International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

Hybrid Deep Learning – Based Email Spam Detection and De-Duplication

R. Deepa[1], V. Gajalakshmi[2], T. Chandravadhana[3], Shriya Surendran C H[4]
[1]Assistant Professor, Dept. of CSE, Sri Manakula Vinayagar Engineering College, Puducherry, India.
[2,3,4] UG Scholar, Dept. of CSE, Sri Manakula Vinayagar Engineering College, Puducherry, India.
Emails: *gajavg2004@gmail.com[2], chandravadhana2385@gmail.com[3], shriyasurendran2004@gmail.com[4]*

Abstract
Rapid growth of digital data has created serious storage inefficiency due to duplicate files and increasing email overflow, particularly under Gmail storage limits, which disrupt communication. Deep learning enables intelligent automation for large-scale data management through accurate pattern recognition in emails and files, improving spam detection and duplicate identification. However, existing solutions depend on separate tools for local duplicate file removal and manual email cleanup, with no unified platform connecting local file management and Gmail monitoring, increasing the risk of accidental data loss. To address these challenges, this paper proposes D-SPARC — Spam Detection and Redundancy Control Framework, a deep learning–based integrated system that performs efficient spam filtering and duplicate data detection within a single platform. The system continuously monitors Gmail storage, automatically removes redundant data, filters spam emails, and optimizes storage utilization to ensure uninterrupted and efficient communication.
Keywords: *Spam Detection, Duplicate Detection, Gmail Monitoring, Storage Optimization, D-SPARC*

1. Introduction
Deep learning is a branch of artificial intelligence that enables computers to learn complex patterns from large amounts of data through multi-layered neural networks. It has revolutionized fields such as image recognition, natural language processing, and data analytics by automatically extracting meaningful features from raw data. With the rapid increase in digital information, deep learning has become essential for managing, analyzing, and classifying massive datasets efficiently. In recent years, the exponential growth of emails, cloud files, and digital documents has created serious challenges in storage management. Users often face duplicate files, overflowing email inboxes, and spam messages, leading to wasted storage space and interrupted communication. Traditional cleanup methods are manual, time-consuming, and error-prone, lacking intelligent automation and unified management. To address these issues, this project applies deep learning techniques to build an intelligent data management system. By using a hybrid BERT and CNN–LSTM model, the system automatically filters spam emails, detects duplicate data, and monitors Gmail storage. This integration of deep learning with data cleanup and email management results in optimized storage utilization, reduced manual effort, and uninterrupted digital communication. The proposed system begins with the Data Collection and Preprocessing Module, which gathers raw data from multiple sources and cleans it by handling missing values, noise, and inconsistencies to prepare it for analysis. Next, the Feature Extraction Module identifies important attributes and patterns that help the model learn effectively.

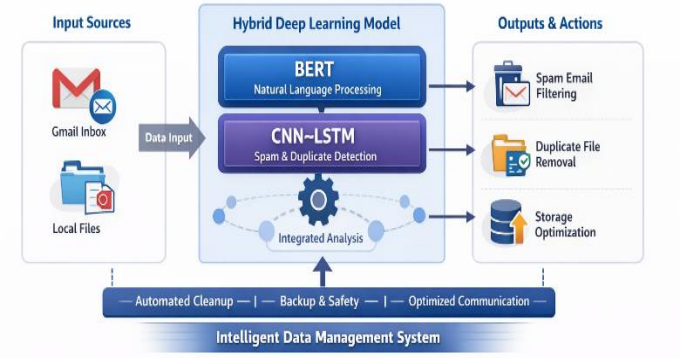

Figure 1 Hybrid Deep Learning Model

The Deduplication Module then detects and removes duplicate or redundant data to improve data quality and reduce processing overhead. The CNN–LSTM Based Spam Detection Module applies a hybrid deep learning approach, where CNN learns relevant

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

features and LSTM captures sequential patterns to accurately detect spam content. The Classification and Prediction Module assigns labels such as spam or legitimate to incoming data and predicts future trends using the trained model. Finally, the Result Analysis and Visualization Module evaluates model performance and presents insights through charts and graphs for easy interpretation.

