Universidade de Aveiro
2024

BRUNO
LOPES DE SOUSA

Desenvolvimento de Software e Avaliação de Modelos de Diagnóstico Nutricional com IA: Combinação de Assistentes Virtuais e Técnicas de Deep Learning

Developing Software and Assessing AI-Enhanced Nutritional Diagnosis Models: Fusion of Virtual Assistant and Deep Learning Techniques

PROPOSTA DE TESE

Universidade de Aveiro
2024

BRUNO
LOPES DE SOUSA

Desenvolvimento de Software e Avaliação de Modelos de Diagnóstico Nutricional com IA: Combinação de Assistentes Virtuais e Técnicas de Deep Learning

Developing Software and Assessing AI-Enhanced Nutritional Diagnosis Models: Fusion of Virtual Assistant and Deep Learning Techniques

PROPOSTA DE TESE

"*An artificial intelligence is only truly beneficial when wielded by an intellect that surpasses its own*"

— Bruno Lopes de Sousa

Universidade de Aveiro
2024

BRUNO
LOPES DE SOUSA

Desenvolvimento de Software e Avaliação de Modelos de Diagnóstico Nutricional com IA: Combinação de Assistentes Virtuais e Técnicas de Deep Learning

Developing Software and Assessing AI-Enhanced Nutritional Diagnosis Models: Fusion of Virtual Assistant and Deep Learning Techniques

Proposta de Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à conclusão da unidade curricular Proposta de Tese, condição necessária para obtenção do grau de Mestre em Engenharia Biomédica , realizada sob a orientação científica do Doutor Professor Fernão Abreu, Professor auxiliar do Departamento de Física da Universidade de Aveiro.

Dedico este trabalho à minha família, que sempre me apoiou em seguir os meus sonhos.

o júri / the jury

presidente / president	Prof. Doutor Luiz Fernando Ribeiro Pereira auxiliar
vogais / examiners committee	Prof. Doutor Fernao Vistulo de Abreu auxiliar
	Prof. Doutor Carlos Alberto da Costa Bastos auxiliar

agradecimentos / acknowledgements

Agradeço à minha família pela paciência que tiveram ao ouvir o computador a correr código de madrugada. Agradeço ao meu orientador pelas grandes sessões de brainstorming. Agradeço imensamente a ajuda da minha namorada Sara Pombo pelo design do meu produto final e de algumas ilustrações deste projeto. /
I'd like to thank my family for their incredible patience while the computer was running code in the early hours of the morning. Thanks to my supervisor for the amazing brainstorming sessions. I really appreciate the help of my girlfriend Sara Pombo in designing my final product and some of the illustrations for this project.

Palavras Chave

Diagnóstico Nutricional, Inteligência Artificial, Modelos de Machine e Deep Learning, Análise em Tempo Real, Recomendações Personalizadas, Reconhecimento Áudio-Textual, Assistente Virtual, ChatGPT-3.5-Turbo, FastAI

Resumo

O desenvolvimento de diagnósticos nutricionais exatos e eficientes é uma necessidade premente dos cuidados de saúde. Atualmente, o processo enfrenta desafios como a avaliação subjetiva, a variabilidade dos dados e a necessidade de uma análise rápida e precisa. Este estudo tem como objetivo abordar estas questões através do desenvolvimento de software baseado em inteligência artificial capaz de fornecer previsões precisas e recomendações personalizadas para o diagnóstico nutricional. O problema específico abordado neste trabalho é a inconsistência e a ineficiência dos actuais métodos de diagnóstico nutricional. Estes métodos baseiam-se frequentemente na análise manual dos dados, que são suscetíveis a erros e a variabilidades. O objetivo deste estudo é desenvolver um sistema de software que integre algoritmos avançados de aprendizagem automática para analisar grandes conjuntos de dados de forma rápida e precisa. Este sistema utilizará algoritmos de reconhecimento de áudio-texto e modelos como o GPT-3.5-Turbo para compreender e prever necessidades nutricionais específicas. Além disso, o projeto desenvolverá e avaliará modelos de aprendizagem profunda utilizando abordagens tradicionais e FastAI, selecionará o modelo com melhor desempenho e integrá-lo-á novamente na assistente virtual. A metodologia inclui várias etapas fundamentais: criação e pré-processamento de bases de dados, formação e avaliação de modelos e integração de modelos numa aplicação de fácil utilização. O processo de desenvolvimento de software está dividido em fases, cada uma delas centrada em tarefas específicas, como a construção de bases de dados, a formação de modelos e os testes de usabilidade e acessibilidade. Serão utilizadas ferramentas como o TensorBoard para analisar as características dos modelos e garantir a exatidão das previsões, fazendo variar certos parâmetros. Os resultados esperados incluem melhorias significativas na exatidão dos diagnósticos nutricionais e na eficiência da análise de dados. O software tem como objetivo fornecer recomendações personalizadas em tempo real, reduzindo a carga dos profissionais de saúde e melhorando os resultados para os doentes. No entanto, as limitações incluem a dependência da qualidade e da exaustividade dos dados disponíveis, potenciais enviesamentos nos dados e os recursos significativos necessários para o desenvolvimento e a manutenção. Garantir a segurança e a privacidade dos dados continua a ser um desafio permanente. O plano de trabalho descreve uma abordagem estruturada do ciclo de vida do desenvolvimento de software, desde o tratamento inicial dos dados até à implantação final da aplicação. Isto inclui fases de teste rigorosas para garantir a usabilidade e acessibilidade, seguidas de um período de teste beta para identificar e resolver quaisquer problemas remanescentes. Após a entrada em funcionamento, são implementados planos de monitorização e apoio contínuos para manter o desempenho da aplicação e a satisfação do utilizador. Em conclusão, este software de diagnóstico nutricional baseado em IA representa um avanço significativo neste domínio, abordando os desafios atuais com soluções inovadoras e eficientes. Ao tirar partido da aprendizagem automática e da inteligência artificial, esta ferramenta promete transformar o diagnóstico nutricional, proporcionando um recurso poderoso tanto para os profissionais de saúde como para os pacientes.

Keywords

Nutritional Diagnosis, Artificial Intelligence, Machine and Deep Learning Models, Real-time Analysis, Personalised Recommendations, Audio-Text Recognition, Virtual Assistant, ChatGPT-3.5-Turbo, FastAI

Abstract

The development of accurate and efficient nutritional diagnostics tools is a pressing healthcare need. Currently, the process faces challenges such as subjective assessment, data variability and the need for rapid and accurate analysis. This study aims to address these issues by developing artificial intelligence-based software capable of providing accurate predictions and personalised recommendations for nutritional diagnostics. The specific problem addressed in this work is the inconsistency and inefficiency of current nutritional diagnostic methods. These methods often rely on manual data analysis, which is prone to error and variability. The aim of this study is to develop a software system that integrates advanced machine learning algorithms to analyse large data sets quickly and accurately. This system will use audio-text recognition algorithms and models such as GPT-3.5-Turbo to understand and predict specific nutritional needs. In addition, the project will develop and evaluate deep learning models using traditional approaches and FastAI, select the best performing model and integrate it back into the virtual assistant. The methodology includes several key milestones: data base creation and pre-processing, model training and evaluation, and model integration into a user-friendly application. The software development process is divided into phases, each focusing on specific tasks such as database construction, model training, usability and accessibility testing. Tools such as TensorBoard will be used to analyse model characteristics and ensure the accuracy of predictions by modifing specific parameters. Expected outcomes include significant improvements in the accuracy of nutritional diagnostics and the efficiency of data analysis. The software aims to provide personalised recommendations in real time, reducing the burden on healthcare professionals and improving patient outcomes. However, limitations include dependence on the quality and completeness of available data, potential biases in the data, and the significant resources required for development and maintenance. Ensuring data security and privacy remains an ongoing challenge. The work plan outlines a structured approach to the software development lifecycle, from initial data handling to final application deployment. This includes rigorous testing phases to ensure usability and accessibility, followed by a beta testing period to identify and resolve any remaining issues. After go-live, ongoing monitoring and support plans are implemented to maintain application performance and user satisfaction. In conclusion, this AI-based nutritional diagnostics software represents a significant advancement in the field, addressing current challenges with innovative and efficient solutions. By leveraging machine learning and artificial intelligence, this tool promises to transform nutritional diagnostics, providing a powerful resource for healthcare professionals and patients alike.

Contents

List of Figures

List of Tables

Glossário

AI	Artificial Inteligence
VAs	Virtual Assistants
ND	Nutritional Diagnosis
ANDS	Assistant Nutritional Diagnosis Systems
NLP	Natual Language Processing
ML	Machine Learning
AIVA	AI Virtual Assistants
ChatGPT	Chat Generative Pre-trained Transformer
OpenAI	Open Artificial Inteligence
NLBVA	Natural Language-Based Virtual Assistant
UI	User Interface
PR	Pattern Recognition
CPS	Complex Problem-Solving
VR	Voice Recognition
IR	Image Recognition
FR	Facial Recognition
AT	Automatic Translators
WAI	Weak Artificial Inteligence
NAI	Narrow Artificial Inteligence
SAI	Strong Artificial Inteligence
GAI	General Artifical Inteligence
LM	Learning Models
NN	Neural Networks
DL	Deep Learning
DTs	Decision Trees
BNs	Bayesian Networks
SVM	Support Vector Machines
TG	Text Generation
LR	Linear Regression
KNN	K-Nearest Neighbours
FE	Feature Engineering
CNN	Convolutional Neural Networks
RNN	Recurrent Neural Networks
GPUs	Graphics Processing Units
SL	Supervised Learning
UL	Unsupervised Learning
RL	Reinforcement Learning
RFS	Random Forests
PCA	Principal Component Analysis
APIs	Application Programming Interfaces
BLUE	Bilingual Evaluation Understudy
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
METEOR	Metric for Evaluation of Translation with Explicit ORdering
CIDEr	Consensus-based Image Description Evaluation
SPICE	Semantic Propositional Image Caption Evaluation
UT	Usability Tests
AT	Accessibility Tests
AIANDS	Artificial Intelligence-Assisted Nutritional Diagnosis Software
UF	User-Friendliness

CHAPTER 1

Introduction

Adequate nutrition is a fundamental aspect of health and well-being, and plays a key role in preventing chronic diseases and promoting quality of life [1]–[3]. However, the field of nutrition faces a number of significant challenges. The traditional approach to nutritional diagnosis is based on the subjective interpretation of health professionals, which can lead to inconsistencies. In addition, inter-individual variability, where different individuals respond in different ways to the same diet plan, introduces an additional layer of complexity [1], [2]. The accurate assessment of an individual's nutritional needs necessitates a comprehensive analysis of a multitude of factors, including genetics, lifestyle, pre-existing health conditions and food preferences. These difficulties can result in incorrect diagnoses and ineffective nutritional interventions.

