



Review

GI meets AI: Glycemic index in the age of AI, computational breakthroughs

N.H. Wanigasingha^{1*}, W.G.L. Harshani¹, M.K.A. Ariyaratne¹, T.G.I. Fernando¹, U. Dikwatta¹, U.S. Samarasinghe²

¹Department of Computer Science, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka

²Department of Information Technology, Faculty of Management Studies and Commerce, University of Sri Jayewardenepura, Sri Lanka

ARTICLE INFO

Article history:

Received 24 August 2025

Received in revised form

10 October 2025

Accepted 26 October 2025

Keywords:

Artificial intelligence, Glycemic index,

Machine learning, Deep learning,

Food analytics

*Corresponding author

Email address:

hirushanethni@sjp.ac.lk

DOI: 10.55670/fpll.futech.5.1.10

ABSTRACT

This review explores the avenues for the application of Artificial Intelligence (AI) techniques in Glycemic Index (GI) related research. The necessity of sophisticated technologies to investigate various GI-related studies in food analytics has been established in recent years. AI technologies have emerged as promising approaches to address these challenges. We identified six major AI technologies applied in GI research: Machine Learning, Reinforcement Learning, Deep Learning, Image Processing, Natural Language Processing, and Explainable AI. Some of our findings include: (a) There have been significant improvements in GI-related studies using AI technologies over the past decade. (b) Machine learning algorithms were widely used (c) Many researchers used custom datasets, with the predominance of research originating from North American countries. (d) Identification of limitations and future directions for GI-related studies employing AI technologies. By embracing AI technologies, the field of food analytics is poised for substantial advancements in understanding and managing glycemic responses. Unlike existing reviews that mainly discuss nutritional or clinical aspects of the glycemic index, this study systematically examines the integration of AI and machine learning technologies in GI-related research. It highlights computational breakthroughs, methodological trends, and future directions for intelligent glycemic analysis.

1. Introduction

Of all the challenges technology seeks to address, human health stands as the most critical and universally compelling. Recent evidence shows that Artificial Intelligence is having a significant impact on the healthcare industry, highlighting how important human health has become for technological advancements. In 2024, a survey by the Berkeley Research Group found that healthcare providers and pharmaceutical professionals are increasingly relying on AI to enhance patient care, streamline processes, and transform the delivery of medical treatment [1]. These breakthroughs are particularly evident in diagnostics and personalized treatments, where AI has shown impressive precision. For example, AI models can now assess cancer aggressiveness more accurately than traditional biopsies, according to a 2024 report from the World Economic Forum [2]. At the same time, the global market for AI-driven healthcare is expected to reach \$70 billion by 2032, fueled by advances in AI for drug discovery and medical imaging. In countries like China, AI plays a vital role in optimizing medical resource distribution and enhancing diagnostic accuracy, especially in areas with limited access to healthcare [3]. Additionally, reviews of AI

healthcare studies from 2023 show that fields like radiology and Gastroenterology are experiencing the greatest impact. Looking ahead, experts predict that AI will have a wider influence across many areas of medicine, including administration and education [4]. Among the leading noncommunicable diseases, diabetes is now one of the primary causes of death worldwide, representing an escalating global health threat [5]. Therefore, extensive research has been conducted, and active research is currently underway, to find solutions to prevent and manage it [6-8]. Diabetes is a chronic, noncommunicable disease that occurs when the body is either unable to produce enough insulin or cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood sugar (glucose) levels, which is crucial for providing energy to the body's cells. There are two main types of diabetes:

- Type 1 diabetes: This form is often diagnosed in children and young adults, though it can occur at any age. It happens when the immune system mistakenly attacks and destroys the insulin-producing beta cells in the pancreas, leading to little or no insulin production. People with Type 1 diabetes require lifelong insulin therapy [9].

Abbreviations	
AI	Artificial Intelligence
ANN	Artificial Neural Network
CGM	Continuous Glucose Monitoring
DL	Deep Learning
GL	Glycemic Load
GI	Glycemic Index
HbA1c	Glycated Hemoglobin
IoT	Internet of Things
ML	Machine Learning
NLP	Natural Language Processing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RL	Reinforcement Learning
SHAP	SHapley Additive exPlanations
XAI	Explainable Artificial Intelligence
IP	ImageProcessing

- **Type 2 diabetes:** This is the most common form and is often linked to lifestyle factors such as obesity, poor diet, and physical inactivity. In this condition, the body becomes resistant to insulin, or the pancreas cannot produce enough insulin to maintain normal blood glucose levels. Type 2 diabetes can often be managed with lifestyle changes, but may also require medication or insulin [9].

Diabetes develops when the glucose in the blood remains elevated over time, leading to serious health complications, such as heart disease, kidney damage, nerve damage, and vision problems. Early detection and management through lifestyle changes, medication, and regular monitoring of blood sugar levels are crucial in preventing or delaying these complications. There are several factors, such as Glycemic Index, Glycemic Load, Fiber Content, Carbohydrate Type, Meal Timing and Composition, Physical Activity, and others, that influence blood glucose levels and diabetes. Among them, the Glycemic Index plays a key role. The Glycemic Index is a scale that measures how fast the carbohydrates in different foods raise your blood sugar after you eat them. Foods are ranked from 0 to 100, with higher numbers meaning they cause a quicker rise in blood glucose, while lower numbers indicate a slower, steadier increase.

- **Low GI foods (GI ≤ 55):** These cause a gradual rise in blood sugar, helping maintain stable levels (e.g., most fruits, vegetables, whole grains, and legumes).
- **Medium GI foods (GI 56–69):** These create a moderate increase in blood glucose (e.g., rye bread, bananas, sweet potatoes).
- **High GI foods (GI ≥ 70):** These lead to rapid spikes in blood sugar (e.g., white bread, sugary drinks, and processed cereals).

For people with diabetes, keeping blood sugar levels under control is essential to managing the condition and preventing complications. Eating high-GI foods can cause sudden blood sugar spikes, which can be dangerous for diabetics who may struggle with insulin production or use. On the other hand, low-GI foods help keep blood sugar stable, making it easier to manage diabetes. The GI is especially important in Type 2 diabetes, where lifestyle and dietary choices play a huge role. By choosing lower-GI foods, people with diabetes can prevent sharp rises in blood sugar, reducing the need for insulin and supporting better long-term blood sugar control. Due to the importance of the GI in the management of blood sugar levels, especially for people with diabetes, extensive research has been conducted to explore its various applications. Research related to GI has been

observed to grow rapidly over the past few decades (Figure 1). Studies have connected GI with various fields, investigating its role in diet, health outcomes, and disease management. Researchers have examined how different types of foods affect blood glucose levels, developed predictive models for GI, and explored the benefits of a low-GI diet in preventing and managing diabetes, obesity, and cardiovascular disease. These efforts aim to deepen the understanding of the significance of GI and provide actionable insights to improve health and wellness.

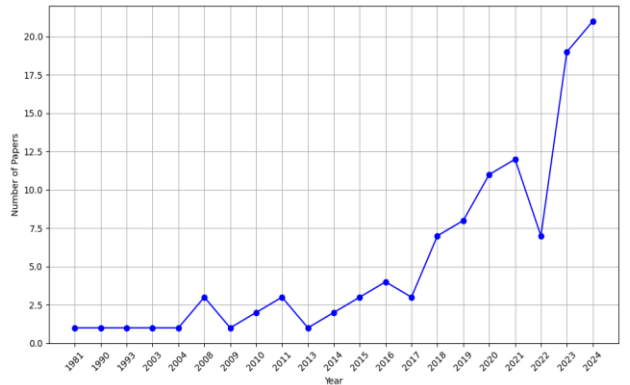


Figure 1. Number of papers by year

It can be quite challenging to determine which technologies are best suited for different applications of the GI and to understand the reasons behind their effectiveness. Whereas a few prior reviews investigated the GI and nutrition science/diabetes research, many of these centered on clinical/dietary applications and did not systematically examine AI/ML contributions in this field. When multiple technologies address the same problem, it becomes even more important to make comparisons. This highlights the need for a systematic review and analysis. The primary goal of this paper is to gather and examine various studies where different approaches have been applied to the GI, with the aim of uncovering useful insights related to human health. By reviewing these approaches in a structured manner, we aim to achieve two main objectives: first, to present and analyze the areas where different technologies have been successfully used with the GI, particularly in predicting GI values; and second, to expand the potential applications of GI for a wider audience.

Here, we provide a comprehensive overview of technologies like deep learning, reinforcement learning, explainable AI, and natural language processing, and how these technologies are utilized to combine viewpoints of food analytics and computational intelligence to predict GI. Unlike prior reviews, which merely described dietary effects, our work systematically charted how AI methodologies evolved for predicting, monitoring, and personalizing nutrition plans for GI. Moreover, our review identifies areas of poor usage of state-of-the-art AI paradigms (e.g., XAI, multimodal data integration) and offers future research prospects for minimizing the gap between food science and computational intelligence.

To enhance the reader’s experience, the rest of this paper is structured as follows: in Section 2, we outline the methodology used to conduct the review. In the next section, we will provide a brief overview of these technologies, offering the reader a foundational understanding of concepts such as Computer Vision, Deep Learning, Food Technologies, Image Processing, Machine Learning, Reinforcement

Learning, Natural Language Processing, Statistical Analysis, and Mathematical Modeling. The research studies, divided into those key concepts, are discussed in Section 4, all related to the use of GI and Information Technology. In Section 5, we discuss the standard datasets used in the selected studies. Finally, we conclude the review with a short discussion.

2. Methodology for systematic review: applying the PRISMA framework

To enhance the transparency of our research reporting, this systematic review was conducted in accordance with the Preferred Reporting Items for systematic reviews and Meta-Analyses (PRISMA) guidelines, ensuring a comprehensive approach to the review [10]. In our initial stages of research, we began by identifying relevant literature using the key terms ‘Glycemic Index’ and ‘machine learning’ to gather foundational insights. In addition to these, we expanded our search scope by incorporating various AI and data prediction-driven keywords such as ‘deep learning’, ‘NLP’, ‘data science’, ‘machine learning’, ‘reinforcement learning’, and ‘statistical mathematics’. Recognizing the importance of contextualizing our findings across different cultural and regional settings, we also included country-specific keywords such as ‘Sri Lanka’, ‘India’, ‘Taiwan’, and ‘Morocco’. Given the research focus on food and health, we employed a range of domain-specific terms such as ‘traditional foods’, ‘breakfast’, ‘diabetes’, ‘glucose’, ‘blood sugar’, and ‘food technology’ to further refine our results.

Table 1. Summary of the searching process

Duration of the search	Used research repositories	Key words	Type of Research works
1st August 2024 to 30th September 2024	Scopus, Semantic Scholar, Scisearch, Scispace, IEEE Xplore, Digital Library, Google Scholar	Glycemic index, Glycemic index predict, machine learning algorithm, Glycemic Index, Glycemic Index for machine learning	Research, Thesis, Review, Articles, Book Chapters, Conference Materials, Reports

We conducted our literature search across well-established academic repositories such as Scopus, ScienceDirect, IEEE Xplore Digital Library, Google Scholar, and Semantic Scholar, ensuring the credibility and diversity of our sources. Our review began in August 2024, concentrating primarily on research articles, review papers, book chapters, and conference proceedings, all published in English. The screening process involved two reviewers. As machine learning technologies gained momentum post-1959, we focused our literature review on publications from 1960 to the present, ensuring that we captured the full breadth of developments in this field. Summarized information is given in Table 1. The reviewed literature was further categorized based on its type, such as journal articles, conference proceedings, and other formats (Figure 2).

The selected papers were categorized based on the key technologies used to derive findings related to the Glycemic Index. The study aims to highlight the significance of the Glycemic Index and assess its value in various research contexts. Our focus includes:

- The diverse technologies applied in Glycemic Index research and how they have been utilized.
- Standard datasets employed in the selected studies.
- Future directions and research opportunities in research based on the prediction of the Glycemic Index.

For the article search, the repositories listed in Table 1 were used. Well-known repositories such as PubMed, Web of Science, and the Cochrane Library were not included, as they often require institutional subscriptions to retrieve full-text papers. Initially, 100 papers were collected in total, and information such as the year, authors, and paper title was added to an Excel sheet. The search query applied was ("Glycemic index" OR "Glycemic index prediction" OR ("machine learning algorithm" AND "Glycemic index") OR "Glycemic index for machine learning"). This search query was utilized to retrieve studies focusing on machine learning algorithms for predicting or analyzing GI. Machine learning was used in the search string, as ML is a key area in AI for predictions and includes deep learning, XAI, and other related technologies.

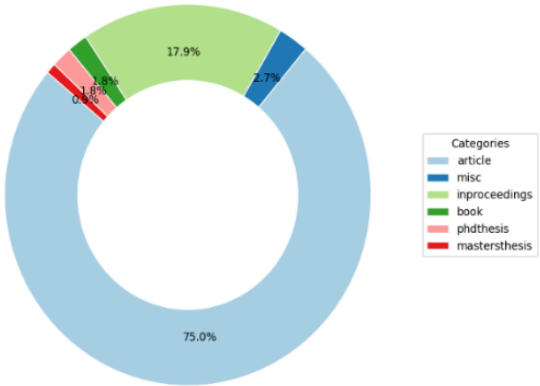


Figure 2. Proportion of papers by category

Of these, 66 articles were directly selected from repositories, while an additional 34 were discovered by reviewing the identified articles or through works by the same authors. Fifteen articles were excluded due to duplication, being written in languages other than English without available translations, or being irrelevant to the study. However, some papers were not related to Computer Science and only described food technology and GI-related content. Such papers were removed after screening. Furthermore, the remaining papers were categorized according to different technologies, and eight major technologies, such as food technology, statistical techniques, NLP, RL, image processing, deep learning, and machine learning, were identified in our survey. Figure 3 provides a detailed summary of the article selection process.

3. Overview of key technologies and concepts

This section provides an overview of the fundamental technologies and concepts pertinent to our literature survey. The discussion will cover key areas including Glycemic Index, Deep Learning, Machine Learning, Food Technology, Image Processing, Reinforcement Learning, Statistical Techniques, and Natural Language Processing.

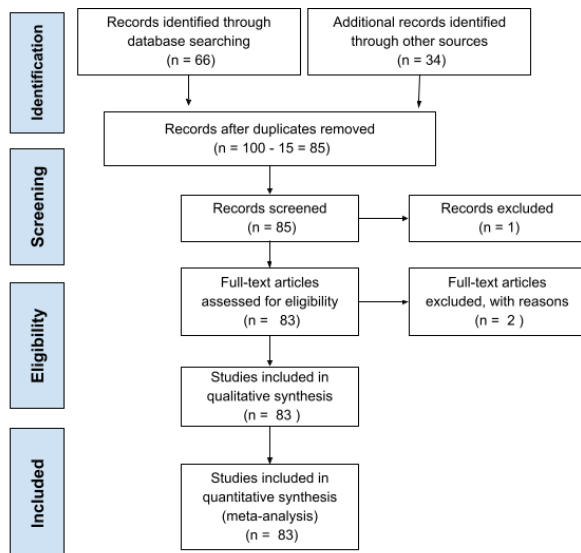


Figure 3. Flow chart of the comprehensive review process based on the PRISMA

Although some areas are closely related, they are discussed and categorized separately to ensure a clearer and more transparent review process. This approach allows for a more structured analysis, making it easier to understand how each technology relates to the Glycemic Index and how it has been applied in various contexts.

3.1 Glycemic index (GI)

Foods and beverages provide the body with energy through carbohydrates, fats, proteins, and alcohol. Among these macronutrients, carbohydrates are the body's preferred source of energy. The Glycemic Index is a system that ranks carbohydrates in various foods and drinks based on their effect on blood glucose levels. Specifically, the GI measures how much and how quickly a particular food raises blood sugar levels after it is consumed. This index typically ranges from 0 to 100, with pure glucose set as the reference point at a value of 100. GI values can be categorized into three ranges:

- Low GI: 55 or less
- Medium GI: 56 to 69
- High GI: 70 to 100

Foods with a high GI value (greater than 70) are rapidly digested and absorbed, causing a rapid increase in blood glucose levels. On the other hand, foods with a low GI value (less than 55) are digested and absorbed more slowly, resulting in a slower and more gradual increase in blood glucose levels. Foods high in refined carbohydrates and sugar are digested more quickly and often have a high GI; whole foods high in protein, fat, or fiber typically have a low GI. Foods that contain no carbohydrates, such as meat, fish, poultry, nuts, seeds, herbs, spices, and oils, are not assigned a GI value [11]. The Glycemic Index is calculated by measuring the blood glucose response of a group of people after they consume a specific food, typically using a standard amount of carbohydrate (usually 50 grams). The Area Under the Curve (AUC) for the blood glucose response over a two-hour period is measured, and the GI is determined by comparing the AUC of the test food to that of the reference food (either glucose or white bread, which are used as reference foods). The formula for calculating the GI is as follows:

$$GI = \frac{(AUC \text{ of test food}) \times 100}{AUC \text{ of Reference food}} \quad (1)$$

To calculate the GI of a meal, one must know the GI values of the individual components. For example, if a meal consists of one cup of cooked brown rice (GI = 50) and one serving of grilled salmon (GI = 25), the total GI for the meal would be 75, which reflects the combined effect of these foods on blood sugar levels. Other factors that affect the GI of a food include its ripeness, cooking method, type of sugar it contains, and the amount of processing it has undergone. Understanding the Glycemic Index and the factors that influence it can help individuals make informed dietary choices, particularly those managing conditions such as diabetes [12].

As GI appeared as one of the promising approaches to identify the levels of carbohydrates, there is extensive literature combining GI with various perspectives, including food technology, nutrition science, and medical research. Although the primary objective of this survey is to explore the technological perspective of GI, it is important to acknowledge relevant past works in the food technology domain. These studies provide valuable insights into the nutritional impact and health benefits of foods with varying GI levels, as well as methods for modifying GI through food processing techniques. By including these works in the GI introduction section, we aim to provide a comprehensive background that contextualizes the technological applications we focus on, even though the main goal of this survey is not to delve deeply into food processing or nutritional studies. This approach helps to emphasize the interdisciplinary nature of GI research, while still keeping our focus on technological advancements and innovations in GI prediction and analysis.

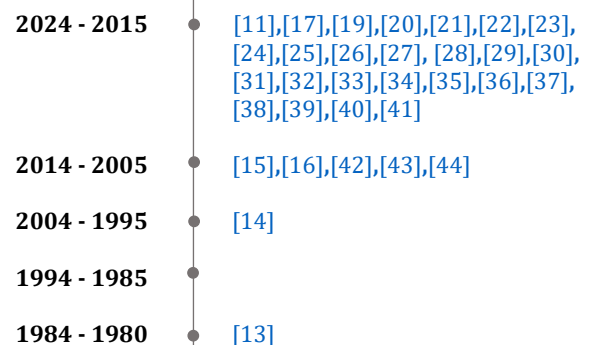


Figure 4. Timeline of research where GI was used in Food Technology

Early such work, mostly focused on calculating the Glycemic Index, by using food with different carbohydrate levels, and measuring blood glucose levels [13-15]. All three studies aim to determine the glycemic impact of foods. The first two studies follow the standard approach of measuring postprandial blood glucose response in humans, while the third study attempts to predict GI using a laboratory-based method. Similar studies, such as [16-19], highlight various approaches to measuring and applying GI, including standardization of measurement techniques, regional adaptations, and predictive modeling. One critical aspect here to address is the variability in methodologies, which can affect the reliability and comparability of GI values across studies. They may lead to inconsistencies in the results. Readers interested in exploring further can refer to Figure 4, which provides an overview of various GI-based approaches applied in food science research.

3.2 Machine learning (ML)

Throughout history, humans have continually developed tools to simplify tasks and meet various needs. The invention of machines was a major leap forward, revolutionizing areas such as transportation, industry, and computing. One of the most significant advancements in recent years is machine learning, a technology that has further extended the capabilities of these machines. Machine Learning focuses on enabling machines to handle data more effectively. In many cases, large datasets are too complex for humans to easily interpret. ML algorithms address this by identifying patterns and extracting meaningful insights from the data. As the availability of vast datasets continues to grow, so too has the demand for machine learning, with industries applying it to uncover valuable information. Unlike traditional programming, where explicit instructions are given, ML allows machines to learn from data and make decisions based on it. This shift has prompted researchers and engineers to develop approaches that allow machines to learn autonomously, without needing detailed programming for every task [45].

It's important to recognize that machine learning is not just about managing data; it is also a crucial part of artificial intelligence. As a subset of AI, machine learning allows systems to discover hidden patterns within datasets, enabling them to make predictions about new data. This ability to generalize from previous experiences is essential for creating systems that can adapt to changing environments. For a system to be considered intelligent, especially in dynamic and unpredictable conditions, it must be able to learn and evolve. If a system can adapt to changes on its own, the designer does not need to foresee and program solutions for every possible scenario. This adaptability is one of the key strengths of machine learning. Machine Learning systems can be categorized based on various criteria. These include:

- How they are trained (e.g., supervised, unsupervised, semi-supervised, self-supervised)
- Whether they can learn continuously in real time (online learning) or process data in batches (batch learning)
- Whether they compare new data points to known data or build predictive models by detecting patterns (instance-based versus model-based learning) [46].

These categories reflect the diversity of machine learning approaches, each tailored to address different types of problems and data environments.

3.3 Reinforcement learning (RL)

Reinforcement Learning is a type of ML in which an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. The agent's goal is to maximize cumulative rewards by choosing actions that lead to favorable outcomes. Unlike supervised learning, where models are trained on labeled datasets, Reinforcement Learning relies on trial and error, allowing the agent to explore and exploit the environment to improve its strategy over time. The core components of RL include:

- Agent: The learner or decision-maker (Single/Multi) [47].
- Environment: The setting in which the agent operates.
- Action: Choices made by the agent to interact with the environment.
- State: The current situation or status of the environment.
- Reward: The feedback the agent receives after taking an action

At each step, the agent observes the current state of the environment, takes an action, and receives a reward based on

the outcome. This process helps the agent learn a policy — a mapping from states to actions that maximizes long-term rewards. Reinforcement Learning is widely used in various applications such as robotics, game AI, autonomous vehicles, and resource management. The Deep Reinforcement Learning (DRL) approach, which combines Deep Learning with RL, has significantly advanced the field by enabling agents to handle high-dimensional, complex environments like images and continuous spaces. The challenge in RL lies in balancing exploration (trying new actions) and exploitation (choosing actions known to yield high rewards), ensuring that the agent learns an optimal strategy efficiently.

3.4 Deep learning (DL)

Deep Learning is a subset of ML that mimics the functioning of the human brain in processing data and creating patterns for decision-making. It uses neural networks with multiple layers to model complex patterns and relationships in large datasets. DL has enabled remarkable advancements in areas such as Computer Vision, Natural Language Processing, and Speech Recognition. This revolutionary approach to machine learning has the potential to reshape various industries, including healthcare, where it is poised to drive significant improvements in medical imaging, disease diagnosis, and drug discovery [48-50]. The distinguishing feature of deep learning is its use of multiple layers of these artificial neurons, often referred to as "deep neural networks". This depth enables the system to automatically extract features from raw data without the need for manual intervention or feature engineering. As a result, deep learning excels in tasks such as image and speech recognition, natural language processing, and even complex game strategies. DL has shown remarkable success in various applications, including self-driving cars, medical diagnostics, and predictive analytics [51,52]. The availability of large datasets, along with significant advancements in computational power (especially through Graphics Processing Units (GPUs) and cloud computing), has contributed to the rapid development and adoption of deep learning techniques. Despite its successes, DL has challenges, such as the need for vast amounts of labeled data and high computational resources. Additionally, the models often act as "black boxes", making their decision-making process difficult to interpret. Nonetheless, the field of deep learning continues to evolve, pushing the boundaries of what machines can achieve in terms of intelligence and automation.

