

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Projects in Machine Learning

INFO8665 - Spring 2025 - Section 1

Prof. Anasuya Bhima

Group 4 - Section 1

Assignment 1

Due on May 22, 2025

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Table of Contents**

[1.1 Project Overview 4](#_Toc198831606)

[1.1.1 Vision 4](#_Toc198831607)

[1.1.2 Mission 4](#_Toc198831608)

[1.1.3 Purpose 4](#_Toc198831609)

[1.1.4 Strategy 4](#_Toc198831610)

[1.1.5 Tactics 4](#_Toc198831611)

[1.2 Project Relevance of Use Cases 5](#_Toc198831612)

[1.2.1 Food Nutritional Analysis 5](#_Toc198831613)

[1.2.1.1 Description 5](#_Toc198831614)

[1.2.1.2 Problems Addressed 5](#_Toc198831615)

[1.2.1.3 Relevance and Impact 5](#_Toc198831616)

[1.2.2 NLP and Deep Learning for Free-Text Diagnosis 5](#_Toc198831617)

[1.2.2.1 Description 5](#_Toc198831618)

[1.2.2.2 Problems Addressed 5](#_Toc198831619)

[1.2.2.3 Relevance and Impact 5](#_Toc198831620)

[1.2.3 Speech-to-Text for Doctor-Patient Consultation 5](#_Toc198831621)

[1.2.3.1 Description 5](#_Toc198831622)

[1.2.3.2 Problems Addressed 6](#_Toc198831623)

[1.2.3.3 Relevance and Impact 6](#_Toc198831624)

[1.3 Literature Review for Use Cases 6](#_Toc198831625)

[1.3.1 Food Nutritional Analysis 6](#_Toc198831626)

[1.3.1.1 Key Contributions 6](#_Toc198831627)

[1.3.1.2 Model Evaluation and Qualitative Insights 6](#_Toc198831628)

[1.3.1.3 Challenges Faced 7](#_Toc198831629)

[1.3.1.4 Our Solution: Addressing the Gaps 7](#_Toc198831630)

[1.3.2 NLP and Deep Learning for Free-Text Diagnosis 7](#_Toc198831631)

[1.3.2.1 Key Contributions 7](#_Toc198831632)

[1.3.2.2 Model Evaluation and Qualitative Insights 8](#_Toc198831633)

[1.3.2.3 Challenges and Future Work 8](#_Toc198831634)

[1.3.3 Speech-to-Text for Elderly Doctor-Patient Video Consultation 8](#_Toc198831635)

[1.3.3.1 Key Contributions 8](#_Toc198831636)

[1.3.3.2 Challenges and Future Work 9](#_Toc198831637)

[1.3.3.3 Our Solution: Addressing the Gaps 9](#_Toc198831638)

[1.4 Use Case Justification 9](#_Toc198831639)

[1.4.1 Food Nutritional Analysis 9](#_Toc198831640)

[1.4.2 NLP and Deep Learning for Free-Text Diagnosis 9](#_Toc198831641)

[1.4.3 Speech-To-Text Analysis for Video Consultation 10](#_Toc198831642)

[1.5 References 11](#_Toc198831643)

SageCare 2.0

# 1.1 Project Overview

SageCare 2.0 is a user-friendly telemedicine platform designed to connect elderly individuals in Canada with licensed healthcare professionals from the comfort of their homes. The application offers virtual consultations, medication tracking, AI-Powered medical assistant, and personalised health monitoring and recommendation features tailored to the needs of seniors. With simplified navigation, multilingual support, and integration with family caregivers, SageCare bridges the gap between the aging population and accessible healthcare.

The platform prioritizes ease of use, security, and inclusivity. It supports video and voice consultations, appointment scheduling, and electronic prescriptions, and integrates health data from wearable devices. It is available on mobile and web platforms.

## 1.1.1 Vision

To provide easily accessible, AI-powered, remote healthcare support and caregiver integration for elderly people in Canada.

## 1.1.2 Mission

Provide an intuitive telemedicine platform enhanced by AI-driven assistance, smart health tracking and recommendations, that deeply integrates caregivers and upholds the highest standards of privacy, accessibility, and senior-specific usability.

## 1.1.3 Purpose

To bridge the divide in accessible healthcare for aging Canadians by enabling convenient, and reliable remote access to medical support, personalized monitoring, and smart health recommendations.

