USC Viterbi School of Engineering

Utilizing AI Generated Images for Object Detection and Classification

Amy Jiang, Dongwook Kim, Harshita Dooja Poojary, Michael Kingsley, Narges Ghasemi, Nealson Setiawan

Advisor: Seon Ho Kim

DataFirst
Fall 2023

Background and Motivation

- Object detection and classification play an important role in smart-city applications as they can replace human surveillance monitoring in identifying critical situations
- However it is often hard to obtain a good training dataset when an object or scenario is rare and costly
- For example: "wounded humans in disastrous situations"
- In such scenarios, artificially generated images can serve as a potential solution
    - Leveraging the advancements of GANs and other data generation techniques, we can create realistic images of rare events, thereby augmenting training datasets
    - Consequently, generated images may help fill the data gaps, ensuring a more comprehensive and diverse dataset for training ML models

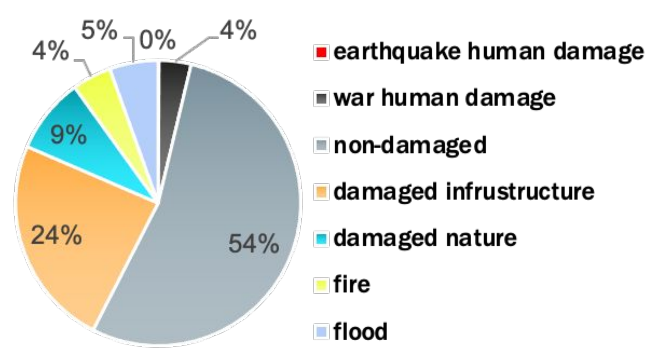
- State-of-the-art generative AI models can create high-quality synthetic images fast
- Carefully crafted generation prompts can serve as labels, reducing workload in labelling and data cleaning process
- The idea: Augmenting existing datasets with synthetic images to improve the accuracy on humans in abnormal postures, particularly in earthquakes
- Primary focus:
    - Replicating post disaster scenarios
        - Injured survivors in debris
        - Damaged buildings
    - Detecting critical objects:
        - Humans, Buildings
    - Damage Assessment
        - Number of injured people
        - Severity of injuries
        - Infrastructure Damage

*Sample image created using DALL-E 2

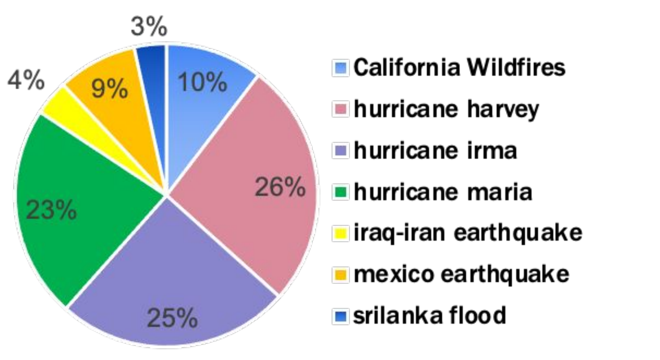
Quality of Existing Datasets

Dataset	Description	Relevant Image Count
South Asia dataset (SAD)	Fire Disaster, Flood Disaster, Infrastructure, Nature Disaster, Human Damage, and Non Damage	240 out of 13720 images
Crisis Image Benchmarks Dataset	Affected, injured, or dead people, Infrastructure, and utility damage	1375 out of 5541 images
Indonesian Disaster Victims 50 (IDV-50)	50 images from the internet and real disaster	50 images
MEDIC dataset	Images sourced from various disaster scenarios, including earthquake, fire, flood, hurricane, and landslide It consists of Crisis Dataset, aidr dataset and asonam dataset	25583 out of 72380 images
Disaster Images Dataset	Images of natural disasters like Cyclone/Hurricane, Earthquake , Flood, Wildfire	1122 out of 4428 images

Crisisnlp Multimodal Damage Identification Dataset



Crisisnlp Dataset



Key Observations in Existing Datasets:

- Existing datasets generalize disaster types, lacking focus on earthquakes.
- There's a significant lack of images showing individuals amid earthquake debris.
- The abundance of non-relevant images reduces the datasets' overall usefulness.
- Many photos are taken too close or includes satellite imagery, missing broader context.
- Numerous pictures are of low quality.

