meets the needs of its users. Stakeholders are analyzed and categorized using the Onion Model, providing a clear view of their involvement in the project. Furthermore, this chapter includes the scope and priority of requirements, supporting the overall development process.

4.2 Rich Picture

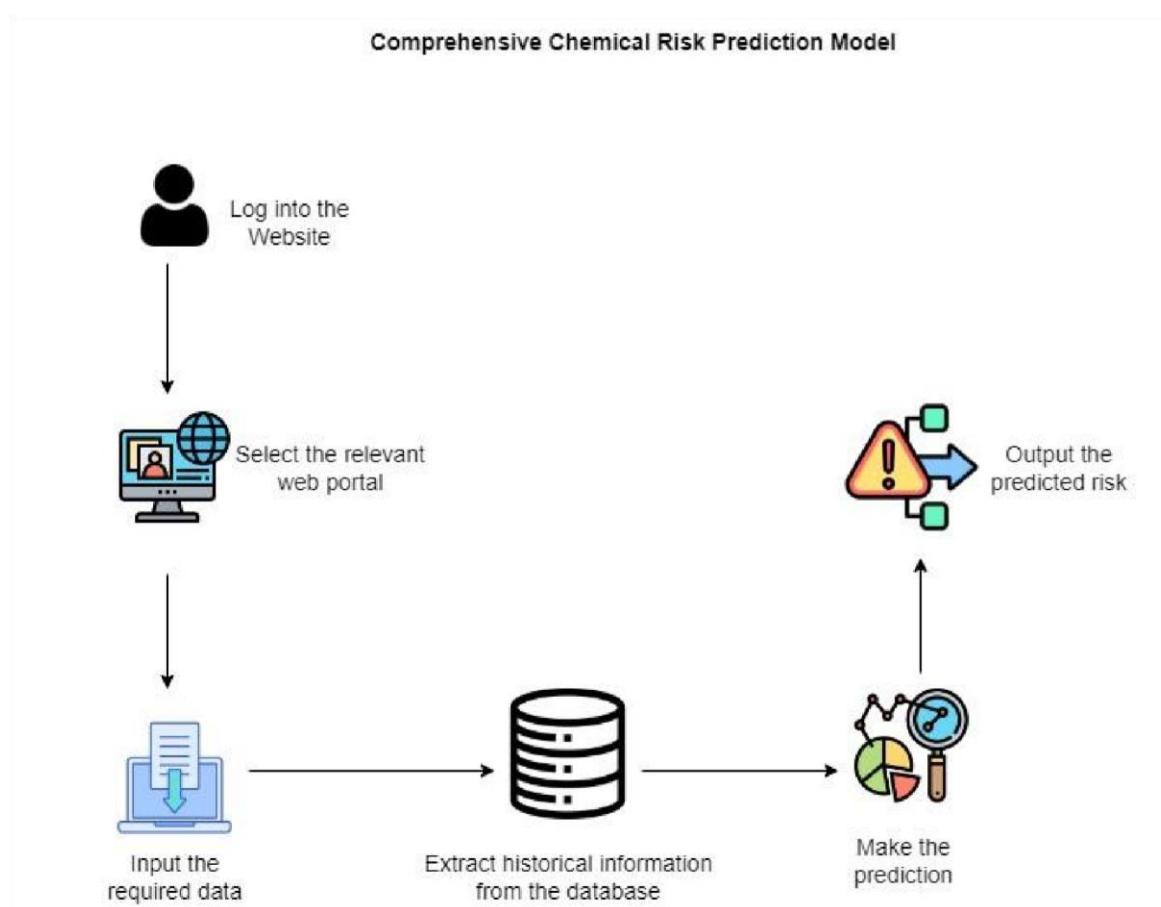


Figure 4 - Rich Picture

4.3 Stakeholder Analysis

4.3.1 Onion Model

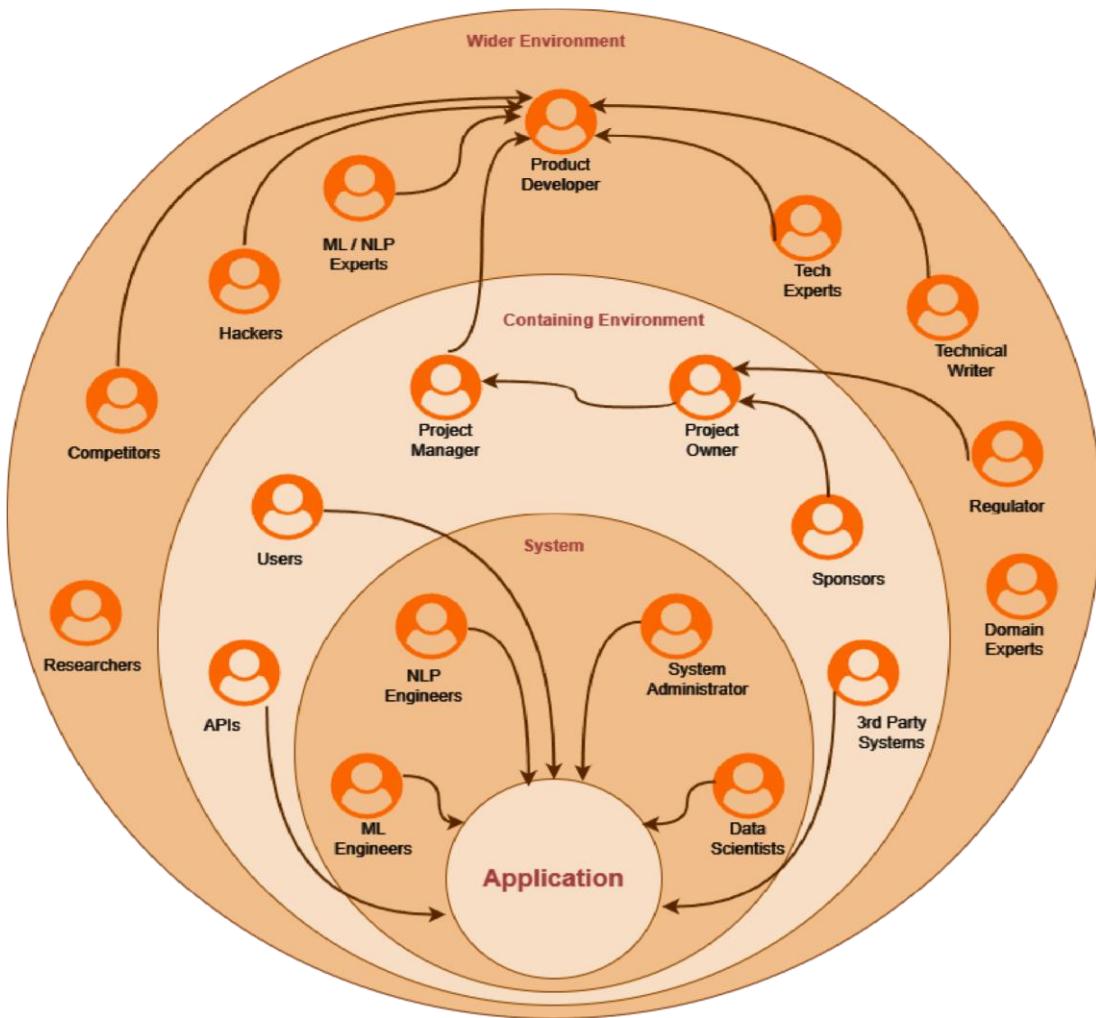


Figure 5 - Onion Model

4.3.2 Stakeholder Viewpoints

The core stakeholders for the Comprehensive Chemical Risk Prediction Model are the National Authority for Implementing the Chemical Weapons of Sri Lanka (NACWC), dual-use chemical importers and end-users like industrialists, educational institutes, researchers and home users, NLP/ML Engineers, data scientists, sponsors, domain expert, tech expert, hackers, etc.

Stakeholder	Role	Benefits
Data scientists, NLP Engineers,	Operational Maintenance	Provide accurate and efficient data processing to enhance the system's predictive capabilities.
System Admin	Operational Administration	Deploying the application and configuring for different environments
Sponsors	Financial Beneficiary	Making key investment decisions and gaining profit from that
Product Owner (NACWC of Sri Lanka)	Functional Beneficiary	Owner of the CCRPM system and Ensures compliance, strengthens monitoring, and aligns the system with national security goals.
Product Manager	Managerial Support	Managing the application process to ensure that the project runs successfully.
Users (NACWC Officials, Importers, End-users), 3rd Party Systems, APIs	Functional Beneficiary	Using the developed CCRPM application via different channels or integrating it with other systems.
Competitor	Negative Stakeholder	Drive continuous improvement to stay ahead and maintain an edge in functionality and reliability.
Hackers	Negative Stakeholder	Emphasize the need for robust security measures to safeguard data and operations.
Product Developer	Developer, Operational Maintenance	Creates and maintains the system.
Regulator	Quality Regulator	Make sure the application does not miss use any data and the data is processed by considering the privacy policies

Table 11 - Stakeholder Viewpoints

4.4 Selection of Requirement Elicitation Techniques

The requirement elicitation process for the CCRPM system was conducted to understand the needs and expectations of stakeholders in chemical risk prediction and management. The existing systems (CHEMWATCH, 2024), (Versik 3E, 2024) were analyzed to identify their features, limitations, and areas in them that can and needs to be improved were identified while performing the literature review.

Apart from the requirement elicitation conducted during the above-mentioned literature review, several interviews were conducted to gain the perspective of the NACWC officials regarding the CCRPM system, whereas an online questionnaire was conducted to collect input from the users' point of view regarding the application.

4.4.1 Interviews

It was decided that interviews would be better suited to collect information from the NACWC perspective as the quality of the information, especially since it is regarding the chemical field, was more important than quantity.

4.4.2 Questionnaire

As a majority of industries and organizations deal with chemicals in some capacity, the target audience for viewpoints related to chemical handling and risk management is extensive. Therefore, a questionnaire was the more preferable option to ensure wider coverage, gather input from diverse stakeholders, and involve more participants in the requirement elicitation process.

Interviews		Questionnaires	
Advantages	Disadvantages	Advantages	Disadvantages
<ul style="list-style-type: none"> • Direct interaction is efficient. • A person can directly ask the interviewer if there is any ambiguity. 	<ul style="list-style-type: none"> • Covering more people and a wider audience is difficult and timeconsuming. • The answers may not be straightforward 	<ul style="list-style-type: none"> • Can cover a wider and larger group of people. • Not time consuming. • Easy to analyze the results using the inbuilt tools. 	<ul style="list-style-type: none"> • Not everyone answers to question as expected manner, which can affect the result • Person who fills the questionnaire may misunderstand a question

Table 12 - Comparison of Interviews and Questionnaires

4.4.3 Followed Requirement Gathering Methods

4.4.3.1 Interview Findings

The interview provided insights into the expectations and challenges faced by NACWC officials regarding chemical import/export monitoring, end-user data management, and risk prediction. Officials emphasized the importance of data accuracy, data security, streamlined processes, and compliance with regulations. Some of the questions asked in the interviews and their answers are given below.

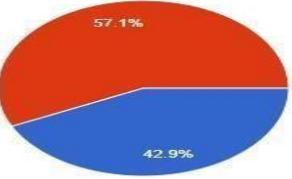
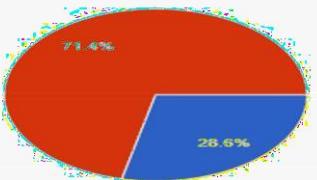
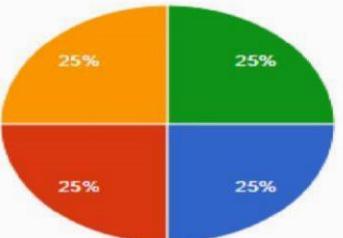
Question Number	Question	Answer
01	What are the current challenges in monitoring chemical imports and exports?	One of the biggest challenges is tracking and verifying chemical shipments in real-time. Currently, we rely on manually updated records, which can lead to delays and errors. A system that automates these processes and cross-checks them with international databases would be very helpful.
02	How would you prioritize the functions of the CCRPM system?	The top priority is risk prediction for chemical usage and movement, followed by compliance monitoring with international regulations and detailed reporting with data visualization.

03	What type of risk prediction features would be most beneficial?	A system that identifies high-risk chemicals and generates alerts for non-compliance or suspicious activity would greatly enhance oversight capabilities.
04	How important is the end-user data analysis in ensuring compliance?	It's vital. Knowing how and where chemicals are used helps us detect misuse or unsafe practices. The system should allow us to classify end-users by risk level and generate reports for enforcement and audit purposes.
05	What additional features would you find valuable in the CCRPM system?	Integration with real-time tracking systems, secure communication channels for sensitive data, and predictive modelling based on historical chemical usage trends would be highly valuable additions.

Table 13 - Questions and Answers – Interview

4.4.3.2 Questionnaire Findings

Question 01	Do you think there is a need to track and regulate the use of chemicals that might be dangerous or used in harmful ways?										
Aim of Question	To gauge public perception of the necessity to monitor and regulate hazardous chemicals.										
Observations											
<table border="1"> <thead> <tr> <th>Frequency</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Daily</td> <td>12.5%</td> </tr> <tr> <td>Weekly</td> <td>0%</td> </tr> <tr> <td>Occasionally</td> <td>87.5%</td> </tr> <tr> <td>Rarely or never</td> <td>0%</td> </tr> </tbody> </table>		Frequency	Percentage	Daily	12.5%	Weekly	0%	Occasionally	87.5%	Rarely or never	0%
Frequency	Percentage										
Daily	12.5%										
Weekly	0%										
Occasionally	87.5%										
Rarely or never	0%										
<p>87.5% of respondents believe that tracking hazardous chemicals is occasionally necessary. 12.5% felt that monitoring and regulation should occur daily.</p>											
Conclusion											
<p>While most participants see occasional regulation as sufficient, a smaller group believes daily monitoring is essential, indicating a spectrum of concern about the frequency of oversight.</p>											
Question 02	Are you aware that certain household or industrial chemicals can be harmful or hazardous if misused?										
Aim of Question	To assess public awareness regarding the potential hazards of household and industrial chemicals.										
Observations											

 <ul style="list-style-type: none"> ● Yes, very aware ● Somewhat aware ● Not aware at all 		<p>57.1% of respondents are somewhat aware of the potential hazards of household or Industrial chemicals. 42.9% reported being very aware of these risks.</p>
Conclusion		
<p>Most participants have a moderate level of awareness, with a significant portion being highly informed. This highlights a generally good understanding but suggests room for improvement in raising awareness about chemical hazards.</p>		
Question 03	<p>Do you think there is a need to track and regulate the use of chemicals that might be dangerous or used in harmful ways?</p>	
Aim of Question	<p>To gauge public perception of the importance of monitoring and regulating hazardous chemicals.</p>	
Observations		
 <ul style="list-style-type: none"> ● Strongly agree ● Agree ● Neutral ● Disagree ● Strongly disagree 		<p>71.4% of respondents agree that there are potential risks associated with chemicals in daily-use products. 28.6% strongly agree with this concern.</p>
Conclusion		
<p>The data indicates widespread concern about chemical risks in daily-use products, with a notable portion expressing strong agreement. This underscores the need for enhanced safety measures and communication regarding these risks.</p>		
Question 04	<p>How often do you check for hazard or safety information on products you buy (e.g., “flammable,” “toxic”)?</p>	
Aim of Question	<p>To evaluate consumer behaviour regarding safety information on product labels</p>	
Observations		
 <ul style="list-style-type: none"> ● Always ● Often ● Sometimes ● Rarely ● Never 		<p>The responses were evenly distributed, with 25% of participants selecting Always, Often, Sometimes, and Rarely. This shows that consumer habits regarding safety information checks vary significantly. Additionally, 10% of respondents indicated that they never check for such information, reflecting a lack of awareness or interest in product safety details.</p>
Conclusion		

The results highlight a diverse range of safety-checking habits, with a concerning minority neglecting hazard labels entirely. Increased efforts to educate the public on the importance of product safety information could encourage more consistent and responsible behaviour.