In this context, the application of Artificial Inteligence (AI) has emerged as a promising tool for nutritionists to assist with the complex task of nutritional diagnosis. AI is capable of processing large amounts of data with remarkable speed and precision, allowing for the identification of subtle patterns that would otherwise go unnoticed [4], [5]. This approach to nutritional diagnosis allows for the analysis of food records, the continuous monitoring of nutritional status, and the suggestion of dietary adjustments based on real-time data. It overcomes subjectivity and provides more consistent diagnoses tailored to the specific needs of each individual [5]–[7]. The integration of AI-assisted diagnostic systems promises to revolutionise the field of nutrition, optimising nutritional care processes and significantly improving health outcomes and overall well-being [5], [7].

Nutrition is the scientific discipline that examines the relationship between food and the human body, with a particular focus on the maintenance of health and well-being. A diet that is both nutritious and balanced provides the body with the essential nutrients that are necessary for optimal functioning [1], [3]. Nutrients include carbohydrates, which are the primary source of energy; proteins, which are essential for maintaining bodily functions, for the production of hormones and enzymes, and for the transport of nutrients in the blood; and fats, which provide energy, store vitamins and minerals, and insulate organs [1], [3].

In addition to providing essential nutrients, a nutritious diet plays a fundamental role in reducing the risk of developing chronic diseases such as diabetes, heart disease and certain types of cancer [1]–[3]. Furthermore, it plays a pivotal role in regulating body weight, preventing obesity and its associated complications. A diet rich in nutrients can improve mood and energy levels, combating fatigue and promoting general well-being [2], [3]. Furthermore, a balanced diet enhances the immune system, enabling the body to respond more effectively to infections and other diseases [2], [3].

The implementation of advanced AI technologies in the field of nutrition also facilitates continuous research and the updating of evidence-based nutritional practices. The capacity of AI systems to learn and evolve enables them to adapt to changing nutritional requirements and new scientific discoveries [4], [5]. This enables them to offer up-to-date and efficacious assistance to healthcare professionals and patients [5], [6].

The convergence of human expertise with the precision and efficiency of AI has the potential to transform nutritional diagnosis, making it more accurate, personalised and effective [5]–[7]. This integrated approach promises not only to enhance individual health outcomes but also to optimise nutritional care processes, thus promoting an enhanced quality of life and overall well-being [4], [5].

1.1 Nutritional Diagnosis

Nutritional diagnosis represents a fundamental process that is designed to identify any potential issues related to an individual's dietary intake. This procedure is of great importance for the maintenance of health and the prevention and treatment of various conditions related to nutrition. A comprehensive evaluation of an individual's nutritional status can ascertain deficiencies or excesses of specific nutrients, the existence of eating disorders, the presence of nutrition-related diseases and other health complications [8], [9].

A qualified nutritionist is tasked with conducting a nutritional diagnosis. This process entails the utilisation of a range of data collection techniques on the individual in question. The preliminary stage of the procedure typically entails an exhaustive interview. During this conversation, the nutritionist attempts to ascertain information regarding the patient's dietary habits, medical history, medication usage, lifestyle, and other factors that may impact their nutritional status. The initial interview is of paramount importance, as it serves as the foundation for the subsequent diagnostic process [8], [9].

Following the interview, the subsequent stage may entail a physical examination. The nutritionist will assess a series of physical indicators of nutritional status, including weight, height, waist circumference and skin condition. These measurements are essential for a preliminary assessment of the individual's general state of health, as they can provide clues to possible nutritional deficiencies [8], [9].

Biochemical tests also play a fundamental role in nutritional diagnosis. In many cases, a nutritionist may prescribe blood tests to assess the levels of nutrients, including vitamins, minerals, cholesterol, and glycaemia. These analyses facilitate a more comprehensive comprehension of the patient's nutritional requirements, enabling the identification of deficiencies or excesses that may not be discernible through alternative assessment techniques [9].

Furthermore, body composition analysis is a commonly employed technique. Techniques such as electrical bioimpedance and anthropometry permit the nutritionist to evaluate the patient's body composition, including the quantity of muscle mass, adipose tissue, and body water. This information is of great importance for the formulation of an appropriate and personalised nutritional plan [9].

Nutritional diagnosis is not merely a process of identifying current issues; it is also a means of preventing future health complications. By continually assessing and adapting dietary recommendations based on changes in the patient's health conditions and lifestyle, the nutritionist can assist in maintaining the individual's health over time [9].

In addition to the techniques previously outlined, nutritional diagnosis may also entail the utilisation of food diaries and food frequency questionnaires. The utilisation of these instruments facilitates a more comprehensive understanding of the patient's dietary patterns over time, thereby enhancing the precision of the nutritional assessment [8].

As the digital age progresses, the incorporation of sophisticated technologies, such as Artificial Inteligence, into the process is becoming increasingly crucial. The potential of AI lies in its capacity to analyse vast quantities of data with remarkable swiftness and precision, thereby offering insights that can enhance the accuracy of diagnosis and the efficacy of nutritional interventions. The next phase in the field of nutrition involves the implementation of systems that promise to revolutionise the way nutritional diagnoses are made and nutritional care is managed [9].

In conclusion, nutritional diagnosis is a complex and multifaceted process that is fundamental to maintaining optimal health and well-being. The process entails the collation and examination of a plethora of data sources, encompassing interviews and physical examinations, biochemical tests and analyses, and body composition assessments. The incorporation of novel technologies, such as AI, is poised to markedly transform the domain of nutrition. This will result in more accurate and personalised diagnoses, as well as improved health outcomes for individuals [9].

1.2 Assisted Nutritional Diagnosis Systems

The integration of Virtual Assistants (VAs) with AI signifies a substantial advancement in the interaction between humans and machines, particularly in the domain of Nutritional Diagnosis (ND). ND is a complex process that aims to identify issues related to an individual's diet. These can encompass nutrient deficiencies or excesses, eating disorders, and nutrition-related diseases. In the past, this diagnosis was conducted by qualified nutritionists through a series of comprehensive interviews, physical examinations, biochemical tests, and body composition analyses [10], [11].

The advent of technology has led to the development of Assistant Nutritional Diagnosis Systems (ANDS) that utilise AI-based VAs, which are transforming this field. These systems utilise sophisticated Natual Language Processing (NLP) and Machine Learning (ML) technologies to interact with users, collect relevant data, and provide personalised recommendations efficiently and accurately [10], [12].

Virtual Assistants represent a specific application of AI, developed to interact with human users and carry out specific tasks, offering support in various areas of daily life. These systems are capable of understanding voice or text commands, analysing large volumes of data, and providing responses that are both relevant and contextualised. Notable examples of AI Virtual Assistants (AIVA) include Chat Generative Pre-trained Transformer (ChatGPT), developed by Open Artificial Inteligence (OpenAI), and other prominent systems such as Google's Gemini, Microsoft's Copilot, Amazon's Alexa, and Apple's Siri [11], [13].

To illustrate, ChatGPT is a Natural Language-Based Virtual Assistant (NLBVA) with the capacity to respond to queries on an extensive range of subjects, including nutrition. Each subsequent iteration of the software, such as GPT-3, GPT-3.5-Turbo, GPT-4, and GPT-4o, represents an advancement in text generation and contextual interaction capabilities. In the context of nutritional diagnosis, these systems can assist nutritionists in collecting detailed information from patients, analysing nutritional data, and providing personalised recommendations based on the specific needs of each individual [10], [12].

In addition to ChatGPT, other virtual assistants, such as Google's Gemini, are able to utilise the power of search and AI to provide context-specific answers and assistance in a range of fields, including nutrition. Copilot is an AI-based assisted coding tool. Nevertheless, the concept can be adapted to assist nutritionists in developing personalised diet plans based on the data collected from patients [11], [13].

Amazon's Alexa and Apple's Siri are other examples of virtual assistants that can be integrated into smart home devices. These assistants provide assistance with everyday tasks, such as reminding users of their meals or recording their food diaries. These assistants are capable of interacting directly with users through voice commands, thereby facilitating the continuous collection of nutritional data and the monitoring of individuals' state of health [10], [12].

The distinction between AI and VAs represents a significant area of investigation. AI encompasses a wide range of intelligent systems, while VAs are specifically designed to interact with humans and provide personalised services. In terms of functionality, AI can be applied in various domains, including medical diagnostics and industrial automation, while VAs are designed to offer direct assistance to the user [11], [13].

In the field of ND, the application of AI through VAs could transform the way in which nutritionists assess and monitor patients' nutritional status. These systems can function independently or in conjunction with healthcare professionals, furnishing information derived from real-time data and continuously adapting to fluctuations in individuals' nutritional requirements. The interactivity capabilities of VAs facilitate more accurate data collection and a faster response to patients' needs, thus increasing the effectiveness of nutritional interventions. A number of examples of VAs for nutritional diagnosis can be found in the literature. The following examples of VAs for nutritional diagnosis are provided for illustrative purposes:

"NutriAI", "Foodvisor", "Lifesum" and "Eat This Much" [12], [13].

NutriAI is a VAs that employs ML to provide personalised dietary recommendations based on the user's health data. It is capable of analysing information such as medical history, food preferences, and health goals in order to create diet plans that are tailored to the individual in question. The system has been developed with the capability to integrate with a range of health devices, including smartwatches and fitness trackers, with the objective of facilitating the continuous monitoring of health data and subsequent adjustment of recommendations as necessary [12], [13].

Foodvisor is an application that employs computer vision to identify foods and automatically calculate their nutritional information. The system is capable of identifying the ingredients of a food item, estimating the portion size, and providing detailed data on the caloric value, macronutrient composition, and other nutrients, all based on a single photograph of the food item. This tool is beneficial for those who wish to regulate their food intake in a precise and practical manner [12], [13].