An enhanced, focused dataset containing images of people in the debris of earthquakes is needed to improve AI-driven response in earthquake human detection

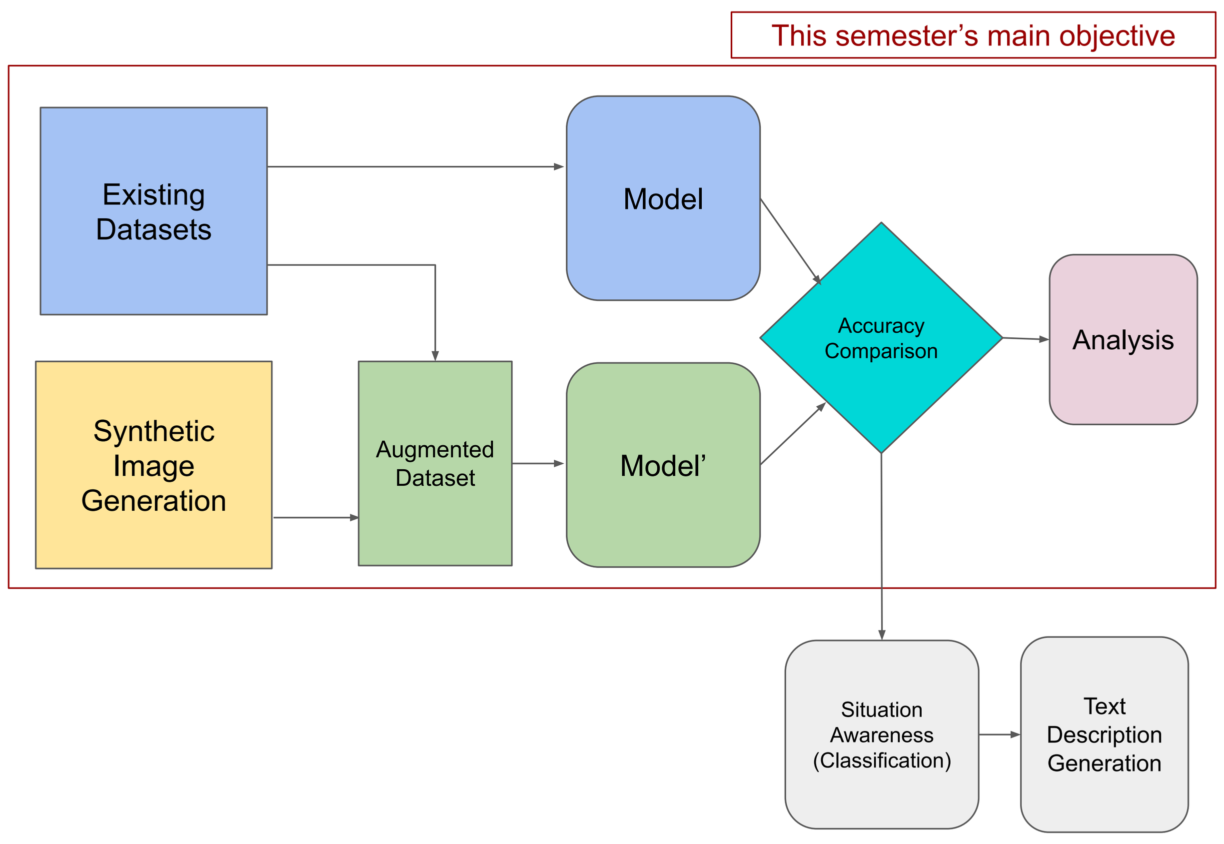
Method

Purpose	Models/Tools Used	Observations
Generate Synthetic Images	Dalle 2, Stable Diffusion	Dalle 2: • Generation time : 10 seconds • Issues: ○ Does not generate images for injured people due to content Policy.
Image Annotation	VoTT, LabelImg, CVAT	VoTT: • manual graphical annotation tool developed by microsoft • annotation data exported to JSON includes: ○ labels ○ height ○ coordinates, etc
Object Detection	Yolov8, FCOS,TTFNet, PAFNet	YOLO : • Pretrained model : ○ Pretrained on COCO dataset. ○ Issues: ■ Difficult to detect survivors who are laying flat on the ground or obscured by dust. ■ Struggles to detect individuals from aerial or drone perspectives. ○ Solution: Finetune Model with the current dataset.
Text Generation	CLIP-Interrogator-2, pix2struct, Mini GPT-4	CLIP-interrogator-2 • pretrained on ViT-H-14 OpenCLIP model to use to create similar image in stable diffusion model 2.0 • text generation time: ○ quality = good: 44.56 s ○ quality = poor: 17.94 s • issues: ○ depends on quality or size of image (poorer quality → faster output times) • example output: ○ a pile of debris sitting on top of a beach next to a body of water, wood pier and houses, wall, beaches, fallen columns

Example Prompts:

- *"A photorealistic, high-resolution image of a city street after an earthquake with clear and detailed humans."*
- *"With the clarity and depth seen in National Geographic photos taken with a Nikon, showcase a city street in the aftermath of an earthquake, where every human figure lying amidst the debris is as distinct and detailed as a portrait."*

