Meal Calorie Prediction

Akanksha Sachin Shah
Texas A&M University
College Station, Texas, USA
shahakanksha@tamu.edu

Gourangi Sanjay Taware
Texas A&M University
College Station, Texas, USA
gourangitaware@tamu.edu

Riddhi Prakash Ghate
Texas A&M University
College Station, Texas, USA
riddhighate.07@tamu.edu

***Abstract*—Estimating caloric intake is a cornerstone of effective dietary monitoring and personalized health management, particularly for chronic conditions like diabetes. This study proposes a multimodal machine learning framework that integrates diverse data sources—continuous glucose monitoring (CGM) readings, microbiome data, demographic information, and meal images—to predict calorie content. Leveraging data from over 40 participants collected across 10 days, the approach captures the complexity of individual dietary and metabolic patterns.**

The framework combines multiple specialized models: a ResNet-50-based convolutional neural network (CNN) for analyzing meal images, a Long Short-Term Memory (LSTM) network to model glucose patterns over time, and a fully connected network for processing tabular demographic and microbial features. These components are fused into a unified architecture that synthesizes inputs from all modalities. Training is conducted in two stages: first, tuning the image analysis module and, subsequently, fine-tuning the integrated model for enhanced prediction accuracy. The results demonstrate significant improvements in calorie estimation by leveraging the complementary strengths of multimodal data.

This research highlights the potential of combining physiological, visual, and demographic data to overcome the limitations of single-modality systems. The findings underscore the value of such frameworks for advancing personalized nutrition and improving dietary assessment tools for real-world health applications.

***Index Terms*—Calorie Estimation, Multimodal Machine Learning, Continuous Glucose Monitoring (CGM), ResNet-50, Long Short-Term Memory (LSTM), Personalized Nutrition, Dietary Monitoring, Macronutrient Prediction, Microbiome Data, Machine Learning in Healthcare.**

I. Introduction

Monitoring caloric intake accurately is a fundamental aspect of promoting overall health and managing chronic conditions such as diabetes and obesity. Traditional dietary tracking methods, like self-reported food diaries, often suffer from significant inaccuracies due to human error, incomplete information, and user fatigue. These shortcomings emphasize the need for automated, reliable, and scalable solutions for estimating dietary intake, particularly in practical, everyday settings.

For individuals with diabetes, precise dietary monitoring is especially critical. Blood glucose levels are closely tied to meal composition, including macronutrient distribution, portion size, and meal timing. Effective glycemic control, crucial for preventing complications, requires detailed insights into how meals influence metabolic responses. However, many existing tools, such as continuous glucose monitoring (CGM) systems or image-based food recognition technologies, often operate in isolation and fail to capture the multifaceted nature of dietary behaviors and their physiological effects.

This research introduces a multimodal machine learning framework to address these limitations by integrating CGM data, microbiome profiles, demographic attributes, and meal images. The strength of a multimodal approach lies in its ability to synthesize complementary data types. For instance, CGM data reflects the dynamic glucose responses to meals, microbiome information provides insights into individual metabolic variability, and food images offer visual details about meal composition. By combining these data sources, the proposed framework delivers a comprehensive and individualized analysis of dietary intake, leading to more accurate predictions.

The primary goal of this study is to develop a robust predictive model for estimating lunch calories, using advanced machine learning techniques. The framework utilizes a ResNet-50-based convolutional neural network (CNN) to extract image features, a Long Short-Term Memory (LSTM) network to analyze temporal glucose trends, and fully connected layers to process tabular demographic and microbial data. This integrated system represents a significant step forward in dietary monitoring, addressing the limitations of single-modality approaches and providing a personalized, data-driven solution for nutrition management.

By leveraging multimodal data, this study contributes to the advancement of dietary monitoring technologies, paving the way for practical applications in personalized nutrition, chronic disease management, and healthcare. This approach offers scalable and user-friendly tools that address individual variability, making them invaluable in improving health outcomes.