3.5 Image processing (IP)

Image Processing is a technique used to perform various operations on images to enhance their quality or extract meaningful information. It is a valuable tool for analyzing and transforming images, making them more suitable for specific applications or interpretations. Whether the goal is to improve visual quality, recover lost or degraded information, or extract critical details, image processing plays a vital role in fields such as computer vision, medical imaging, satellite imagery, and photography. At its core, image processing relies on computational algorithms that analyze the pixel data in images and apply a series of manipulations to achieve the desired outcome. These algorithms can be designed to address different aspects of an image, such as enhancing colors, sharpening details, reducing noise, or highlighting specific features. The complexity of these operations can range from simple tasks, such as adjusting brightness and contrast, to advanced techniques, like edge detection, object recognition, and image segmentation. There are two main types of image processing:

- Analog Image Processing
- Digital Image Processing

Analog Image Processing involves handling images in a continuous signal form (e.g., photographs or X-ray images) and is often used in traditional photography or medical imaging. Digital Image Processing involves converting images into a digital format and then processing them using computers. This type of image processing has widespread applications in fields like computer vision, medical imaging, remote sensing, facial recognition, and more [53]. Key tasks in Digital Image Processing include image enhancement, which focuses on improving the visual quality of an image, such as sharpening or adjusting contrast, and image restoration, which aims to remove noise or distortions to recover the original image. Image segmentation involves dividing an image into meaningful parts, such as identifying objects within the image, while image compression reduces the file size of an image, preserving its quality. Feature extraction is another crucial task, where key patterns or features in an image are identified for further analysis, often used in computer vision and machine learning. Common techniques in digital image processing include filtering, edge detection, histogram equalization, and Fourier transforms.

3.6 Natural language processing (NLP)

Natural Language Processing is a field of artificial intelligence that focuses on the interaction between computers and humans through natural language. The goal of NLP is to enable computers to understand, interpret, and respond to human language in a valuable way. This field involves several tasks, including:

- Text processing: This includes tokenization, stemming, lemmatization, and part-of-speech tagging to prepare text for analysis.
- Sentiment analysis: Determining the emotional tone behind a series of words, used in applications like customer feedback analysis.
- Named entity recognition (NER): Identifying and classifying key entities in text (e.g., names of people, organizations, locations).
- Machine translation: Translating text from one language to another, as seen in tools like Google Translate.
- Speech recognition: Converting spoken language into text, used in virtual assistants like Siri and Alexa.
- Text generation: Creating coherent and contextually relevant text, such as chatbots or story generation.
- Question answering: Developing systems that can answer questions posed in natural language, often used in customer support and search engines.

Machine learning techniques are applied to textual data similarly to how they are utilized in other forms of data, including images, speech, and structured datasets. Supervised machine learning techniques, such as classification and regression methods, play a significant role in various NLP tasks. For instance, an NLP classification task might involve categorizing news articles into specific topics, such as sports or politics. Conversely, regression techniques can predict numeric values, such as estimating the price of a stock based on discussions in social media. Additionally, unsupervised clustering algorithms can be employed to group together similar text documents. Any machine learning approach for NLP, whether supervised or unsupervised, can be characterized by three common steps: extracting features from text, utilizing the feature representation to learn a model, and evaluating and refining the model [54].

3.7 Continuous glucose monitoring (CGM)

Continuous Glucose Monitoring is a technology used to track glucose levels in real-time throughout the day and night. It involves a small, wearable sensor inserted under the skin, typically on the abdomen or arm, which measures interstitial glucose levels at regular intervals. These readings are transmitted to a receiver or smartphone, providing users with a continuous stream of glucose data. This technology not only tracks glucose trends but also provides alerts for hypoglycemia or hyperglycemia, allowing proactive management of diabetes. CGM is widely used in diabetes management, particularly for individuals with type 1 and type 2 diabetes, to improve glycemic control and reduce the risk of complications.

3.8 Explainable AI (XAI)

Explainable AI refers to artificial intelligence systems designed in a way that their decisions, predictions, and behaviors can be understood and interpreted by humans. The goal of XAI is to make AI more transparent, trustworthy, and accountable, especially in critical applications such as healthcare, finance, and autonomous systems [55]. Here's a breakdown of the concept: Key Aspects of Explainable AI:

- Transparency: The AI model provides insights into how it processes input data to produce its output. This might involve revealing the structure of the model, the logic behind decision-making, or the importance of features in a prediction.
- Interpretability: The results or decisions made by the AI are presented in a way that humans can understand. For instance, instead of presenting a decision as a "black box" output, the AI explains why a specific choice or prediction was made.
- Accountability: XAI systems allow developers, users, and regulators to scrutinize and validate the AI's decisions, ensuring ethical and fair outcomes.
- Trustworthiness: By making AI systems understandable, XAI builds confidence in their use, especially in high-stakes scenarios where decisions impact lives.

Traditional AI models, particularly those based on deep learning, often operate as "black boxes," meaning their internal workings are complex and not easily interpretable. XAI addresses this limitation by providing insights into how and why an AI system arrives at specific outcomes, enabling users to trust and validate the model's predictions. By fostering transparency, XAI enhances collaboration between humans and AI while reducing biases and errors in AI applications. Major Techniques in Explainable AI:

- Intrinsic interpretability: Some models, like linear regression or decision trees, are inherently interpretable because their structure is simple and their outputs are easy to trace back to inputs.
- Post-hoc explanations: For complex models like deep neural networks, techniques are applied after the model has made predictions to explain the output. Common methods include:
- SHAP: Quantifies the contribution of each feature to a prediction.
- LIME (Local Interpretable Model-agnostic Explanations): Builds interpretable models around individual predictions.
- Feature importance analysis: Highlights which input features were most influential in a decision.
- Visualization tools: For example, heatmaps in computer vision models show which parts of an image influenced a decision.

SHAP, which stands for Shapley Additive Explanations, is an interpretability method grounded in Shapley values and was introduced by Lundberg and Lee. This approach has become widely adopted in machine learning to explain model outputs by quantifying the contribution of each feature to the final prediction, making complex models more transparent and interpretable. SHAP introduces two key innovations: (1) the identification of a novel class of additive feature importance measures and (2) theoretical results demonstrating the existence of a unique solution within this class that satisfies a set of desirable properties, such as local accuracy, consistency, and additivity [56]. These properties ensure that the feature attributions are both fair and reliable. This framework unifies six existing methods under a common theoretical foundation, offering a more robust and coherent approach to feature importance. Notably, it addresses shortcomings in several recent methods within this class that fail to satisfy the proposed desirable properties. SHAP's interpretability extends beyond theoretical rigor, providing practical tools like visualization plots that enhance understanding of how features influence individual predictions and overall model behavior, thereby empowering users in high-stakes domains like healthcare, finance, and law.

3.9 Analysis of Variance (ANOVA)

ANOVA is a statistical method used to determine whether there are significant differences between the means of three or more unrelated groups. Developed by Ronald Fisher, ANOVA extends the capabilities of the t-test, which is limited to comparing only two groups. The primary function of ANOVA is to analyze how different categorical independent variables influence a continuous dependent variable by partitioning the total variance observed into components attributable to different sources.

ANOVA operates under several assumptions: the samples must be independent, the dependent variable should be normally distributed, and the variances among the groups should be approximately equal (homogeneity of variance). The test statistic for ANOVA is the F-value, calculated as the ratio of variance explained by the treatment (between-group variance) to the variance due to random chance (within-group variance). A significant F-value indicates that at least one group mean differs from the others, although it does not specify which means are different; post-hoc tests are required for that purpose.

There are various forms of ANOVA, including one-way ANOVA, which examines a single independent variable with multiple levels, and two-way ANOVA, which assesses the impact of two independent variables and their interaction on a dependent variable. This flexibility makes ANOVA a powerful tool for researchers looking to understand complex relationships in their data [57].

3.10 Tukey's Q method

The Tukey's Q method, also known as the Tukey HSD (Honestly Significant Difference) test, is a statistical tool used to compare the means of different groups after conducting a one-way ANOVA. It helps identify specific group differences when ANOVA indicates significant variance among groups but does not specify which groups differ. The Tukey HSD test calculates a statistic known as 'q', which is then compared to critical values from the Studentized range distribution. If the calculated 'q' exceeds the critical value, it indicates a significant difference between the group means. This method is particularly useful because it controls the experiment-wise error rate, reducing the likelihood of Type I errors that can

occur when conducting multiple t-tests. By focusing on the largest pairwise differences in means, Tukey's HSD provides a conservative approach to identifying significant differences while maintaining statistical rigor. Researchers often rely on statistical software to perform these calculations due to their complexity, but understanding the underlying steps, such as calculating overall and group means, sum of squares, and mean squares, is crucial for interpreting results accurately. Overall, Tukey's Q method serves as an effective post-hoc analysis tool in research studies where multiple group comparisons are necessary [58].

3.11 T-test

The t-test is a statistical hypothesis test used to determine whether there is a significant difference between the means of two groups or between a sample mean and a known population mean. It is particularly useful when dealing with small sample sizes (typically $n \leq 30$) and when the population standard deviation is unknown. There are three main types of t-tests: the one-sample t-test, which compares a sample mean to a known value; the independent t-test, which assesses the means of two independent groups; and the paired t-test, which evaluates means from the same group at different times or under different conditions. The t-test calculates a t-value based on the difference between group means and their variability, which is then compared to critical values from the t-distribution to assess statistical significance. This method helps researchers understand whether observed differences are likely due to chance or reflect true differences in the populations being studied [59].

4. Harnessing technology in Glycemic index research: innovations and insights

Here, we focus on the main objective of this study: to enlighten the reader on how technology can be effectively utilized to tackle various challenges associated with the Glycemic Index. For convenience, we focus on specific technological aspects one at a time and discuss studies that have utilized them, either fully or partially, to address challenges related to the Glycemic Index, as illustrated in Figure 5, which presents a timeline of research where GI has been used with different AI-based technologies.

4.1 Role of machine learning in research related to the Glycemic index

Machine learning is a branch of AI and computer science that focuses on using data and algorithms to enable AI to imitate the way that humans learn, progressively improving its accuracy. Overall, ML is used to make decisions based on data. By modeling the algorithms on the basis of historical data, they find the patterns and relationships that are hard for humans to detect. These patterns are now further used for future reference to predict solutions to unseen problems in different domains. Biology and Food technology are some of the key domains that have used ML. Numerous studies have focused on the GI, exploring ML techniques to achieve diverse objectives. Given the increasing prominence of ML in GI-related research over the years (Figure 6), it is worth highlighting this category as a central focus of the discussion.

The earliest record in our repository originates from 2017: Glycaemic Index Prediction: a Pilot Study of Data Linkage Challenges and the Application of Machine Learning [63]. They present a ML-based model that predicts the GI of foods based on the biochemical properties. They employed a multiple regression model, which bases its prediction on a weighted linear combination of the independent input variables. These variables include: (1) water (% of mass), (2)

energy [kJ per 100g], (3) protein (% of mass), (4) total carbohydrates (% of mass), (5) sugars (% of mass), (6) fiber (% of mass), and (7) lipids (% of mass). They used GI as a target variable. In addition, the standard five-fold cross-validation methodology has been used for training and testing. Furthermore, they highlighted the need for the adoption of a common standard for recording different types of information on foods so that this information can be cross-linked automatically and without ambiguity.

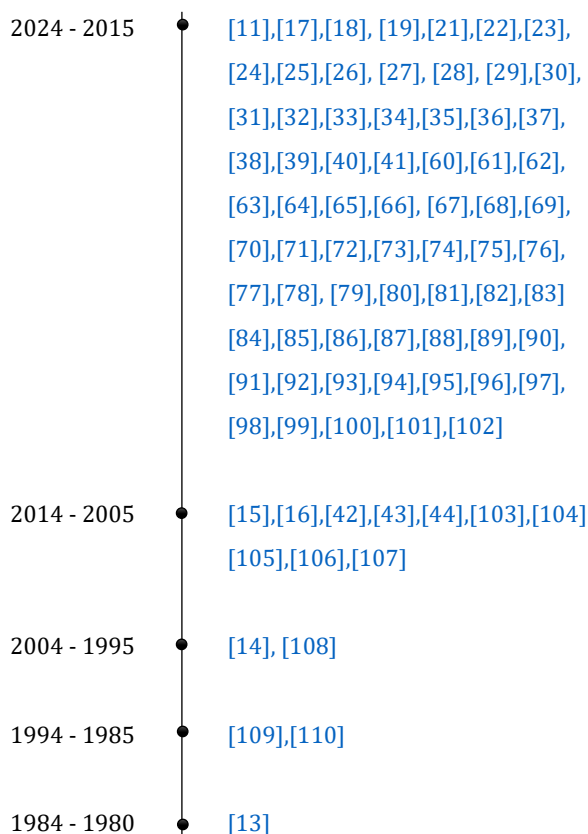


Figure 5. Timeline of research where GI were used with different AI-based technologies

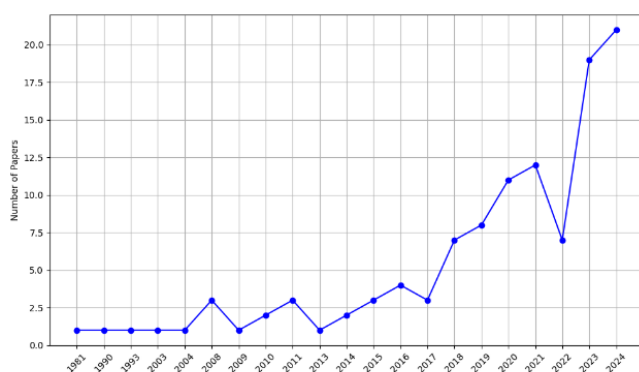


Figure 6. No. of papers by year (overall and ML category)

Colorectal Cancer (CRC) is recognized as the most preventable cancer worldwide. The GI has been used to assess healthy eating in association with CRC. The researchers explored predictors of the Healthy Eating Index (HEI) and GI in multi-ethnic CRC families. In this study, GI served as one of the key measures of diet quality, helping to

realize its role in managing CRC risk. Predicting GI and HEI is a major challenge in the real world. In this study, they employed machine learning techniques for validating and predicting HEI and GI. The validation procedures included the use of ensemble methods and generalized regression models, Elastic Net with Akaike's Information Criterion with correction (AICc), and Leave-One-Out cross-validation methods. Generalized Regression (GR) models were employed with Elastic Net and validation methods (AICc and Leave-One-Out cross-validation) to minimize over-fitting and to optimize prediction models for both HEI and GI. AICc validation and LOO cross-validation methods are effective methods for small data sets. Results obtained revealed that further studies with larger datasets and diverse samples are needed to emphasize findings in diverse groups [66].

In the same direction, another study aimed to validate predictors of healthy eating metrics: HEI, GI, and GL across various modern diets. The researchers examined daily dietary data from 131 diets classified into four primary groups (liquids, convenience foods, ethnic diets, and smoothies) to assess the impact of various diets on GI, GL, and HEI scores. Logistic Regression (LR) was used as a baseline model for initial predictions in this study. Also, Elastic Net Generalized Regression was applied for improved accuracy and to handle complex data with multiple predictors, and Elastic Net combines ridge and lasso regression, employed to minimize over-fitting while allowing selection of relevant predictors [31]. In another work, Partial Least Squares (PLS) regression was applied to predict the GI and amylose content from Near-Infrared (NIR) Spectroscopy to analyze rice characteristics spectral data. Near-Infrared Spectroscopy (NIRS) is a primary technology that is used to collect spectral data from rice varieties in the wavelength range of 740–1070nm. The model mainly used NIR as features to predict the GI and amylose content. Random Forest (RF) was employed for rice varieties classification, while the Principal Component Analysis (PCA) was used for dimensionality reduction to enhance the classification performance. Also Linear Discriminant Analysis (LDA) for classifying rice samples according to parboiling treatments was employed. This research also explored the use of a portable NIR sensor for real-time, on-site evaluation, creating predictive models for identifying rice varieties and estimating amylose content [77].

While the glycemic index may not be one of the most critical factors, it serves an indirect yet important role in certain situations. Intensive insulin treatment is a standard of care for tight glycemic control in people with diabetes to prevent or delay long-term complications of diabetes mellitus. However, insulin therapy may trigger lethal hypoglycemia, and these results show that a number of subjects are prevented by this risk factor from attaining and sustaining near normoglycemia. The forecasting of postprandial hypoglycemia is considered to improve the CGM technology for persons with diabetes using insulin. The GI can also change the Rate of Glucose Increase (RIG), which is a predictor of hypoglycemia. The researchers did not use specific GI values, but used glucose profiles characteristic of high GI foods to help simulate hypoglycemia dangers. This study employed four machine learning models to predict hypoglycemia, including RF, Support Vector Machine (SVM) (with both linear and radial basis functions), K-Nearest Neighbor (KNN), and Logistic Regression. Among the four models, the Random Forest was the best with an average of AUC 0.966 and was quite good at predictive ability. The researchers in this study plan to explore evaluation of their

algorithm on a prospective patient population to clearly establish the clinical use of this system [71].

Another study pursued the usage of machine learning approaches for estimating short-term blood glucose levels of Type 1 Diabetes (T1D) patients. T1D is an autoimmune disease in which the pancreas releases little or no insulin. The traditional method involves having patients administer insulin shots to themselves on a number of occasions each day. In this work, the focus was geared more towards proper identification of suitable ML models depending on glycemic status (hypoglycemia, normoglycemia, and hyperglycemia). This study also addressed the challenge of imbalanced data, which occurs when T1D patients spend the majority of their time in the normoglycemic range. GI affects how the study encompasses the dynamics of glucose related to food intake; it has an indirect effect on how blood glucose is managed and anticipated. Ten different machine learning and deep learning algorithms were used for training regression models, including Linear Support Vector Regression (SVR), Lasso Regression, Decision Trees, Random Forest, KNN, Multilayer Perceptron (MLP), and Gradient Boosting. The prediction models used 24 consecutive CGM sensor measurements obtained every five minutes over a 120-minute period, with the blood glucose level set 30 minutes after the last reading. Overall, the work provided a detailed overview of ML strategies for blood glucose prediction, highlighting the necessity of tailored models and oversampling techniques when dealing with imbalanced glycemic data [73].

Similarly reference [75] aimed at establishing the feasibility of estimating blood glucose levels of T1DM patients through constrained platforms like smartphones and Raspberry Pi. As the research objective, the real-time glucose prediction from aggregated data streams was performed with the help of the Machine Learning models implemented on the local devices excluding the cloud computing that can provide predictions even if there is no internet connection. Data from CGM was used to develop univariate models in which forecasts are based on preceding glycemic values. Despite the fact that the model did not use GI values directly it relied on the CGM records which captured patterns of effects of high-GI foods. The model was able to use the observed glucose changes and calculate the future glucose values without an accurate GI of all foods in the meal. This study employed ML models such as RF, SVM, and Autoregressive Integrated Moving Average (ARIMA). SVMs also perform computations efficiently on devices with restricted computational capabilities, like Smartphones and Raspberry Pi, and this was most evident with tiny sliding window computations. Heading in the same way, another research presented an ensemble machine learning approach to detect unannounced meals (UAM) in type 1 diabetes patients. Maintenance of postprandial glucose level is another considerable barrier in the management of T1D. CGM and hybrid Automated Insulin Delivery (AID) systems depend on patients to alert and predict carbohydrate (CHO) intake, which is frequently ignored, especially in adolescents. This study identified that missing meal announcements increase time to insulin administration, increase postprandial blood glucose variability, and reduce overall glycemic control. Here also, GI is indirectly used for glycemic control by aiming to detect UAM and improve glucose management. The ensemble model was built combining the predictions of three ML models: Artificial Neural Network (ANN), RF, and Logistic Regression. A total of 14 features were used, of which 12 are based on CGM readings and the remaining two are based on insulin data [97].

Maintaining glycemic control in children with type 1 diabetes is a challenging task in clinical practice. A study has focused on predicting glycemic control (measured by glycated hemoglobin levels (A1C)) in children with Type 1 Diabetes using machine learning algorithms. Binary Logistic Regression was applied to predict the probability of poor glycemic control, identifying significant predictors such as A1C at onset and ketoacidosis episodes. GI was not directly used in this study. In the model, the initial A1C level was an essential covariate because it captures historical glycemia that could be influenced by GI. High A1C could indirectly capture patterns associated with frequent intake of high-GI foods if they led to sustained high glucose levels over time. This study used 15 features, including demographic and socioeconomic Factors like family income, living environment, maternal and paternal education. Overall, the indirect impact of GI on long-term glycemia regulation could be conferred in baseline A1C and lipid profile captured by the model [90].

The main types of diabetes are Type 1, which comprises 5-10% of diabetes patients. According to the International Diabetes Federation (IDF, 2017), Type 1 diabetes is caused by an autoimmune reaction in which the body's immune system attacks the insulin-producing beta cells of the pancreas and causes the body to produce very little or no insulin; hence a diabetes patient is required to administer insulin on daily basis to maintain the recommended target blood glucose level. Type 2, which was formerly well known as non-insulin dependent, and which comprises 90-95% of diabetes patients, is caused by the human body's inability to fully respond to insulin (IDF). Focusing on Type 2 diabetes patients, a study aimed to develop a personalized food recommendation system. GI was the core metric to classify foods into high, medium, and low categories, helping the system to recommend foods with a lower likelihood of causing blood sugar spikes. They considered GI, GL, and carbohydrate content as features of the study. These features focused on the glycemic impact and carbohydrate content of food items to classify foods into categories. Naive Bayes Classifier was selected as the primary machine learning model to recognize patterns in glycemic response. Foods with known GI values were sourced from an international Glycemic Index database, which supports the model in recommending foods based on how they are likely to affect blood glucose [85].

In tandem, another study tried to create a model predicting the GI of fruits. They employed a combination of DL and ML methods to predict the GI of fruits. The output of the consequent module can identify three fruits, including Apples, Bananas, and Oranges, but the GI is predicted only for bananas. They used bananas for GI prediction because the GI of bananas varies a large amount according to ripeness cycles and is a simple food for testing. The researchers used Glycemic Load (GL) to assess the overall Glycemic Index of the fruit. A Convolutional Neural Network (CNN) was used to determine the type of food. Then, a simple binarization model was used to characterize the measurements of the fruit length. They used the THRESH_BINARY function from the OpenCV library to binarize the image of the fruit. Two linear regression models were then applied to the prediction length to derive the GL and Carbohydrate content of the fruit, respectively. The pretrained machine learning models were invoked based on the ripeness of the fruit to predict the GL and the carbohydrate content of the fruit. Both linear regression models took length as an input parameter. Once the carbohydrate content and GL values were in, the data

were plugged into the glycemic index formula to determine the Glycemic Index. Clearly, this research employed CNN for fruit classification and linear regression for GL prediction to help Pre-Diabetes patients select healthy fruits according to their glycemic response [30].

