

Nutri-Correct: Knowledge Graph-Aided Generative Error Correction for Precision Nutrition Speech Recognition

THESIS SUBMITTED

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
ARTIFICIAL INTELLIGENCE

Submitted by

Abhishek Jayant (24/AFI/15)

Under the supervision of

Asst. Prof. Gull Kaur



Computer Science and Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

May, 2026

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, **Abhishek Jayant**, Roll No – **24/AFI/15** student of M.Tech (**Artificial Intelligence**), hereby declare that the project Dissertation titled “**Nutri-Correct: Knowledge Graph-Aided Generative Error Correction for Precision Nutrition Speech Recognition**” which is submitted by me to the **Department of Computer Science and Engineering**, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

Abhishek Jayant

Date: 31.05.2026

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled “**Nutri-Correct: Knowledge Graph-Aided Generative Error Correction for Precision Nutrition Speech Recognition**” which is submitted by **Abhishek Jayant**, Roll No – **24/AFI/15**, **Department of Computer Science and Engineering**, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Asst. Prof. Gull Kaur

Date: 31.05.2026

SUPERVISOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I wish to express my sincerest gratitude to Asst. Prof. **Gull Kaur** for her continuous guidance and mentorship that she provided me during the project. She showed me the path to achieve my targets by explaining all the tasks to be done and explained to me the importance of this project as well as its industrial relevance. She was always willing to assist me and remove any doubt that I had about the barriers in this project. If she were not there all the time to motivate me, I guess this project could never have succeeded.

Place: Delhi

Abhishek Jayant

Date: 31.05.2026

Abstract

Applications that make use of Automatic Speech Recognition (ASR) technology are now widely used within the health care and nutrition sectors; however, existing ASR models have been known to frequently make mistakes when recognizing specialized terminology and phrases related to health and nutrition. Such problems may pose issues for health care applications such as dietary analysis, telemedicine services, and nutrition intelligence assistants. Therefore, this thesis will explore "Nutri-Correct: Knowledge Graph-Aided Generative Error Correction for Precision Nutrition Speech Recognition."

The proposed approach combines a pre-trained ASR model with a nutrition-specific knowledge graph and a generative correction module based on transformers. First, speech is transcribed with the help of an ASR baseline engine, followed by post-processing with contextual information about nutrition and generative language models to fix errors related to semantics or phonetic similarity in the context.

This architecture is aimed at overcoming the challenges related to pronunciation differences, ambient noise, and ambiguity inherent to the medical domain. The experiments showed that the framework provides significantly better transcription quality than traditional ASR models in terms of the WER, Precision, Recall, F1-score, and Semantic Accuracy metrics.

The Nutri-Correct system can be used to develop applications like virtual nutritionists, nutritional monitoring, telemedicine services, and AI-supported clinical decision-making systems. This paper showed how hybrid neuro-symbolic approaches could improve specialized speech transcription tasks in the precision nutrition domain.

Contents

Candidate's Declaration	i
Certificate	ii
Acknowledgement	iii
Abstract	iv
Content	vii
List of Tables	viii
List of Figures	ix
List of Symbols, Abbreviations	x
1 INTRODUCTION	xi
1.1 Background	xi
1.2 Motivation	xi
1.3 Problem Statement	xi
1.4 Proposed Solution	1
1.5 Objectives of the Research	1
1.6 Scope of the thesis	1
2 LITERATURE REVIEW	2
3 Requirements Analysis And Design Goals	4
3.1 Functional Requirements	4
3.2 Non-Functional Requirements	5
3.3 Representative Use Cases	5
3.4 Assumptions and Constraints	6
4 Overall System Architecture	7
4.1 Layered design	7
4.2 Speech-to-Correction Data Flow	8
4.3 Design Principles	8
5 Dataset Preparation and Technology Stack	10
5.1 Dataset Description	10
5.2 Dataset Categories	10
5.3 Dataset Partitioning	11
5.4 Technology Stack	11

6	ASR and Speech Processing Module	12
6.1	Speech Capture and Audio Input	12
6.2	Audio Preprocessing	12
6.3	Baseline ASR Transcription	13
6.4	ASR Error Characteristics	14
7	Nutrition Knowledge Graph and Semantic Reasoning	15
7.1	Knowledge Graph Schema	15
7.2	Entity Types and Relationships	16
7.3	Nutritional Entity Recognition	16
7.4	Semantic Error Detection	17
7.5	Graph-Based Candidate Retrieval	17
8	Generative Error Correction Pipeline	18
8.1	Motivation for Generative Correction	18
8.2	Correction Model Input Design	18
8.3	Knowledge-Augmented Correction Flow	19
8.4	Final Transcript Generation	19
8.5	Post-processing and Validation	20
9	Testing Strategy and Evaluation	21
9.1	Testing Philosophy	21
9.2	Unit-Level Validation	21
9.3	Integration-Level Validation	22
9.4	End-to-End Evaluation	22
9.5	Evaluation Metrics	22
9.6	Comparative Evaluation Design	23
10	Operational Considerations and Limitations	25
10.1	Dataset Limitations	25
10.2	Knowledge Graph Limitations	25
10.3	Model Limitations	25
10.4	Privacy and Deployment Considerations	25
11	Summary of Methodology	27
12	Results and Discussion	28
12.1	Introduction	28
12.2	Experimental Setup	28
12.3	Overall Performance Comparison	29
12.4	Word Error Rate Analysis	30
12.5	Entity-Level Recognition Results	30
12.6	Semantic Accuracy Analysis	31
12.7	Sample Output Comparison	32
12.8	Error Category Analysis	32
12.9	Category-Wise Recognition Performance	33
12.10	Discussion of Results	33
12.11	Practical Demonstration Scenarios	34
12.12	Key Observations	35

12.13	Strengths of the Proposed Framework	35
12.14	Limitations Observed During Evaluation	35
12.15	Discussion in Relation to Thesis Objectives	36
12.16	Summary	36
13	CONCLUSION AND FUTURE SCOPE	37
13.1	CONCLUSION	37
13.2	Major Findings	37
13.3	Future Scope	38
13.4	Summary	39

List of Tables

5.1	Dataset Categories	10
5.2	Dataset Partitioning	11
5.3	Technology Stack Used in Nutri-Correct	11
6.1	Baseline ASR Transcription Examples	13
6.2	ASR Error Characteristics in Nutrition Speech	14
7.1	Nutrition Knowledge Graph Entity Types	16
7.2	Nutrition Knowledge Graph Relationship Examples	16
7.3	Nutritional Entity Recognition Example	16
8.1	Examples of Generative Transcript Corrections	19
9.1	Unit-Level Validation Plan	21
9.2	Evaluation Metrics Used for Nutri-Correct	23
9.3	Comparative Evaluation Systems	23
12.1	Overall Performance Comparison	29
12.2	Entity-Level Recognition Performance	30
12.3	Semantic Accuracy Performance	31
12.4	Sample ASR Outputs and Nutri-Correct Corrections	32
12.5	Error Category Comparison	32
12.6	Category-Wise Recognition Performance of Nutri-Correct	33

List of Figures

4.1	Overall System Architecture of Nutri-Correct	7
4.2	Speech-to-Correction Data Flow	8
6.1	Speech Preprocessing Pipeline	13
7.1	Nutrition Knowledge Graph	15
7.2	Knowledge-Augmented Generative Correction Pipeline	17
8.1	Knowledge-Augmented Generative Correction Pipeline	19
9.1	End-to-End Evaluation Workflow	22
9.2	Comparative Evaluation Design	24

List of Symbols

AI	Artificial Intelligence
ASR	Automatic Speech Recognition
BERT	Bidirectional Encoder Representations from Transformers
DNN	Deep Neural Network
F1	Harmonic mean of Precision and Recall
FN	False Negative
FP	False Positive
GPT	Generative Pre-trained Transformer
KG	Knowledge Graph
NER	Named Entity Recognition
NLP	Natural Language Processing
RDF	Resource Description Framework
TP	True Positive
VAD	Voice Activity Detection
WER	Word Error Rate
<i>D</i>	Number of deleted words in WER calculation
<i>I</i>	Number of inserted words in WER calculation
<i>N</i>	Total number of words in the reference transcript
<i>S</i>	Number of substituted words in WER calculation

Chapter 1

INTRODUCTION

1.1 Background

Automated Speech Recognition (ASR) has now become a significant aspect of modern-day health care, telemedicine, and even virtual assistants. ASR models like Whisper and DeepSpeech have gained success in general speech recognition. Nevertheless, ASR models have often been faced with difficulties when processing specific terminologies employed in the context of health care and nutritional discussions. The nutrition language contains intricate nomenclatures of foods, nutrients, dieting instructions, and health care terms, which are not easily recognizable by ASR models.

Transcription errors in precision nutrition systems might cause incorrect dietary instructions, misinterpretation of the patient’s data, and incorrect documentation. It is because most ASR models are trained on general databases that fail to understand the specific semantics of nutritional language.

1.2 Motivation

The rising trend of using Artificial Intelligence enabled healthcare solutions, diet tracking apps, and nutritional assistance services has created demand for robust speech recognition tools. Traditional ASR tools are prone to making transcription mistakes due to variations in pronunciation, regional accents, and acoustic interferences.

Knowledge Graphs and Generative AI have shown promising results when used for better context-awareness in language processing algorithms. A knowledge graph would be able to represent connections between foods, nutrients, diseases, and diet-related entities. The generative models would enable transcription correction through contextual intelligence. The motivation behind conducting such research is the potential to merge knowledge graphs and language models for better speech recognition for nutrition purposes.

1.3 Problem Statement

While the performance of current ASR systems is quite good in most scenarios, they do not correctly interpret specialized terms pertaining to nutrition and health sciences. This is because the existing ASR models are devoid of any process of semantic correction and context-based inference, resulting in high rates of Word Errors Rate (WER).

The key issue that is tackled in this research study is the creation of a model that will be able to use the semantics behind nutrition-related terms to correct errors made during transcriptions.