Table 14 - Questionnaire Findings

4.5 Discussion of Results

The Comprehensive Chemical Risk Prediction Model (CCRPM) requirements are well-aligned with its goal to enhance chemical risk management through functionalities like risk prediction for importers and end-users, hazardous combinations analysis, and future import trends. The use of machine learning models ensures robust, data-driven predictions, but challenges include the lack of existing datasets, the novelty of the concept, and the complexity of model selection. The Agile development approach and iterative testing will help address these challenges, while clearly defining out-of-scope items, like chatbot integration, ensures focus on core functionalities. Effective data collection and continuous feedback are critical for achieving project success.

4.6 Summary of Findings

Findings	Data Analysis	Existing Systems	Expert Insights
Sensitive and real-world data is difficult to access due to privacy, regulatory, and security constraints.	X		X
Lack of existing datasets for dual-use chemicals increases the challenge of data collection and validation.	X		
There are no prior comprehensive systems addressing end-to-end lifecycle risk management for dual-use chemicals		X	
Complexity in selecting suitable machine learning models due to limited benchmarks or prior implementations	X		

Multi-stage risk prediction (importer, end-user, hazardous combinations, and future trends) requires adaptive models and adds technical challenges.	X		X
An adaptive, machine-learning-driven approach is essential for addressing dynamic and evolving risks.		X	

Table 15 - Summary of Findings

4.7 Context Diagram

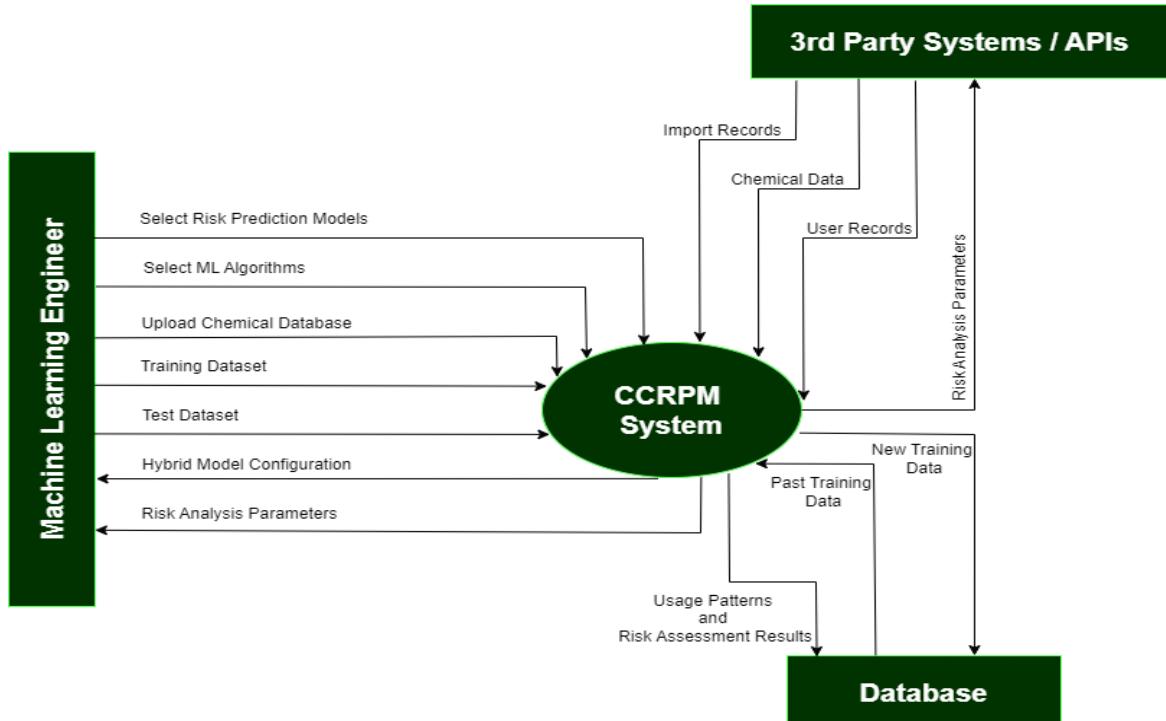


Figure 6 - Context Diagram

4.8 Use Case Diagram

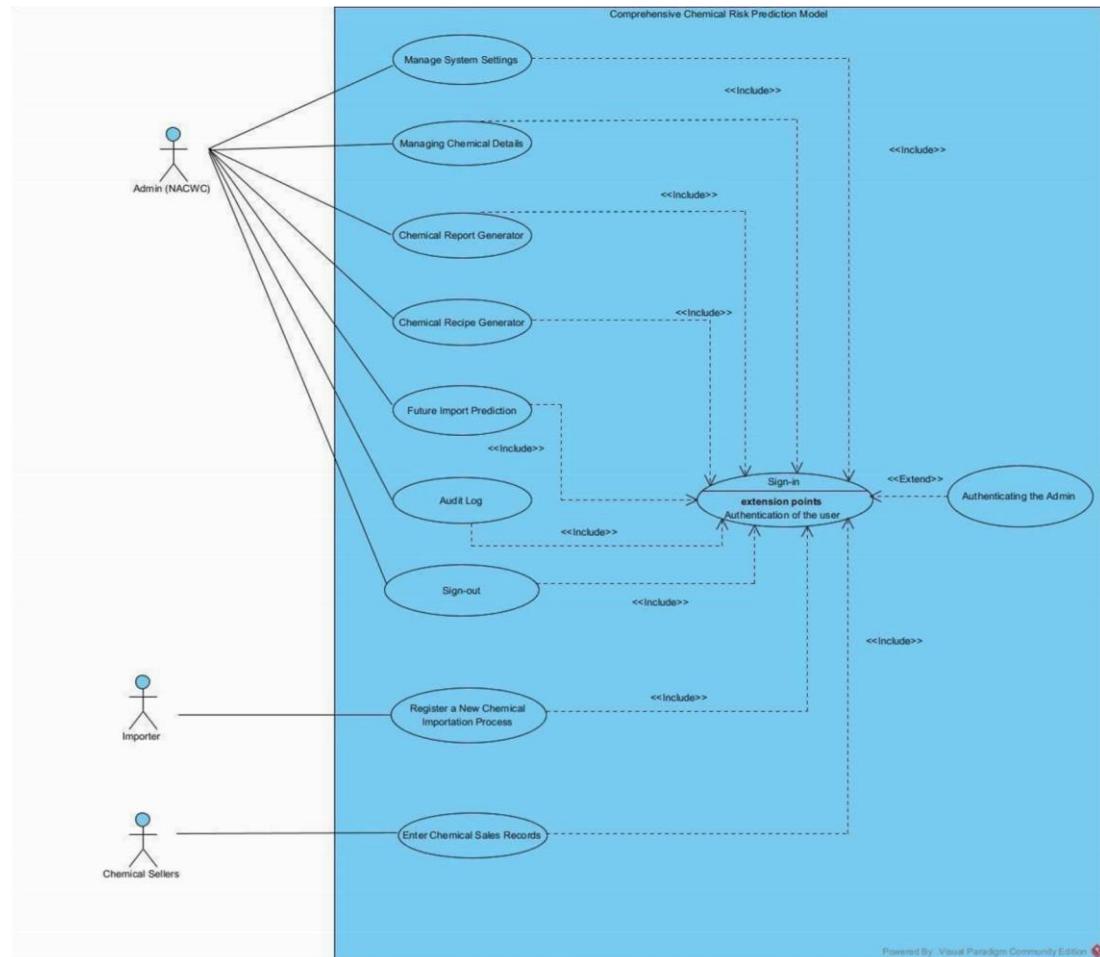


Figure 7 - Use Case Diagram

4.8.1 Use case Descriptions

Use Case 01	
Use Case Name	Manage System Settings
Description	Allows the admin to configure and update system-wide settings.
Participating Actors	Admin (NACWC)
Pre-Conditions	The admin must be signed in and authenticated.
Main Flow	<ol style="list-style-type: none"> 1. Admin selects "Manage System Settings." 2. The system displays current settings. 3. Admin modifies necessary settings. 4. Admin saves the changes. 5. The system updates and confirms the modifications.

Alternative Flow	---
Exceptional Flows	E1. System failure or invalid input.
	Use Case 02
Use Case Name	Managing Chemical Details

Description	Enables the admin to add, update, or delete chemical details in the system database.
Participating Actors	Admin (NACWC)
Pre-Conditions	Admin must be signed in and authenticated.
Main Flow	<ol style="list-style-type: none"> 1. Admin selects "Managing Chemical Details." 2. System displays a list of chemicals. 3. Admin chooses to add, update, or delete a chemical. 4. Admin inputs necessary information. 5. System validates the input and processes the request. 6. System updates the database and confirms changes.
Alternative Flow	---
Exceptional Flows	E1. Invalid input or database error.
	Use Case 03
Use Case Name	Chemical Report Generator
Description	Generates detailed reports on chemical usage, handling, and risks.
Participating Actors	Admin (NACWC)
Pre-Conditions	Admin must be signed in and authenticated.
Main Flow	<ol style="list-style-type: none"> 1. Admin selects "Chemical Report Generator." 2. System prompts admin to select report type (usage, risk, import trends). 3. Admin specifies parameters (date range, chemical type). 4. System processes the request and generates the report. 5. Report is displayed or exported as requested.
Alternative Flow	---

Exceptional Flows	E1. System failure or invalid input.
	Use Case 04
Use Case Name	Future Import Prediction
Description	Predicts future import trends and associated risks.
Participating Actors	Admin (NACWC)
Pre-Conditions	Admin must be signed in and authenticated.

Main Flow	<ol style="list-style-type: none"> 1. Admin selects "Future Import Prediction." 2. System retrieves historical import data. 3. Admin inputs prediction parameters (time frame, chemicals). 4. System processes the data and generates predictions. 5. Predictions are displayed with a risk assessment summary.
Alternative Flow	----
Exceptional Flows	E1. Data retrieval or processing failure.
Use Case 05	
Use Case Name	Audit Log
Description	Allows the admin to review logs of system activities and changes.
Participating Actors	Admin (NACWC)
Pre-Conditions	Admin must be signed in and authenticated.
Main Flow	<ol style="list-style-type: none"> 1. Admin selects "Audit Log." 2. System retrieves and displays the log of recent activities. 3. Admin filters logs based on parameters (date, activity type). 4. Logs are displayed with details (user, timestamp, action). 5. Admin reviews or exports the log as needed.
Alternative Flow	---
Exceptional Flows	E1. Data retrieval or display failure.

Use Case 06	
Use Case Name	Sign-out
Description	Securely ends the admin session.
Participating Actors	Admin (NACWC)
Pre-Conditions	Admin is signed in.
Main Flow	<ol style="list-style-type: none"> 1. Admin selects "Sign-out." 2. System prompts for confirmation. 3. Admin confirms the action. 4. System ends the session and redirects to the sign-in page.
Alternative Flow	---
Exceptional Flows	E1. Session termination failure.
Use Case 07	
Use Case Name	Chemical recipe generator
Description	To analyze and generate recipes for chemical combinations, considering risks
Participating Actors	Admin (NACWC)
Pre-Conditions	Admin must be signed in
Main Flow	<ol style="list-style-type: none"> 1. Admin selects "Chemical Recipe Generator." 2. System prompts for chemical inputs. 3. Admin provides details of the desired recipe. 4. System analyzes inputs and calculates risks. 5. System generates the recipe and displays risk factors.
Alternative Flow	---
Exceptional Flows	E1. System failure

Table 16 - Use Case Description

4.9 Functional Requirements

Priority Level	Description
Critical	Essential features and functionalities that form the core of the system.
Important	Not mandatory but considered beneficial for the system.
Non-important	Requirements that are beyond the project's current scope.

Table 17 - Priority Level

Requirement ID	Functional Requirement / Description	Priority
FR01	Analyze importer and end-user data to predict risks associated with dual-use chemicals.	Critical
FR02	Predict hazardous chemical combinations that could manufacture chemical weapons and explosives	Critical
FR03	Track end-user activities and calculate associated misuse risk	Critical
FR04	Predict future chemical import needs and assess associated risks	Critical
FR05	Maintain comprehensive database of chemical import records, end-user records, and chemical combination data	Important
FR06	GUI and other Interface support. User-friendly UI and API interfaces to connect with the system	Important
FR07	Develop data visualization tools to display risk predictions and chemical import trends	Non-important
FR08	Allow customization of the system for different countries and regulatory frameworks.	Non-important
FR09	Include a chatbot for user support and queries related to the system.	Non-important
FR10	Integrate with external APIs and databases, such as customs databases, chemical registries, and regulations.	Non-important

Table 18 - Functional Requirements

4.10 Non-Functional Requirements

Accuracy: The system must ensure accurate risk predictions by leveraging high-quality data and robust machine-learning models, minimizing false positives and negatives in risk analysis.

Performance: The system should handle large datasets efficiently, providing risk prediction results within a reasonable timeframe, even when managing high-volume imports or chemical data.

Usability: The interface must be intuitive and user-friendly for administrators, government officials, and other stakeholders, ensuring smooth operations for tasks like data entry, report generation, and risk analysis.

Scalability: As chemical data grows and regulations evolve, the system should support increased data loads and additional features without degradation in performance.

Security: The system should prevent unauthorized access, safeguard sensitive importer/end-user data and comply with international cyber security standards to ensure data integrity and confidentiality.

ID	Requirements and Description	Specification	Priority
NFR1	The system must ensure high accuracy in predicting risks, minimizing errors.	Accuracy	Important
NFR2	Risk prediction calculations and data retrieval should be performed promptly.	Performance	Important
NFR3	The system should function efficiently on standard hardware configurations.	Performance	Important
NFR4	Access to sensitive data should be role-based and protected against unauthorized users.	Security	Important
NFR5	All data logs must be encrypted to ensure confidentiality and traceability.	Security	Non-Important
NFR6	A user-friendly graphical interface should simplify risk prediction and report generation workflows.	Usability	Important

NFR7	The system should offer clear error messages and guidance for troubleshooting.	Usability	Non-Important
NFR8	The system must handle increased data volumes and support future functionality upgrades.	Scalability	Non-Important

Table 19 - Non-Functional Requirements

4.11 Chapter Summary

The chapter began with the identification of stakeholders and their roles in the CCRPM system. Various requirement elicitation methods were explored and utilized. The primary use cases and the system's context were defined. Based on these use cases, the functional and non-functional requirements were identified, analysed, and prioritized systematically to align with the project goals.

5 Chapter 5: SOCIAL, LEGAL, ETHICAL AND PROFESSIONAL ISSUES

5.1 Chapter Overview

SLEP analysis is a framework used to evaluate key factors that may impact the product proposed in this thesis. To optimize the product's potential, social, legal, ethical, and professional considerations are carefully examined. This analysis aids in identifying effective strategies to address these concerns, facilitating better decision-making, strategic planning, and problem-solving.