II. Related Work

A. Improving Macronutrient Estimation with Multi-Modal and Biomarker Approaches

Recent advancements in dietary monitoring have explored diverse approaches to improve the accuracy of macronutrient and calorie estimation. Das et al. (2022) investigated the use of dietary biomarkers, including amino acids, triglycerides, insulin, and glucose, to predict the macronutrient composition of mixed meals. Their study demonstrated that incorporating biomarkers significantly enhanced prediction performance compared to using CGM data alone. Using machine learning

models, particularly XGBoost, they achieved normalized root mean squared errors (NRMSE) of 22.9% for carbohydrates, 23.4% for proteins, and 32.3% for fats, compared to 28.9%, 46.4%, and 40.0%, respectively, when using CGM data alone. However, their study relied on invasive blood sampling and controlled liquid meals, limiting generalizability to real-world settings involving solid or mixed meals and dynamic postprandial conditions.

Building on these findings, Zhang et al. (2023) introduced a multi-modal framework that combined CGM data with food image analysis to estimate calorie intake more accurately. This approach addressed key limitations of earlier studies by leveraging non-invasive data collection and advanced machine learning techniques. Zhang et al. used an attention-based Transformer for CGM data and Vision Transformers (ViTs) for analyzing food images, with a late fusion mechanism aggregating embeddings from both modalities. Validated on a dataset of 27 participants consuming both controlled and realistic meals, their model achieved an NRMSE of 0.34 and a correlation of 0.52 for calorie prediction, significantly outperforming models relying solely on CGM data or image data. This framework demonstrated enhanced robustness across varied meal types and better applicability to real-world dietary monitoring scenarios.

While Das et al. effectively highlighted the value of complementary biomarkers for macronutrient prediction, Zhang et al. advanced the field by integrating CGM and food image data in a practical, non-invasive approach. Their work underscores the potential of multi-modal frameworks to provide more accurate and comprehensive dietary insights, marking a significant step forward in automated diet monitoring technologies.

B. Advancements in Diet Monitoring Technologies

Diet monitoring technologies have evolved significantly, enabling automatic tracking of food intake and reducing reliance on manual logging, which can be burdensome and error-prone. Broadly, these technologies can be categorized into wearable sensors, smart devices, and computer vision systems, each contributing unique capabilities to dietary assessment.

Wearable sensors, such as piezoelectric devices and microphones, have been used to detect eating behaviors like chewing and swallowing through acoustic or motion analysis. These approaches allow continuous monitoring of intake patterns but are often limited in their ability to quantify or classify consumed foods. Smart utensils, such as forks and chopsticks equipped with sensors, have been employed to identify food types and measure portions. Although effective in specific settings, these devices can be intrusive and rely heavily on user compliance.

Among these advancements, computer vision technologies have emerged as a transformative tool for diet monitoring. By leveraging advances in image recognition, these systems analyze food photographs to estimate nutritional content, including calorie counts and macronutrient composition. Modern image-based systems utilize deep learning models such as convolutional neural networks (CNNs) and vision transformers (ViTs), which extract features like texture, color, and shape to accurately classify food items and assess portion sizes. These capabilities are increasingly integrated into mobile applications and cloud-based services, such as CalorieMama and FoodAI, offering scalable and user-friendly solutions for dietary assessment.

Recent research focuses on enhancing the accuracy of computer vision systems by integrating complementary data sources, such as contextual information or user-specific health profiles, to refine nutritional estimates. Advances in multimodal learning, where image data is combined with additional inputs like continuous glucose monitor (CGM) data or dietary biomarkers, further enhance the potential of these technologies. These methods not only provide precise nutritional breakdowns but also adapt to individual variations in diet and metabolism, making them valuable for personalized nutrition and chronic disease management.

Such innovations in diet monitoring technology underscore the potential for a future where dietary assessment is not only automated but also tailored to individual needs, offering a powerful tool for improving health outcomes and promoting sustainable dietary behaviors.

