# Food Recognition & Nutrition Estimator

Authors

Ahmed Umer Farooq

Mubashir Hussain

Muhammad Hamza Khan

Supervisor: Dr. Ijazullah

FYP-1 Mid Term Report submitted in partial fulfillment of the requirements for the Degree of BS Data Science (Hons.)

INSTITUTE OF MANAGEMENT SCIENCES, PESHAWAR

PAKISTAN

Session: 2021-2025

# Certificate of Approval

I certify that I have examined the FYP-1 Mid Term Report titled: **Food Recognition & Nutrition Estimator**, by Ahmed Umer Farooq, Mubashir Hussain, and Muhammad Hamza Khan, and in my judgment, this work fulfills the criteria for approving the mid-term report submitted in partial fulfillment of the requirements for BS Data Science (Hons.) at Institute of Management Sciences, Peshawar.

Supervisor: Dr. Ijazullah
Assistant Professor
Signature: ____________________

Coordinator BS DS: Dr. Bahar Ali
Assistant Professor
Signature: ____________________

Coordinator FYP: Mr. Omar Bin Samin
Lecturer
Signature: ____________________

# Declaration

We, Ahmed Umer Farooq, Mubashir Hussain, and Muhammad Hamza Khan, hereby declare that the FYP-1 Mid Term Report titled: **Food Recognition & Nutrition Estimator** submitted to FYP Coordinator and R&DD is our own original work. We acknowledge that should our work be identified as plagiarized or fraudulent, the FYP Coordinator and R&DD reserve the complete authority to invalidate our Final Year Project, and we shall be subject to disciplinary action.

Ahmed Umer Farooq
BS Data Science (Hons.)
Session: 2021-2025

Mubashir Hussain
BS Data Science (Hons.)
Session: 2021-2025

Muhammad Hamza Khan
BS Data Science (Hons.)
Session: 2021-2025

# Dedication

We dedicate this Final Year Project to our parents and teachers, who have consistently provided support and assistance in every facet of our lives.

# Acknowledgement

All praise belongs to Allah, who endowed humanity with intelligence, knowledge, and wisdom. Peace and blessings of Allah be upon the Holy Prophet, who encouraged his followers to pursue knowledge from the cradle to the grave.

First and foremost, we extend our sincere gratitude to our supervisor, Dr. Ijazullah, for his unwavering support, patience, motivation, enthusiasm, and extensive knowledge. His guidance has been instrumental throughout this project. Finally, we express our heartfelt thanks to our parents for their moral and spiritual support throughout our academic journey.

# Abstract

Monitoring dietary intake remains a persistent challenge due to the labor-intensive nature of manual food logging and systematic underreporting issues. This project develops an intelligent solution leveraging computer vision and deep learning to automate food identification and nutritional analysis using Convolutional Neural Networks through transfer learning with EfficientNet and ResNet architectures trained on the Nutrition5k dataset.

The system accepts food photographs as input and generates comprehensive nutritional profiles. The Nutrition5k dataset provides laboratory-measured nutritional ground truth, enabling direct training of models that predict both food categories and nutritional values. This mid-term report documents progress through conceptualization, literature review, and methodological framework development including dataset preparation and preprocessing pipelines.

# Contents

# List of Figures

# Chapter 1
# Introduction

## 1.1 Overview

The escalating prevalence of diet-related chronic conditions represents a significant public health concern. Manual food logging systems demand considerable temporal investment and cognitive effort, frequently culminating in premature abandonment. Research demonstrates that individuals systematically underreport caloric consumption by twenty to fifty percent on average, compromising the reliability of self-reported dietary data.

Recent technological advancements in artificial intelligence and computer vision have created opportunities for revolutionizing dietary monitoring. Deep learning methodologies, particularly Convolutional Neural Networks, have demonstrated remarkable capabilities in image classification challenges. The application of these techniques to food recognition represents a promising avenue for automating dietary assessment while eliminating the burden of manual logging.

This research develops an intelligent system capable of automatically identifying food items from photographs while estimating their nutritional composition. Through transfer learning with CNN architectures, our system provides instantaneous nutritional intelligence by photographing meals, transforming dietary monitoring into an effortless experience.

## 1.2 Project Motivation

### *1.2.1 Inadequacy of manual tracking methods*

Traditional dietary monitoring requires individuals to meticulously record meals, approximate portions, and manually search nutritional databases. This methodology proves temporally intensive and imposes cognitive loads creating adherence barriers. Research demonstrates that most individuals abandon manual tracking within weeks due to excessive effort requirements, fundamentally undermining potential health benefits from sustained self-monitoring.

