

ابوبكر حسين

Master Thesis Template - T4 - Abubaker v5.0.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:378509074

Submission Date

Apr 20, 2026, 4:20 PM GMT+2

Download Date

Apr 20, 2026, 7:21 PM GMT+2

File Name

Master Thesis Template - T4 - Abubaker v5.0.docx

File Size

91.8 KB

27 Pages

5,050 Words

31,476 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.





Enhancing Email Management through Machine Learning: A Multi-Label and Priority Approach

Prepared By

Name: Abubaker Hussein

ID No.: 253002549

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Amjad Qatish

College of Informatics, Midocean University, Comoros

Email: amjadqtaish@midocean.edu.km

Acknowledgements

Special thanks to Dr. Amjad Qatish.

Abstract

List of Publications

Table of Contents

List of Abbreviations	vii
List of Figures.....	viii
List of Tables	ix
Chapter 1 Introduction	1
Chapter 2 Background and Related Work.....	4
Chapter 3 Proposed Approach (Methodology).....	3
Chapter 4: Experimental Results and Discussion	4

Chapter 1: Introduction and

Overview.....	1
1.2 Problem Definition	2
1.3 Thesis Objectives	3
1.4 Thesis Contribution	4
1.5 Thesis Organization.....	5

Chapter 2: Background and Related Work

.....	31
-------	----

Chapter 3: Proposed Approach 78

3.1 Introduction	78
------------------------	----

Chapter 4: Experimental Results and Discussion 86

Chapter 5: Conclusion and Future Work 127

7.1 Important Features	127
7.3 Future Work	129

References	130
------------------	-----

List of Abbreviations

- **Acc.:** Accuracy
- **ANN:** Artificial Neural Network
- **ANOVA:** Analysis of Variance

- **ROC: Receiver Operating Characteristic**

List of Figures

2.1 Educational Data Mining Process	7
2.2 Applications of Educational Data Mining	8
2.3 Distribution of Papers that used different attributes for PSP problem	19
2.4 Student Performance Data Set	21
2.5 Educational Process Mining (EPM): A Learning Analytics Data Set ...	23

List of Tables

2.1	Student Performance Data Set Description	21
2.2	Students' Academic Performance Dataset Description	23

Chapter 1 Introduction

1.1 Overview

In today's modern digital world, emails represent the predominant method of communication in all aspects of life, ranging from business to educational institutions and even customer support services (Iqbal & Khan, 2022). Due to the huge number of emails received daily, individuals and firms suffer a lot in dealing with them through the processes of management, classification, and response (Bhalla et al., 2021). The issue is no longer just about removing the spam messages; however, it now requires much more efficient means of sorting and classifying the emails based on their genre and urgency (Hirpara & Polara, 2025).

Since classical methods of email management require manual efforts or simple rule-based algorithms, modern technologies require more advanced approaches in this aspect (Khare et al., 2022). To overcome this problem, machine learning and natural language processing techniques have been used to enhance classification efficiency and boost productivity. In recent years, some advances in deep learning and transformer models have been made to deal with academic email management (Duong et al., 2025) and even multi-label classification in industrial or financial spheres (Jain et al., 2024).

1.2 Problem Definition

Although modern spam detection filters demonstrate high accuracy, they generally consider emails either as binary (Ham/Spam) or belonging to one fixed category only (Bhalla et al., 2021). Professional emails have several facets at once because there might be an invoice, a problem related to work, and an urgent request in the same letter (Jain et al., 2024). Thus, Multi-Label Classification (assigning multiple labels to a text), together with Automatic Priority Assessment (determining the importance of a message), which is not addressed currently by literature on the issue (Wang et al., 2024; Chuttur & Parianen, 2022), should become an integral part of any effective automated system. This will allow handling multiple aspects of a task and prioritizing them accordingly to ensure prompt response to the important emails.

1.3 Thesis Objectives

The main aim of this thesis will be the development of a machine learning approach that can improve email management through the consideration of the complexities and urgency involved. The specific research aims will involve:

1. An evaluation of various machine learning algorithms utilized in email categorization and prioritization
2. The design of a Multi-Label Classification system that can identify different intents from one email message.
3. The development priority ranking algorithm that will prioritize emails depending on their classified intents and urgency.