Lifesum is a digital nutritional assistant that offers personalised diet plans based on users' health goals, such as losing weight, gaining muscle, or maintaining a healthy lifestyle. The application's user-friendly interface allows users to select from a diverse range of diet plans that can be adapted to their preferences and dietary restrictions. Nevertheless, certain advanced features are only accessible via a paid subscription [12], [13].

Eat This Much is an automated meal planner that generates personalised meal plans based on the user's preferences and goals. The application allows users to specify their goals in terms of calories, food preferences, and dietary restrictions, and then generates a balanced meal plan automatically. It is a practical tool for those who seek to streamline their daily meal planning processes [12], [13].

Table 1.1: Comparison of Virtual Assistants for Nutritional Diagnosis

Assistant	**Advantages**	**Disadvantages**	**Data Structure**	**Key Features**
NutriAI	High customisation, integration with health devices	Requires extensive personal data	Relational, Documental	Personalised recommendation, integration with wearables
Foodvisor	Easy to use, provides data quickly	Potential recognition errors	Documental	Computer vision, rapid nutritional calculation
Lifesum	User-friendly interface, wide range of plans	Some features are paid	Relational	Custom diet plans, intuitive interface
Eat This Much	Customisation, ease of use	Less focus on specific needs	Documental	Automatic meal planning, based on preferences

The potential of VAs for ND is considerable, although there are also areas for improvement. NutriAI offers a high degree of personalisation, but it is essential to guarantee the security of personal data. Although Foodvisor is accurate in its computer vision, it could be improved if it incorporated more context into its food recommendations. The User Interface (UI) of Lifesum is straightforward to navigate, yet there is potential for further personalisation without the necessity of paid subscriptions. Despite its automated recommendations, Eat This Much could integrate data on physical activity to provide a more personalised service [13].

The use of VAs in ND systems represents a promising application of AI. Such systems collect and analyse nutritional data, provide personalised recommendations and monitor individuals' health. Such systems have the potential to transform the field of ND and care, enhancing patients' quality of life and general well-being. Nevertheless, it is essential to enhance data security, the precision of nutritional analysis, the personalisation of analysis, the personalisation of recommendations and the integration with other data sources. This will ensure that these systems can evolve further and provide even more effective assistance to users. The subsequent stage of this process entails the broader implementation of AI technologies [13]. This will be explored in greater detail in the following sections, with a particular focus on its potential impact on ND and in other areas of health, namely through the personalisation of a single database, ML algorithms with certain prediction and decision models, and the application of text detection and personalised voice recognition techniques.

1.3 Artificial Intelligence

Artificial Inteligence is a field of computer science that focuses on the development of systems capable of performing tasks that, when carried out by human beings, require intelligence. AI has demonstrated significant potential in a number of areas, becoming a crucial field of study with practical applications in multiple spheres of modern life. Among the most notable tasks that AI can perform are Pattern Recognition (PR), NLP, ML, reasoning, and Complex Problem-Solving (CPS) [14], [15].

In the field of PR, AI has been employed to identify patterns within data, including in Image Recognition (IR) and Voice Recognition (VR). This capability has numerous practical applications, including security systems that use Facial Recognition (FR) and VAs that respond to voice commands. In the field of NLP, AI enables systems to comprehend and manipulate human language, which is a fundamental aspect of the functionality of VAs and Accessibility Tests (AT) [15], [16]. Another key area in which systems are developed to improve their performance in specific tasks is ML. This is based on data. Examples of applications include product recommendation on e-commerce sites and fraud detection in financial transactions [15]. Finally, AI is also employed in reasoning and CPS, with the capacity to make decisions and resolve intricate issues. This includes the ability to play chess and make medical diagnoses [16].

Artificial Inteligence can be categorised in a number of ways, reflecting different levels of capability and complexity. Weak Artificial Inteligence (WAI), or Narrow Artificial Inteligence (NAI), refers to systems that are designed and trained to perform a specific task, such as FR or the operation of VAs such as Siri and Alexa. These systems are highly specialised yet lack the capacity for awareness and deep understanding. Conversely, Strong Artificial Inteligence (SAI), or General Artifical Inteligence (GAI), is a theoretical concept that refers to systems that have the full range of human cognitive abilities. Such systems are capable of undertaking any intellectual task that a human being can perform, with genuine understanding and awareness. Nevertheless, this level of AI has not yet been achieved and remains largely hypothetical. The concept of AI superintelligence describes a level of intelligence that significantly exceeds human intelligence in virtually every aspect, including creativity, CPS, and wisdom. While the concept of superintelligence is undoubtedly intriguing, it remains, at this time, a mere theoretical proposal [16].

In the field of Artificial Inteligence, ML represents a fundamental sub-discipline. The field of ML concerns the development of algorithms that facilitate the learning and improvement of systems through the analysis of data. The development of Learning Models (LM), including Neural Networks (NN) and classifiers, represents a central aspect of this field of study. NN, which are inspired by the structure of the human brain, are utilised in Deep Learning (DL), a subcategory of ML that employs large volumes of data and multiple processing layers to perform complex tasks such as IR and NLPions such as spam detection and medical diagnosis [15].

A further significant concept is Decision Trees (DTs), which are predictive models that facilitate decision-making based on rules derived from the data. These models are particularly useful in situations where decisions must be made based on multiple criteria. Furthermore, techniques such as Bayesian Networks (BNs) and Support Vector Machines (SVM) are instrumental in the construction of robust and effective AI systems [15].

The creation of an advanced VAs, such as ChatGPT-3.5-turbo, necessitates the incorporation of a multitude of AI techniques. Chat-GPT-3.5-turbo, a model based on the GPT-3 architecture, employs advances in NLP to facilitate coherent and relevant interactions with users. The application of DL techniques facilitates comprehension and Text Generation (TG), thereby enabling a more natural and effective interaction. The selection of methodologies and the meticulous implementation of these ML and DL models, when combined, are of paramount importance for the success of VAs that can provide accurate, useful, and contextually appropriate responses to users [14], [16].

1.3.1 Machine Learning Vs Deep Learning

Machine Learning is a subfield of Artificial Inteligence that has the potential to transform the way computers process and learn from data [17]. Unlike explicitly programmed instructions, ML algorithms are able to learn from previous examples and build mathematical models that can be used to make predictions or decisions. One of the key advantages of ML is the diversity of algorithms available. From Linear Regression (LR) to DTs, SVM and K-Nearest Neighbours (KNN), there is a wide range of techniques that can be applied, depending on the problem in question. Another crucial aspect of ML is Feature Engineering (FE). This stage of the process involves generating new features from existing data. In many cases, the quality of the extracted features can have a considerable influence on the model's performance [18]. Consequently, experts are often involved in a comprehensive process of selecting and extracting relevant features in order to ensure the effectiveness of the model [19].

In comparison to Deep Learning, Machine Learning is less complex in terms of algorithms. The interpretation of models is generally simpler and the training process is less computationally demanding. However, ML remains a powerful tool with a wide range of applications, including predictive analysis and PR, fraud detection and medical diagnosis [20]–[22].

DL represents an advanced branch of ML that has gained prominence in recent years due to its capacity to process vast quantities of data and to model intricate patterns in an efficacious manner [23], [24]. In contrast to traditional ML, which is based on specific algorithms, DL utilises artificial NN with algorithms. The term "deep" refers to the use of multi-layered artificial NN to extract features and learn representations of data in a hierarchical way [25].

One of the most distinctive features of DL is the utilisation of deep NN, which can comprise several hidden layers between the input and output layers [26]. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are two of the most prevalent architectural models employed in the field of DL [27]. Each has specific applications in domains such as image processing and NLP [28].

One of the advantages of DL is its capacity to automate FE [29]. In contrast to traditional ML, where experts must manually select the relevant features, deep NN are capable of automatically extracting features from raw data [30]. This markedly diminishes the necessity for human intervention in the modelling process, with the potential to enhance performance in numerous instances [31].

Nevertheless, DL also presents a number of distinctive challenges [32]. The models are typically more intricate and computationally intensive. The training of deep NN is contingent upon the availability of a substantial quantity of data and processing capacity, which is often augmented by the utilisation of Graphics Processing Units (GPUs) to accelerate the process [33], [34].

In terms of applications, DL has been successfully employed in a diverse range of complex tasks, including VR, IR, NLP, strategy games (e.g. AlphaGo), art and music generation, and the operation of autonomous vehicles. Its capacity to process vast quantities of data and to model intricate patterns renders it a formidable instrument in a plethora of domains [35]–[37].

Table 1.2: Differences between Machine Learning (ML) and Deep Learning (DL)

Characteristic	**Machine Learning (ML)**	**Deep Learning (DL)**
Definition	Subfield of AI involving algorithms for learning from data.	Subfield of ML using deep neural networks for learning.
Algorithms	Regression, decision trees, SVM, KNN, etc.	Deep neural networks, CNNs, RNNs, etc.
Feature Engineering	Manual	Automatic
Complexity	Less complex, more interpretable	More complex, less interpretable
Data Requirements	Less data needed	Large volumes of data needed
Computational Requirements	Less intensive	Highly intensive (requires GPUs)
Applications	Predictive analytics, fraud detection, etc.	Speech recognition, image processing, NLP, autonomous vehicles, etc.

1.4 Learning Models

In the field of Machine Learning, a conceptual hierarchy is employed to structure the development and evaluation of models [38]. This hierarchy commences with the learning models, which represent the general approaches employed to train models [39]. Three principal types of learning can be identified: Supervised Learning (SL), Unsupervised Learning (UL) and Reinforcement Learning (RL) [18].

In SL, the model is trained using labelled data. This implies that for each input in the data set, there is a known corresponding output [40]. This method is frequently employed in the context of prediction and classification tasks. In the field of SL, prediction models can be classified into two main categories: regression and classification [41]. Regression models predict continuous values, such as an individual's weight based on their height and age. Classification models predict discrete categories or classes, such as the classification of a patient with diabetes based on their symptoms [42].

Classifiers represent a subset of predictive models that are used exclusively for classification tasks [43]. They ascribe labels to data instances in accordance with their characteristics [44]. Classifiers can be classified into several types, including SVM, DTs, Random Forests (RFS), KNN and NN [42]. SVM are particularly effective for high-dimensional classification, DTs are relatively simple and transparent, RFS are ensembles of DTs that increase accuracy and reduce overfitting, KNN classify based on the proximity of the data in the feature space, and NN are models inspired by the human brain, capable of picking up complex patterns [45].