### *1.2.2 Prevalence of reporting inaccuracies*

Even among motivated individuals maintaining consistent food diaries, substantial inaccuracies pervade recorded data. Nutritional research indicates caloric intake experiences systematic underestimation averaging twenty to fifty percent. These errors arise from difficulties estimating portion sizes, incomplete nutritional literacy, potential misreporting driven by social desirability bias, and memory lapses when recording retrospectively.

### *1.2.3 Growing health consciousness*

Public awareness regarding diet-health relationships has expanded considerably. Growing numbers of individuals express interest in understanding nutritional intake patterns and making evidence-informed dietary selections. However, this heightened awareness encounters practical obstacles in traditional tracking methodologies, creating a gap between health intentions and actual behaviors.

### *1.2.4 Enabling technologies*

Technological developments have converged to make automated food recognition achievable. Smartphones with high-resolution cameras have achieved ubiquitous adoption. Advances in deep learning have yielded powerful computational models demonstrating accurate visual recognition. Cloud computing infrastructure enables deployment of demanding models accessible from resource-constrained devices.

### *1.2.5 Potential for behavioral change*

By reducing friction in tracking procedures, automated systems create opportunities for sustainable behavioral modifications. When nutritional feedback becomes instantaneous and effortless, individuals can make informed real-time decisions rather than retrospectively analyzing incomplete records. This temporal immediacy may prove substantially more effective for promoting healthier eating patterns.

## 1.3 Project Vision, Scope and Glossary

### *1.3.1 Project vision*

Our vision encompasses creating an intelligent, universally accessible food recognition platform transforming how individuals monitor dietary intake. We envision nutritional awareness becoming as straightforward as capturing photographs, eliminating barriers preventing consistent tracking.

The system targets multiple user populations: individuals pursuing weight management, patients managing diet-sensitive medical conditions, fitness enthusiasts optimizing nutritional intake, and anyone interested in understanding eating patterns. By delivering accurate, immediate nutritional information without requiring specialized knowledge, we aim to democratize access to dietary insights.

### *1.3.2 Scope in*

The following functionalities constitute core scope:

**Food Recognition Model Development:** Complete CNN-based development pipeline utilizing Nutrition5k dataset including data acquisition, preprocessing, transfer learning implementation using EfficientNet or ResNet, model training and validation, and performance optimization.

**Nutritional Estimation Integration:** Leveraging Nutrition5k's laboratory-measured nutritional ground truth to train models predicting nutritional content directly from images rather than database lookups.

**User Interface Development:** User-friendly Streamlit interface enabling straightforward interaction, supporting image upload and immediate feedback displaying identified foods with nutritional content.

**Performance Evaluation Framework:** Comprehensive evaluation metrics assessing classification accuracy, precision, recall, F1-scores, confusion matrices, and nutritional estimation accuracy against laboratory measurements.

### *1.3.3 Scope out*

The following functionalities are excluded:

**Portion Size Estimation Beyond Dataset:** While Nutrition5k includes portion variation, explicit estimation from arbitrary user images remains outside scope, requiring additional depth estimation techniques.

**Complex Multi-Ingredient Meals:** The dataset focuses on individual dishes rather than complex composite meals with multiple distinct ingredients mixed together.

**Real-Time Video Processing:** The system operates on static images only, not continuous video streams.

**Personalized Nutritional Recommendations:** While providing objective nutritional information, the system will not offer personalized dietary advice or customized meal planning.

### *1.3.4 Glossary*

**CNN:** Convolutional Neural Network - deep learning architecture for image processing
**Transfer Learning:** Technique using pre-trained models as starting points
**EfficientNet:** Modern CNN architecture optimizing accuracy and efficiency
**ResNet:** Residual Network architecture using skip connections
**Nutrition5k:** Dataset pairing food images with laboratory-measured nutritional values
**Streamlit:** Python framework for creating web applications

## 1.4 Objectives

### *1.4.1 Primary objectives*

**Develop Accurate Food Classification System:** Construct and train CNN-based models capable of accurately classifying food items from photographs using Nutrition5k dataset. This encompasses architecture selection, transfer learning implementation, training optimization, and comprehensive validation.

**Implement Nutritional Prediction System:** Leverage Nutrition5k's laboratory-measured values to train models predicting nutritional content directly from images. The system must provide accurate predictions for calories, protein, fat, and carbohydrates, evaluated against ground truth measurements.

**Deliver User-Accessible Interface:** Design and implement intuitive, responsive interface enabling diverse users to interact with the system regardless of technical expertise, supporting image upload, displaying results clearly, and presenting nutritional information comprehensibly.

**Establish Rigorous Performance Evaluation:** Implement comprehensive evaluation incorporating multiple metrics providing nuanced understanding of capabilities and limitations, assessing classification accuracy and nutritional prediction accuracy through MAE and RMSE metrics.

