

Phishing Email Detection with Multimodal Deep Learning

The Power of Fusion: Detecting Phishing from Every Angle

Group 3

Vishal Patil, Akash Vora, Srihari Narayan, Anila Sai Namburi

THE PROBLEM: AN EVOLVING, MULTIMODAL THREAT

Responsible for billions in financial losses annually.

Traditional Defenses



Traditional text-only systems fail to analyze the complete picture

Modern Attacks



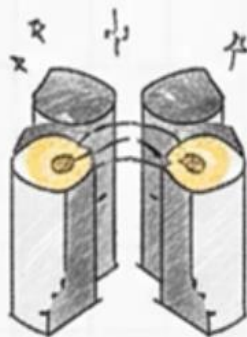
Attackers combine deceptive language with legitimate-looking brand logos.

Our Goal: A system that analyzes both what an email *says* and what it *shows*.

Input: Raw email content (text, images, HTML structure).

Output: Binary classification: Phishing or Legitimate.

Four Project Phases

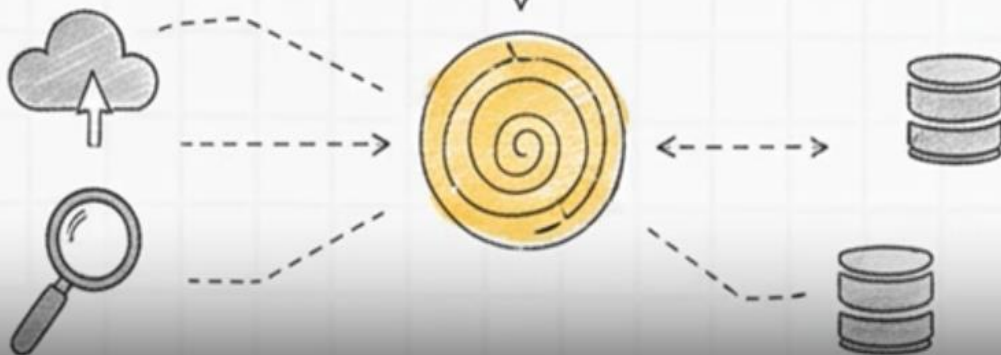


Phase 1

Pre-training the text and image 'specialist' models independently.

Phase 2

Integrating all data into a single 'golden dataset'.



Phase 3

Training the final dual-tower fusion model with all data types.

Phase 4

Deploying the model for real-time inference on raw HTML emails.



OUR APPROACH VS. RELATED WORK

Previous Works (e.g., Doshi et al.)

- ✓ Content-based features & URL analysis.
- ✗ Modalities treated independently (no true fusion).

Other works focus only on text (Zhang & Wallace's CNNs) or only on images.

Our Approach

- ✓ Deep text analysis via Custom CNN.
- ✓ Deep logo analysis via Custom CNN.

Key Innovation:

A jointly trained, multimodal fusion architecture that learns to dynamically weigh text, image, and metadata signals in a single end-to-end system.

OUR FOUNDATION: A DIVERSE, LARGE-SCALE DATASET



Email Text & Metadata

76,346

Emails

From 4 public sources (CEAS_08, Enron, Nazario, etc.)
Balanced 50/50 Phishing-Legitimate Split

20-dim

Engineered Metadata Vector



Brand Logos

72,652

Images

From OpenLogo Dataset
Resized to 224x224 pixels and normalized.

352

Brand Classes

Loss Function: Cross-Entropy Loss | Evaluation Metrics: Accuracy, F1-Score, and ROC-AUC

Single-Modality Detectors Have Clear Limitations

Text-Only Analysis

Custom Text CNN

98.96%

Strong, but can be fooled by carefully crafted language.

Logo-Only Recognition

Custom Image CNN

76.30%

A useful signal, but insufficient on its own.

Building the Foundation for Fusion

1. Train Specialist Models

Text CNN

Trained on 76,346 emails
(CEAS_08, Enron).

Image CNN

Trained on 72,652 logos
(OpenLogo).