III. Methodology

This section offers a detailed overview of the experimental dataset and the methodology used in this study. It outlines the data preprocessing techniques utilized to extract meaningful insights from postprandial responses. Additionally, we describe the implementation model, detailing its architecture, feature extraction methods, and the processes for training and validation. We developed modality-specific models that contained three feature extractors that extract image embeddings from the food images, CGM embeddings from the CGM data, and tabular embeddings from the viome and demographic data. We then concatenated all these embeddings and passed them through a fully connected network to generate calorie estimates.

A. Dataset Description

The dataset for this study is sourced from the research titled "Predicting the Macronutrient Composition of Mixed Meals From Dietary Biomarkers in Blood."[2] It includes nutritional information for breakfast and lunch, meal photographs, motion data, demographic details, and postprandial biomarker measurements. Blood biomarkers were collected via venipuncture at specified intervals post-meal, and continuous glucose levels were monitored using CGM devices. This multimodal data collected from 40 participants over 10 days contained 324 samples with 39 total features.

B. Data Preprocessing and Feature Extraction

The data preprocessing steps were crucial for ensuring the quality and consistency of the dataset used for model training. To maintain dataset integrity, missing values were imputed, while invalid or erroneous data, such as outliers and incorrect entries, were removed to prevent potential biases.

Furthermore, all relevant features were standardized to ensure uniformity, with each feature adjusted to have a mean of 0 and a standard deviation of 1. This step was particularly important for ensuring that all features contributed equally, thereby improving the convergence and overall performance of the machine learning models. Additionally, all the features were converted into tensors to ensure compatibility with deep learning frameworks.

1) CGM Data Preprocessing: The Continuous Glucose Monitoring (CGM) data underwent several preprocessing steps to prepare it for model training and analysis. For each participant, glucose readings were extracted and organized into sequences that included the time difference from both breakfast and lunch, along with the corresponding glucose values. By calculating the time difference between breakfast and the time of CGM reading, we can observe how glucose levels rise and fall after the initial meal, which is crucial for understanding the body's response to different macronutrients. Similarly, the time difference from lunch provides insights into how glucose levels evolve after the second meal. This temporal sequence helps capture the dynamic nature of glucose responses and offers valuable features for training machine learning models to predict postprandial effects and macronutrient composition. This sequence is given as a time step input to the LSTM.

Additionally, the hour of meal times (breakfast and lunch) was extracted, providing temporal context for each glucose measurement. The meal interval, representing the time between breakfast and lunch, was also calculated. Furthermore, the sequences were padded to a consistent length to ensure uniform input size for downstream models.

2) Image Preprocessing: For the extraction of image embeddings, we utilized the pre-trained ResNet50 model, which requires input images to be resized to a standard size of 224x224 pixels. This size is commonly used in deep learning models as it provides a balance between computational efficiency and the ability to capture important image features.

Due to the limited amount of available image data, several augmentation techniques were applied to make the model more generalizable and reduce the risk of overfitting. These augmentations included random horizontal and vertical flipping, random rotation up to 45 degrees, and the conversion of images into tensors for compatibility with PyTorch-based models. This augmentation strategy helped to increase the variability of the image data and allowed the model to better generalize to new, unseen images.

C. Data Analysis

To gain deeper insights into the relationship between glucose dynamics, meal composition, and calorie predictions, we analyzed the Continuous Glucose Monitoring (CGM) data across participants. This analysis clearly showed how postprandial glucose patterns correspond to meal timings, nutrient composition, and caloric intake. By examining these trends, we were able to understand the variations in glucose levels and their modulation by individual characteristics such as demographic details and microbial abundance. This analysis highlights the temporal and metabolic relationships that are critical for modeling accurate calorie predictions.