In the context of UL, the term "unlabelled" is used to describe data that has not been assigned a classification. The objective of this methodology is to discern patterns or structures within the data. Predictive UL models encompass clustering techniques such as k-means and hierarchical clustering, which group data based on similarities [46], [47]. Another prominent example is the use of dimensionality reduction techniques such as Principal Component Analysis (PCA), which facilitate the simplification of data while retaining the majority of its variance [48].

RL is a type of ML in which a model learns to make decisions through interactions with a dynamic environment, receiving rewards or penalties [38], [49]. This type of learning is prevalent in domains such as games and robotics, where the model must learn optimal strategies through experimentation and continuous feedback [50].

Within the field of ML, there is a subcategory known as DL, which focuses on deep NN [23]. Deep Learning employs NN comprising multiple layers between the input and output, enabling the modelling of intricate data and sophisticated abstractions [51]. The principal DL architectures encompass CNN, RNN and transformers. CNN are particularly effective for identifying patterns in images, RNN are advantageous for processing sequential data and time series, such as NLP, and transformers are models based on attention mechanisms that are highly effective in natural language tasks [52].

Building on this theoretical foundation, we can now proceed to analyse in greater detail the methodology and technical components that will be employed in the development of this project. This will entail investigating the specifics of the learning algorithms, network architectures and optimisation techniques that enable the achievement of cutting-edge performance in Artificial Inteligence applications.

1.4.1 Hierarchical Structural Scheme

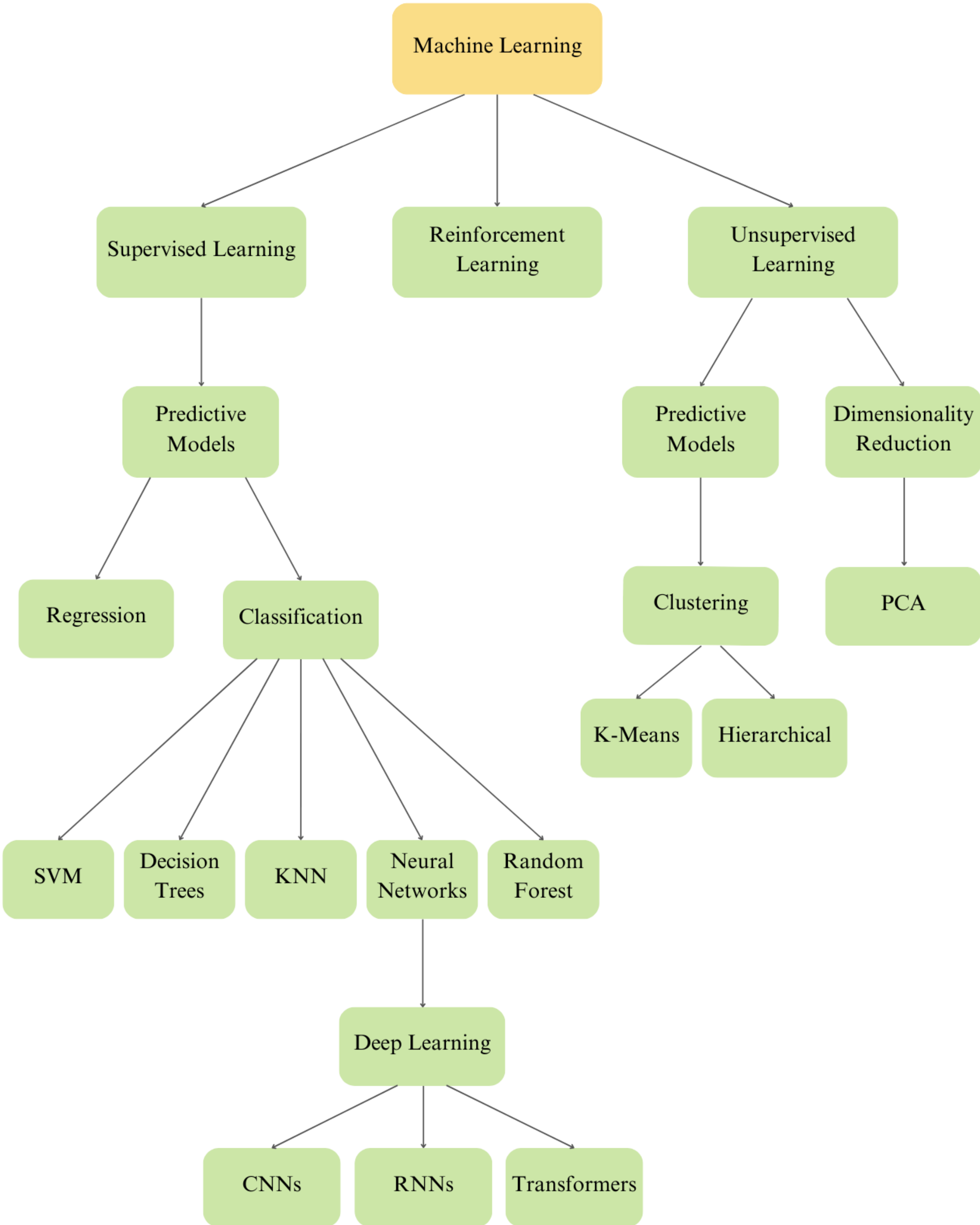

Figure 1.1: Hierarchical Structural Scheme

CHAPTER 2

Methodology

In order to develop software to support Nutritional Diagnosis, it is essential to follow a structured and methodical process, divided into five milestones. The initial phase of the project is divided into two stages. The first stage involves the creation of two relational databases based on the ChatGPT-3.5-turbo model, with the subsequent stage comprising the prediction of either a disease or a nutritional goal. The databases are related to diseases and their associated symptoms, as well as nutritional conditions and goals. The second milestone is to create and train a deep learning model with the inputs and outputs of the first milestone, without recourse to artificial intelligence. The third milestone entails the development of a deep learning model that makes use of the Fastai library. The fourth milestone entails utilising the Tensor Board tool to test the models under diverse conditions and with varying parameters. Subsequently, the models are evaluated using metrics such as Blue to assess their performance, among others. Finally, the fifth milestone corresponds to the implementation of the automatic learning model and the subsequent validation of its usability and accessibility. The software will facilitate the provision of accurate and reliable nutritional and medical diagnoses, thereby becoming an indispensable tool for health and nutrition professionals. Upon completion of the methodology, a projected timeline for the project's implementation is presented, taking into account the preceding stages and the remaining tasks. For the purposes of this visual representation, I have utilised a Gantt chart derived from my work plan.

2.1 MILESTONE 1

The initial stage of the project involves the creation of a relational database and the utilisation of the ChatGPT-3.5-turbo model. For this, it is imperative to design, structure, and implement a robust relational database that will serve as a repository for the data and facilitate its effective management. Subsequently, it is necessary to create a dictionary structure to store the information from the nutrition session. Additionally, it is important to build an algorithm to recognise the textual and auditory input from the client and the nutritionist. Ultimately, it is essential to integrate the ChatGPT-3.5-turbo model into another algorithm, training it to process the data from the database and the data from the nutrition session in real time. Ultimately, the post-trained ChatGPT-3.5-turbo model algorithm generates an output that most closely resembles the disease or goal indicated by the data in the database. Subsequently, an examination of the tools utilised to construct the algorithms that enabled the completion of the initial milestone will be presented.

2.1.1 Data Base

Two databases were initially created, comprising two main fields. The first database, entitled "The 100 Most Common Pathologies", comprised the 100 most common symptoms for each of the diseases included. The second database, entitled "The 100 Most Frequent Nutritional Conditions for Which Someone Goes to a Nutritionist", comprised the 100 most frequent goals related to these conditions. The data was obtained through the utilisation of the ChatGPT-3.5-turbo model. The model was trained to identify the pathology or goal that most closely resembles the data in the databases, based on the client's conversations during the session.

```
Diseases = {
    "Acne": ["Espinhas", "Cravos", "Inflamação da pele"],
    "Anemia": ["Fadiga", "Palidez", "Falta de ar"],
    "Artrite": ["Dor nas articulações", "Rigidez", "Inchaço"],
    "Asma": ["Falta de ar", "Chiado no peito", "Tosse"],
    "Diabetes Mellitus": ["Aumento da sede", "Frequente urinar", "Fadiga"]
}


Physical_Conditions = {
    "Anti-inflammatory diet": ["Redução da inflamação", "Melhoria da digestão", "Redução da dor"],
    "Blood sugar stabilization": ["Controle da glicemia", "Prevenção de picos de açúcar", "Melhoria da energia"],
    "Bone density improvement": ["Fortalecimento ósseo", "Prevenção de fraturas", "Saúde geral dos ossos"],
    "Heart health improvement": ["Redução do colesterol", "Melhoria da circulação", "Redução do risco cardíaco"],
    "Immune system strengthening": ["Aumento da imunidade", "Prevenção de doenças", "Melhoria da recuperação"]
}
```

Figure 2.1: Inicial Data Base Structure

Furthermore, the databases must be expanded by a factor of ten. This will permit the software to implement a more precise filter, correlating the symptoms mentioned by the patients with the most likely pathologies. To enhance the precision of pathology identification, I intend to utilise decision tree techniques. However, I am currently exploring the optimal approach for their integration.

The creation and continuous improvement of these databases is of fundamental importance to the software's effectiveness. The databases not only store crucial information, but also facilitate the efficient operation of the system, thereby ensuring that the nutritional diagnosis is as accurate and useful as possible for nutritionists and their patients.

2.1.2 GPT-3.5-Turbo

The GPT-3.5-turbo model, developed by OpenAI, employs an advanced form of machine learning known as "transformer neural networks," which represents an application of deep learning, a subcategory of machine learning. Transformers represent a deep learning architecture that was first introduced in the 2017 paper "Attention is All You Need" by Vaswani et al

[51]. They are particularly effective in the field of natural language processing (NLP) due to their ability to model long-range dependencies in text sequences.

Transformers are distinguished by their attention mechanism, which allows the model to direct its attention to specific elements of the input sequence during processing. This is of great importance in the context of the complex structure of natural languages. In the context of transformer architectures, self-attention enables each word or token in the input to consider all the other tokens, weighting them according to their relevance [53]. Transformers lack the sequential structure inherent in recurrent neural networks (RNN), which necessitates the use of positional coding to determine the order of words in the input sequence. Moreover, in contrast to RNNs, which process the input sequentially, transformers are capable of processing all the tokens simultaneously, thereby facilitating more expedient and economical training [53].