5.2 SLEP Issues and Mitigation

5.2.1 Social Issues

- Users and stakeholders may question how the system makes decisions, especially if a license application is denied or an importer is flagged. Lack of transparency in AI decision-making can lead to distrust. To mitigate this; provide detailed risk assessment reports to users.
- The system processes importer and end-user data, which includes company details, chemical usage, and personal information. Unauthorized access or misuse of this data could lead to privacy violations. To mitigate this; Use strong encryption, role-based access control, and comply with data protection laws.
- If the system is too strict, it could prevent universities, scientists, and industries from accessing chemicals needed for legitimate research and manufacturing. This may slow down innovation in pharmaceuticals, engineering, and scientific research. To mitigate this; introduce a special licensing system for accredited research institutions and industries.
- Importation risk predictions could lead to delays or bans on essential chemicals, affecting industries like agriculture, pharmaceuticals, and manufacturing, which rely on chemical imports. To mitigate this; create a fasttrack approval system for critical industries like healthcare, agriculture, and manufacturing.

5.2.2 Legal Issues

- Since the system is inspired by the National Authority for Implementation of the Chemical Weapons Conventions (NACWC) in Sri Lanka, it must comply with national laws regulating chemical use
- The system collects and processes sensitive importer and end-user data, raising concerns about privacy violations and legal accountability. To mitigate this; Implement strict data protection measures, including encryption, role-based access, and obedience to Sri Lanka's Data Protection Act.

- If the system fails to detect a high-risk importer or recipe, leading to an explosion, chemical spill, or illegal activity, the responsible could face legal action. To mitigate this; Use a hybrid approach combining AI risk assessment with human expert review to minimize false negatives.
- If hackers gain access to sensitive chemical data, the system operators could be held legally responsible for inadequate security measures. To mitigate this; Implement advanced cyber security measures like multi-factor authentication.

5.2.3 Ethical Issues

- Overly strict regulations might prevent researchers and educational institutions from accessing necessary chemicals. To mitigate this; develop an exemption process for legitimate researchers.
- The system could be used to track individuals or businesses beyond chemical safety concerns. To mitigate this; Implement clear usage policies that restrict the system to chemical safety.
- AI models may unfairly flag certain businesses or individuals due to biased training data, leading to discrimination in approvals. To mitigate this; Train AI models using diverse datasets to reduce bias.

5.2.4 Professional Issues

- The dataset was collected in a professional manner, with the owner's awareness and consent obtained for its use in this project. Permission was also secured from the original author of the dataset.
- Participants who completed the questionnaire and evaluation were given advance notice with enough time to prepare their responses. They were also informed about how their data would be used in this project, and permission was requested.

5.3 Chapter Summary

This chapter examines the social, legal, ethical, and professional challenges encountered and anticipated during the development of the product. Addressing these issues is essential to minimizing potential risks and obstacles that could emerge during deployment.

6 Chapter 6: SYSTEM ARCHITECTURE AND DESIGN

6.1 Chapter Overview

This chapter presents the system architecture design of the AI-Driven Chemical Handling and Risk Analysis System, which aims to predict and assess risks in chemical importation and usage. It explains the system's structure, key components, and how different modules interact to ensure secure and intelligent decision-making using component diagrams, sequence diagrams, class diagrams and the UI/UX design.

6.2 Design Goals

Design Goal	Description
Risk Prediction Accuracy	Ensure high accuracy in predicting importers' and end-users' risk levels using AI algorithms.
Automation	Automate license approval, rejection, and suspension processes based on risk assessment.
Security	Detect and prevent illegal chemical combinations and suspicious activities.
Real-time Alerts	Notify government authorities immediately about any suspicious activity.
Scalability	Design the system to handle large data volumes and future expansions.
Data Privacy	Ensure the confidentiality of importers' and users' information.
Efficiency	Optimize system performance to provide fast and accurate risk assessments.

Table 20 - Design Goals

6.3 System Architecture Design

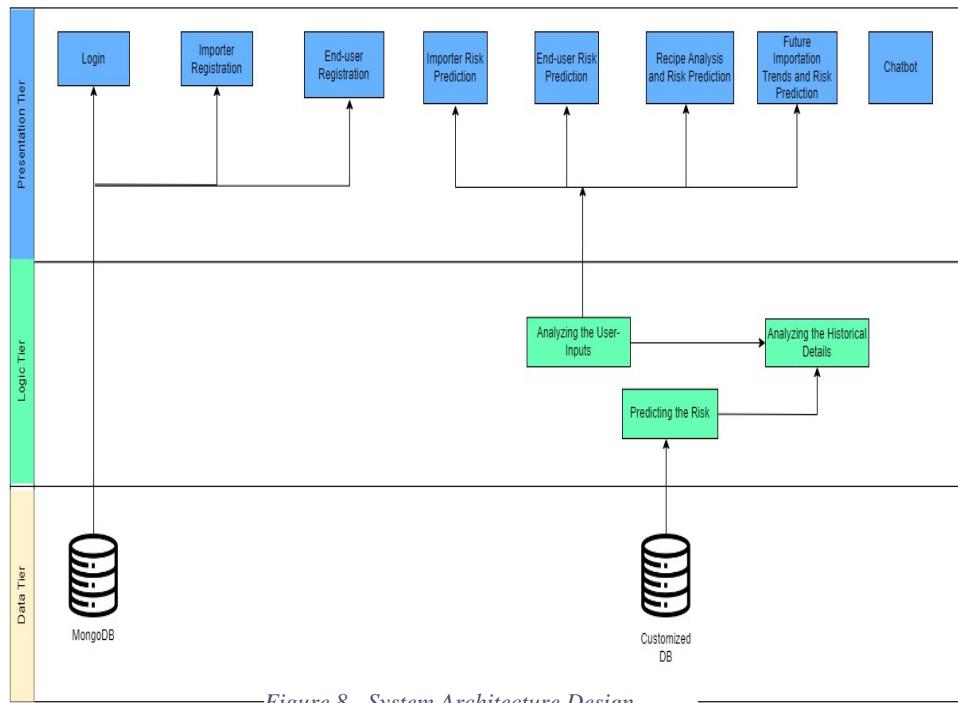


Figure 8 - System Architecture Design

The system architecture consists of three primary layers: Presentation Tier, Logic Tier, and Data Tier, which collaboratively manages the functionalities of the AI-driven chemical handling and risk analysis system

Presentation Tier

The Presentation Tier serves as the interface between users and the system, allowing users to interact with the system's functionalities.

- Login: Authenticates users before accessing the system.
- Importer Risk Prediction: Predicts the risk level of the importer based on their past importation patterns and risk factors.
- End-user Risk Prediction: Analyzes the end-user's purchase patterns to predict the likelihood of illegal chemical activities.
- Recipe Analysis and Risk Prediction: Analyzes chemical combinations to detect if they can be used for illegal purposes.
- Future Importation Trends Prediction: Predicts future importation patterns and risk assessments based on historical data.
- Chatbot: Provides automated assistance for users' queries.

Logic Tier

The Logic Tier handles the core business logic and data processing tasks, including:

- Analyzing the User Inputs: Processes user-provided data such as registration information and purchase patterns.
- Predicting the Risk: Uses AI algorithms to predict risks based on user data.
- Analyzing the Historical Details: Analyzes historical importation and purchase patterns to identify suspicious behavior.

Data Tier

The Data Tier is responsible for data storage and retrieval.

- MongoDB: Stores user registration data, chemical importation details, and risk prediction results.
- Customized DB: Stores historical data and processed risk predictions, used for further risk analysis and future trend prediction.

6.4 System Design

There are two types of system design methodologies, which are:

1. SSADM: Structured System Analysis and Design Methodology
2. OOAD: Object-Oriented Analysis and Design

6.4.1 Choice of Design Paradigm

Object-Oriented Analysis and Design (OOAD) is the design methodology adopted for the Comprehensive Chemical Risk Prediction Model (CCRPM). This approach is well-suited for object-oriented projects, ensuring a structured and efficient development workflow. By clearly defining the responsibilities of each component, OOAD enhances maintainability and scalability. It promotes encapsulation of functionality within objects, improving code readability and reducing complexity.

Given the object-oriented nature of the CCRPM system, Structured Systems Analysis and Design Methodology (SSADM) is not a suitable design paradigm. OOAD provides modularity and reusability, which are essential for developing machine learning models, as algorithms and techniques often need to be reused across different modules. Additionally, the dynamic nature of machine learning systems, such as the CCRPM, requires a flexible and adaptable design approach. OOAD is particularly effective in handling this dynamic

nature, making it the most appropriate methodology for the CCRPM system. Its ability to adapt to changing requirements and integrate new functionalities ensures the system remains robust and scalable as it evolves.