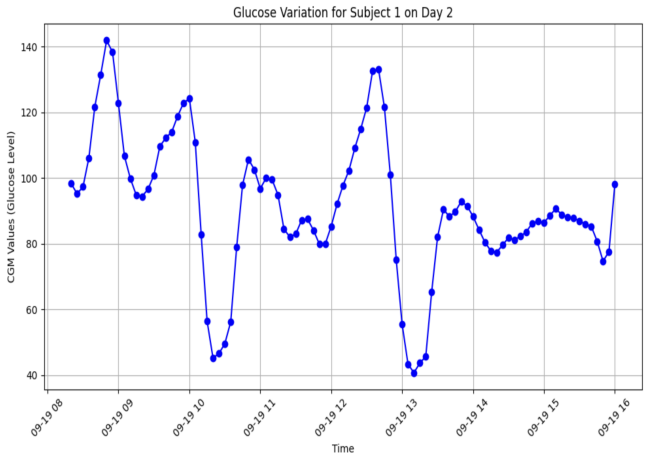

Fig. 1. Glucose Variations

1) Glucose Variation and Meal Timings: The breakfast and lunch times for participants were analyzed to better understand the relationship between meal timing and postprandial glucose responses. For instance, after breakfast, which was typically consumed around 09:00 AM, glucose levels show a sharp increase, followed by a gradual decline. This reflects the body's metabolic response to the carbohydrate content of the meal, as carbohydrates are quickly converted into glucose, causing a rapid increase in blood glucose levels. The gradual decline post-spike represents the body's efforts to return glucose levels to normal.

A similar glucose response is seen after lunch, typically consumed around 12:00 PM. There is a pronounced glucose spike followed by a sustained elevation, which is more prolonged compared to breakfast. This response is influenced by the higher caloric and macronutrient content of the lunch, particularly the increase in carbohydrates, fats, and proteins. The fat content, in particular, may slow down glucose absorption, leading to a delayed and sustained glucose elevation. These variations, as demonstrated in Fig. 1. align closely with the meal timings, which are influenced by the caloric density and macronutrient breakdown of the meals, emphasizing the link between meal composition, timing, and postprandial glucose dynamics.

2) Meal Composition and Glucose Dynamics: The breakfast and lunch meals consumed by participants varied in caloric content and macronutrient composition. Breakfast meals typically contained moderate amounts of carbohydrates, fats, and proteins, with a significant proportion of calories derived from carbohydrates. Lunch meals were often more calorie-dense, with higher amounts of carbohydrates, fats, and proteins. The carbohydrate content in both meals played a significant role in the glucose spikes observed in the CGM data, as carbohydrates are rapidly converted into glucose, causing immediate increases in blood glucose levels after consumption.

The sustained glucose elevation following lunch meals suggests a delayed glucose clearance, which is likely influenced by the combined effect of higher caloric intake and the macronutrient composition, particularly the fat content. Fats, in particular, are known to slow down glucose absorption, contributing to a more prolonged elevation in glucose levels. This relationship highlights the importance of understanding the breakdown of macronutrients in predicting postprandial glucose trends and their contribution to accurate calorie predictions.

3) Analysis of the Demographic Data: The demographic profiles of participants in the study varied, with key attributes such as age, gender, BMI, and diabetes status collected. The dataset included individuals with a wide range of BMI values, encompassing both normal and overweight categories. It also captured individuals with and without diabetes, as indicated by A1C levels below 5.7% for those without diabetes. Baseline fasting glucose and insulin levels were largely within normal ranges, suggesting generally healthy glycemic control and insulin sensitivity across the study population. Lipid profile parameters such as triglycerides, HDL, and LDL were also within healthy ranges, further indicating balanced metabolic states across participants.

4) Viome Data Analysis: The viome data, representing microbial abundance, provided valuable insights into participants' gut microbiota profiles. Gut microbiota plays a significant role in nutrient metabolism and postprandial glucose responses, affecting how calories are processed and absorbed. Including microbial features in the analysis allowed the model to account for individual variations in glucose dynamics and calorie assimilation, improving the model's accuracy in predicting caloric intake. For instance, certain microbial profiles explain why two participants with identical meals exhibit different glucose trends.

