UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO
DEPARTAMENTO DE COMPUTAÇÃO E MATEMÁTICA

BRENO AUGUSTO GUERRA ZANCAN

Numeração dentária e identificação de dentes cariados em radiografias panorâmicas por meio de redes neurais convolucionais

Teeth numbering and identification of decayed teeth in panoramic radiographs through convolutional neural networks

Ribeirão Preto–SP

2023

BRENO AUGUSTO GUERRA ZANCAN

Numeração dentária e identificação de dentes cariados em radiografias panorâmicas por meio de redes neurais convolucionais

Teeth numbering and identification of decayed teeth in panoramic radiographs through convolutional neural networks

Versão Corrigida

Versão original encontra-se na FFCLRP/USP.

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP) da Universidade de São Paulo (USP), como parte das exigências para a obtenção do título de Mestre em Ciências.

Área de Concentração: Computação Aplicada.

Orientador: Profa. Dra. Alessandra Alaniz Macedo

Ribeirão Preto–SP

2023

Breno Augusto Guerra Zancan

Numeração dentária e identificação de dentes cariados em radiografias panorâmicas por meio de redes neurais convolucionais. Ribeirão Preto–SP, 2023.

87p. : il.; 30 cm.

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras
de Ribeirão Preto da USP, como parte das exigências para
a obtenção do título de Mestre em Ciências.
Área: Computação Aplicada.

Orientador: Profa. Dra. Alessandra Alaniz Macedo

1. Diagnóstico Oral. 2. Sistemas de Visão Computacional. 3. Aprendizado profundo. 4. Radiografia Panorâmica. 5. Numeração de dentes. 6. Cárie.

Este trabalho é dedicado aos meus pais,
à minha esposa e à minha avó (in memoriam),
que sempre me apoiaram, me deram forças
e acreditaram em mim.

Agradecimentos

Agradeço a todos que, de alguma forma, participaram desta etapa tão importante na minha vida. Agradeço à Profª Drª Alessandra Alaniz Macedo, pela oportunidade proporcionada e pela orientação durante o curso. Sou grato também pela colaboração, troca de experiências e de aprendizado proporcionado por todos os participantes do grupo de pesquisa InReDD. Agradeço, em especial, ao meu colega de mestrado, parceiro de projeto e amigo Me José Andery Carneiro, que tanto me ajudou, me ensinou e sempre transmitiu otimismo e positividade. Muito obrigado, também, aos meus pais, Maria Aparecida Guerra Zancan e Luiz Zancan, que sempre me apoioram e me incentivaram a estudar e a correr atrás dos meus objetivos. Agradeço também à minha avó, Aparecida Nhoque Guerra (*in memoriam*), por todo carinho, incentivo e por todos os ensinamentos. Obrigado, Letícia Angélica Guerra, minha esposa, que ficou sempre ao meu lado nas minhas escolhas, teve paciência e foi compreensiva mesmo nos momentos mais complicados. Agradeço aos meus irmãos, Ricardo Alexandre Guerra Zancan e Evelin Caroline Guerra Zancan, por serem meus modelos, exemplos e por terem ajudado na minha criação. Muito obrigado também ao meu grande amigo, Guilherme Azambuja Castanho, que esteve presente durante esta jornada, transmitindo entusiasmo e alegria o tempo todo. Agradeço, ainda, ao meu amigo Orlando Mota, por me incentivar, me aconselhar e sempre acreditar em mim. Muito obrigado aos meus colegas de trabalho que sempre estiveram dispostos a me ajudar; em especial, muito obrigado ao meu amigo Alcino Silva, que me ajudou a conciliar o curso de mestrado com o trabalho e sempre me encorajou a seguir em frente.

"Creio firmemente em uma
lei de compensação. As verdadeiras
recompensas são sempre proporcionais
ao esforço e aos sacrifícios feitos."
(Nikola Tesla)

Numeração dentária e identificação de dentes cariados em radiografias panorâmicas por meio de redes neurais convolucionais

Resumo

As radiografias dentárias desempenham um papel crucial na Odontologia, permitindo a detecção precoce de problemas dentários e de uma vasta gama de problemas de saúde dentária e oral, que podem não ser visíveis apenas durante um exame clínico. As radiografias dentárias panorâmicas pertencem à categoria de imagens extraorais, capturando uma visão abrangente de todos os dentes e estruturas anatômicas adjacentes específicas. Apesar do inestimável suporte diagnóstico oferecido pelas radiografias, podem surgir desafios na interpretação precisa das condições bucais dos pacientes. Fatores como falta de experiência, julgamentos subjetivos, estresse e fadiga entre profissionais de Odontologia podem impactar suas avaliações. Neste contexto, as Redes Neurais Convolucionais podem oferecer substancial aprimoramento para análise de exames de pacientes. Essas ferramentas utilizam Aprendizado Profundo, uma subárea de Aprendizado de Máquina, e dispõem de um grande potencial para automatizar aspectos da análise de imagens dentárias, extraindo informações cruciais e facilitando avaliações completas e complexas, ao mesmo tempo que reduzem a subjetividade e economizam uma quantidade significativa de tempo. Levando esses fatores em consideração, este trabalho tem como objetivo empregar Redes Neurais Convolucionais para numeração dentária e identificação de dentes cariados, em radiografias panorâmicas. Essas tarefas contribuem para o desenvolvimento de um sistema modular construído em colaboração com outro aluno de mestrado. Utilizando um conjunto de dados de 18.836 imagens dentárias para a tarefa de numeração, foram alcançados os seguintes resultados: precisão de 0,979, revocação de 0,978, acurácia de 0,998 e F1-score de 0,978. Na tarefa de identificação de dentes cariados, uma combinação de dois conjuntos de dados produziu os seguintes resultados: precisão de 0,963, revocação de 0,914, acurácia de 0,939 e F1-score de 0,937. Esses resultados demonstram um promissor potencial de aplicação de Redes Neurais Convolucionais em tarefas que envolvem numeração de dentes e identificação de dentes cariados, permitindo assim aos dentistas mais tempo para tratamentos clínicos. Além disso, a automatização da análise de radiografias panorâmicas pode auxiliar na geração de relatórios e no preenchimento de registros odontológicos, ao mesmo tempo que serve como avaliação secundária para identificação dentária, reduzindo, em última análise, a probabilidade de erros.

Palavras-chave: Diagnóstico oral. Sistemas de visão computacional. Aprendizado profundo. Radiografia panorâmica. Numeração de dentes. Cárie.

Teeth numbering and identification of decayed teeth in panoramic radiographs through convolutional neural networks

Abstract

Dental radiographs play a crucial role in dentistry, allowing early detection of dental problems and a wide range of dental and oral health problems, that may not be visible during a clinical examination alone. Panoramic dental radiographs belong to the category of extraoral imaging, capturing a comprehensive view of all teeth and specific adjacent anatomical structures. Despite the invaluable diagnostic support offered by radiographs, challenges may arise in the accurate interpretation of patients' oral conditions. Factors such as lack of experience, subjective judgments, stress, and fatigue among dental professionals can impact their assessments. In this context, Convolutional Neural Networks hold substantial promise in enhancing the analysis of patient exams. These tools use Deep Learning, a subarea of Machine Learning, and they possess the potential to automate aspects of dental image analysis, extracting crucial information from images and facilitating thorough, intricate assessments while reducing subjectivity and saving a significant amount of time. Taking these factors into account, this work aims to employ Convolutional Neural Networks for tooth numbering and the identification of decayed teeth in panoramic radiographs. These tasks contribute to the development of a modular system in collaboration with another master's student. Utilizing a dataset of 18,836 tooth images for the numbering task, the following results were achieved: precision of 0.979, recall of 0.978, accuracy of 0.998, and an F1-score of 0.978. In the task of identifying decayed teeth, a combination of two datasets yielded the following results: precision of 0.963, recall of 0.914, accuracy of 0.939, and an F1-score of 0.937. These results illustrate the significant promise of employing Convolutional Neural Networks in tasks involving tooth numbering and identifying decayed teeth, thus allowing dentists more time for clinical treatments. Furthermore, automating the analysis of panoramic radiographs can aid in generating reports and populating dental records while serving as a secondary assessment for tooth identification, ultimately reducing the likelihood of errors.

Keywords: Oral diagnosis. Computer vision systems. Deep learning. Panoramic radiography. Teeth numbering. Tooth decay.

List of Figures

List of Tables

List of abbreviations and acronyms

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
FDI	Federation Dentaire Internationale
FFCLRP	Faculty of Sciences and Letters of Ribeirão Preto
FORP	Faculty of Dentistry of Ribeirão Preto
ML	Machine Learning
PAN	Panoramic Radiography
USP	Universidade de São Paulo

Contents

Introduction

The Ministry of Health, under the National Oral Health Survey - SB Brasil Project (BRASIL, 2012), regularly carries out surveys to evaluate the prevalence of major oral health concerns, such as dental caries, periodontal disease, malocclusions, and their associated impacts. The most recent data, published in 2010, was obtained from a study involving a sample of individuals from 177 cities, covering all 26 capitals and the Federal District. This comprehensive study involved approximately 2,000 SUS (Sistema Único de Saúde) professionals who conducted examinations on a total of 37,519 individuals across various age groups, specifically 5, 12, 15 to 19, 34 to 45, and 65 to 74 years old (BRASIL, 2012). Despite a reduction in the incidence of dental caries observed between 2003 and 2010, it remains the predominant oral health concern affecting the Brazilian population.

According to the World Health Organization (WHO) [1], tooth decay is a significant public health concern that can adversely impact an individual's quality of life and, in advanced stages, lead to pain, discomfort, and chronic systemic infections, often necessitating the removal of affected teeth. The consumption of free sugars in an individual's diet is metabolized by bacteria, resulting in the production of acids that demineralize enamel [2] and dentin [3], contributing to the development of dental caries.

Early diagnosis of dental caries is crucial for preventing its progression and minimizing harm to patients. However, identifying caries in its initial stages is often a challenging task for dental professionals (Ramana Kumari; Nagaraja Rao; Ramana Reddy, 2022; SALUNKE et al., 2022). In such cases, dental radiographs play a pivotal role not only in caries detection but also in diagnosing various other oral conditions like bone loss, fractures, and cysts, which may be prone to errors during purely clinical assessments (BAYDAR et al., 2023; SALUNKE et al., 2022). Therefore, radiographic images serve as indispensable data sources for enhancing diagnostic accuracy (WANG; HUANG;

[1] https://www.who.int/news-room/fact-sheets/detail/sugars-and-dental-caries

[2] Tooth enamel constitutes the outermost layer of a tooth, serving the crucial role of enveloping and safeguarding the underlying dentin. Remarkably, it stands as the most mineralized tissue within the human body.

[3] The layer beneath the enamel in a tooth, known as dentin, exhibits a comparatively less rigid structure. Dentin is a mineralized tissue characterized by a connective, avascular, and acellular nature, constituting the predominant part of the dental structure. Its primary role is to safeguard the initial protective layer by absorbing external impacts resulting from food consumption or wear.

LEE, 2016).

In dentistry, radiographic images aid in assessing the condition of teeth, gums, jaws, fractures, and bone structures (JADER; OLIVEIRA; PITHON, 2018; KUMAR; BHADAURIA; SINGH, 2021). Without access to radiographic images, dentists would often fail to detect many dental issues before they reach advanced stages. Consequently, radiographic images are fundamental for clinical diagnoses, treatment planning, and surgical procedures, enabling the identification of concealed dental structures, benign or malignant masses, bone loss, periodontal issues, impacted teeth, and caries (BAYDAR et al., 2023; WANG; HUANG; LEE, 2016). Oral radiographs can be broadly categorized as intraoral and extraoral, with the three primary types being bitewing, periapical, and panoramic. Bitewing and periapical radiographs fall under the intraoral category, as they are captured within the patient's mouth. In contrast, panoramic radiographs are extraoral, taken outside the patient's mouth, and offer a comprehensive view of all teeth in a single image.

Despite the invaluable diagnostic assistance offered by radiographs, several factors, such as lack of experience, subjective perception, stress, and fatigue among dental professionals, can pose challenges in accurately interpreting patients' oral conditions (PADALIA; VORA; SHARMA, 2022; BAYDAR et al., 2023; SHENG et al., 2022). Furthermore, radiographs are typically grayscale images characterized by high noise levels and limited resolution, further complicating their analysis.

In this context, the incorporation of Deep Learning (DL) techniques, particularly Convolutional Neural Networks (CNNs), holds great promise for enhancing the analysis of patient exams. CNNs have the potential to automate aspects of dental image analysis, extracting information from images, enabling comprehensive and intricate assessments, reducing subjectivity, and significantly saving time. They provide a valuable second opinion for professionals in both Dentistry and Radiology. These tools autonomously detect patterns, creating a supportive platform for technicians, researchers, and specialists.

This approach has proven successful in various areas of healthcare where images serve as a vital information source, such as the automatic identification and classification of pulmonary nodules (HOSNY et al., 2018), mammogram interpretation for cancer screening (SHIMIZU; NAKAYAMA, 2020), detection of liver diseases (ZHOU et al., 2019), and identification of melanomas and malignant carcinomas (SCHMIDT-ERFURTH et al., 2018), among others. Despite the strides made and the effective incorporation of deep learning into mainstream medical practices, its utilization in dentistry is at an early developmental phase. The initial breakthroughs in automated caries detection have primarily been observed in intraoral radiographs (KUNZ et al., 2020). .

This work presents the practical implementation of two convolutional neural network architectures, Inception-v3 and InceptionResNet-v2, for the following objectives:

1. Identifying the number of teeth in panoramic radiographs using FDI World Dental Federation notation (FDI notation) - ISO 3950 notation. To achieve this, two distinct strategies were employed for each architecture to determine the optimal strategy and architecture that yielded the best results.

2. Identification of decayed teeth using CNN and combining two different datasets.

The main goal was the identifying the most efficient strategy for the numbering task and to select the most suitable architecture from two predetermined options for both tasks. These choices played a pivotal role in shaping the development of two modules designed by this investigator, which will be integrated alongside two sequential additional modules created by a fellow researcher—a master's student affiliated with the Postgraduate Program in Applied Computing at FFCLRP-USP. These modules were slated to constitute essential components within a jointly devised system, serving the purpose of modular segmentation, numbering, and classification of decayed teeth. Each module will specialize in a distinct task to collectively fulfill the overarching objective.

Within the outlined context, this study aimed to utilize convolutional neural networks for the analysis of panoramic radiographs of teeth. Specifically, the objective was to employ CNNs for the numbering and classification of decayed teeth to establish a more comprehensive modular system. The concept of a modular system is to enable the development of specialized modules for specific tasks, facilitating the system's expansion to undertake various tasks associated with the analysis of panoramic radiographs of teeth. To contribute to the development of this system, two distinct CNN architectures were tested, Inception-v3 (SZEGEDY et al., 2016) and InceptionResNet-v2 (SZEGEDY et al., 2017). This study will cover the following main aspects: (i) a comparison between two different strategies for numbering teeth in panoramic radiographs, focusing on the use of FDI notation; (ii) an implementation of two CNN architectures, with customized hyperparameters for each of the mentioned strategies; (iii) an in-depth evaluation, comparing the two strategies and the two architectures, with the aim of identifying the combination that provides the most favorable results in terms of accuracy, precision, recall and F1-score; (iv) a comparison between the results of applying two CNN architectures to identify decayed teeth. By pursuing this research, it is intended to make a valuable contribution to the development of a modular system capable of segmenting, numbering and distinguishing decayed and non-decaying teeth effectively, paving the way for better analysis and diagnosis of oral health. Furthermore, this modular system is intended to support dental professionals in increasing the accuracy of patients' dental diagnoses and their treatment planning, minimizing errors in radiographic analyzes resulting from factors such as inexperience, fatigue, stress or subjectivity. In this context, it is also possible to automate tasks that include generating reports and completing dental records, for example, in hospital and forensic contexts.

The structure of this document is as follows: Chapter 1 presents the main categories of radiographs, in addition to the concept of FDI notation and fundamental concepts for the preparation of this work; In Chapter 2, studies are presented that served as references for the development of this work; Chapter 3 provides a description of the materials and methods used in this work; Chapter 4 presents and discusses the results obtained in this study, comparing them with related works; and Chapter 5 concludes by resuming the activities of this work and suggests future work.

1

Theoretical Foundation

As mentioned in the introduction to this study, the application of Deep Learning (DL), a subfield of Artificial Intelligence (AI), is still in its early stages in Dentistry. However, as highlighted by Hung et al. (2019), the number of studies using images from dental and maxillofacial clinics to develop AI models has increased significantly since 2006. Moreover, as indicated by (CARNEIRO et al., 2023b), there was a substantial increase in the volume of research focused on employing DL techniques for the analysis of dental radiographs during the period spanning from 2017 to 2022.

In the process of advancing the tasks involving teeth numbering and the identification of decayed teeth in panoramic radiographs through the application of AI, there was a requirement to draw upon concepts and definitions originating from Dentistry and Radiology, alongside the utilization of Computing technologies within the context of Computer Vision. The application of Artificial Intelligence has demonstrated considerable promise in aiding the analysis of radiographic images of teeth, and as such, it has garnered growing interest in research endeavors geared toward this objective (CARNEIRO et al., 2023b).

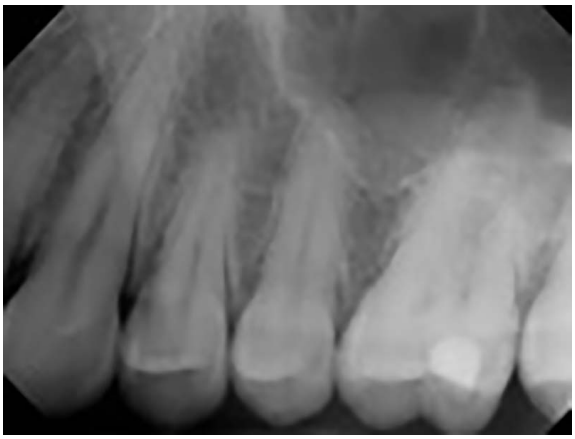
In order to advance this project, the utilization of concepts and definitions from the fields of Dentistry and Radiology, coupled with the application of Computer Vision and DL technologies within the realm of Computing, became imperative for the tasks of teeth numbering and identification of decayed teeth. In the following sections, will present the primary categories of radiographic images of teeth, along with an explanation of FDI notation and concepts related to tooth decay. Additionally, essential insights will be provided on key concepts such as Artificial Intelligence, Machine Learning, Deep Learning, Convolutional Neural Networks, and Computer Vision.

1.1 Teeth Radiographs

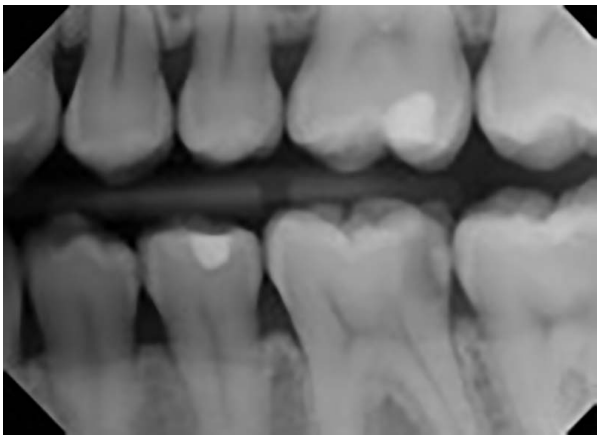
There are three main types of dental radiographs: bitewing, periapical and panoramic (KUMAR; KHAMBETE; PRIYA, 2011; WANG et al., 2016). The bitewing and periapical radiographs are of the intraoral type, which means that they are obtained inside the patient's mouth. Typically, a periapical radiograph shows two to four teeth on individual film and provides detailed information about them, in addition to showing the surrounding alveolar bone. This type of radiography allows dentists to examine conditions below the gum line and provides a clear view of the roots of teeth (MARK, 2020). As for the bitewing radiographs, they show, in an individual film, the crowns of the premolars and molars on one side of the mandible (WHAITES, 2002). They are a valuable tool for dentists to inspect for tooth decay between back teeth or below tooth fillings. Illustrative examples of these intraoral radiographs can be found in Figures 1(a) and 1(b).