6.4.2 Component Diagram

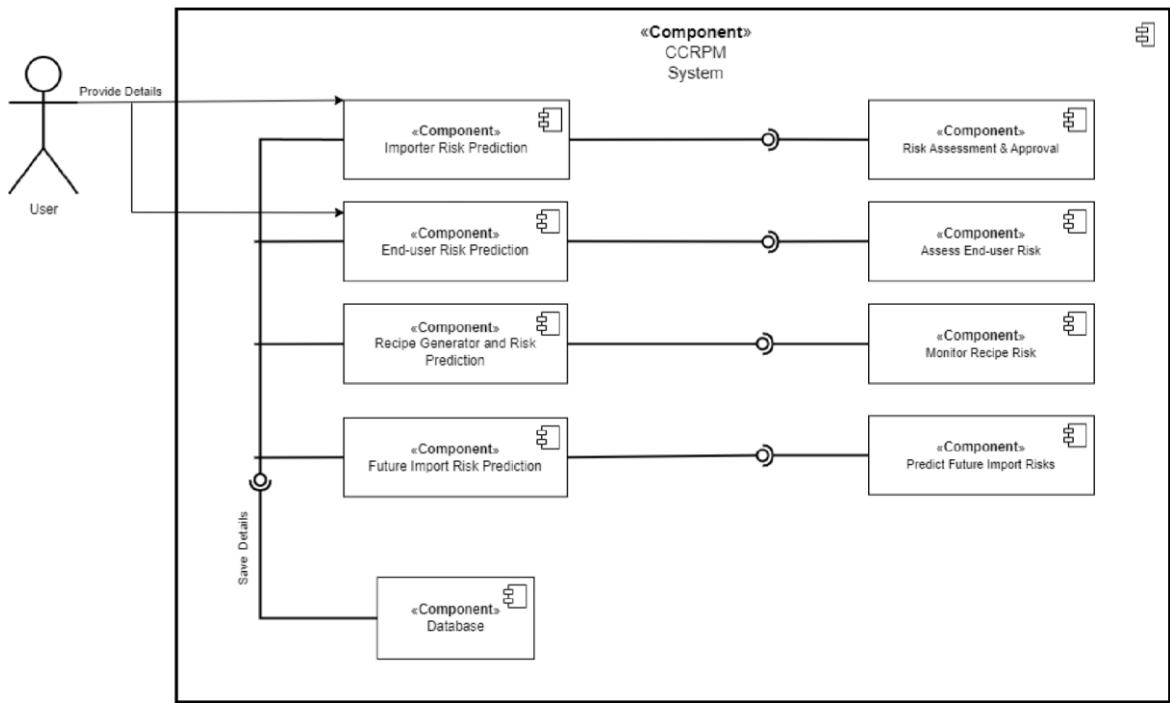


Figure 9 - Component Diagram

6.4.3 Class Diagram

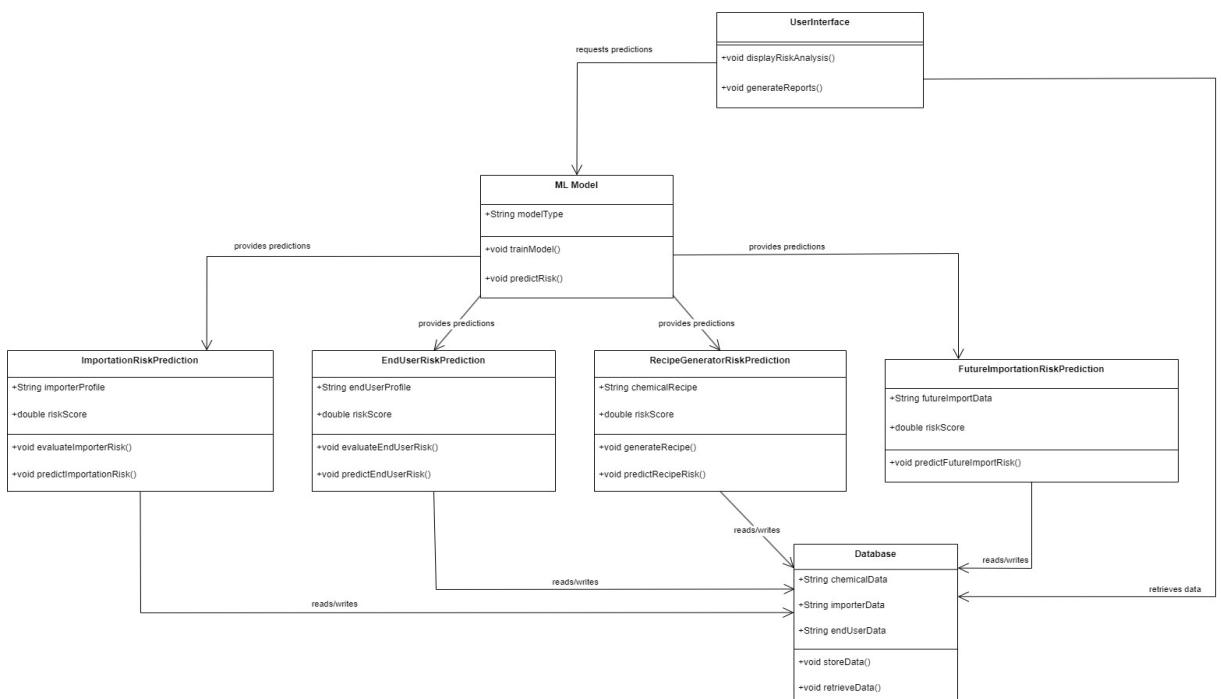


Figure 10 - Class Diagram

6.4.4 Sequence Diagram

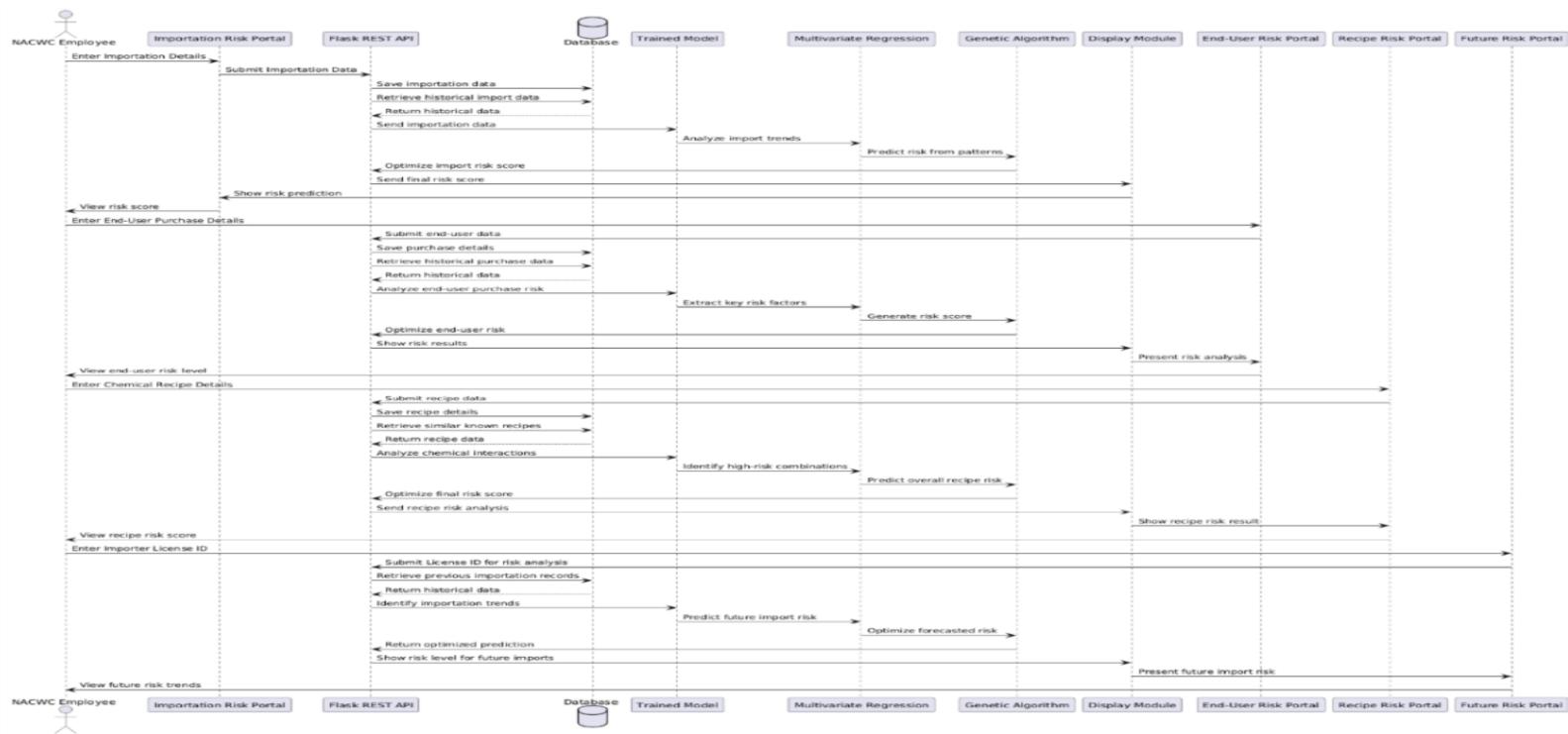


Figure 11 - Sequence Diagram

6.4.5 UI Design

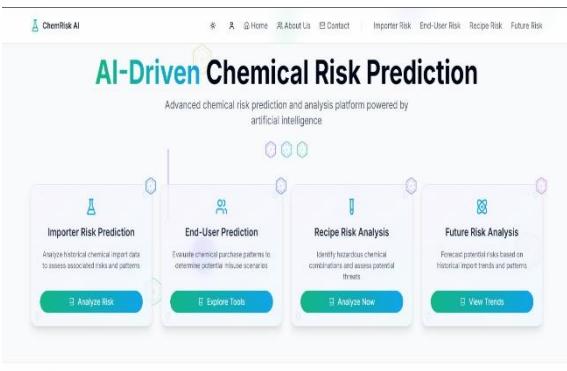


Figure 15 - Main page

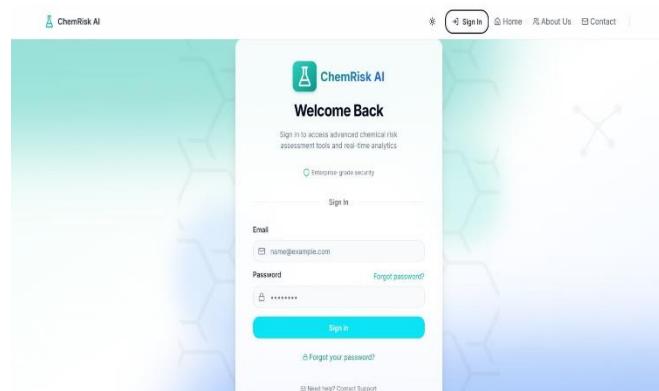


Figure 14 - Sign in page

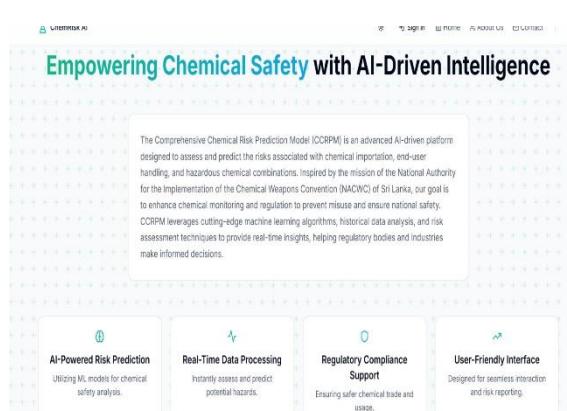


Figure 13 - Home page

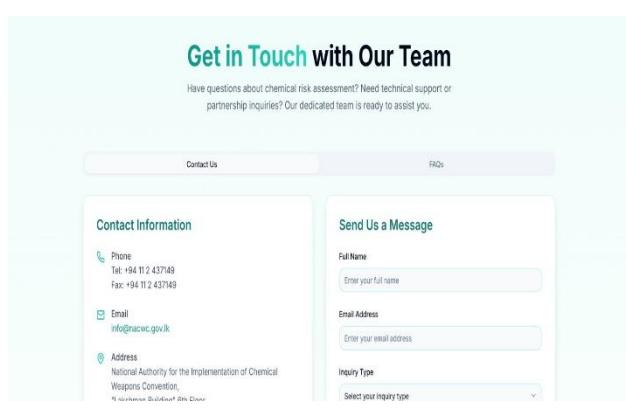
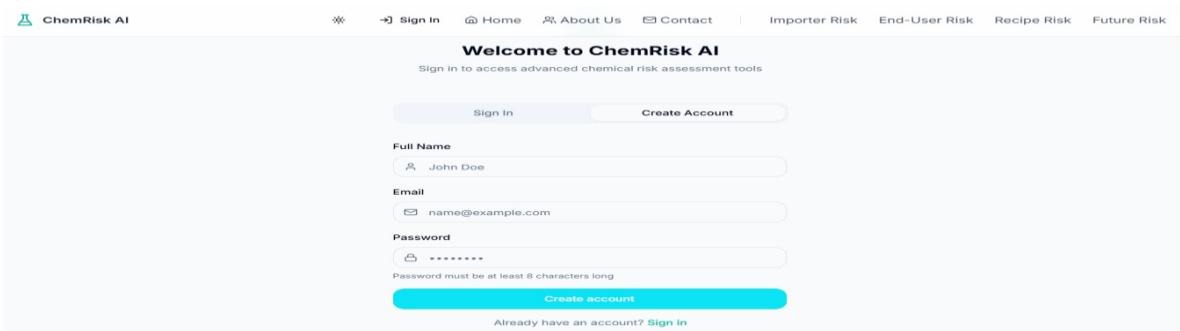


Figure 12 - About us page



Welcome to ChemRisk AI

Sign in to access advanced chemical risk assessment tools

Sign In Create Account

Full Name
John Doe

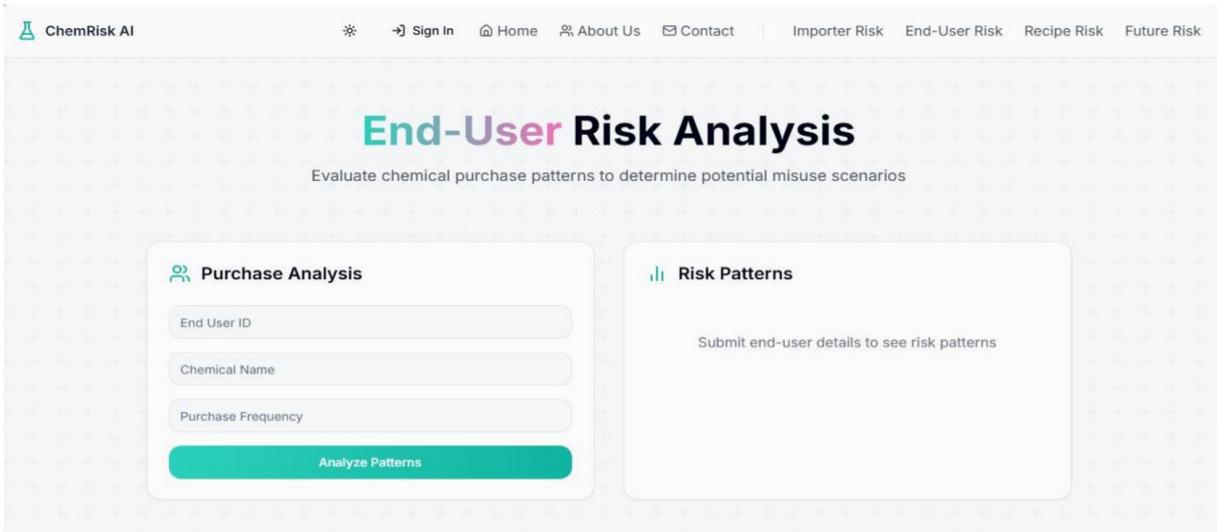
Email
name@example.com

Password

Create account

Already have an account? [Sign in](#)

Figure 16 - Sign up page



End-User Risk Analysis

Evaluate chemical purchase patterns to determine potential misuse scenarios

Purchase Analysis

End User ID

Chemical Name

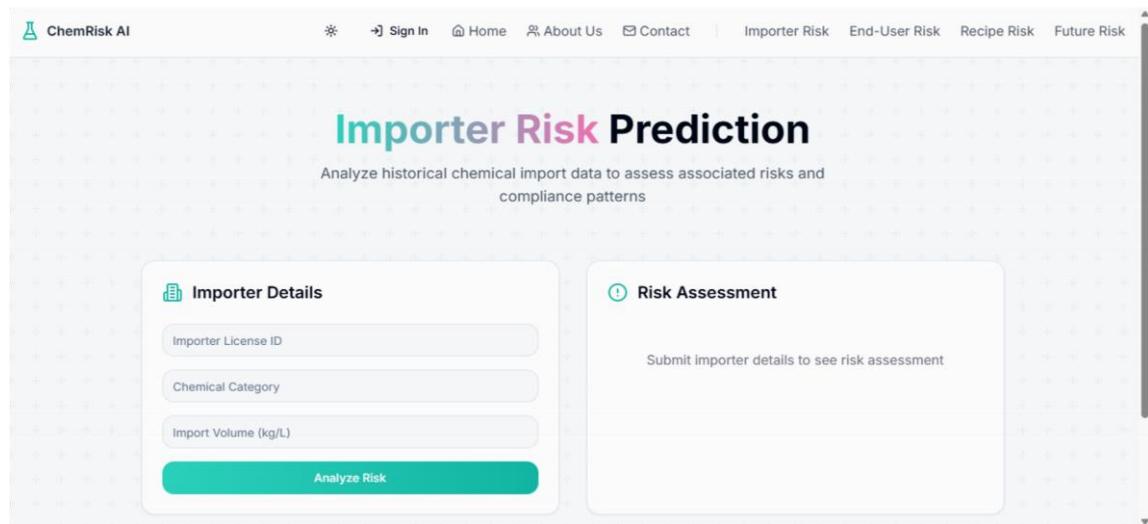
Purchase Frequency

Analyze Patterns

Risk Patterns

Submit end-user details to see risk patterns

Figure 18 - End user risk prediction



Importer Risk Prediction

Analyze historical chemical import data to assess associated risks and compliance patterns

Importer Details

Importer License ID

Chemical Category

Import Volume (kg/L)

Analyze Risk

Risk Assessment

Submit importer details to see risk assessment

Figure 17 - Importer Risk Predict

ChemRisk AI * Sign In Home About Us Contact Importer Risk End-User Risk Recipe Risk Future Risk

Recipe Risk Analysis

Identify hazardous chemical combinations and assess potential threats

Chemical Combination

Chemicals + Add Chemical

Chemical 1

Chemical Name: e.g., Sodium Hydroxide

Quantity: Amount Unit: g (grams)

Analyze Combination

Hazard Assessment

Submit chemical details to see hazard assessment

Figure 19 - Recipe Risk Analysis

ChemRisk AI * Sign In Home About Us Contact Importer Risk End-User Risk Recipe Risk Future Risk

Future Risk Prediction

Forecast potential risks based on historical import trends and patterns

Trend Analysis

License ID
Chemical Category
Time Period (months)

Analyze Trends

Future Projections

Submit details to see future projections

Figure 20 - Future Trend Prediction

6.4.6 User Experience

The compressive chemical risk management system offers a user-friendly experience with a straight-forward flow. Users sign in, then select an analyses type, and see the results in a few simple steps.

Flow and interaction,

- **Sign up & sign in:** Safe login procedure.
- **Homepage:** The main place to choose analysis option.
- **Risk Prediction pages:** User enters data, and depending on the type of analysis, the results are displayed on the same page, either as graphs, figures, numbers, or summaries.

Design Priorities: The system is designed to be simple, with clear navigation and labels to guide users through each step. Risk outcomes are displayed immediately, and risk levels are communicated effectively using visual indicators like graphs or numerical summaries.

This design ensures that users can easily assess chemical import risks without unnecessary complexity. Future enhancements will be based on user feedback and testing.

6.4.7 Process Flow-Chart

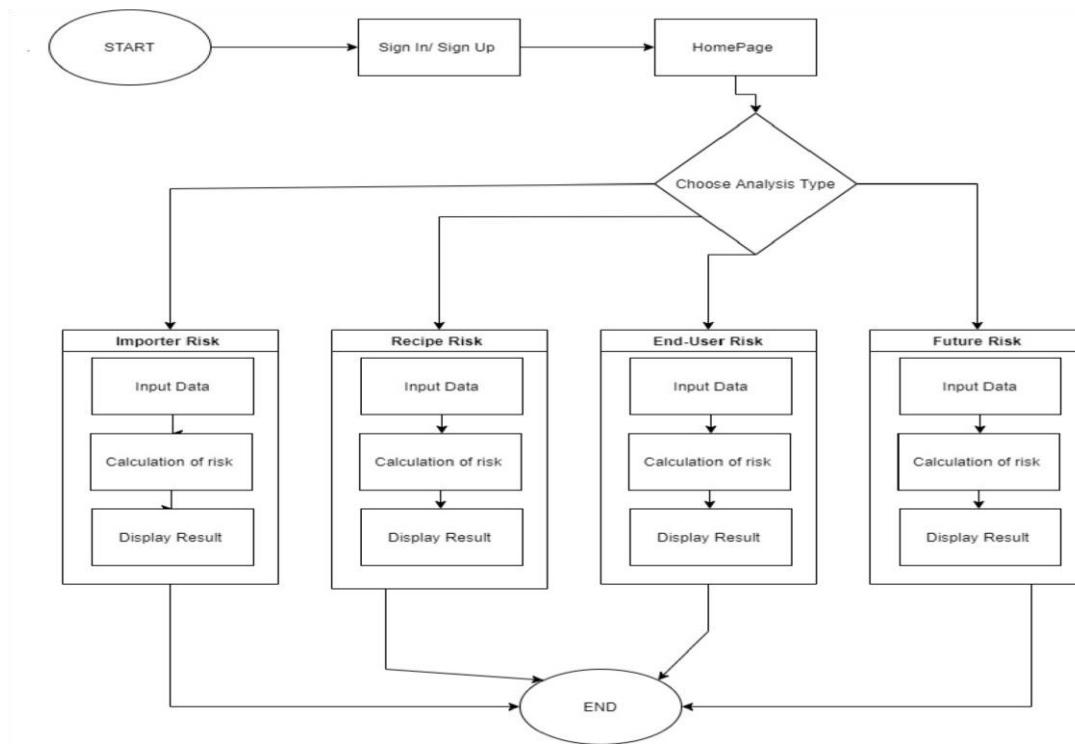


Figure 21 - Process Flow-Chart

6.5 Chapter Summary

This chapter explains the System Architecture & Design of the Chemical Risk Management System. It focuses on the main goals, like making the system efficient, improving risk prediction accuracy, and providing an easy-to-use experience. The section also covers the System Architecture Design, explaining why the chosen design was selected. Key diagrams like the Component Diagram, Class Diagram, and Sequence Diagram are included to show how the system is structured. The chapter wraps up with details on the UI Design and User Experience, ensuring a smooth and simple interaction for users.