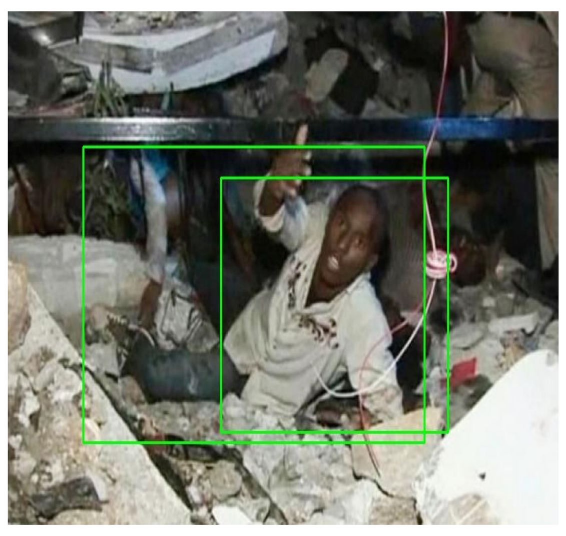
Workflow



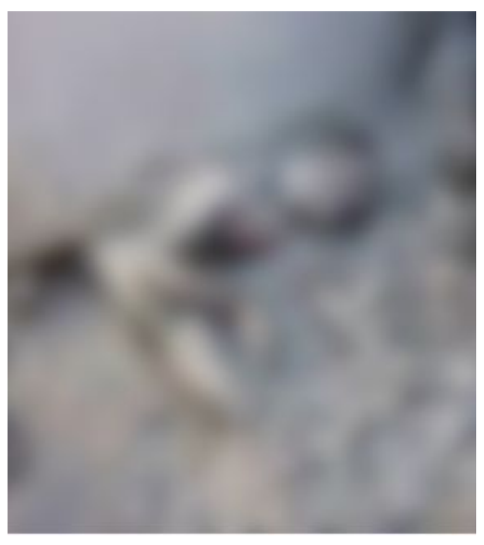
- **Model:**
    - Identifies critical objects and poses.
- **Situation awareness**:
    - Distinguishes the scenarios based on the severity
- **Text Generation**:
    - Creates text summary description of the image
    - Could generate the keywords for the scenario.
- **Datasets**:
    - Existing Dataset: Preprocessed and filtered images with post earthquake scenarios
    - Synthetic Image Generation: Using Generative AI to replicate post earthquake images using prompts.

Key Observations

- ***Dalle 2***:
    - Dalle 2 restricts users to create some images of injured people due to the content policy.
- ***Pre-Trained Yolov8***:
    - Difficult to identify humans in the rubbles of the earthquake
    - Pose estimation becomes difficult when the person is not identified



Person identified in the debris



Person misclassified as bird

People not identified by Yolov8

Next Steps

- **Annotate** existing dataset for human detection.
- **Finetune** Yolov8 with existing Real image datasets to improve human detection in post earthquake scenarios. This will work as a benchmark for comparison.
- **Merge** the existing dataset with the images generated by AI and train Yolov8 and **compare** the model performance with the benchmark based on evaluation metrics.
- Once we confirm that training dataset augmented with synthetic images improves performance in object detection, the pose estimation, text generation on parallel.
- The detections and summary would be processed to analyze the severity of the situation.

Auditing Elon Musk's Impact on Hate Speech and Bots

Daniel Hickey[1] (hickeyda@oregonstate.edu), Matheus Schmitz[2], Daniel Fessler[3], Paul Smaldino[4], Goran Murić[2], Keith Burghardt[2]

[1]Oregon State University, [2]USC Information Sciences Institute, [3]University of California Los Angeles, [4]University of California Merced

Introduction:

- October 27th, 2022: Elon Musk purchased Twitter, becoming its new CEO.
- Musk promised less moderation & fewer social bots.
- Decreased moderation may increase hate on Twitter; the promise of fewer bots has been untested.
- We measured levels of hate and bots on the platform before & after the purchase.

Methods:

- We extract tweets from 2022 that contain:
  - Hateful keywords & are sufficiently toxic according to the Perspective API (4.4K)
  - Common keywords (4.9M baseline tweets)
- We extract 2 months of tweets from a random subset of users who wrote hateful tweets (4.2K)
- Collect bot scores for a subset of accounts (15K) using Botometer
- Analyze the results over time to determine the impact of Elon Musk buying Twitter