## 1.1.4 Strategy

Build and integrate three AI-powered features to support accessible healthcare for elders: a computer vision model for food nutritional analysis and recommendation, an NLP/GenAI-based tool for diagnosis and specialty matching based on free-text symptom description, and a speech-to-text system for transcribing video consultations. Ensure accessibility, accuracy, and seamless integration. Ensure data privacy and compliance with healthcare standards.

## 1.1.5 Tactics

Adopt agile development with sprint planning, daily stand-ups, and backlog grooming using Azure Boards. Collect and annotate high-quality food image datasets to train computer vision models. Apply advanced NLP and GenAI to extract embeddings from free-text symptom descriptions and perform diagnoses. Implement real-time speech-to-text transcription using deep learning models. Perform user testing with elderly participants. Collaborate with healthcare experts for validation, and to ensure data privacy and regulatory compliance.

# 1.2 Project Relevance of Use Cases

## 1.2.1 Food Nutritional Analysis

### 1.2.1.1 Description

SageCare 2.0 provides accurate nutritional insights and personalised dietary recommendations by analysing food images uploaded by users.

### 1.2.1.2 Problems Addressed

Unmanaged/inefficient diets tracking and inadequate nutritional knowledge by providing nutritional content and diet recommendations to seniors, and reducing hospitalisation rates and increasing quality of life for seniors.

### 1.2.1.3 Relevance and Impact

**People:** Enhances quality of life for senior citizens by assisting with healthier and personalised diets.

**Industry:** Helps with remote dietary support, modernising how nutritional care is delivered.

**Economy:** Helps to reduce hospital admissions and related costs by identifying and addressing nutritional risks early.

**Culture:** Adapts to diverse dietary preferences and languages, so that all seniors from any part of the world can get nutritional care remotely.

## 1.2.2 NLP and Deep Learning for Free-Text Diagnosis

### 1.2.2.1 Description

SageCare’s AI assistant will help users by analysing free-text symptom descriptions, producing accurate diagnoses, and mapping them to appropriate specialties.

### 1.2.2.2 Problems Addressed

Communication gap between care provider and patient by interpreting symptom descriptions, which helps in faster and more accurate diagnosis and timely specialist referrals.

### 1.2.2.3 Relevance and Impact

**People:** Bridges communication gaps by interpreting symptom descriptions for faster and clearer health communication.

**Industry:** Automates triage and speeds up referrals, supporting efficient care delivery.

**Economy:** Reduces unnecessary clinic or ER visits by streamlining diagnosis.

**Digital Equity:** Makes access to healthcare more intuitive for underserved or marginalised seniors.

## 1.2.3 Speech-to-Text for Doctor-Patient Consultation

### 1.2.3.1 Description

SageCare delivers real-time transcription and captioning of video consultations, creating accessible records for seniors and their caregivers to review and reference.

### 1.2.3.2 Problems Addressed

Aids understanding and patient engagement by transcribing consultations in real-time, improving communication and documentation.

### 1.2.3.3 Relevance and Impact

**People:** Improves understanding and engagement by transcribing consultations in real-time, especially for seniors.

**Industry:** Enhances communication quality and reduces documentation burden for healthcare providers.

**Economy:** Minimises costly medical errors, miscommunication, and unnecessary readmissions.

**Culture:** Allows caregivers to review transcripts, supporting better coordinated care.

# 1.3 Literature Review for Use Cases

## 1.3.1 Food Nutritional Analysis

Recent research shows that machine learning and computer vision are helping to make dietary assessment easier and more accurate. These technologies can now be used for ingredient identification, food type recognition, portion size estimation, and caloric content calculation. Approaches include deep learning (FoodAI), boosting regression algorithms, and advanced image processing methods aimed at enhancing dietary tracking and nutritional evaluation (Chakraborty et al., 2022; Lee et al., 2019; Chaurasia & Saini, 2021; Khandelwal et al., 2023; Rahman et al., 2024; Yuan et al., 2021).

### 1.3.1.1 Key Contributions

Each study offers a different way to improve how we understand food through images. Chakraborty et al. (2022) built a machine learning system that uses food photos from websites like Yelp to predict ingredients and nutrition with 85% accuracy. Lee et al. (2019) created FoodAI, a system trained on over 400,000 images of 756 different foods, which works well with mobile apps.