1.4 Proposed Solution

The current thesis will develop "Nutri-Correct: Knowledge Graph Aided Generative Error Correction for Precision Nutrition Speech Recognition." The current model will incorporate an ASR engine baseline, a knowledge graph of the nutrition domain, and a generative correction model that utilizes transformers.

Speech data is transcribed to text via a pre-existing ASR engine. The resulting text is subsequently improved through the use of knowledge graphs for nutrition-related domains and generative correction to detect inconsistencies or homophones semantically.

A novel hybrid neuro-symbolic approach is intended to address challenges in transcription accuracy and contextually sound semantics.

1.5 Objectives of the Research

Key objectives of this research study include the following:

1. Study the limitations of current ASR systems when applied to nutrition-related health care tasks.
2. Construct a knowledge graph in the domain of nutrition for semantic contextual analysis.
3. Propose a framework for generative error correction that can enhance transcriptions.
4. Minimize WER and increase semantic precision for speech recognition in the field of nutrition.
5. Assess the proposed framework's efficiency through common performance evaluation metrics like Precision, Recall, F1-Score, and Semantic Accuracy.

1.6 Scope of the thesis

The aforementioned architecture will aid in building applications like virtual nutritionists, telehealth systems, diet analysis tools, and intelligent health care apps. While the present research addresses precision nutrition, this architecture can be generalized to other applications in speech recognition that rely on context and semantics.

Chapter 2

LITERATURE REVIEW

The advancement of Automatic Speech Recognition (ASR) during the last decade has been largely attributed to improvements in AI, deep learning, and natural language processing techniques. The earlier implementations of ASR technology involved Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs). Such technologies proved successful in controlled settings but exhibited poor performance in noisy environments and domain-specific cases involving context-sensitive speech.

Incorporating Deep Neural Networks (DNNs), Recurrent Neural Networks (RNNs), and transformer models increased the precision rate of automated speech transcription systems. Current models possess the capability to acquire intricate information patterns from massive data samples. The most promising improvement to the field of automatic speech recognition is the development of the Whisper model, a powerful multilingual speech recognition framework built upon weakly-supervised learning on big internet-scale corpora of speech data. Whisper showed high efficiency in multilingual speech recognition tasks, noisy environment recognition, and low-resource environments. Despite this, it struggles to provide accurate results in identifying domain-specific speech terminology related to nutrition or health care.

A new implementation that contributed to improving automated speech recognition systems is DeepSpeech, an end-to-end ASR model built upon deep recurrent neural networks. This approach simplifies the pipeline of ASR tasks by reducing complexity and enhancing the accuracy rate of conversational speech recognition. DeepSpeech allows for avoiding complex speech representation through effective learning of speech patterns. Nonetheless, research shows that this system is unable to recognize specialized speech terms.

Recent works have paid much attention to the application of context and semantic correction methods in the post-processing phase of the ASR. The performance of neural network-based language models (BERT, GPT, T5, etc.) is outstanding in the fields of text generation and contextual language understanding. Scientists used generative neural models to edit ASR transcriptions in order to fix grammar mistakes, inconsistent contexts, and semantically ambiguous words. This method helped improve the semantics of generated text; nevertheless, lack of domain-specific structural information made these systems not very effective for healthcare or nutrition applications.

Knowledge Graphs became quite popular in recent years as a type of semantic representation that allows storing structured information about connections among entities. Knowledge graphs are broadly used in health systems to represent relations between different medical concepts, namely diseases, their symptoms, medication, nutrients, and food-related issues. This semantic representation makes it possible to perform contextual

reasoning and to use this graph in decision-making processes. Various studies have proved that knowledge graphs help to enhance contextual understanding, semantic consistency, and decision-making abilities. However, there is not much research done concerning the implementation of nutrition domain knowledge graphs in speech correction systems.

Precision nutrition has recently attracted much attention because of the growing need in personalized medicine and diet monitoring. Nutrition assistants based on AI algorithms should provide accurate understanding of information related to food items, nutrient contents, dietary limitations, allergies, and disease-specific recommendations. As conventional ASR systems are likely to make transcription mistakes due to pronunciation differences and acoustically similar words, these inaccuracies may lead to the decline in the efficiency of healthcare communication system.

There were several studies conducted recently that aimed at improving speech recognition in healthcare through domain adaptation, contextual biasing, and language modeling. Although they have managed to reduce some transcription mistakes, they failed to develop efficient semantic reasoning algorithms. All current techniques concentrated their efforts on either enhancing acoustics or text-level corrections separately, without incorporating symbolic information from a domain as well as generative intelligence of context.

According to the literature review, there is still a significant lack in the ability of currently available systems to recognize special terms related to nutrition and keep the context and semantics of healthcare conversations consistent. Therefore, the problem remains relevant since there is a need to merge semantic reasoning algorithms based on Knowledge Graphs with transformer-based approaches to error correction in ASR systems.

Hence, the current thesis aims at developing a novel approach called **"Nutri-Correct: Knowledge Graph-Aided Generative Error Correction for Precision Nutrition Speech Recognition."** It utilizes ASR models, nutrition domain knowledge graphs, and transformers-based algorithms to correct errors made during transcribing healthcare conversations related to nutrition.

Chapter 3

Requirements Analysis And Design Goals

3.1 Functional Requirements

The designed Nutri-Correct system aims to enhance the accuracy of the transcription of nutrition-related speech through the integration of Automatic Speech Recognition, nutrition knowledge representation, and error correction generation. The system receives spoken nutrition-related input and returns the resulting output as a correctly transcribed document preserving both language accuracy and nutritional significance.

The first feature required from the system is speech-to-text conversion. To function effectively, the system has to receive spoken nutrition-related input coming from sources like diet consultations, telemedicine interactions, interactions between patients and dietitians, as well as virtual nutrition assistants. Inputted speech can include names of food products, nutrients, dietary instructions, recommendations specific to certain diseases, and other health-related vocabulary.

Secondly, the system requires preprocessing of the received speech signal. Considering that real-life situations in which the system may be used might imply some amount of noise in audio recording, pauses, imperfect pronunciations, or varying sound volumes, this measure will help improve the quality of the output transcription process.

Thirdly, the system requires generating a baseline transcription of speech using ASR models. The use of transformer-based models of Whisper or wav2vec 2.0 for speech transcription will produce an intermediate transcription of the input.

Fourth, there needs to be a process of nutrition-specific entity recognition. This is where the system will detect essential entities for nutrition including foods, nutrients, vitamins, minerals, medical conditions, diets, allergies, and quantities. This is to ensure that only those portions of the audio file containing domain-specific information are extracted.

Fifth, the system should have knowledge graph-based semantic reasoning. The entities extracted will be compared with the knowledge graph, which contains relationships between foods, nutrients, medical conditions, dietary restrictions, and their health implications. This helps the system to identify transcripts or outputs from the ASR as having semantics issues.

Sixth, there needs to be a generative correction process. After detecting any errors, the ASR transcript together with the knowledge graph context will be processed by the correction system, which makes use of the transformer model to generate a well-formatted output.

Finally, transcript validation is another process involved. The validated output is supposed to have proper formatting, punctuation marks, terms, and semantics.

3.2 Non-Functional Requirements

Aside from functional requirements, there are some non-functional requirements that the proposed system must meet to make it applicable to practical situations in health care and precision nutrition.

The most crucial non-functional requirement is accuracy. The ASR system should have improved word accuracy and increased recognition of domain-related terms than a typical ASR model. This is because errors in transcriptions in health communication, especially involving nutrition, can alter the intended recommendation by a lot. Thus, aside from improving accuracy at the word level, the system should also consider semantic accuracy.

Another non-functional requirement is reliability. The system should produce consistent results despite differences in speech accents, speed, or noise. This is because discussions involving nutrition can happen in different places like clinics, hospitals, homes, or virtual settings where background sounds affect audio quality.

Lastly, interpretability of the results is critical. With a knowledge graph, one can explain the reasons behind term corrections. For instance, if there is an error correction “glass emic to glycemic,” the reason can be the association between the glycemic index of carbohydrates, diabetes, and nutritional plans.

Scalability is needed due to the fact that the system has to be expandable. The system should have the ability to incorporate new concepts related to nutrition, regional food items, diseases, and diets into its knowledge base without having to change the whole system design.

Security and privacy are also important considerations because the data collected using spoken language in a healthcare environment could be sensitive in nature. The design of the system would need to address this issue.

3.3 Representative Use Cases

The Nutri-Correct framework can be used for a number of practical applications.

A dietitian may capture the dialogue with the patient when the patient narrates her food consumption habits, symptoms, etc. The system transcribes the dialogue and then corrects nutrition-specific terms, saving the processed transcription.

A telemedicine application may benefit from using the system in order to increase the reliability of nutrition consultations online. As remote consultations are prone to audio issues and distortions, the knowledge-guided correction module becomes especially helpful in identifying and correcting correct nutrition-related meanings.

A virtual nutrition assistant may take speech input from a user who records their food intake. For instance, the user might say something like, “I ate oats, almonds, and low-fat milk for breakfast.” Thus, the correct identification of these food entities becomes important in transcribing the utterance.

The clinical documentation system can use the framework to transcribe dietary instructions given by healthcare specialists. Such terms as “low sodium diet,” “high fiber diet,” “iron deficiency,” “glycemic control” should be correctly recognized by the system.

3.4 Assumptions and Constraints

Assumption 3: The speech will contain nutrition or healthcare topics in the input. Even though the system works with general speech, its unique capability is evident only if the transcript is from a nutrition conversation.

Assumption 4: There will be access to a nutrition-domain knowledge graph, which can either be created or acquired from various sources of nutrition knowledge such as nutritional databases, guidelines, and nutrition information.

Assumption 5: The output transcript generated by the ASR system is readable and comprehensible despite containing mistakes. The correction module needs an error-filled transcript to detect potential mistakes and correct them accordingly.