### *1.4.2 Secondary objectives*

**Optimize Computational Efficiency:** While maintaining accuracy targets, systematically optimize models minimizing computational requirements and inference latency, ensuring predictions generate rapidly for satisfactory user experiences.

**Maintain Comprehensive Documentation:** Establish and maintain thorough documentation capturing architectural decisions, experimental results, challenges encountered, and lessons learned.

**Demonstrate Real-World Applicability:** Through systematic testing with diverse image samples and user feedback collection, demonstrate adequate performance under practical deployment scenarios.

## 1.5 Tools

### *1.5.1 LaTeX*

LaTeX serves as the primary tool for generating formal documentation, ensuring strict adherence to institutional formatting guidelines and producing publication-quality documents. The declarative nature separates content creation from formatting concerns.

### *1.5.2 Python*

Python serves as the primary programming language, selected due to its dominant position within machine learning ecosystems, offering extensive library support, readable syntax facilitating collaborative development, and strong community resources.

### *1.5.3 TensorFlow and Keras*

TensorFlow represents a widely adopted deep learning framework. Keras, functioning as TensorFlow's high-level API, provides intuitive interfaces for constructing neural networks. These frameworks will implement CNN architectures, conduct transfer learning, perform GPU-accelerated training, and deploy trained models.

### *1.5.4 NumPy, Pandas, and OpenCV*

NumPy provides fundamental numerical computing infrastructure for array operations. Pandas supplies data manipulation capabilities for managing nutritional measurements and organizing datasets. OpenCV handles image preprocessing operations including resizing, normalization, and augmentation.

### *1.5.5 Matplotlib, Seaborn, and Scikit-learn*

Matplotlib and Seaborn enable creation of informative plots for visualizing training curves, confusion matrices, and performance metrics. Scikit-learn offers machine learning utilities for calculating evaluation metrics and implementing stratified data splits.

### *1.5.6 Streamlit*

Streamlit constitutes a Python library for creating interactive web applications with minimal code, powering the user-facing interface with image upload functionality and nutritional information display.

### *1.5.7 Jupyter Notebook and Google Colab*

Jupyter Notebooks provide interactive computing environments combining code, visualizations, and explanatory text. Google Colaboratory provides free GPU computational resources through cloud-based environments, enabling computationally intensive model training without local hardware investments.

# Chapter 2

# Background Study/ Literature Review

## 2.1 Introduction

The convergence of computer vision and nutritional science has attracted substantial research interest, producing investigations exploring automated food recognition and dietary monitoring systems. This chapter examines relevant prior work, analyzing methodologies, results, limitations, and lessons informing our approach.

The review progresses through traditional dietary assessment methods, evolution of automated food recognition from early computer vision through deep learning revolution, major food recognition datasets with emphasis on Nutrition5k, various nutritional estimation approaches, and persistent challenges limiting practical deployment.

## 2.2 Related Work

### *2.2.1 Traditional dietary assessment methods*

**24-hour dietary recalls**

This assessment method requires trained interviewers guiding participants through structured recollection of foods consumed during the preceding 24-hour period using standardized protocols with multiple passes maximizing completeness.

While benefiting from professional administration and standardized protocols, this approach suffers fundamental limitations. Memory accuracy diminishes as time increases from consumption, introducing systematic recall bias. Substantial costs associated with trained personnel limit scalability. Social desirability bias may influence reporting, with participants potentially underreporting unhealthy foods.

**Food frequency questionnaires**

These instruments ask participants to indicate typical consumption patterns over extended periods for predetermined food lists, selecting from categorical frequency options rather than exact quantities.

Food frequency questionnaires provide useful information about habitual dietary patterns valuable for epidemiological research. However, they lack specificity and precision required for detailed nutritional analysis supporting clinical applications. Reliance on generic frequency categories rather than precise quantities limits nutritional calculation accuracy.

#### Prospective food diaries

This method involves individuals prospectively recording all consumed foods as consumption occurs, typically over periods ranging from days to weeks, requesting information about specific foods, preparation methods, portion sizes, and timing.

Prospective recording theoretically offers the most accurate dietary assessment by eliminating memory-dependent recall. However, empirical research demonstrates that compliance diminishes rapidly over time with recording quality degrading. Portion size estimation remains inaccurate even with detailed instructions. Most problematically, reported intake systematically underestimates actual consumption, with underreporting rates documented between 20-50 percent across diverse populations.

### *2.2.2 Evolution of automated food recognition*

#### Early computer vision approaches

Initial attempts at automated food recognition predated deep learning, relying on traditional computer vision techniques and handcrafted feature extraction designed by domain experts. Research groups explored methods based on color histograms, texture descriptors, shape features, and combinations thereof.