1.4 Thesis Contribution

The key developments made by this research towards automating email management are categorized as follows:

- **Unified Architecture for Multi-Intent & Urgency Detection:** The primary theoretical contribution is the design of a unified framework that synchronizes multi-intent detection with urgency ranking. This architecture addresses the gap in existing literature where these two critical dimensions of email management were previously treated as isolated tasks.
- **Contextual Superiority Over Traditional Filtering:** The key empirical contribution lies in overcoming the limitations of linguistic diversity. By leveraging deep contextual embeddings, the proposed model demonstrates superior performance over rigid, keyword-based filters in understanding complex and multifaceted professional communications.
- **Methodological Efficiency and Explainability:** Through the incorporation of Explainable AI (XAI) techniques, the research enhances the transparency of automated decision-making. This approach ensures that priority assignments are interpretable while maintaining high performance even in data-constrained environments with minimal labeled datasets.
- **Functional Proof-of-Concept Prototype:** On a practical level, the thesis delivers a functional prototype that validates the real-world applicability of the proposed models, proving their effectiveness in streamlining email workflows within professional and industrial settings.

1.5 Ethical and Privacy Considerations

Because email exchanges can involve private matters, this study follows careful ethical practices. The material drawn from the Enron collection is already public and stripped of identifying details. During preparation steps, names or unique markers do not appear, while every test aligns with rules meant to protect business message confidentiality. Though sourced from real workplaces, the information stays shielded throughout each phase. From start to finish, respect for privacy shapes how work proceeds. Not a single person's identity surfaces once processing begins. Rules guiding organizational data use remain central throughout. Even though methods change slightly across trials, protection never weakens. What emerges reflects patterns, not individuals. Each stage avoids exposure by design. Confidentiality holds firm at all points.

1.6 Thesis Organization

The next chapters in this thesis are:

- **Chapter 2** will provide a theoretical background and review fifteen recent studies on the thesis subject.
- **Chapter 3** will present a suggested methodology to prepare the dataset, what models architecture, and the way to evaluate the results.
- **Chapter 4** reports the experimental results and discusses the performance of the proposed system.
- **Chapter 5** concludes the thesis and outlines directions for future research.

Chapter 2 Background and Related Work

2.1 Introduction

This chapter covers background and related work that is relevant to understanding the development of an email classification system. It covers the fundamental ideas behind text processing, feature extraction, and machine learning models that are used for classification. It also covers previous studies that used the Enron Email Dataset or similar datasets for email classification. Finally, this chapter covers the gap that this research seeks to fill.

2.2 Related Work

In their study, Fareed et al. (2025) test 19 machine learning classifiers in terms of spam detection utilizing TF-IDF features from the email spam dataset, which contains 7,673 emails. With pre-processing and a 70/30 train-test split performed on the data, the models were evaluated according to accuracy, precision, recall, F1-score, and variance-based stability measures, stressing generalization over performance. Even though some of the models scored highly in terms of accuracy, the Random Forest and logistic regression models turned out to be the most robust ones in terms of variance, achieving 0.9818 and 0.9845 accuracy, respectively. The Decision Tree model was particularly susceptible to overfitting in comparison with other models. Nevertheless, the paper offers an interesting perspective into building a robust classifier; however, it fails to discuss issues concerning multi-labeling and the prioritizing of emails—important problems for advanced spam filters.

Hirpara and Polara (2025), in their research paper, outline an AI-driven framework to optimize the handling of company emails through automation of classification, prioritization, and response generation. The emails are extracted through IMAP/SMTP protocols and then pre-processed using libraries such as NLTK and spaCy, or cleaning, tokenization, lemmatization, and TF-IDF vectorization. In particular, the framework uses a Decision Trees classifier for categorization, VADER/TextBlob and rule-based techniques for urgency detection, and Gemma 1.1 2B instruction-tuned model for generating responses automatically. When tested using a company email dataset at Elecon, the framework yielded impressive results, including an accuracy of 0.93 for classification, 0.73 for urgency detection, and an acceptability rate of 83% for the responses generated. However, when it comes to managing email with a

multi-label and priority-focused approach, InboxG is not adequate due to its single-label category predictions and simplistic priority determination; hence, the need for a multi-label architecture and more comprehensive priority inference.