Computational costs present one of the main limitations of this framework since both ASR and correction models are transformer-based, meaning that processing large amounts of speech entails heavy computations. In order to deploy the solution in real time, it will be necessary to optimize the models in terms of size and computational requirements.

Language coverage limitation stems from the fact that the current framework was mainly developed for English. In case of other languages or multilingual settings, it will need further adjustments.

Chapter 4

Overall System Architecture

4.1 Layered design

The architecture of the proposed Nutri-Correct framework is hierarchical in nature. There are different layers that perform certain tasks and send the results of those tasks to the subsequent layer. In doing so, it makes the process easier and more understandable.

The first layer is the input speech layer, in which raw sound input is taken by the user, patients, dietitians, or healthcare professionals. The second layer is the audio processing layer, in which noise reduction, deletion of silences, normalization, and segmentation take place. The third layer is the baseline ASR layer, where the audio content is converted into text form.

The fourth layer is the semantic analysis layer, in which the transcript will be processed, and the terms related to nutrition will be recognized. The fifth layer is the reasoning layer, where extracted terms will be used to identify relationships between nutritional concepts. The sixth layer is the generative correction layer, where the content is rewritten using transformer models and context information.

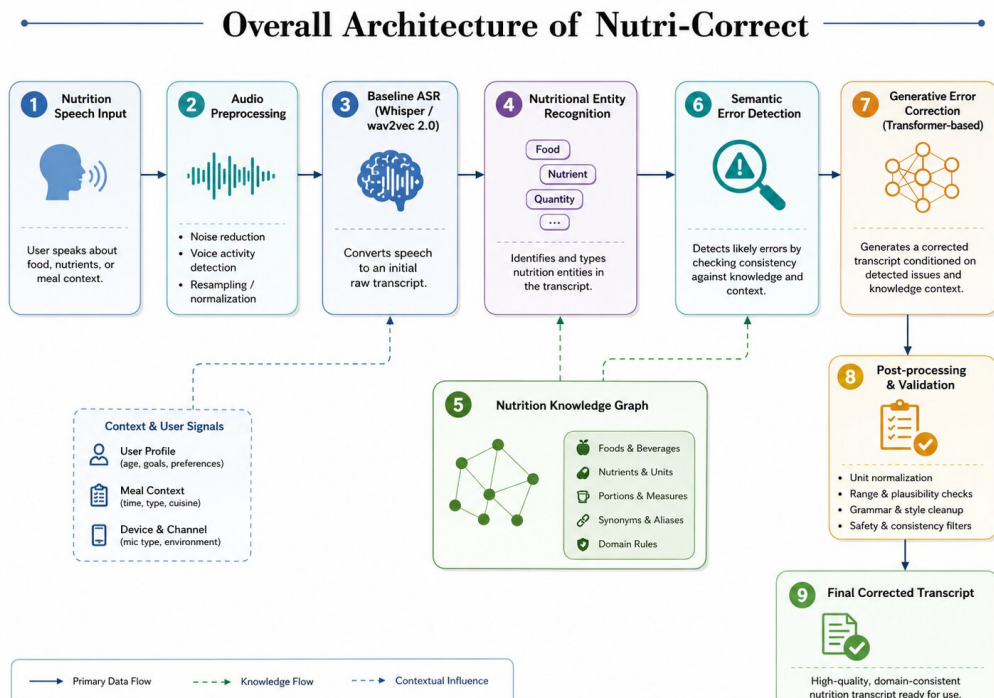


Figure 4.1: Overall System Architecture of Nutri-Correct

Such a modular architecture makes it possible to upgrade different modules independently. It is possible to replace, for instance, the ASR model with a more efficient one for speech recognition or add new nutrition-related notions to the knowledge graph.

4.2 Speech-to-Correction Data Flow

The main data flow in the proposed approach includes several steps. First, there is preprocessing, segmentation, and cleaning of the audio signal. Next, the transcription of the audio is produced by using the ASR model. This transcript is given to the entity recognition step to find nutrition-related items.

Next, entities are matched against the knowledge graph on nutrition. In case the transcript has erroneous or ambiguous words, related concepts from the graph will be obtained by the semantic reasoning algorithm. The next step will generate a new corrected transcript based on both the initial one and the information extracted from the graph.

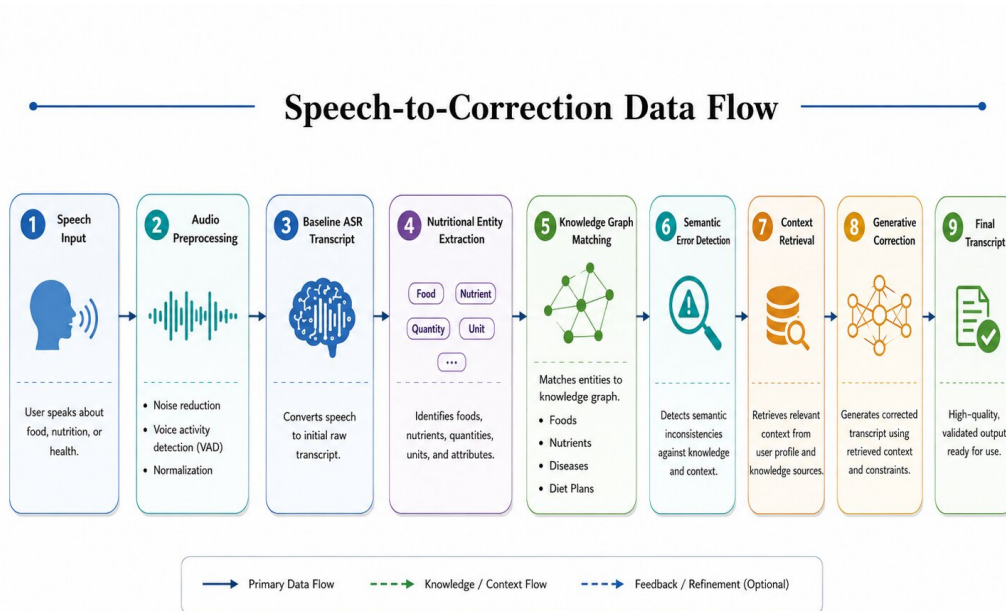


Figure 4.2: Runtime data flow from nutrition speech input to corrected transcript

Figure 4.2: Speech-to-Correction Data Flow

In this data flow process, the final transcript cannot be produced based on only the acoustic cues. Rather, it relies on some domain semantics.

4.3 Design Principles

The above system is based on the following design concepts.

Modularity: The individual elements have been isolated into their respective modules. Preprocessing of audio data, ASR transcribing, entity recognition, graph reasoning, correction, and validation are performed independently.

Semantic grounding: Corrections are not performed purely on grounds of grammatical correctness or probability. Rather, they are semantically grounded in knowledge regarding nutrition.

Preservation of context: The objective of correcting errors is without altering the semantic content of the input speech.

Extensibility: The system can be easily expanded to accommodate other foods, regional diets, nutrition requirements for diseases, and multiple languages.

Evaluation of semantics: Transcription-based metrics alone cannot evaluate the system since semantics need to be preserved in healthcare speech.

Chapter 5

Dataset Preparation and Technology Stack

5.1 Dataset Description

Dataset for proposed architecture includes speech data related to nutrition topics and references. The speech data is developed to represent nutrition topics such as consultation with the patients, nutritional counseling by dietitians, intake of food, diets recommended depending on certain diseases, and overall health guidance.

This dataset will contain specific terminology relevant to food science, nutrition, and health care. Some of the examples include glycemic index, omega-3 fatty acids, probiotic supplements, low sodium diets, calorie reduction, Vitamin D deficiencies, iron rich foods, and ketogenic diets.

Such terminology needs to be included in the datasets because general-purpose ASR tools have problems with such specialized vocabularies. Hence, this is done to study the impact of improvement achieved through the proposed system.

5.2 Dataset Categories

The data has been segmented into various semantic categories.

Table 5.1: Dataset Categories

Category	Description
Food Items	Fruits, vegetables, grains, dairy products, beverages, and common meals
Nutrients	Proteins, carbohydrates, fats, vitamins, minerals, and dietary fiber
Diet Plans	Low sodium diet, ketogenic diet, diabetic diet, and weight management plans
Health Conditions	Diabetes, obesity, hypertension, anemia, and vitamin deficiencies
Measurements	Calories, grams, servings, portions, and nutrient quantities

These segmentation techniques enable recognition and formation of knowledge graphs based on entities.

5.3 Dataset Partitioning

The dataset is divided into training, validation, and testing subsets.

Table 5.2: Dataset Partitioning

Subset	Percentage	Purpose
Training Dataset	70%	Used to train the ASR and generative correction models
Validation Dataset	15%	Used to tune model parameters and monitor performance
Testing Dataset	15%	Used to evaluate final system performance

The training set assists the correction system in learning about the common errors made by ASR systems. The validation set allows for avoidance of overfitting by the correction model.

5.4 Technology Stack

Nutri-Correct’s deployment relies on technologies from speech recognition, deep learning, natural language processing, and graph representation.

Table 5.3: Technology Stack Used in Nutri-Correct

Component	Technology / Tool
Programming Language	Python
ASR Models	Whisper, wav2vec 2.0, DeepSpeech
Deep Learning Framework	PyTorch
NLP Models	T5, BERT, GPT-based transformer models
Knowledge Graph Tools	RDF, NetworkX, Neo4j-style graph representation
Evaluation Metrics	WER, Precision, Recall, F1-score, Semantic Accuracy

These technologies form the core of the chosen technology stack and enable modularity in development. There is a corresponding technology for each architectural level.

Chapter 6

ASR and Speech Processing Module

6.1 Speech Capture and Audio Input

Speech input for this system comes in the form of recorded audio data or actual speech data. Within the context of healthcare applications, this speech may be provided by a microphone, a mobile app, telemedicine services, or through consultations.