In some research, the GI is used as a feature to evaluate and optimize dietary impacts on blood glucose levels. One such study focused on developing a machine learning model to predict postprandial blood glucose responses in patients with Gestational Diabetes Mellitus (GDM). A gradient boosting algorithm was used, and they extracted data from various resources like mobile app diaries, CGM, and individual patient characteristics to improve blood glucose control. The GI was used as one of the features in this model. The model also used GL, which combines the GI and the quantity of carbohydrates consumed to estimate the impact of food on blood glucose levels. They used random grid search and cross-validation for hyperparameter tuning. Values from the SHAP method were employed to understand the influence of different features on model prediction. This study used explainable AI methods through SHAP to evaluate the impact of features on the predictions [78].

On the same line, reference [39] employed GI as a feature. The study focused on increasing the prediction accuracy of the Postprandial Glycemic Responses (PPGR) in women with Gestational Diabetes Mellitus by incorporating GI and GL. CGM & food diaries of pregnant women were used for the development. They compared models with and without GI/GL data to determine whether or not GI/GL data can improve the prediction of PPGR. The study was focused on finding the effect of GI/GL information to enhance the prediction of PPGR outcomes. This study used a total of 124 participants (90 GDM & 34 controls) from the prospective multi-center GEM-GDM (Genetic and Epigenetic Mechanisms of Developing Gestational Diabetes Mellitus) clinical trial. Each of 1,489 meal records was associated with glucose measurements. GI values were derived directly from the University of Sydney database (available until October 2023) [110] and matched to foods in the DiaCompanion app food database. ML methods such as linear regression and regularized regression (Lasso, Ridge, Elastic-Net, LARS Lasso, Orthogonal Matching Pursuit) were used in the study. Another study addressed the need for accurate, automated dietary monitoring by analyzing the Post-Prandial Glucose Response (PPGR) to predict meal macronutrient content. They also used GI as a key factor influencing PPGR, especially for carbohydrate-heavy foods. But the model did not directly calculate GI values. Instead, it used CGM-based PPGR data to infer macronutrient compositions of meals. The model implicitly considered the impact of carbohydrates (via PPGR patterns) in estimating these compositions, which are indirectly related to GI effects. They evaluated the sparse-coding approach against two baseline techniques: (1) ridge regression (RR), as a representative of regularization methods, and (2) a Nearest-Neighbor classifier operating in a Linear Discriminant Analysis subspace (LDA-kNN), as a representative of distance-based classifiers [82]. In another similar work, machine learning models were used to predict the progress of the glycemic values of six patients with diabetes. Eight different algorithms were compared, i.e., ANN with Multilayer Perceptron, Probabilistic Neural Network (PNN), Polynomial Regression, Gradient Boosted Trees Regression, Random Forest Regression, Simple Regression Tree, Tree Ensemble Regression, and Linear Regression. The algorithms were classified based on the ability to minimize four statistical errors, namely: Mean Absolute Error, Mean

Squared Error, Root Mean Squared Error, and Mean Signed Difference. Direct use of GI is not presented. Instead, it aimed to predict overall glycemic status using historical glucose readings from patients [89].

In reference [88], researchers used wearable device data to attempt to predict future glycemic control among adults with prediabetes. In this study, they have 16 features, including physical activities, heart rate, and sleep. They aimed to predict longitudinal continuous changes in hemoglobin A1C and assess worsening, improvement of glycemic control among non-diabetic and prediabetic adults using various features obtained from wearables. Directly calculated or predicted GI was not specifically measured in the study. Instead, changes in glycemic control were monitored using hemoglobin A1C levels (which reflect long-term blood glucose levels rather than short-term postprandial responses to foods).

Diabetic Retinopathy (DR) is one of the major complications of diabetes. A recent study integrates ML models to predict the risk of diabetic retinopathy [99]. SHAP was established to increase the accuracy of risk prediction for diabetic retinopathy, explain the rationality of the findings from model prediction and improve the reliability of prediction results. The features that used in the model were extracted from a diabetes complication dataset. The CatBoost model was employed and optimized for the prediction task. The GI itself is not directly used as a feature in this study; instead, this study focused on glycemic measures like glycated hemoglobin (HbA1c) and fasting blood glucose (GLU_2H) as significant indicators in diabetic retinopathy risk assessment. GI measures how quickly carbohydrate-containing foods raise blood glucose, which is particularly relevant for diabetes management. ML has yielded stunning success in predicting the importance of diabetes risk using health indicators and pattern analysis.

"Diabetes Prediction Using Machine Learning Classification Algorithms" [111] reveals the effectiveness of several classification algorithms, including SVM, Extreme Gradient Boosting (XGB), Decision Trees (DT), and RF, in predicting diabetes. Implementing ML models in GI predictions would lead to identifying the effect of foods on blood glucose levels, thus helping to formulate dietary recommendations and support glucose management within both diabetic and pre-diabetic populations. The glycemic variability metric is an additional measure available to the clinician as a potentially useful tool for estimating overall glycemia. Here, the researchers employed a new measure, Consensus Perceived Glycemic Variability (CPGV), for how much a patient's blood glucose levels fluctuate, as evaluated by doctors. ML models were used to forecast blood glucose levels for 30 and 60 minutes in the future. This study focused on blood glucose variability and predicting blood glucose levels based on CGM data, which are different from the GI. Glycemic variability measurement and blood glucose prediction were modeled with 26 features. The CPGV metric was created using Linear Regression, whereas the future glucose levels were predicted using SVR and MLP [107].

"Application of Machine Learning Algorithms to Predict Uncontrolled Diabetes Using the All of Us Research Program Data" [91] used ML techniques to effectively predict uncontrolled diabetes using clinical markers such as serum electrolytes, body weight, and other physiological indicators. While the GI was not applied directly as a feature in this study, it demonstrated the use of ML as a potential tool for diabetes control using various predictors of glycemic status. This gives GI prediction an additional dimension, making it possible to

adapt ML not only to the amounts of food we are consuming but also to provide a more holistic perspective on factors related to the regulation of blood glucose levels. This suggested the need for including GI as a supplemental dietary characteristic to improve personalized diabetes management in such models. To emphasize this, they employed techniques such as RF, Extreme Gradient Boosting (XGBoost), Logistic Regression, and Weighted Ensemble Model (WEM) that combines RF, XGBoost, and LR models. A similar study explored the relationship between noninvasive wearables and glycemic metrics and demonstrated the feasibility of using non-invasive wearables to estimate glycemic metrics, including hemoglobin A1c and glucose variability metrics [87]. The application was in real-time for people with pre-diabetes or high-normal glucose. ML approaches such as random forest models were used to estimate HbA1c levels. The study did not directly address the concept of GI but accentuated the potential for continuous, noninvasive monitoring of glycemic metrics. For this purpose, digital biomarkers were used to obtain physiological data, including skin temperature, electrodermal activity, heart rate, and accelerometry. The metrics correlate with glucose levels. Along with this, they provide insight into glucose variability without requiring traditional invasive measurements like blood samples.

The maintenance of glycemia in range is one of the biggest challenges in the treatment of patients with diabetes. In a recent comparative study, the focus was to compare different diabetes management therapies, with an idea of the effectiveness of a machine-learning-trained closed-loop artificial pancreas system. Diabetes Type 1 (DT1) patients' data were used in this study. They tried to measure improvements in glycemic control when switching from traditional therapies to the ML-trained system. The Low Blood Glucose Index (LBGI) and High Blood Glucose Index (HBGI) were employed as blood glycemic indices that measure the likelihood of hypo and hyperglycemia events, respectively. Specific features used to predict glycemic control include time in range (TIR%), mean and median blood glucose levels, percentages of hypoglycemia and hyperglycemia, LBGI, HBGI, and glycated hemoglobin. The closed-loop artificial pancreas algorithm was trained using machine learning techniques to optimize insulin dosage based on collected glucose data. Two ML regression techniques were tested by them in the R environment. This study concluded that "hybrid closed-loop" artificial pancreas with control algorithm trained with machine learning technology provides very significant improvement in glycemia control compared to the multi-daily injection (MDI), insulin pump without CGM, and sensor-assisted insulin pump therapies [76].

A study investigated the relationship between glycemic control, hyperhomocysteinemia, and microalbuminuria, which is an early marker of kidney and cardiovascular complications in diabetics [74]. They defined glycemic control by Fasting Blood Glucose (FBS) and glycosylated hemoglobin to assess their relations with microalbuminuria. The analysis for the association of urinary microalbumin with age, gender, HbA1c, FBS, and diabetic status was performed by using a multiple linear regression model. In another recent study, the GI is used as a conceptual foundation for understanding the glycemic impact of food. Estimation of the glycemic impact of cooking recipes using online crowdsourcing and machine learning is a novel approach. This study focused on glycemic impact, which refers to how a recipe affects blood sugar levels post-consumption. The

researchers considered the sugar-to-fiber (S/F) ratio as a proxy for the glycemic impact during the initial stages of recipe selection and modeling. Several ML models were developed, including Logistic Regression and LightGBM. Furthermore, the study experimented with various NLP techniques, such as bag-of-words (BoW) and word embeddings (e.g., Word2Vec, GloVe, FastText). They used both textual features and 20 nutrition features. Textual features included recipe titles, ingredients, and cooking directions. Nutrition features included carbohydrates, protein, fat, and dry weight. As limitations, they highlighted that the models trained on small datasets are prone to overfitting [72].

In addition, there were several review papers related to ML and GI. They were helpful in identifying the existing methods and the gaps related to the field. One of the recent studies has investigated the role of machine learning in nutrition science and diabetes management [102]. The article reviewed machine learning methods for screening Food Bioactive Compounds (FBCs) with bioactivities like antioxidant, anti-inflammatory, antihypertensive, and hypoglycemic effects. It presents an ML model development process, covering data preparation, molecular representation, ML algorithm selection, and evaluation methods. SVM, RF, and KNN are commonly used for initial screening. And also CNN and Recurrent Neural Networks (RNN) are applied to complex data. This study further accentuates the importance of model interpretability. Techniques such as feature importance were used to understand the contribution of each molecular descriptor to the prediction of bioactivity. There was another synthesis that examined the role of AI and ML in outcomes to improve glucose control [94]. They fixated on predictive modeling development in the space of not only automated insulin delivery systems but also CGM. The study specifically worked on challenges faced in terms of data consistency, clinical accuracy, interpretability, and personalization. The study appears as a guide for ML practitioners on diabetes data, including best practices, feature engineering, standardizing datasets, and evaluating models. Overall, the findings of this review describe improvements due to ML in diabetes management. The paper also discussed the difficulties of applying ML and AI techniques, including the data processing inhomogeneity, metrics evaluation for models, and the usage of multiple data sources accounting for glycemic control interpatient variability. Going along the same direction as a literature analysis, another paper investigated the incorporation of GI into smartphone-based food classification and nutritional estimation [95]. Most of the systems discussed in this review use computer vision to categorize foods and predict portion volumes to help facilitate dietary monitoring in diabetes management. Further, the paper reviewed possible future GI integration supportive technologies. ML and DL techniques have been thoroughly reviewed for food identification and volume estimation. CNNs such as AlexNet, VGG, ResNet, and EfficientNet are commonly used as baseline models to classify food images. Classification based on extracted features was performed using SVM and RF. According to these review studies, preliminary knowledge on ML has been integrated for food bioactivity screening, automated dietary assessment, and GI control. Research gaps still exist because no predictive modeling dedicated to the GI was found. Recently, machine learning approaches have been developed to estimate GI or inter-individual glycemic responses than previous studies. These

models are based on complex patterns of food composition variables in larger and more diverse datasets.

Machine Learning technologies play a crucial role in research related to the GI. GI prediction, glycemic control, and how glycemic indices affect a diabetic patient were identified using ML algorithms. For such a task, identifying complex relationships between food composition, blood glucose levels, and health outcomes was taken into account. Regression models and classification models played a major role. Regression models such as multiple linear regression, elastic nets, and ridge regression were used to predict GI by analyzing food characteristics. These models provide a basis for calculating how different foods affect blood glucose levels. Classification algorithms, such as random forests, SVM, and logistic regression, enable food classification based on GI levels and have proven useful in creating personalized dietary recommendations. Time series models such as ARIMA and autoregressive neural networks are used in real-time glycemic monitoring. By analyzing CGM data, it predicted blood glucose trends, which is especially helpful for diabetics who need dynamic management of blood sugar levels. Ensemble models, such as gradient boosting and random forest ensembles, improve predictive accuracy by combining multiple algorithms, often for applications such as identifying unreported foods or adjusting insulin in artificial pancreas systems. Finally, explainable AI models, using methods such as SHAP, provide insight into feature importance, allowing clinicians and researchers to understand which factors most influence glycemic outcomes. Together, these ML models form a robust framework for predicting, monitoring, and managing glycemic responses, making them central to the advancement of personalized glycemic control. Collectively, these approaches highlight how ML techniques can be applied not only to GI prediction but also to broader applications such as diabetes management, dietary assessment, and personalized nutrition by advancing computational nutrition science.

4.2 Role of reinforcement learning (RL) in research related to the Glycemic index

Reinforcement Learning is a machine learning paradigm. Agents learn to make decisions by interacting with an environment to maximize cumulative rewards. Unlike supervised learning, which is based on labeled data, RL involves trial-and-error exploration. This dynamic learning approach is particularly effective in problems that require sequential decision-making, such as feature selection, control systems, and real-time predictions. In our repository, the study "Impartial Feature Selection Using Multi-Agent Reinforcement Learning for Adverse Glycemic Event Prediction" [20] represents a pioneering effort in applying RL to feature selection in the context of blood glucose prediction. They presented a model for predicting adverse glycemic events (normoglycemia, hypoglycemia, hyperglycemia) using CGM, Electronic Medical Record (EMR), Multi-agent Reinforcement Learning (MARL), and Time2Vec (T2V). EMR data were used for feature selection. MARL employed optimal feature selection and selected optimal EMR features for better model performance. Although the study does not directly predict or calculate the GI of foods, it utilizes CGM-derived blood glucose levels and EMR data to predict adverse glycemic events. MARL evaluated individual feature contributions and derived the optimal feature set by dynamically assigning rewards proportional to the performance change each feature contributed. It has been observed that, aside from this study, few significant efforts

have been made to apply Reinforcement Learning (RL) to Glycemic Index research. This highlights an open area of exploration, presenting an opportunity to go deeper into the potential of RL in GI-focused studies.

4.3 Role of deep learning in research related to the Glycemic index

Deep Learning is a specialized area of artificial intelligence that utilizes neural networks with multiple layers to analyze and interpret complex data patterns. By mimicking the way the human brain processes information, deep learning models can automatically extract features from raw data, making them highly effective for tasks such as image and speech recognition, natural language processing, and more. These models are trained on large datasets, adjusting their internal parameters to improve accuracy and performance. The rise of deep learning has been fueled by advancements in computational power and the availability of vast amounts of data, leading to significant breakthroughs in various fields, including healthcare, finance, and autonomous systems. One emerging area where deep learning is making a considerable impact is in health-related research, specifically the prediction and recommendation of foods based on their GI, making it a critical tool for managing diabetes and other metabolic conditions. Through deep learning techniques, researchers can analyze food images, predict GI values, and recommend lower-GI alternatives to support personalized dietary plans. These applications combine computer vision with nutritional science, demonstrating deep learning's potential to support better health outcomes through dietary management.

Reference [44] is the earliest study in our survey, linking deep learning with GI prediction, specifically an artificial neural network, which is employed to predict the GI of foods. The process involves simulating human digestion, where samples undergo enzyme digestion, and their sugar content is analyzed using HPLC (High-Performance Liquid Chromatography). The ANN takes the compositional data (such as protein, fat, dietary fiber, and sugar content) from the HPLC results and predicts the GI by learning from known GI values. The ANN model achieved a high correlation ($r^2 = 0.93$) between predicted and actual in vivo GI values, which shows that it could predict GI values closely matching those obtained from conventional human testing. Another way to predict GI is to analyze captured signals from chewing and swallowing, leading to our next research work, [67]. This proposes a method for managing diabetes by monitoring food intake behavior (chewing, swallowing, and saliva secretion) and its impact on blood glucose levels. The study uses a Microelectromechanical System (MEMS) acoustic sensor to capture signals from chewing and swallowing, analyzing these signals to predict and control postprandial GI. Convolutional neural networks are used for feature extraction from acoustic signals generated during chewing and swallowing, focusing on spatial and frequency patterns. Additionally, Deep Belief Networks (DBNs) are employed to further analyze non-linear relationships in chewing signals, helping to generalize patterns and link them to blood glucose levels. When we talk more about health monitoring and predicting GI, the previous work [46] also focuses on predicting the GI of fruits, primarily focusing on bananas. The CNN model is applied for fruit recognition (apples, bananas, and oranges) and ripeness (raw-green, ripe, overripe) detection. Ripeness is critical because the GI of bananas varies significantly with ripening. After that, the model uses image binarization to estimate banana length, exploring OpenCV's

Thresh_Binary and boundingRect functions. Length serves as an essential feature for GI and carbohydrate content prediction. The GI is derived from GL and carbohydrate content. The model assesses whether the fruit is safe to consume based on GI thresholds, providing dietary recommendations if necessary.

Another fruit-based study [93] uses deep learning techniques, specifically an improved Faster R-CNN model with a Squeeze-and-Excitation (SE) attention module, to estimate the GL index of fruits, not the GI. The model identifies the fruit type (using R-CNN) and estimates its volume (based on fruit size relative to a reference object (e.g., thumb) for accurate GL calculation), which are then used to calculate the GL based on existing GI values. The GI of each fruit type is a known input in this process. Instead, the model uses the fruit's GI, volume, and carbohydrate content to compute the GL, providing a measure of how consuming that fruit might affect blood sugar levels. Just like fruits, bread is also explored as a research area for GI prediction through deep learning. [26] utilizes deep learning and computational intelligence methods to predict the digestion kinetics and GI (with the help of sample concentration of Euryale Ferox Seed Shell Extract (EFSSE), digestion time, and hydrolyzed starch concentration after digestion) of bread fortified with EFSSE. The Swarm Intelligence Supervised Neural Network (SISNN), specifically using Particle Swarm Optimization (PSO), simulates digestion kinetics more accurately than traditional mathematical modeling, aiding in predicting the glycemic response of fortified bread. The model demonstrates improved performance in predicting the GI of bread samples by integrating the optimization strengths of PSO with neural network modeling.

Rather than predicting GI, reference [84] aimed to predict heart disease risk in diabetic patients using deep learning techniques. GI is used as part of the input data for predicting heart disease risk among diabetic patients, providing insights into how certain foods affect blood glucose levels. In this study, LSTM (Long Short-Term Memory) was tested to determine its effectiveness in predicting heart disease based on diabetic patients' data. Gated Recurrent Unit (GRU) is another RNN variant designed to handle sequential data, but with a simpler architecture compared to LSTM. GRU outperformed LSTM, providing better results in terms of accuracy and efficiency. It optimizes the learning rate through backpropagation, adjusting parameters to improve prediction accuracy. While the earlier works utilized deep learning to analyze food intake behaviors, the study [83] shifts focus toward comparing predictive models for continuous glucose monitoring. The comparison was carried out between ARIMA models for auto-adaptive parameter tuning with statistical tests for real-time GI prediction and LSTM-based RNNs to capture long-term dependencies, trained with backpropagation through time. This presents a novel method for parameter optimization in ARIMA and evaluates these models in a practical online learning scenario, with specific applications in health monitoring systems.

Health monitoring alone is insufficient for maintaining good health; it is essential to consume appropriate food varieties and quantities to achieve a healthy lifestyle. In reference [25] GI is incorporated into a recommendation system that suggests healthier food alternatives for users with specific health conditions like diabetes. When a user inputs a food image, the model identifies the food item and retrieves its nutritional content, including GI. If the identified food has a GI over 55, the system recommends three similar foods with a lower GI, suitable for users who need to manage

blood sugar levels. The study uses the InceptionV3 deep learning model, a CNN, for food image recognition, which achieved an accuracy of 75%. Another similar food image recognition application is used in the reference [11] aims to develop a Moroccan food dataset for food image recognition and nutritional analysis, specifically focusing on estimating the GI of various Moroccan dishes using deep learning techniques. The study employed transfer learning using pre-trained CNN models, specifically evaluating DenseNet, MobileNet, and EfficientNet. The GI and GL were calculated based on recognized food items, utilizing established GI databases and the carbohydrate content of the dishes. Foods were categorized into low, medium, and high GI and GL based on their values.

In our collection related to DL, we found one review article [95] combining GI and deep learning techniques. The GI is used in these systems as a benchmark to guide dietary recommendations, particularly for diabetics. Once the system classifies a food item, it uses its estimated GI to assess potential blood glucose impact, and it can offer lower-GI alternatives if needed. This approach aims to help diabetic patients manage post-meal blood glucose levels by suggesting healthier food options. Here are some inputs used to predict GI and Nutritional Estimation in the previously mentioned studies done with DL and GI.

- Food images: Captured by smartphone cameras and processed through CNNs for classification.
- Volume estimation: Uses either single or multi-view images to approximate portion sizes, which are critical for calculating nutrient intake.
- Nutritional database: Contains data on each food item's GI, carbohydrates, protein, and fats, enabling the model to provide personalized dietary advice.

DL has become a transformative approach in GI prediction and dietary management by combining computer vision, signal processing, and recommendation systems. DL models such as CNNs, LSTMs, and GRUs can be used to understand the nonlinear behavior of food composition, eating behavior, and glycemic responses. DL-based applications span estimating GI using food images and detecting ripeness to provide personalized meal plans and continuous glucose monitoring. The integration of attention mechanisms, transfer learning, and optimization algorithms has further enhanced the accuracy of predictions. Summarizing the studies related to DL with GI, it can be stated that deep learning techniques can be effectively applied to predict the Glycemic Index of various foods, offering promising insights for more accurate and scalable GI estimation. By exploiting complex patterns within the data, deep learning models provide a robust approach that can be adapted to various types and characteristics of food, contributing significantly to advances in personalized nutrition and dietary recommendations.