These traditional approaches achieved moderate success on carefully constrained datasets featuring controlled lighting and simplified backgrounds. However, they struggled fundamentally with enormous variability inherent in real-world food imagery. Handcrafted features, while intuitive, proved insufficiently robust handling this variability.

#### Deep learning revolution

The emergence and advancement of deep learning fundamentally transformed food recognition capabilities, introducing methods learning appropriate feature representations directly from data rather than relying on manually engineered features.

**Foundational CNN Architectures:** Early CNN applications to food recognition adapted architectures proven successful in general image classification. AlexNet, introduced by Krizhevsky, Sutskever, and Hinton in 2012, pioneered deep CNN architectures for image recognition. VGGNet demonstrated that network depth significantly impacts performance. GoogLeNet introduced the inception module enabling parallel convolutional operations at multiple scales. When adapted for food recognition through fine-tuning on food-specific datasets, they yielded substantial improvements over traditional methods, achieving accuracies exceeding 70 percent on benchmarks.

**Transfer Learning Paradigm:** A critical realization accelerating progress involved recognizing that training deep neural networks from scratch requires enormous labeled datasets. However, features learned by networks trained on large general image datasets like ImageNet transfer effectively to specialized domains including food recognition. Transfer learning involves initializing network weights from models pre-trained on ImageNet, then fine-tuning on relatively modest food-specific datasets. This approach has become standard practice, enabling strong performance even with limited food-specific training data. Our project adopts this proven methodology.

**Modern Efficient Architectures:** Recent architectural innovations have emphasized achieving better accuracy-efficiency trade-offs. ResNet (Residual Networks) pioneered residual connections enabling training of very deep networks. EfficientNet systematically scales network depth, width, and input resolution using compound scaling methods, achieving state-of-the-art results while maintaining computational efficiency. For our project, we selected EfficientNet as the primary architecture based on its superior accuracy-efficiency trade-off, with ResNet serving as a comparison baseline.

### *2.2.3 Food recognition datasets*

Large-scale, well-annotated datasets have proven crucial for advancing food recognition capabilities, providing standardized benchmarks and sufficient training data for deep learning methods.

#### Food-101 dataset

Food-101 represents one of the most widely adopted benchmarks in food recognition research, containing 101,000 images distributed uniformly across 101 food categories. A distinctive characteristic involves its intentional inclusion of images exhibiting substantial variation in quality, viewpoint, lighting, and presentation to better reflect realistic conditions.

The dataset's advantages include substantial size, well-balanced category distribution, and widespread adoption facilitating comparisons with prior work. However, the predominant focus on Western cuisines limits applicability to diverse global dietary patterns.

#### UNIMIB2016

This dataset specifically targets dietary monitoring applications, containing images captured by users in naturalistic settings using smartphones and wearable cameras. The realistic capture conditions make UNIMIB2016 particularly valuable for assessing practical performance under deployment conditions, though its moderate size limits utility as a primary training dataset.

### Nutrition5k dataset

Nutrition5k represents a groundbreaking contribution by specifically addressing nutritional estimation through pairing food images with detailed nutritional measurements obtained through rigorous laboratory analysis. This dataset distinguishes itself fundamentally from predecessors by providing empirically measured nutritional ground truth rather than relying on database lookups or estimations.

**Dataset Composition and Structure:** The Nutrition5k dataset contains approximately 5,000 unique dish samples, each photographed from multiple standardized viewpoints to capture comprehensive visual information. Crucially, each dish was physically prepared, photographed, and then subjected to laboratory nutritional analysis following established protocols. This rigorous methodology ensures that nutritional values represent actual measured quantities rather than theoretical database values.

Each sample includes RGB images captured from multiple angles, depth information from specialized sensors, and comprehensive nutritional measurements including calories, protein, fat, carbohydrates, and additional micronutrients determined through laboratory analysis. The multi-view imaging approach enables learning robust visual representations invariant to viewpoint variations.

**Advantages for Nutritional Estimation:** The laboratory-measured nutritional values provide genuine ground truth unavailable in most datasets, enabling assessment of nutritional estimation accuracy beyond simple database lookups. This direct empirical grounding represents a significant methodological advancement. Models trained on Nutrition5k can learn direct statistical relationships between visual food appearance and nutritional content, potentially capturing subtle visual cues correlating with nutritional composition that database approaches miss.

The dataset's inclusion of portion size variation within training samples enables models to learn relationships between apparent food quantity and nutritional values, providing a foundation for portion-aware nutritional prediction without requiring explicit portion size estimation modules.