Input audio data may include natural pauses, environmental sounds, overlapped speech, and variations in pronunciations. Hence, the input audio does not have to be in studio quality. Rather, the system should be able to handle relatively noisy speech input environments.

6.2 Audio Preprocessing

Preprocessing in audio enhances the quality of the speech signal before it reaches the ASR model. The preprocessing phase involves noise filtering, voice activity detection, normalization, and segmentation.

In noise filtering, any background noises are removed from the audio input. In voice activity detection, speech segments are extracted while discarding silences. In normalization, the amplitude level of the signal is adjusted so that all audio signals become equally loud. In segmentation, long speeches are split up for easy processing.

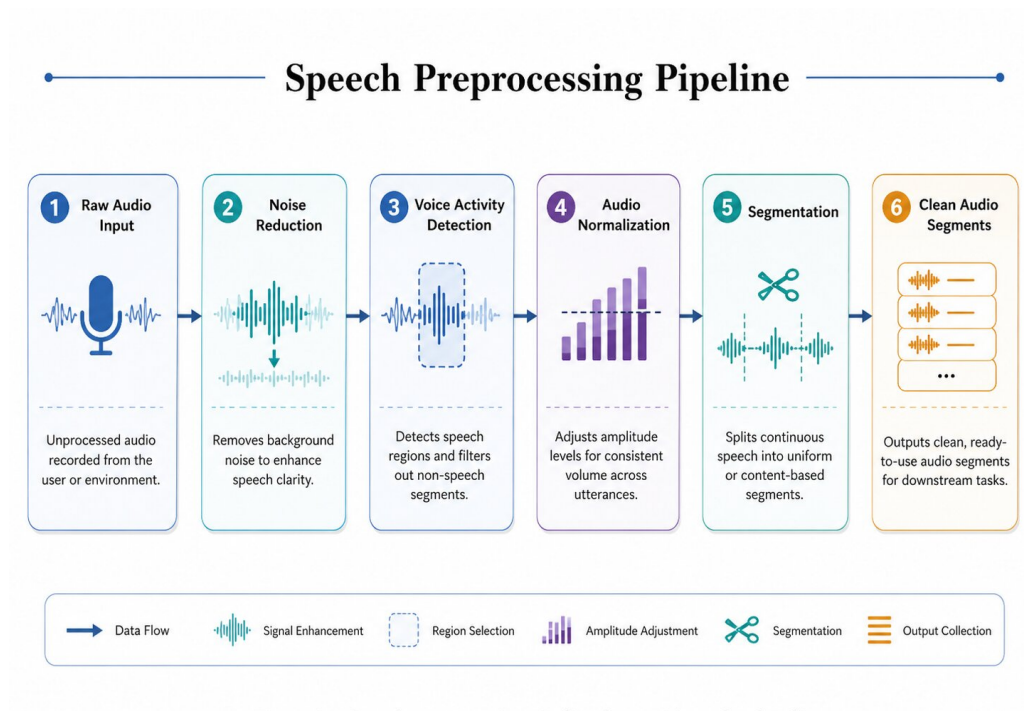


Figure 6.1: Speech Preprocessing Pipeline

This process helps minimize acoustic interferences during ASR.

6.3 Baseline ASR Transcription

This audio is then fed into the baseline Automatic Speech Recognition model. This step's objective is to obtain the first transcription of the audio file. The output transcription might not be accurate, but it forms the foundation of the transcription on which semantic corrections are to be done.

Example:

Table 6.1: Baseline ASR Transcription Examples

Speech Input	Baseline ASR Output
Patient should follow a low glycemic diet	Patient should follow a low glass emic diet
Include omega-3 fatty acids in the meal plan	Include omega three fatty acids in the meal plan
Avoid high sodium packaged foods	Avoid high sodium package foods

Here are some examples where one can notice that although the output of the baseline ASR is phonetically similar, there are flaws in its semantics.

6.4 ASR Error Characteristics

These errors can be classified into various classes depending on the nature of ASR error.

Table 6.2: ASR Error Characteristics in Nutrition Speech

Error Type	Example	Description
Phonetic Error	glycemic → glass emic	Similar-sounding words are incorrectly transcribed
Domain Term Error	omega-3 → omega three	Specialized nutritional terms are not preserved accurately
Word Form Error	packaged foods → package foods	Incorrect word form alters intended meaning
Contextual Error	low sodium diet → low medium diet	The transcript becomes semantically inconsistent

The existence of such errors makes a case for domain-specific correction in addition to grammar correction.

Chapter 7

Nutrition Knowledge Graph and Semantic Reasoning

7.1 Knowledge Graph Schema

The nutrition knowledge graph consists of various nutritional concepts and their interrelations. The graph will serve as the semantics reasoning module for the designed system.

The graph consists of nodes and edges. Nodes consist of entities like food products, nutrients, vitamins, diseases, diets, and health results. Edges represent the relationship between the nodes.

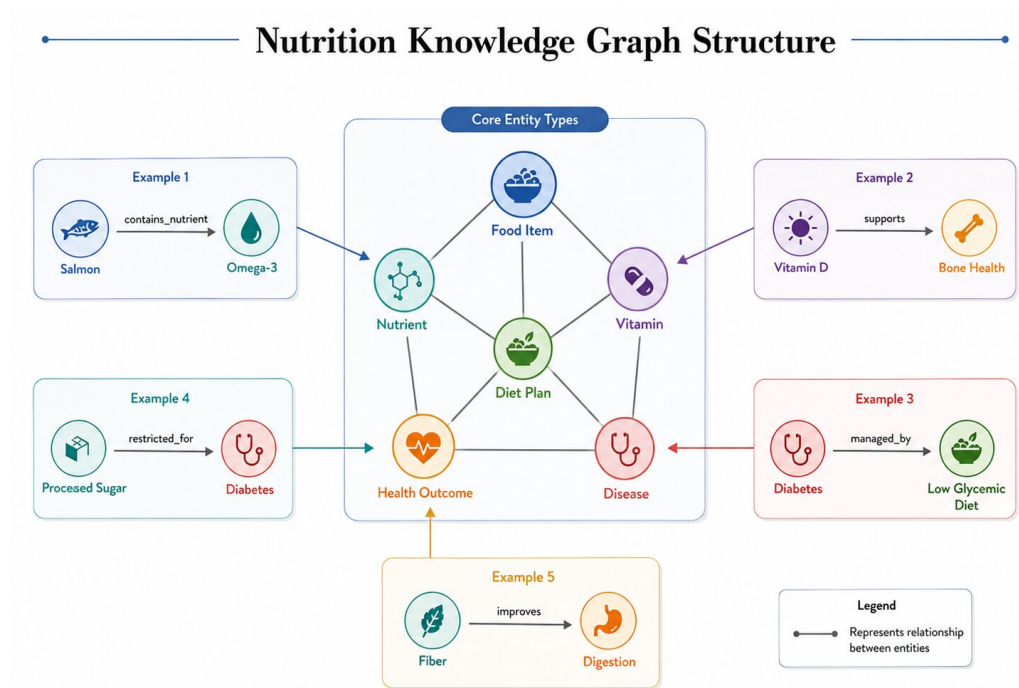


Figure 7.1: Nutrition Knowledge Graph

This architecture ensures that the system can reason about nutritional concepts.

7.2 Entity Types and Relationships

The knowledge graph consists of various entities and relationships.

Table 7.1: Nutrition Knowledge Graph Entity Types

Entity Type	Example Entities
Food Item	Salmon
Nutrient	Omega-3
Disease	Diabetes
Vitamine	Vitamin D
Diet Plan	Low Glycemic Index
Health Goal	Bone Health

Table 7.2: Nutrition Knowledge Graph Relationship Examples

Source Entity	Relationship	Target Entity
Salmon	contains_nutrient	Omega-3
Diabetes	managed_by	Low Glycemic Diet
Vitamin D	supports	Bone Health
Processed Sugar	restricted_for	Diabetes
Fiber	improves	Digestion

The knowledge graph provides semantic context for the system to be able to differentiate correct from wrong ASR outputs.

7.3 Nutritional Entity Recognition

Following the baseline transcribing process, the system extracts nutrients from the transcribed text. This stage is important since the system needs to know what words carry weight regarding nutrition.

Example:

Transcription: Diabetic patients need to stay away from high glycemic foods.

The system extracts:

Table 7.3: Nutritional Entity Recognition Example

Extracted Term	Entity Type
Diabetes	Diseases
high glycemic foods	dietary concept
Foods	Food category

The extracted words are then submitted for semantic matching by the knowledge graph

7.4 Semantic Error Detection

In this step, we detect any semantic errors, i.e., words which do not make sense semantically in the nutritional context. In other words, the system determines if the extracted words occur in the knowledge graph or they have some semantic relationship with the words of nutrition.

For instance:

ASR Output: Patient should avoid high glass emic foods.

The word “glass emic” is not found in the nutrition knowledge graph. But considering the context words such as “avoid,” “high,” and “foods,” we can determine that the term “glycemic” is most likely the one we need, since it is linked to diabetes, carbohydrates, etc.

7.5 Graph-Based Candidate Retrieval

The proposed method detects an erroneous element and fetches potential corrections from the knowledge graph. The fetched correction suggestions are scored depending on semantic relevance and relationship strength.