Chaurasia and Saini (2021) developed CalorieCam, which uses a reference object in the photo (like a spoon or card) to help estimate portion size and calories. Khandelwal et al. (2023) used a special kind of machine learning called boosting to estimate food weight very accurately, with only a small error of 3.73%. Rahman et al. (2024) went further by linking calorie estimates to possible health risks. Lastly, Yuan et al. (2021) built a mobile system that takes real-world food photos and labels them to help train better food recognition models.

### 1.3.1.2 Model Evaluation and Qualitative Insights

Tests show that these models work well when using clear and organized data. For example, Chakraborty et al. (2022) reached 85% accuracy by using a mix of food photos. Lee et al. (2019) pointed out that having lots of different images helps the model work better. Khandelwal et al. (2023) used a method called boosting to guess food weight from photos and got very accurate results.

CalorieCam, made by Chaurasia and Saini (2021), is easy to use and estimates calories by looking at food size in pictures, but it works best when the photo is taken in good lighting and with the right setup. Yuan et al. (2021) created a mobile system that collects real-world food photos, helping the model learn from everyday situations. Still, all models can struggle when pictures are taken in bad lighting, at odd angles, or when foods look very similar to each other.

### 1.3.1.3 Challenges Faced

There are a few main problems with these systems.

* Collecting and labelling lots of food photos from different countries and cuisines takes a lot of time and effort (Chakraborty et al., 2022).
* Some foods, like different kinds of curries, look very similar, which makes it hard for the system to tell them apart (Lee et al., 2019).
* It's still difficult to figure out portion sizes from flat (2D) photos because the pictures don’t show depth (Chaurasia & Saini, 2021).
* Other issues include bad lighting, blurry images, or poor camera quality, which can make the system less accurate (Yuan et al., 2021).
* Moreover, many tools use basic nutrition databases that don’t adjust for local ingredients or recipes (Rahman et al., 2024).
* Lastly, while some systems try to connect food information with health data, this feature is not yet commonly used (Rahman et al., 2024)

### 1.3.1.4 Our Solution: Addressing the Gaps

* To fix the problem of limited food photo data, researchers can use smart learning methods that need fewer labelled pictures or get help from lots of people online to label images (Chakraborty et al., 2022).
* When foods look alike, extra information, like where the user is, what time they’re eating, or what they usually eat can help the system make better guesses (Lee et al., 2019). Also, using both pictures and extra input like text or voice can make the system even more accurate.
* To better estimate how much food is in a picture, phones with depth sensors or using a known object (like a coin or card) in the photo can help (Chaurasia & Saini, 2021).
* Training the model with lots of different types of images can also help it work better under different lighting or camera quality (Yuan et al., 2021).
* Instead of using the same old nutrition charts, future systems can update food info in real time based on local ingredients and recipes (Rahman et al., 2024).
* These systems could also connect to smartwatches or fitness trackers to give personalized advice and feedback based on the user’s health and eating habits (Rahman et al., 2024).

## 1.3.2 NLP and Deep Learning for Free-Text Diagnosis

This study explores the use of NLP and Deep Learning to interpret patient-generated free-text symptom descriptions, suggest possible diagnoses, and intelligently match patients to appropriate medical specialists using semantic similarity. The system has practical applications such as a triage assistant in telehealth portals, a patient intake bot for hospitals and clinics, and a symptom checker on health platforms to guide users toward relevant care.

### 1.3.2.1 Key Contributions

This system combines cutting-edge NLP and deep learning techniques to bridge the gap between patient-reported symptoms and accurate clinical guidance. Leveraging domain-specific transformer models like BioBERT and Bioformer, it encodes free-text symptom descriptions into dense, context-aware embeddings that preserve semantic meaning and handle linguistic variability, such as misspellings, synonyms, and colloquial expressions (Lee et al., 2020; Zhang et al., 2020). These embeddings serve as the foundation for a GenAI-powered classification engine that maps symptoms to potential diagnoses and intelligently recommends suitable medical specialists. The model demonstrates robustness in processing vague or ambiguous inputs, making it highly effective for real-world use cases like digital triage (Soni & Roberts, 2020). To enhance trust and interpretability, the system incorporates the RETAIN architecture which is a reverse-time attention mechanism that identifies the most influential symptoms contributing to each prediction, enabling clinicians and patients to understand the diagnostic rationale (Choi et al., 2016). Furthermore, by aligning patient language with standardized clinical vocabularies through the UMLS ontology, the system ensures interoperability with electronic health records and facilitates seamless integration into clinical workflows (Bodenreider, 2004).