Figure 1 – Examples of intraoral radiographs. Figure 1(a) shows a periapical radiograph. Figure 1(b) shows a bitewing radiograph.

(a) Periapical radiograph example.



(b) Bitewing radiograph example.



Source: School of Dentistry, Michigan University.

In the case of panoramic radiographs (PANs), they belong to the extraoral category, meaning they are taken from outside the patient's mouth. These radiographs provide a comprehensive view, encompassing all teeth and certain neighboring anatomical structures, in contrast to intraoral radiographs, which focus on isolated portions of the teeth (SCHWENDICKE et al., 2019). The technique for acquiring panoramic radiography images involves the simultaneous rotation of both the X-ray source and the image receptor around the stationary patient (KAMBUROGLU et al., 2012). An illustrative example of this panoramic radiography can be found in Figure 2.

When compared to panoramic radiographs, intraoral methods have the disadvantage of requiring greater patient cooperation and tolerance (ABDINIAN et al., 2015). In addition, patients with gag reflexes or great discomfort with intraoral methods, such as

Figure 2 – Example of a panoramic radiography.

Source: Faculty of Dentistry of Ribeirão Preto.

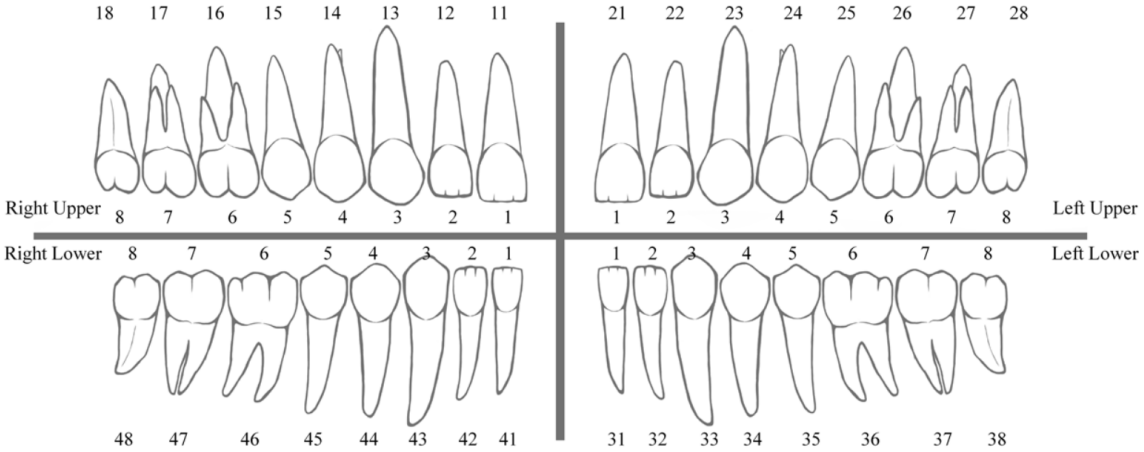
children or people with disabilities, may benefit from extraoral radiographs (CLIFTON; TYNDALL; LUDLOW, 1998). These factors ended up contributing to the increased popularity of extraoral radiographs (AKKAYA et al., 2006).

However, factors such as morphological differences in the mandible between patients and the position of the patient's head inside the X-ray machine can contribute to the resulting image being distorted and blurred (ABDALLA-ASLAN et al., 2020; DU et al., 2018). According to (AYSE, 2018), the great distortion and overlapping of teeth commonly found in panoramic radiographs sometimes make it difficult to segment the contours of the teeth, representing another challenge when working with radiographs of teeth. In addition, (FUKUDA; INAMOTO; SHIBATA, 2019) highlight the complexity of the relationships between anatomical structures in a panoramic image, this complexity can pose challenges, particularly for less experienced observers, potentially leading to the oversight of illnesses and inaccurate diagnoses. In this context, the application of AI methods can help to obtain more accurate analyses and diagnoses, thereby mitigating such issues.

To identify teeth on a radiography, the numbering system defined by the International Dental Federation (FDI - from the French *Fédération Dentaire Internationale*) also known as FDI World Dental Federation notation (FDI notation) or ISO 3950, is usually used. This notation uses two arabic numerals to represent the location of the teeth and divides the mouth into four quadrants of eight teeth (CHIN et al., 2019). Each tooth is represented by a two-digit number, with the left digit indicating the quadrant number, while the right digit referring to the tooth number in the quadrant. In FDI notation,

quadrants 1 (one) and 2 (two) represent the upper arch; quadrants 3 (three) and 4 (four), in turn, represent the lower arch. Additionally, quadrants 1 and 4 contain the teeth on the right side of the mouth, while quadrants 2 and 3 contain the teeth on the left side of the mouth. The tooth number 12, for example, denotes the second tooth in the upper right quadrant. Figure 3 illustrates the FDI numbering notation. Teeth 11, 12, 21 and 22 represent the upper incisors; teeth 13 and 23 represent the upper canines; teeth 14, 15, 24 and 25 represent the upper premolars; teeth 16, 17, 18, 26, 27 and 28 represent the upper molars. Similarly, teeth 31, 32, 41 and 42 represent the lower incisors; teeth 33 and 43 represent the lower canines; teeth 34, 35, 44 and 45 represent the lower premolars; teeth 36, 37, 38, 46, 47 and 48 represent the lower molars. Table 1 displays the arch and tooth type based on the FDI notation.

Figure 3 – FDI teeth notation. The mouth is divided into four quadrants and each tooth is represented by a two-digit number. The left digit indicates the quadrant number, while the right digit refers to the tooth number in the quadrant.



Source: Adapted from (CHIN et al., 2019).

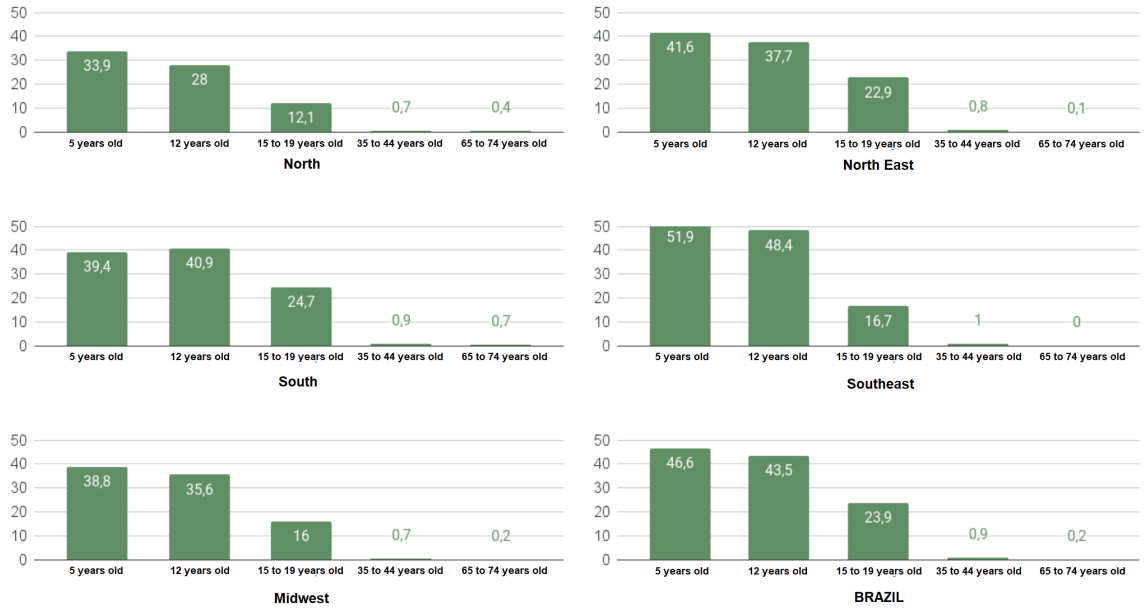
Table 1 – Types of teeth and respective quadrants, according to FDI notation.

FDI number	Tooth and respective quadrant
11, 12, 21, 22	Upper incisor
31, 32, 41, 42	Lower incisor
13, 23	Upper canine
33, 43	Lower canine
14, 15, 24, 25	Upper premolar
34, 35, 44, 45	Lower premolar
16, 17, 18, 26, 27, 28	Upper molar
36, 37, 38, 46, 47, 48	Lower molar

1.2 Tooth Decay

Classified as an infectious disease, tooth decay (cavity or caries) targets the organic material of the tooth, resulting in pain and degradation of this tissue (SINGH; SEHGAL, 2021). It stands as the ailment with the most significant impact on global human populations (FRENCKEN et al., 2017). Risk groups, as identified by the (SELWITZ; ISMAIL; PITTS, 2007) study, encompass children, the elderly, individuals experiencing poverty, those with limited education or socioeconomic resources, and ethnic minority groups, such as immigrants. Figure 4 illustrates the percentage of the population free from cavities in various regions of Brazil and the nation as a whole (BRASIL, 2012). The population was segmented into age groups (specifically 5 years, 12 years, 15-19 years, 35-44 years, and 65-74 years). The measurement utilized the CEO-D (deciduous teeth) and CPO-D (permanent teeth) incidences, encompassing decayed, filled, or missing teeth. Notably, the incidence of such occurrences escalates with age. Regarding the Brazilian average, the research indicates that 99.1% of individuals aged 35-44 exhibited cavities, fillings, or tooth loss.

Figure 4 – Percentage of the population with CEO-D or CPO-D equal to zero by age groups and regions of Brazil.



Source: Adapted from (BRASIL, 2012).

Cavities can emerge due to various factors, including insufficient exposure to fluoride, sugar consumption, or inadequate oral hygiene practices. These events can provoke microbiological alterations within the biofilm encircling the teeth (SELWITZ; ISMAIL; PITTS, 2007). Such changes may lead to the proliferation of bacteria within this biofilm, which ferment sugar and carbohydrate residues lodged between the teeth (LAUDEN-

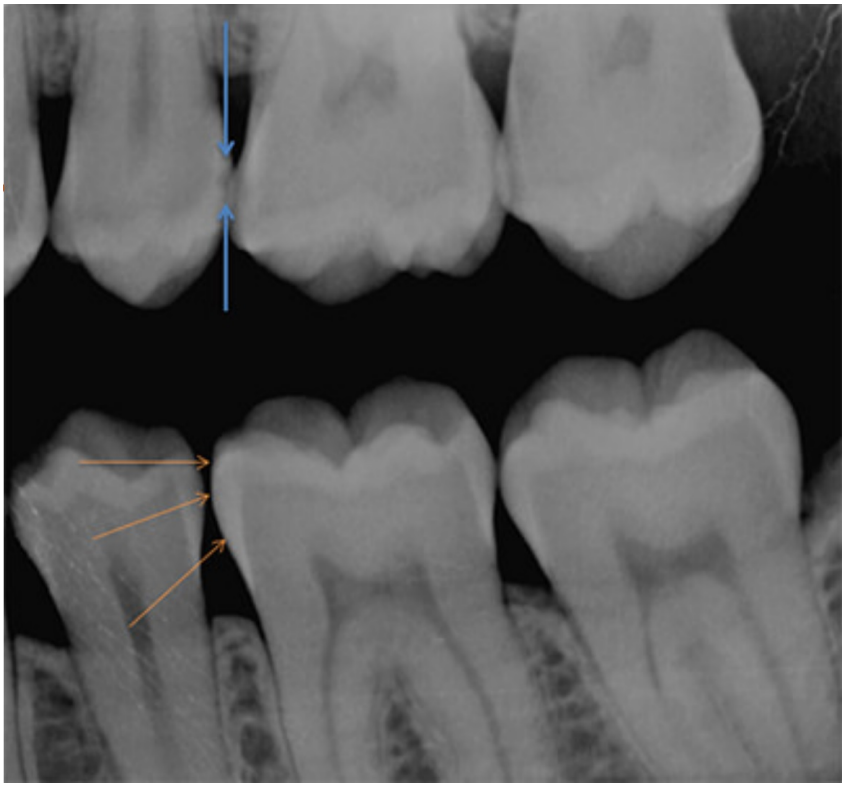
BACH; SIMON, 2014). This process converts these molecules into acids capable of dissolving minerals such as calcium and phosphate, constituents of tooth structures (SINGH; SEHGAL, 2017), resulting in cavity formation. Caries can manifest in the crown or root of the tooth, affecting the enamel (the outer and visible surface of the crown), cementum (the outer surface of the root), or dentin (the tissue situated beneath the previous layers) (SELWITZ; ISMAIL; PITTS, 2007).

Historically, the risk of acquiring cavities was lower in developing countries compared to wealthier nations. Presently, due to more effective policies promoting oral health, caries rates are decreasing in developed countries. Conversely, shifts in diets in developing nations, characterized by increased sugar consumption, are driving the emergence of more cases of the disease (MOREIRA, 2012; PETERSEN, 2003). Despite the tendency toward stability in cavity risk, complete eradication of this disease remains a distant goal (PETERSEN; BAEZ; ORGANIZATION, 2013). To keep it under control, diagnosis and interventions must occur during its initial stages (OSTERLOH; VIRIRI, 2016).

Caries diagnosis can be executed using various methods. The most common approach involves clinical, visual, and tactile examination, aided by tools like a light source, probes, and mirrors. This examination, conducted during a dental appointment, primarily focuses on detecting caries on the visible tooth surfaces and may not identify cases on proximal surfaces (between two teeth) or internal tooth structures (OLSEN et al., 2009). Visual inspection alone can lead to incorrect diagnoses, resulting in many cavities going unnoticed (ISMAIL, 2004). Alternative evaluation techniques include fiber optic transillumination, dyes, fluorescent methods, and electrical conductance meters. However, since the advent of radiographs examinations, this has become the standard method for detecting cavities, particularly interproximal cavities (BOOSHEHRY et al., 2010; AFFAIRS, 2006). Figure 5 shows a cavity located between two teeth in the upper arch, in contrast to a healthy tooth located in the lower arch. Note that the cavity appears as a dark spot.

Treatment for caries depends on its stage of advancement. When caries is identified at an early stage, characterized by mild enamel demineralization, fluoride in toothpaste may suffice to impede its progression and facilitate tooth remineralization. In cases where decay has already impacted the enamel, crown or dental restoration procedures may be necessary. This involves the removal of the affected tissue and surrounding areas, followed by the filling of the void with a restorative substance or the placement of a crown to restore the tooth's form and function. In more advanced cases, where the infection has reached the dentin or pulp (the inner portions of the tooth), root canal treatment (filling the pulp and root canal with sealing material) or tooth extraction may be recommended in instances where infection from tooth decay has spread extensively (BOUCHAHMA et al., 2019).

Figure 5 – Demonstration of tooth decay evident in an intraoral X-ray. Notably, a cavity is discernible between two teeth in the upper arch, indicated by two arrows. Meanwhile, in the lower arch, three arrows pinpoint a sound and healthy tooth.



Source: Mint Hill Dentistry.

1.3 Artificial Intelligence, Machine Learning and Deep Learning

Artificial Intelligence (AI), a field within Computer Science, aims to automate cognitive tasks that are typically performed by humans (CHOLLET, 2017). As suggested by (RUSSELL; NORVIG, 2010), humanity has engaged in a millennia-long quest to comprehend the workings of intelligent entities. However, AI goes beyond mere understanding; it aspires to create and replicate these entities. In essence, Artificial Intelligence endeavors to imbue machines with cognitive capabilities, drawing inspiration from human intelligence in realms such as reasoning, language processing, and visual perception (RUNGTA, 2019).

In 1943, Warren McCulloch and Walter Pitts introduced a model of artificial neurons, where each neuron could exist in one of two states: ”on” or ”off” (RUSSELL; NORVIG, 2010). This seminal work, acknowledged as one of the pioneering contributions to the field of AI, postulates that a neuron can transition from the ”off” to the ”on” state in response to stimuli originating from a specific number of neighboring neurons. In the nascent era of AI research, there was a shared belief that human-level intelligence could

be attained through explicit rules capable of predicting and simulating all human actions. This intellectual lineage laid the foundation for a branch of AI known as Symbolic AI (CHOLLET, 2017).

Symbolic AI is suitable for solving well-defined logical problems, where the rules are clear, and it is possible to apply these rules to model computational systems, as occurs in the creation of expert systems. However, the limitations of symbolic AI become more evident when trying to discover explicit rules to solve more complex and imprecise problems, such as image classification, speech recognition and language translation. For these tasks, another approach has emerged, known as Machine Learning.

Machine learning (ML) is a subset of Artificial Intelligence that allows computers to learn without being explicitly programmed. ML algorithms are trained on data, and they can then use that data to make predictions or decisions. A way to define ML is "a computer program that learns from experience E with respect to some class of tasks T and performance measure P, if its performance in tasks at T, as measured by P, improves with experience E" (MITCHELL, 1997). This definition highlights the three key components of ML:

- Experience E: the data that the ML algorithm is trained on.
- Task T: the problem that the ML algorithm is trying to solve.
- Performance measure P: the way in which the performance of the ML algorithm is measured.

ML is a powerful tool that can be used to solve a wide variety of problems, such as image recognition, natural language processing, and fraud detection.

A central point of Machine Learning is that a computer analyzes a set of training examples, which are the input data, seeking to understand and identify patterns from them, to learn and automate the execution of a task (LOURIDAS; EBERT, 2016; RUNGTA, 2019). It is customary to represent an example, in ML, as a vector x, in attribute-value format, where each entry x_{i} belonging to the vector is a feature. In an image, for example, the values of features are usually the value of each pixel that makes up that image. In general, in ML, a set of examples provides information to an inducer that seeks, from logical inferences, to obtain conclusions and generate a classifier that allows labeling new examples. In this way, a learning problem can be seen as a hypothesis selection problem. Some common ML algorithms include linear regression, logistic regression, and decision trees.

There are three main types of machine learning: supervised learning, semi-supervised learning, and unsupervised learning. In supervised learning, training examples are labeled.

Furthermore, when labels take on discrete values, it is a classification problem; on the other hand, when the labels assume continuous values, it is a regression problem. Regarding unsupervised learning, this approach revolves around unlabeled data and encompasses clustering algorithms. These algorithms operate on datasets spanning various dimensions, aiming to partition the data into groups based on specific criteria (LOURIDAS; EBERT, 2016). Finally, in semi-supervised learning, there is a combination of the two types of learning already presented, supervised and unsupervised. This means that there is a combination of labeled and unlabeled examples. According to (CAMPESATO, 2020), a technique to work with semi-supervised learning involves using labeled data to label those that do not have labels, and then apply a classification algorithm.

Deep Learning (DL) is a subset of Machine Learning that allows computer systems to be improved with experience and data (HEATON, 2017). According to (WAN, 2019), DL is based on Deep Neural Networks (DNNs). DNNs are modeled with Multilayer Perceptrons (MLP) and trained with algorithms to learn representations of data sets, allowing the automation of feature extraction. Basically, for the implementation of a neural network, first it is necessary to collect a set of data, which will be divided into two subsets: one for training and another for testing; after that, the network is trained, with the subset separated for training, generating a prediction model; then, the other subset is used to perform tests on the generated model; finally, the trained network is used to predict information (SHRESTHA; MAHMOOD, 2019).

According to (HOSNY et al., 2018), the basic methods of Deep Learning have been around for decades, but it was only in recent years that the computational power necessary for its growth became available, allowing DL to improve. This allowed the automatic image feature extraction process to use information from the original pixel as input, making it possible to obtain features that are difficult or even impossible for humans to recognize, and consequently significantly reducing the workload of experts (Chen et al., 2019). Deep Learning algorithms have multiple levels of representation, which are derived from a set of non-linear but interconnected layers (LECUN; BENGIO; HINTON, 2015).

A convolutional neural network (CNN) is a class of deep neural network, of the *feedforward* type (in which information flows from input to output in a single direction, without loops), which has been successfully applied in the processing and analysis of digital images. Feedforward CNNs are trained using the backpropagation algorithm. Backpropagation is an optimization algorithm that determines the direction and magnitude of necessary changes in neural network weights in order to minimize a loss function. Like DNNs, a CNN uses a variety of multilayer perceptrons designed to require as little pre-processing as possible (JAE-HONG et al., 2018). A CNN can calculate millions of parameters and learn image representations effectively, being able to identify, recognize and classify objects, as well as detect and segment objects in images (TAYE, 2023). The victory of the

AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) architecture in the Imagenet Large-scale Visual Recognition Challenge 2012[1] helped to popularize the use of CNNs for different fields of computer vision, natural language processing and others (GHOSH et al., 2020).