2. Engineer a Unified Dataset

The Golden Dataset

- Integrated text, logo IDs, and labels.
- Engineered 20 metadata features (URL counts, sender patterns, etc.).

A SYSTEMATIC EVALUATION ACROSS SIX ARCHITECTURES



Traditional ML Baselines

K-Nearest Neighbors: **81.71%**

Logistic Regression: **80.00%**



Unimodal Deep Learning

Custom Text CNN: **98.96%**

Custom Image CNN: **76.30%**



Transfer Learning

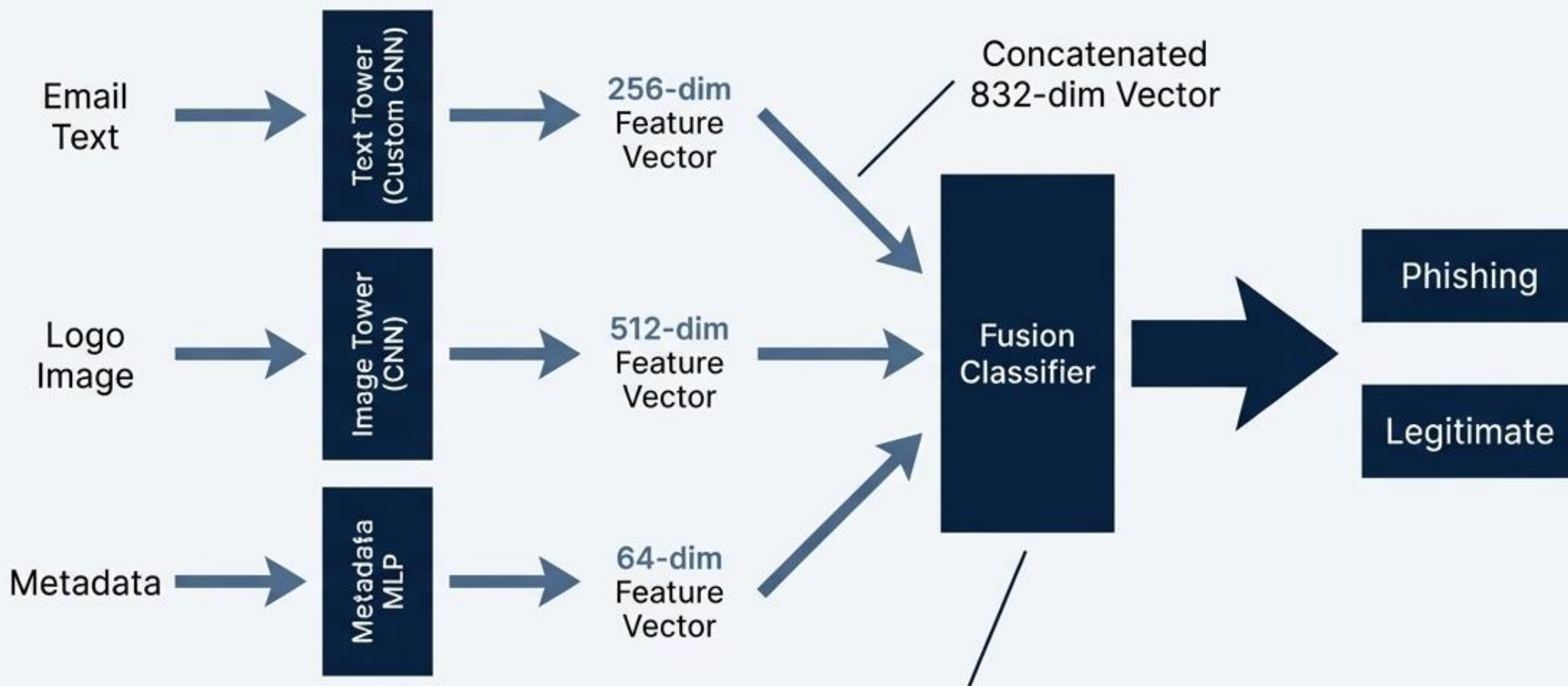
Pretrained ResNet18
(Image): **97.43%**



Multimodal Fusion

**Dual-Tower Fusion
Model: 99.45%**

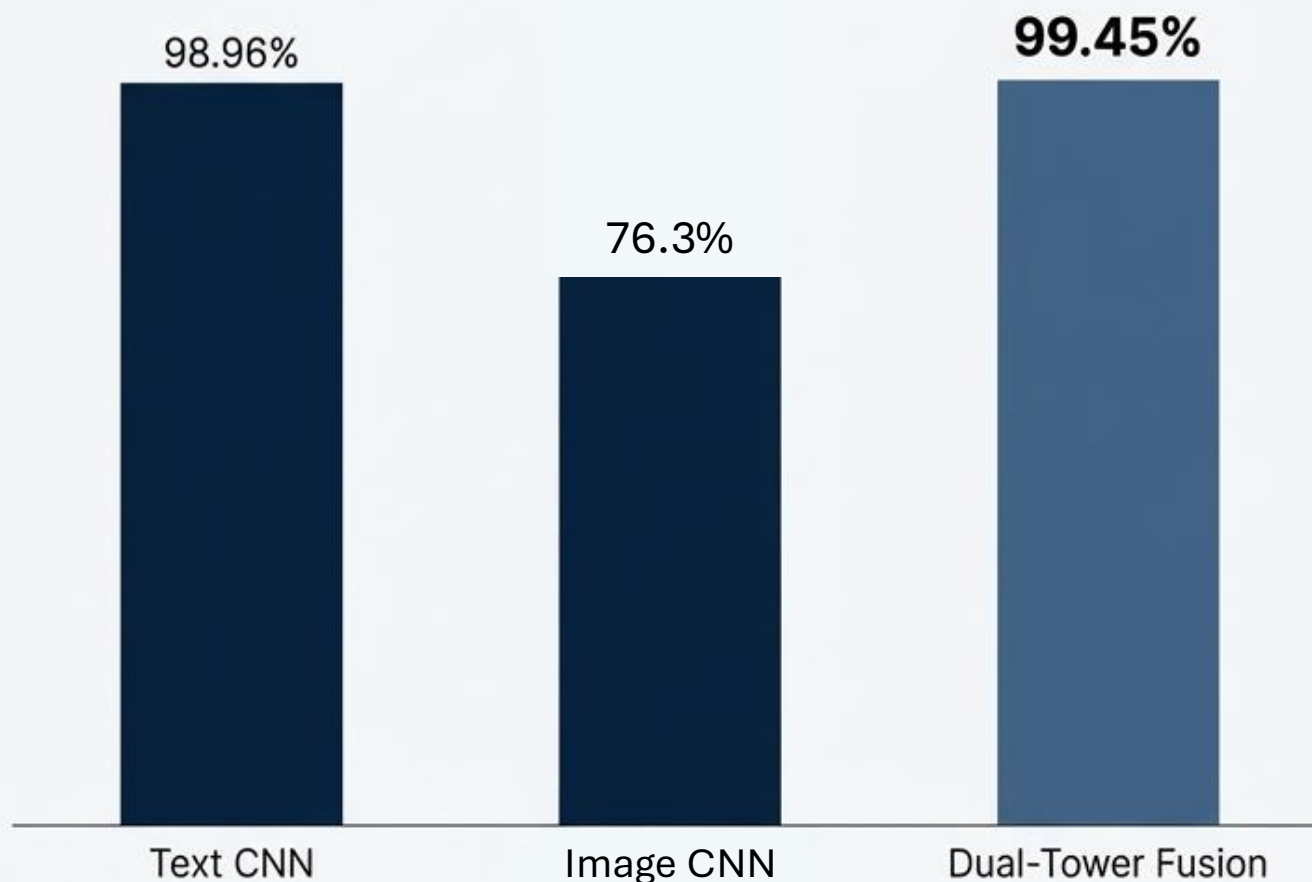
THE INNOVATION: A DUAL-TOWER FUSION ARCHITECTURE



Catches inconsistencies a single-modality approach would miss, such as a mismatched sender domain and brand logo.