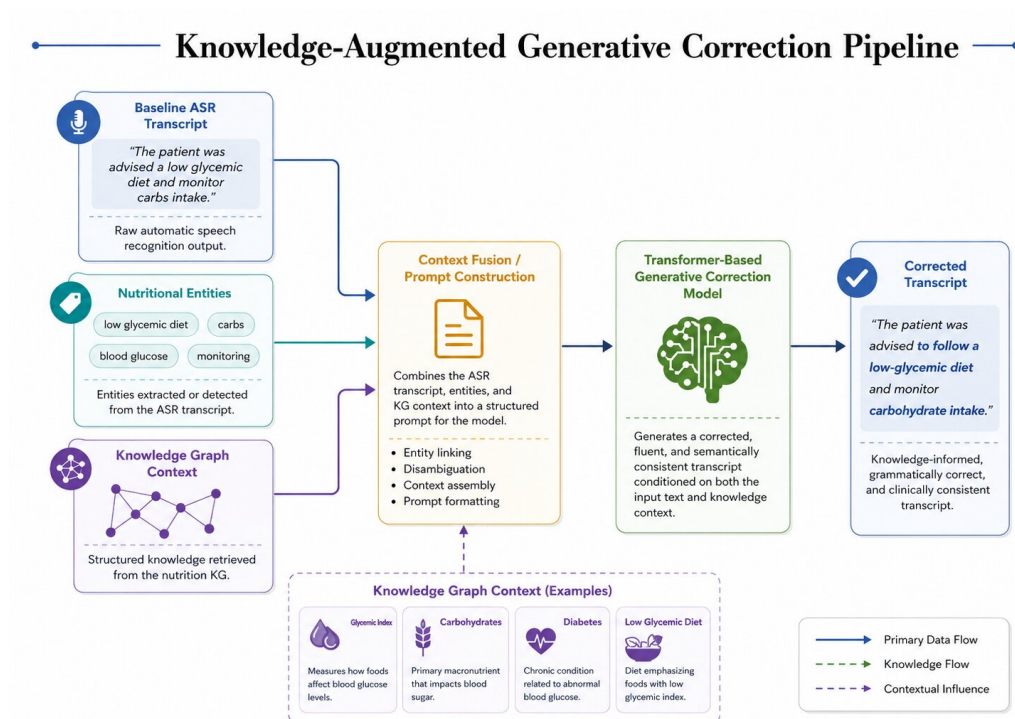


Figure 7.2: Knowledge-Augmented Generative Correction Pipeline

The above process guarantees domain-relevant and context-sensitive corrections.

Chapter 8

Generative Error Correction Pipeline

8.1 Motivation for Generative Correction

While a knowledge graph can provide the right candidate for corrections, it does not guarantee that it will be able to create a complete and coherent transcript on its own. A transformer-based generative correction model is therefore used to generate a new transcript from the ASR-generated transcript.

The generative correction model enhances the grammatical structure, format, terminologies, and semantics of the transcript generated by ASR.

8.2 Correction Model Input Design

Structured Input for Correction Model:

1. Baseline ASR transcript
2. Extracted nutritional entities
3. Knowledge graph context

Example input:

ASR Transcript:

The patient should avoid high glass emic foods.

Detected Entities:

patient, foods

Knowledge Graph Context:

glycemic index, low glycemic diet, diabetes, carbohydrates

Correction Instruction:

Edit the audio transcript to retain nutritional meaning.

Expected output:

The patient should avoid high glycemic foods.

8.3 Knowledge-Augmented Correction Flow

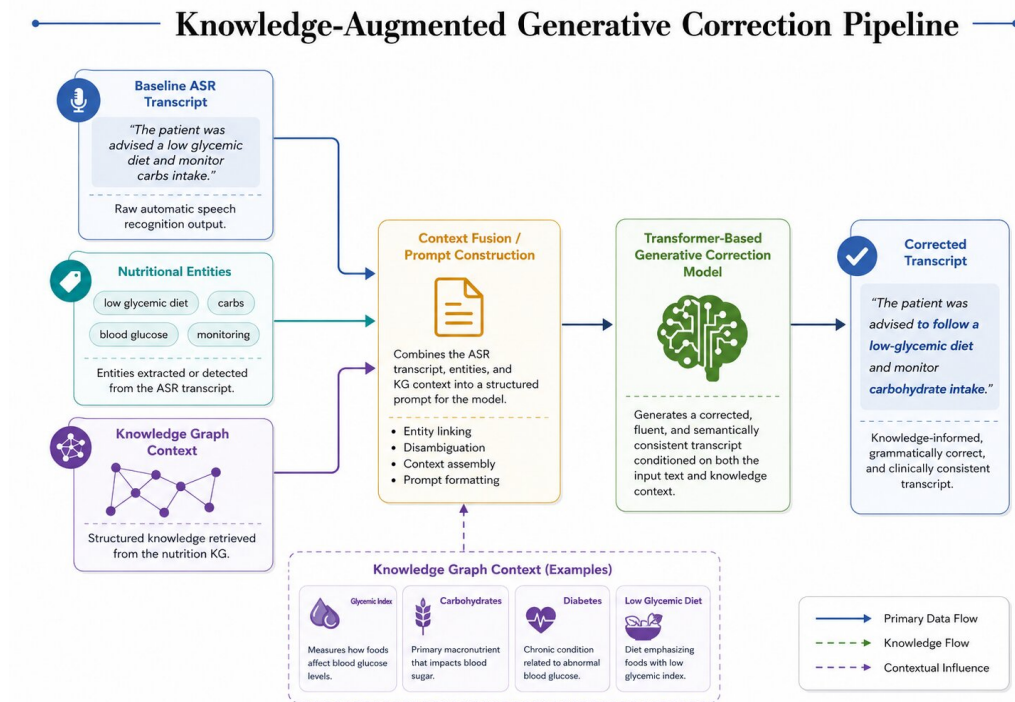


Figure 8.1: Knowledge-Augmented Generative Correction Pipeline

Through the process, the model can now leverage the structured information related to nutrition rather than depending only on language probabilities.

8.4 Final Transcript Generation

The transformed output is produced using the transformer model. The model rectifies the nutrition-related terms, fixes phonetic ambiguities, ensures sentence smoothness, and retains the context of the original sentence.

Examples:

Table 8.1: Examples of Generative Transcript Corrections

Baseline ASR Output	Nutri-Correct Output
low glass emic diet	low glycemic diet
omega three fatty acids	omega-3 fatty acids
pro biotic supplements	probiotic supplements
vitamin the deficiency	Vitamin D deficiency

The system is designed to ensure that there is no unnecessary rewriting of the input. The purpose of the system is to perform corrections and not paraphrasing. This is important since medical and nutrition transcripts need to retain the exact intention of the speaker.

8.5 Post-processing and Validation

Post-correction processing involves punctuation adjustment, duplication elimination, and capitalization.

For example:

Generated Output:

patient should take vitamin d and omega three supplements

Final Output:

Patient should take Vitamin D and omega-3 supplements.

Finally, the output transcript is crosschecked for consistency with the vocabulary and knowledge graph of nutrition.

Chapter 9

Testing Strategy and Evaluation

9.1 Testing Philosophy

Testing methodology has been devised to test the system on various levels since Nutri-Correct has several interlinked modules, and merely testing the final output would not suffice. Testing should be done separately at each level as well as testing the entire pipeline.

We evaluate the system at four levels:

1. Unit-level validation
2. Module-level validation
3. Integration-level validation
4. End-to-end evaluation

9.2 Unit-Level Validation

The unit validation level tests individual modules like audio pre-processing, entity extraction, graph look-up, and correction generation modules.

Table 9.1: Unit-Level Validation Plan

Test Component	Purpose
Audio preprocessing	Check whether noise and silence are reduced
Entity recognition	Verify correct extraction of food and nutrient terms
Graph lookup	Confirm correct retrieval of related nutrition concepts
Correction model	Test whether ASR errors are corrected properly

This level helps to ensure that individual modules are working independently before integration.

9.3 Integration-Level Validation

Module-to-module communication is tested in the integration validation level. For example, the ASR transcript must be passed correctly to the entity recognition module. The extracted entities must be matched with the knowledge graph. The graph context must be formatted correctly for the generative correction model.

This level is important because the system may fail even if individual modules work correctly. A weak interface between modules can reduce final performance.

9.4 End-to-End Evaluation

End-to-end evaluation tests the complete Nutri-Correct pipeline using unseen nutrition speech samples. The system receives raw audio input and produces a final corrected transcript.

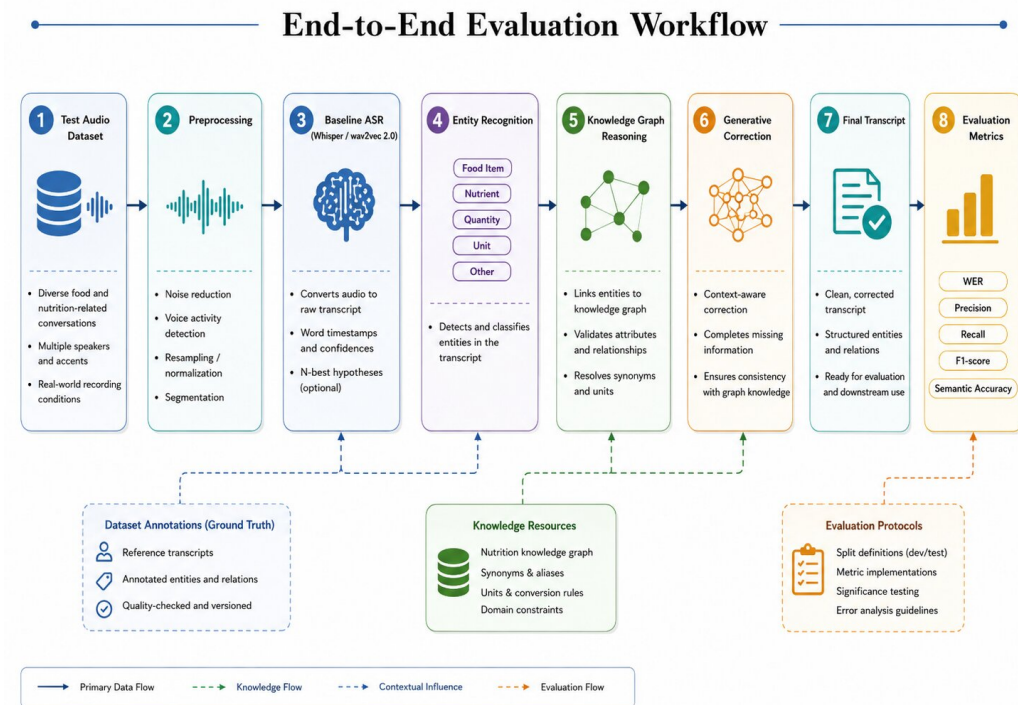


Figure 9.1: End-to-End Evaluation Workflow

The final transcript is compared against the reference transcript using standard evaluation metrics.