Convolutional neural networks are composed of an input layer, hidden layers, and an output layer. The input layer of a CNN model is organized in three dimensions: height, width, and depth (or channel). In a color image, which uses the RGB (Red, Green, Blue) color model, there are three channels, each representing a primary color. In the case of a grayscale image, there is only one channel, each pixel representing a certain light intensity (ALZUBAIDI et al., 2021; GONZALEZ; WOODS, 2008). A typical CNN architecture consists of several layers (or building blocks), which stand out: convolution, pooling, activation function, and fully connected. These layers will be briefly explained below:

- Convolutional layer, is a fundamental element in the architecture of a CNN, playing a pivotal role in its overall structure. This layer consists of a collection of convolutional kernels, each characterized by its width, height, and weight. These kernels serve as key tools for feature extraction from the input data. Initially, the weights of these kernels are assigned randomly, but they undergo gradual updates through the process of backpropagation during training. In practice, a kernel is convolved with an input, which can be an image, resulting in a dot product operation that produces an output feature map. This process is crucial for feature learning because the kernel's weights adapt and evolve over the course of training to extract meaningful features from the input data (LECUN et al., 1998).

- Pooling layer: responsible for downsampling (subsampling) the feature maps, akin to the downsampling process in image processing. Their fundamental purpose is to compress the information within feature maps while retaining the most crucial features at each pooling step. This reduction in dimensionality aids in improving computational efficiency and reducing overfitting (SCHERER; MüLLER; BEHNKE, 2010). There are a variety of pooling techniques available, each with its own unique characteristics and applications. These techniques include tree pooling, gated pooling, average pooling, min pooling, max pooling, global average pooling and global max pooling. Among these techniques, max pooling is the most popular and widely adopted technique in various deep learning applications. This is because max pooling is effective in capturing dominant features in the input data (BOUREAU; PONCE; LECUN, 2010).

[1] https://www.image-net.org/challenges/LSVRC/2012

- Activation function is a pivotal component in neural networks, tasked with mapping an input to an output. In this process, the input value is derived by computing the weighted sum of the neuron's inputs and incorporating a bias term when applicable (GOODFELLOW; BENGIO; COURVILLE, 2016). Essentially, the activation function plays a critical role in determining whether a neuron should activate, based on a particular input, thus generating the corresponding output. Activation functions are used to introduce non-linearity into the neural network, which allows it to learn more complex relationships between the input and output data. Some of the most commonly used activation functions include: Sigmoid, Tanh, ReLU (Rectifier Linear Unit), Leaky ReLu, Noisy ReLu e Parametric Linear Units.

- Fully connected layer (FCL): is usually the last layer of a Convolutional Neural Network. In this layer, each node is connected to each node in the previous layer. A vector resulting from a flattened feature map serves as input to this layer, which is positioned after a pooling or convolution layer. The FCL works as a classifier, that is, its output shows the final CNN classification (GOODFELLOW; BENGIO; COURVILLE, 2016).

In the training process of CNNs, data is typically divided into batches, referred to as the batch size (LIANG; HU, 2015). During each training iteration, a batch of data is passed through the network, and the internal model parameters, including weights and biases, are updated using backpropagation, effectively concluding one iteration. Following the completion of an iteration, another batch of data is selected from the dataset to continue the process of training and fine-tuning the internal model parameters of the network. This process iterates until the entire training dataset has been used for training, at which point one epoch is considered complete (GOODFELLOW; BENGIO; COURVILLE, 2016). Subsequent epochs may be carried out to further refine the model's performance.

1.3.1 Inception-v3

Deep convolutional neural networks (DCNNs) are artificial neural networks with multiple layers. As the number of layers increases, so does the computational complexity, which can lead to longer training times, the need for more powerful hardware, and the risk of overfitting.

To address these challenges, the Inception architecture was developed to improve the performance of DCNNs without necessarily increasing the number of layers. Inception originated as a module of another DCNN, GoogLeNet (SZEGEDY et al., 2015), and was first introduced in the 2014 ImageNet Large-Scale Visual Recognition Challenge

(ILSVRC14) [2]. The main idea behind Inception-v1 was to improve classification performance by increasing the number of parallel layers rather than the number of deep layers. To do this, the authors created Inception modules, which consist of convolutional layers and a pooling layer, and inserted kernels of different sizes into the same layer level. In addition, two auxiliary classifiers were inserted into the split part of the network and auxiliary losses were calculated. The total loss function is a weighted sum of the auxiliary loss and the actual loss.

The third version of the Inception architecture was originally released in the 2016 study (SZEGEDY et al., 2016). It is an evolution of the two previous versions that uses different optimization techniques. It is deeper than previous versions, with 42 layers, but it is still efficient and achieves results greater than 78.1% on the ImageNet dataset. Some of the main improvements in Inception v3 include: (i) Factoring,a technique that reduces the number of connections/parameters without decreasing network efficiency. In Inception-v3, three factoring modules A, B, and C were created, which helps to reduce the number of parameters and the possibility of overfitting, enabling the creation of a deeper network; (ii) Auxiliary classifier, that acts as a regularizer (which also helps to avoid overfitting). In Inception-v3, only one auxiliary classifier is used, unlike the first version of Inception, where two were used; (iii) Efficient grid size reduction that aims to reduce the feature map more efficiently by using a convolution block and a pooling block, which are then concatenated. This results in a less expensive and still efficient network; (iv) Label smoothing, another way to prevent overfitting by making the values of the labels more uncertain. This helps to distribute the probabilities among the other labels more evenly, which prevents the network from learning them too accurately and becoming overconfident in its predictions.

1.3.2 ResNet

Introduced by He et al. (HE et al., 2016), the ResNet architecture is known for its use of residual connections, which help to overcome the vanishing gradient problem in deep neural networks. In this architecture there is a connection or identity mapping between layers, which allows the backpropagation signal to be transferred from the output layers, which are the last, to the input layers, which are the first. This connection is known as a skip connection or shortcut connection. Its application in the ResNet architecture is called residual connection. The main idea of the skip connection is to introduce a direct path (shortcut) that skips some layers of the neural network, so that the information (input data and resources extracted by the layers) that leaves a layer flows and serves as input to another layer not adjacent. Using skip connection can help preserve information that could

[2] https://www.image-net.org/challenges/LSVRC/2014

be lost or diluted when passing through multiple layers; in addition, it can contribute to improving the generalization capacity of a network, as it allows the combination of features from different levels of abstraction and resolution. By using skip connections, the ResNet architecture does not increase its computational complexity and makes it possible to increase depth without increasing the training error, as in previous architectures, which allows gains in accuracy and better results.

1.3.3 InceptionResNet-v2

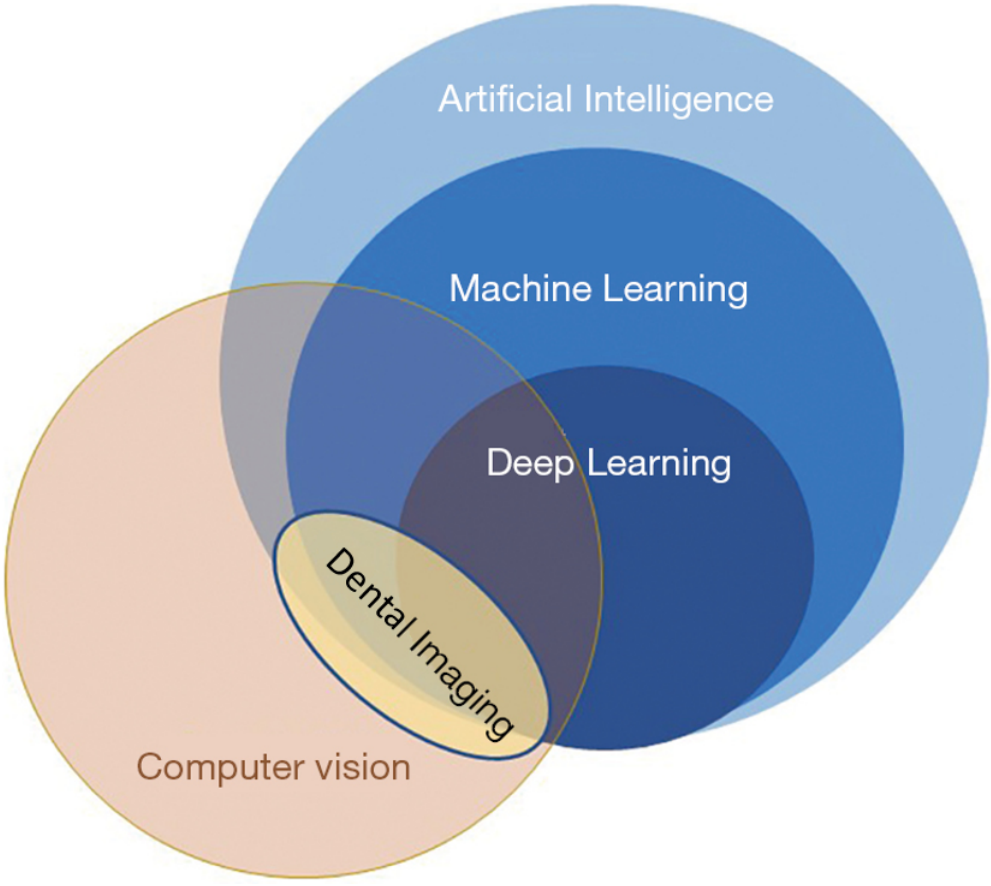
The Inception-ResNet architecture is a fusion of elements from the Inception family and the inclusion of residual connections inspired by the ResNet architecture. This hybrid approach comprises two versions, namely Inception-ResNet-v1 and v2, both adhering to the same underlying principles. Notably, Inception-ResNet-v2 has demonstrated superior accuracy but comes at a higher computational cost compared to its counterpart, version 1. This increased computational demand is a direct consequence of its heightened depth, rendering it particularly well-suited for tasks necessitating high-quality feature representations, such as object recognition in high-resolution images. Both versions made their debut in the same research paper (SZEGEDY et al., 2017), which also introduced the fourth iteration of the Inception architecture. The authors explored the potential advantages of merging the Inception architecture with residual connections. Given that Inception networks tend to exhibit considerable depth, the authors reasoned that substituting the filter concatenation stage of the Inception architecture with residual connections would allow Inception to harness the benefits of the residual approach while preserving computational efficiency. Empirical evidence was then presented to illustrate that training with residual connections significantly expedites the training of Inception networks. The Inception-ResNet-v2 architecture employs Inception-ResNet blocks, which are founded on Inception blocks fused with residual connections. Moreover, the computational load is mitigated by the incorporation of multiple convolutional layers, encompassing both 2D convolutions and channel-separable convolutions, which are a type of convolution that can be more efficient than traditional convolutions.

1.4 Computer Vision

Computer Vision (CV) is a field of Artificial Intelligence that aims to develop artificial systems to deal with visual problems of interest, making use of computers to extract useful information from digital images, videos and other visual inputs (IBM, 2008; FERNANDES; DóREA; ROSA, 2020; PRINCE, 2012).

Computer vision seeks to replicate the functions of human vision. It does this by training machines to understand images and videos, enabling tasks such as detection, segmentation, and classification of objects. Deep Learning and neural networks are widely used in computer vision, but they require a large amount of images and processing power to achieve good results (Esteva et al., 2019; IBM, 2021). In recent years, it has become easier to obtain these requirements, which has allowed for significant improvements in the areas of Machine Learning and Computer Vision. This is due to the fact that there has been a great increase in the availability of datasets combined with a great evolution of graphic processing units (GPUs) (Joe et al., 2017). The figure 6 illustrates a form of relationship between Artificial Intelligence, Machine Learning, Deep Learning, dental imaging and Computer Vision.

Figure 6 – Relationship between Artificial Intelligence, Machine Learning, Deep Learning and Computer Vision.



Source: Adapted from (OLVERES et al., 2021) .

2

Related Work

In this chapter, studies that leveraged Deep Learning to analyze dental radiographs will be presented, serving as fundamental references for the development of this research.

A secondary study was conducted to understand the efforts made to apply Artificial Intelligence, in particular, convolutional neural networks, to the task of classifying radiological images of teeth. Primary studies that could answer the following research question: "What are the existing studies on classifying content extracted from radiographic images of teeth?". The research question is comprehensive, and in-depth analysis techniques, such as meta-analysis and narrative synthesis, were not used in this research. Therefore, this secondary study can be classified as a systematic mapping study (SM) (BA; CHARTERS, 2007).

The SM process was conducted during the first half of 2020 as a part of the Software Engineering discipline within the Applied Computing master's program. It encompassed studies dating back to 2012. Subsequently, in the first half of 2021, a follow-up search was performed using the same search string employed in the prior year's research. The search string used was: ((*artificial intelligence OR neural network*) *AND* ((*decayed teeth OR tooth caries OR teeth caries*) *OR* (*restored teeth OR repaired teeth*) *OR* (*panoramic teeth radiography OR dental x-ray examination*) *OR dental periapical OR* (*teeth numbering OR teeth type identification*)). To create this string, the previously defined research question was taken as a basis. Once defined, the search string was used in the following databases: IEEE[1], PubMed[2], Scopus[3].

Next, the main research used as a basis for this research project will be briefly commented.

Zhang et al. (2018) used a dataset of 1,000 periapical radiographs to propose a method that combines a tree of labels and DL to identify the position of each of the 32

[1] https://ieeexplore.ieee.org/Xplore/home.jsp
[2] https://pubmed.ncbi.nlm.nih.gov
[3] https://www.scopus.com/home.uri

teeth. The label tree decomposed the tooth detection task into subtasks: (i) finding a tooth candidate; (ii) identifying whether the tooth belongs to the upper or lower arch; (iii) identifying the position of the tooth in the arch, from 1 to 8; (iv) identifying the side to which the tooth in belongs, thus classifying the tooth analyzed. The method also had a module to identify the sequence of teeth. The system achieved a precision of 0.958 and a recall of 0.961.

In another study, an application was performed by Chen et al. (2019),with a Deep Learning approach for automatic detection and numbering of teeth, using an implementation of the Faster R-CNN network architecture (REN et al., 2017) and periapical radiographic images. The system uses with three post-processing techniques to supplement the baseline Faster R-CNN to improve detection precisions. In order to evaluate the performance level of the developed tooth detection system, three specialist dentists, with different levels of experience, were invited to carry out the annotation work on the test data set. The performances of the proposed system were close to the level of the less experienced dentist, achieving 0.917 and 0.914 for precision and recall, respectively. However, the automatic system failed in certain cases that could be easily solved by human experts. Thus, the authors suggest that other neural networks and architectures should be tested in future research, aiming at a better method to increase the accuracy of teeth detection results.

TUZOFF et al. (2019) tried to build a CNN-based solution to automatically detect and number teeth using panoramic radiographs, according to the FDI notation. In the study, a dataset of 1,352 radiographs was used for training and another set of 222 were used to evaluate the system's performance and compare it to an expert level. Experts were provided with high resolution panoramic radiographs to delineate bounding boxes around each tooth and assign a class label simultaneously, indicating the tooth number according to the FDI system. Images with implants and bridges were excluded. The authors created a module for detection and another for classification, with each module using a CNN architecture. For detection, a Faster R-CNN architecture was used; for numbering, the authors used a VGG-16 (SIMONYAN; ZISSERMAN, 2014) architecture, together with a heuristic algorithm. The system performance level has come close to expert level. In the context of the teeth detection task, the system achieved 0.9941 for recall and 0.9945 for precision. Regarding teeth numbering, a recall of 0.9800 and a specificity of 0.9994 were achieved. The main reasons for numbering errors included missing teeth close to the target tooth and very small remaining tooth fragments (root remains or decayed teeth).

The review of Schwendicke et al. (2020) describes applications, limitations and perspectives related to AI-assisted dental planning, conduct and diagnostics. The authors emphasize that AI can accelerate the routine activities of professionals, providing them more time to interact with their patients and enabling more humanized care. The

investigators also suggest that AI technologies can reduce the costs of patients' diagnosis and treatment. However, Dentistry has not yet fully embraced the use of AI mainly because of: (1) the lack of data availability caused by data protection concerns and organizational obstacles; (2) the lack of clarity in obtaining and preprocessing data added to the difficulty in creating a single gold standard, which makes it hard to replicate research results in clinical environments; (3) results from the use of AI that are not immediately applicable in dentistry, as they came from studies involving the previously mentioned drawbacks. Finally, the paper emphasizes that AI has great potential to help the dental area and that the current decade will show this is possible.

In a study with panoramic radiographs, Muramatsu et al. (2020) proposed a deep learning model, with the objective of detecting and classifying teeth. Classification involved tooth type (incisor, canine, premolar, and molar) and tooth condition (non-metal restored, partially restored, and fully restored). For detection, the DetectNet (TAO; BARKER; SARATHY, 2016), a CNN architecture designed for object detection tasks, was used. For the classification task, both types and conditions of teeth, a ResNet-50 (Residual Network with 50 layers) was used. As a result, this research obtained a recall of 0.964, in the detection of teeth, and more than 0.900 of accuracy for the classification of dental conditions. However, the study used a low number of panoramic radiographs (100 in total) and did not include third molars in the study. In addition to the results obtained in this research, a point to be highlighted was the researchers' attempt to automate the process of removing structures not related to the mouth from the panoramic image, which are usually present in this type of radiography.

In the 2021 Krois, Schneider, Schwendicke study, the authors aim to use CNN to classify teeth. They hypothesize that the classification performance of CNNs can be improved by providing image context information around the region of interest (ROI). To show this idea, the authors used a ResNet-34 network to classify teeth using the FDI notation. The teeth were annotated by a team of 50 specialists, who drew bounding boxes around 124,314 teeth, distributed in 5,008 panoramic radiographs. These teeth were cut, at the points referring to the BBs, and served as input for training and testing the ResNet-34. The authors then expanded the dimensions (width and height) of the original BBs, annotated by experts, by 150, 200, 250, and 300%. Consequently, the F1-score of the results increased from 0.767 to 0.933, comparing the original BBs (100%) with the expanded BBs (300%).

In the study conducted by Lee et al., 2018, an approach was introduced for the classification of decayed and healthy teeth in periapical radiographs of molars and premolars. The dataset used for this purpose comprised 3,000 radiographs, specifically excluding non-permanent teeth. These radiographs were categorized into two groups: dental caries and non-dental caries, and the categorizations were performed by four trained and cer-

tified dentists. Each tooth was manually cropped to be used as input for the proposed CNN and was resized to a uniform size of 299x299 pixels. To augment the dataset and increase its diversity, various transformations were applied, including rotation (within a range of 10°), width and height shifting (within a range of 0.1), zooming (within a range of 0.8–1.2), shearing (within a range of 0.5), and horizontal flipping. The CNN used for teeth classification was based on a pre-trained Inception-v3 network. The fine-tuning technique was employed, reusing pre-trained weights from the network and calibrating only the final layers for the specific classification task. The study reported the following results: for premolars, the achieved accuracy was 0.89, and the recall was 0.84; for molars, the accuracy reached 0.880, and the recall was 0.92; for both premolars and molars, the diagnostic accuracy was 0.820, and the recall was 0.810.

