ENPM703 Fundamentals of AI and Deep Learning

Project Contribution Report

Group 3 - Phishing Email Detection with Multimodal Deep Learning

Version #:	1.0
Version Date:	12/18/2025
Team Name:	Group 3
Contributing Author(s):	Vishal Patil, Srihari Narayan, Anila Sai Namburi, Akash Vora

Team Collaboration and Shared Learning

Although specific responsibilities were divided among team members, the project was executed in a highly collaborative manner, and everyone gained visibility into - and understanding of - each major component. We held weekly scrum calls with professor Zachary Hanif, where we discussed progress, clarified doubts, and jointly planned upcoming tasks. During these sessions and our internal working meetings, we regularly walked through each other's code and design decisions, so that all four of us became familiar with the full workflow: data preprocessing, individual CNN models, the unified multimodal dataset, the fusion architecture, and the end-to-end inference pipeline and UI. As a result, every member not only contributed to their primary area, but also developed a strong conceptual understanding and practical learning across the entire system.

Individual Member Contributions

- ### Vishal Patil - Data Collection and Preprocessing

Vishal led the end-to-end data preprocessing effort for the project. He was responsible for consolidating emails from multiple public sources into a unified schema, handling tasks such as deduplication, label cleaning, and removal of corrupted or low-quality samples. He engineered the final text dataset split (training/validation), standardized fields such as subject, body, sender, receiver, and URL lists, and ensured that class balance was preserved across splits. In addition, Vishal implemented key text preprocessing steps, including token normalization, maximum sequence length selection, and vocabulary construction, which directly fed into the Text CNN pipeline. His work ensured that all downstream models were trained on clean, consistent, and well-curated data.

- ### Srihari Narayan - Text and Image CNN Model Development

Srihari was primarily responsible for designing and implementing the individual convolutional neural network architectures for both text and image modalities. For the text branch, he adapted a CNN classification architecture, defining the embedding layer, convolutional block configuration, pooling strategies, and fully connected layers for phishing vs. legitimate classification. For the image branch, he implemented a VGG-style Custom Image CNN for logo recognition and set up the training and evaluation pipeline, including data augmentation, metric tracking, and model checkpointing. Srihari also contributed to hyperparameter tuning (learning rates, batch sizes, optimizer settings) and performed detailed analysis of training/validation curves and error cases for the standalone CNN models, which later served as strong backbones for the fusion architecture.

- **Akash Vora - Unified Multimodal Dataset and Fusion Architecture**

Akash led the multimodal integration aspect of the project. He designed and constructed the unified dataset that links textual emails with corresponding logo images and engineered the metadata feature vector used in the fusion model. This involved mapping brand mentions in the email text to logo IDs, normalizing metadata fields, and creating the unified_multimodal_text.csv dataset that serves as the central training resource for the dual-tower model. Akash then implemented the Dual-Tower Fusion architecture, including the text and image feature extractors, the metadata MLP, and the 832-dimensional fusion classifier. He integrated these components into a full HTML-to-decision inference pipeline that parses raw HTML emails, extracts text and images, applies preprocessing, and produces phishing predictions with calibrated confidence scores. He also coordinated experiments comparing single-modality baselines to the fusion model and analyzed their performance.

- **Anila Sai Namburi - Unified Dataset, Web UI and Streamlit Deployment**

Anila contributed both to the data layer and to the user-facing deployment of the system. She worked closely with Akash on the mappings and creation of the unified multimodal dataset, helping to align text samples with their corresponding logo images and ensuring that the dataset structure was compatible with the fusion model's input requirements. This included verifying label consistency, checking brand-to-logo mappings, and validating sample integrity across the combined text, image, and metadata features. In addition, Anila was responsible for making the system accessible through an end-to-end web application. She designed and implemented the user interface in Streamlit, allowing users to upload raw HTML email files, trigger inference, and view model outputs in an interpretable way. The UI presents the predicted label (phishing vs. legitimate), confidence score, and a structured summary of detected issues such as suspicious URLs, urgency indicators, and metadata anomalies. Anila also deployed the complete pipeline - models, preprocessing code, and UI on Hugging Face Spaces, handling environment configuration and dependency management. Her contributions ensured that the research prototype was packaged into a practical, interactive tool suitable for demonstrations and further evaluation.