The code implements the integration of the ChatGPT-3.5-turbo model using the `Agent` class, which is responsible for managing interactions with the OpenAI and Ollama language models. The class is configured to use the API available from both OpenAI and Ollama, depending on the API key provided, and optimises the context of the conversation. The `Agent` class contains methods such as `tell` and `tell_pro`, which facilitate the transmission of questions to the model and the processing of answers in real time and in JSON format, respectively.

```
    ai=Agent(model="gpt-3.5-turbo",
             api_key="My_API_Key",
             system_msg="You are an assistant to help nutritionist diagnose
                 ↪ patients, whether they have pathologies or goals to
                 ↪ achieve. You will record the session history and give
                 ↪ objective and simple solutions.",
             few_shot=custom_few_shot)
```

In addition, the code defines a list of examples (`custom_few_shot`) to train the model on how to provide answers based on the symptoms and goals described by the user. The examples include typical interactions, such as identifying hypertension from the term "high blood pressure" or insomnia from the term "I can't sleep well". This teaches the model to correctly format its responses, indicating whether the response is a pathology or a goal.

The Agent class is configured to utilise the 'gpt-3.5-turbo' model with a specific API key. Furthermore, the system is programmed to function as an assistant to assist nutritionists, recording the history of sessions and providing objective and straightforward solutions.

```
custom_few_shot=[
    {'role': 'user', 'content': 'I have high tension. What could be the
        ↪ pathology closest to the given data?'},
    {'role': 'assistant', 'content': '{"Pathology": Hypertension}'},
    {'role': 'user', 'content': 'I have seizures. What could be the pathology
        ↪  closest to the given data?'},
```

```
    {'role': 'assistant', 'content': '{"Pathology": Epilepsy}'},
    {'role': 'user', 'content': 'I have trouble breathing. What could be the pathology closest to the given data?'},
    {'role': 'assistant', 'content': '{"Pathology": Asthma}'},
    {'role': 'user', 'content': 'I can not sleep properly. What could be the pathology closest to the given data?'},
    {'role': 'assistant', 'content': '{"Pathology": Insomnia}'},
    {'role': 'user', 'content': 'I want to improve my sleeping health. What could be the objective closest to the given data?'},
    {'role': 'assistant', 'content': '{"Objective": Insomnia management}'},
    {'role': 'user', 'content': 'I am obese and I would like to lose some weight. What could be the objective closest to the given data '},
    {'role': 'assistant', 'content': '{"Objective": Weight loss management}'}
]
```

This configuration permits the model to be integrated with algorithms that process data from relational databases and nutrition sessions in real time. This integration enables the conduct of contextual and historical analyses of clients. Following the processing of the data, the model generates precise recommendations based on natural language processing (NLP) and deep learning, thereby ensuring the efficient and accurate management of nutrition sessions. As part of my master's thesis project, which involves the development of software to support nutritional diagnosis, GPT-3.5-turbo plays a pivotal role in the initial milestone.

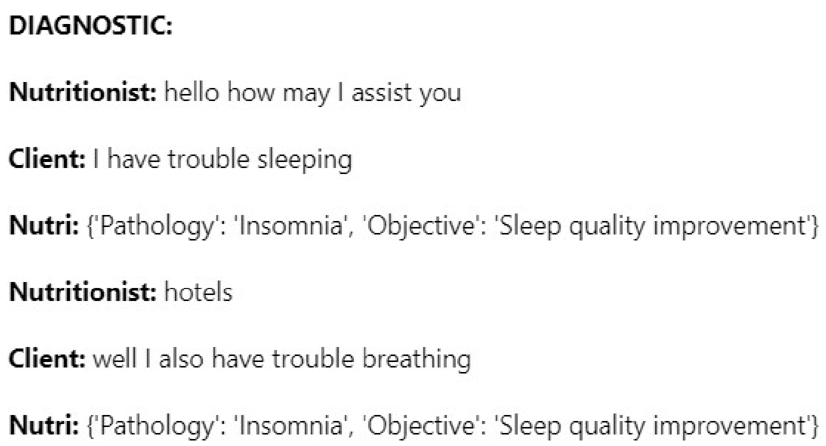

Figure 2.2: Software Output Simulation of the first MILESTONE. "hotels", should be "others", wich pointed out some problems on voice recognition.

2.2 MILESTONE 2

The second milestone is the creation of an automatic learning model based on the data set generated in the initial stage. The objective is to replicate the previous one, while eliminating the necessity for the ChatGPT-3.5-turbo API. This milestone is divided into four stages. The first stage involves the creation of two dataframes. The first contains the diseases obtained from the output of the previous stage, along with the inputs provided by patients. The second contains the nutritional objectives obtained from the previous stage, along with the inputs provided by clients. Additionally, the inputs and outputs of the dataframes must be processed and cleaned to ensure that the data is standardised textually. In the second stage, the inputs and outputs will be tokenised and numericalised, and vector normalisation will be performed on both. Subsequently, the dataframes will be divided into training and test. In the third stage, an algorithm will be constructed that maps the indices of the vectors so that they are associated. In the fourth stage, an embedding matrix will be created and the model will be trained with the numerical vectors.

2.2.1 Dataframes Generating and Pre-Processing

The outputs of the initial milestone were employed to generate two dataframes, comprising the inputs and outputs provided at an earlier stage. Data pre-processing represents a fundamental stage in the data analysis process, as it ensures that the data is in a suitable format for interpretation by the algorithms. This process entails the removal of null values and the standardisation of the data frames. The subsequent stage entails the tokenisation and numbering of the datasets. Subsequently, the vectors are normalised in order to ensure that they are of an identical size and can therefore be processed. Once the vector normalisation has been completed, the data frames will be divided into test and training vectors, with a stratification parameter of 0.2.

2.2.2 Developing and Model Training

In this section, the data will be prepared for the creation of the vocabulary, with the objective of mapping both vectors. This will be achieved through the utilisation of a mapping algorithm, specifically the (`word_to_index`) and (`index_to_word`) algorithms.

Once the pickle vocabulary has been developed, we will proceed to develop an embedding matrix using (`Glove100`) and an autonomous learning model, without reliance on external Application Programming Interfaces (APIs). The model will be developed using libraries such as Scikit-learn, TensorFlow or PyTorch. Firstly, the structure of the model will be defined by selecting an appropriate algorithm for the task, such as a deep neural network or a regression model.

The model will be trained in accordance with the following procedure: the training set will be employed to adjust the model's parameters, while the test set will be used to evaluate the model's performance. The hyperparameters, which are of critical importance to the functioning of the model, will be configured. These include the learning rate, the number of layers and the batch size. The initial training of the model will be conducted, during which the internal parameters will be adjusted in order to minimise the prediction error. Techniques such as the identification of the optimal learning rate will be employed to optimise the training process. Subsequently, the model will be trained for a number of epochs, and its performance will be monitored on a validation set. Should the performance on the validation set fail to improve after a certain number of consecutive epochs, the training will be terminated utilising techniques such as early stopping.

2.3 MILESTONE 3

In this milestone, we will employ the same methodology as in the second milestone, with the exception of the tokenisation and numericalisation of the data frame vectors. Here, we will construct a model that depends on an artificial intelligence library, Fastai. The Fastai library is renowned for its effectiveness and ease of use in machine learning projects. The tokenisation process involves converting textual data into smaller units, such as words or sub-words, which can then be manipulated by the model. Subsequently, numbering transforms these tokens into numbers, thus creating a numerical representation of the textual data. The creation of a vocabulary from the data ensures that the model is able to understand and process the information effectively. The subsequent process is identical to that described in the previous milestone.

2.4 MILESTONE 4

With regard to the fourth stage, a more in-depth analysis of the characteristics of each model will be carried out. In order to accomplish this, it will be necessary to vary the characteristics and parameters of each model in order to assess the different performances in different scenarios. This analysis will be conducted using a tool designed for the analysis of machine learning models, namely `TensorBoard`. Subsequently, the models created will be evaluated using textual evaluation metrics, including BLEU, ROUGE, METEOR, CIDEr and SPICE. Finally, the optimal model will be selected and integrated into the virtual assistant's software.

2.4.1 Models Evaluation

Once the model has been trained, it will be submitted to an evaluation using the test set. This evaluation will facilitate the assessment of the model's accuracy and effectiveness in relation to data that was not previously encountered during the training phase. This will guarantee that the model is not merely memorising the training data, but rather generalising its learning to new situations.

In order to gain a deeper understanding of the characteristics of each model, we will vary the characteristics and parameters of each model in order to assess the different performances in different scenarios. In order to conduct this analysis, we will utilise a tool designed for the analysis of machine learning models, namely `TensorBoard`. The utilisation of `TensorBoard` will facilitate the visualisation of training and evaluation metrics, thereby enabling the identification of patterns and the implementation of necessary adjustments to the models.

Subsequently, the models created will be evaluated using a variety of textual evaluation metrics. The metrics in question compare the model's predictions with real references and include:

BLEU (Bilingual Evaluation Understudy) Bilingual Evaluation Understudy (BLUE) is a metric used to evaluate the quality of textual predictions, such as subtitles, by comparing them with one or more real references. The general formula for calculating BLEU is based on n-gram accuracy and length penalisation [54]:

$$\text{BLEU} = BP \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

where BP is the length penalty, w_n are the weights assigned to each n-gram, and p_n is the precision of the n-grams.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is a set of metrics designed to evaluate the quality of summaries and other automatically generated texts. The most common forms of ROUGE are as follows [55]:

$$\text{ROUGE-N} = \frac{\sum_{\text{sentence}\in\text{references}} \sum_{\text{n-gram}\in\text{sentence}} \text{Count}_{match}(\text{n-gram})}{\sum_{\text{sentence}\in\text{references}} \sum_{\text{n-gram}\in\text{sentence}} \text{Count}(\text{n-gram})}$$

$$\text{ROUGE-L} = \frac{LCS(\text{generated}, \text{reference})}{\text{length of reference}}$$

METEOR (Metric for Evaluation of Translation with Explicit ORdering) Metric for Evaluation of Translation with Explicit ORdering (METEOR) is a metric that measures the degree of precision and retrieval of a given dataset, taking into account the presence of synonyms and word inflections. The formula is expressed as follows [56]:

$$\text{METEOR} = F_{mean} \times (1 - Penalty)$$

where F_{mean} is the harmonised average of precision and recall, and $Penalty$ is a penalty based on fragmentation.