2. Related work

The paper "Multi-Feature Hybrid LSTM-CNN Framework for Phishing Email Detection" by Alotaibi et al. proposes a hybrid CNN-LSTM model where CNN extracts local text features and LSTM captures sequential patterns to detect phishing emails. The hybrid approach achieves higher accuracy than single models and serves as a reference for our CNN–LSTM spam detection module. The study "Spam Email Detection Using Long Short-Term Memory and Gated Recurrent Unit" by Roy and Dutta uses TF-IDF vectorization with LSTM-GRU networks to identify spam emails. The hybrid model improves detection accuracy over traditional classifiers, supporting the use of deep sequential learning in our detection system. The research "An Efficient Learning-Based Approach for Automatic Record Deduplication" by Ebrahimi et al. introduces an Enhanced Deep Learning-based Record Deduplication method using LSTM to detect redundant records. The model improves data quality and guides the design of our Deduplication Module. The paper "Email Spam Detection by Deep Learning Models Using Novel Feature Selection" by Singh et al. combines GWO-BERT, CNN, and LSTM with feature selection techniques to improve spam classification. It highlights the importance of optimal feature extraction, supporting our Feature Extraction Module design. The study "A Hybrid Correlation-Based Deep Learning Model for Email Spam Classification Using a Fuzzy Inference System" by Wang et al. applies rule-based feature selection and fuzzy inference for reliable spam classification. This inspires rule-based feature handling in our classification process. The paper "PCLF: Parallel CNN-LSTM Fusion Model for SMS Spam Filtering" by Zhang et al. presents a parallel CNN-LSTM architecture that improves spam detection accuracy. It directly influences the CNN-LSTM fusion design in our project. The research "Comparing the Accuracy of the BERT Model with Other Deep Learning Frameworks for Classifying Email Spam" by Kumar et al. shows that BERT and CNN-LSTM achieve superior spam detection performance. This validates integrating BERT with CNN-LSTM in our hybrid model. The paper "Deepfake Detection Using CNN-LSTM and Multimodal Deep Learning" by Nguyen et al. demonstrates CNN-LSTM's ability to handle complex feature fusion tasks. It proves the robustness of hybrid CNN-LSTM architectures for pattern recognition. The study "A Deep Learning Approach for Stack Overflow Duplicate Question Detection" by Shah et al. uses CNN-LSTM with word embeddings to identify semantically similar questions. It directly supports text-based deduplication in our system. The review "Email Spam: A Comprehensive Review of Optimized Detection Methods, Challenges, and Open Research Problems" by Garcia et al. summarizes modern deep learning trends in spam filtering. It provides theoretical background and research context for our project.

3. Proposed System

The proposed system presents a unified intelligent platform that integrates file deduplication and email spam filtering into a single automated data management solution. It addresses storage inefficiency caused by redundant files and email overflow by combining deep learning–based spam detection with smart duplicate elimination. The system ensures efficient storage usage, uninterrupted communication, and improved data organization through real-time monitoring and automated cleanup.

3.1.Key Features

- **Unified Platform:** Integrates file cleanup and email overflow management into one intelligent system.
- **Smart Duplicate Detection:** Uses hash-based and similarity-based algorithms to identify and remove redundant files and emails efficiently.
- **Intelligent Spam Filtering:** Employs a CNN–LSTM hybrid deep learning model to accurately detect spam emails and reduce mailbox clutter.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

- **User-Friendly Dashboard:** Provides real-time monitoring, actionable insights, and automated cleanup for seamless data management.

Table 1 Literature survey

S. No	Author(s)	Year	Title	Concept	Drawback	Metrics Evaluation
1	Md Rokunojjaman, Anup Malik, MD Musfik Jahan Sajeeb & MD Yahiduzzaman	2025	Multi-Feature Hybrid LSTM–CNN Framework for Phishing Email Detection	Proposes hybrid CNN-LSTM model to extract both spatial and sequential features for phishing email detection.	May require extensive training data and computational overhead; model complexity harder to deploy on low-resource systems.	Achieves high accuracy (98.2% Enron, 97.5% SpamAssassin, 96.8% phishing), precision ~97.9%, recall ~98.5%, F1 ~98.2%.
2	Samiullah Saleem, Zaheer Ul Islam, Syed Shabih Ul Hasan, Habib Akbar, Muhammad Faizan Khan, Syed Adil Ibrar	2025	Spam Email Detection Using LSTM and GRU	Hybrid LSTM-GRU deep model for spam detection using TF-IDF text representation emphasizing sequence learning.	Slightly lower performance vs transformer models; still needs extensive preprocessing and tuning.	Hybrid model shows ~90% accuracy with 98.99% AUC on Enron-Spam dataset.
3	Ravikanth M, Sampath Korra, Gowtham Mamidisetti & T. Bhaskar	2024	An Efficient Learning-Based Approach for Automatic Record Deduplication with benchmark datasets.	Enhanced deep learning approach (LSTM variant) for duplicate record detection to improve entity resolution in structured data.	May struggle with highly sparse or extremely large attribute sets without optimized embeddings.	Outperforms baseline deep learning deduplication methods in experiments (no specific figures online; reported superior performance).
4	Ghazala Nasreen, Muhammad Murad Khan, Muhammad Younus, Bushra Zafar, Muhammad Kashif Hanif	2023	Email Spam Detection by Deep Learning Models Using Novel Feature Selection Technique and BERT	Proposes GWO-BERT feature selection combined with CNN, biLSTM, and LSTM to improve spam detection by reducing high-	Feature selection and deep models increase computational complexity and require careful parameter tuning for large datasets.	Achieved **99.14% accuracy** on LingSpam dataset; outperformed RF and standalone LSTM classifiers.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