7 Chapter 7: IMPLEMENTATION

GIT URL - <https://github.com/Shabnallmi/Data-Science-Group-Project>

7.1 Chapter Overview

This chapter presents the implementation of the AI-driven chemical handling and risk analysis system. It discusses the main tools and technologies used, the methodologies for data selection, the programming languages and frameworks utilized, and a high-level implementation in pseudo code.

7.2 Technology Selection

7.2.1 Technology Stack

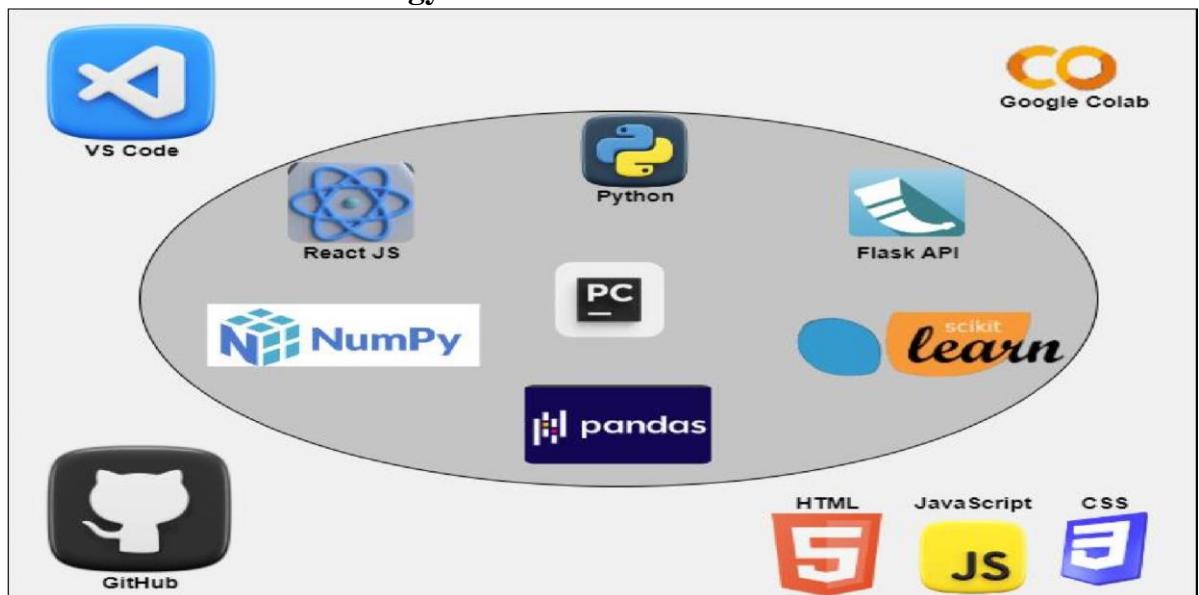


Figure 22 - Technology Stack

7.2.2 Data Selection

To build the dataset for the CCRPM system, we sourced data from the National Authority for the Chemical Weapons Convention (NACWC) and Sri Lanka Customs. The chemical importer dataset was obtained from Sri Lanka Customs, while information on hazardous chemicals was gathered from NACWC. However, we encountered a significant limitation in acquiring real-world data on chemical end users, as only a small amount of end-user data was available from NACWC. Additionally, stakeholders provided new requirements for both the importer risk prediction and end-user risk prediction components. To address these constraints and align with the updated requirements, we generated a synthetic dataset for both the importer risk assessment and end-user risk prediction. This approach ensured a

more comprehensive foundation for risk prediction and analysis in the CCRPM system, allowing for a more effective evaluation of chemical handling risks.

Domain	Dataset	Description
Chemical Importers	Sri Lanka Customs and Synthetic Data	The dataset includes chemical importation activities, including information on importers, transaction timelines, chemical values, product classifications, quantities, and detailed descriptions,
Chemical End Users	NACWC and Synthetic Data	The dataset includes chemical purchase details, with synthetic data supplementing limited end-user records from NACWC.
Hazardous chemicals	NACWC	The dataset includes chemical properties and their potential reactions.

Table 21 - Data Selection

7.2.3 Selection of Development Framework

7.2.3.1 Frontend

React was chosen for developing the web application due to its streamlined and effective approach to building modern, interactive user interfaces. Its component-based structure and declarative syntax enable developers to break down intricate UIs into reusable components, enhancing code maintainability and reusability. Moreover, React's extensive ecosystem of libraries and tools facilitates the development of dynamic, scalable, and responsive web applications. Additionally, React's virtual DOM (Document Object Model) improves performance by efficiently updating the UI without reloading the entire page. With strong community support and seamless integration with other frameworks, React remains a preferred choice for front-end development, allowing developers to build high-performance applications while maintaining clean and manageable code structures.

7.2.3.2 Backend

Python with Flask was selected for backend development because of its reliability and efficiency in managing data processing, routing, and API interactions. Python's clear syntax, simplicity, and vast library support make it a strong choice for backend solutions. Flask, being lightweight and highly adaptable, enables developers to handle routing, manage HTTP requests, and process data effectively. It facilitates seamless communication between

the frontend and backend, making it well-suited for building scalable and maintainable web applications while offering flexibility for integration with various tools and frameworks.

7.2.4 Programming Languages

Python served as the primary programming language for developing the Comprehensive Chemical Risk Prediction Model (CCRPM), alongside JavaScript, HTML, and CSS.

Python was chosen due to its simplicity, readability, and maintainability, making it a highly productive language for this project. As an interpreted language, Python executes code line by line, allowing for easy debugging and error identification. Its extensive ecosystem of libraries and frameworks played a crucial role in implementing the machine learning (ML) components of the CCRPM system.

Libraries such as Scikit-learn and Random Forest were utilized for robust risk analysis and prediction, while Flask, a Python web framework, facilitated the integration of ML components into the web application. Flask enabled the creation of RESTful APIs to handle user inputs and process data efficiently. For the frontend development, React, a popular JavaScript library, was used to build interactive and responsive user interfaces, leveraging its component-based architecture.

Together, these technologies ensured a seamless and efficient development process, combining Python's backend capabilities with JavaScript's frontend strengths to deliver a comprehensive and user-friendly system.

7.2.5 Libraries

Libraries	Version
TensorFlow	2.15.0
NumPy	1.26.3
Scikit-learn	1.4.1.post1
Pandas	2.2.1
Flask	3.0.2
Matplotlib	3.8.2
joblib	1.3.2

Table 22 – Libraries

7.2.6 IDE

VS Code was used as the primary Integrated Development Environment (IDE) for developing the Comprehensive Chemical Risk Prediction Model (CCRPM). It provided a seamless development experience with support for multiple programming languages, extensions, debugging tools, and Git integration, ensuring efficient project management and deployment.

Google Colab was utilized for model training due to its access to GPU resources, which significantly enhanced the speed of training machine learning models.

7.2.7 Summary of Technology Selection

Component	Technology/Tool	Version
Programming Language	Python	3.13
UI Frameworks	React	18.3.1
IDE	VS Code, Google Colab	-

Table 23 - Summary of Technology Selection

7.3 Implementation of Core Functionalities

7.3.1 Importer Risk Prediction

```

START
    DISPLAY
        'Enter inputs'
        chemical_Nam
        e ←
        USER_INPUT
        quantity ←
        USER_INPUT
    
```

```

        MODEL_INPUT ← PrepareModelInput(
            chemical_Name, standard_quantity)
    
```

```
RISK_LEVEL ←
TrainAndPredictRisk(MODEL_INPUT)
```

```
DISPLAY 'Predicted Risk Level: ', RISK_LEVEL
END
```

7.3.2 Future Trend Prediction

```
START
DISPLAY
‘Enter inputs’
chemical_Nam
e ←
USER_INPUT
quantity ←
USER_INPUT
```

```
MODEL_INPUT ← PrepareModelInput(
chemical_Name, standard_quantity) RISK_LEVEL ←
TrainAndPredictRisk(MODEL_INPUT)
```

```
DISPLAY 'Predicted Risk Level: ', RISK_LEVEL
END
```

7.3.3 Recipe Generator

```
START
DISPLAY 'Enter
Chemical Name and
Quantity'
chemical_name ←
USER_INPUT
quantity ←
USER_INPUT
```

```
MODEL_INPUT ←
PrepareModelInput(chemical_name, standard_quantity)
```

```
RISK_LEVEL,EXPLOSIVE_STATUS←TrainAndPredict
RecipeRisk(MODEL_INPU
T)
```

```
DISPLAY 'Risk Level: ', RISK_LEVEL
```

```
DISPLAY 'Explosive Status: ', EXPLOSIVE_STATUS
```

```
END
```

7.3.4 End-User Risk Prediction

```
START
```

```
DISPLAY 'Enter User Profile,
Chemical ID, and Usage Quantity'
user_profile ← USER_INPUT
chemical ← USER_INPUT
usage_quantity ← USER_INPUT
```

```
MODEL_INPUT ←
PrepareModelInput(user_profile,chemical_id,
standard_quantity)
```

```
RISK_LEVEL ←
TrainAndPredictUserRisk(MODEL_INPUT)
```

```
DISPLAY 'Predicted End-User Risk Level: ',
RISK_LEVEL
```

```
END
```

7.4 Chapter Summary

Mostly, this chapter focuses on the technologies and tools used in the Chemical Risk Management System. It outlines the pseudocode implementation of key components, including Importer Risk Prediction, Future Risk Prediction, Recipe Generation, and End-User Risk Prediction. Each component follows a structured flow where user inputs are collected through the web portal, passed to the backend, processed by trained models, and results are returned to the frontend for display. This ensures accurate risk assessments, helping users make informed decisions about chemical safety and risk management.

8 Chapter 8: TESTING

8.1 Chapter Overview

This chapter focuses on evaluating the performance of the Comprehensive Chemical Risk Prediction Model (CCRPM). It details the assessment of the four predictive models using key evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices. Additionally, this chapter discusses functional testing, including module-level and integration testing, to ensure seamless system operation. Benchmarking compares model efficiency, while non-functional testing evaluates load balancing, system robustness, and real-time performance under varying conditions. The chapter also includes an in-depth accuracy assessment to validate CCRPM's risk prediction and decision-making reliability.

8.2 Objectives and Goals of Testing

The primary aim of testing the Comprehensive Chemical Risk Prediction Model (CCRPM) is to validate that the system operates as expected, meeting all functional and performance requirements while ensuring reliability and security. Through thorough testing, the system becomes more reliable, reducing the chances of errors and making it more trustworthy for users.

The testing process is designed to:

- Detect and address potential errors and weaknesses that may impact the accuracy and efficiency of chemical risk assessment.
- Verify that the system functions as intended and meets user expectations, ensuring that any overlooked issues from the development phase are identified and resolved.
- Evaluate system stability, security, and performance, addressing concerns such as inaccurate predictions, potential security threats, and operational inefficiencies.

By achieving these objectives, testing plays a crucial role in reducing risks associated with AI-driven risk assessment, strengthening system dependability, and ensuring that CCRPM delivers precise and secure predictions for chemical risk management.

8.3 Testing Criteria

8.3.1 Functionality Testing Criteria

- **Importer Risk Prediction:** Validate that the system accurately analyzes past importation patterns and assesses the risk level of importers.
- **End-User Risk Prediction:** Ensure that the system correctly evaluates purchase patterns to detect potential risks associated with chemical misuse.
- **Recipe Risk Analysis:** Test the system's ability to analyze chemical combinations and identify potential risks, including illegal applications.
- **Future Import Trend Prediction:** Verify that the system accurately forecasts future chemical importation trends based on historical data.

8.3.2 Non-Functionality Testing Criteria

- **System Performance:** Evaluate how quickly the system processes risk assessments and generates predictions while ensuring it can handle large datasets efficiently.
- **Consistency and Reliability:** Verify that the system consistently delivers accurate and stable results under various conditions without unexpected failures.
- **Data Security:** Ensure that all sensitive information, including importer records, transaction history, and chemical usage data, is securely handled and safeguarded against unauthorized access or breaches.
- **Ease of Maintenance:** Check that the system's structure and code are well-organized and documented, making future updates, troubleshooting, and improvements straightforward.

8.4 Model Evaluation

The effectiveness of the Chemical Comprehensive Risk Prediction Model (CCPRM) was assessed using a confusion matrix and a classification report. These evaluation methods were selected due to their ability to provide in-depth insights into the model's predictive capabilities.

Confusion Matrix Analysis,

A confusion matrix was employed to compare predicted risk classifications against actual outcomes. This analysis quantified the following:

- True Positives (TP) – High-risk cases correctly identified.

- True Negatives (TN) – Low-risk cases correctly identified.
- False Positives (FP) – Low-risk cases mistakenly classified as high-risk.
- False Negatives (FN) – High-risk cases incorrectly predicted as low-risk.

Classification Report Metrics,

In addition, a classification report was generated to evaluate the following key performance indicators:

- Precision – The proportion of correctly predicted high-risk cases among all high-risk classifications.
- Recall – The percentage of actual high-risk cases that were correctly detected by the model.
- F1-Score – A metric that balances precision and recall.
- Accuracy – The overall percentage of correctly classified cases.

8.4.1 Importer Risk Prediction

For the Prediction of a particular import various Machine Learning Models were trained including the (Random Forest, Gradient Boost, Hybrid Model (LSTM + Gradient Boost), Light Boost, Cat Boost, XG Boost, etc,) using a dataset which contained importation details corresponding to various chemical importers and their past importation activities. The model tested focused on a binary classification whether a particular import is risky or not risky.

Performance Comparison

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Observations
Random Forest (10 trees)	100.0%	-	100.0%	-	Overfitting detected
Gradient Boosting Classifier (GBC)	59.0%	-	59.45%	-	The Predictions were fair when compared with other models.

Hybrid Model (LSTM + Gradient Boost)	91.38%	-	92.0%	-	A slight overfitting is detected on a particular column of the dataset.
Cat Boo st, Light Boost & XG Boost	100.0%	-	99.0%	-	Overfitting is detected.

Table 24 - Importer risk model comparison

In the model training, even though the accuracy was low, the Gradient Boost Classifiers and the hybrid Model (LSTM + Gradient Boost) provided accurate predictions when compared with the other models.