CIDEr (Consensus-based Image Description Evaluation) The Consensus-based Image Description Evaluation (CIDEr) metric evaluates the degree of similarity between the generated description and the reference descriptions, assigning weights to the words based on their contextual importance [57].

$$\text{CIDEr} = \frac{1}{m} \sum_{i=1}^{m} \frac{\sum_{j=1}^{n} \text{IDF}(g_j) \cdot \text{IDF}(r_j)}{\sum_{k=1}^{n} \text{IDF}(r_k)^2}$$

where IDF is the inverse frequency of the document, g_j is the generated n-gram, and r_j is the reference n-gram.

SPICE (Semantic Propositional Image Caption Evaluation) Semantic Propositional Image Caption Evaluation (SPICE) evaluates the semantic alignment of the generated descriptions, taking into account the interrelationships between the referenced objects [58].

$$\text{SPICE} = \frac{\sum_{\text{tuple}\in\text{shared tuples}} \text{Count}(\text{tuple})}{\sum_{\text{tuple}\in\text{reference tuples}} \text{Count}(\text{tuple})}$$

These metrics and visualisations are of significant value in understanding the behaviour of each model and in making well-informed decisions regarding the necessary adjustments and improvements for each machine learning model. The objective is to assess the model's capacity to predict subtitles in comparison to the actual subtitles.

Ultimately, the optimal model will be selected for the purpose of supporting nutritional diagnosis. If the model is able to make successful predictions, it can be concluded that the training and validation process was effective. In the event that the model is unable to generate satisfactory predictions, it will be necessary to undertake a further evaluation and potentially implement adjustments to the pre-processing, tokenisation, modelling or training processes.

Upon completion of these steps, a functional machine learning model will have been developed, capable of making specific predictions. This will enhance our comprehension of sophisticated machine learning methodologies and augment our capacity to devise personalised and efficacious solutions without recourse to external APIs.

2.5 MILESTONE 5

To guarantee the efficacy and inclusivity of the machine learning model, it is imperative to conduct a series of comprehensive and systematic tests. In the final phase of the methodology, the fifth milestone, Usability Tests (UT) are initiated. These are indispensable for corroborating the model's efficacy and ease of use in genuine scenarios. Subsequently, AT are conducted to guarantee that the software is accessible to individuals with diverse abilities, thereby fostering an inclusive experience. The final stage of the methodology is the implementation of the software, which comprises ensuring its seamless technical integration and preparing it for public release. This phase is accompanied by an ongoing monitoring and maintenance plan, with the objective of ensuring the satisfaction and continued optimal performance of the model and application.

2.5.1 Usability Tests

Once the machine learning model has been created and trained, it is essential to conduct Usability Tests to ascertain the model's efficacy and efficiency in actual contexts. The objective of usability testing is to identify potential issues and areas for improvement. Firstly, it is necessary to define the objectives of the tests. This entails determining the aspects of the model that should be assessed, such as accuracy, response time and ease of use. Subsequently, criteria for success must be established for each objective. It is recommended that the selection of participants include a representative group of end users with varying levels of knowledge and experience, thus facilitating a comprehensive view [59].

In the context of UT, realistic scenarios are developed in order to reflect the tasks that users may potentially encounter when utilising the model. The scenarios encompass a variety of tasks, encompassing different aspects of functionality. It is of the utmost importance that the test environment accurately reflects the conditions in which the model will be used in real life. In order to achieve this, it is essential to ensure that all the necessary tools and resources are available. During the testing phase, participants are informed of the objectives of the test and the tasks they are expected to complete. Their actions and reactions are observed and detailed notes are taken [59].

The process of evaluating and analysing the results commences with the collection of feedback from users through the utilisation of interviews or questionnaires. The feedback should concentrate on three principal areas: the ease of use, the comprehension of the functionalities and any difficulties encountered. To evaluate the efficiency of the system, quantitative metrics are employed, including the time taken to complete tasks, the success rate and the number of errors made. Qualitative metrics are used to assess the user experience, including user satisfaction and clarity of instructions. This process allows problems and opportunities for improvement to be identified. The process involves an analysis of user observations and feedback, with the aim of prioritising problems based on their severity and impact on the user experience [59].

Once usability issues have been identified through testing, solutions are proposed and implemented in order to resolve them. This may necessitate the implementation of adjustments to the model, user interface, or documentation, as appropriate. Subsequently, new UT are conducted to ascertain whether the proposed enhancements have indeed resolved the issues that were previously identified. This process of iterative model development continues until the success criteria are met [59].

Finally, the results of the UT are documented in a detailed report, which includes a description of the problems encountered, the solutions implemented and the performance metrics. This report is made available to interested parties in order to ensure transparency and effective communication of the results and improvements. A continuous feedback and improvement process has been established, incorporating UT into future iterations of the model and maintaining an open channel for end users to provide continuous feedback on the model's performance and usability. The implementation of structured and iterative UT ensures that the machine learning model functions technically and offers a positive and productive experience to end users [59].

2.5.2 Accessibility Tests

In addition to UT, it is of the utmost importance to carry out Accessibility Tests to ensure that the model and the software that integrates it are usable by all people, including those with disabilities. The objective of accessibility testing is to identify any potential barriers that may prevent or hinder individuals with diverse needs and abilities from utilising the software. Firstly, the objectives of AT must be defined, and the specific aspects of the software that should be evaluated must be determined. Such considerations may include compatibility with screen readers, keyboard navigability, and the clarity of instructions for users with cognitive disabilities [60].

In order to obtain a comprehensive view of accessibility needs, it is essential to select participants with different types of disabilities for these tests. It is of the utmost importance to develop realistic usage scenarios that reflect the tasks that users with disabilities may encounter when utilising the model. It is of the utmost importance that the scenarios encompass a diverse range of tasks and functionalities in order to guarantee a comprehensive evaluation [60].

During the AT phase, participants are informed of the objectives of the tests and the tasks they are expected to perform. Their actions and reactions are meticulously documented, and any difficulties or barriers encountered are duly noted. It is of the utmost importance that the test environment accurately simulates real conditions of use, including the use of assistive technologies such as screen readers, magnification software and alternative input devices [60].

In order to evaluate the results of accessibility tests, it is essential to collect feedback from users through interviews or questionnaires. The feedback should concentrate on the difficulties experienced by users and any recommendations for enhancements. Quantitative metrics, such as the time taken to complete the tasks and the success rate, are employed in addition to qualitative metrics, including user satisfaction and perceived accessibility [60].

The findings of the study facilitate the identification of issues and potential avenues for enhancement. It is recommended that solutions be implemented to remove the identified accessibility barriers, with adjustments being made to the model, user interface, or documentation as necessary. Further AT are conducted to ascertain whether the implemented improvements have resolved the issues identified. This process is repeated until the desired results are achieved [60].

The results of the AT are documented in a comprehensive report, which includes a detailed description of the problems encountered, the solutions implemented, and the performance indicators. The report is made available to the relevant stakeholders in order to ensure transparency and effective communication of the results and improvements. A continuous feedback and improvement process is established, incorporating AT into future iterations of the model and maintaining an open channel for end users to provide continuous feedback on the model's accessibility [60].

The implementation of AT in a structured and iterative manner ensures that the machine learning model and the software that integrates it are accessible and usable by all users, thereby promoting inclusion and guaranteeing a positive and productive experience for all [60].

2.5.3 App's Final Implementation

Once the UT and AT have been completed, the final stage of the process is to implement the software and launch it to the public. This phase comprises the technical integration of the ML model with the application, ensuring that all dependencies and technical requirements are met. Integration tests are conducted to ascertain the functionality of the model within the application, verifying communication between the various components of the system.

Following the technical integration, the application is subjected to a period of beta testing with a selected group of end users. This enables any residual issues to be identified and rectified prior to the official launch. During the launch phase, a monitoring and support plan is implemented to ensure the optimal performance of the application and to provide users with the necessary assistance. Subsequently, the entire implementation process is documented, and an ongoing maintenance plan is established to ensure that the application remains current and operational, meeting users' needs and expectations.