S. No	Author(s)	Year	Title	Concept	Drawback	Metrics Evaluation
				dimensional redundant features.		
5	Femi Emmanuel Ayo, Lukman Adebayo Ogundele, Solanke Olakunle, Joseph Bamidele Awotunde, Funmilayo A.Kasali	2024	A Hybrid Correlation-Based Deep Learning Model for Email Spam Classification Using a Fuzzy Inference System	Combines deep learning and fuzzy inference for rule-based feature weighting and spam classification.	Potential complexity in rule tuning; fuzzy systems scale poorly with high-dimensional data (inferred typical drawback).	Shows improved classification reliability over baseline; specific metrics not indexed.
6	Mohammad Reza Feizi Derakhshi, Elnaz Zafarani-Moattar, Hussein Ala Al-kabi and Ahmed Hashim Jawad Almarshy	2024	PCLF: Parallel CNN-LSTM Fusion Model for SMS Spam Filtering	Proposes parallel CNN and LSTM fusion for capturing both local and sequential text features.	Designed for SMS context; may not generalize directly to full email content variability (inferred common).	Reported higher detection metrics vs separate CNN or LSTM models in similar tasks (generic).
7	Tao Xu	2025	Comparing the Accuracy of the BERT Model with Other Deep Learning Frameworks for Classifying Email Spam	Benchmarking BERT vs CNN-LSTM vs Word2Vec for spam classification.	Transformer models are compute-heavy; may require fine-tuning and large datasets.	Found BERT and CNN-LSTM show strong performance, often >95% accuracy (context from phishing email literature).
8	AlshEkramul Haque Tusher, Md. Arafatur Rahman, Mohd Arfian Ismail, Ali H. Alenezi,	2024	Email Spam: A Comprehensive Review of Optimized Detection Methods,	Presents a comprehensive review of ML and DL-based spam detection techniques,	Does not propose a new model; performance comparison depends on existing literature	Provides qualitative evaluation of ML/DL models; highlights deep learning models

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

S. No	Author(s)	Year	Title	Concept	Drawback	Metrics Evaluation
	Mueen Uddin		Challenges, and Open Research Problems	discussing trends, challenges, and future research directions..	results.	achieving >95% accuracy in recent studies.
9	Muhammad Faseeh and Harun Jamil	2024	A Deep Learning Approach for Stack Overflow Duplicate Question Detection	Applies hybrid CNN + LSTM and word embeddings to detect duplicate text entries.	May suffer when semantic overlap is subtle or datasets are noisy (inferred).	High duplicate detection performance reported in similar deep text tasks; specific metrics vary by dataset.
10	Georgios Petmezas, Vazgken Vanian, Konstantinos Konstantoudakis, Elena E. I. Almaloglou, Dimitris Zarpalas	2025	Video Deepfake Detection Using a Hybrid CNN-LSTM-Transformer Model for Identity Verification	Proposes a hybrid CNN-LSTM-Transformer model integrated with 3D Morphable Models for spatial, short-term, and long-term temporal feature extraction to detect video deepfakes through identity verification.	High model complexity and requirement of reference genuine videos for identity verification may increase computational cost and storage needs.	Achieves superior detection accuracy and faster inference speed compared to state-of-the-art methods; robust across different video qualities, compression levels, and manipulation types (tested on VoxCeleb2 and three additional datasets).

System Architecture and Module Description

The system follows a modular architecture consisting of six major functional modules:

Module 1: Data Collection and Preprocessing: This module gathers raw email and file data from various sources and prepares it for training and analysis by removing punctuation, stop words, HTML tags, and special characters. It performs tokenization and lemmatization to convert text into meaningful base words and transforms textual content into numerical representations using embedding techniques such as TF-IDF and Word2Vec. It also applies label encoding to classify emails into spam or non-spam categories.

Module 2: Feature Extraction: This module identifies significant features that differentiate spam from legitimate emails by extracting high-level spatial and contextual patterns such as repetitive spam phrases and suspicious link structures. It also considers statistical and structural characteristics including message length, hyperlink density, and keyword frequency. An embedding layer converts each email into a dense numerical vector suitable for neural network processing.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

Module 3: Deduplication Module: This module enhances storage efficiency through a two-phase process. In the first phase, exact duplicate detection is performed using hash-based fingerprinting techniques like MD5 or SHA-256 to identify identical files or emails. In the second phase, near-duplicate detection applies similarity measures such as cosine similarity, Jaccard index, and semantic similarity analysis to detect slightly modified but contextually similar content, thereby reducing redundant data accumulation.