THE VERDICT: MULTIMODAL INTEGRATION ACHIEVES SUPERIOR PERFORMANCE

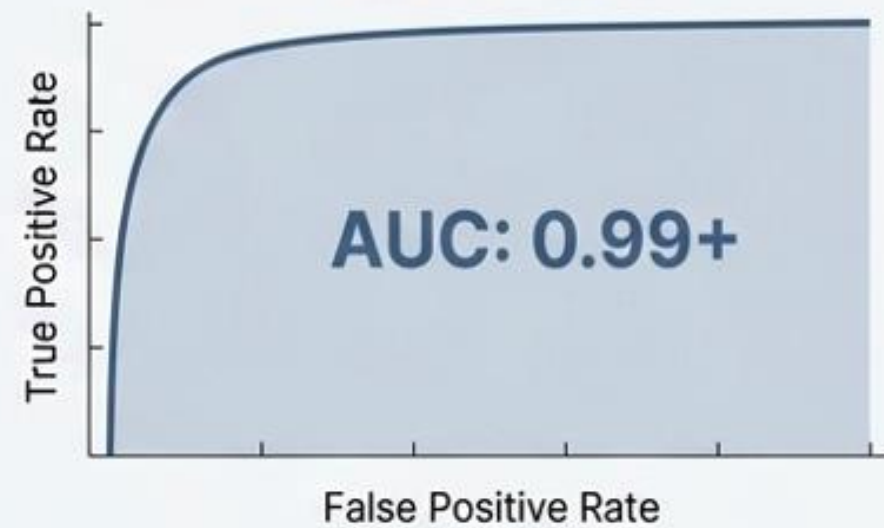
Validation Accuracy Comparison



99.45 %

Fusion Model Accuracy

ROC Curve



FROM MODEL TO APPLICATION: A REAL-WORLD DEPLOYMENT

The Pipeline



The Application

Multimodal Phishing Email Detector

Upload an HTML email file (.html/.htm). The model extracts text + images + metadata and predicts phishing vs legitimate.

Upload .html email

Drag and drop file here
Limit 200MB per file • HTML, HTM

Browse files

C11 (1).html 2.9KB

Filename: C11 (1).html

Analyze email

Result

⚠️ PHISHING — 100.00% confidence

Device: cpu

Detected Issues / Signals

- Multiple call-to-action phrases detected

Download report

Download report (.txt)

Extracted Email Content

Subject

Dear customer,

Body (preview)

Body

account may be permanently restricted.