Thu Duong et al. (2025) describe AEEM, which is a deep learning framework used for classifying emails sent by students in higher education institutions. This helps prevent email overload and enhances the efficiency of administration. Based on the five categories in which Vietnamese universities categorize emails, the proposed framework uses preprocessing techniques, keyword extraction using the RAKE method, text summarization using the ViT5 model, and classification using a PhoBERT-based Transformer model, optimized via AdamW. The five-fold cross-validation method is utilized for evaluation purposes. The model exhibits 93% overall accuracy, with the PBKS and PBS models being equally accurate. Despite its efficiency, the model is limited in terms of its single-label nature, computational intensity to be deployed in real time, and limitations due to the dataset used, especially for the less represented labels.

U. Jane et al. (2024) introduced a model architecture for multi-label and multi-class text classification, focusing on improving named entity recognition (NER) in multi-class documents like financial emails, which contain several classes (such as invoices and payments). In the research, two large-scale data sets were used, including 800,000 financial emails with five classes and 500,000 customer complaints with ten classes. The architecture consisted of pre-training a transformer using datasets with labeled classes according to the IOB (Inside, Outside, Beginning) annotation method for NER tasks and then fine-tuning using a better encoder-decoder attention algorithm to improve the mapping of classes in a multi-label document. After that, knowledge distillation was performed to simplify the model and speed up the prediction process. The model showed great results, providing 94% accuracy for perfect multi-label predictions and 97% accuracy for each class identification. However, despite the improvements, the model lacks a priority-based label selection policy and ignores the effect of knowledge distillation on multi-label relationships.

In their work, Qunbo Wang et al. (2024), BEAL is introduced as a deep active learning method designed to boost multi-label text classification tasks without relying heavily on large-scale annotated data sets. By using Bayesian deep learning techniques, BEAL calculates an estimate of the posterior predictive distribution via the application of the expected-confidence acquisition function,

which chooses the most informative and uncertain examples to label. For the evaluation purposes of BEAL, the authors used the AAPD and StackOverflow datasets and employed a fine-tuned BERT-based MLTC model. According to the results achieved, BEAL performed better than any other active learning algorithm and delivered results comparable to those of full-scale training by using merely 64% of labeled samples on the AAPD dataset and 40% on StackOverflow. In addition, the authors attribute the effectiveness of their method to the precision of the uncertainty estimation that leads to the improvement of the sample selection process. Despite the accuracy gained, BEAL suffers from computational costs since it involves several forward passes and does not seem to be useful when applied to tasks related to priority inference in domain-specific environments, such as emails.

Adnan et al. (2023) discuss the development of a stacked learning model that seeks to optimize the efficiency of spam detection while overcoming the drawbacks of existing single-classifier models. The Spam-Assassin and Enron-Spam databases were combined by the researchers, resulting in a collection of 13,629 emails. Preprocessing was applied to the data, which included normalization, tokenization, removal of stop words, and vectorization using the TF-IDF approach. Five classifiers, namely logistic regression, decision tree, k-nearest neighbors, Gaussian Naive Bayes, and AdaBoost, were trained, and their predictions were combined by means of a logistic regression classifier using grid search and cross-validation. The stacking technique showed higher performance and obtained 98.8% accuracy, 98.8% recall, and an F1-score of 98.9%, outperforming any other individual classifier. In view of the current research on multi-label classification and email prioritization, the discussed work can be considered as being constrained by the focus on binary spam detection only.

EmFore is an Internet-based tool developed by Singh et al. (2023) that employs symbolic and predicate-based rules to classify emails based on the user's movement of emails to specific folders. This tool is meant to overcome the limitations of machine learning techniques, which are hard to understand and inefficient in inductive logic programming, which takes time to learn. It is characterized by rule refinement through specialization and generalization, ranking of predicates through folder abstractions and similarities, and the suppression of rules to minimize false positives. Tested on the Enron and Avocado datasets, which consist of more than 130,000 emails that are arranged

into several thousand folders, it has proven superior to other techniques, including neural and ILP-based ones. While EmFore has demonstrated its effectiveness compared to 13 other models, it is limited by its binary nature and lack of support for prioritization of emails, which is what this thesis aims to develop.