5) Correlation Between Different Biomarkers and Demographic Features: Fig.2. illustrates the relationships between various demographic and biomarker variables, providing insights relevant to research on predicting meal calorie intake. Notable positive correlations, such as between BMI and weight (0.87), insulin and HOMA-IR (0.95), and baseline fasting glucose with A1C (0.87), suggest potential interdependencies between body composition, metabolic markers, and glucose regulation. These relationships play a critical role in calorie prediction models by highlighting how biomarkers like insulin resistance or glucose levels are influenced by dietary intake. Conversely, strong negative correlations, such as HDL with triglycerides (-0.48) and VLDL (-0.50), underscore the inverse relationships in lipid metabolism, which might help refine calorie predictions based on lipid profiles. Understanding these interactions is crucial for developing accurate, personalized calorie estimation tools, particularly for individuals managing conditions like diabetes or obesity.

Therefore, the combination of glucose variation, microbial abundance, and demographic features significantly enhances the accuracy of calorie predictions. Glucose variation, particularly the postprandial glucose spikes and recovery patterns, directly reflects how the body processes and metabolizes the calories consumed. These glucose dynamics, influenced by meal composition and macronutrient content, provide crucial temporal data that helps predict meal caloric content. Microbial abundance plays an equally important role, as gut microbiota affects nutrient digestion and glucose regulation, ultimately influencing how calories are absorbed and utilized. Different microbial profiles can lead to varying glucose responses, adding a layer of personalization to calorie predictions. Demographic features, such as age, BMI, and insulin sensitivity, further modulate these dynamics by affecting metabolic rates, insulin resistance, and glucose clearance. Incorporating these features enables the model to account for individual differences, ensuring more precise and tailored calorie estimates. Together, these factors provide a comprehensive understanding of how calories are processed in the body, allowing for more accurate predictions based on both physiological and individual characteristics.

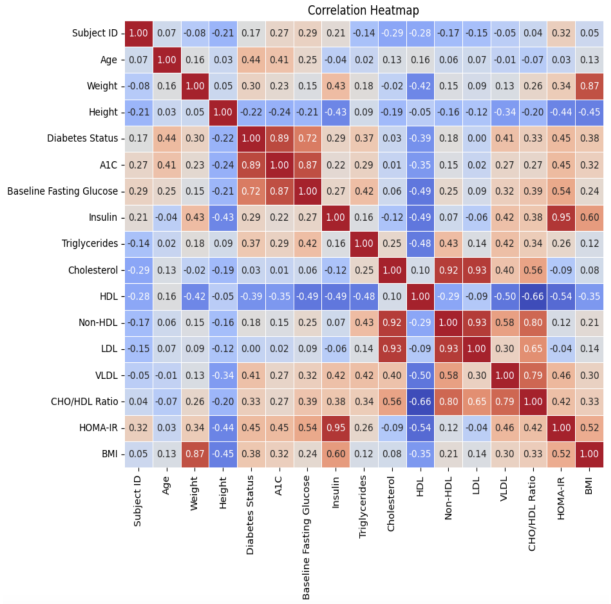

Fig. 2. Correlation Matrix

D. Model Architecture

The model we proposed in our approach combines the heterogeneous data sources including images, continuous glucose monitoring (CGM) data, and tabular demographic data into a unified dataframe. This multi-modal architecture addresses the challenges of dissimilar data types by using customized pipeline for each data source followed by a shared dense network for final prediction. The detailed description of each component is as follows:

- A convolutional neural network (CNN) based on ResNet-50 for processing image inputs.
- A Long Short-Term Memory (LSTM) network for capturing temporal dependencies in CGM data.

- A fully connected (FC) network for transforming metadata into a compact feature representation.

The breakfast and lunch image inputs are processed using a pre-trained ResNet-50 network. ResNet-50 is a deep convolutional neural network with 50 layers trained on the ImageNet dataset. It uses residual connections (skip connections) that bypass one or more layers and address the vanishing gradient problem. The final classification layer of ResNet is replaced with an identity layer to use the network as a feature extractor. We used the pre-trained model to reduce the computational overhead of training and leverage the knowledge from large-scale datasets like ImageNet. The ResNet parameters are initially frozen in the model to allow unfreezing only the last few layers during the training process. The extracted 2048-dimensional feature vectors from the lunch and breakfast image inputs are concatenated and passed through a custom fully connected network. This network refines the combined image features through custom layers of nonlinear transformations, using dropout to further regularize the model. The custom layers are added on top of ResNet-50 layers to transform the 4096 output features(2048 for breakfast and 2048 for lunch) from ResNet into a compact feature representation that adapts to the generic features learned during pre-training specific to breakfast and lunch images.