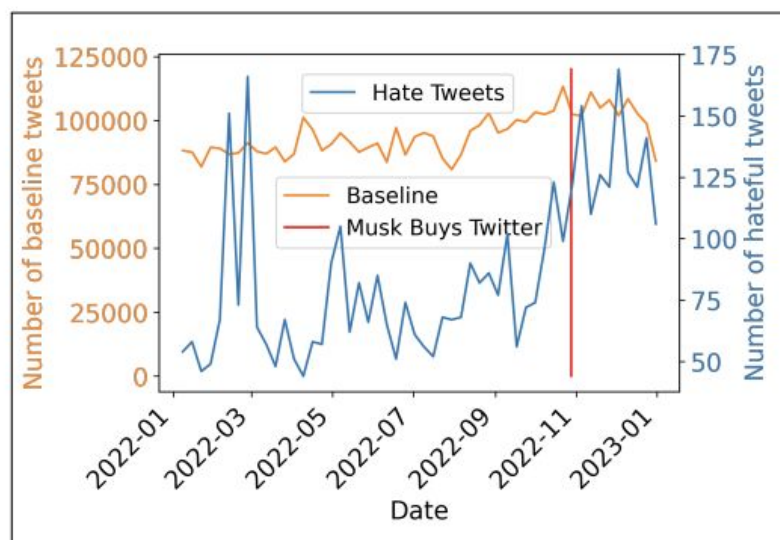
Conclusions:

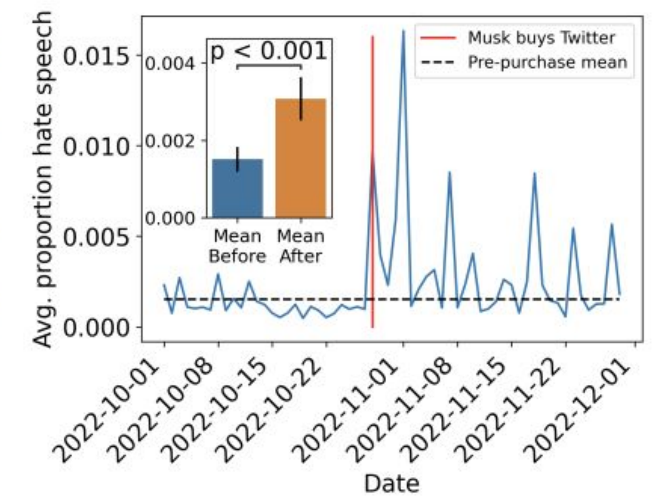
Following Musk's purchase…

- Overall volume of hateful tweets increased
- Hateful users used more hate speech
- Overall bot prevalence did not change significantly

Acknowledgements:

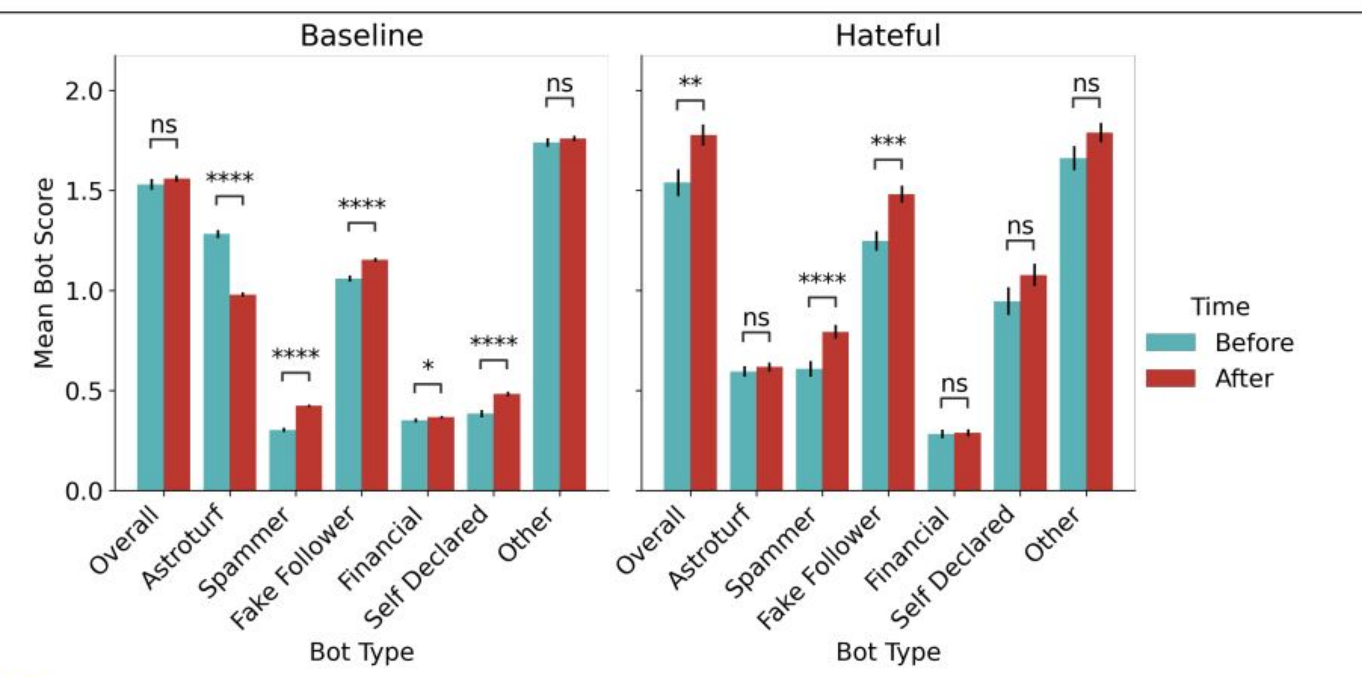
Funding for this work is provided through the USC-ISI Exploratory Research Award, NSF (award #2051101), and through DARPA (awards #HR0011260595 and #HR001121C0169).