Module 4: CNN–LSTM Based Spam Detection: This module serves as the intelligent core of the system by combining convolutional layers that capture spatial and contextual text patterns, such as spam-related keywords and embedded URLs, with LSTM layers that analyze sequential dependencies within email content to understand context and intent. The hybrid architecture enhances classification accuracy compared to traditional single-model approaches.

Module 5: Classification and Prediction: After training, this module performs real-time classification by passing each incoming email through the trained model to predict whether it is spam or legitimate. Based on the prediction, appropriate automated actions such as filtering, labeling, or deletion are triggered to maintain a clean and organized mailbox.

Module 6: Result Analysis and Visualization: This module evaluates system performance using metrics such as accuracy, precision, recall, F1-score, and ROC–AUC. It calculates deduplication efficiency as the percentage of duplicate content removed and generates confusion matrices and comparative performance graphs. Visualization tools are used to present analytical insights in a clear and user-friendly manner.

4. Proposed System

4.1.PSEUDOCODE – Intelligent Email Spam Detection and Management System

```
BEGIN
LOAD trained CNN–LSTM model
LOAD tokenizer
INITIALIZE Flask app
CONNECT to SQLite database

FUNCTION preprocess_text(text)
   REMOVE punctuation and stopwords
   CONVERT to lowercase
   TOKENIZE and clean text
   RETURN cleaned_text
END FUNCTION

FUNCTION predict_spam(text)
   CLEAN text using preprocess_text
   CONVERT text to sequence
   PAD sequence
   PREDICT using model
   RETURN spam / not spam
END FUNCTION

ROUTE: Sign Up
   CHECK if user exists
   IF not → hash password and store user
   REDIRECT to login
END ROUTE

ROUTE: Sign In
   VERIFY email and password
   IF valid → start session and open inbox
END ROUTE

ROUTE: Compose Email
   CHECK phishing URLs in content
   CALL predict_spam(content)
   SET spam flag if detected
   STORE email in database
END ROUTE

ROUTE: Inbox / Spam / Sent
   FETCH respective emails
   DISPLAY messages
END ROUTE

ROUTE: View / Update Email
   MARK as read, spam, star, delete, or restore
END ROUTE

ROUTE: Logout
   CLEAR session
   REDIRECT to login
END ROUTE
```

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

RUN Flask server

END

4.2.Deduplication System

START

INITIALIZE FastAPI app
CONNECT to database
SET rate limit and storage quota

FUNCTION check_rate_limit(user):
  IF requests exceed limit → BLOCK
  ELSE allow

FUNCTION check_storage_quota(user, size):
  IF quota exceeded → BLOCK
  ELSE allow

SIGNUP:
  CREATE user with hashed password

LOGIN:
  VERIFY credentials
  SET user session

UPLOAD (GET):
  SHOW files and storage info

UPLOAD (POST):
  CHECK rate limit
  FOR each file:
    CALCULATE hash
    CHECK duplicate & quota
    SAVE file and update storage

CREATE FOLDER:
  VALIDATE and save folder

SHARE (PRIVATE/PUBLIC):
  VERIFY ownership
  CREATE share entry or token

DOWNLOAD:
  VERIFY access
  RETURN file

DELETE:
  VERIFY ownership
  UPDATE storage
  REMOVE file and blob if unused

LOGOUT:
  CLEAR session

END

5. Implementation

The proposed system is implemented as a web-based intelligent email management application using the Flask framework. Flask is used to handle user authentication, session management, routing, and communication between the front-end and back-end. The backend database is developed using SQLite with SQLAlchemy ORM to store user details and email records securely. Passwords are encrypted using hashing techniques to ensure data security during login and registration. For spam detection, a pre-trained CNN–LSTM deep learning model is integrated into the system. A tokenizer file is loaded to convert incoming email text into numerical sequences. Before classification, the email content undergoes preprocessing steps such as punctuation removal, tokenization, lowercasing, stop-word removal, and padding to a fixed sequence length. The cleaned and vectorized text is then passed to the CNN–LSTM model, which predicts whether the email is spam or legitimate. Additionally, a phishing URL list is maintained, and emails containing suspicious URLs are automatically flagged as spam.