In the research conducted by Moran et al. (2019), CNNs were proposed for the identification of decayed teeth in bitewings. Additionally, the cavities were categorized based on lesion severity, with classifications including superficial or advanced. The created classes encompassed normal, beginner, and advanced. Two architectures, Inception-v1 and ResNet-50, were implemented for this task. A dataset consisting of 480 tooth images extracted from 112 bitewings was used. Pre-processing techniques were applied to enhance the separation of tooth areas from the background, and this enabled the creation of bounding boxes to isolate individual teeth. These image processing steps were implemented using the Python language and the scikit-image library, which offers a collection of image processing algorithms. Annotations were provided by an expert. The best results were obtained with the Inception-v1 model, using a learning rate of 0.001, achieving an accuracy of 0.733 on the test set. For the normal class (no cavities), the precision, recall, specificity, negative predictive value (NPV), and area under the curve of the receiver operating characteristic (AUC-ROC) values were 0.818, 0.600, 0.933, 0.823, and 0.643, respectively. For the incipient class (superficial lesions), the precision, recall, specificity, NPV, and AUC-ROC values were 0.722, 0.866, 0.833, 0.926, and 0.861, respectively. For the advanced class (advanced lesions), the precision, recall, specificity, NPV, and AUC-ROC values were 0.687, 0.733, 0.833, 0.862, and 0.810, respectively. The study compared the best results with annotations provided by two dentists who were familiar with oral radiology but had substantially less experience than the expert whose annotations were considered the ground truth (gold standard). The findings indicated that the CNN performed at a level similar to that of the less experienced dentists.

Silva et al. (2020) conducted instance segmentation and tooth numbering (utilizing the FDI system) on panoramic radiographs by employing Deep Learning techniques. They assessed the performance of Mask R-CNN, HTC, ResNeSt, and PANet networks using a dataset consisting of 543 images. The authors reported the most favorable outcomes achieved with the PANet network, yielding a segmentation accuracy of 71.9 mAP and a numbering accuracy of 74.0 mAP.

Pinheiro et al. (2021) conducted a comparison between two networks: Mask R-CNN (HE et al., 2017) and Mask R-CNN with the PointRend (KIRILLOV et al., 2019) module, for the purpose of segmenting and numbering both permanent and deciduous teeth in panoramic radiographs. They employed a modified version of the UFBA-UESC Dental Images dataset, encompassing 450 images with teeth segmented individually (SILVA; OLIVEIRA; PITHON, 2018). The Mask R-CNN with PointRend configuration achieved noteworthy outcomes, achieving a segmentation accuracy of 77.3 mAP and a numbering accuracy of 75.3 mAP.

Finally, a systematic mapping study was conducted in partnership with master's student José Andery Carneiro (CARNEIRO et al., 2023b). This secondary study aimed to map primary research on the detection or segmentation of teeth, caries and restorations, and the classification of teeth on dental radiographs. The study compiled the characteristics of the datasets, techniques, and evaluation criteria used in the articles. It also discussed patterns, gaps, and opportunities that could help future research in the area. The results of the study can be reproduced over time to explore potential research gaps. The systematic mapping protocol included the following steps: (i) automatic searches of relevant information in four sources, ACM Digital Library[4], IEEE Xplore, PubMed, and Scopus; (ii) application of inclusion/exclusion criteria to articles published between 2012 and 2023; (iii) formulation and use of four search strings (S1, S2, S3, and S4) for searches in the chosen databases; (iv) elimination of duplicated articles, application of inclusion/exclusion criteria to the title and abstract of the articles, and application of the same criteria to the complete article, in addition to a manual search. As a result, 69 studies were accepted.

In the context of this systematic mapping study, several notable trends and observations have emerged:

- Popular Network Architectures: among the various neural network architectures explored, U-Net, Mask R-CNN, and Faster R-CNN have emerged as the most widely utilized choices.

- Common Evaluation Metrics: the evaluation of results consistently relies on metrics such as accuracy, recall, and precision, which are frequently employed to assess the performance of the developed models.

- Data Preference: most of the research articles surveyed predominantly focus on employing panoramic radiographs for their segmentation tasks, indicating a prevalent data preference in this domain.

[4] https://dl.acm.org

- Growing Emphasis on Segmentation: there is a noticeable trend towards an increased application of segmentation tasks in Dentistry-related research.

However, to facilitate the practical implementation of these promising proposals in the field of Dentistry, certain critical gaps need to be addressed:

- Quality Public Databases: the creation and availability of high-quality public databases specifically tailored to Dentistry are essential to support further research and development efforts.
- Methodology Dissemination: clear and comprehensive documentation and dissemination of the methodologies created are crucial for ensuring the replicability and understanding of research findings.
- Reproducibility Solutions: the provision of accessible solutions and resources for reproducibility is essential to enable the broader adoption of the proposed methods in everyday dental practice.
- Addressing these gaps will be pivotal in translating innovative research into practical applications that benefit the field of Dentistry.

This study highlights a crucial issue related to the absence of standardization in the examined research. While the articles provide details about the annotation process, the authors' descriptions of these procedures lack consistency. As a result, the data sources, annotation protocols, and presentation of selected configurations frequently appear incomplete or unclear. All these limitations increase the risk of bias and raise doubts about the robustness, reliability and replicability of the study results. (SCHWENDICKE et al., 2021). A solution suggested by the authors is a better structuring of the planning, conduction and reporting of paper methodologies, using standard reporting guidelines or checklists. There are several checklists for research in the field of AI available in the literature, such as: CONSORT-IA (SCHULZ; ALTMAN; MOHER, 2010), STARD (LIU et al., 2020), CLAIM (MONGAN; MOY; JR, 2020) and the checklist published by the authors of (SCHWENDICKE et al., 2021).

The Checklist for Artificial Intelligence in Medical Imaging (CLAIM) (MONGAN; MOY; JR, 2020) was followed in this research, when applicable. CLAIM is a set of guidelines designed to assist researchers in transparently reporting their work on Artificial Intelligence (AI) in medical imaging, promoting clarity and reproducibility. It serves as a valuable tool, fostering improved transparency, enhanced reproducibility, and better clinical translation of research in AI for medical imaging. By adhering to CLAIM, researchers contribute to building trust in the field, facilitating the adoption of AI into clinical practice. CLAIM ensures clear and understandable reporting, facilitates study reproducibility, and aligns research outcomes with practical applications in patient care.

3

Materials and Methods

The objective of this research is to employ Convolutional Neural Networks (CNNs) to automate the tasks of tooth numbering and identification of decayed teeth. The overarching objective was to harness the findings from this research, in conjunction with (CARNEIRO et al., 2023a), to develop a versatile system capable of seamlessly integrating advanced technologies for oral health analysis. The design of this system allows for potential expansion by creating specialized modules dedicated to tasks associated with the analysis of panoramic radiographs of teeth.

For the execution of teeth numbering and decayed teeth identification tasks, an extensive evaluation was conducted utilizing two distinct CNN architectures: Inception-v3 and InceptionResNet-v2. In the context of teeth numbering, meticulous application of two distinct strategies was undertaken for each of the aforementioned network architectures, with the primary goal of determining the most effective combination of strategy and architecture that yields optimal results. In the pursuit of identifying decayed teeth, the Inception-v3 and InceptionResNet-v2 architectures were similarly employed. Notably, to enhance the results in alignment with the chosen metrics, the initial dataset was augmented with a publicly available dataset.

The core objective behind the utilization of these two architectures was to assess the feasibility of constructing specialized modules tailored to the tasks of tooth numbering and decayed teeth identification. These modules were envisioned as integral components of a collaborative project between this researcher and José Andery Carneiro, a master's student enrolled in the Postgraduate Program in Applied Computing at FFCLRP-USP. This collaborative initiative further encompassed researchers affiliated with the Faculty of Dentistry of Ribeirão Preto (FORP-USP). This work was part of a multidisciplinary research group at the University of São Paulo (USP) called InReDD (Interdisciplinary Research group in Digital Dentistry)[1].

The goal of the modular system, collaboratively devised by the author and mas-

[1] InReDD. Acessible at https://sites.usp.br/inredd/mestrado

ter's student José Andery Carneiro, centers on the efficient segmentation, numbering, and identification of decayed teeth in panoramic dental radiographs. The assessment of performance primarily focuses on the following metrics: (i) F1-score and average precision (AP), pertinent to the tasks of mouth detection and teeth segmentation, which were developed by the aforementioned student but are not the subject of this work; (ii) for the tasks of tooth numbering and decayed tooth classification, the evaluation encompasses precision, recall, F1-score, and accuracy, and the outcomes of these assessments are expounded upon in this study. To realize these objectives, a systematic methodology was adopted. This approach entails the utilization of a modular technique wherein a combination of distinct deep learning networks is strategically deployed to address the multifaceted challenges inherent in these sequential tasks:

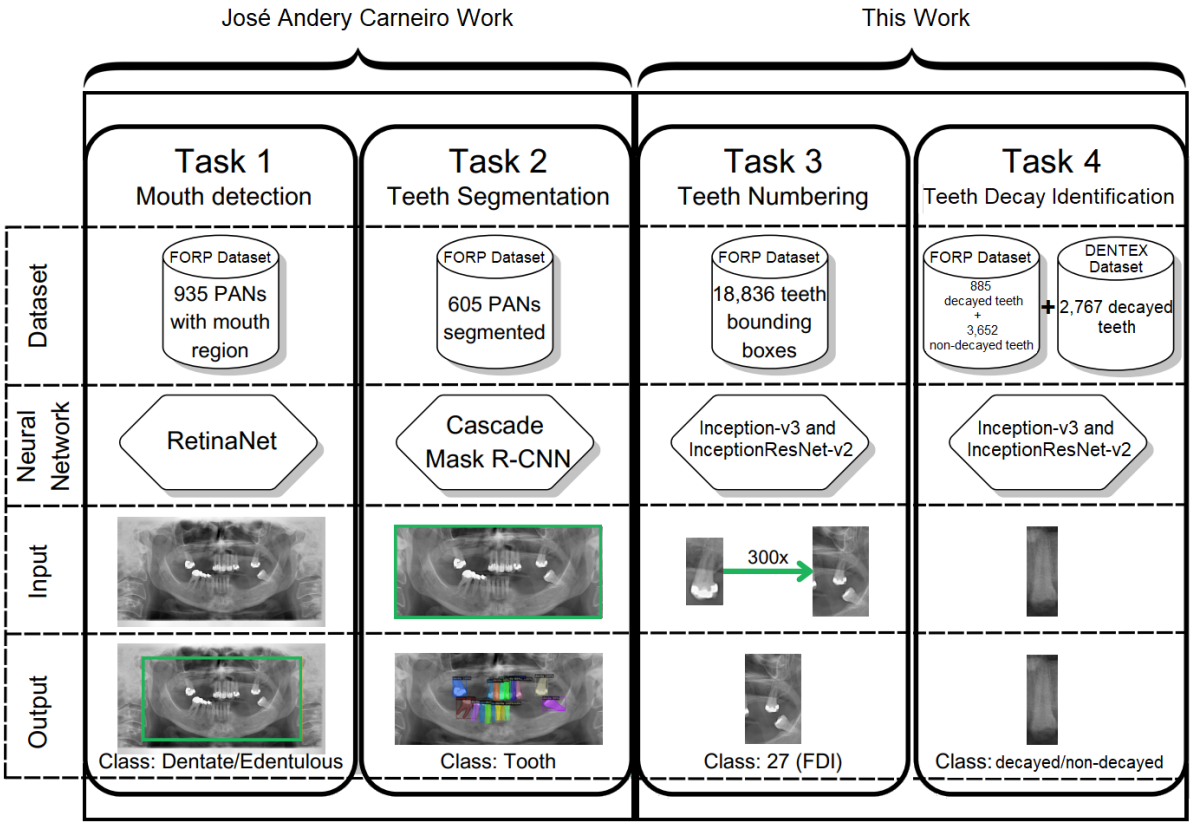
- *Task 1: Detection of the mouth region.* A PAN stands out for including all the teeth of the mandible and maxilla, as well as part of the facial bones in just one image (SILVA; OLIVEIRA; PITHON, 2018). The mouth detection task aims to reduce the radiographic region to be analyzed by removing irrelevant areas, considering the purpose of the modular system. To accomplish this task, almost 1,000 PANs and the RetinaNet (LIN et al., 2017) network were used.

- *Task 2: Segmentation of each tooth.* The segmentation task delimits the edges of each tooth. The regions created can be used for various applications, such as diagnosis of dental pathologies. The Cascade Mask R-CNN (CAI; VASCONCELOS, 2017) network and approximately 600 segmented PANs were applied to perform this task.

- *Task 3: Assignment of FDI numbers to the segmented teeth.* In the proposed modular system, the regions created from the segmentation are cut to serve as input for the numbering task. The output is a segmented panoramic numbered with the notation FDI. A total of 18,836 tooth images were used for this task. An overview of this task, which comprises the numbering module, is presented in Figure 8. More details of this task will be presented in this chapter.

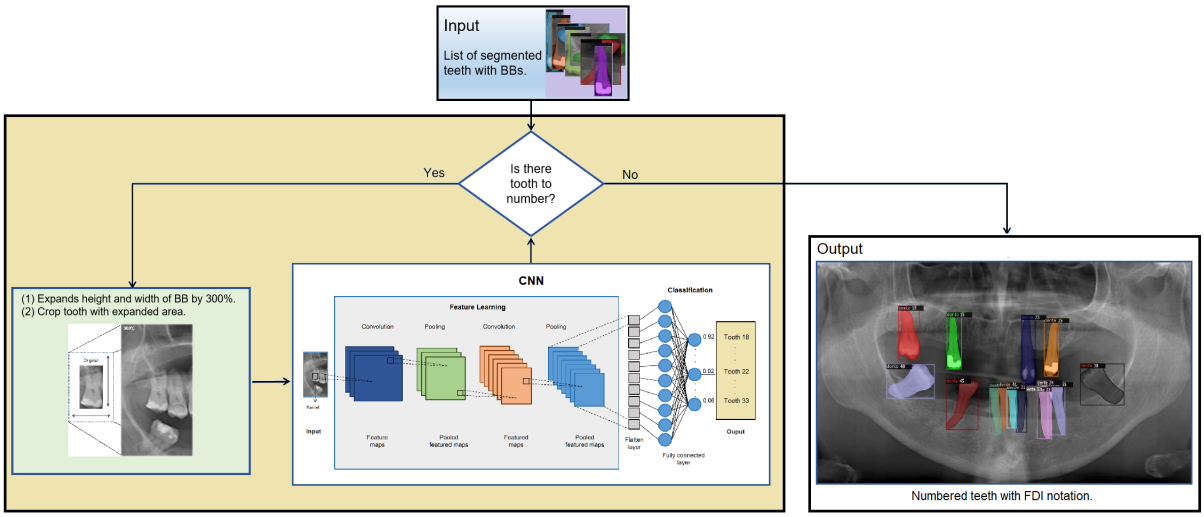
- *Task 4: Identification of decayed teeth.* In the proposed modular system, the identification of decayed teeth is a subsequent step to tooth numbering, that is, right after task 3. The teeth numbered in the previous task serve as input for this task. This allows showing whether a given numbered tooth is decayed or not. A total of 7,304 images of decayed and non-decayed teeth were used for this task, coming from two datasets. An overview of this task, which comprises the teeth decay identification module, is presented in Figure 9. More details of this task will be presented in this chapter.

Figure 7 – Modular system overview. Tasks 1 and 2: developed by master's student José Andery Carneiro and are detailed in (CARNEIRO et al., 2023a). Tasks 3 and 4: tooth numbering and identification of decayed teeth, detailed in this work.



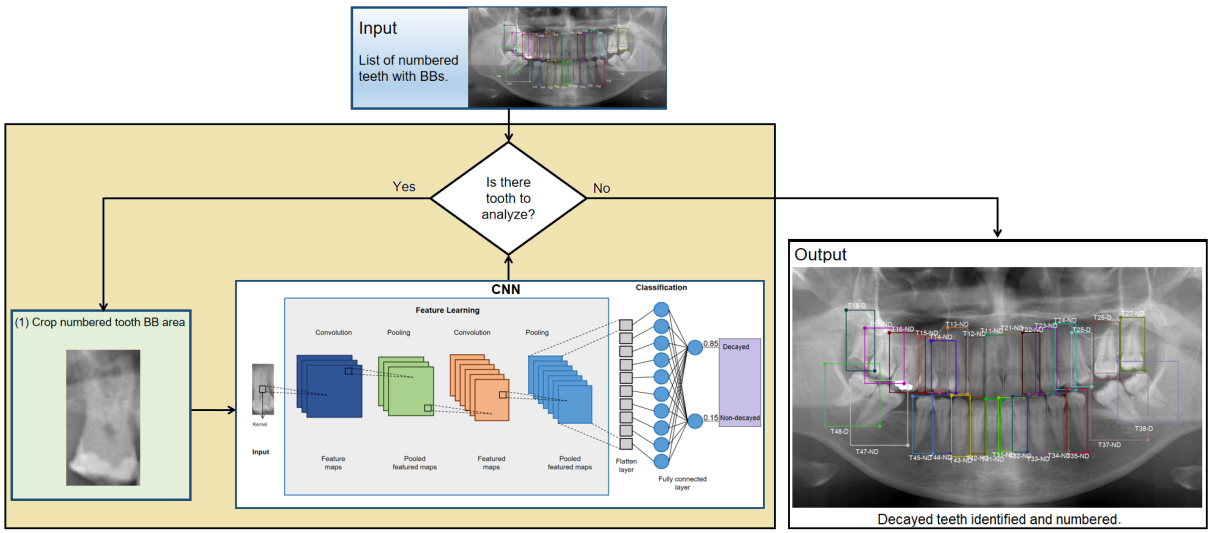
Source: Author' collection.

Figure 8 – Flowchart for Assignment of FDI Numbers Teeth - Task 3.



Source: Author' collection.

Figure 9 – Flowchart for Identifying Decayed Teeth - Task 4.



Source: Author's collection.

For future applications, it is expected that the segmented teeth from Task 2 will be used as input for Task 3; furthermore, the teeth numbered in task 3 are expected to be used as a road to task 4. For the experimental procedures, two datasets were employed, namely, a dataset supplied by the Faculty of Dentistry of Ribeirão Preto (FD) and a separate publicly available dataset (PD). The FD dataset was utilized for both the tooth numbering experiments and the decayed teeth identification experiments, whereas the PD dataset was exclusively employed for the identification of decayed teeth. Further elaboration on each dataset will be provided in the subsequent sections.

3.1 FORP-USP Dataset - FD

The initial dataset was comprised of 935 dental panoramic radiographs (PANs), meticulously curated by an Interdisciplinary Group for Research in Digital Dentistry (InReDD). This dataset was subject to rigorous ethical scrutiny and received approval from the local Research Ethics Committee (Plataforma Brasil, CAAE: 51238021.2.0000.5419). Notably, the dataset, as described in reference (COSTA et al., 2023), encompasses PANs exclusively from adults aged 18 and above, so that each radiograph referred to a different patient. These radiographs were randomly selected by a qualified radiologist from the extensive clinical image database housed at the Faculty of Dentistry of Ribeirão Preto (FORP-USP). This study contains the first results of experimentation using this dataset (FORP-USP Dataset - FD).

The radiographs were initially obtained as part of routine patient care, with explicit consent from the patients for research purposes. These images were sourced from the same clinical setting and were generated using the Veraviewepocs device manufactured

by J. Morita in Japan. It's worth noting that different exposure factors were applied, tailored to the specific characteristics of each patient. The panoramic radiographs collected encompassed the complete maxillary and mandibular arches, maintaining a high level of diagnostic quality characterized by adequate sharpness and contrast. Notably, the dataset intentionally excluded radiographs featuring deciduous teeth or mixed dentition, supernumerary teeth, fractures, periodontal diseases, any history of surgical treatment, the presence of orthodontic appliances, cysts, or congenital anomalies in the maxillofacial region. Furthermore, radiographs exhibiting low contrast, poor sharpness, or motion artifacts were rigorously screened and omitted from the dataset. This selective process was implemented to ensure that the final image dataset remains free of outliers and extraneous noise, thereby maintaining its integrity for research purposes.

The selected images were anonymized and stored in a PACS system (Picture Archiving and Communication System) called LyriaPacs - I-Medsys - Innovative Medical Informatics software [2], in JPEG format, with dimensions of 2903 x 1536 pixels and 300 dpi. In this manipulation, no brightness, contrast, cropping, or resizing adjustments were made.