4.4 Role of image processing (IP) in research related to the Glycemic index

Image processing has revolutionized the way we interpret and analyze visual data, becoming an essential tool across various fields, from healthcare diagnostics to autonomous systems. By transforming raw images into valuable information, image processing techniques allow machines to perceive, interpret, and act on visual inputs, pushing the boundaries of innovation. With a blend of mathematics, algorithms, and creativity, this domain continuously opens up new possibilities for automating complex tasks and unlocking insights that are often invisible to the human eye. In several research that connect with

Glycemic Index, Image Processing techniques were used for different purposes. The earliest work from our repository appears in 2023; A Novel Machine Learning and Deep Learning Driven Prediction for Pre-diabetic Patients [30]. They present a Machine Learning and Deep Learning- based prediction model that predicts the Glycemic Index of fruits, specifically bananas, using image recognition and classification. The model aims to determine if a fruit is safe for consumption by pre-diabetic patients based on its glycemic properties. The system also provides food recommendations with high dietary fiber to help users maintain a balanced diet. From the point of image processing, techniques such as binarization and boundingRect from OpenCV are used for predicting the length of fruits. These images were then used to train a CNN for fruit recognition and classification. Similarly, aiming at diabetic patients [93], focuses on recognizing fruits and estimating the GL index. The primary goal is to help diabetic individuals make informed decisions about their daily diet by identifying fruits with high sweetness but low GL values. The Glycemic Index is indirectly used in estimating the GL of fruits. The GL considers both the GI value and the carbohydrate content of the food, providing a measure of how a particular food affects blood sugar levels. By identifying fruits and estimating their volume, the study calculates the GL index, helping diabetic patients determine whether a fruit is suitable for consumption based on its expected impact on blood glucose levels. The authors developed a custom dataset called DODP, containing 54,000 images of fruits captured from various angles and environments where image preprocessing and data augmentation have been utilized. Automatic White Balance (AWB) and histogram equalization were used to improve the color and contrast of the images to improve the consistency of the input images. To generalize the images, data augmentation techniques such as adding Gaussian white noise, Pretzel noise, image flipping, rotation, and panning were applied, making a more diverse dataset for training. Going along the same direction of implementing data sets, [11], have created and used a Moroccan food dataset containing 72 dishes and 8,300 images. The primary goal is to use this system for nutritional analysis, for estimating the Glycemic Index and GL of Moroccan dishes, for dietary planning and management for chronic conditions such as diabetes. As the main AI technology, DenseNet deep learning models with an attention mechanism were used to improve the accuracy of food image classification. From the image processing perspective, mainly as in the previous, data augmentation has been done with the use of techniques like flipping, rotation, cropping, and noise removal in order to expand the dataset and improve model generalization.

Another recent research work presents a system that recognizes food items from images uploaded by users and predicts their nutritional values, including GI, proteins, carbohydrates, and fats [25]. They have used a custom Inception-V3 model for food image recognition and classification. As images are the main source, and users are responsible for uploading them, different image processing techniques such as noise reduction, histogram equalization, and data augmentation were utilized. Specifically, a 3x3 median filter is applied to reduce "salt and pepper" noise, Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to standardize lighting and color profiles, improving generalization across varied food images. Apart from them, data augmentation techniques such as shearing, zooming, rotation, and horizontal flipping are used to artificially expand the dataset for smoother training.

In addition to such direct use of Image Processing techniques, we found one review article summarizing the current state of mobile-based Food Image Recognition Systems (FIRS) designed for the dietary management of diabetics [95]. The focus is on evaluating technologies for classifying food, estimating food volume, and calculating the nutritional content of foods using smartphone cameras and computer vision techniques. The paper reviews approaches for using deep learning, machine learning, and image processing to automate food classification and support diabetic diet management. In the presented papers, the following image processing techniques were used appropriately.

- Preprocessing: Enhancing food images by correcting lighting, scaling, cropping, and applying contrast adjustments.
- Segmentation: Separating different food items within an image using methods such as manual segmentation, thresholding, color/texture-based segmentation, CNN-based segmentation, and clustering-based techniques.
- Feature extraction: Extracting visual features from the images, such as color, texture, shape, and edges, using methods like Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), Gabor filters, and Local Binary Patterns (LBP).
- Volume estimation: Using geometric modeling, pixel counting, and 3D reconstruction techniques from multi-view images to estimate the volume of food items. Depth map fusion techniques and shape-fitting methods (e.g., cylinders, spheres) are also reviewed.

Although image processing techniques were not the primary focus of the research discussed, they emerged as essential tools for facilitating the effective application of other AI techniques. Image processing techniques have contributed to enhancing the quality of AI-powered systems that predict GI values or dietary plans by transforming images into more insightful information sources. Image processing techniques such as noise reduction and histogram equalization were employed to enhance the images.

4.5 Role of natural language processing (NLP) in research related to the Glycemic index

Natural Language Processing, which reshapes how researchers handle vast amounts of unstructured text, enables deeper insights and more efficient data interpretation. The subject has transformed the way we understand and interact with text-based data, becoming a crucial component in fields ranging from healthcare to customer service automation. By converting raw text into structured information, NLP techniques empower machines to comprehend, interpret, and generate human language, enhancing our ability to analyze large volumes of textual content. With the fusion of linguistics, machine learning, and algorithms, NLP is unlocking new opportunities to automate tasks such as translation, sentiment analysis, and information retrieval, offering deeper insights into language patterns that are often difficult for humans to detect. In many studies related to sentiment and behavior analysis, NLP techniques have been applied for various purposes, demonstrating their growing importance across disciplines. Research that is based on Glycemic index-related studies also benefited from this technology and has opened many research avenues as well. One of the first studies in 2019 related to NLP explores the estimation of the glycemic impact of cooking recipes using a data-driven approach, combining online crowd-sourcing and machine learning [72]. The goal was to classify recipes as

healthy or unhealthy for diabetics based on their glycemic impact. They used Amazon Mechanical Turk (AMT) workers to crowdsource glycemic impact labels for recipes and used machine learning models to predict whether a recipe was unhealthy for diabetics (UD) or healthy (HD). The study used a data set of 55,102 All-recipes recipes, narrowing it down to 990 recipes (Recipe990) for a detailed analysis based on the range of glycemic impact and the difficulty of classification. AMT was used to gather human judgments on the glycemic impact of these recipes. Due to the impracticality of using the Glycemic Index for large datasets, the researchers relied on the sugar-to-fiber ratio (S/F) as a proxy for the glycemic impact. To process textual data, NLP techniques such as Bag-of-Words, word embeddings (word2vec, GloVe, fastText), sentence embeddings (skip-thought vectors), and paragraph embeddings (doc2vec) were used. The best performance among models using only textual features came from the NB-BoW + Logistic Regression model with an F1 score of 0.817. When nutritional features were added, the GloVe + LightGBM model achieved the highest overall F1 score of 0.854, highlighting that the combination of textual and nutritional data improves the precision of estimating the impact of glycemics for diabetics. In addition to the primary focus above on dietary management technologies, one review article systematically evaluates mobile-based Food Image Recognition Systems aimed at dietary management for diabetics [95]. Relevant articles published over the last two decades are evaluated. The paper discusses the importance of integrating the Glycemic Index and GL into food classification systems, which is crucial to predicting the impact of food on blood glucose levels. It emphasizes that diabetic patients can benefit from technologies that predict GI and GL. Although the review discusses the potential use of GI in future applications, it primarily reviews current methods that focus on food recognition and volume estimation, without directly using GI data in the reviewed systems. The paper assesses various methodologies for classifying food, estimating portion sizes, and calculating nutritional content using smartphone cameras and advanced NLP techniques.

These papers mainly follow some of the NLP techniques and concepts with the goal of improving food classification and nutritional estimation.

- Textual analysis: Evaluating the nutritional information associated with food items through text data from various sources, such as recipes and nutritional databases. This involves extracting relevant information about macronutrients, ingredients, and portion sizes from written content.
- Ingredient recognition: Using NLP techniques to parse and recognize food ingredients from text descriptions, which could complement image classification by providing additional context regarding food items that may not be visually distinguishable.
- Recipe parsing: Developing methods to extract nutritional data and estimates from ingredient lists found in recipes. This includes recognizing quantities and types of ingredients to calculate potential carbohydrate content, which is crucial for diabetic diet management.
- Machine learning for text classification: Employing text classification algorithms (e.g., logistic regression, SVM) to categorize recipes or food items based on their healthiness scores or glycemic impact, derived from textual data
- Textual data extraction: NLP techniques were utilized to extract and process textual information from food labels,

menus, or recipes associated with food images, enabling a more comprehensive analysis of nutritional content. NLP techniques have been used to analyze text-related data such as recipes, nutritional descriptions, and ingredient lists to extract features that require further processing to reveal insights such as carbohydrate content, sugar-to-fiber ratios, and overall dietary healthiness. These results can be combined with visual features for more accurate glycemic impact prediction. Beyond classification, NLP also supports ingredient recognition, recipe parsing, and textual data extraction from diverse sources, enabling automated and scalable dietary assessment.

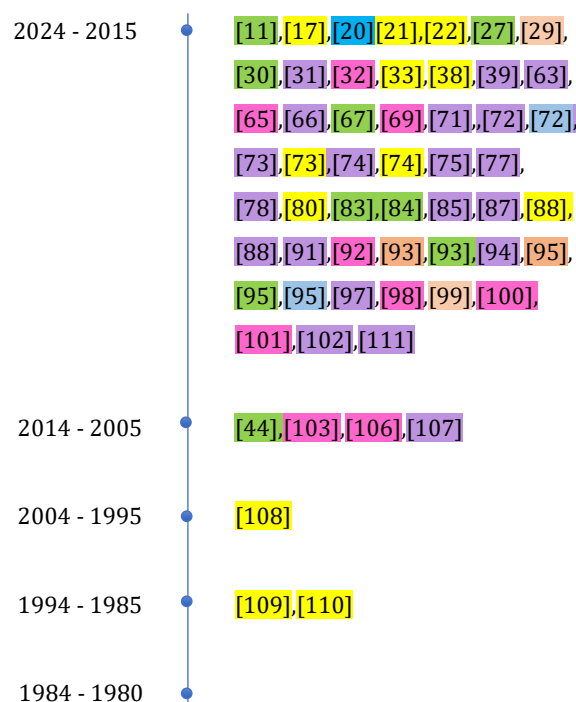


Figure 7: Timeline of research with categories; Background colors represent different categories of researches:

Machine Learning, Deep Learning, NLP, Reinforcement Learning, Image Processing, Statistical Learning, Explainable AI, Others

4.6 Role of explainable AI (XAI) in research related to the Glycemic index

Only a few studies in our repository have directly applied Explainable AI (XAI) techniques in their research. The earliest example identified was published in 2023 [112]. This study employed Shapley Additive exPlanations (SHAP), a widely used XAI method, to evaluate the influence of various meal-related factors on predicting postprandial blood glucose levels at different time intervals. SHAP assigns importance values to the features, effectively highlighting their contributions to the predictions of the model. By utilizing SHAP, the researchers provided valuable insights into the effects of specific nutritional components, such as carbohydrate intake, protein, lipids, and Glycemic Index, on blood glucose levels in individuals with type 1 diabetes. This approach not only validated clinical hypotheses but also enhanced the interpretability of predictive models, fostering more transparent and informed decision-making in diabetes management. Another research has used SHAP to enhance the interpretability of its predictive models [53]. After developing a Diabetic Retinopathy (DR) risk prediction model using the CatBoost algorithm, the researchers applied SHAP to

interpret the model's outputs. SHAP values helped in understanding the contribution of each feature to the model's predictions, thereby elucidating the relationship between various risk factors and the likelihood of developing DR. SHAP was employed to rank features based on their impact on the prediction model. This ranking identified which factors were most influential in predicting DR risk, providing insights into the relative importance of different clinical indicators. Also, the study used SHAP to analyze the correlations between renal function indices and other measures. This analysis offered a deeper understanding of how different health indicators interact and contribute to the risk of developing DR. The integration of SHAP into the predictive modeling process revealed several key risk factors for diabetic retinopathy:

- **Positive correlations:** Factors such as albumin-to-creatinine ratio (ACR), glycated hemoglobin, 24-hour urinary protein, presence of nephropathy, and serum creatinine (SCR) were found to be positively correlated with DR. This means that higher values of these indicators are associated with an increased risk of developing DR.
- **Negative correlations:** Conversely, factors like C-peptide (CP), hemoglobin (HB), albumin (ALB), direct bilirubin (DBIL), and C-reactive protein (CRP) were negatively correlated with DR, indicating that higher levels of these factors might be associated with a reduced risk.
- **Non-significant factors:** The study found that characteristics such as height, weight, and erythrocyte sedimentation rate (ESR) did not have a significant relationship with the development of DR.

By employing SHAP, the researchers were able to provide a transparent and interpretable analysis of the predictive model, making the findings more understandable and actionable for clinical applications. This approach improves the reliability of the prediction results and helps identify critical factors for early prevention and clinical diagnosis of diabetic retinopathy. These two significant contributions, incorporating the latest machine learning advancements such as Explainable AI methods, highlight that the path is open for exploring Explainable AI in future research on the Glycemic Index. Using XAI techniques in glycemic and diabetes-related research has bridged the gap between model accuracy and interpretability. Past studies have shown that XAI techniques can be used to unveil nutritional and clinical factors influencing blood glucose dynamics and diabetic complications. XAI enables researchers to quantify each feature's contribution to predictions, thereby improving trust in AI-assisted decision-making. Although only a few studies have incorporated XAI techniques, these studies have laid a strong foundation for future research to use XAI methods more extensively to analyze predictions.

4.7 Role of statistical techniques in research related to the Glycemic index

Statistical Techniques are essential tools for analyzing data and drawing conclusions in various fields, including social sciences, health, and business. These methods can be categorized into descriptive statistics, which summarize and visualize data through measures such as mean, median, and mode, and inferential statistics, which enable researchers to make predictions or generalizations about a population based on sample data. Key components of statistical techniques include hypothesis testing, where null and alternative hypotheses are formulated to evaluate the significance of results using p-values and confidence intervals; correlation and regression analysis, which examine relationships

between variables to determine how one may predict or influence another; and t-test and ANOVA, which compare means across different groups to assess whether observed differences are statistically significant. Overall, statistical techniques provide a robust framework for making informed decisions based on empirical evidence, ensuring that conclusions drawn from data are reliable and valid.

Most probably, the earliest work of such appears in 1990: "Glycemic Index of Foods in Individual Subject" [109]. The study involved 12 diabetic subjects consuming mixed meals (bread, rice, spaghetti) to determine common Glycemic Index values. The researchers calculated GI based on the area under the glycemic response curve, using white bread as a reference. They used analysis of variation with repeated measures (ANOVARM) to assess differences in glycemic responses between meals and subjects, along with the Tukey's Q method for adjustment of multiple comparisons and Chi-square analysis to compare observed rankings with expected outcomes. The findings indicated that while individual responses varied, mean GI values for each food type were consistent across subjects, validating the predictive capability of GI in dietary studies among diabetic patients. Building on this foundation, in 1993, "Prediction of Glycemic Index for starchy foods" [113] analyzed 18 starchy foods to identify predictive factors for GI based on food components such as protein, fat, and total dietary fiber (TDF). GI was calculated using the area under the glucose response curve for each food relative to white bread. Statistical techniques included Regression Analysis to explore correlations between GI and food components, t-tests for comparing means between legumes and non-legumes, and calculating Correlation Coefficients to quantify relationships between GI and food components. The study highlighted that while certain food components correlate with GI, preparation methods and starch characteristics significantly influence glycemic responses.

A study by M. Mayo et al. [73] investigated the thresholds of fasting blood glucose and glycosylated hemoglobin associated with microalbuminuria. They enrolled 975 subjects, including 873 diabetic patients and 102 non-diabetic controls, to analyze the impact of glycemic control on microalbumin levels. The study does not specifically use the Glycemic Index but focuses on glycemic control measured by FBS and HbA1c levels. Explore how poor glycemic control affects microalbuminuria, which is relevant to understanding the impacts of diet on health. They used the Student t-test for comparing means between two groups and the Analysis of Variance to assess differences among multiple groups. Multiple Linear Regression is used to identify the relationship between variables and urinary microalbumin levels. Chi-squared analysis was employed for comparing prevalence rates. These techniques were used to evaluate differences in FBS, HbA1c, and homocysteine levels among different groups and to develop predictive models for microalbuminuria.

Further expanding the understanding of GI, the [108] study on composite breakfast meals involved 28 healthy young men testing 13 different meals. Researchers used Regression Analysis and Multivariate Analysis to develop prediction equations for GI based on meal components. The findings indicated that energy density and fat/protein ratios were more reliable predictors of GI than carbohydrate content alone. Together, these studies illustrate how statistical techniques are crucial for accurately determining and predicting glycemic responses, ultimately aiding in better dietary choices for individuals managing blood sugar levels. The study by Mohan et al. [74] explores the association

between poor glycemic control, hyperhomocysteinemia, and microalbuminuria, utilizing both traditional statistical methods and machine learning algorithms. Although not directly focused on the Glycemic Index, this research highlights the importance of statistical techniques in understanding health outcomes related to glycemic control. The researchers focused on glycemic control as measured by fasting blood sugar and glycosylated hemoglobin levels. The main technologies used included biochemical analysis equipment and computational software for machine learning. The dataset consisted of 975 subjects, including diabetic patients and non-diabetic controls, with features such as age, gender, FBS, HbA1c, and diabetic status. Statistical techniques employed included Student t-tests, ANOVA, multiple linear regression, and Chi-squared analysis to assess differences and model relationships between variables.

Another study introduces a hybrid model designed to predict GI and GL based on the macronutrient composition of foods [33]. This model combines deterministic calculations for glycemic carbohydrates with empirical coefficients for non-glycemic nutrients like proteins, fats, and fibers. By quantifying both the impact of glycemic carbohydrates and the GI-lowering effects of non-glycemic components, the model aims to facilitate the development of packaged foods and beverages with lower glucose responses. The model was validated using a dataset of 42 breakfast cereals and 60 in vivo trials, employing statistical techniques like Ordinary Least Squares (OLS) regression and Bland-Altman plots to achieve high correlation coefficients ($r = 0.90$ for GI and $r = 0.96$ for GL). The transparency of the model, with explicit coefficients for each nutrient, makes it interpretable and useful to guide the development of the product.

The study "Gluten-free cookies with low Glycemic Index and glycemic load: Optimization of the process variables via response surface methodology and artificial neural network" [80] focused on optimizing the production of gluten-free cookies with low Glycemic Index and glycemic load. It evaluates the impact of baking temperature and time on resistant starch (RS), GI, and GL using cardaba banana flour modified with citric acid to enhance RS content. The study employs Response Surface Methodology (RSM) and Artificial Neural Networks to model and optimize these parameters. RSM is used to understand the relationships between baking conditions and outputs, while ANN provides more accurate predictions. The GI is estimated through in vitro starch digestibility tests using a non-linear model. Experimental data from 13 baking trials, designed using a central composite design, were analyzed using techniques like ANOVA and regression analysis to assess model quality and optimize baking conditions. Overall, the study aims to develop gluten-free cookies with improved nutritional profiles by optimizing production parameters.

The study [83] uses historical glucose data from continuous glucose monitoring devices to predict future GI levels, with Autoregressive Integrated Moving Average adapting in real-time and RNN predicting trends over 30-60 minutes. The models are hosted on Google Cloud, utilizing technologies like Google Cloud PubSub, Functions, and BigQuery for real-time training and data management. The DINAMO dataset, which includes glucose readings from diabetic and non-diabetic patients, is used for model training and validation. Statistical techniques such as ADF tests, ACF/PACF, and Akaike Information Criterion are employed for model optimization and validation. The study aims to monitor glucose levels in bus drivers, providing alerts for

dangerous trends, and exploring the inherent interpretability of ARIMA models to understand forecast dependencies.

Reference [38] focused on determining the Glycemic Index of a complete nutrition drink formulated with retrograded starch and identifying factors influencing the glycemic response. This was achieved through a randomized crossover controlled trial involving 18 healthy participants who consumed the nutrition drink, glucose solution, and white bread as test foods. Normality tests, such as the Shapiro-Wilk test, are employed to evaluate the distribution of data. For comparative analysis, a one-way ANOVA is used for normally distributed data, while the Kruskal-Wallis and Friedman tests are applied for non-parametric comparisons. In scenarios involving repeated measures, repeated measures ANOVA with Tukey's test for post-hoc analysis is performed. For correlation analysis, the Spearman rank correlation is utilized to identify relationships between baseline characteristics and glycemic response. To control Type I errors in multiple comparisons, the Bonferroni correction is applied. Statistical techniques were primarily used to compare postprandial glucose and insulin levels across test foods, analyze baseline characteristics among groups, and identify correlations between predictors (e.g., baseline insulin) and glycemic response (Table 2).

"Predicting Changes in Glycemic Control Among Adults with Prediabetes from Activity Patterns Collected by Wearable Devices" [88]. The study explores the use of wearable devices to predict changes in glycemic control among adults with prediabetes, focusing on comparing wrist-worn and waist-worn devices. Participants were monitored over six months using Fitbit devices that tracked physical activity, sleep, and heart rate. The study did not involve calculating the GI but instead focused on predicting changes in hemoglobin A1c using wearable data and machine learning models. Traditional statistical regression models and machine learning techniques like random forest and ensemble methods were employed to analyze baseline demographic, clinical, and wearable data. Features included demographics, clinical data, physical activity, heart rate, and sleep patterns, which were reduced to 16 principal components using PCA. The study used techniques such as multiple imputations for missing data and hyper-parameter tuning with cross-validation. While transfer learning and explainable AI were not explicitly used, ensemble methods provided insights into predictive factors. The goal was to enhance predictive models for glycemic control changes, leveraging wearable data to potentially inform interventions for preventing diabetes progression.

Enhancing insights into GI, the study [17] aimed to enhance the Venezuelan Food Composition Table by integrating Glycemic Index values to aid in dietary assessments and research. It employed a systematic six-step methodology to assign GI values to 624 food items across 14 categories. This approach included direct assignment from international GI tables, mapping to similar foods, recipe-based calculations, and using subgroup median values for unassignable items. Key features influencing GI assignments were available carbohydrates, nutritional profiles (including fat, protein, and fiber), food preparation methods, and other nutrient compositions. The study utilized IBM SPSS for data analysis and adhered to the ISO 26642:2010 standard for GI determination. Statistical analyses were utilized to stratify the results by food group and to perform calculations such as mean, standard deviation, and percentile distribution of Glycemic Index.

Table 2. Summary of statistical techniques in the glycemic index research

Paper Reference	Statistical Tools Used	Purpose of Statistical Test
[109]	ANOVA with Repeated Measures (ANOVARM), Tukey's Q method, Chi-square analysis	Assess differences in glycemic responses between meals and subjects; adjust for multiple comparisons; compare observed rankings with expected outcomes.
[113]	Regression Analysis, t-tests, Correlation Coefficients	Explore correlations between GI and food components; compare means between legumes and non-legumes; quantify relationships between GI and food components.
[73]	Student t-test, Analysis of Variance, Multiple Linear Regression, Chi-squared analysis	Compare means between groups; assess differences among multiple groups; identify relationships between variables and urinary microalbumin levels; compare prevalence rates.
[108]	Regression Analysis, Multivariate Analysis	Develop prediction equations for GI based on meal components; identify reliable predictors of GI.
[74]	Student t-tests, ANOVA, Multiple Linear Regression, Chi-squared analysis	Assess differences and model relationships between variables; explore associations between glycemic control, hyperhomocysteinemia, and microalbuminuria.
[76]	Ordinary Least Squares Regression, Bland-Altman plots	Develop a hybrid model to predict GI and GL based on macronutrient composition; validate model accuracy.
[80]	Response Surface Methodology, Artificial Neural Networks, ANOVA, Regression Analysis	Optimize production of gluten-free cookies with low GI and GL; model and optimize baking conditions; assess model quality.
[83]	Autoregressive Integrated Moving Average, Recurrent Neural Networks (RNN), ADF tests, ACF/PACF, Akaike Information Criterion	Predict future GI levels using historical glucose data; adapt models in real-time; optimize and validate models.
[38]	Shapiro-Wilk test, One-way ANOVA, Kruskal-Wallis test, Friedman test, Repeated Measures ANOVA, Tukey's test, Spearman Rank Correlation, Bonferroni Correction	Evaluate data distribution; compare postprandial glucose and insulin levels across test foods; analyze baseline characteristics; identify correlations between predictors and glycemic response; control Type I errors in multiple comparisons.
[88]	Principal Component Analysis, Random Forest, Ensemble Methods, Multiple Imputation, Hyper-parameter Tuning, Cross-validation	Predict changes in glycemic control using wearable data; enhance predictive models; manage missing data; optimize model performance.
[17]	Descriptive Statistics, Proportional Analysis	Stratify results by food group; calculate mean, standard deviation, and percentile distribution of GI; classify foods into GI categories; determine the percentage of foods assigned GI values at each methodological step.
[21]	Correlation Analysis, Multiple Regression, One-way ANOVA	Explore relationships between blood glucose data and physiological/nutritional factors; assess postprandial glucose dynamics; analyze data from wearable devices.
[22]	Additive Main Effect and Multiplicative Interaction (AMMI) Analysis, GGE Biplot Analysis, Linear Mixed Models (LMM)	Evaluate grain yield, quality traits, and genotype-environment interactions; predict GI in rice varieties; develop low-GI rice suitable for specific ecosystems.