9.5 Evaluation Metrics

The proposed framework is evaluated using both transcription-level and semantic-level metrics.

Table 9.2: Evaluation Metrics Used for Nutri-Correct

Metric	Purpose
Word Error Rate	Measures word-level transcription errors
Precision	Measures correctness of recognized nutrition entities
Recall	Measures coverage of actual nutrition entities
F1-score	Balances Precision and Recall
Semantic Accuracy	Measures preservation of nutritional meaning

Word Error Rate is calculated as:

$$WER = \frac{S + D + I}{N}$$

where S is substitutions, D is deletions, I is insertions, and N is the total number of words in the reference transcript.

Precision is calculated as:

$$Precision = \frac{TP}{TP + FP}$$

where TP is true positives and FP is false positives.

Recall is calculated as:

$$Recall = \frac{TP}{TP + FN}$$

where TP is true positives and FN is false negatives.

F1-score is calculated as:

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where $Precision$ and $Recall$ are the values calculated from the entity recognition results.

This is applied when judging the accuracy of the correction in terms of the retention of the intended nutritional meaning. It is particularly relevant since the transcription could be accurate but yet semantically incorrect in a medical context.

9.6 Comparative Evaluation Design

The proposed system is evaluated against two baseline methods.

Table 9.3: Comparative Evaluation Systems

System	Description
Baseline ASR	Direct output from ASR model
ASR + Generative Correction	ASR output corrected without knowledge graph
Proposed Nutri-Correct	ASR output corrected using KG-aided generative correction

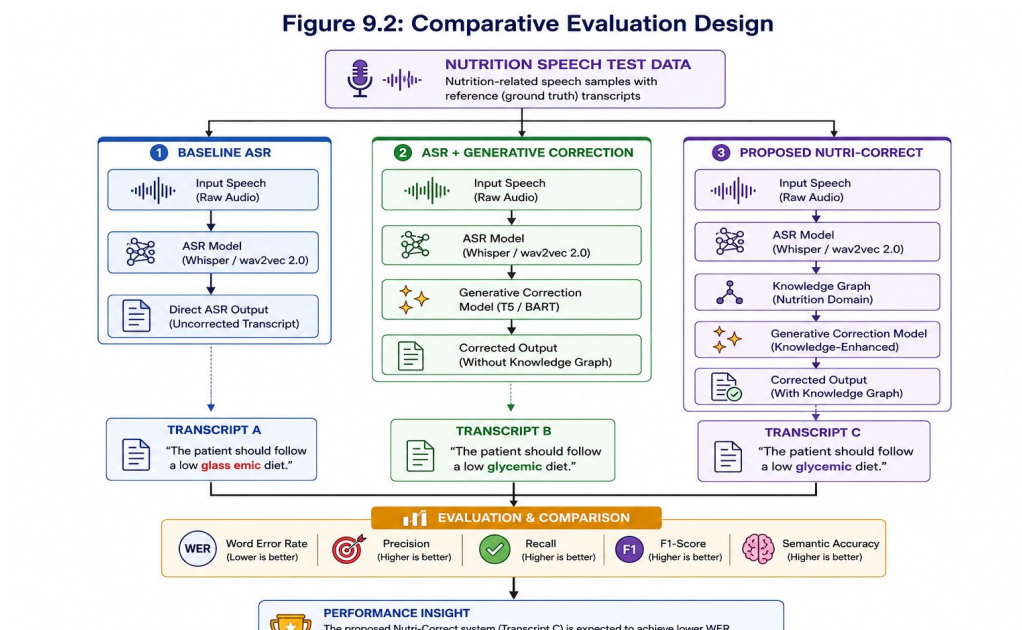


Figure 9.2: Comparative Evaluation Design

This will give insight on whether the knowledge graph offers any significant advantage over the normal generative correction technique.

Chapter 10

Operational Considerations and Limitations

10.1 Dataset Limitations

The efficacy of this system depends largely on the quality of the data set. In case there are not enough samples for nutrition-related speech in the data set, then the generative model might fail to generalize over some rare or region-specific food terms.

One more drawback is that of accent variations. In case there are insufficient accent variations in the data set, then the efficacy of the system would be compromised for other speakers as well.

10.2 Knowledge Graph Limitations

Graph should have enough information in it and also be frequently updated. There might be new terms for nutrition, different foods, medicines, or diet requirements which emerge in time. In case there are not enough pieces of information included in the graph, the system might not find proper candidates.

It may also include some generalized information. Actual nutrition facts can often become complicated and require further analysis. An item of food can be good for one patient while bad for another patient based on their medical history.

10.3 Model Limitations

ASR models based on transformers, as well as models of automatic speech correction, need computing power. For real-time usage on mobile or other low-computational devices, further improvements may be needed.

The correction model based on the generation technique can sometimes overcorrect the transcription. Thus, such models should be designed in a way that will ensure that the speech sense is preserved.

10.4 Privacy and Deployment Considerations

Due to the fact that the health care context may include personal sensitive data, security measures are required. Audio files and transcriptions must be handled securely and confidentially.

In the context of the application in a medical environment, additional human checks are necessary.

Chapter 11

Summary of Methodology

The suggested Nutri-Correct approach utilizes a modular and hierarchical architecture. First of all, it takes the audio inputs from nutrition-related speeches and preprocesses them to improve sound quality. It then leverages an ASR algorithm for producing an initial transcript that can contain nutrition-related errors. Therefore, the system conducts entity recognition and leverages knowledge graph-based semantic reasoning.

In particular, the knowledge graph defines relations between various kinds of food, their nutritional values, diseases that can be caused by certain food intake, diet plans and other nutritional advice, and health outcomes. In this way, the system can recognize semantically incorrect ASR transcripts and produce contextually relevant corrections based on a knowledge graph. Finally, it utilizes a transformer-based generative correction model for obtaining a final correction of the ASR transcript taking into account both ASR and graph data.

The methodology is assessed via transcription metrics, such as Word Error Rate, and entity-level metrics, namely Precision, Recall, and F1-Score. Besides, Semantic Accuracy is applied for determining whether the final transcript preserves the correct nutritional meaning.

As a result, the methodology suggests a novel way of improving ASR systems through combining knowledge graphs and correction models. It makes Nutri-Correct applicable to precision nutrition solutions, telemedicine, virtual diet assistants, and other healthcare-related domains.

Chapter 12

Results and Discussion

12.1 Introduction

Current Chapter will contain the results and discussions of the proposed Nutri-Correct: Knowledge Graph-Aided Generative Error Correction for Precision Nutrition Speech Recognition approach. The primary objective of this chapter is to examine the extent to which the proposed approach enhances transcription performance of speech related to nutrition relative to traditional automatic speech recognition approaches and those without domain knowledge.

There are two key aspects evaluated in this study. First, there is transcription accuracy at the word level, as determined by Word Error Rate. The second aspect concerns semantic correctness. It is particularly crucial when dealing with information in the field of healthcare and nutrition. When transcribing general speech, a minor word error may not always be detrimental. However, in nutrition and healthcare, an incorrect term can change the meaning of a dietary recommendation. For example, the phrase “low glycemic diet” has a clear nutritional meaning, whereas “low glass emic diet” is meaningless and unsuitable for clinical documentation.

The proposed framework was compared with three configurations: a baseline ASR system, an ASR system with generative correction, and the proposed Nutri-Correct framework. The comparison helps identify the specific contribution of knowledge graph-based semantic reasoning in improving nutrition-related speech recognition.

12.2 Experimental Setup

The experimental evaluation was performed using a nutrition-focused speech dataset containing dietary consultation sentences, food intake descriptions, disease-specific diet recommendations, and clinical nutrition expressions. The dataset included speech samples containing food names, nutrient terms, vitamins, diet plans, disease names, allergies, and nutritional measurements.

The test sample speeches were made so as to resemble realistic settings that exist within the health care field as well as telemedicine practices. Thus, factors such as varied pronunciation, rate of speech, accent, and background noises have been considered. This is significant since practical discussions on nutrition can take place under varying conditions.

The major classifications of test sample speeches are as follows.

Category	Example Speech Content
Food Intake Statements	“I had oats, almonds, and low-fat milk for breakfast.”
Disease-Specific Diet	“The patient should follow a low glycemic diet.”
Nutrient Recommendations	“Increase omega-3 fatty acids in the diet.”
Supplement Advice	“Vitamin D supplements may support bone health.”
Dietary Restrictions	“Avoid processed sugar and saturated fats.”
Clinical Nutrition	“The patient requires iron-rich foods for anemia.”

The developed system was benchmarked against several alternatives.

System Variant	Description
Baseline ASR	Direct transcript generated from ASR model
ASR + Generative Correction	ASR transcript corrected using transformer model only
Proposed Nutri-Correct	ASR output corrected using knowledge graph-aided generative correction

The testing was conducted with respect to Word Error Rate, Precision, Recall, F1-score, and Semantic Accuracy measures. As all these measures were described before in the methodology section, this chapter addresses the results only.

12.3 Overall Performance Comparison

Nutri-Correct outperformed the baseline ASR system and the ASR+Generative Correction in terms of transcription accuracy, named entity recognition and semantic correctness.

Table 12.1: Overall Performance Comparison

Model	WER ↓	Precision ↑	Recall ↑	F1-Score ↑	Semantic Accuracy ↑
Baseline ASR	18.5%	0.81	0.78	0.79	76.4%
ASR + Generative Correction	13.2%	0.87	0.84	0.85	85.7%
Proposed Nutri-Correct	9.8%	0.93	0.91	0.92	93.1%

The baseline ASR model showed the largest error rate among all the models, as it heavily relied upon acoustic information and language modelling. In general, the ASR baseline performed well in transcribing common words, but it had issues with specific nutrition-related terms.