The CGM data is a time series marked data of glucose levels and it is processed using an LSTM network. The LSTM is chosen for this task due to its ability to retain long-term dependencies while mitigating vanishing gradient issues. The network processes input sequences with three-dimensional features per time step and outputs a hidden representation of a 32-dimensional vector that represents the entire sequence. To summarize the temporal relations, only this 32-dimensional output corresponding to the final time step is used, as it encodes the information learned across the entire sequence.

Along with images and CGM data, demographic information and viome data are used to add more context. The demographic data is in the tabular form and hence they are just passed through a series of fully connected layers. The input dimensionality depends on the number of metadata features, which includes demographic and viome information. The fully connected layers dynamically adapt their sizes ensuring a gradual reduction in dimensionality, in this model it starts with 128 and gradually reduces as it passes through the layers. For example, if configured with three intermediate layers, the first layer reduces the input size to a specified dimension, followed by two additional layers that successively reduce the feature size. Each layer uses ReLU activations to model complex feature interactions and dropout layers to prevent overfitting, especially when the metadata contains sparse or correlated features. This component produces a consolidated metadata representation suitable for integration with features from other modalities, ensuring that all input dimensions are processed without any bias.

The final component of the model integrates features from all three data modalities: image features (256 dimensions), LSTM-processed CGM data (32 dimensions), and metadata features (32 dimensions). These features are concatenated into a single vector, which serves as the input to a series of fully connected layers designed for feature joining and dimensionality reduction. The intermediate layer sizes are again determined dynamically, ensuring a smooth reduction to the final output dimension. This model involves four fully connected layers, each followed by ReLU activations for non-linear transformations and dropout layers for regularization. The last fully connected layer outputs a single scalar value, representing the regression target “Lunch Calories”.

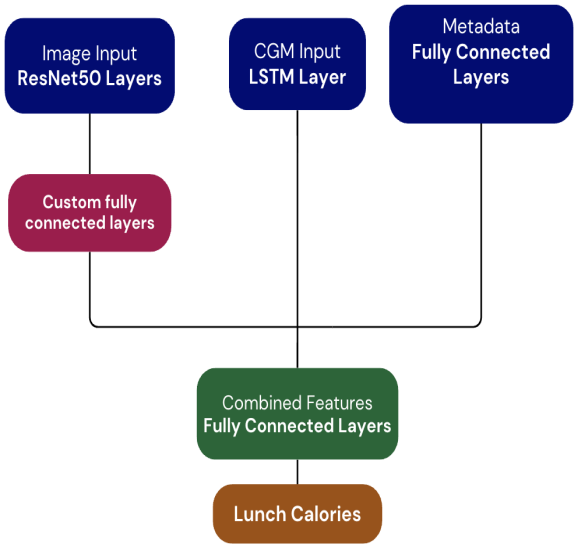

Fig. 3. Model Architecture

E. Meal Calorie Estimation

The training process for the meal calorie estimation model was conducted in two phases to effectively leverage pre-trained features while adapting the model to the specific task of predicting meal calorie content. These phases were designed to balance feature extraction from pre-trained networks with task-specific learning using multimodal data.

1) Tailoring ResNet to Image Dataset: In the first phase, the ResNet backbone was fine-tuned to adapt it to the specific characteristics of the meal image dataset. Only the last few layers of the ResNet model were unfrozen, allowing these layers to be updated during training, while the earlier layers retained their pre-trained weights to preserve generalized feature extraction. The ResNet outputs were passed through a fully connected layer to extract high-level image embeddings tailored to the task.

During this phase, the remaining components of the model, including the LSTM layers for CGM data processing and the metadata processing layers for tabular data, were kept frozen to prevent unnecessary updates and ensure computational efficiency. This phase focused solely on adapting the image-processing component of the model, ensuring that the

extracted embeddings were task-specific and optimized for calorie prediction.