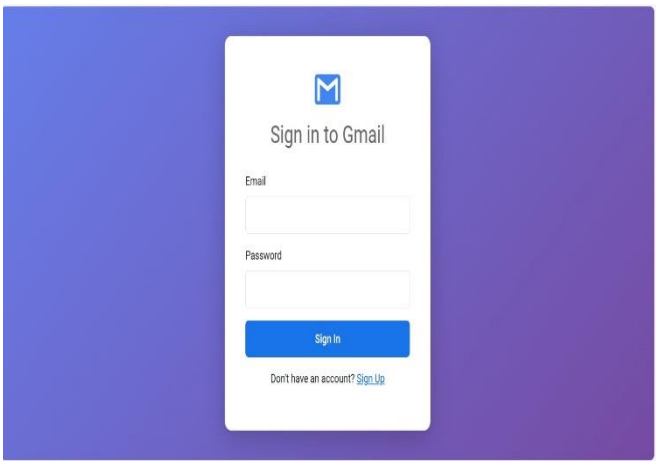

Figure 2 Sign in Page

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

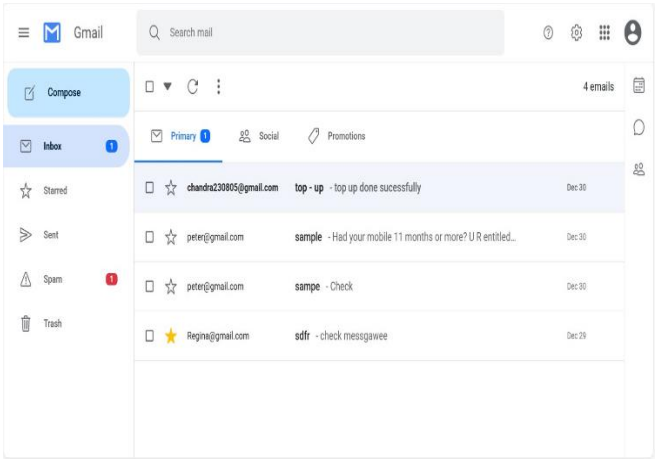

Figure 3 Inbox Page

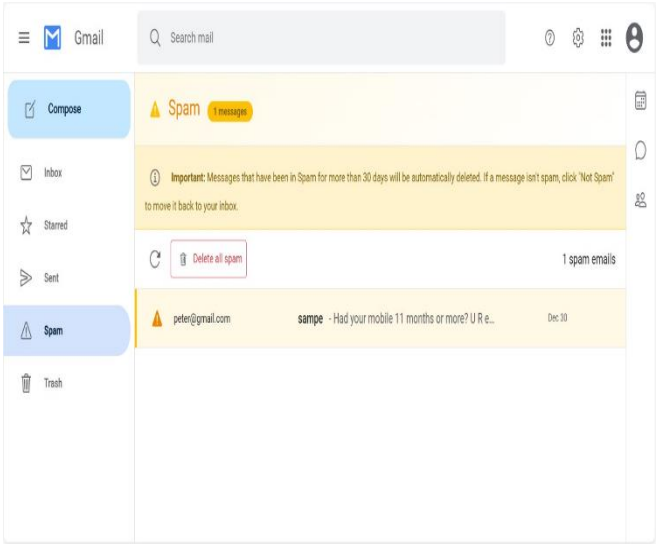

Figure 4 Spam Identification

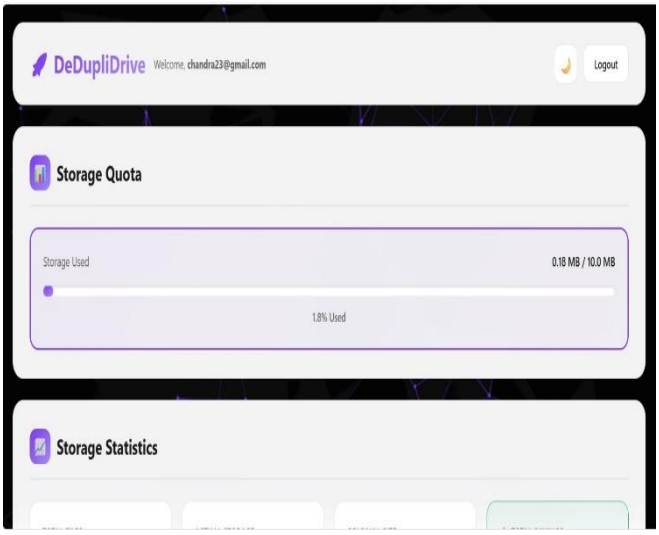

Figure 5 Deduplication Home Page

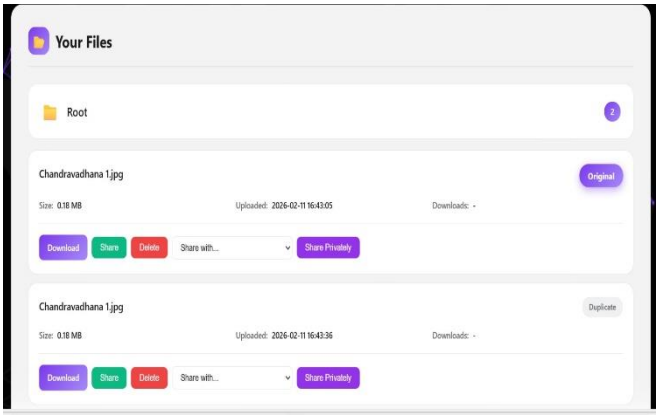

Figure 6 File Upload

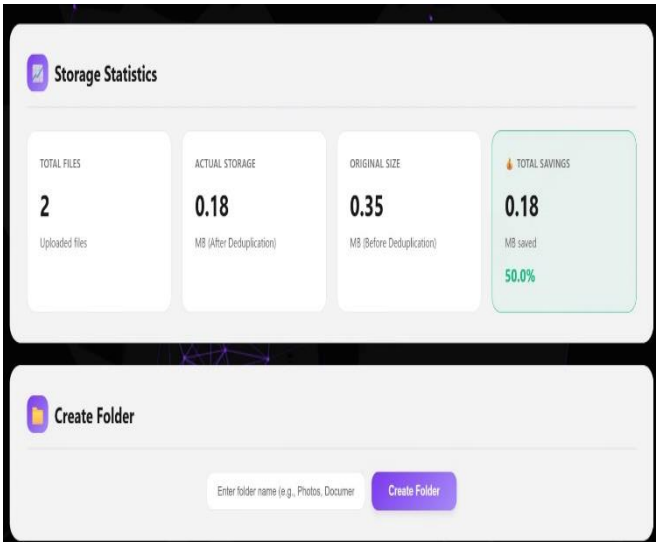

Figure 7 Storage Optimization

6. Performance

The performance comparison graph illustrates the evaluation of nine machine learning algorithms—Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Extra Trees (ET), AdaBoost (Ada), Gradient Boosting (GB), and XGBoost (XGB)—based on accuracy and precision metrics. From the results, it is evident that ensemble learning methods outperform traditional single classifiers. XGBoost achieves the highest performance, showing nearly perfect accuracy and precision, making it the most effective model among all. Extra Trees, Gradient Boosting, and AdaBoost also demonstrate very high performance, closely following XGBoost.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

Random Forest performs strongly as well, indicating the effectiveness of bagging techniques in improving classification stability and reducing variance. In contrast, traditional models such as KNN and Decision Tree show comparatively lower accuracy and precision, possibly due to sensitivity to noise and overfitting issues. Naïve Bayes and SVM perform moderately well but still fall behind ensemble-based approaches. Additionally, the accuracy and precision values for each model are closely aligned, indicating balanced classification performance with minimal false positives.