RQ1: Did overall hate speech on Twitter increase? **YES**

RQ2: Did hateful users become more hateful? **YES**



RQ3: Did the prevalence of bots decrease after Musk's purchase? **NO**

We are using *digital menu data* and *Machine Learning* to predict and characterize *food environment nutritional quality*

Pilot Project: Analyzing Digital Menu Data to Characterize Food Outlet Nutritional Quality in the City of Los Angeles

Abigail Horn[a,b]* [Project PI], Alex DongHyeon Seo[a], Kayla de la Haye[c], Andrés Abeliuk[d], Keith Burghardt[a]
[a] Information Sciences Institute, University of Southern California
[b] Department of Industrial and Systems Engineering, University of Southern California
[c] Department of Population and Public Health Sciences, University of Southern California
[d] Department of Computer Science, University of Chile
*Project PI, Corresponding author: Abigail.Horn@usc.edu

Research Reported In:

What's On the Menu? Towards Predicting the Nutritional Quality of a Restaurant Menu
Under Review at ML4Health Conference

A Continuous Indicator of Food Environment Nutritional Quality
Data-driven Humanitarian Mapping, 27th ACM SIGKDD Conf

Project Support:

National Institute of Minority Health and Health Disparities (NIMHD) of the NIH under award number P50MD017344, to the Southern California Center for Latino Health

Motivation and Knowledge Gap

Poor diet is heavily driven by environmental factors

Low-quality food environments:
- Impact diet and related diseases through (lack of) **access and psychological cues**
- Are frequently **concentrated among low-income communities** and **communities of color**

Food environments:
A key **Social Determinant of Health***

Characterizing food environment nutritional quality

Grocery = Healthy?

Fast food = Unhealthy?

How do we characterize food outlet nutritional quality in food environment research?
- Typically we consider broad categories of outlets
- How many categories do we need?
- What is nutritional distribution within categories?

Fast food outlet categorization example

Typically: **Fast Food outlets** measured as NAICS-coded "limited service restaurants":
Establishments primarily engaged in providing food services where patrons generally order or select items and pay before eating.

Includes outlets we generally think of as unhealthy…

Shifts in foods offered at limited-service restaurants to include:

…to what we think of as healthy

Significance and Impact

Food Environment Research
Index informs nutrition access studies: etiologic **associations** between **food environment access/exposure, diet, & nutritional health**

Longer-Term Vision:
Incentivize healthier food retail
- Integrate nutritional scores into delivery app platforms, search engines
- Like calorie labeling, work to incentivize restaurants to provide healthier foods

Public Health Program Interventions
Use index to identify outlets for food and nutrition insecurity programs and interventions

Stakeholders:
Dipa Shah-Patel, MPH, RD, Director - Nutrition & Physical Program, LAC Dept. PH
Robert Graham, MD, Menus of Change

Idea and Aims

Our Research: Driving Idea

Data is available at scale to help characterize nutritional quality of food outlets …
- *Digital menus*
- *Ingredient & nutrient data*

Can we use these data to develop an approach to directly evaluate nutritional quality of foods offered?

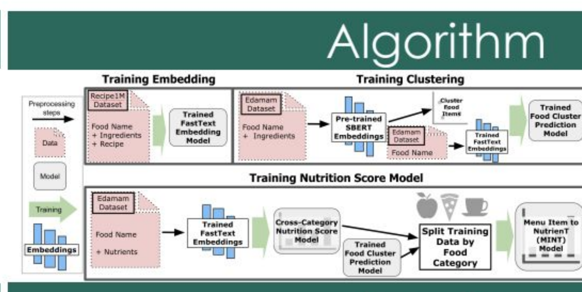
AIMS

Develop Algorithm
Develop algorithm capable of estimating nutritional quality of restaurant menus – across a continuum

Apply to LA Food Environment
Use algorithm to analyze disparities in the nutritional quality of food outlets and food environments in City of LA

