Universidad Politécnica
de Madrid

Escuela Técnica Superior de
Ingenieros Informáticos

Master of Science in Digital Innovation.
Specialization in Health and Medical Data Analytics.

Master Thesis

A Multimodal Deep Learning Classifier for Dental Implant Model and Diameter Identification in Periapical Radiographs

Author: Paula Gisbert Llorca
Supervisor: Raúl Alonso Calvo
Co-supervisor: Mónica Bonfanti Gris

Madrid, July 2025

Este Trabajo Fin de Máster se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Master Thesis
Master of Science in Digital Innovation: Health and Medical Data Analytics.
Title: A Multimodal Deep Learning Classifier for Dental Implant Model and Diameter Identification in Periapical Radiographs
July 2025

Author: Paula Gisbert Llorca

Supervisor:
Raúl Alonso Calvo
ETSI Informáticos
Universidad Politécnica de Madrid

Co-supervisor:
Mónica Bonfanti Gris

Resumen

La identificación correcta de implantes dentales (DIS) en la práctica clínica se convierte en un reto cuando se producen complicaciones y no se dispone de los registros dentales del paciente. Actualmente, los odontólogos reconocen los implantes basándose en la comparación con catálogos comerciales y/o bases de datos en línea, lo cual consume mucho tiempo y es propenso a errores. La necesidad de métodos automáticos para el reconocimiento de implantes es demandada por clínicos y pacientes que se han enfrentado a esta situación.

Por lo tanto, el objetivo de este trabajo de fin de máster es evaluar la aplicación de enfoques de aprendizaje profundo multimodal para clasificar el modelo de implante (3 clases) y el diámetro (8 clases) utilizando radiografías periapicales. Los modelos se entrenaron con 914 radiografías periapicales de pacientes adultos tomadas en la misma clínica dental, que contienen un total de 1.426 implantes dentales anotados que se utilizaron para el entrenamiento y la evaluación. Para cada implante se dispone de máscaras de segmentación y etiquetas de referencia. Además, se realizó un paso de extracción de características mediante técnicas clásicas de procesamiento de imágenes para identificar puntos clave anatómicos, los cuales se estructuraron posteriormente en representaciones basadas en grafos. Estos puntos, junto con los datos de imagen y las características tabulares derivadas de las imágenes, se emplearon para entrenar y evaluar cuatro modelos de clasificación.

Tras aplicar técnicas de aumento de datos, se evaluó el rendimiento de un modelo unimodal y de cuatro arquitecturas multimodales que combinan información de imagen, datos tabulares y estructuras de grafos. Los modelos incluidos fueron: EfficientNet-B2, EfficientNet-B2 combinado con TabNet, EfficientNet-B2 integrado con una red neuronal de grafos (GNN) y ResNet-18 combinado con GNN. Los resultados experimentales muestran que, entre todos los modelos, la combinación de EfficientNet con GNN supera a las demás arquitecturas evaluadas en términos de exactitud, precisión, sensibilidad e índice F1. Este modelo alcanzó una precisión global de 93%, sensibilidad de 92%, precisión de 92% e índice F1 de 92%. A pesar de las limitaciones inherentes al estudio, la arquitectura ha mostrado un rendimiento superior en la clasificación del modelo y el diámetro, mientras que ha demostrado la eficacia de combinar modalidades de datos complementarias para una identificación robusta de implantes.

Abstract

Accurate identification of dental implant systems (DISs) in clinical practice becomes challenging in cases in which complications occur, and dental records are unavailable. Currently, dental practitioners recognize implants based on observation of commercial catalogues and/or online databases, which is time-consuming and prone to error. The need for automatic methods for implant recognition is understood by clinicians and patients who have encountered this situation.

Therefore, the aim of this master thesis is to evaluate the application of multimodal deep learning approaches to classify DISs model (3 classes) and diameter (8 classes) using periapical radiographs. The models were trained on 914 periapical X-ray images from adult patients collected in the same dental clinic, containing a total of 1,426 annotated dental implants that were used for training and evaluation. Segmentation masks and ground-truth labels were available for each implant. In addition, a feature extraction step was performed using classic image processing techniques to identify relevant anatomical keypoints, which were later structured into graph-based representations. These keypoints, along with image data and tabular features derived from the images were used to train and evaluate four classification models.

After applying data augmentation, the performance of a single-modal model and four multimodal architectures combining image, tabular and structural information was evaluated. The models included EfficientNet-B2, EfficientNet-B2 paired with TabNet, EfficientNet-B2 integrated with a Graph Neural Network (GNN), and ResNet-18 combined with GNN. The experimental results show that among all models, the combination of EfficientNet with GNN outperforms the other evaluated architectures in terms of accuracy, precision, recall and F1-Score. This model achieved an accuracy of 93%, precision of 92%, recall of 92% and F1-score of 92%. Even though the limitations inherent to the study, the architecture showed superior performance in classifying DIS model and diameter, while demonstrating the effectiveness of combining complementary data modalities for robust implant identification.

Content

Figure Index

Table Index

1 Introduction

1.1 Motivation

Radiographic imaging, also referred to as X-ray imaging, is widely regarded as a cornerstone of modern dental practice. This technique, which involves capturing images of dental structures using electromagnetic radiation, provides essential information that is not possible to obtain by clinical examination alone, as it enables both a proper diagnosis and a personalized treatment plan. Dental radiographs facilitate the visualisation of hard tissues, such as teeth and bone structures, enabling precise diagnostic imaging of pathologies, assessment of anatomical relationships, and guidance of surgical procedures with a high precision degree [1].

A variety of radiographic techniques are habitually employed in the field of dentistry, each with a distinct set of applications. Intraoral radiographs, including bitewing and periapical images, are the most employed and provide detailed views of individual teeth and adjacent structures. While extraoral techniques, such as panoramic radiographs and 3-dimensional cone-beam computed tomography (CBCT), are used to provide broader views [2]. Of these, periapical (PA) radiographs stand out due to their capacity to produce high-resolution and detailed images of a tooth and its surrounding bone, thus making them useful for the detection of periapical lesions, the assessment of bone structure, and the evaluation of implant integration [3].

Today, dental implants represent a standard treatment modality for the replacement of missing teeth [4]. Their reliability and widespread adoption have led to a proliferation of implant systems, each with unique designs, materials, and dimensions. This variety aims to cover different clinical scenarios and patient needs, improving the versatility and applicability of implant-based treatments [5].

In many clinical situations, especially with patients who have complications (e.g. broken dental prostheses), it becomes necessary to identify the implant system to facilitate the replacement of the component, or the provision of new prosthetics that are compatible with the existing implant. In the latter scenario, determining the implant model and manufacturer is of vital importance, as it dictates the both the connection type and screwdriver to use.

Nevertheless, when dental records are inaccessible due to factors such as the closure of dental practices or their relocation, the identification of the implant often depends on the analysis of information visible in dental radiographs. Dental practitioners often rely on visual inspection, comparison with manufacturer catalogues and personal experience, but this process is often inconsistent, time-consuming and prone to variability - especially with unfamiliar brands and unexperienced professionals [6]. This challenge is amplified by the fact that many implants share similar shapes and features in radiographs, making them difficult to differentiate [7].

This provides the potential for technological support through artificial intelligence (AI), with deep learning-based methods playing a significant role [8]. The automation of the classification of dental implants employing radiographic images promises numerous benefits. Specifically, it has the potential to reduce the time required for identification, improve consistency and accuracy, and support clinicians in making informed decisions [9].

In this context, the development of models that are capable of reliably identifying implant brands and dimensions in radiographs is of particular significance. This advancement would result in an augmentation of both the efficacy and safety of the implantology practice. This thesis contributes to this aim by exploring the use of multimodal deep learning methods for automatic implant classification based on brand and size in periapical radiographs.

1.2 Objectives

The main objective of this master thesis is to develop a dental implant classification system that can predict both the model and diameter of implants from periapical radiographs, using multimodal deep learning approaches.
In order to achieve this general objective, the following specific objectives are proposed:

- Prepare a clinical dataset consisting of periapical radiographs with their respective segmentation masks and annotations.
- Extract key anatomical points using classical image processing techniques to represent the implant geometry and transform it into a graphical structure useful for training graph-based models.
- Build a multimodal dataset by combining the original image, graph-structured anatomical information, and tabular variables derived from image information.
- Design, train, and compare various multimodal deep learning architectures that integrate different combinations of input data (including keypoint-based graphs, tabular features, and masked radiographic images) as well as evaluate single-modality models to establish comparative baselines.
- Evaluate the performance of the models in the task of model and implant diameter classification using metrics such as precision, recall, F1 score, and overall accuracy.
- Make an analysis of the impact of combining different types of information (visual, structural, and tabular) on system performance, highlighting the advantages or disadvantages of a multimodal approach over a unimodal one.

1.3 Thesis Structure

This master thesis is organized as follows:
Chapter 2 provides an overview of the background in which the work is situated, giving relevant context of both dentistry and deep learning fields. Furthermore, it presents the current state of the art of deep learning-based applications for implant recognition in radiographs.
Chapter 3 describes the methodology that has been carried out in this master thesis, from the initial image feature extraction to the final deep learning architectures implemented.
Chapter 4 reports the performance obtained with the implemented classifiers: EfficientNetB2, EfficientNetB2 + GNN, ResNet-18 + GNN and EfficientNet + TabNet.
Chapter 5 analyses the results reported in the previous chapter, identifying the main strengths and limitations of the work.
Chapter 6 presents the conclusions drawn from the entire body of work, while identifying potential areas for future development.

2 State of the art

2.1 Dental Imaging and Periapical Radiographs

Dental imaging is defined as the process of creating digital, print or film representations of anatomic structures for the purpose of diagnosis. In this context, dental radiography is considered an integral part of contemporary clinical dentistry, involving the production of radiographs of teeth and adjacent structures by the exposure of an image receptor to X-rays [10]. Dental images have been shown to be of great value for dental clinicians, as they facilitate the identification of conditions that could not be detectable with a clinical examination alone [11].

The discovery of X-rays by Röntgen in 1895 revolutionized diagnosis by allowing the visualization of internal structures of the human body. This breakthrough was embraced in early 1896 in the dental field by Walkoff, who produced the first dental radiograph [12]. These early adopters made the foundation for modern dental imaging and, over time, both technological advancements and a better understanding of radiation safety, turned dental radiography into an indispensable diagnostic tool in dental practice.

Traditionally, dental radiographs were produced by the X-rays passing through an object (the patient) and interacting with the photographic emulsion on a film, resulting in a blackening of the film. Nowadays, film has been replaced by digital sensors that facilitate the creation of the image in a computer. This development has made digital dental radiography a better option for general practitioners, offering a speeding process, along with a reduction in radiation exposure to the patient [13].

In diagnostic radiology, an X-ray beam of energy has the power to penetrate substances and record image shadows on photographic film. The degree of darkening in an X-ray image is determined by the number of X-ray photons that reach the detector, which is influenced by the attenuation properties of the tissues they traverse. Denser tissues, such as bone, have higher atomic numbers and greater mass density, leading to increased attenuation of X-rays. As a consequence, fewer photons reach the detector in these areas, rendering them lighter or white on the radiograph. In contrast, less dense tissues, like muscle or fat, attenuate fewer X-rays, allowing more photons to reach the detector and resulting in darker image regions [14]. The final image obtained is a two-dimensional representation of a three-dimensional object, made up of a variety of black, white and grey superimposed shadows [11].

Radiographic techniques can be classified into intraoral and extraoral modalities, each serving specific diagnostic purposes.

Intraoral radiography is the most common type of 2D radiography. It is obtained by placing the image receptor inside the oral cavity of the patient, providing high-resolution images of teeth and surrounding structures [15]. The intraoral imaging examination requires the use of intraoral receptors, the most commonly used being solid-state charge-coupled device (CCD), complementary metal-oxide semiconductor (CMOS), and photostimulable phosphor (PSP) [16]. The three primary types of intraoral radiography include:

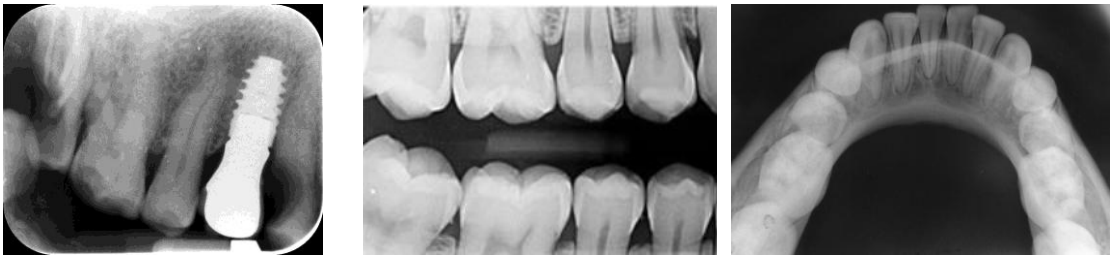
- **Periapical radiographs (Figure 1.a)** show the entire tooth, from the crown to the apex, including all the surrounding bone. These are

instrumental in diagnosing periapical pathologies, assessing root morphology, and evaluating endodontic treatments, among other applications. [10]

- **Bitewing radiographs (Figure 1.b)** display the crown of the upper and lower teeth on a single image, allowing examination of the adjacent tooth surfaces and crestal bone. They are used in interproximal examination (e.g. detection of interproximal caries) and assessment of alveolar bone levels. [17]
- **Occlusal radiographs (Figure 1.c)** provide a view of a larger area of the maxilla or mandible, over the occlusal plane. They aid in the detection of impacted teeth, jaw fractures, and pathological lesions. [10]

Extraoral radiographic techniques involve positioning the imaging receptor outside the mouth, allowing the examination of larger anatomical areas of skull and jaws. Extraoral films are more common in specialty practices, such as orthodontic, paediatric, and oral surgery practices [18]. The three main modalities are:

- **Panoramic radiographs (Figure 1.d)** provide a two-dimensional image of the maxilla and the mandible on a single projection, including both dental arches, temporomandibular joints, and adjacent structures. It is particularly beneficial for assessing tooth development, impacted teeth, and maxillofacial anomalies. [10], [18]
- **Cephalometric radiographs (Figure 1.e)** are sagittal projections of the skull that include both the hard and soft tissues. These are applied in orthodontics and oral surgical procedures to evaluate craniofacial morphology and growing patterns, facilitating treatment planning and outcome assessment [18], [19].
- **Cone Beam Computed Tomography (CBCT) (Figure 1.e)** offers three-dimensional imaging with high spatial resolution. A CBCT X-ray unit generates a beam of X-rays (in the shape of a cone) that rotates around the patient's head, while the image detector moves on the opposite side of the skull in a synchronously. Since its introduction, it has been considered as a mandatory examination to carry out when assessing implant planning, root canal treatments, or maxillofacial pathologies [19].



a b [20] c [21]

d [20] e [22] f [22]

*Figure **1**. Dental X-ray Modalities. **a**, periapical. **b**, bite-wing. **c**, occlusal. **d**, panoramic. **e**, cephalometric. **f**, cbct*

In this study, periapical radiographs (PA) are employed due to their high resolution, accessibility, and cost-effectiveness, but most importantly, because they offer detailed information regarding the shape and dimension of the patients' implants.

PA radiographs are routinely taken in the phases that are covered on dental implantology [23]:

- **Pre-surgical Phase:** They assist in the evaluation of implant placement, as they offer an optimal analysis of the residual alveolar bone and adjacent anatomical structures.
- **Surgical Phase:** While the implant is being placed, periapical radiographs help verify the position, depth, and angulation of the implant. This ensures an accurate implant placementeand alignment.
- **Post-prosthetic Phase:** PA radiographs are taken to follow the appropriate integration of the implant with the surrounding bone and to detect if there is any complication (e.g. peri-implant radiolucencies or mechanical failures).

As specified, periapical radiographs enable professionals to detect bony defects surrounding dental implants, making early implant-related pathology identification possible [24]. Nevertheless, even if this can be possible when the radiograph is acquired under ideal conditions (paralleling technique), misdiagnosis often occur due to unproper acquisition methods [25]. Thus, paralleling technique should always be used in this case, as it minimizes geometric distortion and provides an accurate representation of implant dimensions. [25], [26], [27]

In order to do so, positioning and stabilization devices can be found to enhance image acquisition consistency and reproducibility across different sessions, which is advantageous for longitudinal studies centred on monitoring changes in teeth or implants over time.

Despite their advantages, periapical radiographs also show some limitations: Their two-dimensional nature restricts the assessment of the spatial

relationships of implant placement, potentially affecting the characterization of implant geometry and orientation. Besides this, and as stated before, inadequate positioning or movement can also lead to image distortion, affecting the reliability of implant evaluations [26].

2.2 Dental Implant Systems (DISs)

Dental Implant Systems (DISs) are prosthetic devices made of alloplastic materials used for the replacement of missing teeth. DISs have been widespread adopted over the last decades due to significant developments in terms of implant design, materials and components.

Consequently, they have become a standard treatment, well-known for its endurance, practicality, and capacity to maintain oral health through osteointegration [28], [29] - which is considered essential for the long-term stability of the treatment, being defined as a biologic process in which a direct structural and functional connection between living, haversian bone, and the loaded implant surface is formed [28].

Furthermore, dental implantology has improved significantly the overall quality of life of patients with single, partial or complete edentulism - as showed by several studies that resulted with patients' enhanced masticatory function, aesthetics, and psychological well-being after implant treatment [30], [31]. However, dental implants are not free to suffer complications. Some of the most frequently found ones in the daily practice are both biological (peri-implantitis) and mechanical (prosthetic complications) [32].

A DIS is composed of three elements, which are illustrated in **Figure 2**:

- **Fixture or Implant:** It serves as a restoration's foundation or anchor. It provides a fixed platform for the screwing of an abutment (prosthesis). By growing around the implant, bone tissue can reinforce and regenerate the jaw, preventing the bone loss that comes with losing natural teeth. Titanium is typically used in its construction to avoid foreign body reactions.
- **Abutment**: It gives the crown a foundation to be placed on. Additionally, it serves as the connector piece between the implant and the crown. They are mostly made of titanium and zirconia. The lugs shaped on the stem of the abutment control rotation (twist). To create gradual steps, these lugs limit the abutments' rotational positioning. It is fixed to the implant by a screw.
- **Crown**: It is the upper portion of a restoration. To give a biting surface and a pleasing appearance, they are handmade by a technician. The crown's supporting substructure can be machined (on-site or off-site) or handcrafted. The finished crown is attached to an abutment by screwing or cementing. Implant crowns are made different materials such as metal, porcelain fused to metal (PFM), zirconia and ceramics like lithium disilicate (eMax).

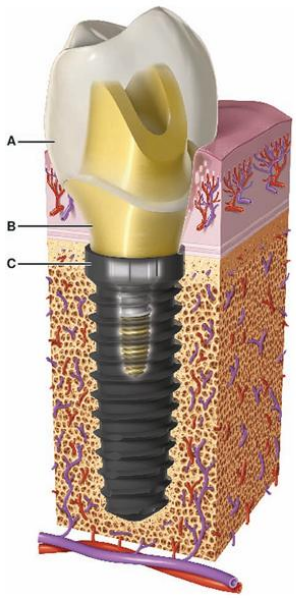
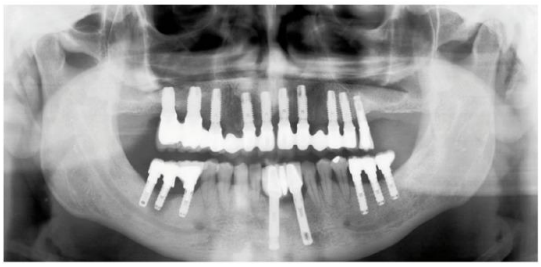


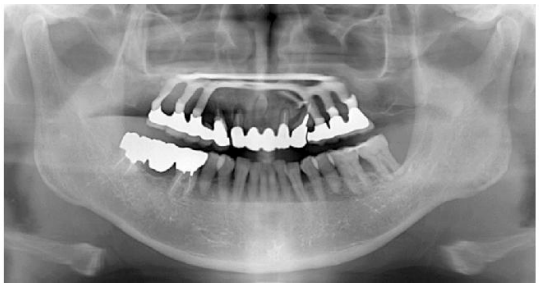
Figure 2. *Components of an implant. A, Implant crown. B, Abutment. C, Implant fixture [33]*

Overall, implants can be classified based on these characteristics. According to **anatomical placement**, implants can be categorized as:

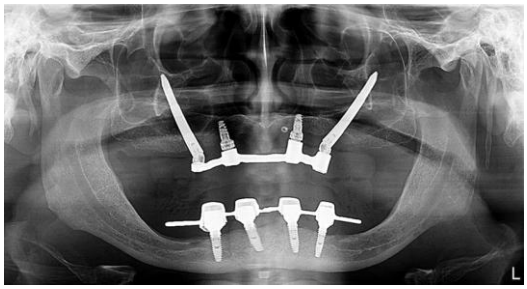
- **Endosteal Implants (Figure 3.a)** are the most common type, since they are indicated for patients with sufficient bone density and volume. These are surgically placed into the patient's upper or lower jawbone. Their design is typically shaped like screws or cylinders. Endosteal implants are recognised in literature for their high success rates and long-lasting support for the implant, which is attributed to a successful osseointegration process [33], [34].
- **Subperiosteal Implants (Figure 3.b)** are placed above the bone and under the gum tissue. A metallic framework is positioned on the bone, with posts protruding through the gums to hold the implant in place. They are indicated for patients with insufficient bone height or density in the jaw or maxilla, who are not advised to go under bone augmentation procedures. Before the emergence of advanced bone grafting techniques, subperiosteal implants were a more common procedure. [35], [36], [37].
- **Zygomatic Implants (Figure 3.c)** are anchored in the zygomatic bone rather than the maxilla. Their design consists of extra-long implants designed to reach the dense zygomatic bone, providing support for prostheses in the maxilla. Their application is indicated in specific cases of severe maxillary bone loss, as it is a complex procedure with potential risks, including sinus complications. Despite this, they offer stability and support for fixed prosthetics [38], [39].



a [40]



b [37]



c [38]

Figure 3. *Types of implants according to anatomical placement.* ***a****, endoseous.* ***b****, subperiosteal.* ***c****, zygomatic*

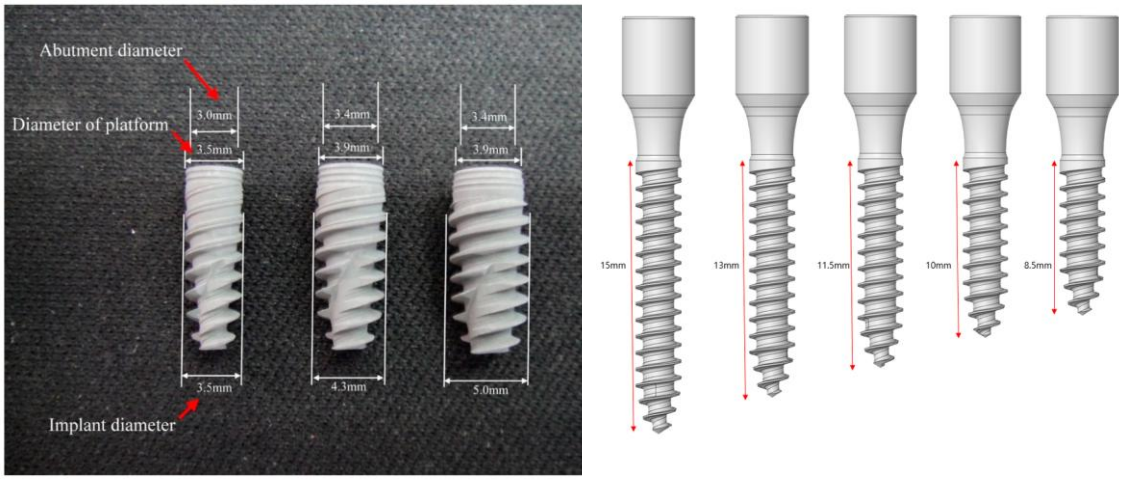
It is estimated that more than 1300 endosteal types of dental implants exist. Differentiating them in terms of size and shape can be challenging both for unexperienced and experienced operators [34].