The dataset was annotated in two cycles, as follows:

1. Initially, all th 935 radiographs were annotated with bounding boxes by three radiologists with 10 years of experience. Radiologist 1 labeled the mouth (maxilla and mandible) and each tooth on the radiography (including crown and root). Radiologist 2 independently verified the annotation. There were 163 discrepancies between the two radiologists, which is just 0.7% of the total annotations. For these cases, Radiologist 3 was consulted for decision-making. Mouths were labeled as dentate, edentulous, edentulous mandibles, or edentulous maxillae. Teeth were labeled based on the FDI notation. Decayed teeth were also noted at this stage. The resultant dataset comprises a comprehensive collection of annotations, including 935 bounding boxes delineating the mouth regions, encompassing 741 dentate, 120 edentulous, 60 edentulous maxillae, and 14 edentulous mandibles. Furthermore, the dataset encompasses 20,250 bounding boxes meticulously outlining individual teeth, including 885 decayed teeth. This initial phase of annotation has laid the foundation for *Task 1*, *Task 3* and *Task 4*, as depicted in Figure 7, utilizing the 935 annotated radiographs featuring delineated mouth regions.

2. Subsequently, in a follow-up step, two radiologists undertook the individual tooth segmentation task for a subset of the dataset. This subset consisted of 605 radiographs, randomly selected from the initial pool, with the exclusion of images featuring edentulous mouths (totaling 120 images). The experts performed pixel-

[2] Available on http://lyria.i- medsys.com/lyriaViewer-web/

wise segmentation of the tooth structures using the LabelMe open-source annotation software [3]. This segmentation process closely adhered to the labels generated during the first cycle of the annotation process. In this segmentation task, spanning the 605 selected images, the radiologists labeled a total of 14,582 teeth. The resulting set of 605 segmented radiographs, generated in this second cycle, forms the basis for the execution of *Task 2*, as illustrated in Figure 7.

The annotation protocol was created by a team of radiologists, dentists, and computing professionals from the InReDD group. The team met regularly to discuss the protocol and ensure that it was comprehensive and accurate. The protocol was designed to calibrate the annotators and ensure that the data was robust, manipulable, workable, and easy to use. The protocol defines the different types of annotations that will be made, such as the location of the mouth and teeth, the condition of the teeth, and the presence of any abnormalities. In addiction, the protocol also specifies the level of detail that is required for each annotation.

3.2 Public Dataset - PD

The public dataset (PD) [4] was specifically curated for the Dental Enumeration and Diagnosis on Panoramic X-rays Challenge (DENTEX) held in partnership with the Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) in 2023. DENTEX's primary objective is to stimulate the development of algorithms capable of accurately identifying abnormal teeth through dental enumeration and associated diagnosis. This endeavor seeks to contribute to treatment planning and the execution of more precise dental procedures. It's important to note that all requisite permissions for using these images were duly obtained from an ethics committee, and the dataset was made available under a Creative Commons Attribution license (CC-BY).

The panoramic radiographs constituting the dataset originate from three distinct institutions, each with its unique clinical practices, equipment, and imaging protocols, consequently leading to images of varying quality. The images included in this dataset pertain to patients aged twelve years or older and were randomly selected from the respective hospitals' databases, preserving the privacy and confidentiality of individuals.

To ensure the quality, accuracy, and reliability of the panoramic radiographs annotations, a team of dentistry specialists undertook this crucial task. Initially, each image received annotations from a final-year Dentistry student, followed by meticulous review and correction by one of three other experts, each boasting over fifteen years of experience.

[3] http://labelme.csail.mit.edu
[4] available at https://github.com/ibrahimethemhamamci/DENTEX

The annotations provided by these professionals encompass the FDI number and four categories of dental abnormalities: caries, deep caries, periapical lesions, and impacted teeth. Teeth marked as having caries or deep caries were instrumental in composing the dataset dedicated to decayed teeth. For experiments pertaining to the decayed teeth classification task, a total of 2,767 images were incorporated into the FORP-USP decayed teeth dataset, encompassing 3,652 decayed teeth instances. Conversely, an equal number of non-decayed teeth, amounting to 3,652, were sourced exclusively from the FORP-USP dataset for comparison and analysis.

3.3 Models and eligibility

In this section, the eligibility criteria for the network architectures employed are provided, along with a detailed exposition of the development environment created to ensure the successful execution of the tasks of tooth numbering and decayed tooth identification.

For the tasks of *assignation of FDI numbers to teeth* and *identification of decayed teeth*, the second version of the InceptionResNet architecture, InceptionResNet-v2, was implemented. To compare the performance of InceptionResNet-v2, the third version of the Inception architecture, Inception-v3, was chosen. Both architectures were implemented by modifying code available on GitHub that use TensorFlow's high-level Application Programming Interface (API)[5], Keras[6], written in Python programming language, which facilitates the creation and training of deep learning models.

The study was conducted while considering eligibility criteria that support the selection of networks. Muramatsu et al. (MURAMATSU et al., 2020) used a network with ResNet-50 architecture to classify teeth into four types: incisors, canines, premolars and molars. Using images with an input resolution of 224x224 and applying data augmentation and four-fold cross-validation, the result was an average accuracy of 0.932. Lin & Chang (LIN; CHANG, 2021) proposed a way to number teeth and classify six different dental conditions, including orthodontics, endodontic therapy, dental restoration, impaction, implant, and dental prosthesis, using a ResNet-152 network. After data augmentation was applied, accuracy improved significantly. In particular, the mean accuracy for numbering teeth, with input images with a resolution of 227x227, reached 0.956. Based on the excellent results obtained by the studies presented, equivalent architectures were chosen in terms of precision and number of parameters, for *assignation of FDI numbers to teeth* task, since the objective is not a direct comparison with other investigations, but rather to propose ways to contribute to future efforts. The eligibility criteria were discovered in a systematic mapping (CARNEIRO et al., 2023b).

[5] https://www.tensorflow.org
[6] https://keras.io

Table 2 shows accuracy and the number of parameters in the ImageNet dataset, from different versions of the ResNet, Inception and InceptionResNet architectures. The "Top-1 Accuracy" column refers to the accuracy of the model when the most likely prediction is the expected response; the "Top-5 Accuracy" column, on the other hand, refers to the accuracy of the model when any one of the five most likely responses predicted by the model matches the expected response. Finally, the "Parameters" column refers to the total parameters, in millions, of the architectures presented in the table. One may observe that the Inception-v3 architecture, implemented in the present work, has higher accuracies (Top-1 and Top-2) than the ResNet-50 implemented in (MURAMATSU et al., 2020), but with a smaller number of parameters; likewise, it can be noted that the architecture InceptionResNet-v2, implemented in present work, has higher accuracies (Top-1 and Top-2) than ResNet152, implemented in (LIN; CHANG, 2021), but with a smaller number of parameters. The accuracies shown in Table 2 suggest a possible improvement in the quality of the results, as well as a smaller number of parameters that can contribute to saving training time and computational power.

To choose network architectures for the *decayed teeth identification* task, inspiration was drawn from the works presented in the "Related Work" Section. In previous studies conducted by (MORAN et al., 2021) and (LEE et al., 2018), which focused on detecting decayed teeth in radiographs, they employed the Inception-v1 and Inception-v3 architectures, respectively, achieving excellent results. Consequently, the initial choice was the Inception-v3 architecture. Furthermore, in an endeavor to enhance the results and pinpoint the optimal model for identifying decayed teeth, the information presented in Table 2 once more served as a guide, leading to the selection of the InceptionResNet-v2 architecture.

Table 2 – CNN Classification Networks Comparison

Model	Top-1 Accuracy	Top-5 Accuracy	Parameters
ResNet50	74.9%	92.1%	25.6M
ResNet50-v2	76.0%	93.0%	25.6M
ResNet152	76.6%	93.1%	60.4M
Resnet152-v2	78.0%	94.2%	60.4M
Inception-v3	77.9%	93.7%	23.9M
InceptionResNet-v2	80.3%	95.3%	55.9M

The development environment and programming language used were, respectively, Google Collaboratory (Colab [7]) and Python. The environment was configured with the Ubuntu 20.04.5 LTS operating system, with 12.7 GB of RAM (Random-Access Memory)

[7] Colab is available on https://colab.research.google.com and it is a cloud platform, developed by a team belonging to Google's scientific research area, Google Research (https://research.google), which offers a programming environment and allows anyone with a Google account to develop and run codes in Python and R, using the browser.

and 168.8 GB of disk memory. In addition, Colab provided a free NVIDIA GPU (Graphics Processing Unit), model Tesla T4, with 15 GB of RAM. This GPU can be used for Deep Learning implementations in Colab through a computing platform that allows programmers to take advantage of the processing power of the GPU, the CUDA (Compute Unified Device Architecture). The CUDA used was version 12.

3.4 Experiments

This section describes the training and testing of the Deep Learning models, as well as the input datasets and parameters used for each model. It also covers the data augmentation techniques employed and the anticipated output for each task.

Five-fold cross-validation was adopted as a standard procedure to evaluate the performance of training and testing of machine learning models. To establish the folds, let **E** represent the original set of images adapted for each task. Five folds (**k=5**) were generated—namely **F1, F2, F3, F4,** and **F5**—through the division of **E** into roughly equivalent portions. This division ensured no image repetition across the folds while preserving the distribution between various classes (such as different tooth types and positions). The cross-validation process was executed five times. In each iteration, a distinct fold (**FT**) was designated for validation (**V**) and testing (**T**), while the remaining four folds were utilized for training (**Tr**). The partition between **V** and **T** was arranged so that 50% of the images from the total **FT** set constituted the validation set, while the remaining 50% comprised the testing set. As a result, distinct datasets were generated, allocating 80% of the images for training, 10% for validation, and 10% for testing.

Data augmentation was performed using the class ImageDataGenerator from Keras. This class allows, at each epoch, the original images to be transformed based on predefined data augmentation parameters, and then used as input to train a CNN. This allows training without repeating images, avoiding overfitting and helping to better generalize the generated model.

For configuring hyperparameters in both the tasks of *assignation of FDI numbers to teeth* and *decayed teeth identification*, initial assignments were made using default values from the Keras library; specifically, a learning rate of 0.001 and a batch size of 32. The selected optimization algorithm was Adam, a first-order gradient-based optimization algorithm designed for stochastic objective functions, leveraging adaptive estimates of lower-order moments. Recognized for its adaptive learning rate properties, Adam was chosen due to its widespread usage in Deep Learning, as emphasized in (GOODFELLOW; BENGIO; COURVILLE, 2016). It has demonstrated effectiveness across various deep learning tasks, including CNNs used for classifying images, object detection, and

segmentation. However, it's essential to note that while Adam is a widely employed optimizer, its performance may vary depending on the specific task and dataset. During model execution, the monitoring of the loss function, precision, recall, and accuracy values was initiated for both training and validation. Subsequently, an empirical early stopping strategy was employed: if there was no improvement in the validation loss function value after three epochs, the training was concluded. Furthermore, if, after one epoch, the validation loss function value did not increase, the learning rate value was halved. Once these criteria were defined, batch size values were empirically adjusted, referencing values published in the foundational works related to this research. This iterative process led to the final set of values used in this study.

The training for the tasks proposed by this study will be detailed below.

3.4.1 Assignation of FDI Numbers to Teeth

The accomplishment of task of *assignation of FDI numbers to teeth*, followed the algorithm presented in Algorithm 1 composed by two distinct strategies of classification, namely, Strategy 1 (S1) and Strategy 2 (S2). S1 and S2 were created to comparative analysis of accuracy, recall, precision and F1-score.

In S1, the algorithm attaches each tooth into one of the 32 unique FDI identifiers, as illustrated in Lines 2 to 4. Alternatively, S2 attributes FDI numbers to teeth, while dividing this classification task into two sequential steps (Step1 and Step2), as depicted in Lines 5 to 32. Step 1 (Lines 5 and 6) employs a categorization process wherein each tooth was grouped into one of eight categories: Upper Incisor (UI), Lower Incisor (LI), Upper Canine (UC), Lower Canine (LC), Upper Premolar (UP), Lower Premolar (LP), Upper Molar (UM), and Lower Molar (LM). Subsequently, in Step 2 (Lines 7 to 33), for each category, one network algorithm was trained to determine a potential FDI identifier for the respective tooth. In this way, each numbering was done with different training. Therefore, eight different models, for each architecture, were generated in Step 2. For example, a tooth initially classified as an Upper Incisor in Step 1 could only be further categorized as 11, 12, 21, or 22, which are the precise FDI numbers corresponding to upper incisors. Similarly, a tooth identified as a Lower Canine could only be assigned the FDI numbers 33 or 43. By breaking down the FDI numbering task into two successive steps, S2 aimed to enhance the precision and accuracy of the FDI numbering process, thereby simplifying the overall task. This idea is similar to (ZHANG et al., 2018), whose authors proposed a method to identify the position of each of the 32 teeth, in periapical radiographs, through the combination of a label tree and DL.

Training and evaluation were conducted using the InceptionResNet-v2 and Inception-v3 models via Keras API. During the training, model weights pretrained on the ImageNet

Algorithm 1: Assignation of Numbers to the Segmented Teeth

Input: Segmented tooth image in task 2.

```
switch Strategy do
    case Strategy 1 do
        Classify the tooth into one of 32 possible numbers with the FDI system.
    end
    case Strategy 2 do
        Classify the tooth into UI, LI, UC, LC, UP, LP, UM or LM;
        switch Classified tooth do
            case UI do
                Classify the tooth in 11, 12, 21 or 22.
            end
            case LI do
                Classify the tooth in 31, 32, 41 or 42.
            end
            case UC do
                Classify the tooth in 13, 23.
            end
            case LC do
                Classify the tooth in 33, 43.
            end
            case UP do
                Classify the tooth in 14, 15, 24 or 25.
            end
            case LP do
                Classify the tooth in 34, 35, 44 or 45.
            end
            case UM do
                Classify the tooth in 16, 17, 18, 26, 27 or 28.
            end
            case LM do
                Classify the tooth in 36, 37, 38, 46, 47 or 48.
            end
        end
    end
end
```

Output: Tooth image numbered using FDI notation.

data set were used. The training and testing procedures involved utilizing cropped images of teeth from panoramic radiographs as input data. The idea was to simulate the entry of already segmented teeth with bounding boxes, resulting from tasks 1 and 2 presented in Figure 7. The base points for cropping were determined using coordinates from bounding boxes (BB) that had been previously annotated by specialists. For training, the BB of teeth annotated by experts in the first cycle of annotation were used (see Section 3). This ensured that the networks were trained with ground truth data. For future applications, the segmented teeth from Task 2 are expected to be used as input.

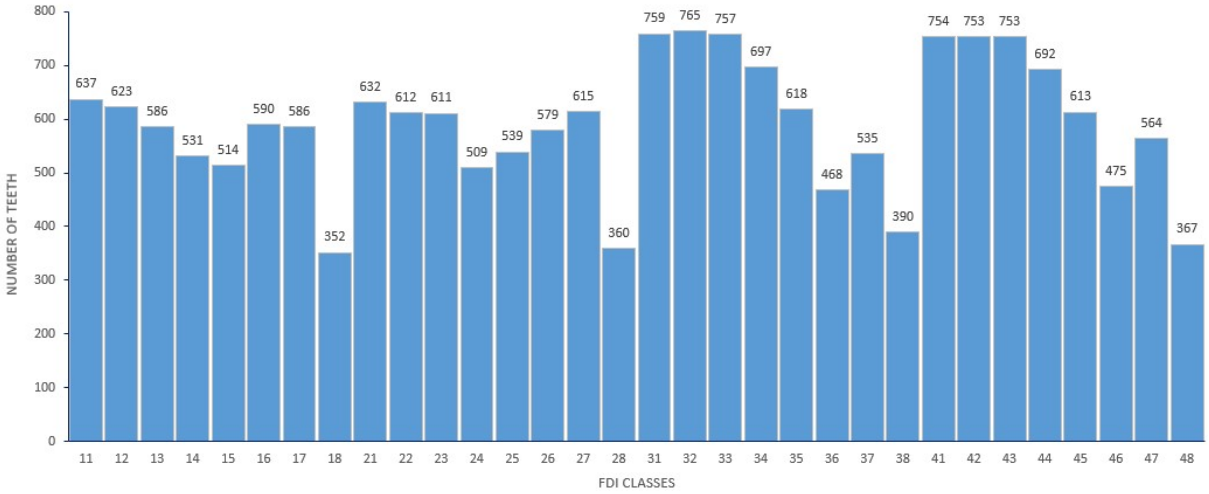
Prior to training, a pre-processing method was performed to exclude images with excessively blurred teeth or those featuring restorations that hindered the identification of boundaries between neighboring teeth because they would not significantly contribute to the training of the neural networks. Additionally, images containing only a pin in place of a tooth were also discarded, considering the specific demands of this classification task. As a result, a total of 18,836 teeth images remained for both training and testing purposes. The quantities of these teeth categorized according to the FDI numbering system is displayed in Figure 10. The dental group with the highest number of images was the molar group (5,881), which corresponds to the sum of teeth 16, 17, 18, 26, 27, 28, 36, 37, 38, 46, 47 and 48; followed by the incisor group (5,535), which corresponds to the sum of teeth 11, 12, 21, 22, 31, 32, 41 and 42; premolar group (4,713), which corresponds to the sum of teeth 14, 15, 24, 25, 34, 35, 44 and 45; and canine (2,707), which corresponds to the sum of teeth 13, 23, 33 and 43. Furthermore, Figure 11 shows the distribution of teeth according to the arch and type of tooth, after pre-processing. For lower premolars (LP), the number of resulting samples was 3,563; for upper premolars (UP), was 3,262; for lower incisors (LI), was 3,031; for upper incisors (UI), was 2,504; for lower molars (LM), was 1,856; for upper molars (UM), was 1,913; for lower canines (LC), was 1,510; finally, for upper canines (UC), the number of samples was 1,197.

Table 3 shows the sizes of the datasets for each of the iterations belonging to the five-fold cross-validation. The first column, Str(Stp), indicates the strategy number; the value between parentheses, Stp, indicates the step number (if any) of the adopted strategy. The second column, "Classes", shows the classes used in each strategy; the third column, "Train/test (1 - 4)" presents the training (including validation) and test set sizes of four iterations of the cross-validation. Likewise, the fourth column "Train/test (5)" shows the sizes of the training (including validation) and test sets, of the fifth iteration cross-validation.

To avoid overfitting in the training performed, as well as to improve the performance and results in the tests of the trained networks, the images of the dataset were diversified using different data augmentation (DA) strategies. For all training, a preprocessing that can be considered as a type of DA was adopted, which consisted of increasing

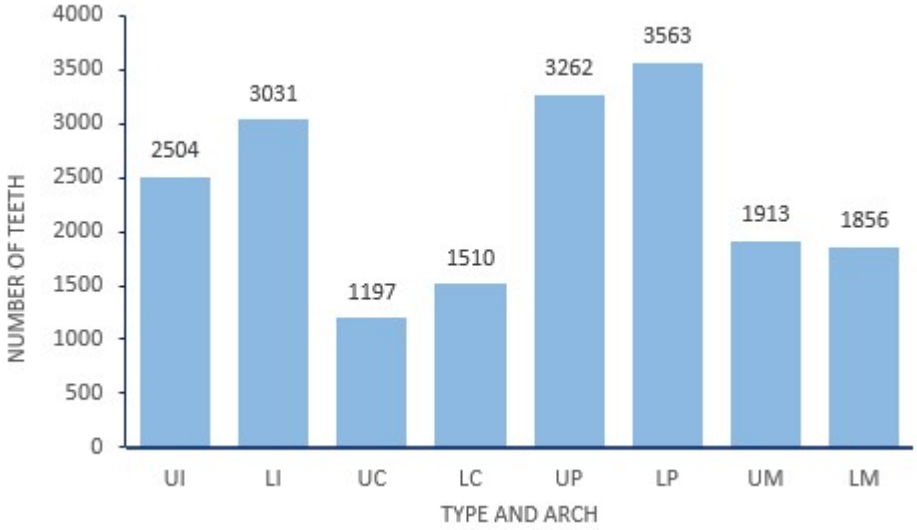
the size of the bounding boxes annotated by specialists around each tooth. This idea was presented by (KROIS; SCHNEIDER; SCHWENDICKE, 2021).

Figure 10 – Number of teeth by FDI number, after preprocessing.



