

ACKNOWLEDGEMENT

I am greatly indebted to our project guide **Dr. A. M. Rajurkar** for her able guidance throughout this work. It has been an altogether different experience to work with her and I would like to thank her for help, suggestions and numerous discussions

I gladly take this opportunity to thank **Dr. A. M. Rajurkar (Head of Computer Science & Engineering, MGM's College of Engineering, Nanded)**. I am heartily thankful to **Dr. Lathkar G. S. (Director, MGM's College of Engineering, Nanded)** for providing facility during the progress of the project and for her kind help, guidance, and inspiration.

Last but not least I am also thankful to all those who help directly or indirectly to develop this project and complete it successfully.

Ambarish Udgirkar
[B.Tech CSE-A]

ABSTRACT

In today's digital communication landscape, email remains a dominant medium, yet users are often overwhelmed by the sheer volume and variety of messages they receive. This project presents a deep learning-based email classification system designed to automatically categorize emails into four major classes: Spam, Promotions, Social, and Updates. The proposed solution is built upon a hybrid CNN-BiLSTM architecture that combines the local feature detection strength of Convolutional Neural Networks (CNN) with the contextual memory and sequence processing capabilities of Bidirectional Long Short-Term Memory (BiLSTM) networks. This design enables the model to capture both keyword patterns and deeper semantic relationships across email content.

Real-world data was extracted directly from Gmail using the Gmail API, and further enhanced using Optical Character Recognition (OCR) to extract text from image-based emails. A comprehensive preprocessing pipeline was implemented to clean, normalize, and tokenize the email data, ensuring linguistic consistency and reducing input noise. With rigorous training, regularization, and hyperparameter tuning, the model achieved a test accuracy of 95.3%, outperforming earlier versions based on LSTM and CNN-LSTM architectures. The system was deployed via a Flask-based API, supporting real-time classification and seamless integration with other applications. Overall, the project demonstrates a reliable, scalable, and production-ready approach for intelligent email sorting, addressing both academic challenges and real-world user needs.comprehension.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	I
ABSTRACT	II
TABLE OF CONTENTS	III
LIST OF FIGURES	IV
Chapter 1. INTRODUCTION	1
1.1 Overview of CNN and BiLSTM	5
1.2 Working Mechanism	6
1.3 Application	9
1.4 Challenges	12
Chapter 2. LITERATURE REVIEW	14
2.1 Introduction	14
2.2 Existing Approach	17
2.3 Limitations of Existing Approaches	21
2.4 Statement of the Problem	25
2.5 Conclusion	26
Chapter 3. METHODOLOGY	30
3.1 Introduction	30
3.2 Deep Learning Framework	32
3.3 Proposed CNN-BiLSTM Model	36
3.4 Limitations and Assumptions	40
Chapter 4. IMPLEMENTATION DETAILS	45
4.1 Dataset	45
4.2 Proposed Hybrid Deep Learning Model	48
4.3 Proposed Approach	51
4.4 Fine Tuning	55
Chapter 5. EXPERIMENTAL RESULTS	58
5.1 Generated Results and Accuracy	59
5.2 Complexity	63
CONCLUSION	66
REFERENCES	67

LIST OF FIGURES

Figure No.	Name of Figure	Page No.
3.1	Layered Deep Learning Architecture	32
3.2	CNN Architecture	33
3.3	BiLSTM Architecture	34
4.1	Preprocessed Dataset (Promotions Class)	45
4.2	Project Timeline	52
5.1	Result UI	57
5.2	Actual Class of Resultant Email	58
5.3	Classification Report	59
5.4	Confusion Matrix	60