Share Via Data Visualization Portal
Make results accessible to stakeholders in research and policy through a data visualization portal

Algorithm



- **Word Embedding Model**: Extract food names, ingredients, and recipe from the Recipe1M+ dataset, concatenate, use to train FastText word embedding model
- **Food Category Prediction Model**: Embed Edamam training dataset using pre-trained SBERT model. Cluster sentence embeddings with HDBSCAN and treat as ground truth category in word embeddings from trainedFastText model. Predict food name clusters
- **Nutrition Score Model**: Use menu item word embeddings from FastText to predict *RRR*, by first training on entire dataset then fine-tuning on predicted food categories

Methods

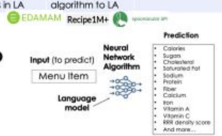
Step 1: Collect Data from nutrition data APIs

Type of Data	Purpose	Count
EDAMAM: Generic menu items	Training main model	81,390 menu items w/ nutrients + ingredients
RECIPE 1M+: Recipes	Training language model	1M recipes w/ nutrients + ingredients
SPOONACULAR: Digital menus for chain restaurants	Method validation analysis	~38,275 menu items w/ nutrients
HYBRID COLLECTION: Digital menus in LA	Applying algorithm to LA	~500 and increasing

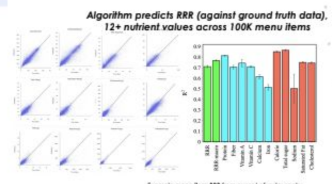
Example food items + ingredients

Meal type	Food Item	Ingredients
Appetizers	Poutine	Beef stock, Flour, Butter, Curd, Black pepper, Salt, White potatoes
Entrées	Ratatouille Provencale	Tomato puree, Herbes de Provence, Eggplants, Zucchini, Pepper, Extra-virgin olive oil, Tomatoes, Garlic, Salt, Onions
Desserts	Cinnamon Swirl Rolls	Rice milk, Baking soda, Dark brown sugar, Yeast, Applesauce, Water, Baking powder, Granulated sugar, Golden Syrup, Xanthan gum, Baking Mix, Ground cinnamon, Salt, Shortening, Ground flaxseed, Raisins

Step 2: Develop ML algorithm to predict menu item nutrient composition



Algorithm predicts RRR (against ground truth data), 12+ nutrient values across 100K menu items



Step 3: Compute menu item level nutrition density (RRR)

$$RRR = \frac{(\%DV\ protein + \%DV\ dietary\ fiber + \%DV\ calcium + \%DV\ iron + \%DV\ vitamin\ A + \%DV\ vitamin\ C)/6}{(\%DV\ calories + \%DV\ sugars + \%DV\ cholesterol + \%DV\ saturated\ fat + \%DV\ sodium)/5}$$

* *RRR-macro is RRR without micronutrients* ***Calcium, Iron, Vitamin A, Vitamin C***

Example **menu item RRR** from our set of restaurants:

Low Nutritional Density	RRR score	Middle Nutritional Density	RRR score	High Nutritional Density	RRR score
Vanilla Creme Donut	0.16	Huevos Rancheros	1.17	Grilled Chicken Breast	2.15
Chips (1 oz)	0.62	Chicken Quesadilla	1.30	Black Beans & Rice	2.80
Five Cheese Tomato Pizza	0.71	Burrito Mojado - Chicken	1.38	Mung Bean Salad	5.84
Mac & Cheese with Chiles	0.83	Sauteed Steak	1.5	Broccoli, 1 cup steamed	6.75

Step 4: Compute restaurant-level nutritional quality index (RNQ)

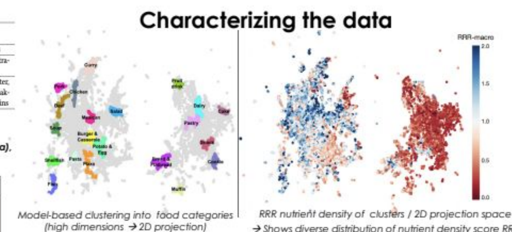
RNQ = Multiple methods:
- Median across menu items
- Median across multiple samples of a la carte orders
- % of items above score of XYZ (e.g. above 1)

Example **restaurant RNQ** from our set of restaurants:

Low RNQ	RNQ score	Middle RNQ	RNQ score	High RNQ	RNQ score
Cold Stone Creamery	0.16	Subway	1.15	Sweetgreen	1.82
Buffalo Wild Wings	0.18	El Pollo Loco	1.18	Sansai Japanese Grill	1.86
McDonald's	0.34	District Taco	1.44	Veggie Patch	2.50