Their classification based on the **size (diameter and length)** does not have a standardized terminology, since each implant manufacturer designs specific implant measurements, not necessarily coinciding with other manufacturer's dimensions. To address this, some authors [41] proposed a standardized classification system in which, bearing in mind the definition for implant diameter - **(Figure 4a)** the measurement from the outermost point of the thread to the same point on the other side of the thread [42] - the proposed classification divides diameters into four categories:

- **Extra-narrow:** Less than 3.0 mm.
- **Narrow:** From 3.0 mm to less than 3.75 mm.
- **Standard:** From 3.75 mm to less than 5.0 mm.
- **Wide:** 5.0 mm or greater.

Similarly, implant **length (Figure 4b)**, is defined as the distance between the platform and the apex of dental implants [42]. According to length, the authors of [41] classified the implants as follows:

- **Short:** Less than 10 mm.
- **Standard:** From 10 mm to less than 13 mm.
- **Long:** 13 mm or greater.



a[43] b[44]

***Figure 4**. Measurement of implant size. **a,** measurement of implant diameter. **b**, measurement of implant length*

Regarding **shape**, there are three main implant designs for endosteal implants: screws, cylinders, and blades [42], which are illustrated in **Figure 5**. The selection between these different designs depends on the implant's primary stability, which is defined as the mechanical stability that the implant reaches immediately after placement [45]. **Screw-shaped** implants are the most frequently used in dental practice because of their improved primary stability and efficient load distribution in area that interfaces with the bone. **Cylindrical** implants have smooth and non-threaded surface. Even though they are easy to place, they offer worse early stability in comparison to screw-shaped implants. Consequently, they are rarely used in contemporary practice. **Conical** cylinders combine features of both, potentially offering a balance between stability and ease of insertion. **Blade-type** implants were designed to maximize bone contact area. Although they are less common today, they are used in specific clinical scenarios like severe ridge resorption [42].

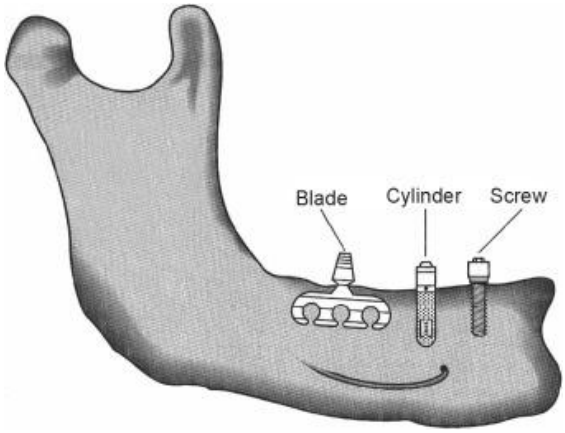


***Figure 5.** Endosteal Dental Implant Shapes [46]*

2.3 AI in Dentistry

AI is a branch of computer science dedicated to creating computer systems able to perform tasks that require human intelligence. This term was first coined by

McCarthy in 1956 at Dartmouth University [47]. High-Level Expert Group on Artificial Intelligence (HLEG AI) define AI as *"software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions."*[48]

Nowadays, AI is experiencing rapid growth and expansion in industries like healthcare, finance, education, and transportation. AI technologies are changing everyday life and industries, from speech and image recognition to driverless cars, smart cities and intelligent robots [47], [49].

The AI landscape has two main approaches: knowledge-based and data-driven. Knowledge-based AI imitates decision-making methods using organized logic and clearly defined expert rules, making it adequate for well-defined issues but limited by the challenges and time needed to formalize and obtain expert knowledge. Nevertheless, data-driven AI identifies patterns in extensive datasets and directly makes decisions, predictions, or recommendations based on large volumes of data, facilitating scalable and adaptable solutions.

Historically, knowledge-driven systems have assisted in modelling clinical expert knowledge. However, since they require high resources, they have gradually been substituted by data-driven approaches. In the context of data-driven techniques, algorithms can be supervised, unsupervised or semi-supervised. In supervised learning the algorithm is trained on a labelled dataset, learning patterns to make predictions or decisions on new, unseen data. Examples of supervised learning popular algorithms are Decision Trees and Support Vector Machines (SVM) for classification tasks; Linear Regression and Logistic Regression for regression tasks.

In contrast, unlabelled datasets are used to train unsupervised learning techniques like Principal Component Analysis (PCA) and k-means clustering. In unsupervised learning, the model learns the data's inherent structure without any specific guidance or instruction. Semi-supervised models are trained with a combination of a small amount of labelled data and a large amount of unlabelled data.

As in many other domains, AI has been widely adopted in the realm of medicine and dentistry. Examples of applications are medical and dental imaging diagnostics, decision support systems, precision and digital medicine, robotic and virtual assistants, drug discovery, wearable devices and hospital monitoring systems. The applications of AI in dentistry are mostly trained on X-ray images and provide support to the dental practitioner in tasks like diagnosis, decision-making, treatment planning, prediction of treatment outcome and disease prognosis. [50]

2.3.1 Deep Learning (DL)

DL is a subset of **Machine Learning (ML)** that uses multiple-layered ANN, which has become a prominent research area that can involve both supervised and unsupervised learning. Deep Neural Networks (DNN) stand out because of their hierarchical structure, that allows examination of intricate data by progressively extracting abstract features from raw input at each hidden layer. A DNN has the capacity to extract relevant features directly from input data

from large datasets, not needing manual feature engineering. Indeed, DNNs simulate how the human brain functions for learning.

As the term "deep" denotes, an ANN architecture is composed of three or more layers: an input layer, one or more hidden layers, and an output layer. Each of these layers contain a variable number of neurons that are interconnected. The input layer in a neural network receives the data, the hidden layers perform mathematical calculations on it, and the output layer returns classification results.

Advancements in fields like computer vision and natural language processing - among others- have been possible due to deep learning models' ability to detect complex patterns in unstructured data. For instance, many image-based dental diagnostics are based on DL trained on large datasets, demonstrating its growing role in enhancing clinical evaluation and decision-making.

2.3.1.1 Application scenarios of deep learning in dentistry

As stated previously, the amount of research of deep learning in dentistry is currently growing and bringing applications that improve the accuracy and reliability of diagnosis. Imaging data is one of the bases for diagnosing oral diseases through both photographic and radiographic images. This way, dentists have the assistance of AI for intelligent recognition and analysis of dental imaging information, leading to an enhanced diagnosis of oral diseases. Based on the oral disease to which they are created for, AI applications in dentistry can be grouped into the following disciplines:

- **Prosthodontics**. Oral implantology, which has been introduced in the previous section, is a branch of prosthodontics. One of the research topics on implantology is the classification of dental types from periapical and panoramic radiographs [51], [52]. Indeed, in [51] a DCNN outperformed most of the participating dental professionals when classifying dental systems. Another type of application of AI in prosthodontics is crown restoration design. In literature, models based on Generative Adversarial Networks learn from extensive datasets of expert-designed crowns to generate personalized prostheses for the patient [53], [54]. Regarding the intersection of AI with medical robotics, there are reports in which successful implant placements have been performed by robots [55], [56], [57].
- **Periodontology.** In this domain there is a focus on periodontitis, which consists of an inflammatory condition affecting the periodontal tissues (the teeth supporting tissues) that results in the loss of periodontal attachment. This pathology is manifested in X-ray imaging with alveolar bone loss, blurring and loss of the top of the alveolar ridge and the hard plate of the bone, and reduction in the height of the alveolar ridge [58]. It has high prevalence, being recognized together with caries as one of the main oral diseases. There is research focusing on measuring bone loss for periodontitis staging using CNNs such as Alexnet, VGG16 and ResNet18 with high accuracy [59]. In other study, the objective of periodontitis staging is achieved with image segmentation: In [60] Multi-Label U-Net and Mask R-CNN models were compared for segmenting digital panoramic radiographs. Other researchers focused on Clinical attachment level (CAL) prediction [61], radiographic bone loss (RBL) classification [62] and diagnosis and prediction of periodontally compromised teeth [63] .

- **Restorative dentistry.** In the field of restorative dentistry most research is intended for caries detection in images, as it stands as the most prevalent oral disease. It is characterised by demineralization and destruction of tooth structure due to acid produced by oral bacteria, leaving spots on the tooth that can be detected by the models. Indeed, it can lead to nerve damage in severe cases. On one hand, there is research literature on predicting the risk of caries based on lifestyle, demographic and environmental factors [64], [65]. On the other hand, there is literature aiming to detect and diagnose dental caries on radiographs through CNNs [66], [67], [68].
- **Endodontics**. AI has already proven to be effective in the field. In literature, there are DL models made for diagnosing periapical radiolucencies and vertical root fractures using radiographs and CBCT [69], [70], [71]. In addition to that, there is research focused on predicting postoperative pain following root canal treatment [72] and support retreatment decisions with high reliability [73].
- **Oral surgery**. In the field of oral surgery, there is a predominant use of CNN for detection and classification of maxilla bone lesion's features obtained in panoramic radiographs (PR) [74], [75] and cone-beam images (CBCT) [76]. There is also research aimed at enhancing orthognathic surgery, which is a surgical procedure aiming to achieve a correct function of the maxillo-mandibular complex. For example, in [77] ,the authors proposed a CNN able to predict the need for orthognathic surgery using cephalogram, achieving an accuracy of 95%.
- **Anatomy**. In the anatomy field there is an extensive use of CNNs for automated tooth detection and numbering [78]. An example of this is the research conducted in [79], in which the performance of the proposed solution is comparable to the level of experts.
- **Orthodontics**: Orthodontics is a field that has had developments regarding available technologies with the availability of cone beam computed tomography (CBCT), the intraoral scanner, and the integration of AI in these devices [80]. In diagnosis, there is research that has reached high performance in identifying anatomical landmarks on cephalograms for classifying skeletal malocclusions[81], and predicting mandibular growth and skeletal maturity after puberty [82]. Regarding treatment planning, the models have shown high performance in predicting orthodontic decisions such as extraction needs [83] and surgical indications [84]. Another example is Dental Monitoring System, which is a tool that enables an accurate remote semi-automatic monitoring of the treatment [85].

In this work, an application in the domain of prosthodontics -specifically implantology- area will be developed.

2.4 Deep Learning Techniques in Computer Vision

Computer Vision (CV) is a subfield of AI that allows machines to interpret, analyse and understand visual information like images and videos, in a similar manner as the human eye does. The first research in this field was made in the decades of 1960 and 1970, with projects that aimed to interpret simple images in controlled environments. The development of more sophisticated algorithms, the increase of computational power and the developments in DL had resulted

in significant progress in this discipline. DL and CV are not isolated research fields, but rather deeply interconnected areas that have evolved synergistically

The practical applications of CV span across various industries: healthcare (e.g., medical image analysis), manufacturing, transportation, security and entertainment, among others.

A fundamental step of every CV pipeline is Feature Extraction, a process in which the most relevant attributes of an image are identified for posterior analysis and classification. These extracted visual features -which can be edges, shapes, textures and patterns- are converted into numerical data for further processing.

The main tasks addressed in the field CV are:

- **Image Classification**: This task involves assigning one or various predefined labels to an image based on its visual content. CNN-based models (e.g., ResNet, VGGNet or EfficientNet) are commonly used for this task.
- **Object Detection:** Involves localizing and labelling specific instances of an object inside an image or video frame. The process typically involves two parts. The first is object localization, in which the location of the object is marked typically with a bounding box. After this, the object is classified into a predefined category. State-of-the-art models such as Faster R-CNN, YOLO (You Only Look Once), and SSD (Single Shot MultiBox Detector) are commonly employed for this purpose.
- **Object Tracking:** Refers to the process in which a specific object is identified and monitored along a sequence of video frames, playing an important role in video analysis. This task is essential in autonomous driving and Algorithms like Deep SORT, SiamMask or Tracktor allow to keep the identity of an object through multiple video frames, even with occlusions or changes in illumination.
- **Segmentation**: Segmentation techniques divide an image in distinct regions based on shared characteristics of pixels, understanding object boundaries. Semantic segmentation involves assigning a category to every pixel of an image, based on the object it belongs. Instance segmentation identifies and isolates each object at a pixel level. Models like U-Net, DeepLabv3+ and Mask R-CNN are popular for segmentation tasks.
- **Image Generation**: This has the purpose to synthetize new visual content from learned distributions, employing generative models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) like StyleGAN and Pix2Pix.

In this work, the type of task that will be explored is Image Classification. Therefore, the present work will develop and evaluate a classification model.

2.4.1 Convolutional Neural Networks (CNNs) for Image Classification

CNNs are a specialized class of deep learning models designed explicitly for CV tasks. They excel at learning complex hierarchies of features, which allows them to distinguish intricate patterns and make image classifications with high performance. CNNs first appeared in the 1980s. However, they became increasingly popular in computer vision over the last ten years, thanks to the massive boost in computational power (GPUs) and an increased availability of

large and labelled datasets. Today, they have become a key technique in image classification tasks.

During the last decade it has been a considerable amount of research for CNNs. Consequently, there are 'core' architectures such as ResNets, ImageNet and MobileNet. Even though they have variations, the general architecture is the same, being composed by an input layer, alternate layers of convolution and pooling layers, one or more fully connected layers, activation functions, and an output layer at the end. This general architecture is illustrated in **Figure 6.** To give a clearer understanding of it, the main components of this architecture are detailed below [86]:

- **Input Layer:** Receives the input image, represented as a matrix of pixels with one or more channels (for example, RGB).
- **Convolutional layers:** Scan the input images applying kernels in order to detect local patterns in data, like textures, shapes, corners, edges or other low-level features. They generate a feature map as output that preserve the spatial relationships between pixels. Preserving the spatial relationships allows the network to maintain information about the relative arrangement of features in the image. As one progresses deeper within the network, convolutional layers are able to identify more abstract and complex representations by combining patterns detected in previous layers. The result of this operation is a set of feature maps that indicate the presence and location of the learned patterns.
- **Pooling layers:** They are usually found in between convolutional layers. These layers preserve the most relevant features while reducing the dimensionality of feature maps. This process, known as downsampling, adds a degree of spatial invariance while lowering the number of parameters and computational cost. The most popular method is max pooling, which highlights the most important activations by choosing the maximum value within a specific area of the map.
- **Flattening Layer:** The convolutional and pooling layers' output is flattened into a one-dimensional vector so that it can be processed in the following layers.
- **Fully Connected Layers**: In these, each neuron is connected to all neurons of the previous layer. In order to generate a prediction, these layers function as a classifier, processing the features that were extracted.
- **Output Layer:** They are usually located at the end of a CNN to make the final prediction using an activation function. Softmax activation function or ReLU are frequently used in classification tasks to translate output values into class-specific probabilities.

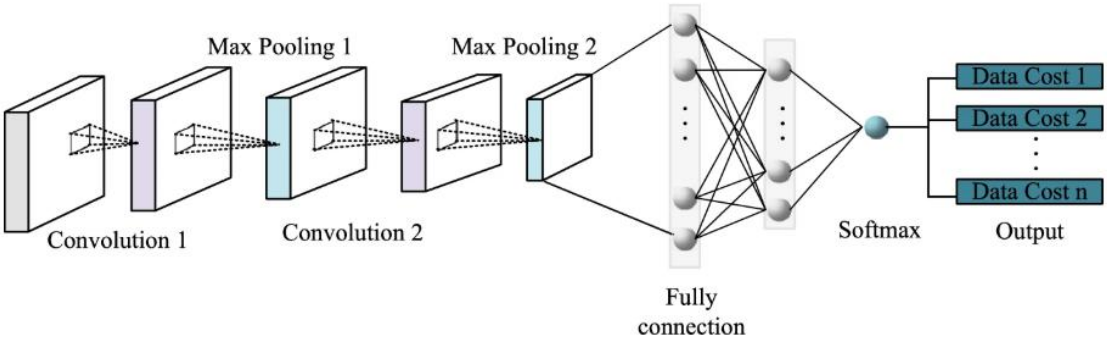


Figure 6. CNN architecture [86]

A notable characteristic of CNNs is that the first layers learn general features that can be transferred between different CV tasks. This phenomenon has given rise to **Transfer Learning (TL)**, a technique that reutilizes previously trained models in general tasks to adapt them to specific tasks. These pretrained models have learned common visual characteristics from large datasets. The idea behind is to use the knowledge gained in the previous task to have better performance on the new one.

Common transfer learning strategies include feature extraction, when the initial layers of the pretrained model are used; and fine-tuning, which implies retraining some layers of the model to adapt them to the new task.

TL is commonly used in the field of computer vision. Since it reuses pretrained models, it reduces training time and resources when training new models. Furthermore, models based on transfer learning achieve superior performance, even when the dataset is small. Indeed, as the model has previously learned general features, the risk of overfitting is reduced. [87]

2.4.1.1 Evolution of CNN architectures

As explained in previous sections, DL has allowed a big progress CV, especially in image classification tasks with CNNs. CNN's architecture evolution has allowed substantial improvements not only in performance, but also regarding computational efficiency. In this subsection the most representative will be outlined, from the first contributions to the most recent developments in the field.

The first advancements in CNN laid the foundations for the development of more complex and efficient models. These architectures introduce key concepts like non-linear activation functions, an increase in network's depth and an accelerated training with GPUs. Even though today they have been replaced by more sophisticated models, their impact in image classification was pivotal.

First, **AlexNet** [88], which was proposed in 2012, marked a turning point by winning the Challenge ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Its architecture consists of five convolutional layers followed by three fully connected layers. The architecture implements ReLU activation function to introduce non-linearity and activate the outputs of convolutional layers. Indeed, it was one of the first models efficiently trained using GPUs, which helped its scalability.

VGGNet [89], developed in 2014, proposed deeper networks by using small 3x3 convolution filters, which allowed to capture finer details in the images. VGG-16 and VGG-19 are popular variants in literature, as they shown that an

increased number of weight layers implies improvements in performance, although at the expense of a considerable increase in the number of parameters.

GoogLeNet [90] and Inception modules introduced an architecture in which convolutions of different sizes were applied in parallel in the same module. This strategy permitted feature extraction at multiple scales without incrementing drastically computational cost. Later versions, such as Inception-v2 and v3, incorporated batch normalization and factorized convolution to further improve efficiency.

As the depth of networks was increased, new challenges like gradient vanishing and redundancy in feature extraction appeared. To address these limitations, new connectivity strategies between layers were introduced. Architectures based on residual and dense connections allowed more stable and deeper networks. Furthermore, they promoted efficient reutilization of features and a more effective gradient propagation.

ResNet [91], proposed in 2016, addressed the degradation problem in deep networks as it introduced residual learning through skip connections. This new apporach made possible the training of extremely deep networks, such as ResNet-152, by facilitating gradient flow and mitigating vanishing gradients.

DenseNet [92], presented in 2017, took connectivity one step further as it connected each layer with all the previous, which favours feature reutilization and gradient propagation. Furthermore, it achieves a considerable reduction in parameter number without compromising performance.

With the increasing demand of CV applications in smartphones, embedded systems and environments with memory and energy restrictions, faster and lighter models were needed. This research line has boosted the development of optimized architectures that sacrifice minimum of precision in change of a significant reduction in computational task, without losing competitiveness in classification tasks.