2.6 WorkPlan

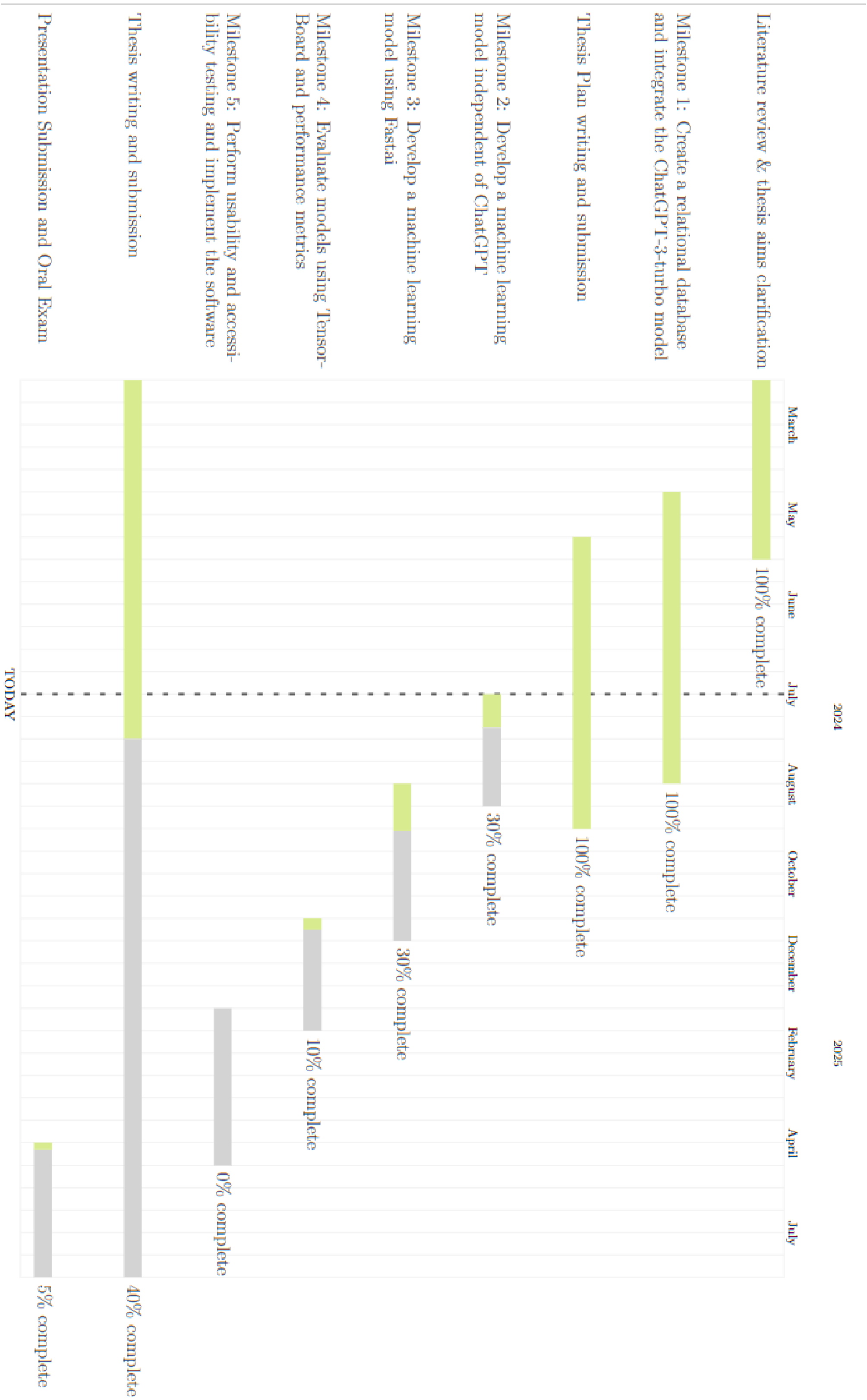

Figure 2.3: WorkPlan

CHAPTER 3

Conclusion

The principal objective of this research is to develop Artificial Intelligence-Assisted Nutritional Diagnosis Software (AIANDS) that is capable of making accurate predictions and offering personalised recommendations. This objective has been identified as a means of improving the accuracy and effectiveness of nutritional diagnoses, a field that is currently facing several challenges, including the subjectivity of assessments, the variability of nutritional data and the need for rapid and accurate analysis.

The current challenges in nutritional diagnosis include the difficulty in analysing large volumes of nutritional data, the variability of nutritional recommendations from different professionals and the lack of automated tools that can offer real-time support. The absence of a standardised, automated system results in inconsistent diagnoses and, consequently, in inappropriate treatments for patients. Audio-text recognition algorithms and models, such as GPT-3.5-Turbo, will be employed to train the software to comprehend and forecast specific nutritional necessities based on a vast database. This process comprises several stages, including the construction and pre-processing of data frames, model creation and training, and an in-depth analysis of the models' characteristics utilising tools such as TensorBoard. Furthermore, the project encompasses the development and assessment of deep learning models employing both traditional and FastAI methodologies, with the objective of identifying the optimal model for integration into the virtual assistant.

The development of this software is comprised of a series of stages, each of which is associated with a specific objective. For instance, the usability testing phase is pivotal for validating the model's efficacy in authentic contexts, identifying shortcomings and potential avenues for enhancement. It is of the utmost importance to conduct accessibility testing at this stage, in order to guarantee that the software is accessible to individuals with diverse abilities, thus promoting an inclusive experience for all users.

Despite the numerous advantages it offers, the software is not without its limitations. One of the principal constraints is the reliance on the quality and comprehensiveness of the available data. The veracity of forecasts and recommendations may be compromised by the presence of incomplete or biased data. Furthermore, the development and maintenance of such a sophisticated system necessitates a significant investment of time and technology. However, the implementation of data security and privacy measures represents a persistent challenge, given the sensitivity of patients' nutritional information.

It is anticipated that the results will demonstrate a significant enhancement in the precision of nutritional diagnoses and a more efficient data analysis process. It is anticipated that the software will be capable of providing personalised nutritional recommendations in real time, thereby reducing the workload of healthcare professionals and improving outcomes for patients. Furthermore, the integration of user feedback throughout the development process and the incorporation of improvements based on Usability Tests and Accessibility Tests ensure that the software remains current and operational.

The work plan delineates a series of phases for the development of the software, commencing with the generation and preliminary processing of the data, continuing with the training and evaluation of the models, and concluding with the final implementation and launch of the software. A comprehensive testing and validation process is conducted for each phase to ensure the system's efficacy and User-Friendliness (UF). The final phase comprises beta testing with selected users and the implementation of an ongoing monitoring and support plan. The aim of this phase is to guarantee the optimum performance of the application after its launch.

In conclusion, the development of this software that utilises artificial intelligence in order to assist in the field of nutritional diagnosis represents a significant advance in this field of study. By addressing current challenges with innovative and effective solutions, the software has the potential to transform the way nutritional diagnoses are carried out, offering a powerful and inclusive tool for both healthcare professionals and patients.

Referências

[1] N. I. of Environmental Health Sciences, «Nutrition, health, and your environment», *NIEHS*, 2023. [Online]. Available: `https://www.niehs.nih.gov/health/topics/nutrition`.

[2] eMediHealth, «How does healthy eating prevent chronic disease», *eMediHealth*, 2023. [Online]. Available: `https://www.emedihealth.com/nutrition/healthy-eating-prevent-chronic-disease`.

[3] U. Health, «When it comes to nutrition and chronic disease, focus on the basics», *UCLA Health*, 2023. [Online]. Available: `https://www.uclahealth.org/news/article/when-it-comes-to-nutrition-and-chronic-disease-focus-on-the-basics`.

[4] T. Miyazawa, Y. Hiratsuka, M. Toda, *et al.*, «Artificial intelligence in food science and nutrition: a narrative review», *Nutrition Reviews*, vol. 80, no. 12, pp. 2288–2300, May 2022, ISSN: 0029-6643. DOI: `10.1093/nutrit/nuac033`. eprint: `https://academic.oup.com/nutritionreviews/article-pdf/80/12/2288/46843610/nuac033.pdf`. [Online]. Available: `https://doi.org/10.1093/nutrit/nuac033`.

[5] H.-C. K. et al., «Applications of artificial intelligence, machine learning, and deep learning in nutrition: A systematic review», *Nutrients*, vol. 16, no. 7, p. 1073, 2024. [Online]. Available: `https://doi.org/10.3390/nu16071073`.

[6] S. Joshi, B. Bisht, V. Kumar, *et al.*, «Artificial intelligence assisted food science and nutrition perspective for smart nutrition research and healthcare», en, *Syst Microbiol and Biomanuf*, vol. 4, no. 1, pp. 86–101, Jan. 2024.

[7] B. N. Limketkai, K. Mauldin, N. Manitius, L. Jalilian, and B. R. Salonen, «The age of artificial intelligence: Use of digital technology in clinical nutrition», en, *Curr. Surg. Rep.*, vol. 9, no. 7, p. 20, Jun. 2021.

[8] M. Holst and A. M. Beck, «Nutritional assessment, diagnosis, and treatment in geriatrics», in *Interdisciplinary Nutritional Management and Care for Older Adults: An Evidence-Based Practical Guide for Nurses*, Ó. G. Geirsdóttir and J. J. Bell, Eds. Cham: Springer International Publishing, 2021, pp. 31–50, ISBN: 978-3-030-63892-4. DOI: `10.1007/978-3-030-63892-4_3`. [Online]. Available: `https://doi.org/10.1007/978-3-030-63892-4_3`.

[9] C. Vanderwall, «Medical nutrition therapy: Managing disease with individualized dietary recommendations and intervention», in *Nutrition Therapy for Urolithiasis*, P. Lowry and K. L. Penniston, Eds. Cham: Springer International Publishing, 2018, pp. 3–15, ISBN: 978-3-319-16414-4. DOI: `10.1007/978-3-319-16414-4_1`. [Online]. Available: `https://doi.org/10.1007/978-3-319-16414-4_1`.

[10] S. Streets, *The rise of virtual health assistants*, Software Streets, 2023. [Online]. Available: `https://softwarestreets.com/blog/virtual-health-assistants/`.

[11] M. Systems, *Ai in healthcare: Transforming patient care and diagnosis*, M3 Systems, 2023. [Online]. Available: `https://m3.systems/the-impact-of-ai-in-healthcare-revolutionizing-patient-care-and-diagnosis/`.

[12] ClinAlly, *Better health outcomes with ai-powered virtual assistants*, ClinAlly, 2023. [Online]. Available: `https://app.clinally.com/better-health-outcomes-with-ai-powered-virtual-assistants/`.

[13] Flatirons, *Artificial intelligence (ai) in healthcare: Revolutionizing patient care*, Flatirons, 2023. [Online]. Available: `https://flatirons.com/blog/ai-in-healthcare/`.

[14] Caltech, *What is artificial intelligence?*, Accessed: 2024-06-25, 2023. [Online]. Available: `https://pg-p.ctme.caltech.edu/blog/ai-ml/what-is-artificial-intelligence`.

[15] M. Sloan, *Machine learning, explained*, Accessed: 2024-06-25, 2023. [Online]. Available: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

[16] U. of Illinois Chicago, *Ai artificial intelligence: What is the definition of ai and how does ai work?*, Accessed: 2024-06-25, 2023. [Online]. Available: https://meng.uic.edu/news-stories/ai-artificial-intelligence-what-is-the-definition-of-ai-and-how-does-ai-work/.

[17] R. Hierons, «Machine learning. tom m. mitchell. published by mcgraw-hill, maidenhead, u.k., international student edition, 1997. isbn: 0-07-115467-1, 414 pages. price: U.k. £22.99, soft cover.», *Software Testing, Verification and Reliability*, vol. 9, no. 3, pp. 191–193, 1999. DOI: https://doi.org/10.1002/(SICI)1099-1689(199909)9:3<191::AID-STVR184>3.0.CO;2-E. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291099-1689%28199909%299%3A3%3C191%3A%3AAID-STVR184%3E3.0.CO%3B2-E. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-1689%28199909%299%3A3%3C191%3A%3AAID-STVR184%3E3.0.CO%3B2-E.

[18] C. M. Bishop, «Pattern recognition and machine learning», *Springer google schola*, vol. 2, pp. 1122–1128, 2006. [Online]. Available: https://cs.haifa.ac.il/~rita/uml_course/lectures/prml8_update_p1.pdf.

[19] P. Domingos, «A few useful things to know about machine learning», *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/2347736.2347755.