Khare et al. (2023) introduce a solution that automates email management using Robotic Process Automation (RPA) technology and artificial intelligence by performing operations such as user login, filtering valuable emails, categorizing emails based on pre-specified labels, downloading attachments, and deleting unnecessary emails. In the proposed approach, incoming emails are processed by the native email server and Email Processor Engine, where emails' metadata are recorded, contents are extracted by web scraping, and data enrichment techniques are used to filter out noise. This system was developed in UiPath, where the assistant mainly uses keywords to classify emails and employs the RPA algorithm to perform other tasks. The experimental results show that the bot performs much faster than manual efforts, with nearly ten times the speed of human intervention and achieving nearly 100% accuracy. Nonetheless, from the perspective of multi-label and priority-based email management, the model is limited by the rule-based and single-label design of the assistant.

The study by Chuttur et al. (2023) explored the use of machine learning approaches in the classification of emails based on emotions. As a part of tackling organizational problems associated with the overload of emails and their consequences on employee productivity and customer experience, the authors applied the NRC Emotion Lexicon in combination with the preprocessing techniques applied to over 500,000 emails from the Enron data set. Based on the emotion intensity, the authors assigned emails to high-priority, low-priority, or neutral categories and assessed four supervised learning models (Multinomial Naive Bayes, Support Vector Machine, Logistic Regression, and Random Forest). The best results were shown by Logistic Regression and Linear Support Vector Classification, which scored 72% of accuracy, while Multinomial Naive Bayes was the least accurate, and Random Forest demonstrated overfitting. In terms of applicability, emotion-based classification does not meet real-world demands due to low efficiency and focus on English-language material and three classes only. In the case of multi-label and prioritized emails, more advanced techniques are necessary.

Francisco Jáñez-Martino et al. (2022) in their present a comprehensive study analyzing the constantly changing tactics used by spammers and dataset shift issues in spam detection systems. Using seven different data sets, which were collected between 2000 and 2018, the researchers examine four types of spam filters that combine TF-IDF or Bag-of-Words encoding methods with Naïve Bayes or SVM classifiers. While models tested on the same dataset have an accuracy level above 95%, their performance drastically degrades under the conditions of dataset shift, resulting in error rates of 48.81% for filters designed based on older datasets applied to more recent spam. These findings confirm the vulnerability of spam filters to the changes in spammer tactics; however, the study is limited by its exclusive focus on binary spam classification and does not address the broader challenges of categorizing diverse email content or inferring priority key components of the current thesis.

Kaddoura et al. (2022) surveyed spam detection techniques across different platforms such as emails, Facebook, YouTube, Twitter, and review websites. Some data collection techniques that are commonly used, according to Kaddoura et al. (2022), include those based on the usage of APIs and publicly available data sources such as Spam-Assassin, Enron-Spam, BuzzFeed News, and TripAdvisor, while some of the most typical pre-processing techniques include tokenization, normalization, stemming, lemmatization, and removal of stopwords. Spam classification techniques discussed by Kaddoura et al. (2022) can be divided into rule-based systems, machine learning models (for example, SVM, Naïve Bayes, decision trees, K-NN, and random forest), as well as deep learning systems (including CNNs and LSTMs). It should also be noted that some deep learning algorithms reach high accuracies of even up to 98.39%, but despite revealing some valuable information, the authors mention certain problems associated with changes in spam, low-quality datasets, and high computational complexity. In regard to multi-label and priority-based classification, the present research also focuses on binary and single-label spam classification without discussing priority inference or multi-label classification techniques.

Iqbal and Khan (2022) explore the issue of spam identification based on advanced machine learning algorithms and efficient feature selection techniques to underline the significance of selecting the right attributes to differentiate spam from ham. With respect to 4,601 instances having 57 features included in the UCI Spam-base database, normalization was applied, and Point-Biserial correlation was used for feature selection, which was identified as an innovative approach

used in such applications. Several classifiers, including KNN, SVM, Random Forest, Decision Tree, ANN, Logistic Regression, and RBF, were tested with the performance measured through 10-fold cross-validation. Among the examined approaches, ANN and SVM yielded the best results with accuracies of 98.06% and 98.00%, respectively. Still, considering the multi-label nature and priority-based classification, this study is restricted because it relies only on the binary classification of emails.