Classification Report

Classification Report:			precision	recall	f1-score	support
Not Risky	0.46	0.65	0.54	37		
Risky	0.62	0.43	0.51	49		
accuracy			0.52	86		
macro avg	0.54	0.54	0.52	86		
weighted avg	0.55	0.52	0.52	86		

Figure 23 - Importer risk classification report

Confusion Matrix

The confusion matrix was created based on 150 test cases, as shown below:

Metric	Description	Value
True Positives (TP)	Correctly classified risky importers	21
True Negatives (TN)	Correctly classified not-risk importers	24
False Positives (FP)	Not-risk importers misclassified as risky	13
False Negatives (FN)	Risky importers misclassified as not risk	28

Table 25 - Importer risk confusion matrix

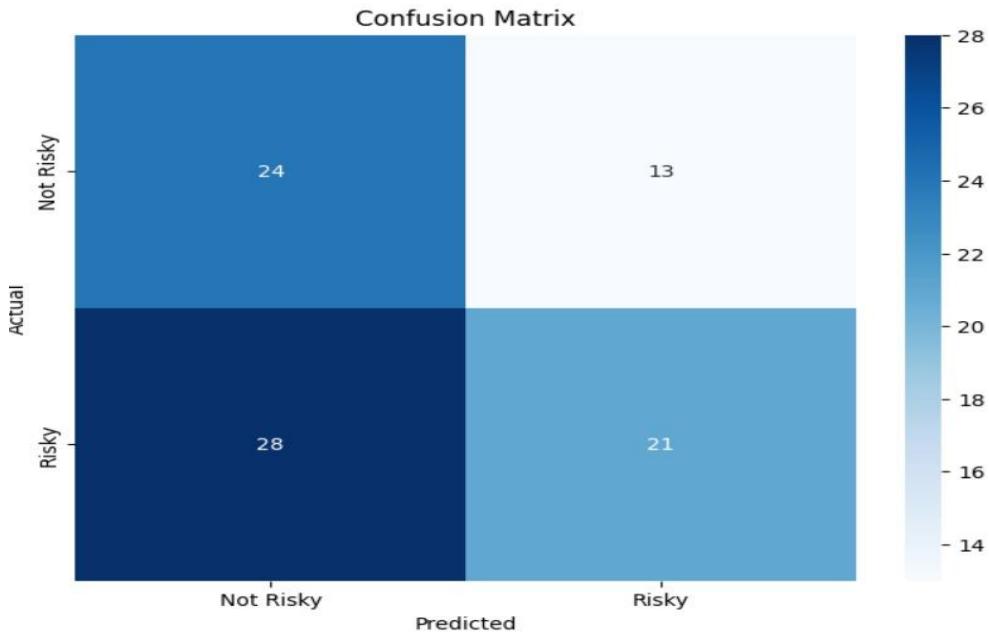


Figure 24 - Importer risk confusion matrix

8.4.2 End User Risk Prediction

To receive the prediction many models were trained including the Random Forest, Gradient Boost and Logistic Regression. The final prediction is a binary classification in which the output would be either risky or not risky.

Performance Comparison

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Observations
Gradient Boosting	100%	0.46%	99%	3.14%	Slight overfitting
Random Forest	100%	1.48%	98%	4.76%	Balanced, less overfitting
Logistic Regression	91%	26%	92%	30%	Slight overfitting

Table 26 - End User model comparison

In this case we chose Random Forest over Gradient Boosting because while both models performed similarly (98% vs 99% validation accuracy), Random Forest offers better computational efficiency and easier interpretability. The marginal accuracy gain from

Gradient Boosting didn't justify its higher complexity and resource requirements for our chemical risk prediction use case.

Classification Report

Classification Report:			precision	recall	f1-score	support
0	0.99	0.99	0.99	367		
1	0.92	1.00	0.96	34		
2	0.88	0.88	0.88	17		
			accuracy		0.98	418
			macro avg	0.93	0.96	0.94
			weighted avg	0.98	0.98	0.98

Figure 25 - End User classification report

Confusion Matrix

Metric	Description	Value
True Positives (TP)	Correctly classified high-risk end-users	362
True Negatives (TN)	Correctly classified low-risk and medium risk end-users	49
False Positives (FP)	Low-risk and medium risk end-users misclassified as high-risk	5
False Negatives (FN)	High-risk end-users misclassified as low risk and medium risk	2

Table 27 - End User confusion matrix

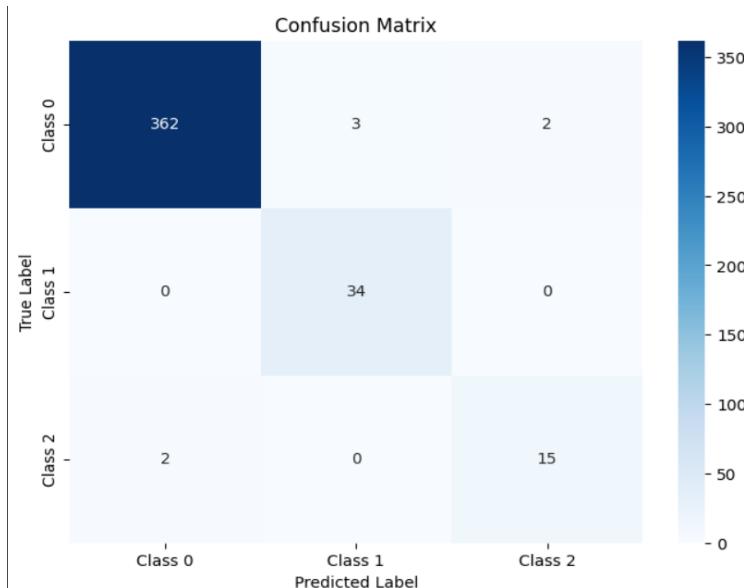


Figure 26 - End User confusion matrix

8.4.3 Chemical recipe risk analysis

For predicting the chemical risk levels, various machine learning models were trained using a dataset containing chemical recipes and their corresponding risk labels. The models tested focused on multi-output classification and regression to predict multiple risk categories (flammability, explosiveness, health risk, etc.).

Performance Comparison

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Observations
Multi-Output Classification	85%	32%	78%	45%	slight overfitting
Multi-Output Regression	-	15%	-	2%	slight overfitting

Table 28 - Chemical recipe model comparison

In this case, the multi-output classification model performed well in terms of accuracy and validation, showing slight overfitting. The multi-output regression model also delivered strong results, performing best for predicting continuous risk scores.

Classification Report

Overall Classification Report for Risk Levels:				
	precision	recall	f1-score	support
High Risk	0.12	0.33	0.18	3
Low Risk	1.00	0.50	0.67	4
Medium Risk	0.33	0.25	0.29	4
Very High Risk	0.82	0.74	0.78	19
accuracy			0.60	30
macro avg	0.57	0.46	0.48	30
weighted avg	0.71	0.60	0.64	30

Figure 27 - Chemical recipe classification report

Confusion Matrix

Metric	Description	Value
True Positives (TP)	Correctly classified high-risk Chemical recipes	14
True Negatives (TN)	Correctly classified low-risk Chemical recipes	4
False Positives (FP)	Low-risk recipes misclassified as high-risk recipes	3
False Negatives (FN)	High-risk recipes misclassified as low-risk recipes	5

Table 29 - Chemical recipe confusion matrix

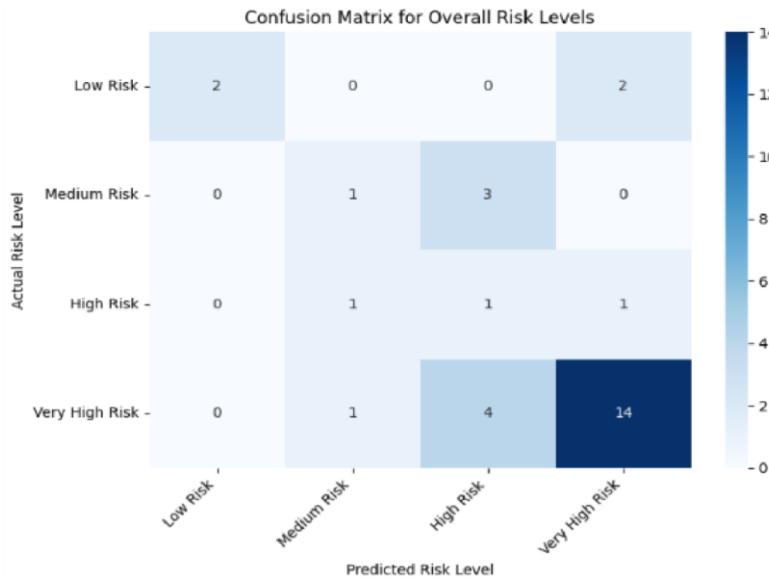


Figure 28 - Chemical recipe confusion matrix

8.4.4 Future Trend Prediction

To receive the prediction many models were trained including the Random Forest and Neural Networks. The final prediction is a binary classification in which the output would be either imports increase or decrease.

Performance Comparison

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Observations
Neural Network	83%	12%	95%	35%	Slight overfitting
Random Forest	100%	-	100%	-	Overfitting detected

Table 30 - Future trend Performance Comparison

Classification Report

Neural Network Accuracy: 95.00%				
Neural Network Classification Report:				
	precision	recall	f1-score	support
0	0.94	0.94	0.94	33
1	0.96	0.96	0.96	47
accuracy			0.95	80
macro avg	0.95	0.95	0.95	80
weighted avg	0.95	0.95	0.95	80

Figure 29 - Future trend Classification Report

Confusion Matrix

Metric	Description	Value
True Positives (TP)	Correctly classified class 1 instances	45
True Negatives (TN)	Correctly classified class 0 instances	31
False Positives (FP)	Class 0 instances misclassified class 1 instances	2
False Negatives (FN)	Class 1 instances misclassified as class 0 instances	2

Table 31 - Future trend Confusion Matrix

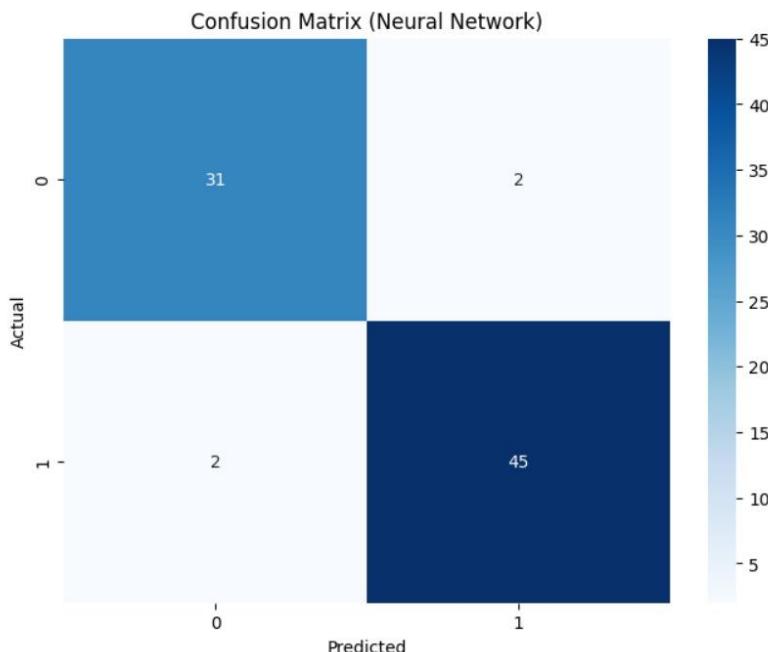


Figure 30 - Future trend Confusion Matrix

8.5 Benchmarking

The purpose of benchmarking is to compare the CCPRM model with similar systems in the industry to assess its effectiveness and identify areas for improvement. Since no complete AI-driven chemical risk prediction models currently exist, comparisons were made using research papers and industry articles.

System Component	Human-Level & State-of-the-Art Performance	Comparison of the Model with Similar Approaches
Importer Risk Prediction	<p>Risk assessments require experts analyzing past import patterns and legal records, which is time-consuming and prone to human error.</p>	<p>CCPRM uses a Gradient Boosting Classifier (GBC) to predict importer risk based on past import patterns and legal history, whereas previous studies have used Logistic Regression or Decision Trees, which may not capture complex dependencies in import trends.</p>
End-User Risk Prediction	<p>Traditional risk assessment methods rely on manual investigations of purchase patterns, which may miss hidden patterns indicating illegal activity.</p>	<p>CCPRM integrates GBC with anomaly detection techniques to identify suspicious purchase behavior. Other models have primarily relied on rule-based approaches or Random Forest, which are less effective for high-dimensional risk factors.</p>
Recipe Risk Analysis	<p>Chemical recipe analysis requires chemical experts to manually evaluate the risks, leading to inconsistencies.</p>	<p>CCPRM uses Gradient Boosting Classifier to evaluate chemical compositions and predict potential risks of illegal synthesis. Previous works have explored neural networks, but they often lack interpretability, whereas CCPRM balances accuracy and explainability.</p>
Future Import Trend Prediction	<p>Predictions are traditionally based on economic forecasting models, requiring extensive data processing.</p>	<p>CCPRM employs a Gradient Boosting Regressor (GBR) to predict future import trends. Similar studies have used Linear Regression and ARIMA models, which may not handle complex non-linear relationships as effectively.</p>

Table 32 – Benchmarking

8.6 Functional Testing

Unit testing has been used to verify the correctness of each functional component.

Test Case	Description	Input	Expected Output	Actual Output	Result
01	Importer risk assessment based on past import history and legal background.	Importer data (import history, legal records)	High/Medium/Low Risk Classification	High/Medium/Low Risk Classification	PASS
02	End-user risk assessment based on purchase patterns and past behaviour.	Purchase history data	High/Medium/Low Risk Classification	High/Medium/Low Risk Classification	PASS
03	Recipe risk analysis: Detect if a given chemical composition is hazardous or could be used illegally.	Chemical recipe	High/Medium/Low Risk Classification	High/Medium/Low Risk Classification	PASS
04	Future import trend prediction: Predict demand and potential increase or decrease in upcoming chemical imports.	Past import data, market trends	Future import volume and decrease or increase in imports	Future import volume and decrease or increase in imports	PASS

Table 33 - Functional Testing

8.7 Module and Integration Testing

Each component was systematically tested to integrate the Comprehensive Chemical Risk Prediction Model (CCRPM) to ensure seamless functionality. The system's efficiency in predicting chemical risks and providing regulatory insights was validated by focusing on critical modules, verifying interfaces, and testing with relevant data. The integration process also emphasized data security to ensure reliable and compliant operation.

Integration Testing Steps:

1. **Test Independently:** Each component of CCRPM, including Importation Risk Prediction, End-User Risk Prediction, Recipe Analyzing, and Future Chemical Importation Risk Prediction, was tested independently for functionality and accuracy.
2. **Start with Important Components:** Priority was given to modules with a critical impact on data analysis and risk prediction, ensuring these were thoroughly tested first.
3. **Check Connections:** Communication and data flow between different modules, including the front-end interface and back-end processing, were tested to ensure seamless integration.
4. **Use Real Data:** Realistic chemical data and trade scenarios were used to evaluate system performance and validate predictions.
5. **Run Tests and Keep Records:** Tests were systematically conducted, with detailed records maintained for analysis and improvements.
6. **Scan for Problems:** All test outcomes were monitored to detect inconsistencies or integration issues.
7. **Fix Errors:** Identified errors were promptly addressed, and the affected modules were re-tested for validation.
8. **Repeat Testing:** Iterative testing cycles were conducted to confirm system robustness and accuracy.

8.8 Non- Functional Testing

8.8.1 Accuracy Testing

Accuracy in CCRPM measures the system's ability to correctly classify and predict the risk levels associated with chemical importation and usage. It is calculated as the ratio of correct predictions to the total predictions made.