MobileNets [93] was a series of architectures designed to be efficient in mobile and embedded devices. Introduced in 2019, they employ depthwise separable convolutions to reduce computational complexity. Later versions like MobileNetV2 and MobileNetV3 integrated lineal bottlenecks, inverted residual connections and neural architecture search (NAS), which improved notably the balance between efficiency and precision.

Proposed in 2019, **EfficientNet** [94] introduced a compound scaling strategy that adjust uniformly depth, width and resolution of the network. This technique allowed to obtain cutting edge results in benchmarks like ImageNet with fewer parameters and a reduction of computational cost in comparison to previous.

In the last years, the field has experimented an acceleration in innovation, with the appearance of hybrid models, multimodal paradigms and architectures inspired in Transformers. These new approaches had overcome the limitations of traditional CNNs and had opened new possibilities in terms of scalability, precision and generalization capabilities. Attention integration, symbolic reasoning, automated architecture search and sequential encoding has enabled cutting-edge performance across various benchmark dataset.

In this context, **Vision Transformers (ViT)** [95] represented a breakthrough with traditional CNNs. Introduced in 2020, these models process images as patch sequences and apply attention mechanisms inspired by NLP. They gave rise to variants like DeiT, MAE and ViT-G, with applications in multiple CV tasks.

ConvNeXt [96], which was proposed in 2022, adopted architectural elements of ViT and incorporated them to modern convolutional networks, achieving to close the performance gap between CNNs and ViTs.

Convolution and Attention Network (CoAtNet) [97] combines convolutions with attention mechanisms, putting together the best of both worlds: spatial inductives of CNNs and global modelling of Transformers.

MedMamba [98], which was proposed in 2024, integrates structured state space with visual data processing, employing models with bidirectional visual blocks. The architecture reduced the computational cost of the attention mechanism and offering cutting-edge results.

Lastly, **ProtoViT** [99] combines deep learning with case-based reasoning. It classifies images through explicit comparations with learned prototypes, providing a more transparent interpretation of the decision-making process.

Over the years, architectures based on convolutional networks have evolved significantly, enabling improvements in both performance and efficiency in image processing. Among all the currently available options, two models are studied in this work: ResNet18 and EfficientNetB2. Each offers specific advantages that make them especially interesting for this study. In the case of ResNet18, its ability to facilitate the training of deep networks thanks to residual connections. While in the case of EfficientNetB2, its balance between accuracy and computational consumption. For this reason, the following sections go deeper into these models, explaining their main contributions.

2.4.1.2 ResNet18

ResNet18 is one of the best-known versions of the Residual Networks (ResNet) family, proposed as a solution to the vanishing and exploding gradients problems. The problems are encountered when trying to train very deep neural networks, that is, DNN with a large number of layers. As more layers are added, not only training becomes more complicated, but instead of improving performance, it starts to degrade. On one hand, vanishing gradient happens when the gradient values tend to diminish during backpropagation. This makes learning slower in the earlier layers, hindering the network's ability to learn useful features as depth increases. On the other hand, exploding occurs when gradients grow exponentially in very deep networks, leading to too large weights and numerical instability. In essence, vanishing results in the gradient becoming zero and exploding results in the gradient being overly large.

To overcome this limitation, ResNet introduced a very simple but effective idea: residual connections, also known as skip connections. These connections allow information to jump from the output of an earlier layer (for instance, n-1) to the output of a later layer (for example, n+1). As n-layer is skipped, gradient flow is facilitated and deep networks are trained with greater stability. The benefit behind this type of skip connection is that if there is any layer hurting the performance of the model or not adding value, it is skipped.

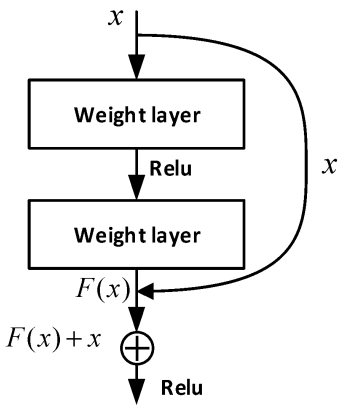


Figure 7. *Residual module structure. [100]*

In this context, ResNet18 consists of 18 layers with trainable parameters, distributed in residual blocks. Each of these blocks allows the network to learn not a complete transformation, but only the difference from the original input, which simplifies the optimization process.

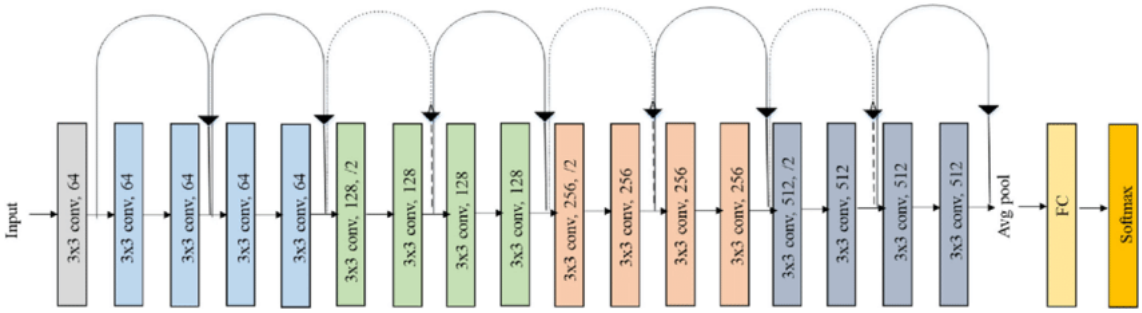


Figure 8. *ResNet-18 Architecture. [101]*

One of the reasons ResNet18 was used in this work is its balance between simplicity and learning capacity. It is a network that does not require a large amount of computational power but offers robust results, making it very useful as a base model or point of comparison in projects where efficiency is needed without sacrificing accuracy.

2.4.1.3 EfficientNet

EfficientNetB2 is part of the EfficientNet family, which as explained previously stand out for having rethought the way neural networks are scaled. Traditionally, there were three options in order to improve a model's performance: increasing its depth (more layers), increasing its width (more filters per layer), or working with higher-resolution images. Increasing each of these features can lead to improve network performance. These three features are directly related with each other, as increasing resolution leads to more features being extracted for training, allowing for a deeper network. Despite this, applying these changes in isolation does not always help to achieve good results. Indeed, it can lead to inefficient or difficult-to-train networks.

EfficientNet's proposal is based on what is called the compound scaling method, which consists of jointly and in a balanced way augmenting the input depth, width, and resolution, following the formula defined in (1). The explained strategy allows for more accurate models, but without increasing computational costs. This formula depends on a scaling factor and three constants:

$$depth: d = \propto^{\Phi}$$

$$width: w = \beta^{\Phi} \quad (1)$$
$$resolution: r = \gamma^{\Phi}$$

where:

- Φ is the global scaling coefficient, which controls how much the model size is increased.
- α, β, γ are positive constants that determine how the scaling is distributed among depth, width, and resolution. The optimal values for the constants are: α=1.2, β=1.1, γ=1.15. These values are the same for all models of EfficientNet family.
- A constraint is imposed to keep computational usage under control:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (2)$$

This restriction makes sure that the total number of FLOPS (floating-point operations per second) doubles when Φ is increased by 1. In this way, a balance between learning capacity and efficiency is maintained.

The base model, EfficientNet-B0, was generated using Neural Architecture Search (NAS) and uses MBConv (Mobile Inverted Bottleneck Convolution) blocks along with attention mechanisms such as squeeze-and-excitation. From this base architecture, different versions (from B1 to B7) were scaled, varying the value of ϕ to adapt to different performance and computing power needs.

In the case of EfficientNetB2, it is an intermediate version of the EfficientNet family. It has improved performance in comparison to smaller models like B0, while being manageable for training when resources are limited. In detail, EfficientNet-B2 corresponds to the case where ϕ=2, which means that it has a deeper, wider, and higher-resolution network than B0 and B1. Nevertheless, it does not reach the computational cost of higher-level versions like B5 or B7. Its architecture incorporates efficient ideas such as depthwise separable convolutions, inverted residual blocks, and attention mechanisms such as squeeze-and-excitation, which help the model focus on the most relevant regions of the image.

In this work, EfficientNetB2 was selected for its good balance between accuracy and efficiency. Its design allows for very good results in classification tasks without the need for a particularly powerful training infrastructure, making it an ideal choice for addressing this type of problem from a practical perspective.

2.4.2 Graph Neural Networks (GNNs)

GNNs have become an increasingly used tool in DL, especially when the data presents a non-regular structure. Unlike CNNs, which are designed to operate on images organized in a two-dimensional mesh, GNNs work with data that is represented on a graph structure. A graph can be defined a data structure formed by two components: of nodes (entities, such as points or regions in an image) and edges (connections or relationships between nodes). In fact, graphs are useful for modelling complex spatial interactions. With graphs, GNNs have the possibility do node-level, edge-level, and graph-level prediction tasks.

Inside the GNN family, one of the broadest applied models is the Graph Convolutional Network (GCN), introduced first by Kipf and Welling in 2016 [102]. This model adapts the idea of traditional convolutions from image domain to graph domain. The main idea behind GCNs is to create a mapping function that can extract interaction-aware features from the node features and their neighbouring nodes, allowing each node to update its representation by considering both its own features and those of its neighbours.

Graph Convolutional Network (GCN) is a type of neural network that operates on data structured as a graph, rather than the regular grid of pixels in an image. In was introduced in 2016 by Kipf et al. [102]. Unlike traditional CNNs that apply convolutions on a grid of pixels, GCNs apply convolution operations on the structure of the graph. This means that each node combines or aggregates information from its neighbouring nodes (according to edges) to update its own feature vector.

In simple terms, a GCN learns to disseminate and integrate information along the connections of the graph, so that each node "learns" the context provided by its neighbours. For instance, when each node has a set of initial features (e.g., local image measurements at that point). In a GCN layer, each node takes the features of its neighbours (and its own), combines them (sum or weighted average) and passes that result through a function (e.g., a linear transformation followed by a nonlinear one). This process is analogous to convolution in imaging, but here the "receptive field" of a node is given by the topology of the graph rather than a fixed neighbourhood of pixels.

Repeating this message-passing at multiple layers allows the network to incorporate information from more distant nodes (not just immediate adjacencies) as the scope of connections with each additional layer is "widened". In addition, GCNs can include pooling or coarsening (node reduction) mechanisms to summarize regions of the network, analogous to pooling in CNNs, thus widening the field of view of the upper layers.

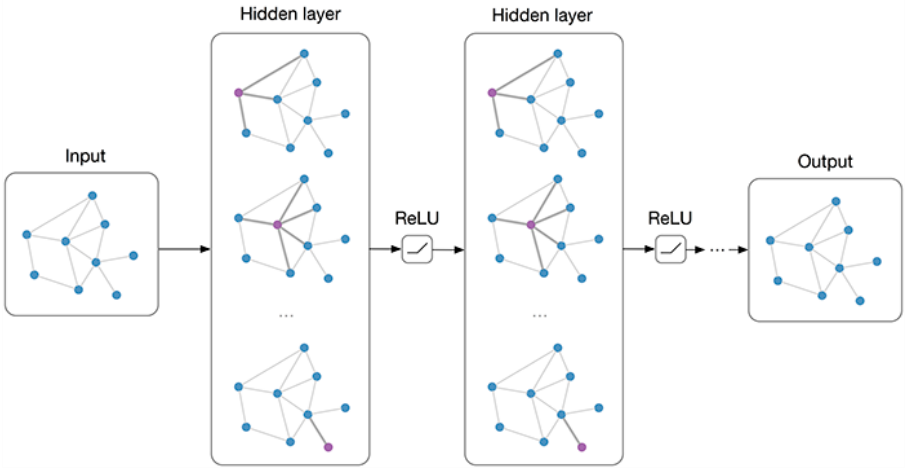


Figure 9. *Multi-layer GCN [103]*

In practice, GCNs are well suited to problems where the data have non-Euclidean or structured relationships (such as anatomical connections in the body, relationships between different regions of an image, or between different images). In summary, a GCN extends the ideas of classical CNNs to data where "pixels" are arbitrarily connected. Instead of sweeping a filter over an entire image, the GCN learns to combine values only from connected nodes in the

graph. This allows modelling complex interactions, for example, how different parts of a body (represented as nodes) influence each other in an X-ray to determine a diagnosis.

Example of this is the work of Li et al. In it, a GCN model estimates bone age from hand X-rays using a graph of key anatomical points and regions as nodes. Using a GCN, the model propagates information between bone regions to improve age prediction. The results show that this approach outperforms traditional methods, reducing error in age estimation. Indeed, it validates that explicitly modelling anatomical regions and their relationships using a GCN can better capture bone maturation signals than a global CNN, aligning with the clinical practice of examining specific points on the radiograph [104].

Other example is the work of Tang et al., in which a graph of lung findings in chest X-rays is used to detect COVID-19. Each node represents a lung region, and a GCN integrates the relationships between them to classify the image. Integrating GCNs added robustness to the model: by capturing how multiple abnormal signals reinforce each other in the image, the model could better distinguish the characteristic patterns of COVID-19 from other pneumonias. The model achieves an accuracy of 96.4%, outperforming previous methods and improving COVID-19 detection by capturing spatial interactions between anomalies [105].

2.4.3 Multimodal deep learning (MMDL)

In the recent years, deep learning has transformed medical image analysis, achieving notable advances in tasks such as classification, segmentation, among others. Nevertheless, most traditional approaches have been applied on a single data modality, without considering the wealth of information that different sources of data can provide. In this context, multimodal deep learning (MMDL) emerged as a solution that create models able to process and link information using various modalities, typically images, video, audio, and text. This integration enables models to capture complex relationships and complementarities between different data modalities, resulting in richer and more robust representations. Indeed, MMDL has become a focus of research in the field of medical image analysis during the last years, due to its potential to improve diagnostic performance and reduce the ambiguity associated with single-modal analysis.

There are two main categories of multimodal fusion strategies: model-agnostic approaches and model-based approaches. On one hand, **model-agnostic approaches** are not based on any specific network architecture. Three main types -which are illustrated in **Figure 10**- can be distinguished:

- **Early fusion:** In this strategy the features extracted from each modality are combined in an early processing phase, generally in feature space. Therefore, they are also known as feature-level fusion. This technique is useful for capturing low-level correlations between the data. However, due to the premature feature integration it can easily become too dependent on the training set, which can increase the risk of overfitting and semantic loss in posterior phases.
- **Late fusion:** In this approach, each modality is processed in an independent way by different models. Then, the final decision is obtained jointly by combining the outputs of each model. Indeed, it is also known as decision-level fusion. The main advantage of this method is its flexibility, as each modality can be trained with the architecture that best

suits its characteristics while keeping independence between them. Despite that, it can limit the interaction between modalities during learning.

- **Hybrid fusion:** This integrates the advantages of early and late fusion, thus forming a multi-level feature fusion system. Preliminary feature-level integration is performed first, followed by higher-level decision combinations. Although it effectively integrates the advantages of early and late fusion, it entails greater computational complexity and resource demands.

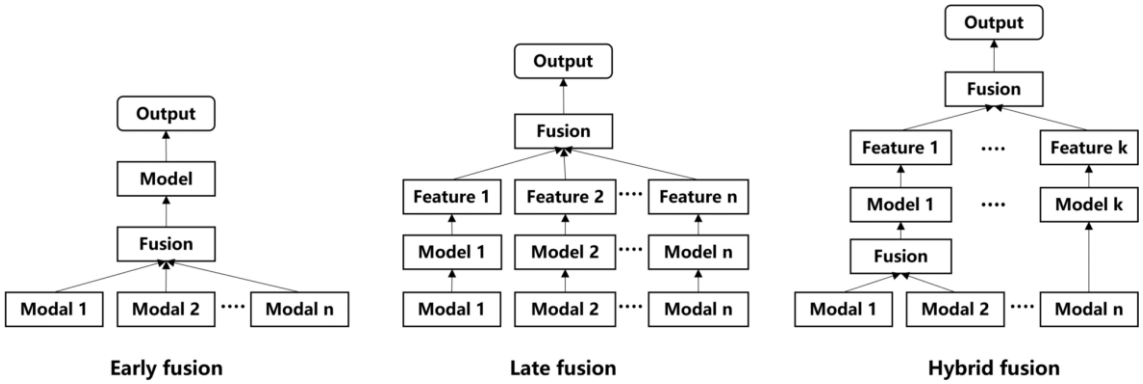


***Figure 10**. Types of model-agnostic approaches [106]*

On the other hand, **model-based approaches** address fusion from an algorithmic and architectural perspective. Common approaches include:

- **Multiple Kernel Learning (MKL**): This approach is based on the support vector machine (SVM) model but introduces an important improvement by allowing the simultaneous use of multiple kernel. Each these kernels acts as a similarity function that captures different perspectives of the data. As the model assigns different kernels to different modalities, it can efficiently integrate heterogeneous information—for instance, structural images and physiological signals—without the need to transform the data into a common space. MKL's main advantage is flexibility, as it allows the relative weight of each kernel to be adjusted based on its relevance to the classification task, making it a robust option when dealing with data of diverse nature and scale. Nevertheless, its effectiveness depends largely on the proper selection of core functions and the system's ability to combine them optimally. This tuning process is not trivial and can require significant computational power.
- **Graphical Models:** These methods exploit the spatial or temporal relationships present in the data—for example, the arrangement of brain regions or the time course of a disease—using techniques such as segmentation, relationship prediction, or the construction of dependency networks. One of their strengths is that they allow prior medical knowledge to be integrated into the model's structure, which can provide a very useful layer of interpretability. However, this same dependence on expert knowledge can be detrimental when it comes to generalizing to other contexts, patients, or different devices.
- **Neural Networks (NNs):** NNs have are widely considered as one of the most powerful and versatile tools for multimodal fusion due to their ability to learn complex hierarchical representations directly from data. These enable effective and scalable fusion of complex features. Their end-to-end training enables the capture of complex nonlinear correlations between modalities, which may go unnoticed in traditional methods. In

addition to that, they have good scalability and adaptability. However, NNs present limitations in terms of poor interpretability -which is important in the medical diagnosis field- and large annotated datasets requirements. Among the most notable models are **CNNs**, recognized for their effectiveness in extracting local patterns in images, and **Vision Transformers (ViTs),** which allow to capture relationships within the image through attention mechanisms. Lastly, **graph neural networks (GNNs)** offer a flexible solution for processing graph structure data by modelling irregular non-Euclidean data.

In this work, MMDL will be applied for DIS classification. The integration of different image and tabular information would allow the complementary between modalities to improve performance in classification.

2.4.3.1 MMDL applications combining image and tabular data

Multiple literature has recently explored multimodal architectures combining visual feature extractors -like CNNs and ViTs- with models designed for processing tabular data -such as MLP and TabNet.

An example of this is the study of Hinterwimmer et al., in which a deep multimodal model that combines bone X-rays images with patient-specific clinical metadata (age, gender, tumor location, etc.) was proposed to classify bone tumors. In the architecture, they used **ResNet** and **NesT** image classification model along with an **MLP** for tabular data, fused in a joint layer. The model outperformed other image-only models. This work demonstrates how a residual CNN (or equivalent) combined with tabular metadata can outperform unimodal models in medical classification tasks.[107]

In the study of Kita et al. a model was developed to classify two types of spinal cord tumors by integrating MRI images with patient clinical data. The architecture is composed by **EfficientNet** as the image feature extractor and **TabNet** (an attention-based neural network for tabular data) for patient information. Then, the features from both branches are combined and fed to fully-connected layers for prediction. In the experimental comparison, the bimodal model outperformed both the image-only model and the model using just tabular data, achieving an AUC-ROC of 0.91. The model even outperformed medical specialists. Furthermore, its key contribution is showing that EfficientNet+TabNet with learned fusion outperforms simpler fusions, demonstrating the effectiveness of TabNet in extracting patterns from clinical data and their integration with deep visual features.[108]

Similarly, the study of Panneerselvam et al. proposes a multimodal model for skin disease classification that combines dermoscopic images with patient clinical metadata. The architecture consists of **EfficientNet-B3** (pre-trained, tuned using transfer learning) to process skin lesion images, and **TabNet** to process tabular clinical variables; the representations of both are fused using a personalized attention mechanism. The resulting model achieved an accuracy of 98.69% in classifying seven skin lesion types in the ISIC 2018 dataset, outperforming unimodal classifiers and other previous approaches. The authors attribute this high performance to EfficientNet's power in vision and TabNet's ability to handle heterogeneous data with interpretability, along with attention fusion that captures the interaction between image and metadata.[109]

In the ViT field, Khader et al.'s work introduced a fully Transformer-based architecture for combining chest X-rays with clinical parameters of ICU patients, with the goal of simultaneously diagnosing up to 25 pathological conditions. For this purpose, they employ a **Vision Transformer (ViT)** as a backbone for the X-

ray images and a Transformer adapted to tabular, which tokenizes numerical/categorical features) to encode the clinical variables. The representations from both modalities are then fused. The results showed that the Transformer multimodal model significantly outperformed unimodal models in detecting multiple diagnoses in critically ill patients. Specifically, for the MIMIC-IV data set, the image with data combination achieved an average AUC of 0.77, higher than image-only or data-only models.[110]

In summary, recent literature shows that combining medical images with tabular clinical data using multimodal architectures can significantly improve diagnostic performance. Models that integrate CNNs or Transformers for images along with MLPs, TabNet, or FT-Transformer for tabular data have proven effective in those studies. Therefore, adopting multimodal strategies in medical data analysis represents a promising step toward more accurate and robust diagnostic support systems.