Source: Author's collection.

Figure 11 – Number of teeth by type and arch, after preprocessing.



Source: Author's collection.

Thus, the height and width of the bounding boxes were increased by 300%, which resulted in the inclusion of teeth close to the tooth under analysis, within these BBs. In addition, three other types of data augmentation were adopted: (i) vertical shift; (ii) horizontal shift; and (iii) resizing. Two of them, horizontal shift (horizontal displacement) and resize (zooming), were inspired by the study (LIN; CHANG, 2021), which showed excellent results in the task of classifying teeth and six different types of dental conditions, using these types of DA. The third type of DA was a vertical shift, added empirically. For horizontal and vertical shifts, a range of 0 to 25% was defined; in the case of zoom, the range was from 0 to 15%. These three types of DA were applied using the class ImageDataGenerator, from Keras.

Table 3 – Datasets for Each Strategy.

Str(Stp)	Classes	Train/test(1-4)	Train/test(5)
1(-)	11-18, 21-28, 31-38, 41-48	16951/1885	16917/1919
2(1)	UI, LI, UC, LC, UP, LP, UM, LM	16957/1879	16924/1912
2(2)	11, 12, 21, 22	2254/250	2249/255
2(2)	13, 23	1077/120	1076/121
2(2)	14, 15, 24, 25	1884/209	1877/216
2(2)	16, 17, 18, 26, 27, 28	2773/309	2770/312
2(2)	31, 32, 41, 42	2730/301	2722/309
2(2)	33, 43	1359/151	1356/154
2(2)	34, 35, 44, 45	2358/262	2353/267
2(2)	36, 37, 38, 46, 47, 48	2518/281	2514/285

For the implementation of the InceptionResNet-v2 network, training was varied from 15 to 25 epochs, using of batch size 16, dropout 0.2 and learning rate of 0.001. For the implementation of the Inception-v3 network, training was varied from 30 to 45 epochs, using batch size of 16, dropout 0.6 and learning rate 0.0001. These values were defined from values recommended in the codes obtained from GitHub and empirically changed through various training sessions. The input resolution of the images was 299x299 pixels.

3.4.2 Identification of Decayed Teeth

For the execution of the *decayed teeth identification* task, Inception-v3 and InceptionResNet-v2 architectures, employing the Keras API along with modified code sourced from GitHub, were implemented. Additionally, this task was conducted with the utilization of two datasets. Both network architectures were implemented for comparative analysis of accuracy, recall, precision and F1-score.

Training and testing were conducted using the InceptionResNet-v2 and Inception-v3 models in Keras via its API. During the training, model weights pretrained on the ImageNet data set were used. For this task, cropped images of teeth extracted from panoramic radiographs were used as the input dataset. The selection of base points for cropping was determined based on bounding box (BB) coordinates, which had been previously annotaded by domain experts. During the initial annotation cycle, as described in Section 2.1, these experts also marked whether or not a particular tooth was decayed. In forthcoming applications, it is anticipated that the numbered teeth obtained from Task 3 will serve as the input data for the network selected in Task 4.

For training and testing for the task of identifying decayed and non-decayed teeth, two datasets were used. One of them was the dataset used in the tooth numbering task, collected at the Faculty of Dentistry of Ribeirão Preto (FORP-USP) and annotated by professionals linked to FORP-USP. The FORP-USP dataset (FD), used for the task of

identifying decayed teeth, was complemented by a public dataset (PD). Further explanations about the two datasets are presented in Chapter 4.

The use of PD was driven by three primary considerations. Firstly, the dataset annotated by experts affiliated with FORP-USP contained a limited number of decayed teeth, specifically 885, up to the point when the decayed teeth identification tests presented in this study were conducted. Secondly, within the FD dataset, a notable proportion of teeth with restorations also exhibited cavities, which has the potential to exert an influence on the network's learning, even when employing data augmentation techniques. Thirdly, the inclusion of datasets sourced from distinct radiography devices and representing diverse communities can contribute to the improved generalization of predictions on images not previously presented to a CNN.

Several strategies were attempted exclusively with the FD dataset. Initially, in an attempt to balance the data set, the same number of images of decayed teeth and non-decayed teeth were separated. Additionally, efforts were made to exclude teeth with restorations from the non-decayed teeth class. Furthermore, different combinations of hyperparameters and data augmentation techniques were tried, in pursuit of enhancing results based on the selected metrics. However, it became apparent that the CNNs exhibited notably better performance in identifying non-decayed teeth. Consequently, PD was turned to in an endeavor to bolster network learning.

A total of 2,767 images of decayed teeth from PD were incorporated. When combined with the 885 from the FD dataset, this resulted in a total of 3,652 decayed teeth samples. As for non-decayed teeth images, only those sourced from the FORP-USP dataset were included. Specifically, 3,652 images of non-decayed teeth from the FD were selected, excluding teeth with extensive restorations, excessively broken teeth and dental pins. Consequently, the dataset employed for the decayed teeth identification task comprises an equal number of decayed and non-decayed teeth images, totaling 3,652 for each category and thus culminating in a grand total of 7,304 images.

For training and testing, the procedure of dividing the dataset into five folds was adopted to apply five-fold cross validation. To create each fold, the proportion between decayed teeth from the FD and PD dataset was maintained. This was done by dividing the total number of images from each dataset, separately, into five parts. Subsequently, the images of decayed teeth from FD and PD were combined. For non-decayed teeth, only FD images were used. Table 4 shows the fold sizes for each of the iterations belonging to the five-fold cross validation.

Table 4 – Dataset for Each Fold - Decayed Teeth Identification Task.

Fold(s)	Classes	Train	Valid	Test
1,2,3 and 4	Decayed, Non-decayed	5844	730	730
5	Decayed, Non-decayed	5840	732	732

Once the dataset was formed, some data augmentation techniques were chosen and adopted, using the class ImageDataGenerator, from Keras: (i) vertical flip; (ii) horizontal flip; (iii) rotate; (iv) brightness variation; (v) zoom variation. Vertical flip and horizontal flip allow you to "mirror" the images to simulate different teeth. Rotations were adopted to randomly rotate the images clockwise or counterclockwise, by 20º, changing the position of the object in the frame. Randomly increasing image brightness, darkening and lightening images was used to simulate the effect of different X-ray machines. For brightness increase, it ranged up to 30%; as for darkening, the brightness was varied by up to 60%. Random zoom changes, varying the size of the images by 30%, decreasing and increasing, were applied, in an attempt to enrich the types of images used as input. The values for each type of data augmentation strategy were obtained empirically.

For the implementation of the InceptionResNet-v2 network, training 35 epochs were used, with batch size of 8, dropout 0.2 and learning rate of 0.001. For the implementation of the Inception-v3 network, training 37 epochs were used, using batch size of 8, dropout 0.6 and learning rate 0.001. These values were defined from values recommended in the codes obtained from GitHub and empirically changed through various training sessions. The input resolution of the images was 299x299 pixels.

3.5 Metrics

The results were evaluated using accuracy, recall, precision e F1-score. These metrics were chosen because they were the most frequently used, according to the second mapping presented in (CARNEIRO et al., 2023b). The equations (3.1),(3.2),(3.3),(3.4) show the definitions of these metrics, in terms of true positives (TP)[8], true negatives (TN)[9], false positives (FP)[10] and false negatives (FN)[11]. A brief explanation of each metric is shown below:

- Accuracy: measures the degree of correct classifications that a trained machine learning model achieves and can be calculated as the number of objects classified correctly by the total number of objects. The (3.1) equation shows how to calculate accuracy.

- Precision: is the rate of correct predictions among all model predictions. The (3.2) equation shows how to calculate precision.

[8] True positives are the positive examples that are correctly classified by a classification model.
[9] True negatives are the negative examples that are correctly classified by a classification model.
[10] A false positive is an example of a negative class that has been incorrectly classified as positive.
[11] A false negative is an example of a positive class that has been incorrectly classified as negative.

- Recall: that is the proportion of ground truth objects correctly predicted by the model. The (3.3) equation shows how to calculate recall.

- F1-score (F-measure): allows the calculation of the harmonic mean between precision and recall. F1-Score values range from zero to 1, and the higher the value, the greater the classification performance(TAHA; HANBURY, 2015). The (3.4) equation shows how to calculate the F1-score.

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

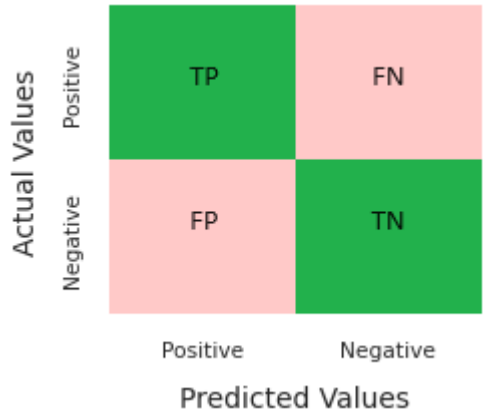
$$P = \frac{TP}{TP + FP} \tag{3.2}$$

$$R = \frac{TP}{TP + FN} \tag{3.3}$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.4}$$

To offer a detailed perspective on the performance of the generated CNN models, confusion matrices were utilized. A confusion matrix represent a tabular format extensively employed for assessing the effectiveness of classification models in machine learning. The matrix provides a comprehensive view of the model's predictions in comparison to the actual values, differentiating between TP, TN, FP and FN classifications (POWERS, 2008; FAWCETT; PROVOST, 2013). To exemplify the construction of a confusion matrix, consider the case of a binary classification. First, a table is created to represent a 2x2 matrix, composed of two rows and two columns. The rows represent the actual classes, while the columns represent the predicted classes. Subsequently, the matrix is populated with the counts of both true and false predictions derived from a specific dataset. Figure 12 displays an example of a 2x2 confusion matrix. For multiclass classification, involving more than two classes, the confusion matrix takes the form of an N x N matrix, where N denotes the number of classes.

Figure 12 – Example of a 2x2 confusion matrix.



Source: Author's collection.

In Chapter 4, the document presents confusion matrices associated with the optimal outcomes achieved during the conducted tests for *decayed teeth identification* task.

4

Results and Discussion

In this section, the results achieved for the FDI tooth numbering task will be presented, using the two strategies explained in the previous section, highlighting the predominant erroneous predictions. The results relating to tests for the task of identifying decayed teeth will also be presented. Furthermore, an in-depth analysis and discussion of the achieved results is conducted.

For the task proposed in this research, *assignation of numbers to teeth*, the 18,836 radiographs of teeth were divided into folds. To assist the computation of the metrics outlined in Chapter 3, a confusion matrix was produced following the conclusion of each test. Table 5 presents the results for Strategy 1 (S1); Table 6 presents the results for Strategy 2 (S2), Step 1; and Table 7 presents the results for Strategy 2, Step 2. The values of Precision (Pre), Recall (Rec), Accuracy (Acc), and F1-Score (F1) are presented, resulting from the tests for numbering of teeth performed with the Inception-v3 and InceptionResNet-v2 networks. Each table shows the arithmetic mean of the tests performed with the models obtained with five-fold cross-validation.

S1 trained the networks directly using thirty-two classes, referring to the positions of each tooth, identified with the FDI numbering. According to the Table 5, the InceptionResNet-v2 architecture achieved better results in all metrics and classes. Furthermore, it can be observed that molars 17, 18, 27, 28, 37, 38, 47, and 48 exhibit the lowest F1-score values for both networks. The overall means of the results are shown in last row of table. For Inception-v3 the metrics precision, recall, accuracy and F1-score were, respectively: 0.923, 0.918, 0.995 and 0.919. For these same metrics, with InceptionResNet-v2, were obtained 0.979, 0.978, 0.998 and 0.978, respectively.

S2, step 1, classified the type of tooth and its arch, without taking into account the FDI number of each tooth. For all eight classes in Table 6, the networks obtained values above 0.9. For the InceptionResNet-v2 architecture, all values were above 0.98. The overall means of the results are shown in last row of table. For Inception-v3 the metrics precision, recall, accuracy and F1-score were, respectively: 0.964, 0.966, 0.992

Table 5 – Results from implementing Inception-v3 and InceptionResNet-v2 for Strategy 1, aiming to distinguish 32 FDI classes in a single step.

	Inception-v3				InceptionResNet-v2			
C	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1
11	0.973	0.985	0.999	0.978	0.997	1.000	1.000	0.998
12	0.965	0.949	0.997	0.957	0.997	0.987	0.999	0.992
13	0.935	0.963	0.997	0.948	0.980	0.990	0.999	0.985
14	0.915	0.883	0.994	0.898	0.981	0.962	0.998	0.971
15	0.912	0.883	0.994	0.898	0.962	0.988	0.999	0.975
16	0.963	0.919	0.996	0.939	0.990	0.969	0.999	0.979
17	0.867	0.946	0.994	0.903	0.957	0.963	0.997	0.959
18	0.925	0.874	0.996	0.892	0.951	0.960	0.998	0.955
21	0.991	0.978	0.999	0.984	0.997	1.000	1.000	0.998
22	0.978	0.967	0.998	0.972	1.000	1.000	1.000	1.000
23	0.959	0.974	0.998	0.966	0.987	0.993	0.999	0.990
24	0.939	0.887	0.995	0.909	0.984	0.973	0.999	0.978
25	0.899	0.912	0.995	0.905	0.993	0.989	0.999	0.991
26	0.941	0.904	0.995	0.920	0.983	0.983	0.999	0.983
27	0.873	0.919	0.993	0.895	0.974	0.952	0.998	0.962
28	0.908	0.906	0.996	0.905	0.937	0.972	0.998	0.954
31	0.891	0.835	0.989	0.861	0.984	0.984	0.999	0.984
32	0.855	0.901	0.990	0.877	0.992	0.982	0.999	0.987
33	0.912	0.948	0.994	0.928	0.992	1.000	1.000	0.996
34	0.929	0.946	0.995	0.937	0.992	0.994	0.999	0.993
35	0.953	0.907	0.995	0.929	0.994	0.981	0.999	0.987
36	0.921	0.920	0.996	0.920	0.983	0.979	0.999	0.981
37	0.877	0.900	0.993	0.885	0.953	0.963	0.998	0.958
38	0.901	0.903	0.996	0.895	0.961	0.959	0.998	0.959
41	0.950	0.830	0.991	0.883	0.979	0.987	0.999	0.983
42	0.899	0.942	0.993	0.919	0.989	0.976	0.999	0.983
43	0.948	0.958	0.996	0.953	0.987	0.997	0.999	0.992
44	0.943	0.922	0.995	0.931	0.997	0.983	0.999	0.990
45	0.908	0.935	0.995	0.920	0.984	1.000	0.999	0.992
46	0.923	0.850	0.994	0.881	0.983	0.979	0.999	0.981
47	0.870	0.922	0.993	0.891	0.944	0.954	0.997	0.949
48	0.924	0.903	0.997	0.911	0.946	0.925	0.997	0.935
-	0.923	0.918	0.995	0.919	0.979	0.978	0.998	0.978

and 0.965. For these same metrics, with InceptionResNet-v2, were obtained 0.994, 0.994, 0.998 and 0.994, respectively.

The results of step 2 of S2 are presented in Table 7. In this table, the first column, "G", shows the group (UI, LI, UC, LC, PS, PI, MS, MI) to which the "C" classes from step 2 belong. The following results can be highlighted:

- The first group in Table 7, UI, shows the results for class 11, 12, 21 and 22, referring to the Upper Incisor class. It can be noted that teeth 21 and 22 achieved the best values for all metrics, for both networks. In contrast, the values for tooth 11, for

Table 6 – Results from implementing Inception-v3 and InceptionResNet-v2 for Strategy 2, Step 1 aiming to distinguish 8 classes, including the arch and type of each tooth.

	Inception-v3				InceptionResNet-v2			
C	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1
UI	0.986	0.986	0.996	0.986	0.999	0.998	1.000	0.999
LI	0.995	0.968	0.994	0.981	0.996	0.996	0.999	0.996
UC	0.933	0.965	0.993	0.948	0.985	0.995	0.999	0.990
LC	0.910	0.943	0.988	0.926	0.989	0.991	0.998	0.990
UP	0.975	0.927	0.989	0.950	0.995	0.982	0.997	0.988
LP	0.975	0.950	0.990	0.963	0.998	0.995	0.999	0.997
UM	0.969	0.994	0.994	0.981	0.992	0.998	0.998	0.995
LM	0.970	0.995	0.995	0.982	0.996	0.999	0.999	0.998
-	0.964	0.966	0.992	0.965	0.9940	0.9942	0.9987	0.9941

Inception-v3, were the lowest.

- The next group, LI, presents the results for the numbering of teeth 21, 22, 31, and 32, referring to the Lower Incisor class. It can be noted that the results obtained for the InceptionResNet-v2 network were much higher than those obtained for InceptionV3, for all classes.

- The posterior group, UC, presents the results for numbering teeth 13 and 23, referring to the Upper Canine class.

- Classes 33 and 43, referring to the LC group, lower canines, appear next. It can be noted that the results obtained for both networks, for the numbering of the upper and lower canines, reached excellent results. In particular, for lower canines, InceptionResNet-v2 predicted correctly all numbers for test images.

- The next group, UP, presents the results for the numbering of teeth 14, 15, 24 and 25, referring to the Upper Premolars class. It can be noted that the performance of InceptionResNet-v2 was considerably higher than that of Inception-v3. In particular, the results obtained by InceptionResNet-v2 reached much higher values of precision, recall and F1-Score.

- The LP group shows the results for numbering teeth 34, 35, 44, and 45, referring to the Lower Premolar class. It can be noted that the InceptionResNet-v2 values were all above 0.97. In contrast, Inception-v3 obtained some values below 0.90; in particular, the recall for tooth 34 was 0.895 and the accuracy for tooth 35 was 0.887.

- The next group, UM, presents the results for the numbering teeth 16, 17, 18, 26, 27, and 28, referring to the Upper Molar class. It can be noted that in Inception-v3, with the exception of tooth 16, all f1-score values were below 0.9. With regard to

InceptionResNet-v2, all values were above 0.9, with emphasis on the values obtained for tooth 16, which were all above 0.94.

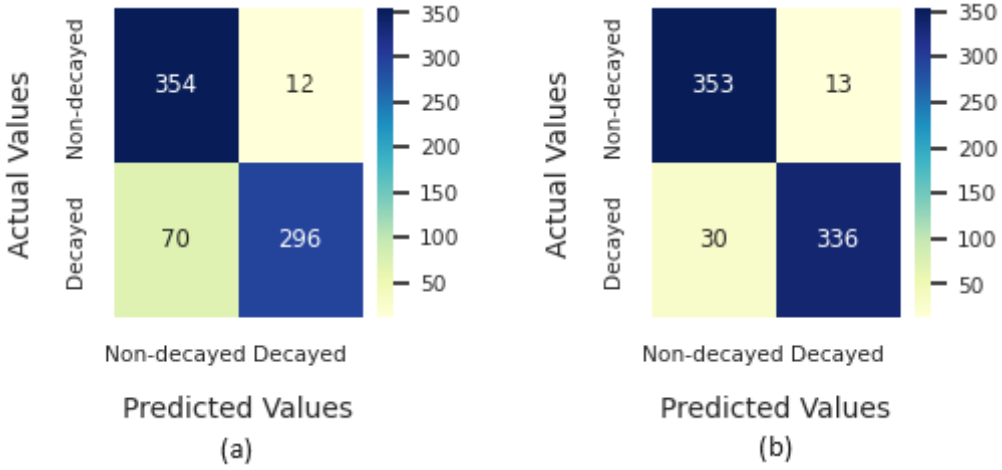
- Finally, the group LM presents the results for numbering teeth 36, 37, 38, 46, 47 and 48, referring to the Lower Molar class. It can be noted that in InceptionV3, with the exception of teeth 36 and 46, all f1-score values were below 0.9. With regard to InceptionResNet-v2, all values were above 0.9, with emphasis on the values obtained for tooth 36, which were all above 0.97.

Table 7 – Results from implementing Inception-v3 and InceptionResNet-v2 for Strategy 2, Step 2 aiming to numbering the teeth categorized in Step 1.