**Justification for Dataset Selection:** For our project, Nutrition5k's unique pairing of images with measured nutritional values provides critical capabilities unavailable with other datasets. The ability to train models predicting nutritional content directly from images, validated against laboratory measurements, represents a substantial methodological advantage. While the dataset size is modest, transfer learning from ImageNet pre-trained models enables leveraging general visual features learned from millions of images, compensating for limited food-specific training data.

### *2.2.4 Nutritional estimation approaches*

While food recognition has received extensive research attention, nutritional estimation presents distinct challenges requiring different methodological approaches.

#### Database lookup methods

The conceptually simplest approach maps recognized food categories directly to authoritative nutritional databases. Upon successfully classifying an image, the system retrieves standardized nutritional information. This method offers advantages including simplicity of implementation and leverage of existing comprehensive databases.

However, database lookup approaches cannot account for preparation method variations that substantially alter nutritional content. Without portion size information, nutritional estimates must assume standard serving sizes that may differ substantially from actual consumption.

#### Portion size estimation

More sophisticated systems attempt estimating portion sizes from images, enabling personalized nutritional calculations. Depth estimation techniques use multiple images or depth sensors to reconstruct three-dimensional food geometry. Reference object methods use objects of known size for scale calibration. Segmentation-based approaches attempt to estimate volume through learned relationships between two-dimensional appearance and three-dimensional structure.

Current state-of-the-art portion estimation methods achieve errors around 20-30 percent, which directly propagates to nutritional estimates. The technical complexity and limited accuracy present significant challenges.

#### Direct nutritional prediction

Recent research has explored end-to-end approaches wherein neural networks directly predict nutritional values from images without explicit food classification as an intermediate step. These models learn statistical relationships between visual appearance patterns and nutritional content during training on datasets pairing images with measured nutritional values.

The Nutrition5k dataset specifically enables this approach through its laboratory-measured nutritional ground truth. Rather than requiring intermediate food classification followed by database lookup, models can learn direct mappings from images to nutritional values. This end-to-end learning potentially captures subtle visual cues correlating with nutritional content that categorical approaches miss.

### Hybrid approaches

Our project adopts a hybrid strategy combining food classification for interpretability with direct nutritional prediction leveraging Nutrition5k's measured values. This approach provides users with both categorical food identification (supporting transparency and trust) and accurate nutritional predictions (leveraging direct learning from measured data).

## 2.3 Limitations

Despite substantial progress, several significant challenges persist in automated food recognition and nutritional estimation.

### *2.3.1 Visual variability*

Food appearances vary dramatically based on cooking methods, cultural preparations, regional variations, and individual preferences. A single food category encompasses countless visual presentations. This enormous intra-class variability complicates classification, as models must learn to recognize items as belonging to the same category despite substantial visual differences.

### *2.3.2 Inter-class similarity*

Conversely, different food categories often appear visually quite similar, creating confusion. Distinguishing between different pasta types or rice varieties can challenge even human observers based solely on visual information. Certain foods prove visually indistinguishable despite having substantially different nutritional profiles.

### *2.3.3 Cultural and regional diversity*

Most existing datasets exhibit pronounced biases toward Western cuisines, limiting generalizability to diverse cultural contexts. Asian, African, Middle Eastern, and Latin American cuisines remain substantially underrepresented, creating systematic performance gaps for these food types.

### *2.3.4 Multi-ingredient meals*

Real-world eating patterns frequently involve meals consisting of multiple distinct components mixed together. Current systems generally perform poorly on such complex dishes. Approaches attempting ingredient segmentation require sophisticated image understanding capabilities beyond current reliable performance.

### *2.3.5 Portion size estimation challenges*

Accurate portion size estimation from single images remains challenging without depth information or reference objects. The Nutrition5k dataset partially addresses this through its laboratory-measured nutritional values corresponding to photographed portions, but generalizing to arbitrary user images with varied portion sizes requires additional techniques.

### *2.3.6 Limited dataset sizes*

Compared to general image recognition datasets containing millions of images, food-specific datasets remain relatively modest in size. This limitation affects the ability of models to learn highly generalizable visual representations, though transfer learning partially mitigates this challenge.

### *2.3.7 Justification for our approach*

Given the current state of research, our project makes deliberate methodological decisions:

**Nutrition5k Dataset Selection:** The dataset's laboratory-measured nutritional values provide unique capabilities for training models that predict nutritional content directly from images. This represents a methodological advancement over database lookup approaches.

**Transfer Learning with Modern Architectures:** Leveraging pre-trained EfficientNet and ResNet models represents current best practices, enabling strong performance with available computational resources and the Nutrition5k dataset.

**Hybrid Classification and Prediction:** Combining food classification for interpretability with direct nutritional prediction leverages the dataset's strengths while maintaining user transparency.

**Streamlit-Based Deployment:** Prioritizing accessibility through web-based interfaces ensures broader usability while avoiding mobile deployment complexities in initial implementation.