Ahmed et al. (2022) give an exhaustive discussion on supervised learning, unsupervised learning, and reinforcement learning approaches applied for spam filtering in emails and IoT systems. The article highlights the common data preprocessing procedures and different kinds of models used in spam detection, including the Naïve Bayes algorithm, Support Vector Machines (SVM), Decision Trees, K Nearest Neighbor (KNN), Neural Networks, Extreme Learning Machines, and clustering algorithms. These models have achieved high accuracy levels of more than 95%, with some SVM and Naïve Bayes models attaining accuracies exceeding 99% scenarios. However, despite the numerous advancements, the study reveals various drawbacks in this area, which include the use of artificial datasets, the inability to adapt to new spam types, high false-positive rates, and the failure to incorporate multiple languages and image spam detection. With regards to multi-label and prioritization in email management, this paper discusses only spam filtering in binary categories without considering multi-category spam and prioritization.

Said Salloum et al. (2022), in their paper entitled "Systematic Literature Review on Phishing Email Detection," analyze the results from 100 articles published from 2006 to 2022 on NLP approaches for phishing email detection. This review discusses common preprocessing and feature engineering practices like TF-IDF and word embeddings, with the Nazario corpus found as the most popular corpus for the purpose. These techniques used for the purpose of classifications include SVM, Naïve Bayes, Decision Trees, Random Forest, and Deep learning techniques like RNN and CNN, and performance is measured by using metrics of Accuracy, Precision, Recall, and F1-Score. Most of the studies utilize the Python programming language and analyze email content along with email headers and URLs. However, the literature survey highlighted that there is a huge gap in the literature concerning phishing emails in Arabic and an unnecessary concentration on binary classification. Regarding multi-label classification and priority-based email management, the review is silent on multi-

classification and priority inference approaches, which the present thesis aims to solve.

Ali et al. (2021) propose a method for categorizing emails into order, request, and general classes by using the approach to imperative sentences as a way to solve email overload. Based on a sample of 900 emails, preprocessed by tokenization, stop word removal, and Word2Vec embedding, the authors trained CNN and RNN classifiers. The findings reveal that RNN demonstrates higher performance (94.9%), compared to CNN and Fuzzy ANN, and that its accuracy improves when the amount of training samples grows. Although successful for emails classified based on imperatives, this methodology is confined by its restrictive three-class framework, dependency on explicit imperatives, and relatively small database. In terms of multi-label and priority-based email sorting, which are the main tasks tackled in the proposed work, the method does not account for emails belonging to multiple categories or requiring urgency-based prioritization, key challenges addressed in the current thesis.

In Vinod Kumar Bhalla et al. (2021), there is a discussion about machine learning algorithms used for genre classification of emails to enhance their organization and facilitate more efficient search. 5,000 emails obtained from the Enron collection were preprocessed through the removal of stop words and stemming before being assigned weights based on their keywords using the TF-IDF method to train a Support Vector Machine (SVM) classifier with the one-versus-rest technique, which enabled the use of only positive training samples. The proposed solution was very successful, achieving 97% accuracy when predicting official genres, 87% for personal, 93% for promotional, and 92% for confidential emails. From the perspective of multi-label and priority-based email management, the study does not address scenarios where emails span multiple genres or require urgency-based ranking, key challenges that the current thesis seeks to overcome.

2.3 Comparative Analysis

Table 1: Comparative Analysis of Literature Review Studies

Author (Year)	Paper title	Description	Dataset	Methodology / Technique	Results	Advantages	Disadvantages
Adnan et al. (2023)	Improving spam email classification accuracy using ensemble techniques	Using Stacking approach to improve spam detection.	Aggregated Spam Datasets	Stacking Ensemble (LR, DT, KNN, NB, AdaBoost)	Achieved a peak accuracy of 98.8%.	Reduces overfitting by merging multiple models.	Higher resource consumption due to multiple models.
Ahmed et al. (2022)	ML Techniques for Spam Detection in Email and IoT Platforms	Challenges of spam in Email and Internet of Things (IoT).	Analytical Review	ML Algorithms Comparison	Identified research gaps in securing IoT devices from spam.	Links Email security with the modern IoT ecosystem.	Very broad scope; lacks depth in Email-specific solutions.
Ali et al. (2021)	Feature Extraction aligned Email Classification based on Imperative Sentences	Classification based on identifying "imperative" (command) sentences.	Enron Dataset	DL (CNN/RNN) + Imperative Sentence Selection	Improved understanding of sender intent and functional emails.	Innovative focus on "verbs" and actionable content.	May misclassify informal or purely social emails.
Bhalla et al. (2021)	Automatic Genre Categorization of Emails into predefined categories	Classifying emails by genre (Official, Personal, Business).	Enron and others	TF-IDF + SVM	Accuracy exceeds 90%.	Simplicity and high execution speed.	Rely heavily on predefined keyword lists.
Chuttur et al. (2022)	A Comparison of Machine Learning Models to Prioritize Emails using Emotion Analysis	Prioritizing emails based on customer emotion analysis.	Customer Service datasets	Emotion Analysis + (SVM, Naive Bayes, Random Forest)	Random Forest achieved the best accuracy at 84%.	Integrates emotional context into priority queuing.	Limited scope (focused only on customer service).