		Inception-v3				InceptionResNet-v2			
G	C	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1
UI	11	0.893	0.957	0.958	0.922	0.988	0.985	0.993	0.986
	12	0.966	0.894	0.966	0.927	0.986	0.987	0.993	0.986
	21	0.959	0.930	0.972	0.944	0.994	0.997	0.998	0.995
	22	0.951	0.971	0.980	0.960	0.997	0.993	0.998	0.995
LI	31	0.881	0.812	0.923	0.842	0.982	0.974	0.989	0.978
	32	0.889	0.935	0.953	0.910	0.979	0.990	0.992	0.984
	41	0.840	0.814	0.914	0.825	0.987	0.971	0.989	0.979
	42	0.873	0.905	0.943	0.887	0.980	0.992	0.993	0.986
UC	13	0.984	0.993	0.988	0.988	0.990	1.000	0.995	0.995
	23	0.993	0.984	0.988	0.988	1.000	0.990	0.995	0.995
LC	33	0.997	0.992	0.995	0.995	1.000	1.000	1.000	1.000
	43	0.992	0.997	0.995	0.995	1.000	1.000	1.000	1.000
UP	14	0.885	0.883	0.941	0.883	0.963	0.974	0.984	0.968
	15	0.885	0.876	0.941	0.879	0.958	0.961	0.980	0.959
	24	0.889	0.836	0.933	0.858	0.977	0.950	0.982	0.963
	25	0.851	0.894	0.931	0.869	0.968	0.978	0.986	0.973
LP	34	0.970	0.895	0.964	0.930	0.979	0.986	0.990	0.982
	35	0.887	0.955	0.961	0.920	0.985	0.974	0.990	0.979
	44	0.954	0.905	0.963	0.928	0.980	0.983	0.990	0.981
	45	0.900	0.957	0.965	0,927	0.980	0.977	0.990	0.979
UM	16	0.944	0.881	0.952	0.911	0.996	0.949	0.985	0.972
	17	0.803	0.868	0.904	0.834	0.919	0.932	0.957	0.924
	18	0.905	0.829	0.956	0.864	0.994	0.943	0.989	0.967
	26	0.941	0.856	0.944	0.893	0.984	0.935	0.977	0.958
	27	0.827	0.845	0.901	0.834	0.919	0.929	0.955	0.922
	28	0.845	0.922	0.957	0.879	0.916	0.961	0.977	0.937
LM	36	0.967	0.919	0.971	0.942	0.988	0.979	0.991	0.983
	37	0.831	0.907	0.917	0.864	0.941	0.930	0.963	0.934
	38	0.952	0.826	0.956	0.880	0.984	0.944	0.985	0.963
	46	0.940	0.921	0.964	0.928	0.968	0.988	0.988	0.977
	47	0.833	0.897	0,914	0.862	0.957	0.943	0.970	0.950
	48	0.913	0.813	0.946	0.853	0.952	0.946	0.980	0.949

For the task proposed in this research, *identification of decayed teeth*, the 7,304 of teeth were divided into five folds. Figure 13 illustrates the confusion matrices corresponding to the optimal outcomes achieved in the tests for the generated models. Specifically, Figure 13 (a) presents the confusion matrix values for the Inception-v3 architecture, while Figure 13 (b) displays the values for the InceptionResNet-v2 architecture. The results depicted in both matrices pertain to the fourth training iteration utilizing five-fold cross-validation.

Figure 13 – Confusion matrices illustrating the outcomes of the most successful training for identifying decayed teeth. Figure 13(a) depicts the results from the model developed with the Inception-v3 architecture, while Figure 13(b) illustrates the outcomes from the model created with the InceptionResNet-v2 architecture.



Source: Author's collection.

Table 8 presents the results of each of the five tests, indicated in the "Test" column, referring to the two classes created, "Non-decayed" and "Decayed", resulting from the implementations of the Inception- v3 and InceptionResNet-v2. The following results can be highlighted:

- In test 1, it can be noted that the precision for non-decayed teeth obtained the highest results for both networks, with the exception of Recall in Inception-v3. On the other hand, all values referring to decayed teeth in the Inception-v3 network reached, at most, 0.85. As for InceptionResNet-v2, all values for decayed teeth above 0.85 were obtained, highlighting the precision of 0.934 for this class. In the case of the non-decayed teeth class, InceptionResNet-v2 reached all values above 0.9.

- In test 2, it is noted that the difference between the values of decayed and non-decayed teeth from Inception-v3 increased, indicating that the network learned to identify non-decayed teeth much better. In the case of InceptionResNet-v2, this

difference between classes was smaller, which indicates a better ability to recognize both classes.

- In test 3, it can be seen that the difference between the values of decayed and non-decayed teeth from Incpetion-v3 decreases (with the exception of the precision value), compared to test 2. However, it can be seen that the values for non-decayed teeth decreased, whereas there was a very small improvement for decayed teeth. As for InceptionResNet-v2, it showed improvement in all values, for both classes.

- In test 4, it is noted that, with the exception of precision, the differences between results for decayed and non-decayed teeth, from Inception-v3, increased in relation to test 3. As for InceptionResNet-v2, it presented all results above 0.9 ; furthermore, the difference between the values obtained for the two classes decreased.

- In test 5, the differences between results from the Inception-v3 network had little change, in relation to test 4. As for InceptionResNet-v2, again all values above 0.9 were obtained.

Table 9 displays the average values of the outcomes across the five folds presented in Table 8, categorized by individual classes as well as both classes combined. The following results can be highlighted:

- On average, in first line, recall was the lowest metric for both the ”non-decayed” and ”decayed” classes. InceptionResNet-v2 had smaller differences in performance between the two classes than Inception-v3, indicating that it can distinguish between the two classes more accurately. The ”non-decayed” class was the best recognized by both networks.

- Finally, in the last line it is possible to check the values of the two implemented models. Inception-v3 obtained 0.95, 0.866, 0.910 and 0.910 for the precision, recall, accuracy and F1-score metrics, respectively. As for InceptionResNet-v2, it obtained 0.963, 0.914, 0.939 and 0.937, for the precision, recall, accuracy and F1-score metrics, respectively.

Table 8 – Five-fold cross validation results - Identification of Decayed Teeth.

		Inception-v3				InceptionResNet-v2			
Test	Classes	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1
1	Non-decayed	0.990	0.817	0.897	0.895	0.991	0.940	0.963	0.965
	Decayed	0.823	0.850	0.845	0.836	0.934	0.855	0.899	0.893
2	Non-decayed	0.992	0.962	0.974	0.976	0.991	0.893	0.940	0.940
	Decayed	0.953	0.781	0.872	0.859	0.895	0.907	0.904	0.901
3	Non-decayed	0.991	0.899	0.940	0.943	0.992	0.936	0.961	0.963
	Decayed	0.891	0.831	0.869	0.860	0.927	0.872	0.903	0.899
4	Non-decayed	0.992	0.970	0.979	0.981	0.992	0.965	0.977	0.978
	Decayed	0.964	0.809	0.889	0.880	0.963	0.918	0.941	0.940
5	Non-decayed	0.991	0.929	0.957	0.959	0.991	0.951	0.970	0.971
	Decayed	0.920	0.817	0.875	0.866	0.948	0.902	0.927	0.925

Table 9 – Average values of the results presented in Table 8.

	Inception-v3				InceptionResNet-v2			
Classes	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1
Non-decayed	0.991	0.915	0.950	0.951	0.991	0.937	0.962	0.963
Decayed	0.910	0.818	0.870	0.860	0.934	0.891	0.915	0.911
Non-Decayed/Decayed	0.951	0.866	0.910	0.910	0.963	0.914	0.939	0.937

The *assignation of numbers to teeth* was approached through two distinct strategies to achieve optimal performance to precision, recall, accuracy, and F1-score metrics. Strategy 1 aimed to directly number teeth using the FDI system, involving training the Inception-v3 and InceptionResNet-v2 networks to classify 32 unique classes. On the other hand, Strategy 2 tried to accomplish the same task through two filtering steps. Regarding the *identification of decayed teeth* task, the same architectures as the numbering task were utilized. Additionally, two distinct datasets were employed, with annotations performed by qualified professionals to ensure the consistency, reliability and quality of the annotations.

Upon analyzing Tables 5, 6, and 7, it becomes evident that InceptionResNet-v2 consistently outperformed or at least tied with Inception-v3 in all tests and for all proposed classes. Remarkably, this was achieved using the same images for training, validation, and testing. Additionally, while training with Inception-v3 spanned between 35 and 50 epochs for the same datasets and classes, InceptionResNet-v2 yielded superior results within a range of 10 to 25 epochs. Another notable aspect was the impact of context information on model training, which involved expanding the tooth analysis area by 300%. This approach was inspired by (KROIS; SCHNEIDER; SCHWENDICKE, 2021), where teeth were classified into incisors, canines, premolars, and molars. Through a 300% increase in bounding boxes, they achieved precision and F1-Score values of 0.940 and 0.934, respectively. This concept parallels the first step of Strategy 2 in the present study. However, in this research, a more thorough investigation into class specification was carried out by

identifying the arch to which the teeth belong, resulting in precision, recall, and F1-Score values of 0.994, 0.994, and 0.994, respectively.

Turning to the combined results of both steps in Strategy 2, precision and recall reached 0.979 and 0.978, respectively. In contrast, (ZHANG et al., 2018) reported precision and recall values of 0.958 and 0.961, respectively. In Table 10, the outcomes attained by InceptionResNet-v2, the architecture that demonstrated the most favorable outcomes across all metrics, are showcased. The column displaying the results for Strategy 2 was computed by multiplying the results from step 1 with the corresponding FDI numbers from step 2. For instance, the precision outcomes for class UI (Upper Incisor) in step 1 were multiplied by the precision results for teeth 11, 12, 21, and 22 in step 2.

Upon scrutinizing Table 10, it's apparent that Strategy 1 consistently outperformed Strategy 2 in most aspects. Thus, Strategy 1 emerged as the superior choice for tooth numbering, yielding average values of 0.979, 0.978, 0.998, and 0.978 for precision, recall, accuracy, and F1-Score, respectively.

The exceptional performance demonstrated by Strategy 1 can be attributed to postive factors such as the random division of images across created folds, the quantity and quality of available images per class, the selection and combination of hyperparameters within chosen network architectures, and the adopted data augmentation strategies. It's imperative to emphasize that the quality of the annotations played a pivotal role in contributing to the results. The fact that these annotations were carried out by experienced professionals following a rigorous and meticulous analysis and numbering protocol ensured a consistency that greatly facilitated the training of the selected CNNs.

As with the numbering task, the InceptionResNet-v2 implementation for the *identification of decayed teeth* task outperformed all values achieved by Inception-v3. Notably, it is observed that, in addition to yielding the best results, InceptionResNet-v2 exhibited smaller differences between the "Decayed" and "Non-decayed" class values. This suggests that this network implementation recognizes each class with greater accuracy. Furthermore, it's evident that for both networks, precision achieved higher average values than recall. This indicates that the generated models excel at correctly identifying true positive cases while being cautious about labeling items as positive, thereby reducing false positive errors. Consequently, these models may not capture all positive cases in the dataset, but when they do label something as positive, it is usually accurate. The values obtained for precision and recall directly impact the F1-score values, as it is the harmonic mean of precision and recall. In the context of identifying decayed teeth, where both precision and recall are crucial, the F1-score serves as a valuable metric for evaluating overall model performance.

In the study conducted by (MORAN et al., 2021), image pre-processing and data augmentation techniques were applied to a set of bitewing radiographs, which had been

Table 10 – Best results achieved through the application of Strategies 1 and 2. Both Strategy 1 and Strategy 2 yielded their most successful results when employing implementations of InceptionResNet-v2.

	Strategy 1				Strategy 2 (Step 1 and Step 2)			
C	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1
11	0.997	1.000	1.000	0.998	0.987	0.983	0.993	0.985
12	0.997	0.987	0.999	0.992	0.985	0.986	0.993	0.985
13	0.980	0.990	0.999	0.985	0.975	0.995	0.994	0.985
14	0.981	0.962	0.998	0.971	0.923	0.945	0.969	0.933
15	0.962	0.988	0.999	0.975	0.954	0.905	0.969	0.929
16	0.990	0.969	0.999	0.979	0.989	0.947	0.983	0.967
17	0.957	0.963	0.997	0.959	0.912	0.930	0.956	0.920
18	0.951	0.960	0.998	0.955	0.986	0.941	0.988	0.963
21	0.997	1.000	1.000	0.998	0.993	0.995	0.997	0.994
22	1.000	1.000	1.000	1.000	0.996	0.992	0.997	0.994
23	0.987	0.993	0.999	0.990	0.985	0.985	0.994	0.985
24	0.984	0.973	0.999	0.978	0.939	0.963	0.978	0.950
25	0.993	0.989	0.999	0.991	0.976	0.927	0.978	0.951
26	0.983	0.983	0.999	0.983	0.976	0.933	0.976	0.953
27	0.974	0.952	0.998	0.962	0.912	0.927	0.953	0.917
28	0.937	0.972	0.998	0.954	0.909	0.959	0.976	0.932
31	0.984	0.984	0.999	0.984	0.978	0.970	0.988	0.974
32	0.992	0.982	0.999	0.987	0.976	0.986	0.991	0.981
33	0.992	1.000	1.000	0.996	0.989	0.991	0.998	0.990
34	0.992	0.994	0.999	0.993	0.920	0.995	0.976	0.956
34	0.994	0.981	0.999	0.987	0.998	0.898	0.976	0.946
36	0.983	0.979	0.999	0.981	0.984	0.978	0.991	0.981
37	0.953	0.963	0.998	0.958	0.937	0.929	0.962	0.932
38	0.961	0.959	0.998	0.959	0.980	0.943	0.984	0.961
41	0.979	0.987	0.999	0.983	0.983	0.967	0.988	0.975
42	0.989	0.976	0.999	0.983	0.976	0.988	0.991	0.982
43	0.987	0.997	0.999	0.992	0.989	0.991	0.998	0.990
44	0.997	0.983	0.999	0.990	0.984	0.966	0.988	0.975
45	0.984	1.000	0.999	0.992	0.966	0.978	0.988	0.972
46	0.983	0.979	0.999	0.981	0.964	0.987	0.988	0.975
47	0.944	0.954	0.997	0.949	0.954	0.943	0.970	0.948
48	0.946	0.925	0.997	0.935	0.949	0.946	0.979	0.947
-	0.979	0.978	0.998	0.978	0.966	0.962	0.983	0.963

annotated by dental professionals. The aim was to identify decayed teeth using Convolutional Neural Networks (CNNs). Additionally, the authors tackled a more challenging task—identifying the stage of cavities in the radiographs. The model that yielded the best results in their study was based on Inception-v1 and achieved an accuracy of 0.733. In contrast, our study focuses solely on determining whether a tooth has cavities or not, without considering the severity of the cavities. This approach is taken because panoramic radiographs, as used in our study, are not typically the primary choice for identifying cavi-

ties, especially in their early stages. Panoramic radiographs do not provide the same level of detail as intraoral radiographs, making it challenging to detect early cavities. Consequently, the task proposed in our study aims to offer a supplementary opinion to assist dentists and radiologists in their diagnoses. It can also be beneficial for automatically generating reports and summaries related to the analysis of panoramic radiographs. Furthermore, a tool that can automatically identify cavities may prove valuable for dental and radiology students and professionals, allowing them to compare their analyses with those generated by CNNs, as proposed in (MORAN et al., 2021).

The study conducted by (LEE et al., 2018) demonstrated excellent results for the identification of decayed teeth using periapical radiographs and Inception-v3, achieving an accuracy of 0.890 and a recall of 0.840 for premolar teeth, an accuracy of 0.880 and a recall of 0.920 for molar teeth, and an accuracy of 0.820 and a recall of 0.810 for both premolar and molar teeth. In contrast, this study, also employing Inception-v3, achieved an overall accuracy of 0.910 and a recall of 0.860. Furthermore, with the implementation of the InceptionResNet-v2 architecture, the current study improved upon these results, achieving an accuracy of 0.939 and a recall of 0.914. It's important to note that despite sharing the objective of identifying decayed teeth, this study and Lee et al.'s study exhibit significant differences: (i) the present study utilized panoramic radiographs, while Lee et al. used periapical radiographs; (ii) Lee et al. focused on cavity identification in molar and premolar teeth, while the current study included various types of teeth. Consequently, both studies offer valuable contributions, making direct comparisons challenging. Another complicating factor in such comparisons is the varying stages of decay among the teeth used in each study.

5

Conclusion

The primary aim of this study was to contribute to the development of a modular system capable of detecting the mouth region, locating, segmenting, and numerically identifying teeth, while also providing information about their decay status. This research specifically addressed the tasks of numbering teeth, using the FDI notation, and classifying them as decayed or non-decayed.

To accomplish the task of *assigning FDI numbers to teeth* on panoramic radiographs, two strategies were compared to determine which one could best achieve the desired objectives in terms of precision, recall, accuracy, and F1-score. Strategy 1 was a direct approach, performed in a single step, aiming to train the CNN to differentiate the 32 classes of teeth as per the FDI notation. Conversely, strategy 2 was designed to potentially yield better results by breaking down the tooth numbering task into two steps, following a divide-and-conquer approach inspired by the work of (ZHANG et al., 2018). For both strategies, two well-established network architectures were implemented, chosen from previous research in the field to determine which one would yield superior results.

To train the networks, radiographs from a set of images sourced from the clinical image database of the Faculty of Dentistry of Ribeirão Preto (FORP-USP) were utilized. These images had bounding boxes annotated by experts in radiology and dentistry. The dimensions of these annotated bounding boxes were increased by 300%, following the approach introduced by (KROIS; SCHNEIDER; SCHWENDICKE, 2021). Consequently, each bounding box served as a reference to extract teeth with an expanded area, which were then used as input for training the selected CNNs.

Among the different combinations of strategies and architectures, strategy 1 coupled with the InceptionResNet-v2 architecture produced the most promising results. This combination achieved precision, recall, accuracy, and F1-score values of 0.966, 0.962, 0.983, and 0.963, respectively. The InceptionResNet-v2 architecture, which amalgamates elements from the Inception and ResNet architectures to form a deeper and more efficient CNN, demonstrated its effectiveness in this context. Furthermore, the quality of the

annotations played a crucial role in the outstanding results obtained. The experts meticulously adhered to a well-thought-out annotation protocol. Additionally, the strategy of expanding the bounding boxes, as suggested by (KROIS; SCHNEIDER; SCHWENDICKE, 2021), proved to be valuable, as it considered the context in which a tooth is situated, contributing to its accurate identification. These factors, coupled with other data augmentation techniques and hyperparameter adjustments, further enhanced the overall performance.

For the *decayed teeth identification* task, a combination of datasets was employed to increase the number of decayed teeth. This approach was necessitated by initial tests conducted solely with the FORP-USP dataset, which indicated that the networks struggled to effectively distinguish decayed teeth from non-decayed ones, even with the application of data augmentation techniques. Considering this, a public dataset, sourced from three different institutions and annotated by dental experts, was integrated with the FORP-USP dataset, resulting in a total of 3,652 decayed teeth images. An equal number of non-decayed teeth images were selected from the FORP-USP dataset, totaling 7,304 images, encompassing both decayed and non-decayed teeth.

Once again, the Inception-v3 and InceptionResNet-v2 networks were employed to compare the results obtained using these two architectures. The InceptionResNet-v2 implementation yielded the best results, achieving precision, recall, accuracy, and F1-score values of 0.963, 0.914, 0.939, and 0.937, respectively. These high precision and accuracy values indicate that the results closely align with the actual values and exhibit consistency among themselves, highlighting a high level of trustworthiness and dependability in the measurements. Moreover, the precision scores exceeded recall, implying that the network might not capture all positive cases in the dataset, but when it identifies something as positive, it is generally accurate. The results demonstrate promising potential for utilizing the InceptionResNet-v2 architecture in tooth numbering and decayed tooth identification tasks. However, cautious consideration is necessary when contemplating its practical application in the dental and radiology fields.