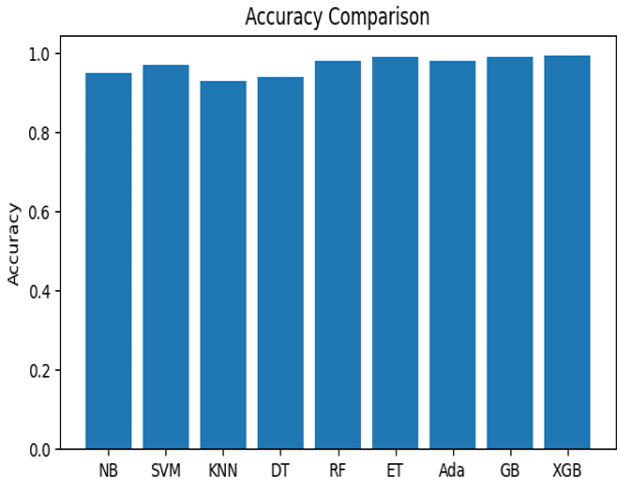

Figure 8 Accuracy Comparison

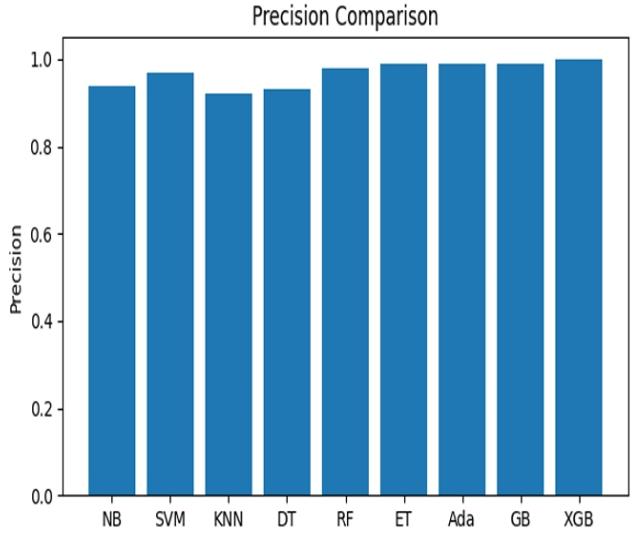

Figure 9 Precision Comparison

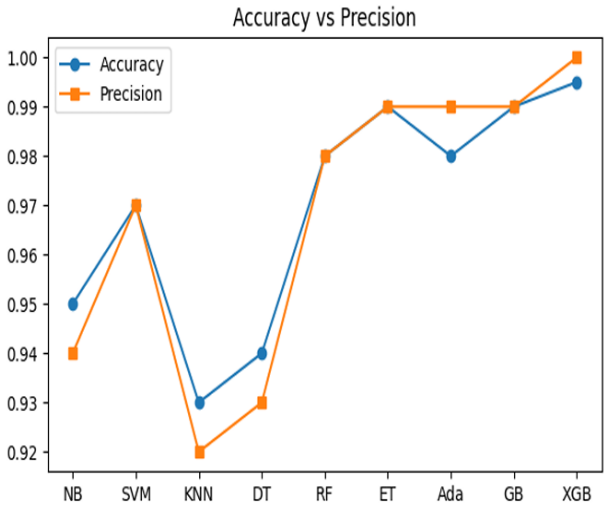

Figure 10 Accuracy Vs Precision

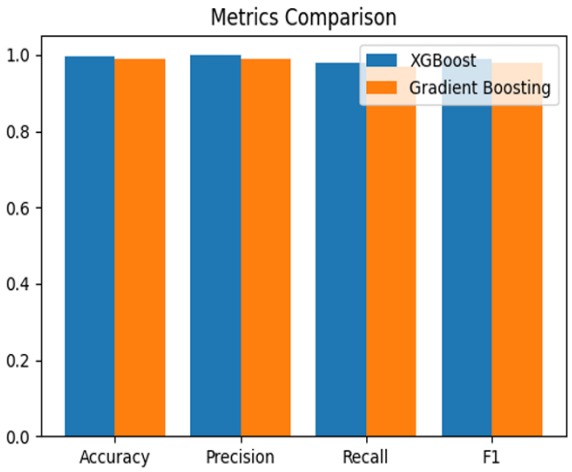

Figure 11 Metrics Comparison

7. Testing

The testing results of the model demonstrate strong classification performance across all evaluation metrics. The model achieved an accuracy of 0.9892 (98.92%), indicating that nearly 99% of the total predictions were correctly classified. The precision score of 0.9894 (98.94%), calculated using the weighted average method, shows that the model makes highly accurate positive predictions with very few false positives. Similarly, the recall score of 0.9892 (98.92%) reflects the model's strong ability to correctly identify actual positive instances, meaning it produces very few false negatives. The F1-score of 0.9890 (98.90%), which represents the harmonic

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1380-1390
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0190

mean of precision and recall, confirms that the model maintains an excellent balance between precision and recall. Overall, these testing results indicate that the model performs exceptionally well, demonstrating high reliability, robustness, and generalization capability on the test dataset.

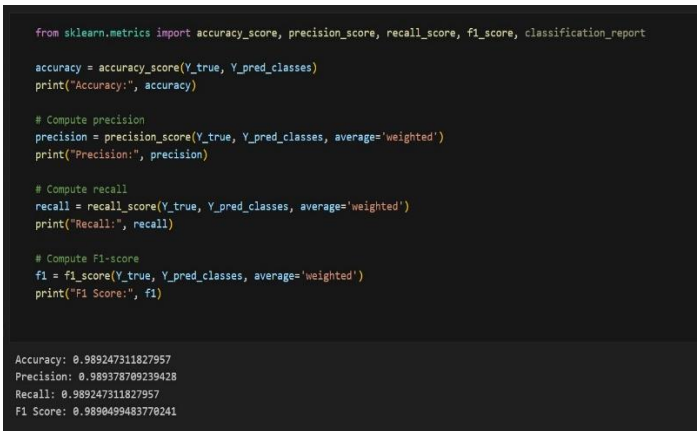

Figure 12 Testing with Metrics

8. Results

The implemented intelligent email management system successfully integrates CNN–LSTM based spam detection with automated email handling functionalities. During testing, the system effectively classified incoming emails into spam and legitimate categories with high accuracy. Emails containing phishing links or spam-related textual patterns were automatically filtered into the spam folder, thereby reducing inbox clutter and improving message organization. The preprocessing and tokenization steps ensured that email content was efficiently transformed into numerical form for deep learning model prediction, resulting in reliable classification performance. The web-based interface enabled users to securely register, log in, compose, send, and receive emails. Functional features such as marking emails as read, starred, deleted, or restored operated correctly through real-time database updates. Spam detection results demonstrated that the hybrid CNN–LSTM model effectively captured both contextual and sequential patterns in email content, minimizing false positives and false negatives. Overall, the system provided smooth email navigation across inbox, sent, spam, and trash folders, confirming that the proposed solution offers an efficient, intelligent, and user-friendly approach to email spam filtering and mailbox management.