2.5 Automatic Implant Classification: Existing Work

The correct identification of DIS represents a growing challenge in clinical practice due to multiple factors. One major reason is that the implant market is saturated with a wide variety of brands and models, with subtle variations between models from the same brand and continuous design updates. The rise of dental tourism is another factor, in which patients receive implant treatments abroad and return without adequate documentation. Furthermore, another key issue is the lack of clinical records in patients who change of dental professionals, leaving dentists with little or no information to base their restorative or diagnostic decisions on [111]. All these circumstances show the necessity for accurate automatic implant identification systems.

In recent years, approaches based on DL had gained relevance in the field of dental imaging, offering promising solutions for DIS automatic identification and classification. Traditionally, this task required clinical experience of the dental practitioner, as well as high-quality radiographs. However, AI has demonstrated to overcome dentists both in speed and precision. Even though literature has focused more on panoramic radiographs, periapical radiographs -with superior resolution and level of detail- are especially useful for identifying finer characteristics of the implant. Such characteristics can be brand and diameter, which are essential for treatment planning and complications management.

Early studies had shown the feasibility of models based on CNN to identify implant brands in periapical radiographs. Lee et al. evaluated the performance of a fine-tuned Inception-v3 model, trained with more than 10.000 cropped radiographs (including 5380 periapical) from three implant brands: Osstem, Dentium and Straumann. The model achieved an AUC of 0.979 in periapical images, overcoming the dentist (AUC=0.959), which validates its clinical applicability for brand classification with high-resolution radiographs. [112]

The study of Kim et al. evaluated the capabilities of pretrained neural networks for classifying four types of common dental implant brands from exclusively periapical radiographs. Using a smaller dataset composed of 801 images, the authors employed transfer learning on five CNN architectures (SqueezeNet, GoogLeNet, ResNet-18, MobileNet-v2 and ResNet-50). Despite the reduced size of the dataset, all models achieved accuracies higher than 90%, with ResNet-50 and ResNet-18 standing out with a 98% accuracy.[113]

In this context, Santos et al. trained a CNN with 1800 periapical radiographs from 3 brands (Straumann, Neodent and SIN). After applying data augmentation techniques, the model achieved 85,29% of accuracy. The authors indicated that with a bigger dataset the system could have the potential to be helpful in clinical practice. [114]

Similarly, Tiryaki et al. used five implant brands and transfer learning techniques with multiple CNN models, including VGG-19 and ResNet. The best model, based on VGG-19 with majority voting of 5 models, achieved a precision of 98,9% and F1-Score higher than 96% for all brands, including periapical radiographs as part of the training set. [115]

In a study made on a larger scale, Park et al. compiled more than 156.000 radiographs (panoramic and periapical) from 27 implant brands. When evaluating exclusively periapical radiographs, the model achieved an AUC of 0.987, confirming the efficacy of this type of image in brand classification. [116]

Another notable contribution was made by Lee et al., who trained a ResNet-50 model in high-quality panoramic and periapical radiographs and evaluated in low-quality or distorted periapical radiographs. Despite the challenging conditions, the model achieved a precision of 95,1%, overcoming five dental practitioners who achieved an average accuracy of 37,2%. [117]

Recently, Benakatti et al. trained a model based on transformers -DEtection Transformer (DETR)- with 1138 radiographs, 868 of them being periapical. The study included 5 implant brands: Osstem TSIII SA, Dentium superline, Noris Tuff TT, MIS V3 and Adin Touareg. After applying data augmentation, the model achieved an AUC of 0.96. The results show a robust performance in multiclass classification even though there were signs of overfitting in the validation set, which indicated the necessity of optimization for better generalization. [118]

Whereas most of the research has focused on implant brand, the diameter of classification has received less attention. However, estimating the diameter is as clinically relevant as estimating the brand. Implants -even from the same brand- can present a wide range of connections depending on the diameter. They can have narrow, regular or wide connections, which are closely related with the compatibility of component of the prosthetic system. Therefore, developing models capable of automatically classifying not only the brand, but also the implant's diameter would enable to determine and prepare compatible components for repair and maintenance in case of mechanical complications

In this context, Park et al. developed a dual approach for implant size classification (diameter and length), using exclusively periapical radiographs. One of the models was based in DL (VGG16) and the other one in clustering, employing manually annotated anatomical features. The DL-based model achieved precision ≥0.994 and AUC-ROC ≥0.975, whereas the clustering model was slightly less accurate in general but shown a better performance for diameter estimation. These findings suggest that a future combination of visual features with anatomical landmarks could improve accuracy and robustness of the classification system. The study also highlighted that periapical images are ideal for diameter estimation due to their higher resolution and lower distortion in comparison to panoramic. [119]

Altogether, the studies confirm that periapical radiographs -despite their limited field of view- are highly effective for DL classification both for implant brand and diameter. Its high resolution allows to distinguish structural differences between different implant systems. Furthermore, models like ResNet-50,

VGG16/19 and Inception-v3 have shown an excellent performance across datasets, even in less ideal conditions.

In clinical practice, identifying correctly both diameter and model is important for managing mechanical complications (such as fracture of dental prothesis) and selecting compatible prosthetic components. Revised literature supports the use of DL models as decision-support tools, especially when medical records are unavailable and image quality is deficient. As datasets become more representative and as hybrid models combining both image features and anatomical landmarks evolve, it is expected that these models will be an essential component of odontological clinical workflows.

2.5.1 Commercial solutions for AI-assisted implant identification

As indicated before, the identification of DIS supposes a relevant challenge in clinical practice, especially in cases of patients with previous treatments made by other professionals or in centres without access to complete clinical records. During the last years, AI-based solutions to solve this problem have emerged not only in research, but also in industry. In this section, the principal commercial solutions integrating AI for DIS identification in dental radiographs will be introduced, as well as be compared in **Table 1**:

- **Spotimplant**: The solution employs deep learning to identify implants from periapical radiographs. Its database includes more than 3500 models from more than 300 brands. The system generates a report with brand, model and approximate size from the detected implant. In 2023 Spotimplant was acquired by Allisone, which has allowed to integrate its capacities inside a broader set of digital dental tools.
- **Implantif.AI**: Developed by the Spanish startup MovumTech, the solution analyses periapical X-rays images and detects dental implants. It offers information about brand, model and size, as well as suggestions of compatible components and contact with commercial representatives. Each image uploaded by users contribute to improve the performance of the model through continuous learning.
- **Dentifier:** Dentifier is a solution developed by ProScan in collaboration with Cegeka, an IT company. They have developed an AI-based solution for automatic implant brand identification. Its self-learning system allows for continuous improvement in accuracy as the database expands. However, it is still in the launch phase.

Name	Company	Year	Country	Brand ident.	Model ident.	Size pred.	Notes
Spotimplant	Allisone	2019	France	Yes	Yes	Yes	Specialized in dental implants; large proprietary database.
Implantif.AI	MovumTech	2022	Spain	Yes	Yes	Yes	Also provides component compatibility and sales contact
Dentifier	Proscan & Cegeka	2024	Belgium	Yes	Yes	No	Still in launch phase

Table 1. *AI-assisted implant identification solutions comparison.*

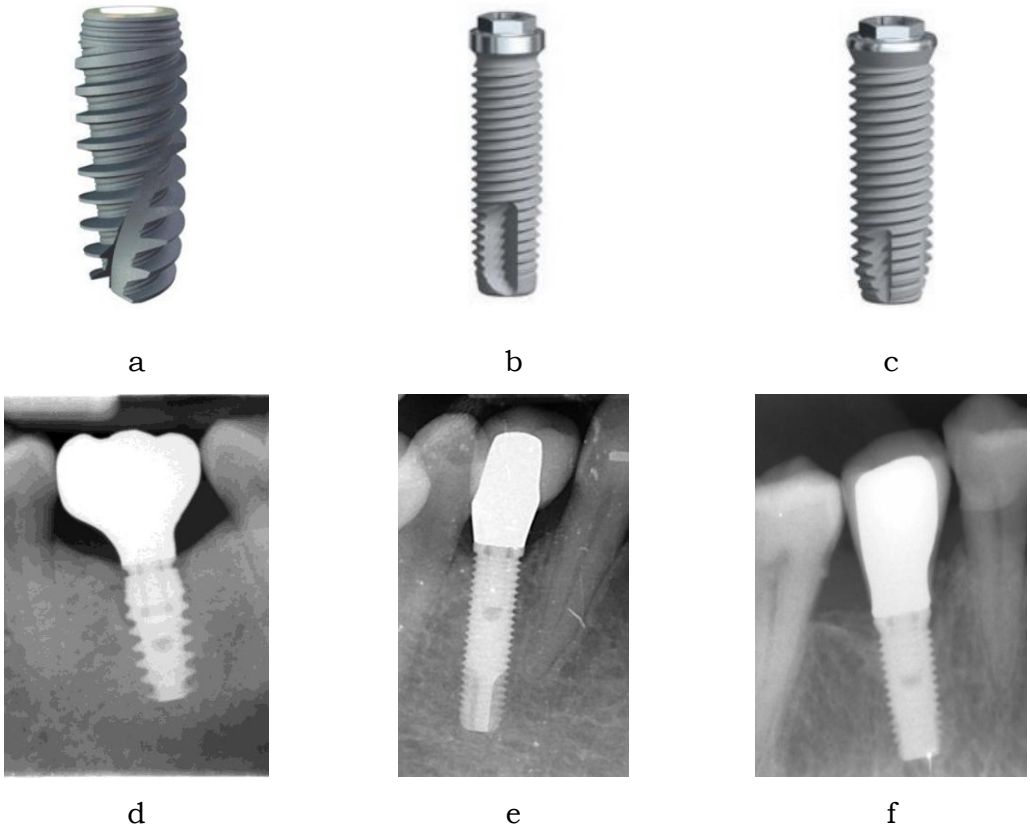
3 Methodology

This section outlines the steps followed for the development of the dental classifier, including the collection, preparation and analysis of the dataset employed. The technical and visual features of the three implant models will be introduced. Furthermore, the criteria employed for the annotation and labelling of images will be explained, as well as the preprocessing techniques applied. Finally, the deep learning models will be described, specifying the architecture and the parameters employed during both training and validation phases.

3.1 Dataset description

The dataset employed in this work is composed of a total of 914 periapical radiographs from adults having at least one implant. In detail, 478 images contain a single implant, 363 images contain two implants, 70 images have three implants, and 3 images contain four implants. Therefore, the set includes a total of 1,426 implants. For each image, segmentation masks for the implants are available, as well as the brand and diameter labels. Having this allows to use the dataset for image classification tasks.

The implants included in the study belong to three different models from the brand Nobel Biocare®: Nobel Active® (**Figure 11a, 7d**), Nobel Brånemark System® MK III (**Figure 11b, 7e**) and Nobel Brånemark® System MK IV (**Figure 11c, 7f**). Each of the models present slightly different technical characteristics, which are described in **Table 2**. As shown in **Table 3**, the distribution by brand shows a clear predominance of the MKIII and ACTIVE models, while the MKIV model is present in a smaller proportion.



a b c

d e f

Figure 11. *Implant models included in the study.* ***a****, Nobel Active®.* ***b****, Nobel Brånemark® System MK III.* ***c****, Nobel Brånemark® System MK IV.* ***d****, Cropped periapical radiograph of Nobel Active®.* ***e****, Cropped periapical radiograph of Nobel Brånemark System® MK III.* ***f****, Cropped periapical radiograph of Nobel Brånemark® System MK IV*

Technical features	**ACTIVE**	**MK III**	**MK IV**
Material	CP Titanium	CP Titanium	CP Titanium
Level	Bone Level	Bone Level	Bone Level
Connection type	Internal	External	External
Connection Shape	Hexagon	Hexagon	Hexagon
Head Shape	Straight	Wide	Flared
Head Microthreads	Yes	No	No
Body Shape	Tapered	Straight	Tapered apex
Body Threads	V shaped	V shaped	V shaped
Apex Shape	Flat	Flat	Flat
Apex Hole	No hole	No hole	No hole
Apex Grooves	Yes	Yes	Yes

Table 2. *Technical features of the implants included in the study*

Implant Brand	Number Of Implants
ACTIVE	706
MKIII	670
MKIV	50

Table 3. *Distribution of implants in the data set according to model.*

Regarding the implant's diameter, the dataset includes eight distinct diameter categories, ranging from 3.00 mm to 5.50 mm, with a significant concentration in the intermediate sizes. **Table 4** shows the frequency of each class.

Diameter (mm)	Number Of Implants
3.00	39
3.30	92
3.50	349
3.75	600
4.00	63
4.30	219
5.00	45
5.50	19

Table 4. *Distribution of implants in the dataset according to diameter.*

To analyze the relationship between brand and diameter, **Table 5** shows how the different diameters are distributed among the models. It can be observed that MKIII is primarily concentrated in 3.75 mm implants, with some cases of 3.3 mm, 4.0 mm, and 5.0 mm implants. ACTIVE has a broader distribution, predominantly in 3.5 mm and 4.3 mm diameters, and additional presence in 3.0 mm, 5.0 mm, and 5.5 mm. MKIV appears exclusively in 4.0 mm implants.

Model/Diameter (Mm)	3.0	3.3	3.5	3.75	4.0	4.3	5	5.5
ACTIVE	39	0	349	0	0	219	44	19
MKIII	0	92	0	600	13	0	1	0
MKIV	0	0	0	0	50	0	0	0

Table 5. *Distribution of implants according to brand and diameter*

It is worth noting that all radiographs were obtained using 3 × 4 cm plates and an average distance of 2cm, with 698 recorded in horizontal orientation and 216 in vertical orientation. Indeed, the average resolution of the images is 156.84 KB, and the average size is 648 x 498 pixels. Furthermore, all images were taken in the same dental clinic.

As explained before, the dataset was initially structured in COCO (Common Objects in Context) format, a standard which has been widely used in object detection and segmentation tasks. The annotations are organized in a .json file containing five principal sections.

- **“licenses”**: list of licenses associated with the dataset images. In this case, the field is void, not specifying any license.
- **“info”**: informative section containing metadata about the dataset, where fields such as ‘contributor’, ‘date_created’, ‘description’, ‘url’, ‘version’ and ‘year’ are included, although in this case they are not filled in.
- **“categories”**: Defines the classes that can appear in annotations. A taxonomy with five classes is defined: ‘**ACTIVE’**, ‘**NO BONE LOSS’**, ‘**BONE LOSS’**, ‘**MKIII’** y ‘**MKIV’**. Each of the classes is identified with ‘**id**’ and ‘**name’**.
- **“images”**: List of images in the dataset. The entries contain metadata about each image of the dataset, including its ‘**file_name’**, ‘**width’** and ‘**height’**. There are additional fields such as ‘license’, ‘flickr_url’, ‘coco_url’, and ‘date_captured’, but they are void.
- **“annotations”**: Contains the annotations associated with the images. Each annotation is composed of an ‘**id’**, the ‘**image_id**’ to which it corresponds, the ‘**category_id**’, the segmentation mask in the form of a list of polygons ‘segmentation’, the area of the object ‘**area**’, the bounding box ‘**bbox**’, a grouping flag ‘**iscrowd**’ and an attribute dictionary ‘attributes’ that includes additional data such as ‘**diameter**’ and ‘**length**’.

This standardized structure facilitates data manipulation and analysis for computer vision applications, especially in image classification contexts such as the present one. **Figure 12** shows an example of an annotated image.

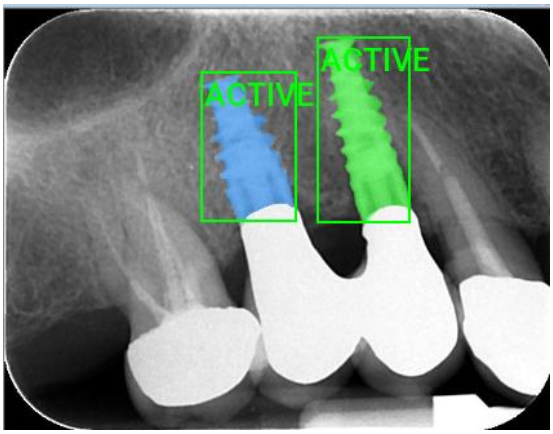


Figure 12. *Example of an annotated periapical radiograph*

3.2 Feature Engineering

Before training the classification models, an automated pipeline for image processing and morphological and geometrical feature extraction was developed. This phase enriches the model’s input with anatomical keypoints obtained from the segmented masks of the model. The process is summarized in **Figure 13**.

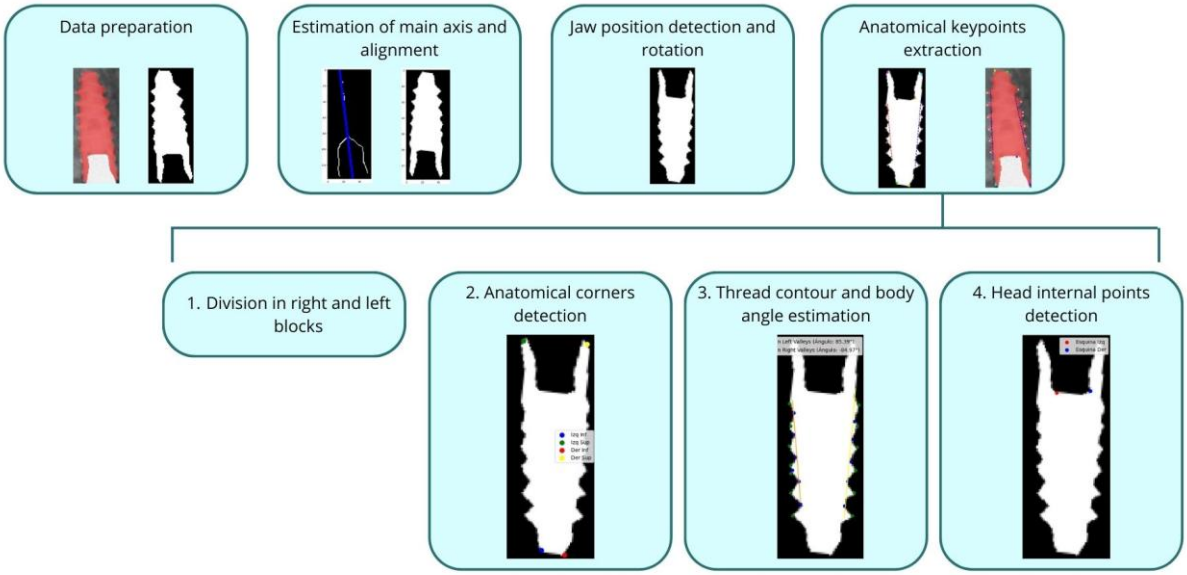


***Figure 13**. Anatomical keypoints extraction pipeline*

3.2.1 Data preparation

As explained before, the periapical radiographs contained labelled implants by means of segmentations in COCO format. For each image, the segmentation masks corresponding to implants with labels "ACTIVE", "MKIII" and "MKIV" were extracted. The original image was cropped using the bounding box with a small additional margin. In this way, there was an image for every implant. Then, binary masks were generated from the corresponding segmentations to these cropped images.

3.2.2 Estimation of main axis and alignment

A morphological skeleton was generated for every binary mask, by using the *thinning* function from OpenCV library. On every skeleton, a Principal Component Analysis (PCA) was applied to estimate the principal axis of the implant and its angular orientations. This angle was used to rotate the implants, leaving them aligned in relation to the image.

3.2.3 Jaw position detection

In this part, the function for extracting upper and lower right and left corners of the implant was used in a preliminary manner, although after it is applied again in its final form. These points were extracted to estimate the distances between left and right corners in the apex and right and left corners in the head of the implant. Based on these distances, it was estimated if the implant belonged to the upper jaw or the lower jaw. If the horizontal distance between the head keypoints was greater than the distance between the apex keypoints, it was assumed that the location was in the lower jaw. In the opposite case -implant being in upper jaw-, the image was rotated 180 degrees so that all masks have the same orientation for posterior analysis.

Once all the masks of the implants were in the same orientation, the contour of the implant head was analysed to determine the shape of the mask. Indeed, two types were distinguished. The first type had a central concavity in the crown interface with respect to the lateral edges. The second type had a straight, horizontal contour at the top of the implant, in direct contact with the crown. An additional alignment step was made just for the second type: the top

boundary of the mask was corrected by identifying the first white pixel in each column (above a predefined threshold) and a straight light was fitted. The slope of the line was adjusted to make a final rotation, leaving the upper contour perfectly horizontal.

3.2.4 Anatomical keypoints extraction

At this point, all the masks were already aligned and in the same orientation. Then, it was possible to proceed to anatomical keypoints extraction from the mask. This step was important for capturing the structural morphology of the implant, making possible its posterior use as input in tabular or multimodal models. The procedure was carried out as follows:

1. **Division in left and right blocks**

The mask was vertically divided into two equal halves. This division allowed a symmetrical analysis of the implant's geometry on each side.

2. **Anatomical corners detection**

For each block of the mask, two keypoints were detected:

- Head corner: The white pixel with the lowest y-coordinate (highest), inside a tolerance of ±5 pixels from the absolute minimum.
- Apex corner: The white pixel with the highest (lowest) y-coordinate, within the same tolerance.

This detection was restricted to the main contours of each block and allowed the anatomical extremes of the implanted body to be precisely defined. The principal contours were obtained from the mask with the function *findContours* from OpenCV. Thus, four keypoints were defined in total.

3. **Body contour extraction and angle estimation**

From the obtained contour for each block, the function *find_peaks* from SciPy.signal library was used to determine the internal and external points of the thread. Each block's contour was treated as a one-dimensional signal function representing the implant thread surface. This analysis identified:

- Thread external points: corresponding to local maxima of the function.
- Thread internal points: corresponding to local minima of the function.