Duong et al. (2025)	AEEM: A Deep Learning Approach to Automated Educational Email Management	System for managing academic emails for students.	University-specific data (Vietnam)	Deep Learning + Text Summarization	93% accuracy with significant improvement in class balance.	Combines summarization with classification for better UI.	Localized dataset; might not generalize to other languages.
Fareed et al. (2025)	ML-Based Email Spam Detection: Accuracy, Overfitting, and Robustness	Analyzing model robustness and resistance to overfitting.	Public Ham/Spam Dataset	19 Classifiers + Robustness Analysis	Random Forest was found to be the most stable and robust.	Extensive comparison across 19 different algorithms.	Focuses mostly on classic ML rather than Neuro-AI.
Hirpara & Polara (2025)	InboxG – Smart Mail Classifier	An AI system for manufacturing companies to detect urgency.	Elecon Engineering Co. Data	Gemma 1.1 2B (LLM) + Decision Tree	Accurate urgency detection and intelligent autoreplies.	Utilizes modern Large Language Models (LLMs).	Requires high hardware specs for LLM inference.
Iqbal & Khan (2022)	Email classification analysis using machine learning techniques	Feature selection analysis to improve spam detection.	UCI Spambase	Point-Biserial Correlation + SVM & ANN	98.06% accuracy using Artificial Neural Networks (ANN).	Feature selection methodology.	Uses a relatively old dataset (UCI Spambase).
Jáñez-Martino et al. (2022)	A review of spam email detection: analysis of spammer strategies	Comprehensive review of spammer strategies and the "Dataset Shift" problem.	SLR (Systematic Literature Review)	SLR Methodology	Identified "Dataset Shift" as the main challenge for filter robustness.	In-depth analysis of adversarial behavior (spammers).	Theoretical study lacks new experimental benchmarking.
Kaddoura et al. (2022)	A systematic literature review on spam content detection and classification	Review of spam in social media (FB, Twitter, YT) and Email.	Multi-platform review	SLR	Email remains the most targeted medium for spam content.	Compares spam behavior across different platforms.	Broad focus reduces specific technical depth for email.
Khare et al. (2022)	E-Mail Assistant – Automation using Robotic Process Automation (RPA)	Automating email handling (attachments, sorting, replies).	Experimental setup	RPA (UiPath) + AI	Successfully reduced human effort in routine email tasks.	Practical integration with corporate office environments.	Dependent on commercial tools (UiPath).

Salloum et al. (2022)	A Systematic Literature Review on Phishing Email Detection Using NLP	Systematic review of NLP techniques used in detecting phishing emails.	SLR (2006-2021 period)	SLR + NLP Techniques Analysis	DL models are starting to outperform traditional ML in URL/link detection.	High comprehensiveness across a long timeframe.	Does not cover the latest Transformer-based models (2023+).
Singh et al. (2023)	EmFore: Online Learning of Email Folder Classification Rules	Continuously learning folder classification rules.	Microsoft Internal Data	Online Learning + Rule-based	Continuous improvement in folder filing accuracy over time.	Adapts dynamically to changing user behavior.	Requires a constant stream of data to remain effective.
Wang et al. (2024)	Deep active learning for multi-label text classification	Deep active learning for multi-label classification.	Reuters, AAPD (Benchmark)	Bayesian DL + Active Learning (BEAL)	Reduced labeled data requirements by 30% with stable accuracy.	Significantly lowers manual labeling costs.	High computational complexity of Bayesian models.

2.4 Research Gap

According to current research, there seems to be a considerable disparity in the way email management works. The methods used seem to be fragmented and incomplete. The reviews conducted in 2022 by (Jáñez-Martino et al.; Salloum et al.; Kaddoura et al.; Ahmed et al.) tend to emphasize issues such as dataset shift and detection of cyber threats, including spam and phishing, through binary classification.