These analyses also facilitated the classification of foods into low, moderate, or high GI categories. Descriptive statistics were applied to summarize the GI distributions, while proportional analysis was employed to determine the percentage of foods assigned GI values at each step of the methodology. Another study investigates the relationship between blood glucose data and various physiological and nutritional factors using wearable devices and publicly available datasets. It utilizes devices like the Dexcom G6 CGM for blood glucose monitoring and the Empatica E4 Wristband for capturing physiological signals such as heart rate and skin temperature.

The analysis employs software tools like SAS 9.4, JMP Pro 16.1.0, and Microsoft Excel, and statistical techniques including correlation analysis, multiple regression, and one-way ANOVA. The study draws data from the PhysioNet BIG IDEAS Lab Dataset, which includes 16 participants with prediabetic HbA1c levels, focusing on glucose, physiological indices (e.g., heart rate, skin temperature), and nutritional indices (e.g., carbohydrates, dietary fiber). Notably, the study does not directly use or predict the Glycemic Index; instead, it focuses on real-time blood glucose data collected via CGM devices to explore correlations between physiological and nutritional indices and blood glucose levels, as well as assess

postprandial glucose dynamics [21]. Extending expertise on GI, Reference [22] aimed to develop low-glycemic-index rice varieties suitable for irrigated ecosystems in Bangladesh. It evaluated the grain yield, quality traits, and genotype-environment interactions (GEI) of three rice genotypes across 27 environments over three years. The GI values were determined using the Howlader and Biswas technique, which involves measuring postprandial blood glucose levels after consuming test and reference foods. Statistical analyses included Additive Main Effect and Multiplicative Interaction analysis, GGE biplot analysis, and Linear Mixed Models (LMM) using R software. Features such as milled rice outturn, head rice yield, amylose content, and protein content were used to predict GI. This research is crucial for developing rice varieties that can benefit health by reducing the impact of carbohydrates on blood glucose levels.

Overall, statistical techniques play a pivotal role in understanding and predicting glycemic responses, which are crucial for managing blood sugar levels. These methods encompass descriptive and inferential statistics, hypothesis testing, correlation and regression analysis, and tests like t-tests and ANOVA. Studies have utilized these techniques to determine Glycemic Index values for various foods, identify predictive factors for GI, and develop models to predict GI based on food components. For instance, regression analysis has been used to explore correlations between GI and food components, while response surface methodology and artificial neural networks have been employed to optimize production parameters for low-GI foods. Additionally, machine learning and wearable devices are being explored to predict changes in glycemic control, further enhancing the application of statistical techniques in this field. These methods provide a robust framework for making informed dietary choices and developing healthier food products.

4.8 Role of other computer-related techniques in research related to the Glycemic index

In the main section, we explored the role of AI techniques such as Machine Learning, Deep Learning (DL), Image Processing, and Natural Language Processing in research related to the Glycemic Index. While these advanced methodologies have transformed the field, there are other computer-related techniques that also contribute significantly to Glycemic Index research. This section focuses on these additional approaches, highlighting their unique applications and the value they bring to advancing our understanding of this critical area. The Internet of Things (IoT) has played a pivotal role in predicting the Glycemic Index. IoT refers to a network of interconnected physical devices ranging from appliances to vehicles embedded with sensors, software, and connectivity. This technology facilitates seamless communication and data exchange between devices, paving the way for more efficient and automated systems. In recent years, challenges such as data mining, machine learning integration, and IoT applications have gained prominence in the healthcare sector. A notable study published in 2019, titled "Internet of Things Based on Electronic and Mobile Health Systems for Blood Glucose Continuous Monitoring and Management" [69], showcased the integration of the Libre flash glucose monitoring sensor with mobile applications, creating a connected and comprehensive environment for glucose monitoring. They used cloud technologies to collect blood glucose data continuously, provide real-time alerts, and perform graphical analysis while monitoring and analyzing patient data remotely via a secure cloud-based platform. However, GI was

not a direct concern; the system focused on real-time monitoring of blood glucose. These data were used to identify patterns related to GI. In general, the study emphasizes the usage of IoT in the health care system to manage diabetes and low-cost alternatives to traditional methodologies for continuous glucose monitoring systems. The utilization of the long-term effect of the Internet of Things on glycemic control is controversial, and Type 2 diabetes is a common problem today. Another study focused on evaluating the long-term effects of an IoT-based approach on glycemic control in people with Type 2 Diabetes (T2D). The personal health records (weights, blood pressure, physical activities) were measured using IoT-enabled devices, and feedback messages were sent to encourage behavioral changes in diet and exercise. Data was shared with healthcare providers via cloud systems. GI was not directly addressed; instead, it focused on glycemic control through HbA1c levels and lifestyle modifications facilitated by IoT technologies [100]. The advancements in wearable glucose monitoring technologies, as reviewed by Mansour et al. [101]. This offers significant implications for Glycemic Index research. The paper reviewed advancements in wearable devices for CGM, including invasive, minimally invasive, and non-invasive methods. The paper highlights the integration of biosensing technologies with wireless communication, energy harvesting, and AI-based predictive analytics for diabetes management. The Glycemic Index is not explicitly discussed or used for predicting or calculating glucose levels. Instead, the focus was on measuring glucose directly from biofluids (e.g., blood, sweat, interstitial fluid) using various biosensor technologies. By integrating AI-driven models like RNNs, researchers can better account for factors such as physical activity, stress, and insulin sensitivity, enhancing the accuracy of GI predictions. The insights from this review pave the way for a more personalized and scalable approach to the management of diabetes and diet.

A healthy and balanced diet is essential for quality of life. Carbohydrates play a crucial role in maintaining a healthy and balanced diet, since they serve as the primary source of energy in the body. Going along with Mathematical approaches, the study, "A robust optimization approach to diet problem with overall GL as objective function" [106], addresses the problem of minimizing the overall GL in daily diets while meeting nutritional and serving size requirements. The authors proposed a mixed-integer programming model that incorporates uncertainties in GL values, allowing for flexible and adaptive diet planning. This study focuses on creating a mathematical framework to optimize daily food selection while minimizing the total GL. This optimization ensures that daily nutritional needs are met, minimizes the impact of foods with a high GL on blood glucose levels, and allows for uncertainty in GL values so that meals can be maintained under different circumstances. Another study investigates the effects of dietary Glycemic Index on β -cell function in adults with prediabetes through a randomized controlled feeding trial. A total of 35 adults with prediabetes underwent a 2-week control diet (GI = 55–58), followed by randomization into a 4-week low Glycemic Index (LGI; GI < 35) or high Glycemic Index (HGI; GI > 70) diet. Meals were carefully designed to meet GI specifications while maintaining consistent macronutrient distribution (55% carbohydrate, 30% fat, 15% protein) and ensuring weight stability. Meal tolerance tests (MTTs) were conducted at baseline and post-intervention to evaluate glucose, insulin, and C-peptide responses. Mathematical models and statistical tools (e.g., SPSS, MATLAB) were used to estimate β -cell

glucose sensitivity, insulin secretion rates, and insulin sensitivity indices (OGIS, Matsuda index). The LGI diet significantly reduced postprandial glucose concentrations ($p < 0.001$) and increased total insulin secretion adjusted for glucose levels and insulin sensitivity ($p = 0.002$). Conversely, the HGI diet showed trends towards higher glucose levels ($p = 0.14$) and reduced insulin secretion. Despite these differences, neither diet significantly affected traditional measures of insulin sensitivity. These findings highlight the potential of LGI diets in improving β -cell function and glucose regulation in individuals with prediabetes [32].

Continuous glucose monitoring is an important aspect for diabetic patients. There are many studies that focus on that in different ways. A study aimed to automate the calculations of the GI using a continuous glucose monitoring system (CGMS). GI was calculated using the incremental area under the blood glucose curve (IAUC). They calculated the IAUC using CGMS. A custom-built Microsoft Excel-based software for the automation process was introduced. This software, called DegifXL, processed CGMS data, calculated IAUC, and computed GI values using pre-defined and custom input parameters. The process reduces the manual processing time and improves the standardization of GI computations [103].

The study “Beyond nutrient-based food indices: a data mining approach to search for a quantitative holistic index reflecting the degree of food processing and including physicochemical properties” [65] explores the relationship between food processing, nutritional quality, and health potential, with a focus on the Glycemic Index as a key indicator of glucose bioavailability. Using data mining techniques such as decision trees, Bayesian networks, and Principal Component Analysis, the study analyzed 117 foods categorized by processing levels (minimally processed, processed, and ultra-processed) to develop a holistic Technological Index (TI). This index integrates functional properties like nutrient density, glycemic glucose equivalents, and physicochemical characteristics such as texture and water activity. Results showed that minimally processed foods generally exhibit lower GI, higher nutrient density, and better satiety profiles compared to ultra-processed foods, emphasizing the adverse nutritional impact of food processing. The study highlighted how GI and other food properties can inform health-focused dietary guidelines and aid in promoting healthier food choices.

Complementing this, the paper “AI4FoodDB: A database for personalized e-Health nutrition and lifestyle through wearable devices and artificial intelligence” [92], established a comprehensive database integrating IoT-enabled wearable devices, food diaries, and biological samples to explore relationships between diet, physical activity, and glycemic responses. While GI is not directly measured, continuous glucose monitoring data and dietary logs provide insights into postprandial glucose variability. Technologies such as FitBit Sense and FreeStyle Libre 2 sensors were used for real-time data collection, with AI and machine learning employed for analysis across 10 domains, including biomarkers, nutrition, and gut microbiome. Similarly, another study developed a comprehensive Glycemic Index and GL database for the United States (U.S.) using NHANES data (1999–2018) to analyze dietary carbohydrate quality and its health implications [98]. AI models, specifically OpenAI’s pretrained embedding tools, were employed to assign GI values to over 7,976 unique food codes, achieving 75% initial accuracy. But after manual review and adjustments based on expert knowledge, only 31.3% of the AI’s predictions were kept. Key

databases, including the International Tables of GI [37] and the Diogenes Study [114], provided reference values. The dietary GL was calculated by combining carbohydrate content and GI values, while statistical and mathematical methods, such as weighted averages and residual adjustments, ensured robust data analysis. Trends in GI and GL were examined across demographics, highlighting disparities in diet quality by sex, race, education, and income. While transfer learning enabled efficient GI assignment, manual adjustments ensured accuracy and interpretability, making this database a critical resource for precision nutrition and public health research. Together, these studies underscore the importance of integrating glycemic metrics, food processing indices, and advanced technologies like AI and IoT to promote precision-based nutrition and health monitoring.

5. Datasets used in the Glycemic index-related research

In the constantly evolving area of Glycemic index-related studies, the dataset a researcher chooses can significantly influence the study’s results and overall impact. Some researchers choose established benchmark datasets to maintain consistency and enable comparability across studies, while others create custom datasets to better align with their specific research goals. This chapter examines both approaches, emphasizing the importance and application of these diverse data sources in Glycemic Index research. Building on the AI-driven techniques introduced in Chapter 4, here we discuss how careful data selection and preparation are essential for pushing the field forward. The graph below highlights the type of dataset utilized in each reviewed paper, categorizing them as either benchmark datasets (pre-existing) or self-developed datasets (created specifically for the study).

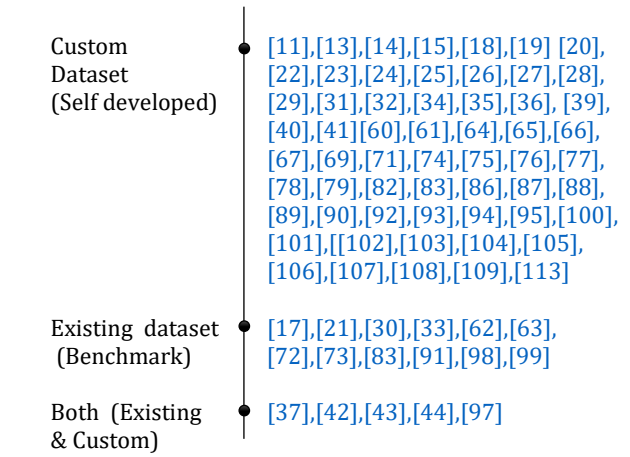


Figure 8: Dataset used by the research (Benchmark/ Custom dataset (Self-developed))

5.1 Utilization of custom datasets in technological approaches for Glycemic index assessment

As illustrated in the graph, the majority of studies utilize self-developed datasets, as these are tailored to the specific purpose of the research, resulting in improved model outcomes. As the earliest record in our repository [13] uses a custom-built dataset that consists of 62 commonly consumed foods and sugars. The foods were tested individually on groups of 5–10 healthy volunteers, totaling 34 individuals (21 male, 13 female). The dataset includes the Glycemic Index of various food items, calculated by measuring blood glucose levels over two hours after consuming the foods. The

glycemic response was expressed as a percentage of the blood glucose area under the curve compared to an equivalent amount of glucose. The study aimed to classify foods based on their physiological effects on blood glucose levels to better guide dietary choices, particularly for diabetics. Our second-earliest research [109] involved 12 diabetic subjects who were given three types of mixed meals: bread, rice, and spaghetti, with Glycemic index values of 100, 79, and 61, respectively. These meals were tested in a randomized block design, with each subject consuming each meal four times. The dataset includes glycemic response measurements over time, expressed as incremental areas under the glycemic response curves. The study investigated how well the GI values predicted glycemic response rankings for individuals. By normalizing responses to a standard reference (bread), the researchers showed that GI values could effectively rank glycemic responses across different subjects, despite individual variability.

A particular research [14], done for starch-rich foods typically consumed in the Mediterranean region with type 2 diabetic patients, measuring the plasma glucose response after consumption of different food portions containing 50 g of available carbohydrates (spaghetti, white bread, potatoes, pizza, potato dumplings, and hard toasted bread). The researchers measured the glycemic response of these foods in terms of blood glucose levels over a specific postprandial period, providing insights into the dietary effects of these foods on glycemic control. For example, they found that spaghetti and potato dumplings had lower glycemic responses compared to bread and potatoes, attributed to differences in food structure and preparation techniques. Reference [15] prepared and analyzed *Amaranthus cruentus* seeds using various processing methods such as cooking, popping, roasting, flaking, and extrusion. The seeds were sourced from a local producer in Brasília, Brazil. Using the Hydrolysis Index (HI) derived from *in vitro* tests, they determined the Predicted Glycemic Index (pGI) for each processed seed sample. But in here, a glycemic response comparison has been conducted with white bread as a benchmark (reference sample). Reference [16] involved 28 laboratories testing the Glycemic Index of foods such as cheese puffs and fruit leather. Each laboratory followed a standardized protocol where 10 healthy participants consumed test foods and reference foods (glucose or white bread) on separate occasions. Blood glucose responses were measured at regular intervals, and the Glycemic Index was calculated based on the incremental area under the curve of glucose response.

Reference [40] aimed to determine the Glycemic Index values of traditional foods and mixed meals from Northern Sri Lanka. They conducted experiments to measure the GI values of various traditional foods and mixed meals consumed in Northern Sri Lanka. This involved selecting specific foods, preparing them according to traditional methods, and then measuring the postprandial blood glucose responses in participants after consumption. The findings provided insights into dietary recommendations, especially for individuals with diabetes or coronary heart disease, by identifying foods with lower GI values that are more suitable for maintaining stable blood glucose levels. The researchers constructed a comprehensive dataset from their own experimental data for reference [60], encompassing continuous glucose measurements, detailed dietary logs, physical activity records, gut microbiota profiles, and various blood parameters from an 800-person cohort. The primary purpose of this dataset was to develop and train a machine-

learning algorithm capable of predicting individualized postprandial glycemic responses to different meals. The researchers evaluated the Glycemic Index of eight rice varieties in Taiwan, including two brown and six white rice types in [61]. The dataset was primarily built using both *in vitro* and *in vivo* methods for different rice varieties. *In vitro* starch digestion tests were conducted to determine the predicted Glycemic Index of the rice samples. This approach provided valuable insights into the glycemic properties of Taiwanese rice varieties, aiding in the development of rice with desired health benefits.

The dataset used in [64] was built by the researchers and comprised blood glucose measurements from healthy human volunteers who consumed bread samples with varying levels: 0%, 10%, 15%, and 20% of cassava flour substitution. Glucose was used as a reference food to calculate the Glycemic Index of each bread variant. Participants consumed the test breads after a 10–12-hour overnight fast, and blood glucose levels were recorded at 30-minute intervals over a 2-hour period following consumption. Then, the GI values of the bread samples were determined. The dataset enabled the assessment of how substituting wheat flour with cassava flour affects postprandial glycemic responses. Also, the study found that increasing cassava flour content led to lower glycemic responses, with GI values ranging from 91 to 94. The authors developed their own dataset for [67] using acoustic signals during chewing and swallowing from 50 diabetic individuals using an acoustic Micro-Electro-Mechanical Systems (MEMS) sensor. These signals were then processed with a deep learning algorithm to analyze eating patterns and formulate a standard procedure aimed at reducing blood glucose levels. Reference [74] used a dataset comprising clinical measurements from diabetic and non-diabetic individuals by collecting data on Fasting Blood Glucose, glycosylated hemoglobin, total plasma homocysteine levels, and urinary microalbumin concentrations. Data were analyzed using multiple linear regression and machine learning algorithms to investigate the relationships between glycemic control, hyperhomocysteinemia, and the presence of microalbuminuria. The study aimed to identify threshold values of FBS and HbA1c associated with microalbuminuria and to explore the concurrent association of microalbuminuria with hyperhomocysteinemia.

The dataset in [22] includes comprehensive data on rice grain yield and quality traits over a three-year period, spanning 27 different environments in Bangladesh. The primary objective was to evaluate the performance of various rice genotypes under different environmental conditions to identify a stable and adaptive variety with desirable traits, including a low Glycemic Index. The data were analyzed using statistical methods such as ANOVA and Additive Main effects and Multiplicative Interaction analysis to assess genotype-environment interactions and stability. This analysis aimed to inform breeding programs focused on developing rice varieties that are both high-yielding and suitable for the irrigated ecosystems of Bangladesh. The study [20] utilized a dataset that comprises continuous glucose monitoring data from 102 patients with type 2 diabetes admitted to Cheonan Hospital, Soonchunhyang University. This data includes blood glucose levels, insulin doses, meal times, and other Electronic Medical Records information. The researchers employed a multi-agent reinforcement learning algorithm to perform feature selection, aiming to enhance the prediction accuracy of adverse glycemic events. The model achieved F1-scores of 89.0% for normoglycemia, 60.6% for hypoglycemia, and

89.8% for hyperglycemia, demonstrating its effectiveness in predicting glycemic events.

Reference [71] utilized retrospective CGM datasets from 104 individuals who had experienced at least one hypoglycemia alert during a three-day CGM session. These datasets were collected from participants who had previously undergone CGM monitoring. Among different ML methods experimented on the dataset, random forest demonstrated the best performance, with an average AUC of 0.966, sensitivity of 89.6%, specificity of 91.3%, and an F1 score of 0.543, concluding that the random forest model holds potential for accurately predicting postprandial hypoglycemia, which could enhance the effectiveness of continuous glucose monitoring and artificial pancreas systems. Likewise, reference [39] also follows CMG data from 124 pregnant women—90 diagnosed with gestational diabetes mellitus and 34 healthy controls. This dataset included records of 1,489 food intakes, which were documented using a mobile application developed for the study. The Glycemic Index values for the foods consumed were sourced from the University of Sydney's database and incorporated into the app's food database. The collected data were analyzed to develop predictive models for postprandial glycemic responses, assessing the impact of incorporating GI and GL information on the accuracy of these models.

Similarly, reference [115] used CGM data from individuals, particularly European, Chinese, younger, and older participants with Type 1 Diabetes (T1D) Mellitus. Benchmark GI values, such as those for glucose and white bread, were used as references for comparison and scaling. The authors employed machine learning algorithms to analyze this data, aiming to predict glycemic levels in real-time using constrained Internet of Things devices. The study concluded that local, on-the-fly forecasting of glycemia is feasible with such devices. The dataset used in [78] was built by the researchers based on data collected from a clinical trial involving 235 participants, including women with gestational diabetes mellitus and healthy pregnant women. The data includes CGM records, meal-related information, patient characteristics, and survey data. Participants recorded their meals in a mobile app while wearing CGM devices to monitor blood glucose levels. The dataset captures meal timing, composition (e.g., carbohydrate content, GL), and pre-meal CGM trends. Flawed records (e.g., underreported meals) were detected and removed to ensure data quality. Models were evaluated using cross-validation and test data from unseen participants. Postprandial glucose responses (PPGRs) from 15 participants who consumed nine standardized meals with known macronutrient compositions in [82]. Participants' PPGRs were recorded using continuous glucose monitors after they consumed the standardized meals. Each meal's macronutrient content, carbohydrates, proteins, and fats were precisely measured. The model's performance was assessed by comparing its macronutrient predictions against the actual known values. The proposed sparse coding approach consistently outperformed baseline systems based on ridge regression and nearest-neighbors in terms of correlation and normalized root mean square error of the predictions. This methodology demonstrates the potential of using CGM data to automatically estimate dietary intake, reducing reliance on self-reported measures.

Similarly, reference [107] used a dataset that comprises continuous glucose monitoring data collected from participants, capturing detailed blood glucose measurements over time. The raw CGM data were processed to extract relevant features indicative of glycemic patterns, such as

mean glucose levels, variability metrics, and trends over time. Machine learning regression models were trained using the extracted features to predict future blood glucose levels. These models aimed to forecast glucose trends and potential hyperglycemic or hypoglycemic events. This approach highlights the effectiveness of using self-collected CGM datasets in developing personalized machine learning models for predicting blood glucose levels, which can be instrumental in managing diabetes. The dataset in [103] also comprises continuous glucose monitoring, which profiles from 20 healthy subjects who consumed 50 grams of glucose or one of four alternative foodstuffs, like chocolate, apple baby food, rice squares, or yogurt, at breakfast and dinner over a one-week period, resulting in 300 CGM glucose profiles. Participants wore CGM devices to continuously record interstitial fluid glucose concentrations. They consumed specified test foods, each containing 50 grams of carbohydrates, at designated meal times, with glucose serving as the reference food. The IAUC values obtained from the test foods were compared to those from the reference food (glucose) to calculate the Glycemic Index for each food item. Reference [69] introduced an integrated environment for continuous blood glucose monitoring. This system utilizes Internet of Things technology to provide real-time data to doctors and caregivers remotely. The researchers developed their own dataset by collecting blood glucose measurements using the Freestyle Libre system. This data was then transmitted through their IoT-based platform, enabling continuous monitoring and management. The dataset facilitated the evaluation of the system's performance by comparing the glucose rates measured with the official Freestyle Libre software during the same period. Comparably, the dataset in [86] includes continuous glucose monitoring readings, records of insulin injections, and carbohydrate intake information. The researchers applied exponential models to the raw carbohydrate and insulin data to simulate absorption processes in the body, aiming to enhance the accuracy of their predictive models. By incorporating these simulated absorption curves into an RNN based on long short-term memory cells, they sought to improve the prediction of future blood glucose levels. However, subsequent analysis revealed flaws in the experimental techniques, particularly in the model validation scheme, which invalidated the reported results and conclusions.