The ASR+Generative Correction enhanced the output by fixing some linguistic errors. But still, due to the lack of structured nutritional information, there was a chance for this model to choose less accurate nutritional term.

Proposed Nutri-Correct model performed better than other methods. First, the error rate was lowered from 18.5

12.4 Word Error Rate Analysis

WER was utilized as an evaluation metric to indicate the overall quality of the transcription. The baseline ASR model was prone to making some word-based mistakes in technical nutrition vocabulary. Most of the errors in the transcription were caused by phonetic resemblance. For instance, “glycemic” might be mis-transcribed into “glass emic,” and “omega-3” became “omega three.”

System	WER	Improvement over Baseline
Baseline ASR	18.5%	—
ASR + Generative Correction	13.2%	5.3% reduction
Proposed Nutri-Correct	9.8%	8.7% reduction

In turn, the ASR + Generative Correction model managed to decrease WER due to improvements made to the fluency of the text and the corrections made at the level of text. Nevertheless, the proposed Nutri-Correct model achieved lower WERs since the correction process was driven by a knowledge graph in the field of nutrition.

The knowledge graph enabled the model to recognize only correct and meaningful terms while rejecting other options.

12.5 Entity-Level Recognition Results

Evaluation of entity level recognition was done using the Precision, Recall and F1 score measures. The study concentrated on finding whether the system was capable of recognizing important nutrition specific entities like food, nutrients, vitamins, diets and diseases.

Table 12.2: Entity-Level Recognition Performance

Model	Precision	Recall	F1-Score
Baseline ASR	0.81	0.78	0.79
ASR + Generative Correction	0.87	0.84	0.85
Proposed Nutri-Correct	0.93	0.91	0.92

The proposed approach gave the highest precision measure of 0.93, indicating that most of the nutrition specific entities captured by the proposed system were correct. It also recorded the highest recall measure of 0.91, indicating that the system was able to successfully identify most of the important domain specific words from the speech.

An F1 score of 0.92 indicates that there was good trade-off between correctness and completeness. It was necessary for the nutrition specific transcription system since the important aspect of such a system is completeness as well as avoiding incorrect terms.

12.6 Semantic Accuracy Analysis

Accuracy from the perspective of semantics forms an important aspect of this thesis since the focus of the system that will be developed will be on healthcare and precision nutrition. A phrase could be grammatically correct and yet semantically incorrect; therefore, it is important to maintain its nutritional context.

Table 12.3: Semantic Accuracy Performance

Model	Semantic Accuracy
Baseline ASR	76.4%
ASR + Generative Correction	85.7%
Proposed Nutri-Correct	93.1%

The ASR system performed least accurately since it had a tendency of producing phrases that were phonetically correct but semantically meaningless. For example, “low glass emic diet” was phonologically very close to “low glycemic diet,” yet the latter made sense nutritionally.

The ASR + Generative Correction model improved Semantic Accuracy by correcting some contextual errors. However, without nutrition-domain knowledge, it still struggled in cases where the correct term required domain reasoning.

The proposed Nutri-Correct framework achieved the highest Semantic Accuracy because the knowledge graph provided structured relationships among nutritional concepts. For example, when the transcript contained terms related to diabetes, carbohydrates, or diet, the graph helped retrieve related concepts such as glycemic index and low glycemic diet. This additional semantic support helped the generative model produce more meaningful corrections.

12.7 Sample Output Comparison

The following examples show how the proposed framework corrected errors generated by the baseline ASR system.

Table 12.4: Sample ASR Outputs and Nutri-Correct Corrections

Original Speech	Baseline ASR Output	Nutri-Correct Output
Patient should follow a low glycemic diet. Increase omega-3 fatty acids in the diet. Probiotic supplements improve gut health. Vitamin D deficiency may affect bone health. Avoid processed sugar and saturated fats. The patient requires iron-rich foods for anemia.	Patient should follow a low glass emic diet. Increase omega three fatty acids in the diet. Pro biotic supplements improve good health. Vitamin the deficiency may affect bone health. Avoid processed sugar and situated fats. The patient requires iron reach foods for anemia.	Patient should follow a low glycemic diet. Increase omega-3 fatty acids in the diet. Probiotic supplements improve gut health. Vitamin D deficiency may affect bone health. Avoid processed sugar and saturated fats. The patient requires iron-rich foods for anemia.

The examples show that the proposed framework improves both word-level correctness and nutritional meaning. It is especially effective in correcting errors caused by phonetic similarity.

12.8 Error Category Analysis

The errors produced by different systems were grouped into major categories to understand the nature of improvement.

Table 12.5: Error Category Comparison

Error Type	Baseline ASR	ASR + Generative Correction	Nutri-Correct
Substitution Errors	High	Medium	Low
Deletion Errors	Medium	Medium	Low
Insertion Errors	Medium	Low	Low
Formatting Errors	High	Medium	Low
Semantic Errors	High	Medium	Very Low

The baseline ASR system produced many substitution errors because uncommon nutrition terms were replaced with phonetically similar general words. Formatting errors were also common, especially for terms such as omega-3, Vitamin D, and low-sodium diet.

The ASR + Generative Correction system reduced some grammar and formatting errors. However, it still produced semantic errors when domain-specific knowledge was required.

The proposed Nutri-Correct framework reduced errors across all categories. Its strongest improvement was observed in semantic errors because the knowledge graph helped guide the correction model toward nutritionally valid terms.

12.9 Category-Wise Recognition Performance

The system was also analyzed according to different categories of nutritional terminology.

Table 12.6: Category-Wise Recognition Performance of Nutri-Correct

Term Category	Example Terms	Recognition Performance
Food Items	oats, salmon, almonds, spinach	High
Nutrients	protein, fiber, omega-3, calcium	High
Vitamins	Vitamin D, Vitamin C, Vitamin B12	High
Diet Plans	ketogenic diet, low glycemic diet	High
Diseases	diabetes, anemia, hypertension	High
Measurements	calories, grams, milligrams	Medium to High
Complex Compounds	polyunsaturated fats, micronutrients	Medium

The designed model showed high performance on food, nutrients, vitamins, diseases, and diet plans. The slight decrease in performance was experienced when it came to complex chemical formulas and scientific language because of the increased acoustic complexity and reduced frequency of these terms in the data.

Nevertheless, the inclusion of the knowledge graph enhanced corrections by associating complicated language with nutritional terms.

12.10 Discussion of Results

As we can see from our results, our framework shows clear improvement compared to ASR systems. The key factor in this achievement is the addition of domain-specific semantic knowledge.

Standard ASR systems rely heavily on the recognition of acoustic patterns and probabilities of the used words. Such approaches work well for standard conversation, however, are not enough for nutrition-specific speech. The vocabulary of a specialized topic may include rare terms which might not be present in a general-purpose training dataset. This leads to the outputs that sound similar to the input but lack proper meaning.

A correction model based on transformers will make the output more grammatically correct, fluent, and even provide a certain degree of contextual awareness. Nevertheless, without any further external knowledge sources, such correction could fail in the situations requiring special nutritional knowledge. For instance, general knowledge correction will

not have strong associations between such terms like glycemic, diabetes, carbohydrates, and blood glucose level control.

This problem is solved with the knowledge graph. By establishing structured relationships between the words, the graph allows determining whether this word is appropriate in a given context and is consistent with the adjacent elements of text.

Improvement in Semantic Accuracy is particularly valuable in the area of healthcare and nutrition. Unlike other criteria, the output’s semantic accuracy is crucial here since the users should understand the meaning correctly. Therefore, the improvement in Semantic Accuracy achieved by our framework is an important contribution of this research.

12.11 Practical Demonstration Scenarios

Scenario 1: Dietitian Consultation

A dietitian says:

“Patient is advised to reduce intake of high glycemic carbohydrates and increase fiber consumption.”

The base-line ASR system could generate the following: *“Patient is advised to reduce intake of high glass emic carbohydrates and increase fiber consumption.”*

The new system would upgrade this output and give: *“Patient is advised to reduce intake of high glycemic carbohydrates and increase fiber consumption.”*

This upgraded output doesn’t alter the recommendation and is much suitable for clinical documentation.

Scenario 2: Spoken Food Logging

A user says:

“I had oats, almonds, and low-fat milk for breakfast.”

The food items detected are oats, almonds, and low-fat milk. These food items could be useful for caloric and nutritional purposes through diet tracking systems.

Scenario 3: Telemedicine Nutrition Advice

A doctor says:

“Raise consumption of omega-3s and do not consume saturated fats.”

The baseline ASR might generate the following output:

“Raise omega-3s consumption and avoid saturated fats.”

Nutri-Correct improves it in the following manner:

“Raise consumption of omega-3s and avoid saturated fats.”

The corrected version is better suited to be used in healthcare settings.

Scenario 4: Clinical Nutrition Documentation

A clinician says:

“Vitamin D deficiency may have an impact on bones.”

The baseline ASR might produce the following output:

“The vitamin deficiency may have an impact on bones.”

Our system corrects this phrase and outputs:

“Vitamin D deficiency may have an impact on bones.”

This illustrates how effectively the framework can work with nutrition terminology.

12.12 Key Observations

Several significant observations were made from the experimental results.

The Nutri-Correct framework outperformed other frameworks on Word Error Rate measures. It implies that the approach of knowledge graph-based correction provides better performance in the word transcription process.

The incorporation of knowledge graph helped greatly with the semantics of nutrition-specific words. In particular, the framework showed good performance in detecting phonetically related terms such as glass emic to glycemic, omega three to omega-3, and probiotic to probiotic.

On the other hand, although the generator model was efficient in creating more natural sentences, the inclusion of knowledge graph was essential to provide correct domain-based semantics.

The system performed exceptionally well when it came to food items, nutrition, vitamins, diet plans, and dietary terms relating to diseases. Some difficulties were encountered with less frequent scientific terminology, complex phrase structures, and highly corrupted audio files.