Results

Characterizing the data

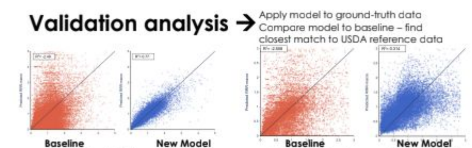


Model-based clustering into food categories (high dimensions → 2D projection)

RRR nutrient density of clusters / 2D projection space → Shows diverse distribution of nutrient density score RRR

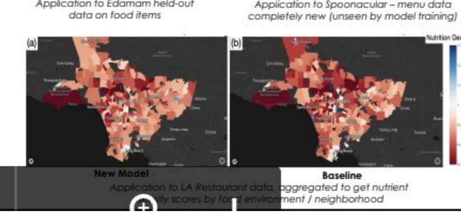
Validation analysis →

Apply model to ground-truth data
Compare model to baseline – find closest match to USDA reference data



Baseline **New Model**
Application to Edamam held-out data on food items

Baseline **New Model**
Application to Spoonacular – menu data completely new (unseen by model training)



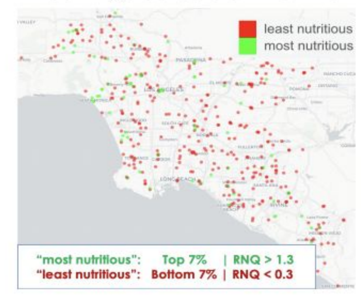
New Model **Baseline**
Application to LA Restaurant data, aggregated to get nutrient density scores by food environment / neighborhood

Prediction results

Ablation study with average R^2 from 5-fold Cross Validation

NLP Model	Prediction Model	R^2 ↑	RMSE ↓	MAE ↓
Pre-trained (base-uncased) BERT	FCWM	0.363±0.013	0.241±0.002	0.338±0.005
	MLP	0.463±0.01	0.204±0.004	0.314±0.005
	MINT	0.529±0.009	0.176±0.004	0.279±0.001
Pre-trained (Common Crawl) FastText	FCWM	0.405±0.02	0.226±0.001	0.326±0.004
	MLP	0.673±0.011	0.121±0.004	0.229±0.002
	MINT	0.757±0.007	0.093±0.004	0.189±0.002
FastText trained w/ Recipe1M+	FCWM	0.406±0.019	0.226±0.001	0.325±0.004
	MLP	0.696±0.008	0.115±0.003	0.223±0.001
	MINT	**0.768 ± 0.006**	**0.088 ± 0.003**	**0.183 ± 0.001**
USDA-matching algorithm (Lin et al., 2021)		-2.49±0.745	1.405±0.314	0.569±0.01