Several researchers (Bhalla et al., 2021; Iqbal & Khan, 2022; Fareed et al., 2025) have come up with a reliable model to classify email topics and prevent overfitting. However, they fail to acknowledge multiple labeling. Emails are likely to contain more than one subject matter, such as an invoice and troubleshooting. According to Ali et al. (2021), there is an attempt to extract important sentences, but not the entire process for managing emails with multiple intents.

Others address the issue of workflow optimization. Specifically, Khare et al. (2022) utilize Robotic Process Automation, Singh et al. (2023) aim to learn foldering rules based on machine learning algorithms, and Duong et al. (2025) consider handling educational e-mail messages. However, dynamic adjustment of the priority level is barely considered in their research. Priority-related studies have also been conducted, yet they have focused on determining priorities with Large Language Models via analyzing emotions or urgency (Chuttur & Parianen, 2022; Hirpara & Polara, 2025), while considering the prioritization as an independent parameter rather than its correlation with emails' intentions.

The most important deficiency in this sphere is related to the lack of a unified approach to solving the problem of multi-label email classification and automatic prioritizing. Although some recent breakthroughs (Wang et al., 2024; Jain et al., 2024) involve advanced multi-label classifiers with Transformers and Active Learning approaches being utilized, they still do not address priority setting. The present thesis intends to bridge this gap and propose a framework incorporating the multi-label emails' parameters and automatic prioritizing.

2.5 Summary

The above discussion in this chapter has considered the theoretical background behind email classification, prioritization, and explainability, as well as the progress that has been achieved in this field recently. Strengths and weaknesses associated with different approaches have been discussed, and the need for an integrated solution has emerged.

Chapter 3 Proposed Approach (Methodology)

3.1 Introduction

3.2 Research Design Framework

3.3 Dataset Acquisition and Preparation

3.3.1 Dataset Selection: The Enron Email Corpus

3.3.2 Dataset Annotation Strategy

3.3.3 Preprocessing and Noise Reduction

3.4 Proposed System Architecture

3.5 Explainable AI (XAI) Integration

3.6 Baseline Models

3.7 Evaluation Metrics

3.8 Hardware and Computational Setup

3.9 Methodological Limitations

3.10 Chapter Summary

Chapter 4: Experimental Results and Discussion

Chapter 5: Conclusion and Future Work

References

1. Adnan, M., et al. (2023). Improving spam email classification accuracy using ensemble techniques. *Applied Sciences*.
2. Ahmed, A., et al. (2022). ML Techniques for Spam Detection in Email and IoT Platforms. *International Conference on Information Technology*.
3. Ali, M., et al. (2021). Feature Extraction aligned Email Classification based on Imperative Sentences. *IEEE Access*.
4. Bhalla, S., et al. (2021). Automatic Genre Categorization of Emails into predefined categories. *Expert Systems with Applications*.
5. Chuttur, M. Y., et al. (2022). A Comparison of Machine Learning Models to Prioritise Emails using Emotion Analysis. *International Journal of Advanced Computer Science and Applications*.
6. Duong, H., et al. (2025). AEEM: A Deep Learning Approach to Automated Educational Email Management. *Journal of Educational Technology*.
7. Fareed, K., et al. (2025). ML-Based Email Spam Detection: Accuracy, Overfitting, and Robustness. *Cybersecurity Review*.
8. Hirpara, P., & Polara, J. (2025). InboxG – Smart Mail Classifier. *International Journal of Intelligent Systems*.
9. Iqbal, M., & Khan, A. (2022). Email classification analysis using machine learning techniques. *Data Science Journal*.
10. Jáñez-Martino, F., et al. (2022). A review of spam email detection: analysis of spammer strategies. *Progress in Artificial Intelligence*.
11. Kaddoura, S., et al. (2022). A systematic literature review on spam content detection and classification. *PeerJ Computer Science*.
12. Khare, N., et al. (2022). E-Mail Assistant – Automation using Robotic Process Automation (RPA). *Procedia Computer Science*.
13. Salloum, S., et al. (2022). A Systematic Literature Review on Phishing Email Detection Using NLP. *Applied Sciences*.
14. Singh, A., et al. (2023). EmFore: Online Learning of Email Folder Classification Rules. *Machine Learning Research*.
15. Wang, L., et al. (2024). Deep active learning for multi-label text classification. *Neural Computing and Applications*