These decisions position our project to deliver practical, reliable results while establishing foundations for future enhancements addressing identified limitations in current approaches.

# Chapter 3

# Methodology

## 3.1 Flow Chart

Figure 3.1 illustrates the complete system workflow from data preparation through model deployment.

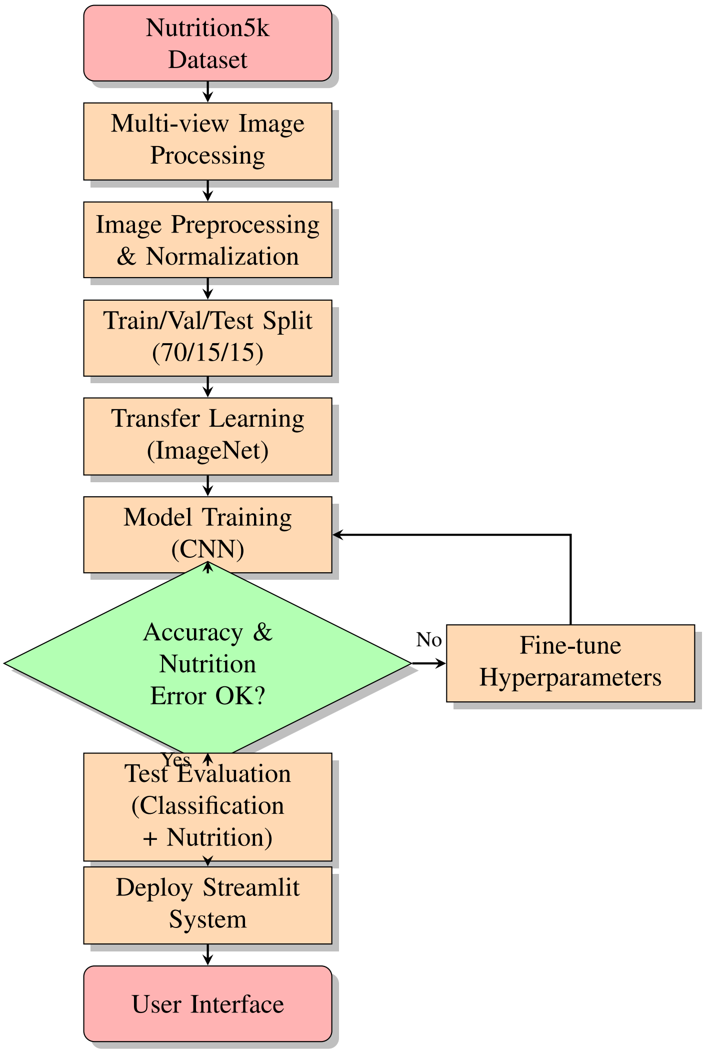
Nutrition5k Dataset
Multi-view Image Processing
Image Preprocessing & Normalization
Train/Val/Test Split (70/15/15)
Transfer Learning (ImageNet)
Model Training (CNN)
Accuracy & Nutrition Error OK?
No
Fine-tune Hyperparameters
Yes
Test Evaluation (Classification + Nutrition)
Deploy Streamlit System
User Interface

**Figure 3.1:** System development workflow from dataset preparation to deployment

## 3.2 Dataset

### *3.2.1 Dataset requirements*

Several criteria guided our dataset selection process. The dataset must contain substantial numbers of images to enable effective deep learning model training. Most critically for nutritional estimation, the dataset must provide accurate nutritional ground truth enabling validation of predicted values. Laboratory-measured nutritional values represent a significant advantage over database-derived estimates.

### *3.2.2 Nutrition5k dataset selection*

After systematically evaluating available options, we selected Nutrition5k as our primary dataset. The dataset contains approximately 5,000 unique dish samples, each photographed from multiple viewpoints and subjected to comprehensive laboratory nutritional analysis.

Critical advantages of Nutrition5k include laboratory-measured nutritional ground truth unavailable in most datasets, multi-view imaging enabling robust visual feature learning, explicit pairing of images with measured calories, protein, fat, and carbohydrates, and inclusion of portion size variation within training samples.

While the dataset size is more modest than Food-101, transfer learning from ImageNet pre-trained models enables leveraging general visual features learned from millions of images, compensating for limited food-specific training data.