The study [76] employed virtual patient models for T1D patients, such as a virtual patient cohort that includes 10 adults and 10 adolescents. These models simulate various physiological responses to insulin treatment, allowing for controlled experimentation without human participants. In-Silico Data: The virtual patient models were used to simulate and evaluate the performance of the machine learning-based artificial pancreas algorithm under various scenarios, providing preliminary insights into its potential effectiveness and safety. In-Vivo Data: The clinical trials with human participants were conducted to validate the algorithm's performance in real-world settings, assessing outcomes such as time-in-range (TIR), hypoglycemic events, and overall glucose control. The study [18] creates a dataset by utilizing details of 12 healthy volunteers (6 men and 6 women) aged between 20 and 30 years. Four traditional Omani rice dishes were selected: white rice, biryani, kabsa, and maqboos. Glucose was used as the reference food for determining the Glycemic Index. Blood glucose levels were measured at intervals of 15, 30, 45, 60, 90, and 120 minutes after consumption. With the data, the GI and GL were calculated.

The study found that white rice had the highest GI value (77.3), while the other rice dishes had moderate GI values. This information is valuable for dietary planning, especially for individuals managing blood sugar levels. The dataset in [19] comprised 20 infant cereal prototypes, each with varying macronutrient compositions, particularly in glycemic carbohydrates (ranging from 51 to 76 grams per 100 grams), and was then utilized to validate a predictive model for estimating glycemic responses based on nutritional composition. The data were collected through four independent studies measuring the postprandial glucose responses of these cereal prototypes in healthy adults. The collected data were then applied to a predictive model previously developed to estimate the GI and GL of breakfast cereals based on their macronutrient composition. This model quantifies both the impact of glycemic carbohydrates and the GI-lowering effects of other macronutrients such as proteins, fats, and fibers. 35 adults with prediabetes (17 females, 18 males; mean age 54.2 years; mean BMI 32.44 kg/m²) have contributed data to [32]. Participants underwent a controlled feeding study, consuming either a low Glycemic Index (LGI) diet (GI < 35) or a High Glycemic Index (HGI) diet (GI > 70) for four weeks, following a two-week control diet (GI = 55–58). The researchers conducted 4-hour meal tolerance tests to assess insulin sensitivity, insulin secretion, and β -cell function. The collected data were analyzed using mathematical modeling to evaluate the impact of dietary Glycemic Index on β -cell function in individuals with prediabetes.

The 1993 study titled “Prediction of Glycemic Index for starchy foods”, reference [113] analyzed 18 starchy foods to examine the relationship between their Glycemic Index and chemical components such as protein, fat, phytic acid, and Total Dietary Fiber. For each food item, the dataset included measurements of protein, fat, phytic acid, and TDF present in portions containing 50 grams of available carbohydrate. The researchers employed regression analysis to explore associations between the GI and the chemical components of the foods. They found significant correlations ($P < 0.05$) between GI and TDF, protein, and phytate, also the analysis suggested that the method of food preparation and the characteristics of starch and starch granules might be more critical in predicting GI among starchy foods than the content of any single component. Reference [108] involved a dataset that comprised glycaemic index measurements from a randomized crossover meal test with 28 healthy young men. Participants consumed 13 different breakfast meals and a reference meal, each containing 50 grams of available carbohydrates but varying significantly in energy and macronutrient composition. Venous blood samples were collected over a two-hour period to analyze glucose and insulin responses. The study aimed to assess whether the GI of mixed meals, calculated using standard GI tables, accurately predicted the measured GI.

The dataset in [104] encompasses various cereal and legume-based food products, with detailed information on their macronutrient compositions, including carbohydrate, protein, fat, and fiber contents. Additionally, the dataset includes measured Glycemic Index values for these foods, obtained through in vivo testing. Researchers then developed predictive models to estimate the GI of foods based on their macronutrient profiles. These models aimed to identify relationships between macronutrient composition and GI, facilitating the prediction of GI for similar foods without the need for extensive in vivo testing. Reference [106] introduced a mixed-integer programming model aimed at minimizing the

total daily glycemic load of foods while satisfying daily nutritional and serving size requirements. The dataset employed comprises 177 foods, with their nutritional information and GL values sourced from the U.S. Department of Health and Human Services and the U.S. Department of Agriculture (USDA) guidelines. This dataset is not a standard benchmark but is constructed by the researchers using publicly available nutritional data. In the study, the dataset is utilized to perform experimental analyses, applying robust optimization techniques to account for uncertainties in GL values. Participants for the research [90] were 79 children diagnosed with type 1 diabetes. For each participant, various factors were recorded, including demographic information, biological markers, and socioeconomic status. Machine learning algorithms were employed to train predictive models using the selected features. The objective was to forecast glycemic control, focusing on achieving an A1C level below 7.5%, as recommended by organizations such as the American Diabetes Association (ADA) and the International Society for Pediatric and Adolescent Diabetes (ISPAD). This approach aimed to enhance the understanding of factors influencing glycemic control and to improve predictive capabilities in clinical settings, ultimately contributing to better management strategies for children with type 1 diabetes. Reference [31] comprised dietary records from 131 participants following various modern diets. Participants’ dietary intakes were recorded and analyzed using the Nutrition Data Systems for Research (NDSR) software. This process involved detailed logging of food consumption to assess diet quality. The collected data were used to calculate three key dietary indices: Healthy Eating Index, GI, and GL, which evaluate diet quality based on adherence to dietary guidelines. Artificial Intelligence and Machine Learning techniques were applied to the dataset to identify predictors of the dietary indices. Factors such as whole fruit and whole grain consumption were found to be significant predictors of HEI, while carbohydrate intake was a common predictor for both GI and GL. The dataset comprises data from laboratory analyses conducted on wheat-based bread samples fortified with varying concentrations (0.25% to 2%) of Euryale ferox seed shell extract in [26]. The researchers assessed the inhibitory effects of EFSSE on α -amylase and α -glucosidase activities, and evaluated the in vitro starch digestibility (IVSD) and predicted Glycemic Index of the bread samples. Advanced computational techniques, including Swarm Intelligence supervised neural network modeling, were employed to simulate digestion kinetics and predict the Glycemic Index, providing insights into the potential of EFSSE as a functional additive for producing lower Glycemic Index bread.

The study titled “Moroccan Food Dataset for Food Image Recognition Towards Glycemic Index Estimation”, [11] introduced the MFOOD-70 dataset, a collection of 70 Moroccan food categories comprising 14,000 images. This dataset was specifically developed by the authors to enhance food image recognition and facilitate Glycemic Index estimation. The images were sourced from web scraping and existing datasets, ensuring a diverse representation of Moroccan cuisine. The dataset was utilized to train and evaluate convolutional neural network models, aiming to improve the accuracy of food recognition systems and support dietary monitoring applications. In reference [28], the dataset comprises data from 10 healthy non-diabetic volunteers (5 males and 5 females). Each participant consumed 50 grams of carbohydrate from different black rice cultivars after an overnight fast. Blood glucose levels were

measured at intervals of 0, 15, 30, 45, 60, 90, and 120 minutes post-consumption. The researchers calculated the Incremental IAUC for each rice cultivar and compared it to the IAUC of a reference food (glucose) to determine the Glycemic Index values. The study found that the GI values of the black rice cultivars ranged from 44.6 to 59.7, indicating that these cultivars have a low to medium GI. The dataset was built by the researchers specifically for this study to assess the GI of selected black rice cultivars in Bangladesh. The dataset described in [24] primarily consists of various rice genotypes used to study their Glycemic Index and associated biochemical properties such as resistant starch content, amylose content, and other indicators of starch digestibility. These genotypes were sourced from different ecologies or traits and analyzed using *in vitro* methods to determine their GI values and RS levels. This dataset aids in understanding how genetic and biochemical variations in rice influence its digestion and postprandial glucose response, with implications for dietary recommendations and crop improvement strategies. Reference [23] developed a custom dataset specifically by formulating various bread samples by substituting wheat flour with chickpea flour, red chicory powder, and different types of resistant starch. They then assessed the predicted Glycemic Index and technological properties of these bread formulations. The dataset includes measurements of pGI, moisture content, volume, specific volume, baking loss, and texture parameters such as hardness, cohesiveness, and chewiness. This comprehensive dataset enabled the researchers to analyze how each ingredient influenced the bread's Glycemic Index and technological characteristics. Reference [65] encompasses various food items, each characterized by both nutrient-based information and non-nutrient physicochemical properties such as texture, water activity, glycemic potential, satiety potential, and shelf life. The researchers employed data mining techniques to analyze the compiled dataset, aiming to establish correlations between the degree of food processing and the physicochemical properties of the foods. By examining these relationships, they sought to develop a comprehensive quantitative index that reflects the extent of food processing, moving beyond traditional nutrient-based indices. This holistic index is intended to provide a more nuanced understanding of how processing affects food quality and health implications. 106 participants, including 53 colorectal cancer cases and 53 family members from diverse ethnic backgrounds, participated in creating the dataset of reference [66]. The data encompassed individual dietary parameters, health outcomes, and demographic information. The researchers employed machine learning validation procedures, such as the ensemble method and generalized regression prediction, to analyze the data. Significant dietary predictors identified included whole fruit, milk or milk alternatives, whole grains, saturated fat, and oils and nuts. These findings highlight the importance of specific dietary components in promoting healthy eating habits among multi-ethnic colorectal cancer families. The dataset employed in [36] comprises various Sri Lankan starchy tubers, including arrowroot, cassava, potato, purple yam, sweet potato, and white yam. The researchers collected these tubers from local sources and prepared them under controlled laboratory conditions to assess their starch hydrolysis indices. Each tuber was subjected to enzymatic digestion to measure the rate and extent of starch breakdown over time. These measurements enabled the calculation of the Hydrolysis Index for each tuber, which serves as an indicator of the potential glycemic response upon consumption. The

findings provide valuable insights into the nutritional properties of these traditional Sri Lankan tubers, particularly concerning their impact on blood sugar levels.

The study [29] utilizes a dataset comprising real-world data from individuals with Type 1 Diabetes. This dataset includes patient-specific information such as blood glucose levels, insulin doses, and nutritional intake. The researchers collected this data to develop a machine learning model capable of predicting postprandial blood glucose levels at various time intervals (15, 30, 45, and 60 minutes) following a meal. By incorporating these nutritional factors, the model aims to enhance the accuracy of blood glucose predictions, thereby supporting better management of T1D. AI4FoodDB is a public database developed by researchers to support personalized e-health nutrition and lifestyle studies in [92]. It was constructed from a nutritional weight loss intervention involving 100 overweight and obese participants over one month. The dataset includes various types of data collected through manual methods, clinical assessments, and digital tools such as wearable devices. The database comprises several distinct datasets: anthropometric measurements, lifestyle and health, nutrition, biomarkers, physical activity, sleep activity, emotional state, etc. These datasets are utilized to analyze the relationships between various lifestyle, biological, and digital factors and health outcomes. By integrating diverse data sources, AI4FoodDB facilitates the development of artificial intelligence techniques aimed at advancing personalized healthcare.

In the study [93] the researchers constructed a custom fruit dataset specifically for their research on diabetic patients' daily diets. This dataset was not sourced from existing benchmarks but was developed to facilitate the identification of fruits with high sweetness and low glycemic load values. The dataset was utilized to train and evaluate an improved Faster R-CNN network, which incorporated an attention mechanism module during feature extraction, adjusted the anchor aspect ratio of the Region Proposal Network (RPN), and implemented a fusion update operation in the fully connected layer. These enhancements aimed to improve the precision and recall rates of fruit recognition, ultimately assisting diabetic patients in making informed dietary choices.

These trials provided empirical evidence supporting the algorithm's efficacy and safety in managing T1D. The authors collected spectral data for the study [77] from rice samples using a portable Near-Infrared sensor operating in the 740–1070 nm wavelength range. The collected spectral data were then analyzed using machine learning techniques, including principal component analysis, linear discriminant analysis, random forest classifier, and partial least squares regression, to develop predictive models for rice quality attributes such as Glycemic Index, amylose content, and viscoelasticity. These models aimed to provide rapid, on-site evaluation of rice quality. Similarly, a dataset collected by the researchers through a randomized trial involving adults with prediabetes using waist-worn or wrist-worn wearables to monitor their activity patterns in reference [88]. Baseline information, including demographics, medical history, and laboratory test results, was also gathered. The study developed predictive models to assess changes in hemoglobin A1c levels, an indicator of glycemic control. The models compared traditional regression methods with machine learning approaches, finding that ensemble machine learning methods provided better predictions. Additionally, incorporating wearable data alongside baseline information improved prediction accuracy. Notably, wrist-worn wearables yielded

more accurate predictions compared to waist-worn devices. These findings suggest that integrating wearable-derived activity data with clinical information can enhance the prediction of glycemic control changes in individuals with prediabetes.

Another research focuses on developing a personal device to assist diabetic patients in managing insulin therapy [105]. The researchers collected data from diabetic patients, including blood glucose levels, insulin dosages, meal information, and other relevant health metrics. This data was gathered using the developed personal device integrated with various sensors and input methods. The collected data was utilized to test and refine the device's algorithms for calculating insulin dosages, considering factors such as patient weight, glucose levels, physician recommendations, and carbohydrate absorption. The dataset supported the integration of the personal device with a Glycemic Index information system, nurses' and physicians' desktop applications, and a patient web portal, facilitating comprehensive diabetes management within an Ambient Assisted Living environment. Reference [79] comprised 92 blood samples collected from individuals in Santa Cruz do Sul, Brazil, with informed consent. The dataset contains mid-infrared spectra of peripheral blood samples, which were analyzed using diffuse reflectance infrared spectroscopy (DRIFTS). The spectral data were processed to quantify biochemical parameters such as total cholesterol, using artificial neural networks. The ANN achieved a correlation coefficient (r) of 0.81 and a root mean square error (RMSE) of 30.14 in the preliminary trial. Future plans include expanding the dataset to 500 samples to enhance accuracy and include other parameters like HDL, LDL, triglycerides, and glucose. 13 qualified individuals (8 men and 5 women) participated in [81]. Blood samples were taken in the fasting state and at 15, 30, 45, 60, 90, and 120 minutes after ingestion. The blood glucose levels measured at the specified intervals were used to assess the body's glycemic response to both the reference and test foods. Results suggest that the nutritional product elicits a lower glycemic response compared to glucose, making it suitable for individuals managing diabetes mellitus. Reference [38] utilized a dataset collected by 18 healthy volunteers with fasting plasma glucose levels below 100 mg/dL. Participants consumed three different test foods in a randomized sequence: a complete nutrition drink containing retrograded starch, a glucose solution, and white bread. Plasma glucose and insulin levels were measured at baseline and at multiple time points up to 180 minutes post-consumption. The dataset facilitated the assessment of postprandial insulin responses, revealing that the complete nutrition drink led to a sustained increase in plasma insulin levels over the 3-hour period, in contrast to the more rapid decline observed with glucose solution and white bread.

The dataset in [89] collected as part of the "Smart District 4.0 Project", supported by the Italian Ministry of Economic Development. The study involved six patients with diabetes. Glycemic values were recorded every 3 minutes using specialized monitoring devices. The number of observations varied among patients; for instance, Patient A had 243 observations, while Patient B had approximately 13,204 observations. Eight different algorithms were employed to predict the glycemic status of the patients: artificial neural network, Probabilistic Neural Network, Polynomial Regression, Gradient Boosted Trees Regression, Random Forest Regression, Simple Regression Tree, Tree Ensemble Regression, and Linear Regression. The models were evaluated based on their ability to minimize four statistical

errors: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error, and Mean Signed Difference (MSD). The study aimed to identify the most efficient algorithm for predicting glycemic status by comparing these errors across the different models. Twenty-six participants, who contributed to the creation of the dataset of [87], also wore non-invasive, wrist-worn wearable devices in conjunction with continuous glucose monitors for 8–10 days after undergoing a clinical HbA1c measurement. The wearables captured physiological data, which were then analyzed to estimate glucose variability metrics and HbA1c levels. The study developed 27 models to estimate glucose variability metrics using data from the non-invasive wearables, achieving high accuracy (mean average percent error (MAPE), of less than 10%) in 11 of these models. Additionally, the HbA1c estimation model achieved a MAPE of 5.1% on an external validation cohort. This proof-of-concept study demonstrated the feasibility of using non-invasive wearables for glycemic monitoring, potentially offering a more convenient and less invasive method for patients to monitor their glucose levels and HbA1c remotely.

Likewise, the dataset in [35] comprises data collected from participants using wearable sensors and mobile devices to monitor food intake, physical activity, and corresponding blood glucose levels. Relevant features, such as meal timing, nutritional content, activity type, duration, and intensity, are extracted from the raw data to serve as inputs for the predictive models. Deep Learning algorithms are employed to analyze the extracted features and predict blood glucose levels based on observed patterns in food consumption and physical activity. This approach aims to develop a non-invasive method for monitoring blood glucose levels by leveraging deep learning techniques to interpret lifestyle data. 1,159 adults aged 20–74 years with type 2 diabetes and HbA1c levels between 6.0–8.9 (42–74 mmol/mol) were involved in [100]. The participants' health metrics were continuously monitored using IoT devices, and this real-time data was analyzed to assess the effectiveness of the IoT-based intervention on glycemic control over a 52-week period. The primary endpoint was the change in HbA1c levels from baseline to the final measurement at 52 weeks. The study concluded that the IoT-based approach did not significantly reduce HbA1c in patients with type 2 diabetes. The authors suggested that incorporating daily glycemic control data and HbA1c levels into the IoT-based intervention may be necessary to improve glycemic control.

5.2 Utilization of benchmark datasets in technological approaches for Glycemic index assessment

The 2016 study [62] utilized datasets from two European Union-funded projects:

- DIAdvisor (EU FP7-funded project): This dataset comprised clinical trial data, including intermittent blood glucose measurements from patients with type 1 diabetes. The data were used to calibrate and test nocturnal hypoglycemia (NH) predictors based on various glycemic control indices (GCI).
- AMMODIT (EU Horizon 2020-funded project): This dataset was employed to validate the portability and effectiveness of the proposed NH prediction approach across different patient populations.

The authors developed a method to predict nocturnal hypoglycemia by aggregating predictors constructed from different GCIs, such as the Low Blood Glucose Index. They applied machine learning techniques to combine these predictors, aiming to enhance the accuracy of NH predictions.

The datasets provided the necessary blood glucose measurements to calibrate and test the performance of the aggregated predictors, demonstrating improved sensitivity and specificity in predicting NH events. The dataset used in [63] includes Glycaemic Index values from an existing comprehensive list (e.g., a glycaemic index corpus) and food nutritional composition data (such as macronutrient content) from organizations like the USDA. The dataset is not created by the authors but sourced from these established repositories. The study highlights challenges in integrating these datasets due to differences in how foods are labeled and categorized. The data is used to predict the GI of foods based on their biochemical properties using machine learning techniques. The authors manually cross-linked a subset of 100 food entries to ensure reliability and used features like water, energy, protein, carbohydrate, sugar, fiber, and lipid content to build predictive models. This approach aimed to explore the feasibility of using widely available food data for practical GI prediction.

Reference [72] exploits a dataset of 55, 102 cooking recipes collected from the Allrecipes website. This dataset was built by the researchers specifically for their analysis. A subset of 1,000 recipes was selected for further analysis. These recipes were annotated for glycemic impact through online crowdsourcing using Amazon Mechanical Turk. From the collected data, both textual features (e.g., ingredients and cooking directions) and nutritional features (e.g., carbohydrate and sugar content) were extracted to represent each recipe. Machine Learning models were trained using the annotated subset to classify recipes as healthy or unhealthy for diabetics based on their glycemic impact. This approach combines online crowdsourcing and machine learning to estimate the glycemic impact of cooking recipes, offering a data-driven method to assist diabetics and pre-diabetics in making informed dietary choices. Reference [17] presented a systematic six-step methodology to assign Glycemic Index values to over 600 foods in the Venezuelan food composition database. The process begins with compilation, where GI values from international sources are gathered. Next, the matching step aligns these values with local foods based on ingredient and preparation similarities. For foods lacking direct matches, estimation is employed by analyzing macronutrient composition and comparing it with analogous foods. This is followed by validation, ensuring accuracy through cross-referencing with literature and expert opinions. The validated GI values are then integrated into the national database. Finally, documentation provides a transparent record of the sources and rationale, ensuring traceability and facilitating updates. This methodology enriches the database with reliable GI values, supporting dietary planning and nutritional evaluation tailored to the Venezuelan context.

In 2019, reference [33] utilized a dataset comprising the macronutrient compositions of 42 breakfast cereals. This dataset was collected from existing nutritional information available for these products. The researchers did not generate new experimental data but instead relied on published macronutrient profiles to develop their predictive model. The dataset was employed to create a model that predicts the Glycemic Index and glycemic load of foods based on their macronutrient content. By analyzing the relationship between the macronutrient composition and the GI/GL values, the model quantifies the impact of glycemic carbohydrates and the GI-lowering effects of other nutrients such as proteins, fats, and fibers. Reference [30] utilizes the PIMA Indians Diabetes dataset, a benchmark dataset

provided by the National Institute of Diabetes and Digestive and Kidney Diseases. The numerical data from the PIMA dataset are transformed into image representations. Each feature is assigned a specific location and size within the image based on its importance, determined using the ReliefF feature selection algorithm. This approach enables the application of convolutional neural networks designed for image data. To enhance the dataset, data augmentation techniques are applied to the generated images, artificially increasing the number of training samples and improving model robustness. The augmented image data are used to train deep learning models, specifically ResNet18 and ResNet50 CNN architectures, for predicting pre-diabetic conditions. The innovative approach leverages image-based deep learning techniques to enhance the prediction of pre-diabetic conditions using a well-established benchmark dataset. Reference [91] employed data from the All of Us Research Program, a comprehensive initiative by the U.S. National Institutes of Health aimed at gathering health data from diverse populations to advance precision medicine. Researchers applied various machine learning algorithms, including random forest, extreme gradient boost, logistic regression, and a weighted ensemble model, to predict uncontrolled diabetes. They identified patients aged 18 and above with diabetes from the All of us dataset and defined uncontrolled diabetes based on specific International Classification of Diseases codes. The models incorporated features such as basic demographics, biomarkers, and hematological indices. Among these, the random forest model demonstrated the highest performance, achieving an accuracy of 80% and an area under the receiver operating characteristic curve of 0.77. Key predictors of uncontrolled diabetes included serum potassium levels, body weight, aspartate aminotransferase, height, and heart rate.