### 1.3.2.2 Model Evaluation and Qualitative Insights

The system was evaluated using simulated patient inputs and benchmark clinical datasets, demonstrating substantial improvements in semantic understanding and classification accuracy. It showed higher recall for informal symptom phrases, better precision in linking symptoms to the appropriate specialists, and fewer misinterpretations of vague or overlapping conditions (Soni & Roberts, 2020). Qualitative analysis revealed strong performance across varied patient language inputs, including slang, misspellings, and non-specific terms such as “burning chest pain” or “pins and needles.” Most remaining errors were attributed to symptoms shared across multiple specialties or rare and emerging conditions that were insufficiently represented in the training corpus.

### 1.3.2.3 Challenges and Future Work

Future enhancements will focus on:

* Ongoing challenges include missed detections in complex or noisy records and errors due to high symptom-diagnosis overlap.
* Temporal modelling of symptom progression.
* Cross-lingual understanding for multilingual patient populations.
* Continuous model fine-tuning using feedback from clinicians and patients.

## 1.3.3 Speech-to-Text for Elderly Doctor-Patient Video Consultation

This section explores real-world implementations of speech-to-text (STT) and voice interactive technologies across healthcare institutions in Ontario, specifically focused on enhancing communication and clinical documentation in elderly care. These technologies are increasingly used in telehealth portals, long-term care settings, and hospital documentation systems to support accessibility, improve physician efficiency, and enhance patient engagement.

### 1.3.3.1 Key Contributions

OntarioMD (2024) launched an AI-powered medical scribe across 150 primary care clinics to automate clinical documentation, reducing physician note-taking time by up to 90% and improving patient interactions. Similarly, Barbaric et al. (2022) developed a voice-enabled version of the Medly heart failure monitoring program using Amazon Alexa, achieving 84% adherence among elderly patients due to its ease of use and accessibility. Baycrest Health Sciences (2021) deployed Alexa-powered telehealth carts in long-term care to combat senior isolation during the pandemic, offering voice activated communication and engagement tools. Trillium Health Partners adopted Dragon Medical One for hospital documentation, cutting charting time by 40% and improving provider focus during consultations (Nuance Communications, 2021). Lastly, CLRI (2022) integrated smart voice assistants in Ontario long-term care homes, enabling seniors to request help, receive reminders, and interact with content handsfree, fostering independence and emotional well-being.

### 1.3.3.2 Challenges and Future Work

Each implementation faced limitations that highlight areas for improvement. OntarioMD’s solution occasionally struggled with noisy environments and domain specific medical terminology. The Medly voice interface faced challenges with speech clarity, accents, and inconsistent recognition across patient demographics. Baycrest residents with cognitive impairments needed assistance using Alexa features, while Trillium’s system was prone to processing lags and confusion with overlapping speech. CLRI’s solution lacked personalization for age-related speech variation and performed inconsistently in high-noise communal spaces. These obstacles emphasize the need for more adaptive, accurate, and elder-friendly STT systems.

### 1.3.3.3 Our Solution: Addressing the Gaps

To address these limitations, our project aims to implement a real-time, domain specific speech-to-text system tailored for elderly patients in telehealth consultations. Key improvements will include:

* Adaptive voice recognition for aging speech patterns and accent variation.
* Robust noise-cancelling and multi-speaker handling in real-time settings.
* Personalized feedback loops to improve accuracy over time.
* Integration with clinical workflows to support documentation and accessibility

# 1.4 Use Case Justification

The selected use cases for this project directly addresses the challenges faced by elderly Canadians in gaining easy access to quality healthcare, which makes them very crucial to the success of the project.

## 1.4.1 Food Nutritional Analysis

Proper nutrition is key to aging well. However, many seniors find it hard to effectively manage their diet due to decline in cognitive ability, limited mobility, or lack of essential nutritional knowledge. Through the use of a computer vision-based food analysis system, seniors can make well-informed dietary decisions that enhance their quality of life and help prevent or manage chronic illnesses. This system offers immediate, accurate nutritional insights and personalised recommendations based on their medical history. This use case proactively promotes good health and reduces hospital visits, providing significant benefits to individuals and the healthcare system as a whole.

## 1.4.2 NLP and Deep Learning for Free-Text Diagnosis

Seniors often face challenges in articulating symptoms clearly or navigating complex healthcare pathways. By employing advanced NLP and generative AI to interpret free-text symptom descriptions, the system bridges communication gaps, accelerates accurate diagnosis, and facilitates timely referrals to the right specialists. This dramatically enhances access to care, reduces diagnostic errors, and mitigates delays that can lead to severe complications.