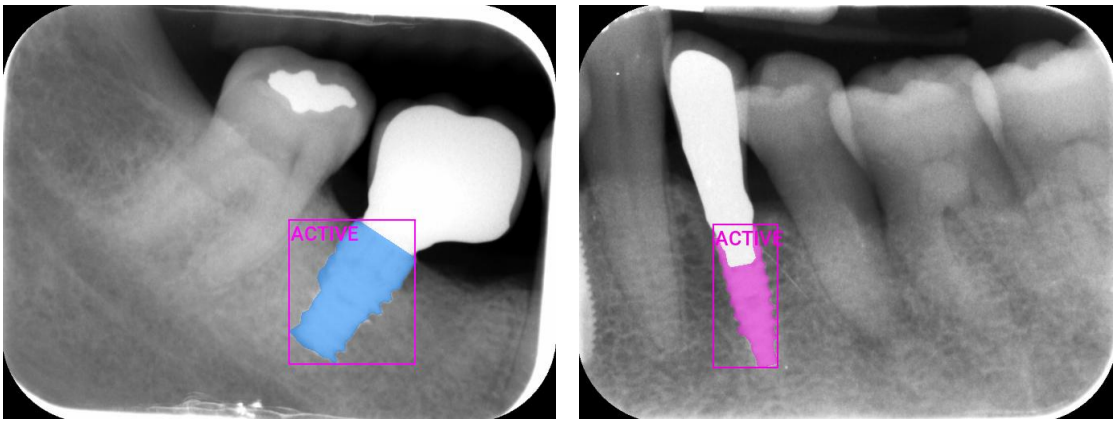
To avoid erroneous points generated by the noise of the mask contour, a minimum distance between adjacent points was determined in proportion to the image width. This guaranteed that only meaningful variations of the thread were detected in the profile.

The points classified as internal points of the thread were used to adjust a linear regression fit on each block, by using LinearRegression function from *sklearn.* The coefficients (slope and intercept) were obtained for each side. The angle of inclination of the line with respect to the horizontal axis was calculated, allowing the implant's body angle to be characterized at both sides.

4. **Head internal points detection**

As explained previously, some masks had a central concavity in the crown interface with respect to the lateral edges, which can be observed in **Figure 14b**. This characteristic is due to the fact that certain implants have an internal connection with the crown, and on radiographs, this connection can appear as an extension into the implant, generating a concave shape on the segmented

mask. However, not all radiographs of implants with an internal connection show this extension, instead they present a uniform shape without the central concavity, as can be seen in **Figure 14a**.



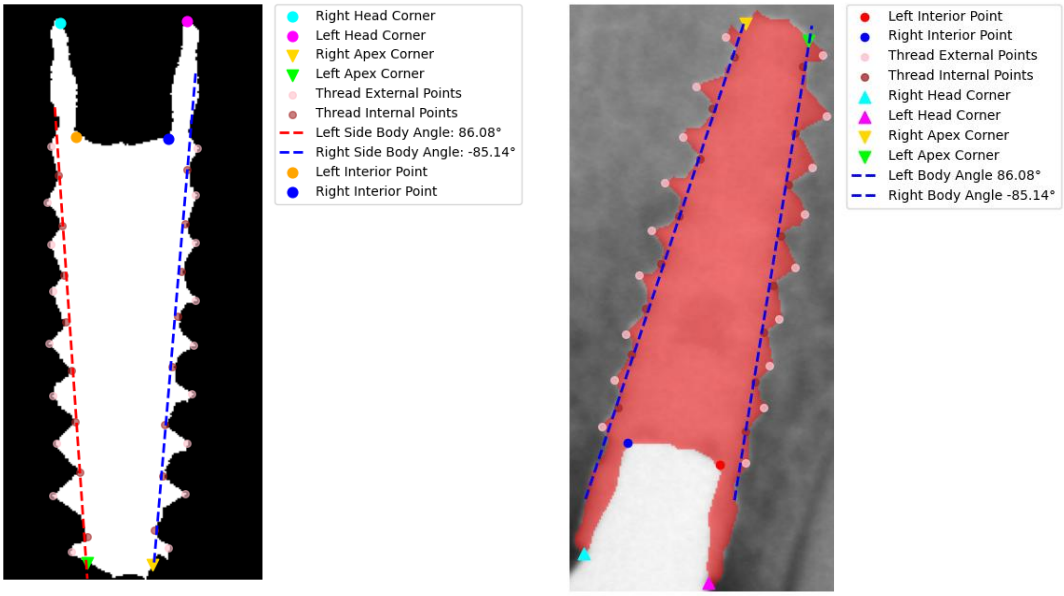
a b

Figure 14. *Types of Segmentation Masks.* ***a****, Straight.* ***b****, With Concavity*

For the masks showing this concavity, a specific procedure was applied to detect the internal points. These points are relevant to measure the diameter of the implant.

To obtain them, the vertical profile of the upper part of the binary mask (corresponding to the upper 30% of the implant) was analyzed. Therefore, a function representing the head profile is obtained. After identifying the two local maxima, which correspond to the lateral edges, the derivative of the function is calculated in the region between them. The minima of this derivative indicates abrupt changes in the profile slope and are interpreted as the two internal inflection points, reflecting a morphological concavity in the implant head. These points are then translated into the original image coordinates. Apart of this, for all implants with a straight-shaped upper mask, the internal points were defined as the previously extracted head corners.

Finally, all the points are obtained. **Figure 15a** shows all the anatomical keypoints in the aligned mask and **Figure 15b** shows all the anatomical keypoints in the original radiograph.



a b

Figure 15. *Anatomical Keypoints Extraction. a, Keypoints in aligned mask. b, Keypoints in the original image*

3.2.5 Radiographic distance estimation

After this, the distance in cm between the X-ray source and the implant plan was calculated. To do this, the **implant diameter in pixels** was first obtained by measuring the horizontal distance between the keypoints associated with "left interior point" and "right interior point". It is important to remember that in the implants with straight-shaped upper mask, the internal points were defined with the same coordinates as the head corners.

On the other hand, the **pixel to centimetre conversion value** was calculated from the original dimensions of the radiograph and the physical dimensions of the radiographic film, which is 3 cm x 4cm. If the image was vertical (height ≥ width), the width in cm was assumed to correspond to 3 cm and the height to 4 cm. Otherwise, the film was considered rotated, and the proportions were reversed. Having this information, two partial scales were calculated:

$$pixel\ per\ cm\ (width) = \frac{width\ in\ pixels}{width\ in\ cm} \quad (3),$$

$$pixel\ per\ cm\ (height) = \frac{height\ in\ pixels}{height\ in\ cm} \quad (4)$$

For an image, both values are approximately the same. However, both were calculated, being the final pixel to centimeter conversion value the average of both.

Knowing the real implant diameter in mm, the **expected implant diameter in pixels** was calculated as:

$$expected\ implant\ diameter\ (px) = implant\ diameter\ (mm) \times \frac{pixel\ per\ cm}{10} \quad (5)$$

Finally, the **radiographic distance in cm** was estimated using the inverse rule of three. The clinical assumption was that the radiograph was acquired with a standard distance of approximately 2.0 cm between the X-ray source and the sensor. If at this distance the implant should be projected with an expected

diameter in pixels, but the image appears with other diameter in pixels, then the actual distance is calculated as:

$$radiographic\ distance\ (cm) = 2.0 \times \frac{expected\ implant\ diameter\ (px)}{real\ implant\ diameter\ (px)} \quad (6)$$

This formula is based on a direct proportional relationship: the greater the actual distance between the focus and the implant plane, the smaller the projected implant diameter in the image. This variable provides complementary geometric information that is potentially useful for classification models.

3.2.6 Description of the processed tabular dataset

The cropped images with their corresponding masks were saved in .png format, keeping the alignment between them. All the extracted keypoints were saved in a .csv file with other characteristics, being useful as input for tabular models and as a complement for multimodal models. In this file, each row represents an instance and contains columns such as the **image name**, the **original width and height of the radiograph**, the **pixel to cm** conversion scale, and the **dimensions (width and height) of the bounding box** corresponding to the implant, as well as its **proportion** (the quotient between the width and the height of the box). Furthermore, data associated with each implant is included, such as the **model**, the **diameter in millimeters** and the **radiographic distance in centimeters**. The coordinates of **six key anatomical points** are also included, with the **lateral left and right angles** formed by the body of the **position** of the implant inside the jaw. Thus, the dataset contains a total of 19 columns that have its content summarized in **Table 6**.

Variable	Data Type	Description
image_name	Text	Image name (without file extension).
orig_width	Float	Original image width in pixels.
orig_height	Float	Original image height in pixels.
pixel_per_cm	Float	Estimated pixel-to-centimeter conversion factor.
bbox_width	Integer	Width of the implant bounding box in pixels.
bbox_height	Integer	Height of the implant bounding box in pixels.
implant_bbox_ratio	Float	Aspect ratio of the bounding box (bbox_width / bbox_height).
model	Text	Implant model (e.g., ACTIVE, MKIII, MKIV).

implant_diameter_mm	Float	Nominal implant diameter in millimeters.
radiographic_distance_cm	Float	Estimated distance from the implant to the X-ray source (in cm).
left_top	Tuple (x, y)	Coordinates of the top-left point of the implant.
right_top	Tuple (x, y)	Coordinates of the top-right point of the implant.
left_bottom	Tuple (x, y)	Coordinates of the bottom-left point of the implant.
right_bottom	Tuple (x, y)	Coordinates of the bottom-right point of the implant.
interior_left	Tuple (x, y)	Inner left point within the implant body.
interior_right	Tuple (x, y)	Inner right point within the implant body.
body_angle_left	Float	Body angle at the left lateral (in degrees).
body_angle_right	Float	Body angle at the right lateral (in degrees).
position	Text	Implant anatomical position (Superior or Inferior).

Table 6. *Variable description of the processed dataset*

3.3 Data augmentation

With the aim to tackle the existing class imbalance and improving the robustness of the classification models, data augmentation techniques were implemented on all instances belonging to minoritary classes of diameter (3.0, 3.3, 4.0, 5.0 and 5.5) and/or model (MKIV). The process consisted of the generation of new images from geometrical transformations applied to the original images, their corresponding masks and the anatomical keypoints.

There were 6 transformations applied to instances with minoritarian diameter, which included: horizontal (**Figure 16c**) and vertical flips (**Figure 16b**), 90º (**Figure 16d**) and 270º rotations (**Figure 16e**), translations within 5% of the image dimensions (**Figure 16f**), as well as random rotations with a range of ±15º (**Figure 16g**). All of these transformations were designed ensuring geometrical coherence between the image, the segmentation mask and the anatomical keypoints, which were transformated using affinity matrixes or specific geometric rules. For instances belonging to MKIV model, 4 additional combinations of transformations were performed: vertical flip followed by 90º

rotation (**Figure 17a**), vertical flip followed by 270° rotation (**Figure 17b**), horizontal flip followed by random rotation (**Figure 17c**) and vertical flip followed by translation (**Figure 17d**).

After each transformation, new instances were saved with a distinctive suffix for identification, and the .csv file was updated with the new coordinates. In essence, this process allowed to increment the representation of minoritarian classes, contributing in this way to a better generalization of the deep learning models.

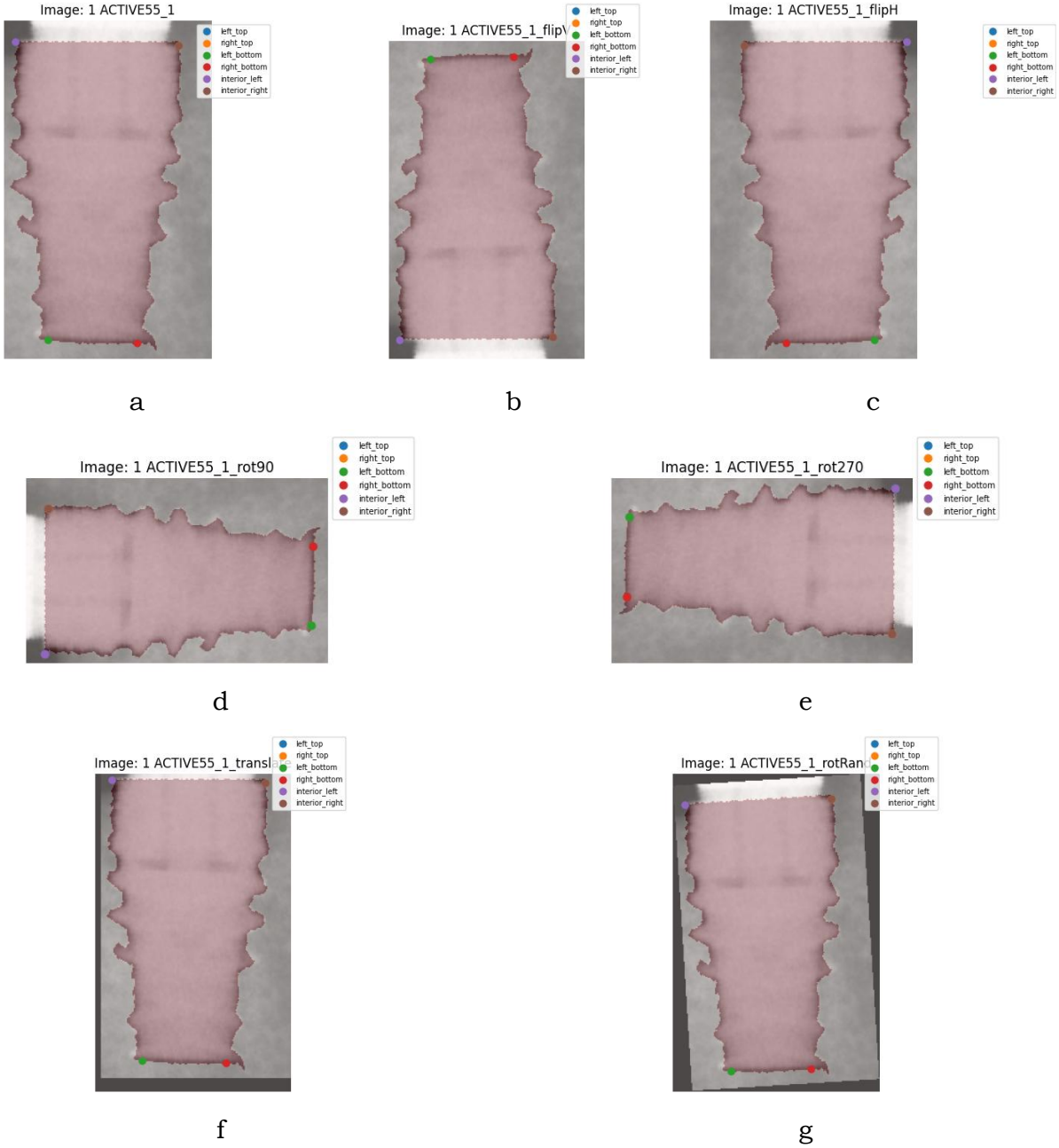


Figure 16. *Example of data augmentation transformations applied to a periapical radiograph, including mask overlay and anatomical landmarks.* ***a****, Original image before transformation.* ***b****, Vertical flip.* ***c****, Horizontal flip.* ***d****, 90º rotation.* ***e****, 270º rotation.* ***f****, Translation.* ***g****, Random rotation*

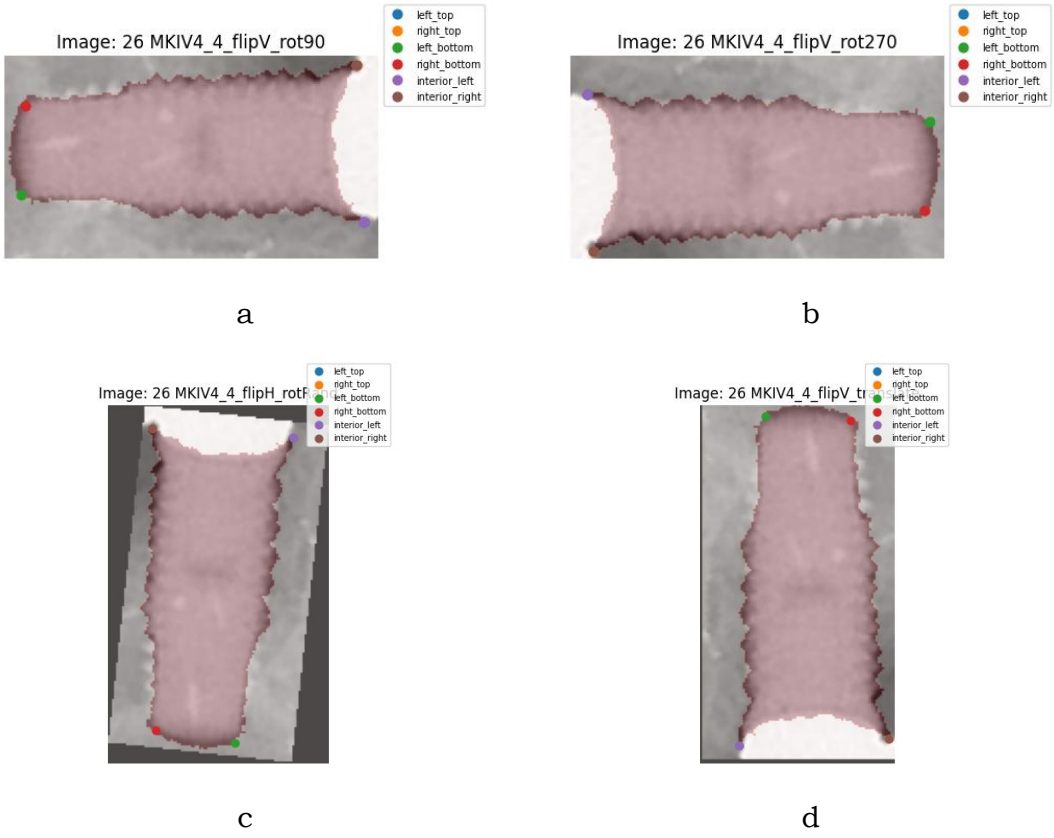


Figure 17. *Additional data augmentation transformations for MKIV model.* ***a****, Vertical flip followed by 90° rotation.* ***b****, Vertical flip followed by 270° rotation.* ***c****, Horizontal flip followed by random rotation.* ***d****, Vertical flip followed by translation.*

As a result of data augmentation, a dataset with average size of 115 x 169 pixels and an average image resolution of 12,04 KB was generated. The final dataset is composed by a total of 3174 images distributed across different implant models: MKIII (n=1342), ACTIVE (n=1282) and MKIV (n=550). Regarding the diameters, the set presents a more balanced distribution with an improved representation of the minority classes: 3.00 mm (n=273), 3.30 mm (n=644), 4.00 mm (n=641), 5.00 mm (n=315) y 5.50 mm (n=133).

3.4 Model architectures

Having the dataset already augmented and structured, next step consisted of developing classification models with the objective of predicting the model and diameter of the implants.

Given that the input comes from heterogeneous sources—radiographic images and tabular variables derived from geometric analyses—a series of multimodal architectures was proposed. These architectures allow the integration of both types of data. This approach seeks to leverage the complementary advantages of deep visual representations and numerical features.

Diverse model architectures are described below. They combine different visual feature extractors (ResNet18, EfficientNet-B2) with tabular branches (TabNet), and also incorporate Graph Neural Networks (GNNs) to model spatial relationships between anatomical keypoints. In all cases, a late fusion strategy was applied. This consisted of concatenating the latent representations from the branches before branching into two independent classification heads. These

architectures were trained in a multi-task learning framework, jointly tuning the model parameters for both tasks. The following subsections detail not only the design of each approach, but also the training procedures used.

3.4.1 Multimodal Models

3.4.1.1 EfficientNetB2 + GNN

The model combines three different types of data: images, segmentation masks and scalar tabular features. Specifically, it uses a CNN for feature extraction, a GNN for processing the anatomical keypoints, and a fully connected layer to incorporate the scalar features.

The model uses **EfficientNetB2** as backbone, for extracting features from the images. EfficientNetB2 is a CNN pretrained on ImageNet. Therefore, the images are resized to 260 x 260 pixels and normalized with ImageNet parameters. The final fully connected layer of EfficientNetB2, which normally is used for classification, is removed and substituted by a fully connected layer of 128 units activated by ReLU. To avoid overfitting, a Dropout of 30% is applied. The images are segmented using the binary masks, which allow that the model focus specifically on the implant region, eliminating irrelevant information.

The substitution of the final layer by the fully connected layer is done to retain the feature maps generated from the convolutional layers. These features maps are then processed by a fully connected layer to generate a compressed image embedding, which captures the high-level visual features of the implant

In addition to image-based features, the model includes a **GNN** to capture spatial relationships between the anatomical keypoints of the dental implant. These keypoints, whose obtainment is explained in Section 3.2, represent relevant positions of the implant structure:

- left_top: upper left corner of the implant.
- right_top: upper right corner of the implant.
- left_bottom: lower left corner of the implant.
- right_bottom: lower right corner of the implant.
- interior_left: inner left point in the implant body.
- interior_right: inner right point in the implant body.

These keypoints are treated as the nodes of a graph. The edges of the graph are defined by connecting every pair of keypoints, in this way forming a fully connected graph. The GNN, process the features of the nodes by aggregating information from the neighbouring nodes with the GCNConv layers. This allows the model to learn the geometrical structure of the implant. Finally, the output of the GNN is grouped with the global mean pool operation, which allow the features at the node level transform into a single feature vector representing the spatial structure of the implant.