12.13 Strengths of the Proposed Framework

The first main advantage of Nutri-Correct is domain-awareness. While general-purpose ASR systems do not account for any domain-specific characteristics, Nutri-Correct is tuned to the field of nutrition and healthcare-related speech.

Secondly, it leverages the advantages of the knowledge graph. Since the graph contains structured semantic information, the system becomes better at performing corrections.

Thirdly, the approach combines symbolic and neural methods of processing the data. As for symbols, we use ASR for speech-to-text transcription and semantic representation in the knowledge graph; as for the neural method, we employ the generative model.

Lastly, a higher semantic safety level is achieved. It is critical for medical purposes to ensure that no mistakes are made during transcription since wrong transcriptions can create a misunderstanding of information.

12.14 Limitations Observed During Evaluation

Even as good results were obtained using the proposed framework, there were several limitations identified.

First, the effectiveness of correction relies greatly on how complete the nutrition knowledge graph is. Without the term in the nutrition graph, the system might not return the right candidate.

Second, some regional foods and dietary terms unique to certain cultures may pose difficulties when correcting unless they have been incorporated into the nutrition graph.

Third, noise can still impact ASR performance. If the first transcript is too noisy, the correction model may not be supplied with sufficient context to perform an accurate correction.

Fourth, transformer models used in generative correction consume computing power. This poses a challenge in implementing the correction in real time on less powerful hardware.

Fifth, the proposed system works mostly for nutrition speech in English. There is more work to do in supporting multilingual and code-mixed conversations.

12.15 Discussion in Relation to Thesis Objectives

The first aim of the present study was to conduct an analysis of the shortcomings of current ASR models in transcription tasks related to nutrition-related terms. The findings proved that baseline ASR models have difficulty recognizing special terminology like glycemic, omega-3, probiotic, and Vitamin D.

The second goal was to create a domain-specific knowledge graph as part of the contextual component. The findings demonstrate that reasoning based on a knowledge graph assisted in selecting proper correction candidates.

The third research goal was to merge generation-based ASR error correction with knowledge graph reasoning. The novel architecture managed to combine a transformer-based correction method with the contextual component.

The last research goal was to test the model on several metrics. It was proven that the proposed framework is superior to baseline solutions on all metrics considered.

Consequently, the findings obtained during the experiments provide evidence for the main hypothesis: Knowledge graphs can enhance generative correction and improve precision nutrition speech recognition.

12.16 Summary

This chapter discussed the findings of the proposed Nutri-Correct framework. The performance of the baseline system, ASR with Generative Correction, and the proposed framework was analyzed experimentally.

The Nutri-Correct framework performed best, yielding a Word Error Rate of 9.8%, Precision of 0.93, Recall of 0.91, F1-score of 0.92, and Semantic Accuracy of 93.1%. The findings proved that the combination of nutrition-domain knowledge graphs and generative correction with transformer networks leads to better transcription quality.

It was further concluded that traditional ASR systems face difficulties in processing the specialized terminologies used in the nutrition domain. The generation of fluent text through generative correction is insufficient without reasoning about semantics using knowledge graphs. Hence, the proposed framework serves as a solid platform for developing intelligent assistants for nutrition, telemedicine applications, and health documentation systems.

Chapter 13

CONCLUSION AND FUTURE SCOPE

13.1 CONCLUSION

This thesis presented Nutri-Correct: Knowledge Graph-Aided Generative Error Correction for Precision Nutrition Speech Recognition, a domain-aware framework designed to improve the reliability of ASR systems in nutrition and healthcare-related conversations. The research addressed a key limitation of general-purpose ASR models: their difficulty in accurately recognizing specialized nutritional terms, food-related expressions, nutrient names, diet plans, and disease-specific dietary instructions.

The proposed framework combines three major components: a baseline ASR model, a nutrition-domain knowledge graph, and a transformer-based generative correction module. The ASR model first generates the initial transcript from speech input. The knowledge graph then provides structured semantic relationships among foods, nutrients, diseases, diet plans, and health outcomes. Finally, the generative correction module uses this domain context to correct semantically incorrect or phonetically similar ASR outputs.

The methodology followed a layered pipeline consisting of audio preprocessing, ASR transcription, nutritional entity recognition, knowledge graph reasoning, generative correction, and transcript validation. This design helped the system move beyond simple word-level correction and focus on preserving the actual nutritional meaning of the spoken content.

The experimental results showed that Nutri-Correct performed better than both the baseline ASR system and the ASR with generative correction model. The proposed system achieved lower Word Error Rate, higher Precision, Recall, and F1-score, and improved Semantic Accuracy. These improvements confirm that integrating structured domain knowledge with generative AI improves both transcription accuracy and semantic reliability.

The major contribution of this thesis is the development of a hybrid neuro-symbolic approach for precision nutrition speech recognition. By combining neural ASR, knowledge graph reasoning, and transformer-based correction, Nutri-Correct provides a strong foundation for intelligent nutrition assistants, telemedicine platforms, dietary monitoring systems, and healthcare documentation tools.

13.2 Major Findings

The study produced several important findings. First, baseline ASR systems are prone to errors when transcribing technical nutrition terms. Words such as glycemic, omega-3,

probiotic, Vitamin D, ketogenic, and saturated fats are frequently misrecognized due to phonetic similarity or lack of domain adaptation.

Second, generative correction improves transcript fluency but is not always sufficient for domain-specific accuracy. A transformer-based correction model can improve grammar and sentence structure, but without nutritional knowledge, it may still select incorrect or incomplete terms.

Third, the nutrition knowledge graph significantly improves semantic correction. By representing relationships among diseases, nutrients, diet plans, and food items, the graph helps identify contextually valid alternatives. This is especially useful for correcting errors such as glass emic $\rightarrow$ glycemic, omega three $\rightarrow$ omega-3, and pro biotic $\rightarrow$ probiotic.

Fourth, the proposed framework improves semantic reliability, which is more important than simple word matching in healthcare communication. The final transcript generated by Nutri-Correct is not only more accurate but also more meaningful for clinical and nutrition-related use.

13.3 Future Scope

Although the proposed Nutri-Correct framework achieved strong results, there are several opportunities for future improvement and extension.

One major area of future work is the development of a larger and more diverse nutrition speech dataset. The current framework can be further improved by including more speakers, accents, regional food names, cultural dietary terms, and real-world telemedicine recordings. A larger dataset would improve the robustness of the ASR and correction modules across different user groups and environments.

Another important direction is multilingual and code-mixed speech recognition. In real healthcare settings, users may speak in regional languages or mix English with local languages. Extending Nutri-Correct to support Hindi, regional Indian languages, and multilingual nutrition conversations would make the system more practical and accessible.

The nutrition knowledge graph can also be expanded in future work. New food items, dietary supplements, regional cuisines, disease-specific nutrition rules, allergies, and medical dietary guidelines can be added to make the graph more complete. A richer knowledge graph would improve the system’s ability to correct rare and complex nutritional expressions.

Future versions of the system may also integrate personalized nutrition profiles. User-specific information such as age, health condition, allergy status, diet preference, and medical history could help the system provide more accurate contextual correction. For example, if the user has diabetes, the system can give more importance to terms related to glycemic index, sugar intake, and carbohydrate control.

Real-time deployment is another important future direction. The current framework can be optimized for faster inference so that it can be used in live telemedicine sessions, mobile health applications, or virtual nutrition assistants. Techniques such as model compression, quantization, caching, and lightweight transformer models can help reduce computational cost.

The framework can also be extended beyond nutrition speech recognition. The same knowledge graph-aided generative correction approach can be applied to other specialized domains such as clinical diagnosis, pharmaceutical consultation, radiology reporting, mental health counseling, and medical transcription.

Another future enhancement is human-in-the-loop validation. Since healthcare transcription can directly affect decision-making, critical corrections may be reviewed by healthcare professionals before being stored in official records. This would improve safety, trust, and clinical reliability.

Multimodal extension is also a promising research direction. Future systems may combine speech input with food images, nutrition labels, wearable device data, and electronic health records. Such a system could support more advanced dietary monitoring and precision nutrition recommendations.

Finally, future research may explore adaptive learning mechanisms. The system can learn from user feedback and correction history to improve over time. If a particular user frequently mentions certain foods, supplements, or medical terms, the model can adapt its correction behavior accordingly.

13.4 Summary

This thesis presented Nutri-Correct, a knowledge graph-aided generative correction framework for improving nutrition-focused speech recognition. The system combines ASR, nutrition-domain semantic reasoning, and transformer-based correction to reduce transcription errors and preserve contextual meaning. The results show that Nutri-Correct performs better than baseline ASR and simple generative correction approaches, making it suitable for applications such as telemedicine, dietary monitoring, virtual nutrition assistants, and healthcare documentation. Future improvements can focus on multilingual support, real-time deployment, larger nutrition datasets, and personalized nutrition integration.

References

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 2019.
- [3] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *Journal of Machine Learning Research*, vol. 21, no. 140, 2020.
- [4] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, “Language models are few-shot learners,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [5] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, “wav2vec 2.0: A framework for self-supervised learning of speech representations,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [6] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen *et al.*, “Deep speech 2: End-to-end speech recognition in english and mandarin,” in *Proceedings of the 33rd International Conference on Machine Learning*, 2016.
- [7] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust speech recognition via large-scale weak supervision,” in *Proceedings of the 40th International Conference on Machine Learning*, 2023.
- [8] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.
- [9] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, “Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.
- [10] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *International Conference on Learning Representations*, 2017.

- [11] H. Paulheim, “Knowledge graph refinement: A survey of approaches and evaluation methods,” *Semantic Web*, vol. 8, no. 3, 2017.
- [12] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. Draft Manuscript, 2024.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [14] C. D. Manning and H. Schutze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [15] World Health Organization, “Healthy diet fact sheet,” <https://www.who.int/news-room/fact-sheets/detail/healthy-diet>, 2020.