[20] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, and M. Data, «Practical machine learning tools and techniques», in *Data mining*, Elsevier Amsterdam, The Netherlands, vol. 2, 2005, pp. 403–413. [Online]. Available: https://sisis.rz.htw-berlin.de/inh2012/12401301.pdf.

[21] F. Provost and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking.* " O'Reilly Media, Inc.", 2013. [Online]. Available: https://books.google.pt/books?hl=pt-PT&lr=&id=EZAtAAAAQBAJ&oi=fnd&pg=PP1&dq=Provost,+F.,+%26+Fawcett,+T.+(2013).+Data+Science+for+Business:+What+You+Need+to+Know+About+Data+Mining+and+Data-Analytic+Thinking.+O%27Reilly+Media.&ots=ymWQRo5SC-&sig=8DcgFX6d9prSDNcLJYJm8C9OsGY&redir_esc=y#v=onepage&q&f=false.

[22] V. L. Patel, E. H. Shortliffe, M. Stefanelli, *et al.*, «The coming of age of artificial intelligence in medicine», *Artificial Intelligence in Medicine*, vol. 46, no. 1, pp. 5–17, 2009, Artificial Intelligence in Medicine AIME' 07, ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2008.07.017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365708000961.

[23] Y. LeCun, Y. Bengio, and G. Hinton, «Deep learning», *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: https://doi.org/10.1038/nature14539.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016. [Online]. Available: https://books.google.pt/books?hl=pt-PT&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=Goodfellow,+I.,+Bengio,+Y.,+%26+Courville,+A.+(2016).+Deep+Learning.+MIT+Press.&ots=MON7dnsAPQ&sig=xbchFMHRdoWInhteTbflxiWdQ5E&redir_esc=y#v=onepage&q&f=false.

[25] J. Schmidhuber, «Deep learning in neural networks: An overview», *Neural Networks*, vol. 61, pp. 85–117, 2015, ISSN: 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2014.09.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608014002135.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, «Imagenet classification with deep convolutional neural networks», *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, ISSN: 0001-0782. DOI: 10.1145/3065386. [Online]. Available: https://doi.org/10.1145/3065386.

[27] S. Hochreiter and J. Schmidhuber, «Long short-term memory», *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, «Gradient-based learning applied to document recognition», *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.

[29] Y. Bengio, A. Courville, and P. Vincent, «Representation learning: A review and new perspectives», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013. DOI: 10.1109/TPAMI.2013.50.

[30] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, «Why does unsupervised pre-training help deep learning?», in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y. W. Teh and M. Titterington, Eds., ser. Proceedings of Machine Learning Research, vol. 9, Chia Laguna Resort, Sardinia, Italy: PMLR, May 2010, pp. 201–208. [Online]. Available: `https://proceedings.mlr.press/v9/erhan10a.html`.

[31] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, «Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.», *Journal of machine learning research*, vol. 11, no. 12, 2010. [Online]. Available: `https://www.jmlr.org/papers/volume11/vincent10a/vincent10a.pdf?ref=https://githubhelp.com`.

[32] J. Dean, G. Corrado, R. Monga, *et al.*, «Large scale distributed deep networks», in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012. [Online]. Available: `https://proceedings.neurips.cc/paper_files/paper/2012/file/6aca97005c68f1206823815f66102863-Paper.pdf`.

[33] J. Pennington, R. Socher, and C. D. Manning, «Glove: Global vectors for word representation», in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: `https://aclanthology.org/D14-1162.pdf`.

[34] N. P. Jouppi, C. Young, N. Patil, *et al.*, «In-datacenter performance analysis of a tensor processing unit», *SIGARCH Comput. Archit. News*, vol. 45, no. 2, pp. 1–12, Jun. 2017. DOI: `10.1145/3140659.3080246`.

[35] G. Hinton, L. Deng, D. Yu, *et al.*, «Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups», *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012. DOI: `10.1109/MSP.2012.2205597`.

[36] D. Silver, A. Huang, C. J. Maddison, *et al.*, «Mastering the game of go with deep neural networks and tree search», *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016. [Online]. Available: `https://doi.org/10.1038/nature16961`.

[37] M. Bojarski, D. Del Testa, D. Dworakowski, *et al.*, *End to end learning for self-driving cars*, 2016. eprint: `1604.07316` (cs.CV). [Online]. Available: `https://doi.org/10.48550/arXiv.1604.07316`.

[38] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018. [Online]. Available: `https://books.google.pt/books?hl=pt-PT&lr=&id=uWV0DwAAQBAJ&oi=fnd&pg=PR7&dq=Sutton,+R.+S.,+%26+Barto,+A.+G.+(2018).+Reinforcement+Learning:+An+Introduction.+MIT+Press.&ots=mjoNq-3Ym1&sig=RtbJ71R9I7l4y9YSgILAhUJIIM0&redir_esc=y#v=onepage&q&f=false`.

[39] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012. [Online]. Available: `https://books.google.pt/books?hl=pt-PT&lr=&id=RC43AgAAQBAJ&oi=fnd&pg=PR7&dq=Murphy,+K.+P.+(2012).+Machine+Learning:+A+Probabilistic+Perspective.+MIT+Press.&ots=ungDcyKs_a&sig=0r4oGc3ICFv_gWsev-RjMrN21H8&redir_esc=y#v=onepage&q&f=false`.

[40] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2. [Online]. Available: `https://doi.org/10.1007/978-0-387-21606-5`.

[41] G. James, D. Witten, T. Hastie, R. Tibshirani, *et al.*, *An introduction to statistical learning*. Springer, 2013, vol. 112. [Online]. Available: `https://doi.org/10.1007/978-1-0716-1418-1`.

[42] L. Breiman, «Random forests», *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [Online]. Available: `https://doi.org/10.1023/A:1010933404324`.

[43] C. Cortes and V. Vapnik, «Support-vector networks», *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995. [Online]. Available: `https://doi.org/10.1007/BF00994018`.

[44] J. R. Quinlan, «Induction of decision trees», *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986. [Online]. Available: `https://doi.org/10.1007/BF00116251`.

[45] T. Cover and P. Hart, «Nearest neighbor pattern classification», *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967. DOI: `10.1109/TIT.1967.1053964`.

[46] S. Lloyd, «Least squares quantization in pcm», *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982. DOI: `10.1109/TIT.1982.1056489`.

[47] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009. [Online]. Available: https://books.google.pt/books?hl=pt-PT&lr=&id=YeFQHiikNo0C&oi=fnd&pg=PR11&dq=Kaufman,+L.,+%26+Rousseeuw,+P.+J.+(2009).+Finding+Groups+in+Data:+An+Introduction+to+Cluster+Analysis.+Wiley.&ots=5CohB2KAAH&sig=A3lVNk3wej7e8Mo_dO5ur0_z4T0&redir_esc=y#v=onepage&q&f=false.

[48] «Principal component analysis for special types of data», in *Principal Component Analysis*. New York, NY: Springer New York, 2002, pp. 338–372, ISBN: 978-0-387-22440-4. DOI: 10.1007/0-387-22440-8_13.

[49] V. Mnih, K. Kavukcuoglu, D. Silver, *et al.*, «Human-level control through deep reinforcement learning», *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015. [Online]. Available: https://doi.org/10.1038/nature14236.

[50] J. Kober, J. A. Bagnell, and J. Peters, «Reinforcement learning in robotics: A survey», *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013. DOI: 10.1177/0278364913495721.

[51] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, «Attention is all you need», in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[52] K. Cho, B. van Merrienboer, C. Gulcehre, *et al.*, *Learning phrase representations using rnn encoder-decoder for statistical machine translation*, 2014. [Online]. Available: https://doi.org/10.48550/arXiv.1406.1078.

[53] D. Rothman, *Transformers for Natural Language Processing: Build, train, and fine-tune deep neural network architectures for NLP with Python, Hugging Face, and OpenAI's GPT-3, ChatGPT, and GPT-4*. Packt Publishing Ltd, 2022. [Online]. Available: https://books.google.pt/books?hl=pt-PT&lr=&id=u9FjEAAAQBAJ&oi=fnd&pg=PP1&dq=ChatGPT+%2B+transformer+neural+network&ots=5DYxxt8dGG&sig=K0siFfjBiQasBQckrGoQ9GvBvr8&redir_esc=y#v=onepage&q&f=false.

[54] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, «Bleu: A method for automatic evaluation of machine translation», in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ser. ACL '02, Philadelphia, Pennsylvania: Association for Computational Linguistics, 2002, pp. 311–318. DOI: 10.3115/1073083.1073135. [Online]. Available: https://doi.org/10.3115/1073083.1073135.

[55] B. Goodrich, V. Rao, P. J. Liu, and M. Saleh, «Assessing the factual accuracy of generated text», in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19, Anchorage, AK, USA: Association for Computing Machinery, 2019, pp. 166–175, ISBN: 9781450362016. DOI: 10.1145/3292500.3330955. [Online]. Available: https://doi.org/10.1145/3292500.3330955.

[56] A. Lavie and A. Agarwal, «Meteor: An automatic metric for mt evaluation with high levels of correlation with human judgments», in *Proceedings of the Second Workshop on Statistical Machine Translation*, ser. StatMT '07, Prague, Czech Republic: Association for Computational Linguistics, 2007, pp. 228–231.

[57] G. O. d. Santos, E. L. Colombini, and S. Avila, «Cider-r: Robust consensus-based image description evaluation», *arXiv preprint arXiv:2109.13701*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2109.13701.

[58] P. Anderson, B. Fernando, M. Johnson, and S. Gould, «Spice: Semantic propositional image caption evaluation», in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 382–398, ISBN: 978-3-319-46454-1. [Online]. Available: https://doi.org/10.1007/978-3-319-46454-1_24.

[59] J. R. Lewis, «Usability testing», *Handbook of human factors and ergonomics*, pp. 1267–1312, 2012. [Online]. Available: https://www.researchgate.net/profile/Cahyono-St/publication/361885118_Handbook_of_human_factors_and_ergonomics_fourth_edition/links/62ca4ecc00d0b4511046af45/Handbook-of-human-factors-and-ergonomics-fourth-edition.pdf#page=1274.

[60] M. M. Eler, J. M. Rojas, Y. Ge, and G. Fraser, «Automated accessibility testing of mobile apps», in *2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST)*, 2018, pp. 116–126. DOI: 10.1109/ICST.2018.00021.