The CCRPM model underwent extensive accuracy testing using historical data and synthetic scenarios to ensure precise and reliable risk categorization. Continuous evaluation and refinement were implemented to optimize prediction outcomes.

8.8.2 Performance Testing

Performance testing assessed the system's responsiveness, stability, and scalability under varying conditions.

- **Output Time:** The system is designed to deliver risk prediction results within seconds of data input, ensuring prompt decision-making support for regulatory authorities.
- **Scalability:** The system's ability to handle increasing data volumes and concurrent user requests was validated.
- **Stability:** Long-duration tests ensured that the system maintains consistent performance without degradation over time.

This approach ensures that CCRPM remains efficient and reliable, even under demanding operational conditions.

8.8.3 Load Balancing

Load balancing divides network requests across multiple servers to improve system performance, prevent overload, and ensure high availability. A few popular techniques include Round Robin, Least Connections, and weighted load balancing. While the current system is on one local server, thus not requiring load balancing, future scalability may require it, especially if the system is moved to cloud hosting or handles high user traffic. For these cases, solutions such as Nginx, HAProxy, or AWS ELB can distribute requests across multiple instances of Flask API for reliability and performance

8.9 Limitations

During the testing process, the following limitations were identified:

- High processor usage during machine learning model training.
- Class imbalance in risk prediction affects model accuracy.
- Lack of publicly available data requires manual data collection.
- The system does not predict risks for unseen chemicals.
- No real-time updates: predictions rely only on past data.

8.10 Chapter Summary

This chapter provides an extensive study of the tests that were conducted to evaluate the functionality and performance of the Comprehensive Chemical Risk Prediction Model (CCRPM). The system was tested using functionality testing, module and integration testing, benchmarking, and non-functional testing in order to ensure it meets project requirements. Performance, accuracy, and risk classification efficiency were thoroughly evaluated. Certain limitations were also found, such as class imbalance, computational complexity, and scalability limitations which were addressed in this chapter

9 Chapter 9: EVALUATION

9.1 Chapter Overview

This chapter presents the evaluations conducted by domain experts and technical specialists on the Comprehensive Chemical Risk Prediction Model (CCPRM). The system has been assessed across multiple dimensions, including functionality, performance, accuracy, usability, and compliance with safety regulations. Both functional and non-functional aspects of the system were critically reviewed, with feedback collected to refine and enhance its capabilities. Additionally, an internal assessment was conducted by the authors to ensure that the system meets its intended objectives and aligns with real-world chemical risk prediction requirements. The insights from these evaluations provide valuable recommendations for improving the system's effectiveness and reliability.

9.2 Evaluation Methodology and Approach

This system predicts risks in chemical handling by analyzing importer details, past importation data, end-user purchase patterns, and chemical recipes. It also forecasts future importation trends using multivariate regression. The goal is to detect potential misuse, prevent illegal applications, and ensure safe chemical distribution.

The system is evaluated by chemical safety experts for accuracy, reliability, and compliance with regulations. Technical experts assess its model performance, usability, user-friendliness, and error handling. Key evaluation areas include the effectiveness of risk prediction, the accuracy of chemical recipe analysis, system response time, and the reliability of future trend forecasts.

9.3 Evaluation Criteria

The system was evaluated based on the following key areas:

- **Risk Prediction Accuracy** – How well the system detects and predicts risks in chemical handling.
- **Scope and Applicability** – Whether the system covers important risks and follows regulations.
- **System Design and Performance** – The efficiency and reliability of the system's structure and models.
- **Usability** – How easy it is for users to interact with the system and understand risk assessments.

- **Compliance and Security** – Ensuring the system follows safety regulations and protects data.
- **Error Handling** – How well the system manages incorrect or missing data.
- **Recommendation Accuracy** – The effectiveness of suggested actions based on detected risks.

9.4 Self-Evaluation

Criteria	Author's Evaluation
Project Concept	The Comprehensive Chemical Risk Prediction Model (CCRPM) is designed to assess and predict potential risks associated with chemical usage, importation, and combinations. This system aids regulatory bodies in making informed decisions, ensuring compliance with safety regulations, and mitigating risks related to hazardous chemicals.
Scope of the Project	The project focuses on chemical risk analysis, including Importer Risk Prediction, End-User Risk Prediction, Recipe Analysis, and Future Importation Risk Prediction. It is designed to support regulatory authorities, traders, and other stakeholders in assessing risks associated with dual-use chemicals.
System Design, Architecture, and Implementation	The system is implemented as a web-based platform integrated with React.js (frontend), Flask (backend), and MongoDB (database). Machine learning models are deployed via Flask APIs to provide accurate risk analysis. The system is built with a user-friendly interface, allowing seamless input of chemical data and retrieval of risk predictions.
Solution and Prototype	A fully functional web application has been developed, featuring chemical risk prediction modules, a secure authentication system, and an intuitive user interface. The prototype has undergone extensive testing to ensure accuracy, security, and performance before deployment.

Table 34 - Self Evaluation

9.5 Selection of Evaluators

Two types of evaluators have been selected:

1. Chemical Inspector
2. Senior Software Engineer

The domain experts are Ms. Radika Madupani and Mr. Amaradewa De Silve, a Chemical Inspectors specializing in chemical safety and risk assessment.

The technical expert is Mr. Akindu Perera, a Senior Software Engineer with expertise in web development

9.6 Evaluation Results

9.6.1 The concept of the project

Question	
What is your perspective on the core concept of this project?	
Person	Feedback
Ms. Radika Madupani	"The project presents an advanced approach to chemical risk assessment, improving monitoring and compliance with safety regulations. Its AI-driven analysis enhances efficiency in evaluating hazardous substances."
Mr. Akindu Perera	"Integrating AI into chemical risk assessment is a forward-thinking approach. The project concept is strong, addressing key challenges in industrial safety and regulatory processes."

Table 35 - Evaluation Results- 1

9.6.2 Scope of the project

Question	
What do you think about the scope of the project?	
Person	Feedback
Ms. Radika Madupani	"The project has broad applications in chemical regulation, customs inspections, and industrial safety. With proper validation, it can significantly aid regulatory authorities in risk mitigation."
Mr. Akindu Perera	"The AI-driven approach allows for scalable and adaptable solutions, making the system applicable across multiple industries, from chemical trade monitoring to environmental safety."

Table 36 - Evaluation Results- 2

9.6.3 System design, architecture and implementation

Question	
What do you think about the system structure, architecture, and deployment?	
Person	Feedback
Ms. Radika Madupani	"The structured risk evaluation process ensures consistent and reliable results. Proper integration with regulatory databases would enhance its effectiveness."
Mr. Akindu Perera	"The web-based deployment with Flask and MongoDB ensures scalability and secure data handling. The AI model integration enhances automation and efficiency in risk analysis."

Table 37 - Evaluation Results- 3

9.6.4 Solution and prototype

Question	
What is your perspective about the proposed solution and prototype development?	
Person	Feedback
Ms. Radika Madupani	"The prototype shows strong potential in streamlining chemical risk assessment. Further testing with real-world data will enhance its reliability and practical impact."
Mr. Akindu Perera	"The prototype effectively demonstrates the system's capabilities. Performance optimization and security enhancements will be key for large-scale deployment."

Table 38 - Evaluation Results- 4

9.7 Limitations

Person	Suggestion/ Feedback
Mr. Amaradewa De Silva	Collaborate with more organizations (e.g., customs, environmental agencies) to gather additional real-world datasets, especially for end-users. While synthetic data is useful, consider improving its quality.
Ms. Radika Madupani	Instead of fixed thresholds use dynamic thresholds based on historical data or industry standards and Assign weights to different risk factors

	(e.g., chemical type, purchase frequency, quantity) to improve the accuracy of risk predictions.
Mr. Akindu Perera	Develop interactive dashboards for users to visualize risk levels, trends, and compliance status in real-time and Allow users to generate custom reports based on specific criteria (e.g., chemical type, time period, risk level).

Table 39 – Limitations

9.8 Evaluation on Functional Requirements

No	Requirement and Description	Evaluation	Priority
FR01	Login and Accessing System: Users must be able to enter username (hospital ID) and password to login and access the system	Implemented	Critical
FR02	Chemical Importation Risk Prediction: The system should evaluate importer profiles and predict risks associated with chemical importation.	Implemented	Critical
FR03	End-User Risk Prediction: predict risks of misuse or accidents.	Implemented	Critical
FR04	Recipe Generation and Risk Prediction: Allow users to input chemical recipes and predict associated risks and Flag recipes that could lead to the creation of illegal or dangerous compounds.	Implemented	Critical
FR05	Future Chemical Importation Risk Prediction: Predict risks associated with future chemical imports based on historical data and trends.	Implemented	Critical
FR06	Reporting and Analytics: Generate detailed reports on risk levels, compliance status, and chemical usage trends.	Not Implemented	Important
FR07	Efficiency: The system must perform efficiently.	Implemented	Critical

FR08	Security: the system must be able to protect data integrity, confidentiality and availability.	Implemented	Important
FR09	Compatibility: The system must be compatible with different devices, operating system and browsers.	Not Implemented	Important
FR10	Scalability: The system should be able to handle large volumes of traffic without any complications.	Not Implemented	Important
FR11	Databases: The system must have a database to save user details for saving login credentials. The user must be able to save patient data for future needs.	Not Implemented	Important
FR12	Interface: The system must have a user-friendly interface to make it simple and easy to use.	Implemented	Important

Table 40 - Evaluation on functional requirements

9.9 Evaluation on Non-Functional Requirements

No	Requirement and Description	Evaluation	Priority
NFR01	Performance – The system should provide real-time risk predictions with minimal latency.	Implemented	High
NFR02	Scalability – The system should handle increasing data volume and multiple user requests.	Not Implemented	Medium
NFR03	Security – The platform must protect sensitive chemical import data from unauthorized access.	Implemented	High
NFR04	Usability – The user interface should be intuitive for regulatory authorities and analysts.	Implemented	High
NFR05	Reliability – The system should function correctly under different conditions.	Implemented	high
NFR06	Compatibility – The application should work across different devices and browsers.	Implemented	Medium

Table 41 - Evaluation on non-functional requirements

9.10 Chapter Summary

This chapter evaluates the Comprehensive Chemical Risk Prediction Model (CCRPM) based on defined criteria. It covers the evaluation methodology, self-assessment, and selection of evaluators. The results highlight system strengths and areas for improvement, while limitations are addressed. Finally, the functional and non-functional requirements are assessed, confirming the successful implementation of key system capabilities.

10 Chapter 10: CONCLUSION

10.1 Chapter Overview

This chapter provides a comprehensive review of the Comprehensive Chemical Risk Prediction Model. It begins by outlining the system's objectives and purpose, followed by a reflection on the team's experience during its development. The challenges encountered throughout the project are also discussed. Additionally, the chapter highlights potential areas for enhancement, limitations of the current model, and ethical concerns related to chemical risk management. The final section summarizes the individual contribution made by each team member.

10.2 Achievements of Research Aims and Objectives

10.2.1 Project Aim

The Comprehensive Chemical Risk Prediction Model is designed to analyze and assess the risks associated with chemical importation, distribution, usage, and formulation. Using historical data, the system identifies high-risk importers, evaluates end-user purchase behavior, and assesses chemical recipe risks to prevent illicit activities. The model not only predicts future import trends and potential hazards using machine learning techniques but also analyzes chemical formulations to detect dangerous or illegal combinations. The AI-driven system can detect suspicious transactions and automatically recommend license suspension, trade restrictions, or chemical bans to regulatory authorities when necessary.

10.2.2 Completion of Objectives of the Project

Task	Completion Status
Literature Review	
Review of existing chemical risk assessment models and AI-driven regulatory approaches.	Completed
System Requirements and Specifications	
Identified system requirements, regulatory constraints, and stakeholder needs.	Completed
System Architecture and Design	

Designed the AI-powered web system and database framework.	Completed
Implementation and Development	
Built the prototype for risk prediction and decision support.	Completed
Testing and Validation	
Conducted accuracy testing, system functionality validation, and scenario-based assessments.	Completed

Table 42 - Completion of Objectives of the Project

10.3 Utilization of Knowledge from the Course

Module	Description
Programming Fundamentals (CM1601)	The module taught basic Python programming concepts, helping to understand the language and use it for this project.
Object Orientated Development (CM2601)	This module provided basic knowledge of Object-Oriented Programming (OOP) using Java. It also covered best practices for developing programs, including writing clean code and using version control to track changes and collaborate effectively.
Machine Learning (CM2604)	The module covered machine learning concepts and taught how to build and assess machine learning models.
Web Technology (CM1605)	This module introduced basic web development skills using HTML, CSS, and JavaScript, which played a key role in developing the web application for this project.
Data Science Group Project (CM2603)	This lesson provided a foundation for a collaborative data science project and included a list of exam questions. It also covered how to conduct a Literature Review, SRS, and other key components, helping to structure research and project development. Prior proposals were presented, which closely resembled the thesis.

Table 43 - Utilization of Knowledge from the Course

10.4 Use of Existing Skills

The project's development was built on existing skills, which played a key role in ensuring its successful and timely implementation.

10.4.1 Machine Learning

Our foundational understanding of machine learning was developed through a structured combination of online courses on LinkedIn Learning, supplemented by practical tutorials on YouTube and hands-on exploration of GitHub repositories. This versatile learning approach provided deep insights into the workings of machine learning models, enabling us to effectively develop and implement AI solutions using tools like Jupyter Notebook, TensorFlow, Keras, and scikit-learn. The blend of theoretical knowledge and applied practice from these resources equipped us with the skills to build and optimize robust machine learning models.

10.4.2 Web Development - Front End

During the web development module in the first year of our degree program, we learned and applied HTML, CSS, and JavaScript to build the application's frontend. These technologies were instrumental in creating a functional, visually appealing, and user-friendly interface, ensuring an engaging experience for users.

10.4.3 Python

The authors developed proficiency in Python through the Programming Fundamentals module of their academic curriculum, complemented by online resources such as interactive tutorials, documentation, and coding challenges. This foundation enabled them to implement key system functionalities, including data processing, machine learning model development, and backend logic. By utilizing Python's libraries (e.g., Pandas, NumPy, Scikit-learn) and frameworks (e.g., Flask), they ensured robust and scalable performance.

10.5 Use of New Skills

10.5.1 Machine Learning

Through this project, we enhanced our machine learning expertise across key areas including data preprocessing, feature engineering, and model deployment. We implemented techniques like PCA for dimensionality reduction and SMOTE for handling imbalanced datasets during preprocessing. For model development, we worked with algorithms such as Random Forest, Gradient Boost, Cat Boost, etc.., and LSTM networks for sequential data analysis. The project also provided hands-on experience in data visualization using libraries like Matplotlib and Seaborn, and we successfully integrated trained models into the application backend via API endpoints for real-time predictions.

10.5.2 Web Development – Backend

For the project's backend development, we leveraged modern frameworks like Flask and React. React's component-based architecture streamlined the creation of interactive user interfaces, while Flask served as a lightweight yet powerful backend framework for developing robust RESTful APIs. Working with these technologies enhanced our backend development skills, enabling us to build scalable and efficient web applications.

10.6 Achievement of Learning Outcomes

10.6.1 Skills developed through collaborating within a team on a software development project

- **Teamwork and Communication** – Effective coordination among team members for dataset preparation, model selection, and API integration.
- **Project Management** – Managing timelines, setting milestones, and dividing tasks efficiently.
- **Problem-Solving** – Addressing challenges like low model accuracy and API integration issues.
- **Technical Proficiency** – Gaining hands-on experience with machine learning, Flask API development, and MongoDB authentication.