In the context of tooth numbering, a limitation arises from the nature and quantity of images employed. The training and testing phases were constrained by a relatively small number of annotated images, especially in terms of dataset balance. Although different data augmentation techniques were implemented and positively impacted the results, a larger image dataset could enhance the reliability of outcomes. Moreover, the reliance on images from a single source, generated by the same panoramic X-ray machine, restricts the representation to a specific community. The scarcity of publicly accessible, high-quality datasets poses challenges for future research endeavors in this domain (SCHWENDICKE; SAMEK; KROIS, 2020; CARNEIRO et al., 2023b). This limitation impedes the development of widely applicable models for dental training and testing. With the goal of aiding

in the improvement of this situation, the dataset utilized in this study will be made accessible to the academic community. Future research can broaden the scope to encompass a wider array of dental conditions, including broken or extensively restored teeth, as well as dental artifacts like appliances, bridges, pins, and other elements that were not considered in the current study. By doing so, it is possible to better represent the diverse range of individuals commonly encountered in dental practice, leading to more comprehensive and applicable results. Future research can broaden the scope to encompass a wider array of dental conditions, including broken or extensively restored teeth, as well as dental artifacts like appliances, bridges, pins, and other elements that were not considered in the current study. Incorporating a greater variety of dental conditions and artifacts aligns research outcomes more closely with real-world clinical scenarios. This increased relevance can facilitate the translation of research findings into practical applications in dental practice.Researchers can explore new avenues of investigation and innovation when armed with a more diverse dataset. This can lead to breakthroughs in dental technology, materials, and treatment methodologies.

Concerning the task of decayed tooth identification, additional limitations should be acknowledged. Initially, while the number of images for ”decayed” and ”non-decayed” categories was equivalent, they did not encompass a balanced representation of all tooth types. Furthermore, the datasets exhibited varying degrees of tooth decay, further compromising the quality of predictive generalization. For more robust generalization, the dataset should ideally encompass each of the 32 tooth types in close to equal proportions. Future research endeavors should aim to utilize datasets that exhibit a more balanced distribution in terms of the number of decayed teeth for each specific type of tooth. Building upon the approach taken in the (MORAN et al., 2021) study, there should be an effort to assess the severity of cavities. This evaluation can greatly assist in recommending the most suitable and personalized treatment options for individual patients, thereby improving the quality of dental care and patient outcomes.

In both the tasks of numbering teeth and identifying decayed teeth, the application of image preprocessing techniques emerges as a valuable strategy for enhancing future work. These techniques offer a range of benefits, including noise reduction, smoothing, contrast enhancement, and the normalization and standardization of datasets. By employing such techniques, the quality and consistency of images can be significantly improved. This, in turn, allows for the accentuation of important characteristics within the images, thereby facilitating seamless comparison and analysis across different datasets(GONZALEZ; WOODS, 2008; PRATT; JR., 2007; JAIN, 1989).

Concerning the tasks within the modular system framework, developed with José Andery Carneiro (CARNEIRO et al., 2023a), some considerations arise. The modular structure and task integration facilitate the inclusion of additional modules, each con-

tributing unique functionality to the system. For instance, it is feasible to develop and integrate a module dedicated to detecting restorations in segmented teeth. Similarly, another module could be introduced to identify impacted teeth or the number of teeth with root canal treatment. This adaptable framework accommodates the utilization of different network architectures, allowing for the selection of the most suitable model for each task, whether it pertains to the tasks proposed in this research or those integrated into the system. Moreover, the system holds potential to support other research initiatives aimed at automating extensive dataset analysis, report generation, and dental record population. This could offer dentists more time for clinical treatments (CHEN et al., 2022). Additionally, the system can serve as a valuable second opinion, mitigating errors stemming from factors such as inexperience, fatigue, stress, subjectivity, and more. It's worth noting that applying CNNs to automate dental radiographs analysis can assist less experienced students and professionals in Radiology and Dentistry by enabling them to compare their analyses with network-generated predictions, thereby facilitating learning and enhancing their professional capabilities.

Bibliography

ABDALLA-ASLAN, R. et al. An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, v. 130, 06 2020.

ABDINIAN, M. et al. Accuracy of digital bitewing radiography versus different views of digital panoramic radiography for detection of proximal caries. *J Dent (Tehran)*, v. 12, n. 4, p. 290–297, 2015.

AFFAIRS, A. D. A. C. on S. The use of dental radiographs: Update and recommendations. *The Journal of the American Dental Association*, v. 137, n. 9, p. 1304–1312, 2006. ISSN 0002-8177. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0002817714643221>.

AKKAYA, N. et al. Comparing the accuracy of panoramic and intraoral radiography in the diagnosis of proximal caries. *Dentomaxillofac Radiol*, v. 35, n. 3, p. 170–174, 2006.

ALZUBAIDI, L. et al. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, v. 8, 03 2021.

AYSE, B. Human identification with dental panoramic radiographic images. *IET Biometrics*, v. 7, n. 4, p. 349–355, 2018.

BA, K.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. v. 2, 01 2007.

BAYDAR, O. et al. The u-net approaches to evaluation of dental bite-wing radiographs: An artificial intelligence study. *Diagnostics*, v. 13, p. 453, 01 2023.

BOOSHEHRY, M. Z. et al. Dental caries diagnostic methods. *Avicenna J Dent Res*, Hamadan University of Medical Sciences, v. 2, n. 1, p. 1–12, 2010. ISSN 20087659. Disponível em: <http://ajdr.umsha.ac.ir/Article/ajdr-20>.

BOUCHAHMA, M. et al. An automatic dental decay treatment prediction using a deep convolutional neural network on x-ray images. In: *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*. [S.l.: s.n.], 2019. p. 1–4.

BOUREAU, Y.-L.; PONCE, J.; LECUN, Y. A theoretical analysis of feature pooling in visual recognition. In: . [S.l.: s.n.], 2010. p. 111–118.

BRASIL, M. d. S. *Sb brazil 2010: Pesquisa nacional de saúde bucal: resultados principais*. [S.l.]: Ministério da Saúde, Secretaria de Atenção à Saúde Secretaria de Vigilância em Saúde, 2012.

CAI, Z.; VASCONCELOS, N. *Cascade R-CNN: Delving into High Quality Object Detection*. 2017.

CAMPESATO, O. *Artificial Intelligence, Machine Learning, and Deep Learning*. Mercury Learning & Information, 2020. ISBN 9781683924661. Disponível em: <https://books.google.com.br/books?id=pqnNDwAAQBAJ>.

CARNEIRO, J. A. et al. A deep learning-powered modular segmentation approach on dental panoramic radiograph images. *Submited to Computer Methods and Programs in Biomedicine*, p. 34, 09 2023.

CARNEIRO, J. A. et al. Deep learning to detect and classify teeth, dental caries and restorations: A systematic mapping. *Submitted to Dentomaxillofacial Radiology (DMFR-D-23-00388)*, 2023.

CHEN, S.-L. et al. Missing teeth and restoration detection using dental panoramic radiography based on transfer learning with cnns. *IEEE Access*, v. 10, p. 118654–118664, 2022.

CHIN, C.-L. et al. Dentition labeling and root canal recognition using ganand rule-based system. In: *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*. [S.l.: s.n.], 2019. p. 1–6.

CHOLLET, F. *Deep Learning with Python*. [S.l.]: Manning, 2017. ISBN 9781617294433.

CLIFTON, T.; TYNDALL, D.; LUDLOW, J. Extraoral radiographic imaging of primary caries. *Dento maxillo facial radiology*, v. 27, p. 193–8, 08 1998.

COSTA, E. D. et al. Development of a dental digital dataset for research in artificial intelligence: importance of labeling by radiologists. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology (Accepted with minor revision)*, 2023.

DU, X. et al. A convolutional neural network based auto-positioning method for dental arch in rotational panoramic radiography. In: . [S.l.: s.n.], 2018. v. 2018, p. 2615–2618.

FAWCETT, T.; PROVOST, F. *Data Science for Business*. [S.l.: s.n.], 2013.

FERNANDES, A. F. A.; DóREA, J. R. R.; ROSA, G. J. d. M. Image analysis and computer vision applications in animal sciences: An overview. *Frontiers in Veterinary Science*, v. 7, p. 800, 2020. ISSN 2297-1769. Disponível em: <https://www.frontiersin.org/article/10.3389/fvets.2020.551269>.

FRENCKEN, J. E. et al. Global epidemiology of dental caries and severe periodontitis - a comprehensive review. *J Clin Periodontol*, v. 44, n. 18, p. 94–105, 2017.

FUKUDA, M.; INAMOTO, K.; SHIBATA, N. Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiol*, v. 7, 2019.

GHOSH, A. et al. Fundamental concepts of convolutional neural network. In: _____. [S.l.: s.n.], 2020. p. 519–567. ISBN 978-3-030-32643-2.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN 9780131687288 013168728X 9780135052679 013505267X. Disponível em: <http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>.

GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. <http://www.deeplearningbook.org>.

HE, K. et al. Mask r-cnn. In: . [S.l.: s.n.], 2017. p. 2980–2988.

HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778.

HEATON, J. Ian goodfellow, yoshua bengio, and aaron courville: Deep learning. *Genetic Programming and Evolvable Machines*, v. 19, p. 305–307, 2017.

HOSNY, A. et al. Artificial intelligence in radiology. *Nature Reviews Cancer*, v. 18, 05 2018.

IBM. *What is computer vision?* 2008. <https://www.ibm.com/topics/computer-vision>. [Online; accessed 18-October-2021].

ISMAIL, A. Visual and visuo-tactile detection of dental caries. *Journal of Dental Research*, v. 83, n. 1_suppl, p. 56–66, 2004. Disponível em: <https://doi.org/10.1177/154405910408301s12>.

JADER, G.; OLIVEIRA, L.; PITHON, M. Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications*, 2018.

JAE-HONG, L. et al. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *Journal of periodontal amp; implant science*, v. 48, n. 2, p. 114–123, 2018.

JAIN, A. K. *Fundamentals of digital image processing*. USA: Prentice-Hall, Inc., 1989. ISBN 0133361659.

KAMBUROGLU, K. et al. Proximal caries detection accuracy using intraoral bitewing radiography, extraoral bitewing radiography and panoramic radiography. *Dentomaxillofac Radiol*, v. 41, n. 6, p. 450–459, 2012.

KIRILLOV, A. et al. Pointrend: Image segmentation as rendering. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 9796–9805, 2019. Disponível em: <https://api.semanticscholar.org/CorpusID:209386851>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. v. 25. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.

KROIS, J.; SCHNEIDER, L.; SCHWENDICKE, F. Impact of image context on deep learning for classification of teeth on radiographs. *Journal of Clinical Medicine*, v. 10, p. 1635, 04 2021.

KUMAR, A.; BHADAURIA, H.; SINGH, A. Descriptive analysis of dental x-ray images using various practical methods: A review. *PeerJ Computer Science*, v. 7, p. e620, 09 2021.

KUMAR, R.; KHAMBETE, N.; PRIYA, E. Extraoral periapical radiography: an alternative approach to intraoral periapical radiography. *Imaging Science in Dentistry*, v. 41, p. 161 – 165, 2011.

KUNZ, F. et al. Artificial intelligence in orthodontics: Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. *J Orofac Orthop.*, v. 81, n. 1, p. 52–68, 2020.

LAUDENBACH, J. M.; SIMON, Z. Common dental and periodontal diseases: Evaluation and management. *Medical Clinics of North America*, v. 98, n. 6, p. 1239–1260, 2014. ISSN 0025-7125. Oral Medicine: A Handbook for Physicians. Disponível em: <https://www.sciencedirect.com/science/article/pii/S002571251400128X>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–444, 2015.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LEE, J. et al. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent*, 2018.

LIANG, M.; HU, X. Recurrent convolutional neural network for object recognition. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 3367–3375.

LIN, S.-Y.; CHANG, H.-Y. Tooth numbering and condition recognition on dental panoramic radiograph images using cnns. *IEEE Access*, v. 9, p. 166008–166026, 2021.

LIN, T.-Y. et al. *Focal Loss for Dense Object Detection*. arXiv, 2017. Disponível em: <https://arxiv.org/abs/1708.02002>.

LIU, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the consort-ai extension. *Nature medicine*, v. 26, p. 1364–1374, 09 2020.

LOURIDAS, P.; EBERT, C. Machine learning. *IEEE Softw.*, v. 33, n. 5, p. 110–115, 2016. Disponível em: <https://doi.org/10.1109/MS.2016.114>.

MARK, A. M. Dental x-rays. *The Journal of the American Dental Association*, v. 151, n. 6R, 2020.

MITCHELL, T. *Machine Learning*. [S.l.]: McGraw-Hill Education, 1997.

MONGAN, J.; MOY, L.; JR, C. E. K. *Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers*. 2020. <https://pubs.rsna.org/page/ai/claim>. [Online; accessed 21-January-2024].

MORAN, M. et al. Classification of approximal caries in bitewing radiographs using convolutional neural networks. *Sensors*, v. 21, 07 2021.

MOREIRA, R. da S. Epidemiology of dental caries in the world. In: VIRDI, M. S. (Ed.). *Oral Health Care*. Rijeka: IntechOpen, 2012. cap. 8. Disponível em: <https://doi.org/10.5772/31951>.

MURAMATSU, C. et al. Tooth detection and classification on panoramic radiographs for automatic dental chart filing: improved classification by multi-sized input data. *Oral Radiology*, v. 37, p. 1–7, 01 2020.

OLSEN, G. et al. An image-processing enabled dental caries detection system. *2009 ICME International Conference on Complex Medical Engineering, CME 2009*, 04 2009.

OLVERES, J. et al. What is new in computer vision and artificial intelligence in medical image analysis applications. *Quantitative Imaging in Medicine and Surgery*, v. 11, n. 8, 2021. ISSN 2223-4306. Disponível em: <https://qims.amegroups.com/article/view/70834>.

OSTERLOH, D.; VIRIRI, S. Unsupervised caries detection in non-standardized bitewing dental x-rays. In: BEBIS, G. et al. (Ed.). *Advances in Visual Computing*. Cham: Springer International Publishing, 2016. p. 649–658. ISBN 978-3-319-50835-1.

PADALIA, D.; VORA, K.; SHARMA, D. An attention u-net for semantic segmentation of dental panoramic x-ray images. In: *2022 5th International Conference on Advances in Science and Technology (ICAST)*. [S.l.: s.n.], 2022. p. 491–496.

PETERSEN, P. E. The world oral health report 2003: continuous improvement of oral health in the 21st century – the approach of the who global oral health programme. *Community Dentistry and Oral Epidemiology*, v. 31, n. s1, p. 3–24, 2003. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j..2003.com122.x>.

PETERSEN, P. E.; BAEZ, R. J.; ORGANIZATION, W. H. Publications. *Oral health surveys: basic methods*. 5th ed. ed. [S.l.]: World Health Organization, 2013. vii, 125 p. p.

POWERS, D. Evaluation: From precision, recall and f-factor to roc, informedness, markedness correlation. *Mach. Learn. Technol.*, v. 2, 01 2008.

PRATT, W. K.; JR., J. E. A. Digital image processing, 4th edition. *J. Electronic Imaging*, v. 16, n. 2, p. 029901, 2007. Disponível em: <http://dblp.uni-trier.de/db/journals/jei/jei16.html#PrattA07>.

PRINCE, S. J. D. *Computer Vision: Models, Learning, and Inference*. 1st. ed. USA: Cambridge University Press, 2012. ISBN 1107011795.

Ramana Kumari, A.; Nagaraja Rao, S.; Ramana Reddy, P. Design of hybrid dental caries segmentation and caries detection with meta-heuristic-based resnext-rnn. *Biomedical Signal Processing and Control*, v. 78, p. 103961, 2022. ISSN 1746-8094. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1746809422004608>.

REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 6, p. 1137–1149, 2017.

RUNGTA, K. *TensorFlow in 1 Day: Make your own Neural Network*. PublishDrive, 2019. Disponível em: <https://books.google.com.br/books?id=Hb7jDwAAQBAJ>.

RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3. ed. [S.l.]: Prentice Hall, 2010.

SALUNKE, D. et al. Customized convolutional neural network to detect dental caries from radiovisiography(rvg) images. *International Journal of Advanced Technology and Engineering Exploration*, v. 9, p. 827–838, 06 2022.

SCHERER, D.; MüLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. In: . [S.l.: s.n.], 2010. p. 92–101. ISBN 978-3-642-15824-7.

SCHMIDT-ERFURTH, U. et al. Artificial intelligence in retina. *Progress in Retinal and Eye Research*, v. 67, p. 1–29, 2018. ISSN 1350-9462. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1350946218300119>.

SCHULZ, K.; ALTMAN, D.; MOHER, D. Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine*, v. 8, p. 18, 03 2010.

SCHWENDICKE, F. et al. Convolutional neural networks for dental image diagnostics: A scoping review. *Journal of Dentistry*, v. 91, p. 103226, 11 2019.

SCHWENDICKE, F.; SAMEK, W.; KROIS, J. Artificial intelligence in dentistry: Chances and challenges. *Journal of Dental Research*, v. 99, n. 7, p. 769–774, 2020. PMID: 32315260.

SCHWENDICKE, F. et al. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *Journal of Dentistry*, v. 107, p. 103610, 2021. ISSN 0300-5712.

SELWITZ, R. H.; ISMAIL, A. I.; PITTS, N. B. Dental caries. *The Lancet*, v. 369, n. 9555, p. 51–59, 2007. ISSN 0140-6736. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0140673607600312>.

SHENG, C. et al. Transformer-based deep learning network for tooth segmentation on panoramic radiographs. *Journal of systems science and complexity*, p. 1—16, October 2022. ISSN 1009-6124. Disponível em: <https://europepmc.org/articles/PMC9976655>.

SHIMIZU, H.; NAKAYAMA, K. Artificial intelligence in oncology. *Cancer Science*, v. 111, 03 2020.

SHRESTHA, A.; MAHMOOD, A. Review of deep learning algorithms and architectures. *IEEE Access*, v. 7, p. 53040–53065, 2019.

SILVA, G.; OLIVEIRA, L.; PITHON, M. Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications*, v. 107, p. 15–31, 2018. ISSN 0957-4174.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.

SINGH, P.; SEHGAL, P. Automated caries detection based on radon transformation and dct. In: . [S.l.: s.n.], 2017. p. 1–6.

SINGH, P.; SEHGAL, P. G.v black dental caries classification and preparation technique using optimal cnn-lstm classifier. *Multimedia Tools and Applications*, v. 80, p. 1–18, 02 2021.

SZEGEDY, C. et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. [S.l.]: AAAI Press, 2017. p. 4278–4284.

SZEGEDY, C. et al. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 1–9.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: . [S.l.: s.n.], 2016.

TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, v. 15, 2015. Disponível em: <https://api.semanticscholar.org/CorpusID:18182893>.

TAO, A.; BARKER, J.; SARATHY, S. *DetectNet: Deep Neural Network for Object Detection in DIGITS*. 2016. <https://developer.nvidia.com/blog/detectnet-deep-neural-network-object-detection-digits>. [Online; accessed 21-January-2024].

TAYE, M. M. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, v. 11, n. 3, 2023. ISSN 2079-3197. Disponível em: <https://www.mdpi.com/2079-3197/11/3/52>.

WAN, H. Deep learning:neural network, optimizing method and libraries review. *2019 International Conference on Robots & Intelligent System (ICRIS), Haikou, China*, p. 497–500, 2019.

WANG, C.; HUANG, C.; LEE, J. A benchmark for comparison of dental radiography analysis algorithms. *Med Image Anal*, v. 31, p. 63–76, 2016.

WANG, C.-W. et al. A benchmark for comparison of dental radiography analysis algorithms. *Medical Image Analysis*, v. 31, p. 63–76, 2016. ISSN 1361-8415.

WHAITES, E. *Essentials of dental radiography and radiology*. [S.l.]: Churchill Livingstone, 2002.

ZHANG, K. et al. An effective teeth recognition method using label tree with cascade network structure. *Computerized Medical Imaging and Graphics*, v. 68, p. 61–70, 2018. ISSN 0895-6111.

ZHOU, L.-Q. et al. Artificial intelligence in medical imaging of the liver. *World Journal of Gastroenterology*, v. 25, p. 672 – 682, 2019.