### *3.2.3 Dataset structure*

The Nutrition5k dataset organizes samples with the following structure:

**Image Data:**

- RGB images from multiple viewpoints per dish
- Depth information from specialized sensors
- Standardized resolution and color space
- Multiple images per unique dish sample (typically 3-5 views)

**Nutritional Measurements:**

- Calories (kcal) - laboratory measured
- Protein (g) - laboratory measured
- Fat (g) - laboratory measured
- Carbohydrates (g) - laboratory measured
- Additional micronutrients

### *3.2.4 Dataset statistics*

- Total Unique Dishes: 5,000
- Images per Dish: Multiple viewpoints (typically 3-5)
- Total Images: 20,000-25,000
- Image Resolution: Variable, standardized during preprocessing
- Color Space: RGB
- Depth Information: Available

- Nutritional Measurements: Laboratory-analyzed

## 3.3 Features Description

Our CNN-based approach automatically learns hierarchical visual features from raw pixel data rather than relying on manually engineered features.

### *3.3.1 Low-level features*

The initial convolutional layers of the network automatically learn to detect fundamental visual elements:

- Edge detection in various orientations (horizontal, vertical, diagonal)
- Color blobs and gradients across different regions
- Basic texture patterns (smooth, rough, granular)
- Simple geometric shapes (circles, rectangles, curves)

### *3.3.2 Mid-level features*

Intermediate layers combine low-level features into more complex representations:

- Food-specific textures (crispy, creamy, fibrous)
- Color combinations characteristic of particular foods
- Structural patterns indicating cooking methods
- Surface properties (glossy, matte, wet, dry)

### *3.3.3 High-level features*

Deeper layers learn abstract representations enabling food recognition and nutritional prediction:

- Overall food shape and form recognition
- Portion size visual cues and quantity indicators
- Plating and presentation patterns
- Semantic food identity representations correlating with nutritional content

## 3.4 Dataset Preprocessing

Raw images require systematic preprocessing before serving as suitable model inputs. Our preprocessing pipeline ensures consistency and compatibility with pre-trained model requirements.

### *3.4.1 Image resizing*

Neural networks require fixed-size inputs. We systematically resize all images to 224×224 pixels, appropriate for both EfficientNet and ResNet architectures. Resizing preserves aspect ratio by first rescaling images ensuring the shorter dimension matches target size, then cropping the center region to achieve exact dimensions. Bicubic interpolation is employed during resizing to maintain image quality.

### *3.4.2 Pixel normalization*

Pixel values are normalized using ImageNet dataset statistics to align with preprocessing used during pre-training of transfer learning base models. The normalization uses mean values $\mu_R = 0.485$, $\mu_G = 0.456$, $\mu_B = 0.406$ and standard deviations $\sigma_R = 0.229$, $\sigma_G = 0.224$, $\sigma_B = 0.225$ for red, green, and blue channels respectively. Each color channel undergoes independent normalization according to:

$$x_{normalized} = \frac{x - \mu}{\sigma} \tag{3.1}$$

This normalization ensures that input images have similar statistical properties to the ImageNet dataset on which the base models were originally trained.

### *3.4.3 Data augmentation*

To improve model robustness and prevent overfitting, we apply data augmentation techniques during training. Our augmentation pipeline includes:

- **Random Horizontal Flipping:** 50% probability of flipping images horizontally
- **Random Rotation:** ±15 degrees rotation range
- **Random Brightness:** ±20% brightness adjustment range
- **Random Contrast:** ±20% contrast adjustment range
- **Random Saturation:** ±30% color saturation adjustment range

These augmentations apply probabilistically during training only, not during validation or testing phases, to simulate realistic variations in food photography conditions.

### *3.4.4 Complete preprocessing pipeline*

The complete preprocessing pipeline executes the following steps sequentially:

1. Image loading from Nutrition5k dataset
2. Aspect ratio preservation (shorter side rescaled to 224 pixels)
3. Center cropping to extract 224×224 region

4. Array conversion to NumPy format (224, 224, 3)
5. Normalization with ImageNet statistics
6. Data augmentation (training only)
7. Tensor conversion to TensorFlow format
8. Batch formation (32 images per batch)

## 3.5 Training, Validation and Test Dataset Split

The Nutrition5k dataset undergoes systematic partitioning into three distinct subsets supporting different phases of model development and evaluation.

### *3.5.1 Split rationale and configuration*

**Training Set (70%):** Comprising approximately 70 percent of available data, the training set serves for model parameter optimization. The model learns visual features and relationships between food appearance and nutritional content through this data. Data augmentation is applied during training to improve generalization.

**Validation Set (15%):** Containing approximately 15 percent of data, the validation set serves to tune hyperparameters and monitor training progress. This set guides decisions about learning rates, network architectures, and early stopping to prevent overfitting. Data augmentation is not applied to validation data.

**Test Set (15%):** Consisting of the remaining approximately 15 percent, the test set is held out completely until final evaluation. This provides an unbiased assessment of model performance on unseen data. Data augmentation is not applied to test data.