The model also incorporates scalar characteristics derived from tabular data. These characteristics give contextual and geometrical information about the implant:

- distance_cm
- pixel_per_cm
- bbox_width and bbox_height
- implant_bbox_ratio
- angle_left_valley and angle_right_valley

These scalar features are processed with a **fully connected layer** that generates a scalar embedding, which is then concatenated with the image and graph features.

Once the features from the three modalities (image, graph and scalar features) are obtained, they are concatenated in a single feature vector. This combined vector is passed through a fully connected layer of 128 units with a ReLU and Dropout activation to reduce overfitting. The final output consists of two separate heads: one for brand classification and one for diameter classification. Each head consists of a fully connected layer that predicts the class probabilities for brand and diameter, respectively. Furthermore, each head is trained separately with its own loss function, using weighted Focal Loss to reduce class imbalance. This is important for the diameter prediction, given the existing class imbalance. The entire model architecture is described in **Figure 18**.

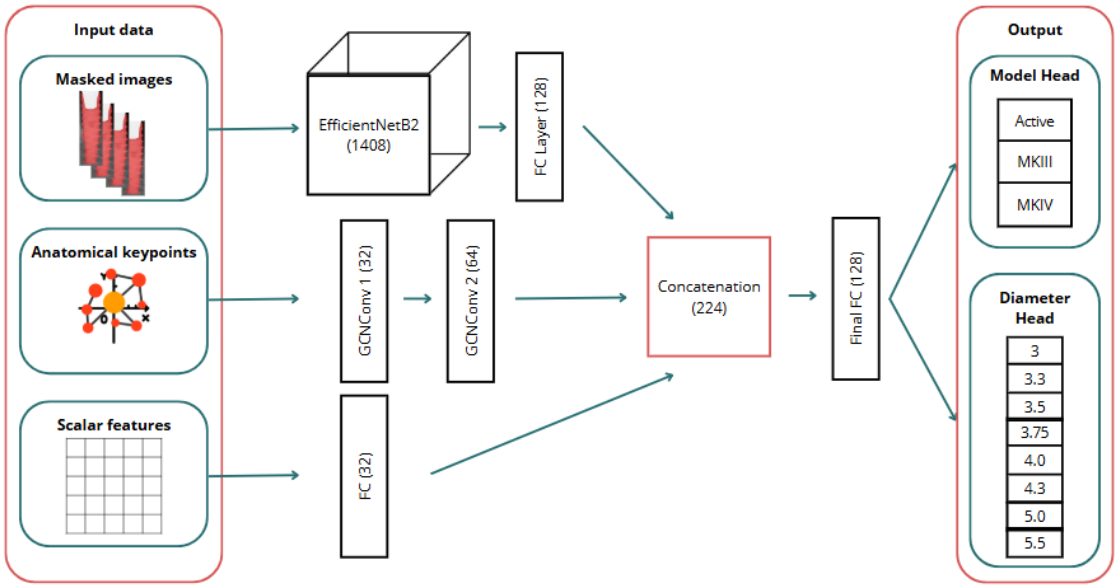


Figure 18. *Model Architecture: EfficientNet + GNN.*

3.4.1.2 ResNet-18 + GNN

This model is based on two principal components, which are the visual feature extractor using ResNet-18 as backbone, and a geometrical feature processing based on a GNN. It also incorporates scalar features derived from tabular data. The structure is equivalent to the GNN + EfficientNetB2 (described previously in Section 3.4.1.1), but in this case ResNet-18 is used instead of EfficientNetB2.

As the visual component is based on **ResNet-18**, which is a CNN architecture pretrained on ImageNet, the images are resized to 224x224 pixels and normalized according to ImageNet parameters. Before passing the images to the ResNet-18 network, the images are multiplied by the segmentation masks. This step is done to make sure that the model only pays attention to the implant region. Like in the previously explained model, the ResNet-18 network is modified to substitute its final classification layer with a fully connected layer with 128 units, activated by ReLU. This modification allows the model to take advantage of the high-level features learned without doing a direct classification.

The GNN and feature fusion follow the same procedure described in Section 3.4.1.1 for the EfficientNetB2 + GNN model. As in the previous architecture, the GNN processes the spatial relationships between the implant's anatomical

points, and its outputs are combined with the image-extracted features and scalar features. This feature fusion is passed through a fusion layer, which then feeds the classification heads for implant brand and diameter. In this way, the structure and feature fusion process remain consistent with the previous model. The whole model architecture is shown in **Figure 19**.

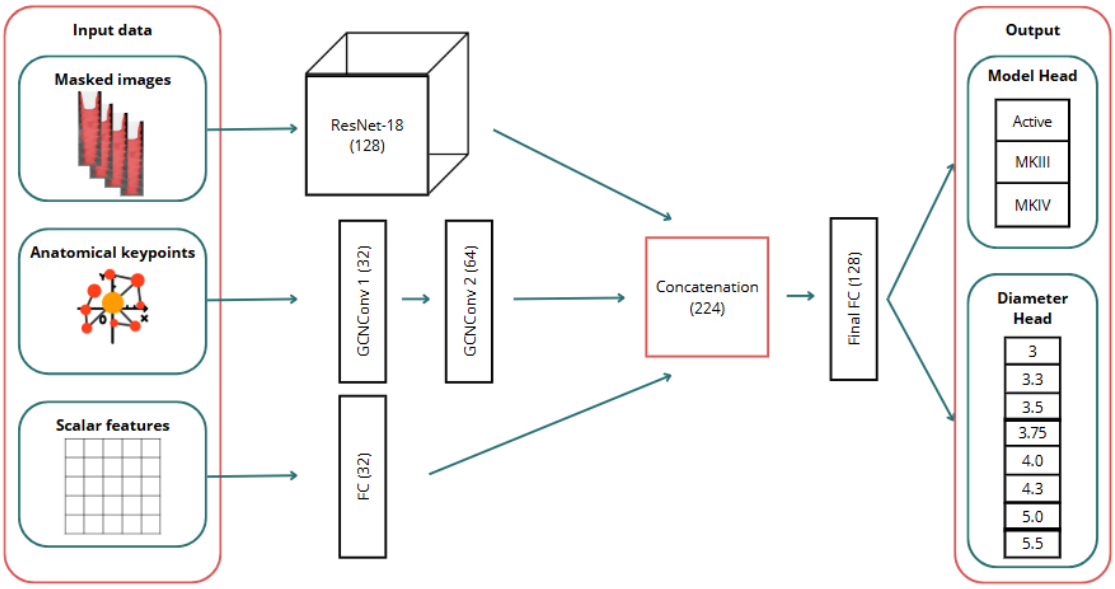


Figure 19. *Model Architecture: ResNet-18 + GNN*

3.4.1.3 EfficientNetB2 + TabNet

In this model, a visual backbone based on EfficientNetB2 is combined with TabNet for processing visual features. The visual branch follows the same approach of GNN + EfficientNetB2 model (previously described in Section 3.4.1.1.) both for image preprocessing and the network tuning.

The main difference with the other models described is the use of TabNet, which uses for processing all the tabular characteristics derived from the implant. TabNet is a sequential attention-based model that allows learning hierarchical relationships between attributes through its decision steps and attention masks. Unlike previous models that used GNNs to process the spatial relationships of anatomical landmarks, TabNet centers exclusively on numerical features, not having the need for graph processing.

In this model, the tabular features of the dataset are preprocessed using StandardScaler. This process is intended to normalize the numerical features, eliminating any bias caused by differences in value scales. First, the original dataset was loaded from the CSV file and the 'position' column was removed, as it was not used in the model. Columns that should not be scaled, such as 'image_name', 'brand', 'implant_diameter_mm', and the coordinates of anatomical points (e.g., 'left_top', 'right_top', etc.), were excluded from the normalization process.

The remaining numeric columns were identified and scaled using the StandardScaler, which scales the data to have a mean of 0 and a standard deviation of 1. This preprocessing is crucial to ensure that the tabular features are on a similar scale and do not disproportionately influence the model's performance, especially when combined with the visual features extracted by EfficientNetB2.

Then, the TabNetEncoder is configured with 64 decision units, 3 decision steps, and an attention hyperparameter of 1.5, allowing the model to capture complex patterns in the tabular features in an efficient manner. The whole model architecture can be observed in **Figure 20**.

The fusion of visual features extracted by EfficientNetB2 and tabular features processed by TabNet follows the same approach as in the previous models, where both sources of information are concatenated and passed through a fusion layer before being processed by the classification heads for implant brand and diameter.

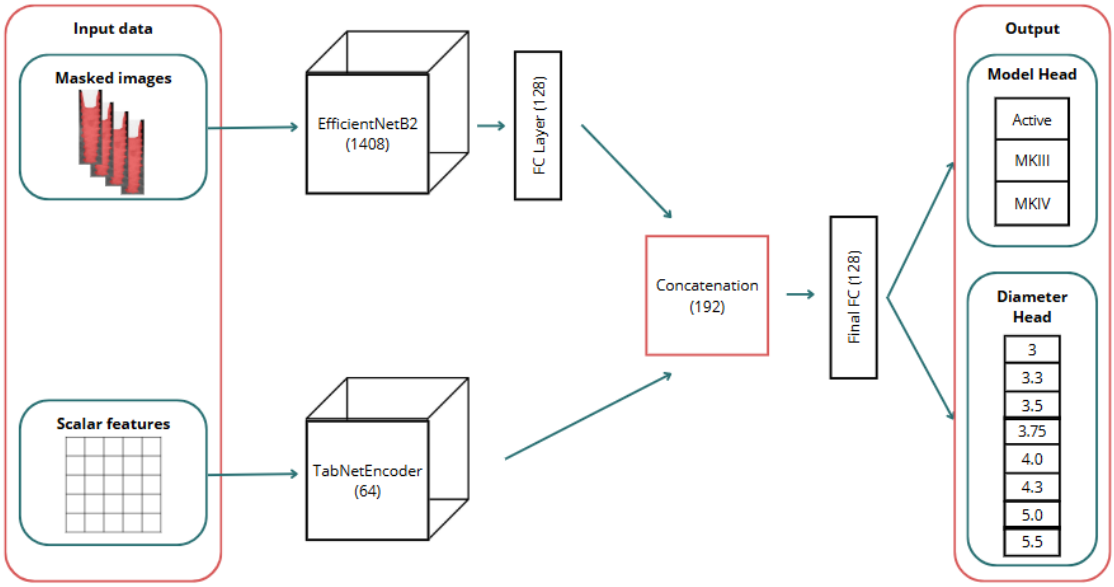


Figure 20. *Model Architecture: EfficientNetB2 + TabNet*

3.4.2 Single-modality models

In this section, two models that process a single type of data are presented: images with corresponding segmentation masks or tabular data. These single-modal models serve as a reference for the multimodal models previously described above.

3.4.2.1 EfficientNetB2

This model focuses solely on image processing and segmentation masks for implant classification. The model takes radiographic images that have been preprocessed to focus exclusively on the implant region using binary masks. These masks allow the model to focus on relevant parts of the image, eliminating irrelevant areas and improving classification performance.

The visual backbone is composed of EfficientNetB2, a convolutional neural network pre-trained on ImageNet, which extracts high-level features from the images. The final classification layer of EfficientNetB2 is replaced by a 128-unit fully connected layer with ReLU activation, and a 30% dropout is applied to prevent overfitting. The architecture can be observed in **Figure 21**.

The extracted features are finally processed in two classification heads: one to predict the implant brand and the other to predict the diameter. Both heads are trained using Weighted Focal Loss.

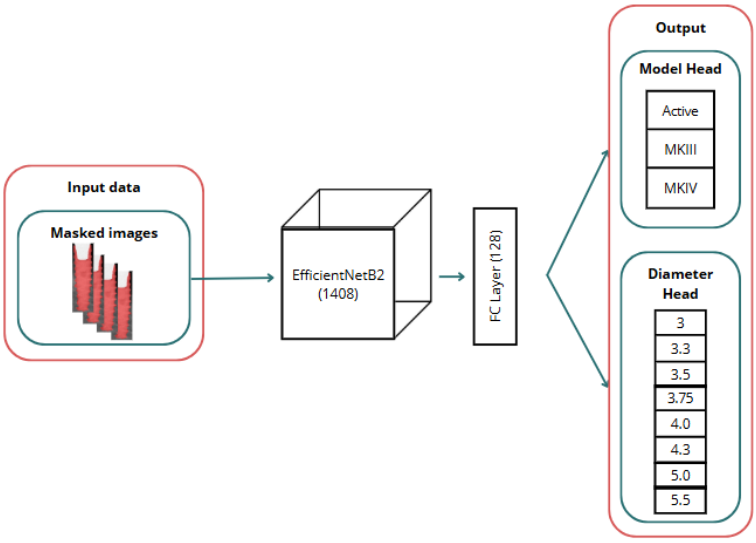


Figure 21. *Model Architecture: EfficientNetB2*

3.4.2.2 TabNet

This model processes only tabular data derived from the geometric characteristics of implants. Unlike multimodal models, TabNet processes only numerical data, without incorporating visual information. Before passing the input to TabNet, the numerical features are scaled following the same approach as in Section 3.4.1.3.

TabNet uses a sequential attention-based approach to learn complex relationships between tabular features. Unlike previous models that use GNNs or CNNs, TabNet does not require image processing, only numerical features.

The TabNet model's configuration include 64 decision units, 3 decision steps, and an attention hyperparameter of 1.5. Similar to the EfficientNetB2 model, the TabNet model has two classification heads: one for predicting the brand and one for predicting the implant diameter, each of them using weight focal loss to address class imbalance.

3.4.3 Training and Validation Settings

The dataset was divided into 80% for training and 20% for validation using a stratified partition with the aim to have a proportional representation of all classes in both partitions. This division was employed in a consistent manner in all the evaluated architectures, so that all models were trained and validated under the same instances. To guarantee reproducibility, a fixed random seed was used during partitioning process.

All architectures were trained during 10 epochs using the Adam optimizer with a learning rate of 1e-4 and a batch size of 16. With the objective of preventing overfitting and improve generalization, an early stopping strategy based on validation loss was employed, with a patience of three epochs.

In addition, regularization techniques such as the use of Dropout layers (with a ratio of 0.3) were applied in the dense blocks. Given the class imbalance present in the target variables, class weights were incorporated for both brand and implant diameter classification, and the Focal Loss function was used to penalize errors in minority classes.

3.5 Evaluation Metrics

In this section the metrics used to evaluate the performance of the brand and diameter classification models are described. The metrics include precision, recall, F1 score, and accuracy. These are commonly used in classification tasks to give a complete view of model performance, allowing to assess the model's capacity to handle imbalanced classes and effectivity in general classification tasks.

Precision measures the proportion of correct positive predictions among all positive predictions made by the model. In other words, it measures the accuracy of positive predictions. Thus, a high precision indicates that the model is making few false positive predictions. Mathematically, it is expressed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

Where:

- TP (True Positives) represents the number of instances correctly classified as positives.
- FP (False Positives) represents the number of instances incorrectly classified as positives.

Recall measures the model's ability to correctly identify true positive predictions among all actual positive instances in the dataset. It indicates how many of the actual positive instances the model is able to correctly identify. Its formula is:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

Where:

- FN (False Negatives) represents false negatives, that is, positive instances that the model did not correctly identify.

F1-score is the harmonic mean of precision and recall. It provides a single metric that combines both precision and recall, balancing the trade-off between the two. This metric is useful for evaluating the performance of a model in imbalanced datasets, where precision and recall may be skewed. It is calculated as:

$$F1\_score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

Accuracy is one of the most common metrics and measures the proportion of correct predictions out of the total number of predictions made. Even though it is an intuitive metric, its usefulness can be reduced in imbalanced data sets, where the majority class can dominate overall performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Where:

- TN (True Negatives) represent instances correctly classified as negatives.

4 Results

4.1 Model performance

In this section, the performance metrics for the models trained for model and diameter classification are presented. The evaluated models include both multimodal and single-modal configurations. Multimodal models incorporate both image data and tabular data. In this case, GNN + EfficientNetB2, GNN + ResNet-18 and EfficientNet + TabNet architectures are evaluated. Their performance metrics for "Model" and "Diameter" are shown in **Table 7** and **Table 8**, respectively.

Model	Precision	Recall	F1-Score	Accuracy
GNN + EfficientNetB2	0.99	0.98	0.98	0.99
GNN + ResNet-18	0.98	0.98	0.98	0.98
TabNet + EfficientNetB2	0.95	0.97	0.96	0.97

Table 7. *"Model" performance metrics for multimodal models*

Model	Precision	Recall	F1-Score	Accuracy
GNN + EfficientNetB2	0.92	0.92	0.92	0.93
GNN + ResNet-18	0.88	0.87	0.87	0.89
TabNet + EfficientNetB2	0.90	0.90	0.90	0.91

Table 8. *"Diameter" performance metrics for multimodal models*

The single-modal architectures include EfficientNetB2 and TabNet. TabNet model's performance was below 70% in all the tested metrics for both model and diameter classification. Nevertheless, EfficientNetB2 model, which is trained only on image data, achieved a considerably better performance. This is shown in **Table 9** and **Table 10** for model and diameter classification, respectively.

Model	Precision	Recall	F1-Score	Accuracy
EfficientNetB2	0.95	0.97	0.96	0.97

Table 9. *"Model" performance metrics for EfficientNetB2*

Model	Precision	Recall	F1-Score	Accuracy
EfficientNetB2	0.90	0.90	0.90	0.90

Table 10. *"Diameter" performance metrics for EfficientNetB2*

Among multimodal models, GNN + EfficientNetB2 shows the best global performance including model and diameter classification together, across all metrics. For model classification, GNN + ResNet-18 shows slightly lower performance than GNN + EfficientNetB2. The same can be said for TabNet + EfficientNetB2 for diameter classification.

Still, the diameter classification performance of the single-modal model EfficientNetB2 is superior to the one achieved by GNN + ResNet-18 and very similar to the one achieved by GNN + EfficientNetB2. For model classification,

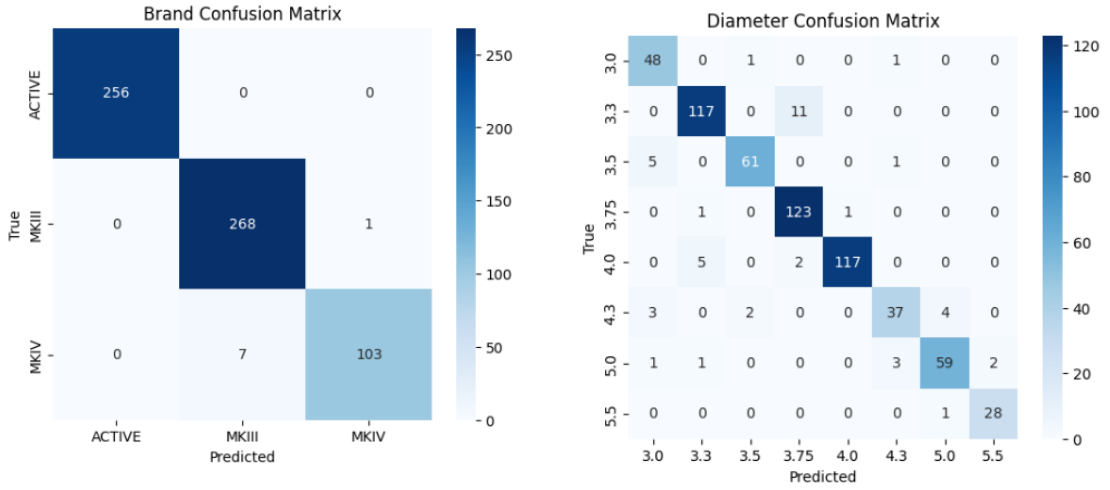
it has one of the lowest metrics, which are equal to the ones obtained with TabNet + EfficientNetB2.

To sum up, all multimodal models outperformed the EfficientNetB2 single-modal for model classification. Actually, GNN + EfficientNetB2 had the best performance for both model and diameter classification tasks. To have a more detailed view of how the model evolves during training, the train loss and validation accuracy graphs for both tasks have been included in **Appendix A**: EfficientNetB2 + GNN Training Figures.

4.2 Visualizations

4.2.1 Confusion Matrixes

As indicated in the previous subsection, the GNN + EfficientNetB2 architecture shows superior performance for model classification and one of the best performance for diameter classification. Its confusion matrices for brand and diameter are shown in **Figure 22a** and **Figure 22b**, respectively. The confusion matrix of the rest of the models are shown in **Appendix B**: Confusion Matrixes.



a b

Figure 22. *GNN + EfficientNetB2 confusion matrixes.* ***a,*** *model confusion matrix.* ***b,*** *diameter confusion matrix*

To give a more global view on the model performance according to its different classes, the detailed classification reports are shown in **Figure 23**.

```
=== BRAND CLASSIFICATION REPORT ===
              precision    recall  f1-score   support

      ACTIVE       1.00      1.00      1.00       256
      MKIII        0.97      1.00      0.99       269
      MKIV         0.99      0.94      0.96       110

    accuracy                           0.99       635
   macro avg       0.99      0.98      0.98       635
weighted avg       0.99      0.99      0.99       635
```

a

```
=== DIAMETER CLASSIFICATION REPORT ===
              precision    recall  f1-score   support

         3.0       0.84      0.96      0.90        50
         3.3       0.94      0.91      0.93       128
         3.5       0.95      0.91      0.93        67
        3.75       0.90      0.98      0.94       125
         4.0       0.99      0.94      0.97       124
         4.3       0.88      0.80      0.84        46
         5.0       0.92      0.89      0.91        66
         5.5       0.93      0.97      0.95        29

    accuracy                           0.93       635
   macro avg       0.92      0.92      0.92       635
weighted avg       0.93      0.93      0.93       635
```

b

Figure 23. *GNN + EfficientNetB2 classification report.* ***a,*** *model classification report.* ***b,*** *diameter classification report*

On one hand, regarding the brand classification, it can be seen in **Figure 22a** and **Figure 23a** that the model classified correctly all “ACTIVE” instances. For "MKIII" class, its precision is 0.97 and its recall is 1.00, indicating that the model was able to identify all implants in this class (except 1), but misclassified a few instances which are false positives. On the other hand, for the "MKIV" class, the precision achieved is 0.99 and the recall is 0.94. Both the macro average and weighted average are 0.99, demonstrating consistent and balanced performance across classes. Overall, the model achieves an accuracy of 0.99 for model classification, demonstrating excellent performance.