Illustrative application to Los Angeles food environment



"most nutritious": Top 7% | RNQ > 1.3
"least nutritious": Bottom 7% | RNQ < 0.3

Result: *Refined Identification of Disparities in Healthy Food Access*

Datasets

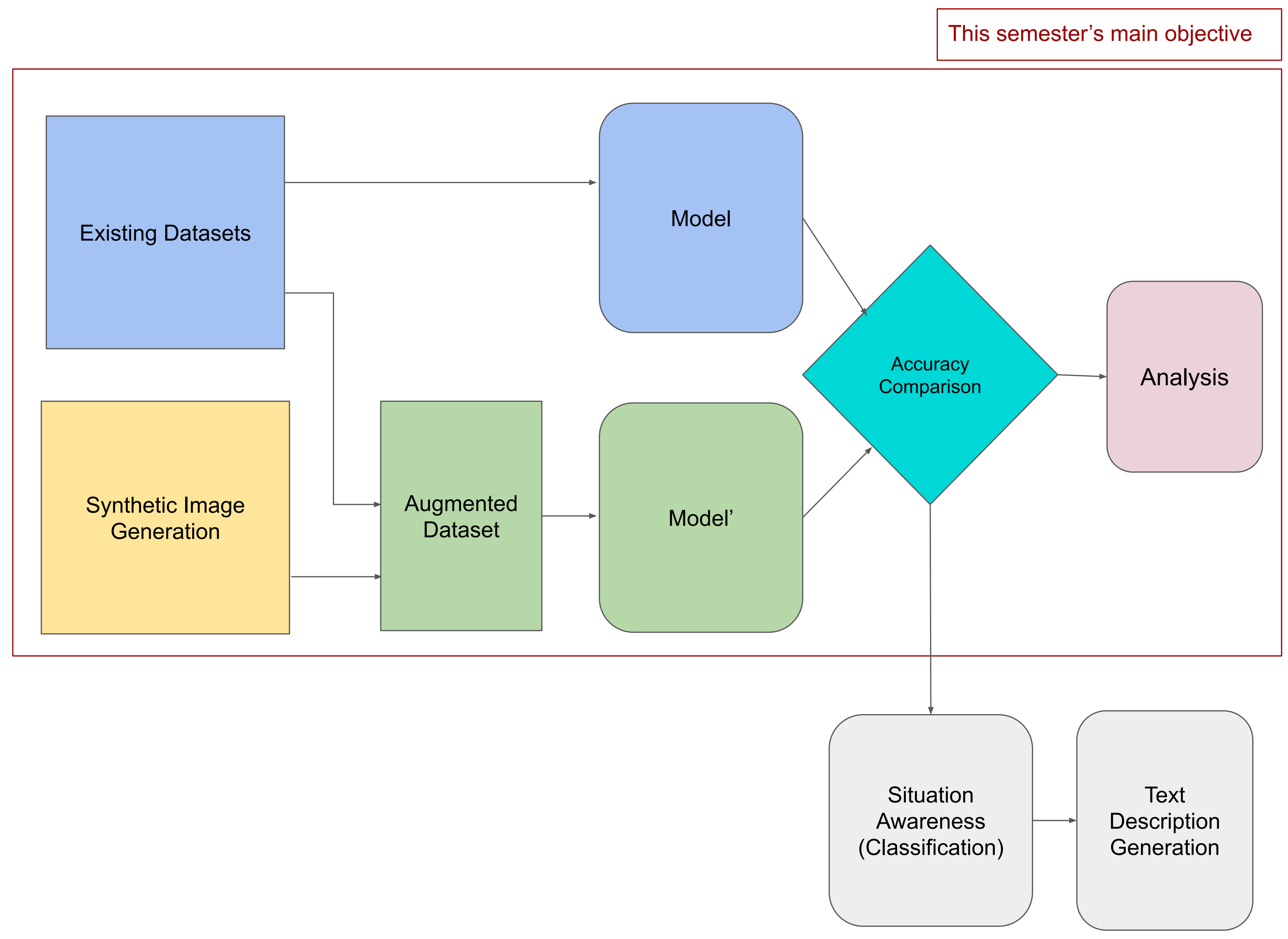
Dataset	Description	Relevant Image Count
South Asia dataset (SAD)	Fire Disaster, Flood Disaster, Infrastructure, Nature Disaster, Human Damage, and Non Damage	240 out of 13720 images
Crisis Image Benchmarks Dataset	Affected, injured, or dead people, Infrastructure, and utility damage	1375 out of 5541 images
Indonesian Disaster Victims 50 (IDV-50)	50 images from the internet and real disaster	50 images
MEDIC dataset	Images sourced from various disaster scenarios, including earthquake, fire, flood, hurricane, and landslide It consists of Crisis Dataset, aidr dataset and asonam dataset	25583 out of 72380 images
Disaster Images Dataset	Images of natural disasters like Cyclone/Hurricane, Earthquake , Flood, Wildfire	1122 out of 4428 images

Methods for various tasks

Purpose	Models/Tools Used	Observations
Generate Synthetic Images	Dalle 2, Stable Diffusion	Dalle 2: ● Generation time : 10 seconds ● Issues: ○ Does not generate images for injured people due content Policy.
Image Annotation	VoTT, LabelImg, CVAT	VoTT: ● manual graphical annotation tool developed by microsoft ● annotation data exported to JSON includes: ○ labels ○ height ○ coordinates, etc
Object Detection	Yolov8, FCOS,TTFNet, PAFNet	YOLO : ● Pretrained model : ○ Pretrained on COCO dataset. ○ Issues: ■ Difficult to detect survivors who are laying flat on ground or obscured by dust. ■ Struggles to detect individuals from aerial or d perspectives. ○ Solution: Finetune Model with the current dataset.
Text Generation	CLIP-Interrogator-2, pix2struct, Mini GPT-4	CLIP-interrogator-2 ● pretrained on ViT-H-14 OpenCLIP model to use to create simi image in stable diffusion model 2.0 ● text generation time: ○ quality = good: 44.56 s ○ quality = poor: 17.94 s ● issues: ○ depends on quality or size of image (poorer quality → faste output times) ● example output: ○ a pile of debris sitting on top of a beach next to a body of water, wood pier and houses, wall, beaches, fallen column

Additional Work

- Once we confirm that training dataset augmented with synthetic images improves accuracy in object detection accuracy, as time allows, we want to also train a classification model that predicts the situations in images
- Using the detected critical objects and situation classification information, we want to develop an application that generates text descriptions and quantify the severity of situations.
- Such an application can be utilized by first-responders to help analyzing the situation and making decisions
- Similar applications could then be implemented for other disasters including floods, hurricanes, etc.


This semester's main objective
Existing Datasets
Model
Accuracy Comparison
Analysis
Synthetic Image Generation
Augmented Dataset
Model'
Situation Awareness (Classification)
Text Description Generation