### *3.5.2 Stratification strategy*

Data splitting is performed with careful consideration of dataset structure to ensure fair evaluation:

- Unique dishes kept within same split to prevent data leakage where the model could memorize specific dishes appearing in both training and test sets
- Balanced representation across nutritional value ranges ensuring all splits contain diverse nutritional profiles
- Random shuffling prevents temporal biases or systematic ordering effects
- Multiple viewpoints of same dish remain together in the same split

This stratification ensures that model evaluation accurately reflects performance on truly unseen food items rather than merely unseen viewpoints of memorized dishes.

## 3.6 Libraries

This section describes the software libraries utilized throughout the project development lifecycle.

### *3.6.1 Core deep learning libraries*

**TensorFlow 2.x** (Version: 2.13+): Primary deep learning framework for model implementation, training, and deployment. Provides GPU acceleration and comprehensive neural network operations.

**Keras API** (Integration: tf.keras): High-level API for building and training neural networks with intuitive interfaces while maintaining access to TensorFlow's backend capabilities.

### *3.6.2 Data processing libraries*

**NumPy** (Version: 1.24+): Fundamental library for numerical computing, handling array operations and mathematical computations throughout the pipeline.

**Pandas** (Version: 2.0+): Data manipulation library for managing nutritional measurements, organizing datasets with metadata, and processing experimental results.

**OpenCV** (Version: 4.8+): Computer vision library for image preprocessing operations including resizing, color space conversions, and augmentation implementations.

### *3.6.3 Visualization libraries*

**Matplotlib** (Version: 3.7+): Primary plotting library for creating visualizations including training curves, performance metrics, and exploratory data analysis.

**Seaborn** (Version: 0.12+): Statistical visualization library built on Matplotlib, providing enhanced aesthetics and specialized plots.

### *3.6.4 Evaluation libraries*

**Scikit-learn** (Version: 1.3+): Machine learning library providing utilities for calculating precision, recall, F1-scores, confusion matrices, and implementing stratified data splits.

### *3.6.5 Deployment libraries*

**Streamlit** (Version: 1.28+): Web application framework for creating the user interface, enabling image upload, displaying predictions, and presenting nutritional information accessibly.

## 3.7 Hardware Description

### *3.7.1 Development environment*

**Google Colab Pro:** Our primary development platform providing:

- Platform: Cloud-based Jupyter notebook environment
- GPU Access: Tesla T4, P100, or V100 depending on availability
- GPU Memory: 16GB
- System RAM: Up to 25GB
- Session Duration: Up to 24 hours continuous
- Storage: Integration with Google Drive for dataset and checkpoint storage

Google Colab eliminates the need for expensive local hardware investments while providing powerful GPU resources essential for training deep learning models efficiently.

### *3.7.2 GPU specifications*

**NVIDIA Tesla T4:**

- CUDA Cores: 2,560
- Memory: 16GB GDDR6
- Architecture: Turing
- Typical Training Speed: 4-6 hours for full training cycle

**NVIDIA Tesla P100:**

- CUDA Cores: 3,584
- Memory: 16GB HBM2
- Architecture: Pascal
- Typical Training Speed: 3-5 hours for full training cycle

**NVIDIA Tesla V100:**

- CUDA Cores: 5,120
- Memory: 16GB HBM2
- Architecture: Volta
- Typical Training Speed: 2-4 hours for full training cycle

The availability of these GPU options through Google Colab enables efficient experimentation and model training without requiring investment in dedicated hardware infrastructure.

# Bibliography

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[3] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114.

[4] T. Thames, D. Wolff, P. Joshi, and S. Palaniappan, "Nutrition5k: Towards automatic nutritional understanding of generic food," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 8903–8911.

[5] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101 – Mining discriminative components with random forests," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 446–461.

[6] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, and Y. Ma, "DeepFood: Deep learning-based food image recognition for computer-aided dietary assessment," in *International Conference on Smart Homes and Health Telematics*, 2016, pp. 37–48.

[7] F. E. Thompson, A. F. Subar, C. M. Loria, J. L. Reedy, and T. Baranowski, "Need for technological innovation in dietary assessment," *Journal of the American Dietetic Association*, vol. 110, no. 1, pp. 48–51, 2010.

[8] S. A. Poslusna, J. Ruprich, J. H. de Vries, M. Jakubikova, and P. van't Veer, "Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice," *British Journal of Nutrition*, vol. 101, no. S2, pp. S73–S85, 2009.

[9] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 3320–3328.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[13] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[14] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Food image analysis: Segmentation, identification and weight estimation," in *IEEE International Conference on Multimedia and Expo*, 2013, pp. 1–6.

[15] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, "A survey on food computing," *ACM Computing Surveys*, vol. 52, no. 5, pp. 1–36, 2019.