PayPal Security Team

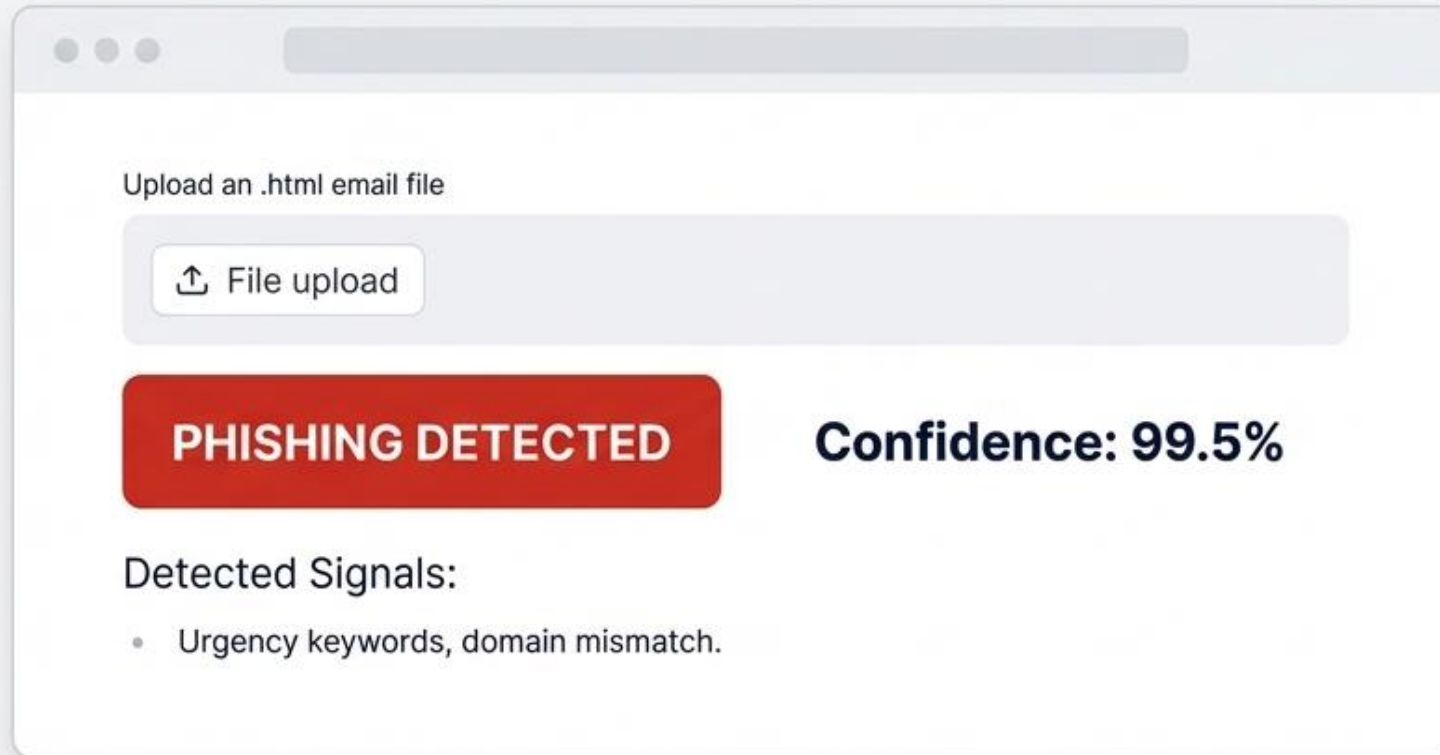
-----=_Part_C11_logo
Content-Type: text/html; charset="utf-8"
Content-Transfer-Encoding: 8bit PayPal Security Alert Dear customer, We noticed unusual activity on your PayPal account and have temporarily limited certain features. To restore full access, please confirm your account information within 24 hours. If you do not verify in time, your account may be permanently restricted. Restore My Account -----
=_Part_C11_logo--

Images found: 1

Successfully identified **100%** of 20 synthetic phishing emails (with and without logos), with **zero** false positives on legitimate test emails.

Live System Demo

<https://huggingface.co/spaces/anilawork/phish-detection-ui-final>



The screenshot shows a web application interface for phishing detection. At the top, there is a text input field with the placeholder "Upload an .html email file". Below this is a "File upload" button with an upward arrow icon. The main result area features a large red button with the text "PHISHING DETECTED" and a "Confidence: 99.5%" label to its right. Underneath, the section "Detected Signals:" is followed by a bulleted list containing "Urgency keywords, domain mismatch."

Upload an .html email file

File upload

PHISHING DETECTED **Confidence: 99.5%**

Detected Signals:

- Urgency keywords, domain mismatch.

An end-to-end inference pipeline deployed as an interactive web application on Hugging Face.

IMPACT, LIMITATIONS, AND THE PATH FORWARD



BROADER IMPACT

- Provides an open-source framework for multimodal threat detection.
- Demonstrates a practical, end-to-end deployable system.



LIMITATIONS

- Static datasets may not reflect new zero-day attacks.
- Scarcity of real-world phishing HTML samples for training.
- Class imbalance exists in the logo dataset for less common brands.



FUTURE WORK

- Evaluate the system on a live feed of phishing emails.
- Incorporate advanced text embeddings like BERT.
- Expand classification to distinguish between spam and phishing.



Thank You

Group 3

Vishal Patil, Akash Vora, Srihari Narayan, Anila Sai Namburi

Phishing Email Detection with Multimodal Deep Learning