## 1.4.3 Speech-To-Text Analysis for Video Consultation

Effective communication during consultations is vital but can be hindered by cognitive issues, or language barriers common among elderly patients. Real-time transcription of video consultations ensures accurate clinical documentation and enables patients and caregivers to review conversations afterward, enhancing understanding and adherence to medical advice. This use case improves care quality, fosters patient engagement, and supports inclusive healthcare delivery.

# 1.5 References

1. Chakraborty, I., Mukhopadhyay, R., & Balakrishnan, V. (2022). *Nutrition information estimation from food photos using machine learning based on multiple datasets*. ResearchGate. <https://www.researchgate.net/publication/364083577>
2. Chaurasia, M. A., & Saini, N. (2021). *Estimation of quantity and nutritional information of food using image processing*. ResearchGate. <https://www.researchgate.net/publication/355888149>
3. Khandelwal, R., Kumar, K., & Saini, N. (2023). A novel approach to estimate the weight of food items based on features extracted from an image using boosting algorithms. *Scientific Reports, 13*, Article 47885. <https://www.nature.com/articles/s41598-023-47885-0>
4. Lee, Y. C., Gan, W. S., & Chua, T. S. (2019). *FoodAI: Food image recognition via deep learning for smart food logging*. arXiv. <https://arxiv.org/abs/1909.11946>
5. Rahman, R., Singh, A., & Dev, A. (2024). Enhancing dietary analysis: Using machine learning for food caloric estimation and health risk assessment. *Journal of Food Science*, 89(1), 12–23. <https://doi.org/10.1111/1750-3841.17421>
6. Yuan, Z., Wang, J., & Liu, M. (2021). *An integrated system for mobile image-based dietary assessment*. arXiv. <https://arxiv.org/abs/2110.01754>
7. Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research, 32*(suppl\_1), D267–D270. <https://doi.org/10.1093/nar/gkh061>
8. Choi, E., Bahadori, M. T., Sun, J., Kulas, J., Schuetz, A., & Stewart, W. F. (2016). RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Advances in Neural Information Processing Systems*, 29. <https://proceedings.neurips.cc/paper_files/paper/2016/file/076a0c97d09cf1a0ec3e19c7f2529f2b-Paper.pdf>
9. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics, 36*(4), 1234–1240. <https://doi.org/10.1093/bioinformatics/btz682>
10. Soni, S., & Roberts, K. (2020). Evaluation of symptom extraction from COVID-19 case reports using BERT-based models. *Journal of Biomedical Informatics, 113*, 103661. <https://doi.org/10.1016/j.jbi.2020.103661>
11. Zhang, Y., Chen, Q., Yang, Z., Lin, H., & Lu, Z. (2020). Bioformer: Efficient transformer-based models for biomedical text mining. *Bioinformatics, 36*(23), 6012–6018. <https://doi.org/10.1093/bioinformatics/btaa956>
12. Barbaric, P., Radhakrishnan, A., Goyal, A., & Seto, E. (2022). Development and usability testing of a voice user interface for the Medly heart failure telemonitoring program. medRxiv. [https://doi.org/10.1101/2022.03.01.22271798](https://doi.org/10.1101/2022.03.01.22271798%20)
13. Baycrest Health Sciences. (2021). Enhancing senior care with Amazon Alexa. AWS Public Sector Blog. [https://aws.amazon.com/blogs/publicsector/baycresthealth-sciences-enhances-senior-care-with-amazon-alexa/](https://aws.amazon.com/blogs/publicsector/baycresthealth-sciences-enhances-senior-care-with-amazon-alexa/%20)
14. CLRI. (2022). Using voice technology in long-term care. [https://clriltc.ca/2022/11/using-voice-technology-in-ltc/](https://clriltc.ca/2022/11/using-voice-technology-in-ltc/%20)
15. Nuance Communications. (2021). Dragon Medical One: Trillium Health Partners case study. [https://www.nuance.com/en-ca/healthcare/clinicaldocumentation/dragon-medical-one.html](https://www.nuance.com/en-ca/healthcare/clinicaldocumentation/dragon-medical-one.html%20)
16. OntarioMD. (2024). AI medical scribe pilot: Final report. <https://www.ontariomd.ca/pages/ai-medical-scribe-pilot.aspx>