Reference [98] involved the creation of a comprehensive database by the researchers themselves. This database integrates Glycemic Index and glycemic load values with dietary data from the National Health and Nutrition Examination Survey (NHANES) spanning 1999 to 2018. The researchers employed an artificial intelligence-enabled model to align GI values from existing databases with NHANES food codes. This process was manually validated to ensure accuracy, resulting in GI values covering 99.9% of total carbohydrate intake. This newly developed database serves as a valuable resource for large-scale epidemiologic studies, enabling researchers to assess the impact of carbohydrate quality on health outcomes within the U.S. population. Reference [99] comprises 1,000 records from the Diabetes Complication Early Warning Dataset provided by the National Clinical Medical Sciences Data Center. The dataset underwent preprocessing to address missing values and outliers. Feature selection was performed using information gain to identify the most relevant variables. Subsequently, the authors developed a diabetic retinopathy risk prediction model employing the CatBoost algorithm, an advanced machine learning technique. To enhance the interpretability of the model's predictions, they applied SHAP values, which elucidate the contribution of each feature to the model's output. This approach enabled the identification of key risk factors associated with diabetic retinopathy, such as poor renal function, elevated blood glucose levels, liver disease, hematonosis, and dysarteriotony. The integration of Machine Learning with interpretable models facilitated a more transparent understanding of the factors influencing diabetic retinopathy risk. PhysioNet, a public database, is used in [21]. This dataset comprises physiological and nutritional

information collected via wearable devices and dietary surveys. The researchers employed this existing dataset to analyze the relationships between blood glucose levels and various physiological and nutritional factors. They conducted correlation analyses, multiple regression analyses, and one-way analyses of variance to explore how different physiological indicators and nutritional values are associated with blood glucose fluctuations.

The public dataset, D1NAMO is used in [83], an open-source collection of real-time Glycemic Index readings. This dataset comprises continuous glucose monitoring data collected every 5 minutes from 9 diabetic patients and six times daily from 20 non-diabetic individuals. The data was gathered under normal conditions using the Zephir Bioharness 3 wearable device. They implemented an auto-adaptive algorithm for optimizing ARIMA model parameters, enabling real-time predictions in an online learning environment. The dataset's comprehensive CGM readings facilitated the training and evaluation of these models, aiming to enhance glycemic control through accurate short-term forecasts. Similarly, reference [73] utilizes benchmark datasets from continuous glucose monitoring devices worn by patients with Type 1 Diabetes. The CGM data is segmented into feature vectors using a sliding window technique, capturing blood glucose readings over specific time intervals. This method generates training examples that reflect the temporal dynamics of glucose levels. Machine learning models, including Support Vector Regression and Multilayer Perceptron, are trained on the processed and balanced datasets.

The article [40] is a consensus statement that reviews existing research and provides expert opinions on the Glycemic Index, glycemic load, and glycemic response. As a consensus statement, it does not introduce new experimental data or utilize a specific dataset. Instead, it synthesizes findings from numerous studies to offer guidance on the application and interpretation of GI and GL in nutrition science and public health. The document serves to consolidate scientific understanding and provide recommendations based on a comprehensive review of existing literature. The article [27] is a review that synthesizes existing research on the Glycemic Index of rice and its products. It does not introduce a new dataset; rather, it compiles and analyzes data from various studies published up to December 2022. The authors conducted a comprehensive literature review, gathering information from the Web of Science and Scopus databases. They categorized the findings into four main sections: basic information about starch digestion and recent advanced measurement methods, the mechanism of the effect of various factors on GI, recent advanced technologies to modulate GI, and a table of the Glycemic Index for rice and rice products in different countries. This compilation provides an overview of the GI values of different rice varieties and discusses the impact of various factors and processing techniques on the GI of rice products.

The review paper [94] provides consensus guidelines for machine learning practitioners in diabetes care. It reviews common features used in machine learning applications for glucose control and offers an open-source library of functions for calculating these features. Additionally, it provides a framework for specifying datasets using data sheets and reviews current datasets available for training algorithms, along with an online repository of data sources. These resources are designed to improve the performance and translatability of new machine learning algorithms developed

in the field of diabetes. Reference [95] is also a systematic review that evaluates various mobile computer vision-based approaches for food classification, volume estimation, and nutrient estimation. As a review, it does not introduce a new dataset but rather examines existing methods and the datasets they utilize. The datasets referenced in the reviewed studies vary; some are proprietary datasets developed by researchers, while others are benchmark datasets commonly used in the field. These datasets are employed to train and validate models that can accurately classify food items, estimate portion sizes, and assess nutritional content, which are crucial for managing dietary intake in individuals with diabetes. The review article [101] published in the Alexandria Engineering Journal in 2024 conducted a comprehensive literature review, synthesizing information from various studies and sources to discuss the advancements and trends in wearable glucose monitoring technologies. This approach involves aggregating and analyzing existing research findings rather than applying a new or benchmark dataset.

The dataset used in reference [28] comprises two parts: The "Indian Images Top (20)" dataset available on Kaggle, containing 3996 images from 20 different Indian food classes, and a custom dataset created for the research, which includes nutritional information like Glycemic Index, protein, fats, and carbohydrates for the food items. The Kaggle dataset serves as the primary image dataset for training and testing the Inception V3 model for food classification. Data augmentation techniques such as rotation, shearing, and horizontal flipping were applied to increase the dataset to 4996 images, ensuring better model generalization. The custom dataset complements this by providing essential nutritional details to enable personalized food recommendations. Thus, the research combines a public benchmark dataset with a custom-built dataset tailored for its objectives. The dataset in the document "Index of foods: A review [34] is derived from existing literature, using databases like MEDLINE, PubMed, Scielo, and Google Scholar. It is not an original dataset created by the authors but rather a compilation of data from prior research studies. This approach allows the authors to summarize findings on the Glycemic Index and its influence on health, using these benchmark sources. The dataset is utilized to analyze patterns and outcomes related to dietary habits, carbohydrate types, and their metabolic impacts, emphasizing their relevance in managing chronic diseases like diabetes and obesity.

5.3 Studies utilizing both benchmark and custom data

The dataset used in [44] combines experimental data and benchmark GI data produced specifically for this research. Experimental data were measured through in vivo methods based on FAO/WHO protocols. These measurements are combined with benchmark GI data (sourced from international tables and scientific literature) to train an artificial neural network. These data are then analyzed using an ANN to establish a predictive model for Glycemic Index values. This approach provides a more cost-effective and faster alternative to in vivo testing, allowing for the prediction of GI with high accuracy, as evidenced by the study's cross-validation results ($R^2 = 0.89$). Another study has included [43] both benchmark values for Glycemic Index from prior in vivo studies and new experimental data derived from in vitro digestion methods combined with High-Performance Liquid Chromatography (HPLC) analysis. Also, in reference [42], the dataset used is drawn from both experimental data and existing GI benchmarks. It includes measured Glycemic Index values for foods like rice and breakfast cereals, using

glucose and starchy reference foods. These were collected through controlled experiments involving European, Chinese, younger, and older participants. Benchmark GI values, such as those for glucose and white bread, were used as references for comparison and scaling. The dataset was employed to explore how reference foods influence observed GI values across populations. It was also used to assess variability in GI results due to participant factors like ethnicity or age, and to evaluate the appropriateness of different reference foods for GI testing.

The review paper [37] presents an updated compilation of Glycemic Index and GL values for various foods. The dataset used in this study is an extensive collection of both published and unpublished GI values sourced from global research conducted between January 1, 2008, and June 30, 2020. The authors systematically reviewed and tabulated these sources, adhering to the International Standards Organization (ISO) methodology to ensure data reliability. The dataset includes over 4,000 food items, representing a 61% increase from the previous edition published in 2008. This comprehensive dataset serves as a valuable resource for researchers and healthcare professionals, facilitating a better understanding of the glycemic impact of various foods and aiding in the development of dietary recommendations. Reference [97] employed datasets from two primary sources: in-silico simulations and the OhioT1DM dataset. The in-silico cohorts comprised 20 and 47 virtual patients, respectively, designed to mimic real-world scenarios. The OhioT1DM dataset is a publicly available collection of data from individuals with type 1 diabetes, including continuous glucose monitoring data. The researchers employed a heterogeneous ensemble method combining artificial neural networks, random forests, and logistic regression to develop a meal detection model. This model was trained and tested on both the in-silico and OhioT1DM datasets to enhance its robustness and accuracy. The ensemble majority voting approach achieved high sensitivity and precision in detecting unannounced meals, thereby improving postprandial glucose control.

A review paper [37] presents an updated compilation of Glycemic Index and glycemic load values for various foods. The dataset used in this study is an extensive collection of both published and unpublished GI values sourced from global research conducted between January 1, 2008, and June 30, 2020. The authors systematically reviewed and tabulated these sources, adhering to the International Standards Organization methodology to ensure data reliability. The dataset includes over 4,000 food items, representing a 61% increase from the previous edition published in 2008. This comprehensive dataset serves as a valuable resource for researchers and healthcare professionals, facilitating a better understanding of the glycemic impact of various foods and aiding in the development of dietary recommendations.

6. Research gaps and future directions

The timeline analysis (Figure 7) reveals that a significant portion of Glycemic Index research was conducted during the 2015-2024 period, reflecting a concentrated effort to explore foundational techniques in GI prediction and analysis. However, much of this work has relied on traditional statistical methods and machine learning, with limited integration of emerging technologies such as Explainable AI, Deep Learning, and Reinforcement Learning. The consistent presence of machine learning (green category) underscores its foundational role in GI research, but future studies should explore ensemble methods and meta-learning to improve

predictive performance, particularly for diverse food types and ripeness stages. Deep learning's growing role since 2015 presents opportunities for leveraging advanced architectures such as transformer-based models or multimodal learning. By combining diverse data sources, including biochemical food properties, imaging data, and textual descriptions, these methods can significantly enhance the accuracy and scope of GI prediction. Statistical learning, while enduringly relevant, can be effectively combined with modern deep learning approaches to create hybrid models for interpretable and robust predictions. Reinforcement Learning, with only one study in the current collection, is an underexplored yet promising direction. RL agents can dynamically predict GI based on ripeness levels by interacting with sensor networks in food supply chains, using iterative feedback to enhance accuracy. Another promising application involves RL-driven dietary recommendation systems that adjust in real-time based on users' blood glucose levels, dietary preferences, and lifestyle factors such as exercise and stress. With advancements in wearable technology and IoT devices, RL-based systems could revolutionize personalized glycemic management by minimizing glycemic spikes and optimizing dietary plans tailored to individual needs.

The mapping of the literature by region revealed notable insights into the geographical distribution of research activities. As shown in Figure 9, the majority of GI-based research utilizing AI technologies has been conducted in North American countries. Additionally, several Asian countries, such as India, have contributed significantly to this field with a substantial number of publications. These findings highlight the global interest in GI-based research while also emphasizing regional disparities in research output, suggesting opportunities for further contributions from underrepresented regions.

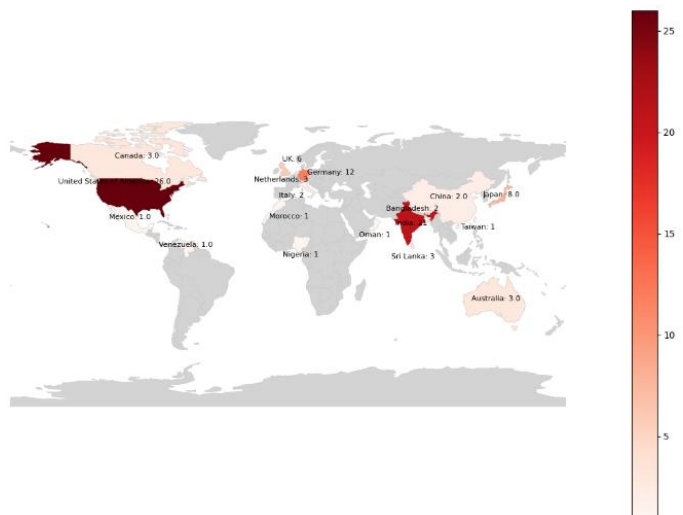


Figure 9: Number of papers occurrences by geographic region

The dataset analysis (Figure 8) highlights the increasing use of custom datasets, which often integrate continuous glucose monitoring (CGM) data, wearables, and IoT devices to monitor food intake, activity patterns, and health metrics. While custom datasets enable granular insights, existing benchmark datasets like NHANES, DIAdvisor, AMMODIT, D1NAMO, and OhioT1DM remain crucial for standardization. Hybrid datasets that combine custom and

benchmark data are emerging as a powerful approach, offering the reliability of standardized data along with the specificity of real-world measurements. Additionally, datasets sourced from recipe platforms like allrecipes provide valuable nutritional insights, complementing structured survey data and enhancing the comprehensiveness of research.

Looking ahead, integrating these diverse datasets with AI-driven models holds immense potential for advancing GI research. Automated image analysis can improve ripeness and portion-size estimation, while wearable data and IoT systems can facilitate real-time and longitudinal studies. Such approaches will enhance our understanding of GI variability across populations and time scales, enabling personalized interventions and population-level dietary recommendations. Finally, the limited application of Explainable AI in GI research represents a critical gap. XAI tools, such as SHAP, have only been applied in two studies, highlighting an opportunity for future work. By integrating XAI into GI prediction models, researchers can improve transparency and trust while uncovering the factors driving glycemic variability, such as food preparation methods, ripeness, and individual metabolic responses. XAI has the potential to bridge the gap between advanced AI techniques and their practical, interpretable application in healthcare, paving the way for more effective and user-centric glycemic management solutions. By addressing these gaps and leveraging advanced AI methodologies, future research can unlock transformative potential in Glycemic Index prediction, management, and personalized healthcare.

The future of Glycemic Index-related studies can lead to providing personalized nutrition advice by integrating CGM devices, wearable sensors, and mobile health applications that allow AI models to learn users' real-time responses. These models can predict individual GI values using each person's physiology, activity level, stress, and circadian rhythm. These AI systems can help individuals maintain a stable glucose level by giving personalized meal plans and precision dietary interventions. Another emerging avenue is multimodal data fusion, where different data sources such as food composition, metabolic responses, environmental context, and even emotional or behavioral cues can be used to build a holistic model for glycemic dynamics. The fusion can include different data types such as image-based meal recognition, nutrient text analysis, and CGM data combined to build context-aware predictive models to monitor glucose level fluctuation.

Figure 10 indicates the main objective of the research covered by the GI-related studies. The significant number of studies dedicated to predicting Glycemic Index using machine learning and artificial intelligence underscores a prevailing trend in current research. This focus reflects the scientific community's commitment to leveraging advanced computational methods to forecast GI values accurately, thereby enhancing dietary recommendations and metabolic health management. In contrast, the relatively limited research on developing low-GI foods, with only one study [26] identified highlights a notable gap in the literature. Addressing this disparity presents a valuable opportunity for future investigations to concentrate on creating and promoting low-GI food options. Such efforts could significantly contribute to dietary interventions aimed at improving glycemic control and reducing the risk of metabolic disorders. Pilot studies have demonstrated the feasibility of implementing low-GI diets in primary care settings, suggesting that with appropriate support and

resources, patients can successfully adopt these dietary changes. However, these studies also indicate the need for larger-scale research to confirm the benefits and practicality of such interventions across diverse populations. In summary, while substantial progress has been made in predicting GI through ML and AI, there is a pressing need for future research to prioritize the development of low-GI foods and to conduct comprehensive studies evaluating the effectiveness of personalized nutrition strategies. Such endeavors will be crucial in advancing dietary recommendations and improving health outcomes related to glycemic control.

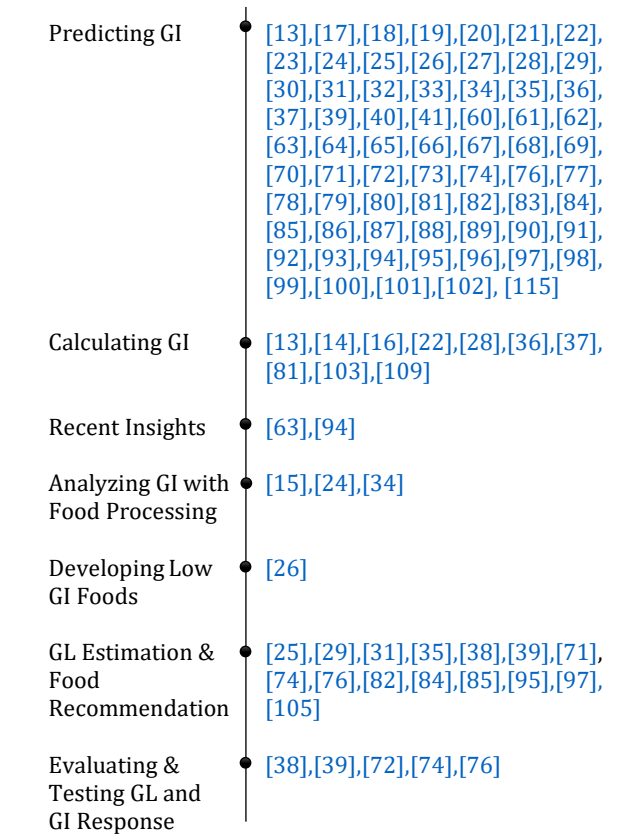


Figure 10. Area covered by the research

7. Conclusion

The Glycemic Index serves as a vital indicator for understanding how foods influence blood glucose levels, playing a key role in managing diabetes and promoting healthy dietary habits. While predictive modeling dedicated to GI remains sparse, recent advances in machine learning approaches have enabled more accurate estimations of GI and inter-individual glycemic responses than traditional methods. Deep learning techniques, in particular, have demonstrated their effectiveness in uncovering complex patterns within data, offering scalable and precise GI predictions for a wide range of foods. This advancement paves the way for significant contributions to personalized nutrition and dietary recommendations. Interestingly, reinforcement learning has not yet been extensively explored in GI-focused research, marking an open area for future investigation. The potential of RL in this domain offers exciting opportunities to expand the scope of AI-driven solutions for dietary management. Additionally, although image processing techniques were not a primary focus in this study, they have emerged as crucial facilitators for enhancing the application

of other AI methodologies in GI research. Only two studies in the reviewed literature directly employed explainable AI techniques, specifically SHAP, to enhance the interpretability of their models. While other studies did not explicitly integrate XAI methodologies, they made efforts to validate and clarify how AI algorithms arrived at specific decisions or predictions. To achieve this, several methods were employed to improve model interpretability. Linear regression models, with their straightforward representation of the relationship between input features and target variables, provided clear insights into feature influence. Decision trees, by creating a series of binary choices, offered an intuitive, tree-like structure to trace decision-making paths. Similarly, rule-based systems used “if-then” rules to form logical and transparent reasoning processes, making them valuable tools for understanding machine learning model outputs. Most studies reviewed utilize continuous glucose monitoring data as inputs for their models, reflecting a trend toward leveraging real-time and highly granular data. Early AI-based GI research predominantly relied on statistical learning techniques; however, with the rise of ML and advanced DL approaches, the field has shifted toward leveraging these powerful tools for deeper insights and improved accuracy. The majority of studies employ self-developed datasets, as these datasets are specifically designed and curated to meet the unique objectives and requirements of the research. By tailoring the data to the problem at hand, researchers can ensure that the models are trained on highly relevant and domain-specific information, which significantly improves their accuracy, reliability, and overall performance. This customized approach also allows for better control over the quality and diversity of the data, addressing potential gaps or biases that may be present in publicly available datasets. Consequently, the use of self-developed datasets not only enhances the precision of the models but also ensures that the outcomes are better aligned with the intended purpose of the study. In conclusion, the intersection of AI and GI research is at an exciting juncture, with deep learning and emerging techniques like reinforcement learning presenting untapped potential. These advancements hold promise for addressing existing gaps in GI prediction, improving scalability, and enabling more personalized approaches to nutrition and healthcare. This review distinguishes itself by connecting glycemic index research with a wide range of AI techniques, emphasizing methodological diversity and revealing underexplored opportunities in explainable and reinforcement learning for food analytics.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The authors declare no potential conflict of interest.