2) Full Model Fine Tuning: In the second phase, the full model was fine-tuned to integrate features from all modalities, including image data, CGM readings, and tabular data. All layers were unfrozen except for the layers of ResNet, which were frozen to retain the generalized pre-trained features. This phase allowed the custom LSTM layers for CGM data and the metadata processing layers to learn task-specific patterns while fine-tuning the remaining trainable layers. The integration of multimodal data at this stage enhanced the model's ability to capture complex relationships between meal composition, glucose response, and calorie predictions.

The model training process involved processing multimodal inputs:

- Image Data: Meal images (breakfast and lunch) were passed through the fine-tuned ResNet backbone and fully connected layers to generate feature embeddings.
- CGM Data: Time-series glucose data, recorded at 5-minute intervals, was processed through LSTM layers to capture temporal glucose patterns.
- Tabular Data: Demographic and microbial abundance data were processed through fully connected layers to extract relevant features.

The multimodal features were combined and processed through subsequent layers to predict meal calorie content. Backpropagation was used to compute gradients for trainable parameters, and updates were performed using the Adam optimizer. The model was trained for 20 epochs in both phases.

Validation was conducted on a 20% holdout split of the original dataset after each epoch. The validation loss was computed without updating the model's parameters to ensure unbiased evaluation. This process provided insights into the model's generalization ability and facilitated adjustments to mitigate overfitting.

F. Evaluation Metrics

To evaluate the performance of the model in predicting meal calorie content, we employed the Root Mean Squared Relative Error (RMSRE) as the primary evaluation metric. RMSRE was chosen for its ability to measure prediction accuracy relative to the true calorie values, offering robustness against variations in meal calorie data. The RMSRE is defined as

$$\text{RMSRE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{y}_i - y_i}{y_i + \epsilon}\right)^2}$$

RMSRE normalizes the error by dividing the absolute error by the true value, providing a relative measure of prediction accuracy. This makes it particularly useful when calorie values vary widely, as it ensures fair evaluation across the entire dataset. A lower RMSRE indicates better model performance, signifying that the predicted calorie values closely align with the true values on a relative scale. RMSRE provides a robust measure of the model's accuracy and consistency, capturing how well the predictions generalize to unseen data.

In this study, RMSRE served as the primary evaluation criterion, guiding the optimization of the model during training and validating its effectiveness in predicting meal calories from multimodal data inputs.