On the other hand, diameter classification shows solid performance despite having more variability between classes compared to model classification. It can be observed that in the most common classes, such as “3.75” and “4.0”, the obtained metrics have the highest F1-score metrics, suggesting that the model correctly identifies instances with these majority diameters in most cases. It should also be noted that the lowest F1-score metrics are obtained for the less represented classes, such as "3.0" and "4.3".

In general, it can be observed that the F1-score values are around 0.93–0.97 for the most represented classes. In contrast, the minority classes perform less well, with class “4.3” for example, being the only class with an F1-score below 0.90. However, as mentioned, the F1-score is high for most classes, exceeding 0.90 for 7 out of 8 classes.

4.3 Model training time

Model training times were evaluated on a system with the following hardware and software specifications:

- Operating System: Windows 10
- Python Version: 3.11.9
- Processor: Intel Core i7-1165G7 (Intel64 Family 6 Model 140 Stepping 1, Genuine Intel)
- RAM: 16,0 GB
- Number of Physical Cores: 4

Training was done using a CPU, having as a consequence longer training times in comparison to using a GPU. CPUs, being sequential processors, are not as efficient as GPUs in deep learning tasks that require massive parallelization. Even though the process was slower, the model's performance in terms of accuracy and other metrics was not compromised. In **Table 11.** Training times of all models, the training times for each architecture can be observed.

Model	Total Training Time	Average Epoch Time
GNN + EfficientNetB2	1h 44 min 7s	10 min 25 s
GNN + ResNet-18	47 min 23 s	4 min 44 s
TabNet + EfficientNetB2	1 h 48 min 9s	10 min 49 s
EfficientNetB2	1h 23 min 15 s	8 min 19 s

Table 11. *Training times of all models*

The model with the longest training time is TabNet + EfficientNetB2, which is closely followed by the GNN + EfficientNetB2 model. These results are consistent with the additional computational complexity that has combining convolutional backbones with other components. Nevertheless, GNN + ResNet-18 is the most computationally efficient, as ResNet-18 has lower computational demand than EfficientNetB2.

5 Discussion

CNNs have demonstrated remarkable effectiveness in the domain of image classification tasks, having the potential to recognize implant types and characteristics in dentistry. This work confirms that, as all the evaluated architectures reached metrics close or above 90% in precision, recall, F1-score and accuracy. This phenomenon happened both in image unimodal architecture and multimodal architectures, confirming the feasibility of combining periapical radiographs with structured information as input.

One of the objectives of the study was to evaluate the impact of MMDL in model performance. The results show that the combination of masked images, tabular data and graph networks improves performance compared to unimodal image-only models. Nevertheless, this effect is not observed when TabNet is used in multimodal settings, since its performance is slightly inferior to the image-only model one. This finding suggests that integrating spatial relationships between relevant anatomical keypoints provides relevant information that have the potential to improve classification, whereas conventional use of tabular data may not be sufficient.

In the evaluated configurations, EfficientNetB2 backbone had better performance in comparison to ResNet-18. This behaviour is consistent with the literature, as EfficientNet employs a compound scaling strategy that optimizes the network's depth, width, and resolution, allowing for more efficient and accurate feature extraction. Consequently, it can be inferred that EfficientNet is more suitable for implant classification tasks in periapical radiographs, at least in contexts with datasets with similar characteristics to the one used in this study.

With regard to the model classification task, it is observed in all architectures that the vast majority of errors happen between models “MKIII” and “MKIV”. This confusion can be explained by the fact that these models belong to the same implant system (Nobel Brånemark) and they share very similar morphological characteristics. Indeed, Nobel Brånemark MKIV is the fourth generation and a direct evolution of Nobel Brånemark MKIII model, which corresponds to the third generation. In contrast, “ACTIVE” implants present a more distinctive design, resulting very rare for an “ACTIVE” implant to be misclassified as “MKIII” or “MKIV”.

It should be highlighted that despite the similarity between MKIII and MKIV, the results obtained have been remained above 90% in all cases for class-specific metrics. This performance reinforces the model's robustness and allows to infer that if implants from other brands (with more distinct morphological characteristics) were included in the dataset, the architecture will be able to classify them accurately.

Furthermore, it can be observed that implant model classification task was easier for all classifiers compared to the diameter estimation task. This is evidenced by the consistently higher metrics obtained. The difference is can be also visualized in the training graphs included in **Appendix A**: EfficientNetB2 + GNN Training Figures, in which validation accuracy for implant model achieves stable values in fewer epochs, indicating faster and more stable convergence in this task.

Regarding the diameter classification task, it must be acknowledged that is a clinically relevant task, specifically in contexts where prosthetic crown substitution is required. Knowing the implant diameter allows for the selection

of appropriate prosthetic components and avoid compatibility errors. Despite its relevance, to the best of our knowledge this issue has been scarcely addressed in literature compared to classification by brand. In fact, our review identified only one study that addressed this task. When examining the results in terms of dataset distribution, it is observed that majoritarian diameter classes reach better performance metrics in comparison to minoritarian classes. This behaviour can be explained because of imbalance in the distribution of samples, which tends to favour those classes with greater representation.

A possible source of bias has been identified from the correlation between certain diameter classes and implant models. In the dataset used in this study, most diameters are present exclusively in a single model, which can be seen in **Table 5**. This could facilitate classification by providing the model with indirect information about the diameter based on morphological features of the brand. Although the classification heads for brand and diameter are independent, both tasks share the same convolutional backbone, so it can be possible that the learned visual representations are influenced by this correlation. However, this situation did not suppose a deterioration in performance for the only class shared between two models: the "4.0" mm diameter, present in "MKIII" and "MKIV". Despite its presence in two different models, the evaluated models obtained metrics superior to 90% for this class. Indeed, "4.0" class had the highest F1-score among all classes, as it was also the second most common class. This result highlights that, at least for the scope of this study, the model has been able to effectively discriminate diameter as an independent category, without relying exclusively on the model as a latent variable.

Even though so, a detailed analysis of the confusion matrices for diameter classification (**Figure 22b**) reveals a pattern suggesting influence of the implant model on diameter prediction. There are cases in which errors do not follow a metric logic (that is, they are not confused with diameters close in numerical value), but rather occur between diameters belonging to the same implant model. For example, instances with an actual diameter of 3.5 mm, which is exclusive in "ACTIVE" model, are most often misclassified as 3.0 mm, also found only in "ACTIVE", rather than as 3.3 mm, which is the closest in size but appears only in the "MKIII" model. Similarly, implants of 3.3 mm, exclusive to "MKIII", are primarily confused with 3.75 mm, another "MKIII" diameter, rather than 3.5 mm, despite the latter being numerically closer but belonging to "ACTIVE". This phenomenon suggests that the architecture, by sharing a common backbone for both tasks, could be implicitly using morphological characteristics associated with the implant model as support to predict the diameter.

Although this bias could limit the ability to generalize to new brand-diameter combinations, it is important to note implant manufacturers do not offer all diameters in all their models. On the contrary, each implant system is available in a limited set of diameters. Therefore, the association between brand and diameter should not necessarily be understood as something undesirable, but instead as a faithful representation of the real clinical context, as there is always going to exist some correlation between model and diameter. However, to ensure robust performance it would be advisable to train and validate the model on less correlated datasets.

A relevant question that arises when analysing the results obtained is why EfficientNetB2, a model that only receives masked images as input, outperformed some multimodal configurations that integrate additional input. To understand this behaviour, it is necessary to consider the image acquisition context. All the radiographs used in this study came from the same dental clinic,

where the imaging equipment, plate size, and acquisition techniques were kept consistent. This technical homogeneity considerably reduces variability in the visual representation of implants, which favours the vision-based model. In this controlled environment, EfficientNetB2 can extract very accurate and consistent visual representations, without the need for complementary data to compensate for variations. However, this same consistency poses a limitation in terms of generalization. In real-world clinical practice, multiple sources of variability (like plate size and the distance between the implant and the sensor) can significantly affect the apparent scale of the image. Of course, these sources of variability are present between different dental clinics. Changes in these parameters can induce magnification or alter the pixel-to-millimeter ratio, which directly impacts diameter prediction. Since EfficientNet lacks explicit metric references, its performance is likely to be compromised in more heterogeneous settings. Therefore, while current results are excellent, it will be essential to evaluate the model on multicenter datasets to determine its true robustness outside of a controlled setting.

In contrast, models that integrate structured tabular information (with features like patient-sensor distance and plate size) might be better equipped to handle this type of variability, as they have complementary data that helps them correctly interpret the physical scale of the implant under different conditions. Therefore, although EfficientNetB2 showed good performance in this homogeneous environment, it is reasonable to assume that multimodal models could have greater robustness than image-only ones with multicenter datasets.

Notwithstanding the high performance metrics obtained, the most relevant limitation of this study is the small size of the dataset. It must be kept in mind that the model was trained on 914 radiographs with 1426 implants coming from the same clinic, which only represent three commercial models. Although this small dataset enabled experimental validation of this approach's feasibility, the system's robustness and generalization capacity might be affected by data scarcity. Furthermore, only three implant models were included in the dataset, when in real clinical practice coexist hundreds of brands, designs, and sizes. This is a barrier to the model's implementation in clinical practice.

6 Conclusion and Future Directions

Although the limited data available, the results obtained in this master thesis show that automatic classification of DISs model and diameter can be successfully achieved through MMDL. The developed models in this work have achieved high performance metrics, validating the technical feasibility of the approach.

Regarding the objectives set out at the beginning of this work, it can be concluded that these have been satisfactorily achieved. Different deep learning architectures have been designed, trained, and compared, including models that integrate multimodal information such as images, tabular features, and anatomical keypoints in the form of graphs. Furthermore, the impact of each modality on system performance has been evaluated, and the models' ability to differentiate morphologically similar implants has been analysed, both in terms of brand and diameter. Among the different configurations evaluated, the multimodal model based on EfficientNetB2 and a graph neural network (GNN) had the best overall performance. In brand classification, it achieved precision, recall, and F1-score values of 98%, with an overall accuracy of 99%. For diameter classification, a more complex task less addressed in the literature, the model achieved a precision of 92%, a recall of 92%, and an F1-score of 92%, with a final accuracy of 93%. These results confirm the usefulness of integrating structured information with visual representation to improve system performance in both tasks.

However, future studies are necessary to determine the feasibility in real-world clinical practice. First, it will be essential to expand the dataset, both in volume and diversity, incorporating a greater variety of implant models and radiographs from multiple centres. A multicentre dataset, collected from multiple centres or institutions, is needed to improve data diversity, better represent the variety of implant systems and confirm the robustness and generalizability of the architecture.

Another key aspect will be the clinical validation of the system, ideally in collaboration with dental practitioners who can compare the model's results with their own diagnostic experience. For the model to be truly useful in clinical practice, it must be integrated into intuitive software tools or interfaces designed for medical professionals. This would enable efficient AI-assisted diagnosis, offering dental professionals support in their decision-making processes.

Furthermore, it would be also interesting to implement explainable artificial intelligence (XAI) techniques into models like the ones developed in this study. This would allow dental professionals to understand the reasonable behind the model’s decisions, while at the same time, would increase trust and acceptance by clinical staff. Consequently, the integration of these types of tools into clinical practice will be facilitated

Finally, this master thesis contributes to the growing development of deep learning-based solutions for dental implant classification. It is hoped that this study will serve as a foundation for future advances in the field, with the aim of supporting dental professionals and, ultimately, improving patient care in the field of implantology.

7 Innovation and Entrepeneurship

7.1 Clinical Need

Dental practitioners frequently encounter patients with existing implants placed elsewhere, lacking documentation of the implant model or size. Given hundreds of implant brands and thousands of models on the market, identifying an unknown implant from an X-ray is a considerable challenge for both experienced and unexperienced dental professionals. Recent advances in artificial intelligence (AI) promise a solution: studies have shown that AI can outperform dental experts in implant identification accuracy and do so much faster – up to 19 times faster than manual identification [8]. By dramatically reducing the time spent researching implant details, an AI-driven system can streamline clinical workflows and allow dentists to focus on patient care rather than on tedious identification tasks. This time-saving potential directly addresses a pain point in implant dentistry and can improve efficiency in dental radiology and implant treatment planning.

7.2 Solution Overview and Competitive Advantages

Building on our research, we propose a **deep learning software** for automatic dental implant identification. The system analyses periapical radiographs in combination with anatomical landmarks (e.g. jaw location, bone levels) and other data to determine the implant's brand and dimensions. This cloud-based AI inference can be integrated into the dental clinic's existing imaging workflow, providing results in near real-time during patient visits.

Competitive Advantages: Our platform offers several key advantages over traditional methods and existing tools:

- **Real-Time Inference:** The AI model provides instant identification (on the order of seconds), enabling point-of-care use. This real-time response is a major improvement over manual workflows that might require consultations or lengthy database searches. Rapid, automated identification accelerates diagnosis and treatment decisions, addressing the clinical need for speed.

- **Validated Performance:** The system will be built upon a research-tested model trained on a large dataset of implant radiographs. Prior studies have demonstrated that AI can reliably classify implant systems even in low-quality or distorted radiographs with much higher accuracy than clinicians. This evidence-based foundation will be a competitive advantage, lending credibility and reducing adoption barriers in an evidence-driven industry.

- **Workflow Integration:** The solution is designed to integrate seamlessly into dental radiography software and workflows (see Integration section below). Competing services exist that attempt implant identification through image databases, but they often function as separate applications or require expert validation of results. In contrast, our AI runs fully autonomously and can be embedded in the imaging software interface, providing a *one-click* identification without disrupting the clinician's routine. This ease of use and integration capability positions our product favourably against any stand-alone tools.

In summary, the combination of real-time AI inference, a multimodal deep learning approach, and proven accuracy gives our system a strong competitive edge. These features directly address the current limitations in implant identification, offering a solution that is faster and more reliable than both manual methods and existing partial solutions.

7.3 Market Trends and Opportunities

The market environment is highly favourable for an AI-driven dental imaging solution. Dentistry is undergoing a digital transformation, with growing adoption of imaging technologies and AI for diagnostics. Market data indicate growth in AI applied to dentistry: for example, the global *AI in dental imaging* market was valued around $1.2 billion in 2023 and is projected to reach $2.4 billion by 2030 (approximately 18–24% annual CAGR) [120]. This growth is driven by the need for enhanced diagnostic efficiency and the increasing volume of dental imaging procedures. The broader *AI in dentistry* market (covering diagnostics, planning, etc.) is expected to expand from about $421 million in 2024 to $3.1 billion by 2034 (22% CAGR), underscoring the strong upward trend in AI adoption across dental practices. [120]

These trends reflect a convergence of factors: the dental profession's push for improved accuracy and time savings, advancements in machine learning algorithms, and a generational shift toward digital solutions in oral health care. Early clinical adoption of AI in dentistry is still in its infancy, but confidence is building as regulatory approvals and success stories accumulate. Major dental organizations and investors have started supporting AI dental startups, validating the market potential. In this context, a solution that addresses a clear clinical pain point – implant identification – is well positioned to capture interest. Time-saving and accuracy-enhancing tools are in high demand, and our product directly contributes to these values. With few direct competitors in the specific niche of implant model identification, we have an opportunity to emerge as a leader in a growing segment. The expanding market and favourable technology trends provide a supportive runway for commercialization, suggesting that a sustainable customer base can be established through early adopter dental centres and scaled as AI in dentistry becomes mainstream.

7.4 Business Model Strategy

Our business model will be centered on a **Software-as-a-Service (SaaS)** subscription platform, with flexibility to offer enterprise licensing for large integrations. The SaaS approach allows clinics and institutions to access the AI software via a secure cloud-based interface or API without significant upfront investment in hardware. Subscribers (e.g. dental practices or imaging centers) would pay a monthly or annual fee (tiered by usage or number of scans analyzed), ensuring recurring revenue. This model also enables continuous updates and improvements to the AI model on the server side, so all users benefit from the latest performance enhancements without complex reinstallations – a critical factor given the fast-evolving nature of AI algorithms.

For distribution and sales, we will target key early adopters (see next section) through direct outreach and partnerships. A free trial period may be offered to encourage adoption and demonstrate value (e.g. allowing a certain number of identifications per month at no cost). Once users integrate the tool into their workflow and see time savings, conversion to paid plans is anticipated. Revenue can be further scaled by offering premium enterprise plans that include on-premise deployment or advanced integrations; for example, a dental hospital

network might prefer a licensing arrangement to host the model internally for data privacy reasons, which we would accommodate at a higher price point. We will also explore OEM licensing opportunities with dental imaging software vendors – embedding our AI engine into popular dental radiography/PACS systems under a licensing fee agreement. This dual model (SaaS + licensing) maximizes market penetration: smaller clinics can quickly adopt via SaaS, while larger entities and software companies can integrate the technology deeply into their products or infrastructure.

7.5 Integration with Dental Software Ecosystems

Seamless integration with existing dental software and workflows is a core part of our strategy. We recognize that for busy dental professionals, an AI tool must fit into their current practice management and imaging systems with minimal disruption. Our system will therefore support integration with common dental imaging formats and platforms, such as DICOM-based PACS (Picture Archiving and Communication Systems) used in radiology, as well as electronic health record (EHR) or dental practice management software. Technically, we will provide open APIs and standards-based interfaces so that the AI can fetch radiographs from a clinic's imaging database and return identification results directly into the software that the dentist already uses to view X-rays. This could take the form of a plugin or add-on module for popular dental radiography software, or a web portal that is launched from within the PACS/EHR environment.

Industry experience has shown that AI imaging software *"must integrate with existing systems like EMRs and PACS"* to be truly effective in real clinical workflows. By adhering to this principle, our tool will allow clinicians to, for example, open a patient's radiograph in their usual viewer and simply click an "Identify Implant" button – the result (implant model and specs) would be displayed on-screen alongside the image and saved into the patient's record. We will ensure compatibility with major digital radiography formats and work closely with early users to refine the integration process.

Data security and compliance are also crucial: our integration will be HIPAA-compliant and follow dental data protection standards. Images can be anonymized and securely transmitted to our cloud for analysis, or where required, the analysis can be deployed on-premises for full control. The integration with EHRs means the identified implant information can be automatically documented in the patient's chart, saving further administrative time. Additionally, integrating with inventory or ordering systems could allow the software to suggest compatible prosthetic components or where to order them, once the implant is identified – adding value to the clinical workflow. This level of integration into the dental software ecosystem not only solidifies our tool's utility but also creates a high switching cost (users will find it seamlessly woven into their practice, making them less likely to abandon it).

7.6 Target Market and Initial Users

Our initial go-to-market focus will be on specialized dental service providers and institutions that have an acute need for implant identification efficiency. These early adopters are most likely to recognize the value and to influence broader adoption. Key target user segments include:

- **Dental Radiology Centres:** Imaging centres and oral radiology practices that handle large volumes of dental X-rays for referrals. They often

receive cases specifically to identify unknown implants or to diagnose implant complications. Providing them with an AI tool will greatly speed up their consultative reports. Radiology specialists can use the software to enhance their analysis, increasing throughput and accuracy in their diagnostic services. Successful use in radiology centres will also generate case studies and endorsements for our product.

- **Oral and Maxillofacial Surgeons:** These surgeons frequently deal with implant removals, revisions, or placements adjacent to existing implants. When planning a surgery, knowing the exact implant system in the patient's jaw is crucial (for example, to obtain the correct removal tool or a matching implant for a contralateral site). An AI that instantly identifies the implant model from a pre-op radiograph can save the surgeon from exploratory surgery or guesswork. We will target oral surgery clinics and hospital departments, highlighting how the system improves surgical planning efficiency and reduces intraoperative surprises.

- **Prosthodontists and Implant Restorative Dentists:** Specialists who restore implants (place crowns, abutments, or overdentures on implants) often see patients with implants placed elsewhere. They need to determine the implant's brand and size to select compatible prosthetic components. This process can otherwise involve sending radiographs to manufacturers or searching online databases. Our solution offers an on-the-spot answer, which is invaluable for prosthodontic practices. By initially marketing to leading prosthodontists (especially those in implant referral networks), we can penetrate the restorative segment of the market.

- **Academic Dental Institutions:** Dental schools and university hospitals will be another strategic early adopter. Academic institutions are often early adopters of innovative technology for both clinical use and research. They have high volumes of diverse cases (useful for further training our AI) and can act as opinion leaders endorsing the technology. We will collaborate with university dental clinics to pilot the system, as their feedback will help refine the product, and their publications or conference presentations can serve as credible, peer-reviewed validation. Integrating the tool into dental practitioners training will familiarize new dentists with the technology, contributing to long-term acceptance.

The go-to-market strategy will involve attending specialty dental conferences (implantology, prosthodontics, radiology), publishing in dental journals (to provide scientific backing), and partnering with implant manufacturers or dental supply companies for referrals. These channels will help reach our target users efficiently and lend authority to our solution by aligning with established dental industry players.

8 Bibliography

[1] B. Vandenberghe, R. Jacobs, and H. Bosmans, “Modern dental imaging: a review of the current technology and clinical applications in dental practice,” *Eur Radiol*, vol. 20, no. 11, pp. 2637–2655, Nov. 2010, doi: 10.1007/s00330-010-1836-1.