9. Future Enhancement

Future works can make modifications to the system by focusing on real-time Gmail storage monitoring . AI capabilities such as zero-shot phishing detection, conversational email clustering, and machine-learning-based priority inbox classification can improve email security, organization, and user productivity. An automated email archiving mechanism can be in emails to Google Drive along with predictive storage warnings within a one-click smart cleanup button in Gmail interface. Advanced concept interventions like extending its CNN-LSTM–based architecture to support real-time Gmail API integration for continuous monitoring of the 15GB mailbox quota, predictive storage alerts, and automated archiving of older emails to Google Drive through a one-click Smart Cleanup feature. A multi-cloud backup orchestrator that enables intelligent data distribution, cross-cloud deduplication, and cost-optimized storage across platforms such as Google Drive, OneDrive, and Dropbox. Can add more privacy-preserving spam detection, multi-model email analysis by self-trained new models.

Conclusion

In conclusion, the proposed D-SPARC framework offers a practical and intelligent solution to the growing problem of email spam and duplicate data management. By combining a hybrid CNN–LSTM deep learning model with efficient duplicate detection techniques, the system successfully reduces inbox clutter and optimizes storage usage. The strong performance results, with nearly 99% accuracy, precision, recall, and F1-score, demonstrate that the model can reliably distinguish between spam and legitimate emails while minimizing errors. Unlike traditional approaches that rely on separate tools, D-SPARC brings spam filtering and data cleanup together into one unified platform, making email management simpler and more efficient. Overall, the system enhances user productivity, ensures smoother communication, and provides a scalable foundation for future advancements in intelligent email and storage management.

Reference

[1]. MD Rokunojjaman, Anup Malik, MD

Musfik Jahan Sajeeb, MD Yahiduzzaman, MD Sajedul Islam,"Multi-Feature Hybrid LSTM-CNN Framework for Phishing Email Detection", International Journal of Electronics and Communications System, vol. 5, no. 1, pp. 93 – 103, 2025.

[2]. S. Saleem, Z. UI Islam, S. A. Ibrar, "Spam Email Detection Using Long Short-Term Memory and Gated Recurrent Unit", Multidisciplinary Digital Publishing Institute, vol. 15, no. 13, pp.7407, 2025.

[3]. M. Ravikanth, C. L. Reddy, "An Efficient Learning-Based Approach for Automatic Record Deduplication with Benchmark Datasets", Scientific Reports (Springer Nature), vol. 14, no. 1, pp.63242, 2024.

[4]. G. Nasreen, M. Murad Khan, Muhammad Younus, Bushra Zafar, M. Kashif Hanif, "Email Spam Detection by Deep Learning Models Using Novel Feature Selection", ELSEVIER, vol. 35, no. 2, pp. 1-13, 2024.

[5]. Femi Emmanuel Ayo, Lukman Adebayo Ogundele, Solanke Olakunle, Joseph Bamidele Awotunde, Funmilayo A. Kasali, "A Hybrid Correlation-Based Deep Learning Model for Email Spam Classification Using a Fuzzy Inference System", ELSEVIER, vol. 35, Article 100390, pp. 1-14, 2024.

[6]. Mohammad Reza Feizi Derakhshi, Elnaz Zafarani-Moattar, Hussein Ala'a Al-Kabi, Ahmed Hashim Jawad Almarashy," PCLF: Parallel CNN-LSTM Fusion Model for SMS Spam Filtering ", BIO Web of Conferences[ISCKU], vol. 97, Article 00136, pp. 7-13 , 2024.

[7]. Tao. Xu, "Enhancing Cyber Security: Comparing the Accuracy of the BERT Model with Other Common Deep Learning Models in Identifying Email Spam", Advances in Engineering and Intelligence Systems, vol. 04, no. 01, pp. 84, 2025.

[8]. Georgios Petmezas, Vazgken Vanian, Konstantinos Konstantoudakis, Elena E. I. Almaloglou, Dimitris Zarpalas, "Video Deepfake Detection Using a Hybrid CNN-LSTM-Transformer Model for Identity Verification", Multimedia Tools and Applications (Springer Nature), vol. 84, pp.40617 – 40636, 2025.

[9]. M. Faseeh, Harun Jamil, "Revolutionizing Duplicate Question Detection: A Deep Learning Approach for Stack Overflow", IgMin Research – STEM, vol. 135, no. 1, pp. 1-6, 2024.

[10]. E. H. Tusher, M. A. Ismail, M. A. Rahman, A. H. Alenezi, M. Uddin, "Email Spam: A Comprehensive Review of Optimize Detection Methods, Challenges, and Open, Research Problems", IEEE Access, vol. 12, Article. 143627, pp.143627- 143657, 2024.