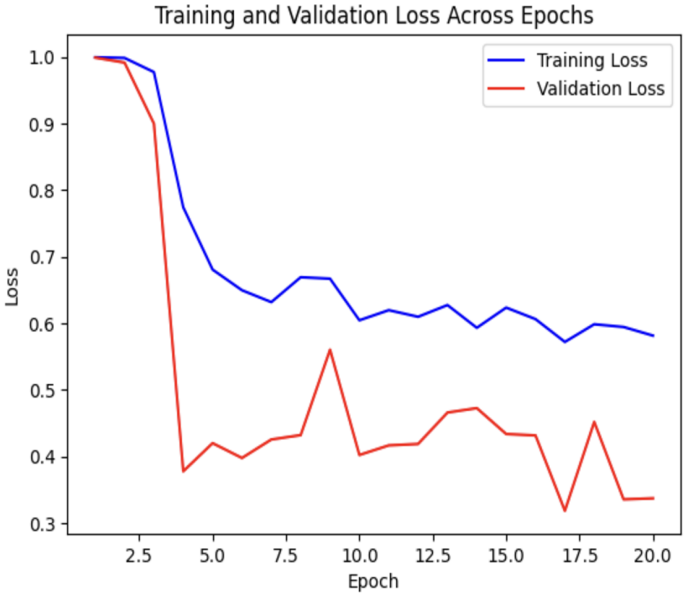

Fig. 4. Train v/s Validation Loss

IV. Result and Analysis

Table I summarizes the results of our experiments and highlights the performance of the proposed model under different configurations. The best performance was achieved with a dropout rate of 0.5, a learning rate of 0.001, 20 epochs, 128 as the input feature vector size of metadata layers, and 128 as the input feature vector size of the intermediate layers in the final fully connected layers, resulting in an RMSRE of 0.33 on the validation dataset for the target prediction task and RMSRE of 0.36 for the test dataset. The experiments reveal that increasing the number of epochs from 10 to 20 improves the model's performance by reducing the error by approximately 40%. Similarly, fine-tuning the learning rate demonstrated that a rate of 0.001 creates a balance between convergence speed and performance, outperforming higher and lower rates. Additionally, varying the input feature vector size of the metadata layers and the combined feature vector size in the final fully connected layers improved the model's ability to process structured data as it maintained the equilibrium between a very complex model taht would overfit and a very simple model that would underfit. Finally, the dropout rate of 0.5 was optimal, effectively mitigating overfitting while preserving model complexity. This analysis highlights the importance of hyperparameter tuning in optimizing the predictive accuracy of multimodal models.

TABLE I
LUNCH CALORIES PREDICTION PERFORMANCE COMPARISON AMONG DIFFERENT SET OF HYPERPARAMETERS

Epochs	Learning Rate	Dropout Rate	Input Feature Vector Size for Metadata Layers	Input Feature Vector Size for Intermediate Layers	RMSRE
10	0.001	0.5	128	128	0.56
10	0.001	0.25	1024	1024	0.561
20	**0.001**	**0.5**	**128**	**128**	**0.33**
20	0.001	0.5	256	256	0.51
30	0.001	0.40	512	1024	0.571
30	0.00001	0.5	128	1024	0.748

V. CONTRIBUTION

The implementation of this project was carried out collaboratively, with each team member contributing to different aspects of preprocessing, model training, and documentation.

A. Preprocessing

The data preprocessing phase was divided into three distinct parts to ensure efficiency and clarity:

- **Image Preprocessing:** Riddhi was responsible for preparing the meal images. This included resizing, data augmentation, and feature extraction using a pre-trained ResNet-50 model.
- **CGM Preprocessing:** Akanksha handled the preprocessing of continuous glucose monitoring (CGM) data. This involved organizing temporal glucose readings, extracting features like meal intervals, and preparing input sequences for the LSTM model.
- **Demographic Data Preprocessing:** Gourangi managed the preprocessing of tabular demographic and microbiome data, ensuring proper feature engineering, standardization, and alignment with other modalities.

B. Model Training

The training process was divided into modular and integrated phases:

- **Individual Modality Training:** Initially, the model components were trained separately for each data modality. Riddhi trained the image-processing module using ResNet-50, Akanksha worked on the CGM module using LSTM, and Gourangi developed the demographic and microbiome model using fully connected networks.
- **Multimodal Model Integration:** After training individual modules, we merged them into a unified framework and trained the combined model to leverage all modalities.
- **Hyperparameter Tuning:** Each member tuned hyperparameters for their respective modules on their machines to optimize performance. We collaborated to consolidate the results and finalize the best-performing model.

C. Documentation and Demonstration

- The report was collectively drafted, with each team member taking equal responsibility for specific sections.
- Riddhi recorded the final demo, showcasing the model's functionality and results.

This collaborative approach ensured that each member contributed their expertise, leading to a well-rounded implementation and a comprehensive project outcome.

REFERENCES

[1] "Joint Embedding of Food Photographs and Blood Glucose for Improved Calorie Estimation" By Lida Zhang1, Sicong Huang1, Anurag Das1, Edmund Do1, Namino Glantz2, Wendy Bevier3, Rony Santiago3,David Kerr4, Ricardo Gutierrez-Osuna1, and Bobak J. Mortazavi1

[2] "Predicting the Macronutrient Composition of Mixed Meals From Dietary Biomarkers in Blood" Anurag Das, Bobak Mortazavi , Member, IEEE, Seyedhooman Sajjadi , Theodora Chaspari , Member, IEEE, Laura E. Ruebush, Nicolaas E. Deutz , Gerard L. Cote, and Ricardo Gutierrez-Osuna