References

- [1] B. R. Group, “AI and The Future of Healthcare: A 2024 Survey,” Berkeley Research Group, 2024. [Online]. Available: <https://media.thinkbrg.com/wp-content/uploads/2024/02/26113735/BRG-Report-AI-and-The-Future-of-Healthcare.pdf>
- [2] World Economic Forum, “World Economic Forum; Annual Report 2023-2024,” World Economic Forum, Geneva, Switzerland, 2024. [Online]. Available: <https://www.weforum.org/publications/annual-report-2023-2024/>
- [3] L. Jiaying, “Cutting-edge tech boosting diagnostic accuracy,” China Dly., 2024, [Online]. Available: <https://www.chinadaily.com.cn/a/202407/24/WS66a06492a31095c51c50fa09.html>
- [4] R. Awasthi et al., “Artificial Intelligence in Healthcare: 2023 Year in Review,” medRxiv, pp. 2024–02, 2024.
- [5] World Health Organization, “Noncommunicable diseases.” Sep. 13, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>
- [6] A. Kumar, R. Gangwar, A. Ahmad Zargar, R. Kumar, and A. Sharma, “Prevalence of diabetes in India: A review of IDF diabetes atlas 10th edition,” Curr. Diabetes Rev., vol. 20, no. 1, pp. 105–114, 2024.
- [7] K. O. Olimjonovna, “UNDERSTANDING THE CAUSES AND RISK FACTORS OF DIABETES,” Biol. VA Kim. FANLARI ILMIY JURNALI, vol. 2, no. 5, pp. 1–7, 2024.
- [8] N. A. ElSayed et al., “Introduction and methodology: standards of care in diabetes—2023,” Diabetes care, vol. 46, no. Supplement_1. Am Diabetes Assoc, pp. S1–S4, 2023.
- [9] N. G. Forouhi and N. J. Wareham, “Epidemiology of diabetes,” Medicine (Baltimore), vol. 47, no. 1, pp. 22–27, Jan. 2019, doi: 10.1016/j.mpmed.2018.10.004.
- [10] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and t PRISMA Group*, “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement,” Ann. Intern. Med., vol. 151, no. 4, pp. 264–269, 2009.
- [11] M. Mansouri, S. Benabdellah Chaoui, S. Jai Andaloussi, and O. Ouchetto, “MFOOD-70: Moroccan Food Dataset for Food Image Recognition Towards Glycemic Index Estimation,” in The Proceedings of the International Conference on Smart City Applications, Springer, 2023, pp. 372–380.
- [12] S. Murillo et al., “Culinary strategies to manage glycemic response in people with type 2 diabetes: A narrative review,” Front. Nutr., vol. 9, p. 1025993, 2022.
- [13] D. J. Jenkins et al., “Glycemic index of foods: a physiological basis for carbohydrate exchange,” Am. J. Clin. Nutr., vol. 34, no. 3, pp. 362–366, 1981.
- [14] G. Riccardi, G. Clemente, and R. Giacco, “Glycemic index of local foods and diets: the Mediterranean experience,” Nutr. Rev., vol. 61, no. suppl_5, pp. S56–S60, 2003.
- [15] V. Capriles, K. Coelho, A. Guerra-Matias, and J. Arêas, “Effects of processing methods on amaranth starch digestibility and predicted glycemic index,” J. Food Sci., vol. 73, no. 7, pp. H160–H164, 2008.
- [16] T. M. Wolever et al., “Measuring the glycemic index of foods: interlaboratory study,” Am. J. Clin. Nutr., vol. 87, no. 1, pp. 247S–257S, 2008.
- [17] P. Hernandez, V. Rojas, and C. Mata, “Methodology for adding glycemic index values to a Venezuelan food

- composition database," *Meas. Food*, vol. 7, p. 100048, 2022.
- [18] A. Ali, M. I. Waly, M. Al-Mahrazi, J. Al-Maskari, and S. AlWaheibi, "Glycemic Index and Glycemic Load of Selected Omani Rice Dishes," *Food Sci. Eng.*, pp. 1–8, 2023.
- [19] C. Monnard et al., "Nutritional Composition of Infant Cereal Prototypes Can Precisely Predict Their Glycemic Index," *Nutrients*, vol. 14, no. 18, p. 3702, 2022.
- [20] S.-H. Kim, D.-Y. Kim, S.-W. Chun, J. Kim, and J. Woo, "Impartial feature selection using multi-agent reinforcement learning for adverse glycemic event prediction," *Comput. Biol. Med.*, vol. 173, p. 108257, 2024.
- [21] T. Miyakoshi and Y. M. Ito, "Association of Blood Glucose Data with Physiological and Nutritional Data from Dietary Surveys Investigated Using Publicly Available Wearable-type Databases," *JMIR Diabetes*, Oct. 2024, doi: 10.2196/62831.
- [22] M. A. Kader et al., "Dataset on developing low glycemic index rice variety suitable for irrigated ecosystem in Bangladesh," *Data Brief*, vol. 55, p. 110757, 2024.
- [23] I. Pasqualoni, R. Tolve, B. Simonato, and F. Bianchi, "The Impact of Selected Ingredients on the Predicted Glycemic Index and Technological Properties of Bread," *Foods*, vol. 13, no. 16, p. 2488, 2024.
- [24] A. Kumar et al., "Glycemic index of Rice: Role in Diabetics," *NRRI Res. Bull.*, no. 44, 2024.
- [25] S. Gopinath, U. S. S. N. Jathin P, P. V. J, and S. Palaniswamy, "Glycemic Index Based Food Recommendation System Using Deep Learning," in 2024 1st International Conference on Communications and Computer Science (InCCCS), Bangalore, India: IEEE, May 2024, pp. 1–6. doi: 10.1109/InCCCS60947.2024.10593438.
- [26] B. D. Maibam, S. Chakraborty, C. Nickhil, and S. C. Deka, "Effect of Euryale ferox seed shell extract addition on the in vitro starch digestibility and predicted glycemic index of wheat-based bread," *Int. J. Biol. Macromol.*, vol. 226, pp. 1066–1078, Jan. 2023, doi: 10.1016/j.jbiomac.2022.11.223.
- [27] T. V. Ngo, K. Konyanee, and N. Luangsakul, "Insights into recent updates on factors and technologies that modulate the glycemic index of rice and its products," *Foods*, vol. 12, no. 19, p. 3659, 2023.
- [28] M. H. Pipil, M. M. Hasan, A. Setu, S. R. Ray, S. Sharmin, and M. J. H. Bhuiyan, "Evaluation of Glycemic Index of Selected Black Rice Cultivars in Bangladesh: Evaluation of glycemic index of selected local rice cultivars in Bangladesh," *J. Bangladesh Agric. Univ.*, vol. 22, no. 2, pp. 249–256, 2024.
- [29] G. Annuzzi et al., "Impact of nutritional factors in blood glucose prediction in type 1 diabetes through machine learning," *IEEE Access*, vol. 11, pp. 17104–17115, 2023.
- [30] S. Shahakar, P. Chopde, N. Purohit, A. Vishwakarma, A. Nite, and A. K. Sharma, "A Novel Machine Learning and Deep Learning Driven Prediction for Pre-diabetic Patients," in 2023 6th International Conference on Information Systems and Computer Networks (ISCON), IEEE, 2023, pp. 1–6.
- [31] Z.-F. Chen, J. D. Kusuma, and S.-Y. P. K. Shiao, "Validating Healthy Eating Index, Glycemic Index, and Glycemic Load with Modern Diets for E-Health Era," *Nutrients*, vol. 15, no. 5, p. 1263, 2023.
- [32] A. T. Sipe, M. L. Neuhouser, K. L. Breymeyer, and K. M. Utzschneider, "Effect of dietary glycemic index on β -cell function in prediabetes: A randomized controlled feeding study," *Nutrients*, vol. 14, no. 4, p. 887, 2022.
- [33] A. Rytz et al., "Predicting glycemic index and glycemic load from macronutrients to accelerate development of foods and beverages with lower glucose responses," *Nutrients*, vol. 11, no. 5, p. 1172, 2019.
- [34] P. H. S. de Rossi, S. M. Barbalho, M. Oshiiwa, and A. A. J. Sampaio, "Glycemic index of foods: A review," *Pubsaúde*, vol. 4, p. a072, 2020, doi: 10.31533/pubsaude4.a072.
- [35] M. A. Samir, Z. A. Mohamed, M. A. A. Hussein, and A. Atia, "Applying Deep Learning to Track Food Consumption and Human Activity for Non-intrusive Blood Glucose Monitoring," in 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE, 2021, pp. 0319–0324.
- [36] U. Gamage, D. Kumari, R. Senavirathna, and D. Rathish, "In vitro Starch Hydrolysis Indices of selected Sri Lankan Starchy Tubers," *J. Agric. Sci. Sri Lanka*, vol. 16, no. 2, 2021.
- [37] F. S. Atkinson, J. C. Brand-Miller, K. Foster-Powell, A. E. Buyken, and J. Goletzke, "International tables of glycemic index and glycemic load values 2021: a systematic review," *Am. J. Clin. Nutr.*, vol. 114, no. 5, pp. 1625–1632, 2021.
- [38] W. Wongniyomkaset, N. Rungraung, N. Muangpracha, T. Winuprasith, and D. Trachootham, "Baseline Insulin Levels may Predict Response to Low Glycemic Index Complete Nutrition Formula: A Randomized Cross-Over Control Trial," 2021.
- [39] E. Pustozarov et al., "The role of glycemic index and glycemic load in the development of real-time postprandial glycemic response prediction models for patients with gestational diabetes," *Nutrients*, vol. 12, no. 2, p. 302, 2020.
- [40] S. Pirasath, S. Balakumar, and V. Arasaratnam, "Glycemic index of traditional foods in Northern Sri Lanka," 2015.
- [41] L. S. Augustin et al., "Glycemic index, glycemic load and glycemic response: an International Scientific Consensus Summit from the International Carbohydrate Quality Consortium (ICQC)," *Nutr. Metab. Cardiovasc. Dis.*, vol. 25, no. 9, pp. 795–815, 2015.
- [42] B. J. Venn, M. Kataoka, and J. Mann, "The use of different reference foods in determining the glycemic index of starchy and non-starchy test foods," *Nutr. J.*, vol. 13, pp. 1–6, 2014.
- [43] N. Gibson, H. C. Schönfeldt, and B. Pretorius, "Development of a rapid assessment method for the prediction of the glycemic index," *J. Food Compos. Anal.*, vol. 24, no. 4–5, pp. 750–754, 2011.
- [44] R. L. Magaletta, S. N. DiCataldo, D. Liu, H. L. Li, R. P. Borwankar, and M. C. Martini, "In vitro method for predicting glycemic index of foods using simulated digestion and an artificial neural network," *Cereal Chem.*, vol. 87, no. 4, pp. 363–369, 2010.
- [45] B. Mahesh, "Machine learning algorithms-a review," *Int. J. Sci. Res. IJSRInternet*, vol. 9, no. 1, pp. 381–386, 2020.
- [46] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc., 2022.

- [47] S. Gronauer and K. Diepold, "Multi-agent deep reinforcement learning: a survey," *Artif. Intell. Rev.*, vol. 55, no. 2, pp. 895–943, 2022.
- [48] H. Askr, E. Elgeldawi, H. Aboul Ella, Y. A. Elshaier, M. M. Gomaa, and A. E. Hassanien, "Deep learning in drug discovery: an integrative review and future challenges," *Artif. Intell. Rev.*, vol. 56, no. 7, pp. 5975–6037, 2023.
- [49] K. Suzuki, "Overview of deep learning in medical imaging," *Radiol. Phys. Technol.*, vol. 10, no. 3, pp. 257–273, 2017.
- [50] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, 2018.
- [51] Q. Rao and J. Frtunikj, "Deep learning for self-driving cars: Chances and challenges," in *Proceedings of the 1st international workshop on software engineering for AI in autonomous systems*, 2018, pp. 35–38.
- [52] T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazelais, "Predictive analytics in education: a comparison of deep learning frameworks," *Educ. Inf. Technol.*, vol. 25, pp. 1951–1963, 2020.
- [53] L. Solorzano, "Image Processing, Machine Learning and Visualization for Tissue Analysis," PhD Thesis, Acta Universitatis Upsaliensis, 2021.
- [54] S. Vajjala, B. Majumder, A. Gupta, and H. Surana, *Practical natural language processing: a comprehensive guide to building real-world NLP systems*. O'Reilly Media, 2020.
- [55] A. Saranya and R. Subhashini, "A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends," *Decis. Anal. J.*, vol. 7, p. 100230, 2023.
- [56] S. Lundberg, "A unified approach to interpreting model predictions," *ArXiv Prepr. ArXiv170507874*, 2017.
- [57] "ANOVA (analysis of variance)," *Statistics Solutions*. Dec. 2010. [Online]. Available: <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/anova/>
- [58] "Tukey HSD (Honestly Significant Difference)." [Online]. Available: <https://real-statistics.com/one-way-analysis-of-variance-anova/unplanned-comparisons/tukey-hsd/>
- [59] XM for Employee Experience, "T-test theory for surveys: An introduction," *Qualtrics*. Sep. 2020. [Online]. Available: <https://www.qualtrics.com/experience-management/research/t-test-analysis/>
- [60] D. Zeevi et al., "Personalized nutrition by prediction of glycemic responses," *Cell*, vol. 163, no. 5, pp. 1079–1094, 2015.
- [61] M.-H. Lai et al., "Predicted glycemic index and glycemic index of rice varieties grown in Taiwan," *Cereal Chem.*, vol. 93, no. 2, pp. 150–155, 2016.
- [62] S. Sampath, P. Tkachenko, E. Renard, and S. V. Pereverzev, "Glycemic control indices and their aggregation in the prediction of nocturnal hypoglycemia from intermittent blood glucose measurements," *J. Diabetes Sci. Technol.*, vol. 10, no. 6, pp. 1245–1250, 2016.
- [63] J. Li and O. Arandjelovic, "Glycaemic index prediction: a pilot study of data linkage challenges and the application of machine learning," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, 2017, pp. 357–360.
- [64] E. N. Okafor et al., "Cassava flour substitution modulates glycemic responses and glycemic index of wheat breads in apparent healthy volunteers," *J. Diet. Suppl.*, vol. 14, no. 4, pp. 446–452, 2017.
- [65] A. Fardet, S. Lakhssassi, and A. Briffaz, "Beyond nutrient-based food indices: a data mining approach to search for a quantitative holistic index reflecting the degree of food processing and including physicochemical properties," *Food Funct.*, vol. 9, no. 1, pp. 561–572, 2018.
- [66] S. P. K. Shiao, J. Grayson, A. Lie, and C. H. Yu, "Predictors of the healthy eating index and glycemic index in multi-ethnic colorectal cancer families," *Nutrients*, vol. 10, no. 6, p. 674, 2018.
- [67] S. K. Kumari and J. Mathana, "Blood sugar level indication through chewing and swallowing from acoustic MEMS sensor and deep learning algorithm for diabetic management," *J. Med. Syst.*, vol. 43, pp. 1–9, 2019.
- [68] M. Ghazal, M. Alhalabi, M. Chabnari, L. Ghazal, and L. Fraiwan, "Design and implementation of a mobile device for blood glucose level assessment," in *2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, IEEE, 2018, pp. 134–139.
- [69] J. J. R. Barata, R. Munoz, R. D. D. C. Silva, J. J. Rodrigues, and V. H. C. De Albuquerque, "Internet of Things based on electronic and mobile health systems for blood glucose continuous monitoring and management," *IEEE Access*, vol. 7, pp. 175116–175125, 2019.
- [70] S. Buranapin, "A STUDY OF THE GLYCEMIC INDEX OF THE MEDICAL FOOD NEO-MUNE," *J. Southeast Asian Med. Res.*, vol. 2, no. 2, pp. 67–75, 2018.
- [71] W. Seo, Y.-B. Lee, S. Lee, S.-M. Jin, and S.-M. Park, "A machine-learning approach to predict postprandial hypoglycemia," *BMC Med. Inform. Decis. Mak.*, vol. 19, pp. 1–13, 2019.
- [72] H. H. Lee, P. Achananuparp, Y. Liu, E.-P. Lim, and L. R. Varshney, "Estimating glycemic impact of cooking recipes via online crowdsourcing and machine learning," in *Proceedings of the 9th International Conference on Digital Public Health*, 2019, pp. 31–35.
- [73] M. Mayo, L. Chepulis, and R. G. Paul, "Glycemic-aware metrics and oversampling techniques for predicting blood glucose levels using machine learning," *PloS One*, vol. 14, no. 12, p. e0225613, 2019.
- [74] I. K. Mohan et al., "Application of multiple linear regression and machine learning algorithms to elucidate the association of poor glycemic control and hyperhomocysteinemia with microalbuminuria," *Indian J. Biochem. Biophys. IJBB*, vol. 56, no. 2, pp. 150–154, 2019.
- [75] I. Rodríguez-Rodríguez, J.-V. Rodríguez, I. Chatzigiannakis, and M. A. Zamora Izquierdo, "On the possibility of predicting glycaemia 'on the fly' with constrained IoT devices in type 1 diabetes mellitus patients," *Sensors*, vol. 19, no. 20, p. 4538, 2019.
- [76] J. C. Peiró, "Comparing artificial pancreas controlled by hybrid" closed-loop" machine learning (ML) trained algorithm to multi-daily injection (MDI), insulin pump without CGM and" sensor assisted" insulin pump therapies for Diabetes Type 1 (DT1) treatment," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, IEEE, 2020, pp. 1–6.

- [77] S. Rizwana and M. K. Hazarika, "Application of near-infrared spectroscopy for rice characterization using machine learning," *J. Inst. Eng. India Ser. A*, vol. 101, no. 4, pp. 579–587, 2020.
- [78] E. A. Pustozarov et al., "Machine learning approach for postprandial blood glucose prediction in gestational diabetes mellitus," *IEEE Access*, vol. 8, pp. 219308–219321, 2020.
- [79] H. Hesse, R. Frozza, V. Corbellini, C. Reuter, and M. Burgos, "The use of Artificial Neural Network for lipid and glycaemic profiles quantification through infrared spectroscopy," in *Anais do XVI Workshop de Informática Médica, SBC*, 2016, pp. 2613–2616.
- [80] B. Olawoye, S. O. Gbadamosi, I. O. Otemuyiwa, and C. T. Akanbi, "Gluten-free cookies with low glycemic index and glycemic load: optimization of the process variables via response surface methodology and artificial neural network," *Heliyon*, vol. 6, no. 10, 2020.
- [81] N. T. D. Anh et al., "Determining the Glycemic Index of Nutritional Product for Diabetes Mellitus-Np through Measuring Glycemic Responses to Reference Food (Glucose) and Test Food (Nutritional product-Np)," *J. Pharm. Res. Int.*, vol. 33, no. 47B, pp. 298–307, 2021.
- [82] A. Das et al., "A sparse coding approach to automatic diet monitoring with continuous glucose monitors," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 2900–2904.
- [83] G. Lazzarinetti, N. Massarenti, and S. Mantsi, "A Cloud-Hosted Online Learning Approach for Glycemic Index Forecasting," in *AIABI@ AI* IA*, 2021.
- [84] S. Narmadha, S. Gokulan, M. Pavithra, R. Rajmohan, and T. Ananthkumar, "Determination of various deep learning parameters to predict heart disease for diabetes patients," in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, IEEE, 2020, pp. 1–6.
- [85] E. M. Kariuki, "Food recommender system for Diabetes Type 2 patients," PhD Thesis, Strathmore University, 2021.
- [86] L. Martínez-Delgado, M. Munoz-Organero, and P. Queipo-Alvarez, "Using absorption models for insulin and carbohydrates and deep leaning to improve glucose level predictions," *Sensors*, vol. 21, no. 16, p. 5273, 2021.
- [87] B. Bent et al., "Non-invasive wearables for remote monitoring of HbA1c and glucose variability: proof of concept," *BMJ Open Diabetes Res. Care*, vol. 9, no. 1, p. e002027, 2021.
- [88] M. S. Patel et al., "Predicting changes in glycemic control among adults with prediabetes from activity patterns collected by wearable devices," *Npj Digit. Med.*, vol. 4, no. 1, p. 172, 2021.
- [89] A. Massaro, N. Magaletti, G. Cosoli, A. Leogrande, and F. Cannone, "Use of machine learning to predict the glycemic status of patients with diabetes," in *Medical Sciences Forum, MDPI*, 2022, p. 11.
- [90] B. Neamtu, M. O. Negrea, and I. Neagu, "Predicting Glycemic Control in a Small Cohort of Children with Type 1 Diabetes Using Machine Learning Algorithms," *Mathematics*, vol. 11, no. 20, p. 4388, 2023.
- [91] T. M. Abegaz, M. Ahmed, F. Sherbeny, V. Diaby, H. Chi, and A. A. Ali, "Application of machine learning algorithms to predict uncontrolled diabetes using the All of Us research program data," in *Healthcare, MDPI*, 2023, p. 1138.
- [92] S. Romero-Tapiador et al., "AI4FoodDB: a database for personalized e-Health nutrition and lifestyle through wearable devices and artificial intelligence," *Database*, vol. 2023, p. baad049, 2023.
- [93] Y. Liu, Z. Han, X. Liu, J. Wang, C. Wang, and R. Liu, "Estimation method and research of fruit glycemic load index based on the fusion SE module Faster R-CNN," *Comput. Electr. Eng.*, vol. 109, p. 108696, 2023.
- [94] P. G. Jacobs et al., "Artificial intelligence and machine learning for improving glycemic control in diabetes: best practices, pitfalls and opportunities," *IEEE Rev. Biomed. Eng.*, 2023.
- [95] A. A. Crystal, M. Valero, V. Nino, and K. H. Ingram, "Empowering Diabetics: Advancements in Smartphone-Based Food Classification, Volume Measurement, and Nutritional Estimation," *Sensors*, vol. 24, no. 13, p. 4089, 2024.
- [96] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Comput. Sci.*, vol. 132, pp. 1578–1585, 2018.
- [97] M. Ibrahim, A. Beneyto, I. Contreras, and J. Vehi, "An ensemble machine learning approach for the detection of unannounced meals to enhance postprandial glucose control," *Comput. Biol. Med.*, vol. 171, p. 108154, 2024.
- [98] K. A. Della Corte, D. Della Corte, S. Titensor, B. Yang, and S. Liu, "Development of a national database for dietary glycemic index and load for nutritional epidemiologic studies in the United States," *Am. J. Clin. Nutr.*, 2024.
- [99] X. Wang, W. Wang, H. Ren, X. Li, and Y. Wen, "Prediction and analysis of risk factors for diabetic retinopathy based on machine learning and interpretable models," *Heliyon*, vol. 10, no. 9, 2024.
- [100] R. Bouchi et al., "Internet of things-based approach for glycemic control in people with type 2 diabetes: A randomized controlled trial," *J. Diabetes Investig.*, 2024.
- [101] M. Mansour, M. S. Darweesh, and A. Soltan, "Wearable devices for glucose monitoring: A review of state-of-the-art technologies and emerging trends," *Alex. Eng. J.*, vol. 89, pp. 224–243, 2024.
- [102] Y. Zhang, X. Bao, Y. Zhu, Z. Dai, Q. Shen, and Y. Xue, "Advances in Machine Learning Screening of Food Bioactive Compounds," *Trends Food Sci. Technol.*, p. 104578, 2024.
- [103] R. Chlup et al., "Automated computation of glycemic index for foodstuffs using continuous glucose monitoring," *J. Diabetes Sci. Technol.*, vol. 2, no. 1, pp. 67–75, 2008.
- [104] S. Al Hamli, M. Holmes, S. Khokhar, and C. Orfila, "Predicting glyceamic index in cereal and legume-based foods from macronutrient composition data," *Proc. Nutr. Soc.*, vol. 70, no. OCE4, p. E122, 2011.
- [105] A. J. Jara, M. A. Zamora, and A. F. Skarmeta, "An internet of things-based personal device for diabetes therapy management in ambient assisted living (AAL)," *Pers. Ubiquitous Comput.*, vol. 15, pp. 431–440, 2011.
- [106] E. Bas, "A robust optimization approach to diet problem with overall glycemic load as objective function," *Appl. Math. Model.*, vol. 38, no. 19–20, pp. 4926–4940, 2014.
- [107] N. Struble, "Measuring Glycemic Variability and Predicting Blood Glucose Levels Using Machine Learning Regression Models," Master's Thesis, Ohio University, 2013.

- [108] A. Flint et al., "The use of glycaemic index tables to predict glycaemic index of composite breakfast meals," *Br. J. Nutr.*, vol. 91, no. 6, pp. 979–989, 2004.
- [109] T. M. Wolever, D. J. Jenkins, V. Vuksan, R. G. Josse, G. S. Wong, and A. L. Jenkins, "Glycemic index of foods in individual subjects," *Diabetes Care*, vol. 13, no. 2, pp. 126–132, 1990.
- [110] GI Group and A. Barclay, "Glycemic index – glycemic index research and GI news." Nov. 2024. [Online]. Available: <http://www.glycemicindex.com>
- [111] P. Rani, R. Lamba, R. K. Sachdeva, P. Bathla, and A. N. Aledaily, "Diabetes prediction using machine learning classification algorithms," in *2023 International Conference on Smart Computing and Application (ICSCA)*, IEEE, 2023, pp. 1–5.
- [112] G. Annuzzi et al., "Exploring Nutritional Influence on Blood Glucose Forecasting for Type 1 Diabetes Using Explainable AI," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 5, pp. 3123–3133, May 2024, doi: 10.1109/JBHI.2023.3348334.
- [113] D. L. Trout, K. M. Behall, and O. Osilesi, "Prediction of glycemic index for starchy foods," *Am. J. Clin. Nutr.*, vol. 58, no. 6, pp. 873–878, 1993.
- [114] T. Larsen et al., "The Diet, Obesity and Genes (Diogenes) Dietary Study in eight European countries– a comprehensive design for long-term intervention," *Obes. Rev.*, vol. 11, no. 1, pp. 76–91, 2010.
- [115] I. Rodríguez-Rodríguez, J.-V. Rodríguez, I. Chatzigiannakis, and M. A. Zamora Izquierdo, "On the possibility of predicting glycaemia 'on the fly' with constrained IoT devices in type 1 diabetes mellitus patients," *Sensors*, vol. 19, no. 20, p. 4538, 2019.



This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).