- **Version Control & Collaboration** – Using GitHub for code management and version control.
- **Testing & Evaluation** – Conducting model performance testing and ensuring system reliability.

10.6.2 Analysis of the User-Centered Design Process and its Impact on Legal, Ethical, Professional, and Social Issues in Data Science Applications

- **Legal Considerations** – Ensuring compliance with chemical regulations, safeguarding sensitive data.
- **Ethical Concerns** – Avoiding misuse of chemical risk prediction, maintaining transparency in AI decisions.
- **Professional Responsibilities** – Developing an accurate and reliable system for regulatory bodies and stakeholders.
- **Social Impact** – Enhancing safety by predicting chemical risks, aiding authorities in monitoring dual-use chemicals.
- **User Experience & Accessibility** – Designing an intuitive interface for traders and regulators to assess chemical risks easily.
- **Bias & Fairness** – Addressing dataset limitations to ensure unbiased predictions in risk assessment.

10.7 Problems and Challenges Faced

Problems/Challenges	Solution/Workarounds
Limited dataset availability for chemical risk prediction	Created a synthetic dataset and structured chemical recipes for effective model training.
Time constraints due to multiple project components	Used task prioritization, version control, and team collaboration to manage workload effectively.

Integration of Flask API with the React frontend	Debugged API requests, optimized data flow, and ensured smooth communication between components.
Handling unexpected technical issues	Adopted a problem-solving mindset, leveraged online resources, and conducted trial-and-error testing.

Table 44 - Problems and Challenges

10.8 Deviations

In the Future Chemical Importation component of the CCRPM system, we deviated from the original plan to predict and display risk levels for future chemical imports. After reevaluating the project's practical needs, we determined that analyzing import trends (e.g., volume increases/decreases) would provide more actionable insights than speculative risk scoring. This shift allowed the system to focus on detectable patterns in chemical demand while avoiding the uncertainties of long-term risk projection.

We were unable to acquire sufficient real-world chemical data, which was originally planned for the model. As an alternative, we generated synthetic datasets to simulate the system, enabling functional development while acknowledging that real-world data would enhance accuracy in future implementations.

10.9 Limitations of the Development

Though effective, the Comprehensive Chemical Risk Prediction Model (CCRPM) has a few shortcomings:

- **Limited Data** – Owing to the sensitivity of chemical import data, stakeholders and the authorities tend to limit access to actual-time records. To counter this, we had to create synthetic data with the aid of AI, which might not entirely represent actual circumstances.
- **Class Imbalance** – The data set had a skewed risk level distribution. Even after applying techniques like SMOTE for balancing, there may be slight biases.
- **No Predictions for Unseen Data** – The model can only do nothing but examine chemicals found in the training set, limiting its ability to examine new substances.

- **Shortage of Real-Time Updates** – The system relies on historical data and doesn't change dynamically in response to sudden regulatory updates or newly evolving risks

10.10 Future Enhancements

Several improvements can be made to enhance the system's effectiveness and usability:

- **Integration with Regulatory Databases** – Integration with governmental and industry databases will improve risk assessment accuracy and ensure compliance with evolving regulations.
- **Performance Optimization** – Optimizing AI models to reduce computational costs and improve real-time risk predictions.
- **Real-World Data Testing** – Expanding data collection through collaboration with regulatory bodies and industries for improving model accuracy.
- **Dynamic Risk Thresholds** – Instead of fixed thresholds, utilize dynamic risk thresholds based on historical trends and industry benchmarks.
- **Interactive Dashboards** – Developing real-time dashboards for users to observe trends, risk levels, and compliance status.
- **Custom Report Generation** – Enabling users to generate reports based on defined parameters such as chemical type, time period, and risk levels.
- **Real-Time Monitoring & Alerts** – Creating a notification system to alert stakeholders of potential risks in real time.
- **Handling Unseen Data** – Enhancing the model to predict risks for chemical imports encountered for the first time.

10.11 Achievement of the contribution to body of knowledge

This research paper suggests a Comprehensive Chemical Risk Prediction Model (CCRPM) that enhances regulatory frameworks by integrating machine learning (ML) techniques for chemical risk evaluation. The contributions are as follows:

- **Technological Advances:** The framework uses supervised ML models, like classification and multivariate regression, to predict importers', end-users', and chemical

recipes' risks. Using a Flask REST API-based platform provides a user-friendly web interface for better usability.

- Risk Avoidance through Data Analysis:** Pattern examination and forecast for prospective trends techniques are applied in the model so that detection of potential risks is enabled before importation or exploitation. It further ensures automation in verification of chemical safety regulations.
- Counteracting Security & Ethical Problems:** As opposed to previous models, the current system accounts for dual-use chemicals and supports regulators' interpretability and transparency. Ethical frameworks for AI and methods for risk assessment in real time may be enhanced through additional research in the future.

By making these contributions, the CCRPM significantly enhances chemical risk management, offering an expandable and flexible solution to regulatory agencies worldwide.

10.12 Individual Contribution

The CCRPM system was developed through coordinated teamwork, with each member contributing specialized expertise to key components as detailed below.

Team Member	Contribution
Shabna Ilmi	<p>Main role – AI Engineer</p> <ul style="list-style-type: none"> • AI model for importer risk prediction • Frontend for the model. • UI Design of the web application.(Importer Part) • Data Handling and Preprocessing • Validations for the entire system • Backend for the model using Flask and backend integration
Senuth Perera	<p>Main role – AI Engineer</p> <ul style="list-style-type: none"> • AI model for predict chemical end-user risk • Backend for the model using Flask. • UI Design of the web application.(End-user part) • Functional and Non-functional testing.

	<ul style="list-style-type: none"> • Data Handling and Preprocessing
Kalana Kannangara	<p>Main role – AI Engineer</p> <ul style="list-style-type: none"> • AI model for Recipe Risk Analysis • UI Design of the web application.(Main and chemical recipe part) • Data Handling and Preprocessing • Backend for the model using Flask.
Loganathan Thusharkanth	<p>Main role – AI Engineer</p> <ul style="list-style-type: none"> • AI model for future trend prediction • Data Handling and Preprocessing • Investigation of Techniques and Algorithms • Backend for the model using Flask. • UI Design of the web application.(Future trend part)

Table 45 - Individual Contribution

10.13 Chapter Summary

This chapter concluded the project by summarizing key achievements, limitations, and possible improvements. It highlighted how the study achieved its objectives, utilized technical skills, and contributed to chemical risk prediction.

Key limitations, such as costly computation, lack of real-time data, and handling unseen chemicals, were addressed with optimizations. Possible future improvements include incorporating real-time data and improving risk assessment models.

Finally, individual-level contributions were emphasized, with an emphasis on data pre-processing, machine learning, and system development work, as a solid foundation for future research.

References

- ✓ Ahmet Murat Erturan, G.K. & H.D., 2023. *Machine learning-based approach for efficient prediction of toxicity of chemical gases using feature selection*. [Online]. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0304389423008993?via%3Dhub> [Accessed 10 Feb 2025].
- ✓ Aldahmani, E.A.A.K.I., 2024. *Demand Forecasting in Supply Chain Using UniRegression Deep Approximate Forecasting Model*. [Online]. Available at: https://www.researchgate.net/publication/383922347_Demand_Forecasting_in_Supply_Chain_Using_Uni-Regression_Deep_Approximate_Forecasting_Model [Accessed 11 Oct 2025].
- ✓ Aliabadi, H.A.K., S.D.-K.M.I.M.F.Z.B., 2022. *Developing a Model for Managing the Risk Assessment of Import Declarations in Customs based on Data Analysis Techniques*. [Online]. Available at: https://journals.iau.ir/article_688300_ff07b750e04f2fe49714307d6b1ea629.pdf [Accessed 10 Feb. 2025].
- ✓ Anuar, N.I.P. & H.A.A.R.A., 2019. *Integrated chemical, technology & equipment process knowledge management system based on risk-based process safety*. [Online]. Available at: <https://iopscience.iop.org/article/10.1088/1757-899X/702/1/012052> [Accessed 11 Oct 2025].
- ✓ Cahn, D. & R.C., 2014. *Best Practices in Chemical Management for Textile Manufacturing*. [Online]. Available at: <https://publications.iadb.org/en/publications/english/viewer/Best-Practices-InChemical-Management-for-Textile-Manufacturing.pdf> [Accessed 11 Oct 2025].
- ✓ Gawlik-Kobylińska, M., G.G., Ł.S.N., K.H.K.-E. & M.U., 2021. *The EU-SENSE System for Chemical Hazards Detection, Identification, and Monitoring*. [Online]. Available at: <https://www.mdpi.com/2076-3417/11/21/10308> [Accessed 10 Oct 2025].

- ✓ Google Scholar, 2024. *Fuzzy logic and neural network-based risk assessment model for import and export enterprises: A review.* [Online]. Available at:
<https://ojs.bonviewpress.com/index.php/jdsis/article/view/1078> [Accessed 11 Oct 2025].
- ✓ He, Z., W.X.Z., L.F.W.Y.C.T.Y., 2023. *Game mechanism and event-triggering based Natural Gas Demand Prediction (GMET-NGDP) model for Chemical and Fertilizer Industry.* [Online]. Available at:
<https://ieeexplore.ieee.org/document/10303108> [Accessed 11 Oct 2025].
- ✓ Jellis, V.M.D.P.B., 2020. *Customs fraud detection.* [Online]. Available at:
https://repository.uantwerpen.be/docman/irua/2bb8d6/163580_2020_10_30.pdf [Accessed 11 Oct 2025].
- ✓ Jeong, K., J.-Y.L., S.W.D., K.Y.J., T.I.R., S.-R.H. & W.-H.J., 2022. *Vapor Pressure and Toxicity Prediction for Novichok Agent Candidates Using Machine Learning Model.* [Online]. Available at:
<https://pubs.acs.org/doi/10.1021/acs.chemrestox.1c00410> [Accessed 11 Oct 2025].
- ✓ Luo, N., H.Y.Z., Y.Y.L., T.Z.Y.J., N.H.C.L., Z.J. & S.Q., 2023. *Fuzzy Logic and Neural Network-based Risk Assessment Model for Import and Export Enterprises: A Review.* [Online]. Available at:
<https://ojs.bonviewpress.com/index.php/jdsis/article/view/1078/510> [Accessed 11 Oct 2025].
- ✓ Ilagan, L.C. & E.P.D., 2024. *Hazardous Chemicals Detection and Classification Through Millimeter Wave and Machine Learning.* [Online]. Available at:https://www.jstage.jst.go.jp/article/jaciii/28/4/28_753/_pdf-char/ja [Accessed 11 Oct 2025].
- ✓ Payne, M.K., A.W.N.W., R.H.C.M.S., 2020. *The Chemical Management System (CMS): A Useful Tool for Inventory Management.* [Online]. Available at:
https://pubs.acs.org/doi/epdf/10.1021/acs.jchemed.9b00905?ref=article_ope_nPDF [Accessed 11 Oct 2025].

- ✓ Riley, P.C., S.V.D.B.S.I.R.D.C.E.D.K.P.O.B.C.H., 2022. *Interpreting chemical detection alarms with live analysis of ML algorithms*. [Online]. Available at: <https://www.spiedigitallibrary.org/conference-proceedingsofspi/12116/2619166/Interpreting-chemical-detection-alarms-with-live-analysisofML-algorithms/10.1117/12.2619166.full> [Accessed 11 Oct 2025].

- ✓ Seair Exim Solutions, 2025. *HS Codes – Chapter 28: Inorganic Chemicals; Compounds of Precious Metals*. [Online]. Available at: <https://www.seair.co.in/hscodes/chapter-28-inorganic-chemicals-compoundsprecious-metals.aspx> [Accessed 1 Feb 2025].

- ✓ Shukran, M.A.M., M.S.I.M. & N.A., 2017. *Enhancing Chemical Inventory Management in Laboratory through a Mobile-Based QR Code Tag*. [Online]. Available at: <https://iopscience.iop.org/article/10.1088/1757-899X/226/1/012093> [Accessed 11 Oct 2025].

- ✓ Sun, M. & H.Y., 2023. *Forecasting Model of Fishery Import and Export Trade Data Using Deep Learning Method*. [Online]. Available at: <https://ieeexplore.ieee.org/document/10387376> [Accessed 14 Feb. 2025].

- ✓ Tsuji, Y., K.T.R.H., 2016. *Chemical substances management system at the University of Tokyo*. [Online]. Available at: <https://www.semanticscholar.org/paper/Chemical-substances-managementsystemat-the-of-Tsuji-Tonokura/16a45d74277ce55695f9b96575991333859553ed> [Accessed 14 Feb. 2025].

Appendix A - Questionnaire for Requirement Gathering



Comprehensive Chemical Risk Prediction Model

The Comprehensive Chemical Risk Prediction Model (CCRPM) was inspired by the National Authority for Implementation of the Chemical Weapons Convention in Sri Lanka (NACWC). It monitors and predicts risks associated with dual-use chemicals, which serve legitimate industrial purposes but can be repurposed for illegal activities, such as making chemical weapons or explosives. The system tracks risks from importers to end-users and predicts future importation trends and associated risks, ensuring proactive and robust chemical risk management.

senuth.20230317@iit.ac.lk [Switch account](#)



Not shared

* Indicates required question

How often do you come into contact with cleaning or chemical products (e.g., household cleaners, fertilizers, etc.)? *

- Daily
- Weekly
- Occasionally
- Rarely or never

Are you aware that certain household or industrial chemicals can be harmful or hazardous if misused? *

- Yes, very aware
- Somewhat aware
- Not aware at all

Do you think there is a need to track and regulate the use of chemicals that might *
be dangerous or used in harmful ways?

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree

How concerned are you about the potential risks of chemicals in products you *
use daily?

- Very concerned
- Somewhat concerned
- Not very concerned
- Not concerned at all

Do you feel that enough information about potential chemical risks is provided on *
product labels?

- Yes, it's sufficient
- Somewhat, but it could be clearer
- No, it's not enough

How much do you trust authorities to regulate and monitor the safe use of *
chemicals?

- Fully trust
- Somewhat trust
- Neutral
- Slightly distrust
- Do not trust at all

Were you aware that technology (like AI) is used to predict risks related to *
chemicals (e.g., misuse, accidents)?

- Yes
- No

How comfortable are you with the idea of using AI to monitor and predict *
chemical risks in products and industries?

- Very comfortable
- Somewhat comfortable
- Neutral
- Slightly uncomfortable
- Not comfortable at all

Do you think that a system to predict chemical risks could help improve public safety? *

Strongly agree
 Agree
 Neutral
 Disagree
 Strongly disagree

How often do you check for hazard or safety information on products you buy (e.g., "flammable," "toxic")? *

Always
 Often
 Sometimes
 Rarely
 Never

Would you be interested in receiving alerts or information about potential chemical hazards in your area? *

Yes, I'd like alerts
 Only for severe hazards
 No, I wouldn't be interested

Please share any comments or suggestions you have regarding our Comprehensive Chemical Risk Prediction Model for the national safety of the country

Your answer

Submit Page 1 of 1 [Clear form](#)

Never submit passwords through Google Forms.

This form was created inside of Informatics Institute of Technology.
Does this form look suspicious? [Report](#)

Google Forms

Figure 31 - Survey form for Students