[2] N. Shah, “Recent advances in imaging technologies in dentistry,” *World J Radiol*, vol. 6, no. 10, p. 794, 2014, doi: 10.4329/wjr.v6.i10.794.

[3] E. Corbet, D. Ho, and S. Lai, “Radiographs in periodontal disease diagnosis and management,” *Aust Dent J*, vol. 54, no. s1, Sep. 2009, doi: 10.1111/j.1834-7819.2009.01141.x.

[4] M. Schimmel *et al.*, “Group 4 <scp>ITI</scp> Consensus Report: Patient benefits following implant treatment in partially and fully edentulous patients,” *Clin Oral Implants Res*, vol. 34, no. S26, pp. 257–265, Sep. 2023, doi: 10.1111/clr.14145.

[5] D. G. K. Hong and J. Oh, “Recent advances in dental implants,” *Maxillofac Plast Reconstr Surg*, vol. 39, no. 1, p. 33, Dec. 2017, doi: 10.1186/s40902-017-0132-2.

[6] J.-H. Park, H. S. Moon, H.-I. Jung, J. Hwang, Y.-H. Choi, and J.-E. Kim, “Deep learning and clustering approaches for dental implant size classification based on periapical radiographs,” *Sci Rep*, vol. 13, no. 1, p. 16856, Oct. 2023, doi: 10.1038/s41598-023-42385-7.

[7] J.-H. Lee and S.-N. Jeong, “Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs,” *Medicine*, vol. 99, no. 26, p. e20787, Jun. 2020, doi: 10.1097/MD.0000000000020787.

[8] J.-H. Lee, Y.-T. Kim, and J.-B. Lee, “Identification of dental implant systems from low-quality and distorted dental radiographs using AI trained on a large multi-center dataset,” *Sci Rep*, vol. 14, no. 1, p. 12606, Jun. 2024, doi: 10.1038/s41598-024-63422-z.

[9] T. Takahashi, K. Nozaki, T. Gonda, T. Mameno, M. Wada, and K. Ikebe, “Identification of dental implants using deep learning—pilot study,” *Int J Implant Dent*, vol. 6, no. 1, p. 53, Dec. 2020, doi: 10.1186/s40729-020-00250-6.

[10] J. Iannucci and L. J. Howerton, *Dental radiography-E-book: principles and techniques.* Elsevier Health Sciences, 2016.

[11] E. Whaites and N. Drage, *Essentials of Dental Radiography and Radiology E-Book*. Elsevier Health Sciences, 2013.

[12] F. Hagi and F. Shamshoon Hagi, “The History of Dental Radiology: A Review,” vol. 141, 2023, [Online]. Available: https://www.researchgate.net/publication/369794222

[13] B. A. Niemiec, “Digital Dental Radiography,” *J Vet Dent*, vol. 24, no. 3, pp. 192–197, Sep. 2007, doi: 10.1177/089875640702400311.

[14] E. Sy, V. Samboju, and T. Mukhdomi, *X-ray Image Production Procedures*. 2025.

[15] V. Tsapaki, “Radiation protection in dental radiology – Recent advances and future directions,” *Physica Medica*, vol. 44, pp. 222–226, Dec. 2017, doi: 10.1016/j.ejmp.2017.07.018.

[16] M. Cash and B. Basrani, “Intraoral Radiographic Principles and Techniques,” in *Endodontic Radiology*, Wiley, 2012, pp. 18–38. doi: 10.1002/9781119421689.ch2.

[17] L. A. Foster Page *et al.*, “The effect of bitewing radiography on estimates of dental caries experience among children differs according to their disease experience,” *BMC Oral Health*, vol. 18, no. 1, p. 137, Dec. 2018, doi: 10.1186/s12903-018-0596-1.

[18] J. Aps, “Radiography in pediatric dental practice,” *Clinical Dentistry Reviewed*, vol. 4, no. 1, p. 5, Dec. 2020, doi: 10.1007/s41894-019-0067-3.

[19] “Extraoral Imaging Techniques,” in *Fundamentals of Oral and Maxillofacial Radiology*, Wiley, 2017, pp. 68–95. doi: 10.1002/9781119411871.ch13.

[20] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, “Deep Instance Segmentation of Teeth in Panoramic X-Ray Images,” in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, IEEE, Oct. 2018, pp. 400–407. doi: 10.1109/SIBGRAPI.2018.00058.

[21] P. Subramaniam, K. Kumar, T. Ramakrishna, and A. Bhadranna, “Bone regeneration with plasma-rich-protein following enucleation of traumatic bone cyst,” *Eur J Dent*, vol. 07, no. 03, pp. 377–381, Jul. 2013, doi: 10.4103/1305-7456.115427.

[22] S. Płotka, T. Włodarczy, and R. Szczerba, “Convolutional Neural Networks in Orthodontics: a review”.

[23] S. Gupta, N. Patil, J. Solanki, R. Singh, and S. Laller, “Oral Implant Imaging: A Review.,” *Malays J Med Sci*, vol. 22, no. 3, pp. 7–17, 2015.

[24] C.-W. Liao, K.-J. Huang, J.-C. Chen, C.-W. Kuo, Y.-Y. Wu, and J.-T. Hsu, “A Prototype Intraoral Periapical Sensor with High Frame Rates for a 2.5D Periapical Radiography System,” *Appl Bionics Biomech*, vol. 2019, pp. 1–9, Apr. 2019, doi: 10.1155/2019/7987496.

[25] P. Sahrmann, S. Kühl, D. Dagassan-Berndt, M. M. Bornstein, and N. U. Zitzmann, “Radiographic assessment of the peri-implant site,” *Periodontol 2000*, vol. 95, no. 1, pp. 70–86, Jun. 2024, doi: 10.1111/prd.12577.

[26] A. Gupta, R. Srivastava, and B. Jyoti, “Intra oral periapical radiography-basics yet intrigue: A review.”

[27] M. Yen and A. W. K. Yeung, “The Performance of Paralleling Technique and Bisecting Angle Technique for Taking Periapical Radiographs: A Systematic Review.,” *Dent J (Basel)*, vol. 11, no. 7, Jun. 2023, doi: 10.3390/dj11070155.

[28] A. Hudecki, G. Kiryczyński, and M. J. Łos, “Biomaterials, Definition, Overview,” in *Stem Cells and Biomaterials for Regenerative Medicine*, Elsevier, 2019, pp. 85–98. doi: 10.1016/B978-0-12-812258-7.00007-1.

[29] R. Gupta, N. Gupta, and D. K. K. Weber, *Dental Implants*. 2025.

[30] Y. Wang, D. Bäumer, A.-K. Ozga, G. Körner, and A. Bäumer, “Patient satisfaction and oral health-related quality of life 10 years after implant placement,” *BMC Oral Health*, vol. 21, no. 1, p. 30, Dec. 2021, doi: 10.1186/s12903-020-01381-3.

[31] H. Duong, A. Roccuzzo, A. Stähli, G. E. Salvi, N. P. Lang, and A. Sculean, “Oral health-related quality of life of patients rehabilitated with fixed and removable implant-supported dental prostheses,” *Periodontol 2000*, vol. 88, no. 1, pp. 201–237, Feb. 2022, doi: 10.1111/prd.12419.

[32] A. D. Pye, D. E. A. Lockhart, M. P. Dawson, C. A. Murray, and A. J. Smith, “A review of dental implants and infection,” *Journal of Hospital Infection*, vol. 72, no. 2, pp. 104–110, Jun. 2009, doi: 10.1016/j.jhin.2009.02.010.

[33] D. W. Fain, *Introduction to implant dentistry: a student guide*, 2nd ed., vol. 75. WB SAUNDERS CO-ELSEVIER INC 1600 JOHN F KENNEDY BOULEVARD, STE 1800, 2017.

[34] M. Esposito, Y. Ardebili, and H. V Worthington, “Interventions for replacing missing teeth: different types of dental implants,” *Cochrane Database of Systematic Reviews*, Jul. 2014, doi: 10.1002/14651858.CD003815.pub4.

[35] J. Łoginoff, A. Majos, and M. Elgalal, “The Evolution of Custom Subperiosteal Implants for Treatment of Partial or Complete Edentulism in Patients with Severe Alveolar Ridge Atrophy,” *J Clin Med*, vol. 13, no. 12, p. 3582, Jun. 2024, doi: 10.3390/jcm13123582.

[36] G. Timarcioglu, “The Role of Custom Subperiosteal Implants in Edentulous Patients,” *Int Dent J*, vol. 74, pp. S248–S249, Oct. 2024, doi: 10.1016/j.identj.2024.07.137.

[37] T. Watanabe *et al.*, “Squamous cell carcinoma around a subperiosteal implant in the maxilla and the association of chronic mechanical irritation and peri-implantitis: a case report,” *Int J Implant Dent*, vol. 8, no. 1, p. 10, Dec. 2022, doi: 10.1186/s40729-022-00409-3.

[38] F. H. Abdel-Rahman, S. A. Hegazy, M. A. Nagib, A. M. Ibrahim, and A. Habib, “Evaluation of Quality of Life and Satisfaction with Fixed Prostheses on Zygomatic Implants vs All-on-Four Concept: A Randomized Clinical Study,” *J Contemp Dent Pract*, vol. 25, no. 2, pp. 141–147, Mar. 2024, doi: 10.5005/jp-journals-10024-3632.

[39] W. D. Polido, A. Machado-Fernandez, W.-S. Lin, and T. Aghaloo, “Indications for zygomatic implants: a systematic review,” *Int J Implant Dent*, vol. 9, no. 1, p. 17, Jul. 2023, doi: 10.1186/s40729-023-00480-4.

[40] M. Penarrocha-Diago, S. Galan-Gil, C. Carrillo-Garcia, D. Penarrocha-Diago, and M. Penarrocha-Diago, “Transcrestal sinus lift and implant placement using the sinus balloon technique,” *Med Oral Patol Oral Cir Bucal*, pp. e122–e128, 2012, doi: 10.4317/medoral.17268.

[41] S. S. Al-Johany, M. D. Al Amri, S. Alsaeed, and B. Alalola, “Dental Implant Length and Diameter: A Proposed Classification Scheme.,” *J Prosthodont*, vol. 26, no. 3, pp. 252–260, Apr. 2017, doi: 10.1111/jopr.12517.

[42] P. Rahmanivahid and M. Heidari, “Design parameters of dental implants: A review,” *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, vol. 38, no. 1, 2022, doi: 10.23967/j.rimni.2022.03.002.

[43] H. Lung, J.-T. Hsu, A. Y.-J. Wu, and H.-L. Huang, “Biomechanical Effects of Diameters of Implant Body and Implant Platform in Bone Strain around an Immediately Loaded Dental Implant with Platform Switching Concept,” *Applied Sciences*, vol. 9, no. 10, p. 1998, May 2019, doi: 10.3390/app9101998.

[44] S. Ahn, J. Kim, S. C. Jeong, M. Kim, C. Kim, and D. Park, “Stress Distribution Analysis of Threaded Implants for Digital Dentistry,” *Int J Environ Res Public Health*, vol. 19, no. 19, p. 12674, Oct. 2022, doi: 10.3390/ijerph191912674.

[45] N. Quispe-López, S. Martín-Martín, C. Gómez-Polo, O. Figueras-Alvarez, M. I. Sánchez-Jorge, and J. Montero, “Primary and Secondary Stability Assessments of Dental Implants According to Their Macro-Design, Length, Width, Location, and Bone Quality,” *Applied Sciences*, vol. 14, no. 11, p. 4841, Jun. 2024, doi: 10.3390/app14114841.

[46] M. SULEIMAN, “Testing the performance of dental implants,” in *Dental Biomaterials*, Elsevier, 2008, pp. 360–427. doi: 10.1533/9781845694241.360.

[47] H. Sheikh, C. Prins, and E. Schrijvers, “Artificial Intelligence: Definition and Background,” 2023, pp. 15–41. doi: 10.1007/978-3-031-21448-6_2.

[48] European Comission and Joint Research Center, *AI watch, defining artificial intelligence 2.0 – Towards an operational definition and taxonomy for the AI landscape*. Publications Office of the European Union, 2021.

[49] Y. Jiang, X. Li, H. Luo, S. Yin, and O. Kaynak, “Quo vadis artificial intelligence?,” *Discover Artificial Intelligence*, vol. 2, no. 1, p. 4, Mar. 2022, doi: 10.1007/s44163-022-00022-8.

[50] H. Ding, J. Wu, W. Zhao, J. P. Matinlinna, M. F. Burrow, and J. K. H. Tsoi, “Artificial intelligence in dentistry—A review,” *Frontiers in Dental Medicine*, vol. 4, Feb. 2023, doi: 10.3389/fdmed.2023.1085251.

[51] J.-H. Lee, Y.-T. Kim, J.-B. Lee, and S.-N. Jeong, “A Performance Comparison between Automated Deep Learning and Dental Professionals in Classification of Dental Implant Systems from Dental Imaging: A Multi-Center Study,” *Diagnostics*, vol. 10, no. 11, p. 910, Nov. 2020, doi: 10.3390/diagnostics10110910.

[52] J.-H. Lee and S.-N. Jeong, “Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs,” *Medicine*, vol. 99, no. 26, p. e20787, Jun. 2020, doi: 10.1097/MD.0000000000020787.

[53] H. Ding *et al.*, “Morphology and mechanical performance of dental crown designed by 3D-DCGAN,” *Dental Materials*, vol. 39, no. 3, pp. 320–332, Mar. 2023, doi: 10.1016/j.dental.2023.02.001.

[54] S. Tian *et al.*, “DCPR-GAN: Dental Crown Prosthesis Restoration Using Two-Stage Generative Adversarial Networks,” *IEEE J Biomed Health Inform*, vol. 26, no. 1, pp. 151–160, Jan. 2022, doi: 10.1109/JBHI.2021.3119394.

[55] C. Li *et al.*, “Autonomous robotic surgery for zygomatic implant placement and immediately loaded implant-supported full-arch prosthesis: a preliminary research,” *Int J Implant Dent*, vol. 9, no. 1, p. 12, May 2023, doi: 10.1186/s40729-023-00474-2.

[56] M. Sun, C. Yue, S. Stuhr, X. Fu, and H.-L. Wang, “Autonomous Dental Implant Robotic System Utilization for Implant Placement and Transcrestal Sinus Elevation Using Osseodensification: A Case Report,” *Int J Periodontics Restorative Dent*, Oct. 2023, doi: 10.11607/prd.6832.

[57] Z. Li, R. Xie, S. Bai, and Y. Zhao, “Implant placement with an autonomous dental implant robot: A clinical report,” *J Prosthet Dent*, vol. 133, no. 2, pp. 340–345, Feb. 2025, doi: 10.1016/j.prosdent.2023.02.014.

[58] E. Ferrara, B. Rapone, and A. D’Albenzio, “Applications of deep learning in periodontal disease diagnosis and management: a systematic review and critical appraisal,” *J Med Artif Intell*, vol. 8, pp. 23–23, Sep. 2025, doi: 10.21037/jmai-24-241.

[59] F. Dai *et al.*, “Convolutional neural networks combined with classification algorithms for the diagnosis of periodontitis,” *Oral Radiol*, vol. 40, no. 3, pp. 357–366, Jul. 2024, doi: 10.1007/s11282-024-00739-5.

[60] R. Widyaningrum, I. Candradewi, N. R. A. S. Aji, and R. Aulianisa, “Comparison of Multi-Label U-Net and Mask R-CNN for panoramic radiograph segmentation to detect periodontitis,” *Imaging Sci Dent*, vol. 52, no. 4, p. 383, 2022, doi: 10.5624/isd.20220105.

[61] V. P. Kearney *et al.*, “A generative adversarial inpainting network to enhance prediction of periodontal clinical attachment level,” *J Dent*, vol. 123, p. 104211, Aug. 2022, doi: 10.1016/j.jdent.2022.104211.

[62] J. Chang *et al.*, “Application of deep machine learning for the radiographic diagnosis of periodontitis,” *Clin Oral Investig*, vol. 26, no. 11, pp. 6629–6637, Jul. 2022, doi: 10.1007/s00784-022-04617-4.

[63] J.-H. Lee, D. Kim, S.-N. Jeong, and S.-H. Choi, “Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm,” *J Periodontal Implant Sci*, vol. 48, no. 2, p. 114, 2018, doi: 10.5051/jpis.2018.48.2.114.

[64] L. Pang, K. Wang, Y. Tao, Q. Zhi, J. Zhang, and H. Lin, “A New Model for Caries Risk Prediction in Teenagers Using a Machine Learning Algorithm Based on Environmental and Genetic Factors,” *Front Genet*, vol. 12, Mar. 2021, doi: 10.3389/fgene.2021.636867.

[65] L. A. Zanella-Calzada *et al.*, “Deep Artificial Neural Networks for the Diagnostic of Caries Using Socioeconomic and Nutritional Features as Determinants: Data from NHANES 2013–2014,” *Bioengineering*, vol. 5, no. 2, p. 47, Jun. 2018, doi: 10.3390/bioengineering5020047.

[66] H. Chen, H. Li, Y. Zhao, J. Zhao, and Y. Wang, “Dental disease detection on periapical radiographs based on deep convolutional neural networks,” *Int J Comput Assist Radiol Surg*, vol. 16, no. 4, pp. 649–661, Apr. 2021, doi: 10.1007/s11548-021-02319-y.

[67] R. Esmaeilyfard, H. Bonyadifard, and M. Paknahad, “Dental Caries Detection and Classification in CBCT Images Using Deep Learning,” *Int Dent J*, vol. 74, no. 2, pp. 328–334, Apr. 2024, doi: 10.1016/j.identj.2023.10.003.

[68] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, “Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm,” *J Dent*, vol. 77, pp. 106–111, Oct. 2018, doi: 10.1016/j.jdent.2018.07.015.

[69] M. Fukuda *et al.*, “Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography,” *Oral Radiol*, vol. 36, no. 4, pp. 337–343, Oct. 2020, doi: 10.1007/s11282-019-00409-x.

[70] K. Orhan, I. S. Bayrakdar, M. Ezhov, A. Kravtsov, and T. Özyürek, “Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans,” *Int Endod J*, vol. 53, no. 5, pp. 680–689, May 2020, doi: 10.1111/iej.13265.

[71] M. G. Endres *et al.*, “Development of a Deep Learning Algorithm for Periapical Disease Detection in Dental Radiographs,” *Diagnostics*, vol. 10, no. 6, p. 430, Jun. 2020, doi: 10.3390/diagnostics10060430.

[72] X. Gao, X. Xin, Z. Li, and W. Zhang, “Predicting postoperative pain following root canal treatment by using artificial neural network evaluation,” *Sci Rep*, vol. 11, no. 1, p. 17243, Aug. 2021, doi: 10.1038/s41598-021-96777-8.

[73] L. Campo *et al.*, “Retreatment Predictions in Odontology by means of CBR Systems,” *Comput Intell Neurosci*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/7485250.

[74] H. Yang *et al.*, “Deep Learning for Automated Detection of Cyst and Tumors of the Jaw in Panoramic Radiographs,” *J Clin Med*, vol. 9, no. 6, p. 1839, Jun. 2020, doi: 10.3390/jcm9061839.

[75] W. Poedjiastoeti and S. Suebnukarn, “Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors,” *Healthc Inform Res*, vol. 24, no. 3, p. 236, 2018, doi: 10.4258/hir.2018.24.3.236.

[76] K. Orhan, I. S. Bayrakdar, M. Ezhov, A. Kravtsov, and T. Özyürek, “Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans,” *Int Endod J*, vol. 53, no. 5, pp. 680–689, May 2020, doi: 10.1111/iej.13265.

[77] W. Shin *et al.*, “Deep learning based prediction of necessity for orthognathic surgery of skeletal malocclusion using cephalogram in Korean individuals,” *BMC Oral Health*, vol. 21, no. 1, p. 130, Dec. 2021, doi: 10.1186/s12903-021-01513-3.

[78] P. C. Maganur *et al.*, “Development of Artificial Intelligence Models for Tooth Numbering and Detection: A Systematic Review,” *Int Dent J*, vol. 74, no. 5, pp. 917–929, Oct. 2024, doi: 10.1016/j.identj.2024.04.021.

[79] D. V. Tuzoff *et al.*, “Tooth detection and numbering in panoramic radiographs using convolutional neural networks,” *Dentomaxillofacial Radiology*, vol. 48, no. 4, p. 20180051, May 2019, doi: 10.1259/dmfr.20180051.

[80] G. Dipalma *et al.*, “Artificial Intelligence and Its Clinical Applications in Orthodontics: A Systematic Review,” *Diagnostics*, vol. 13, no. 24, p. 3677, Dec. 2023, doi: 10.3390/diagnostics13243677.

[81] H.-J. Kim, K. D. Kim, and D.-H. Kim, “Deep convolutional neural network-based skeletal classification of cephalometric image compared with automated-tracing software,” *Sci Rep*, vol. 12, no. 1, p. 11659, Jul. 2022, doi: 10.1038/s41598-022-15856-6.

[82] T. Wood, J. O. Anigbo, G. Eckert, K. T. Stewart, M. M. Dundar, and H. Turkkahraman, “Prediction of the Post-Pubertal Mandibular Length and

Y Axis of Growth by Using Various Machine Learning Techniques: A Retrospective Longitudinal Study," *Diagnostics*, vol. 13, no. 9, p. 1553, Apr. 2023, doi: 10.3390/diagnostics13091553.

[83] T. Mason, K. M. Kelly, G. Eckert, J. A. Dean, M. M. Dundar, and H. Turkkahraman, "A machine learning model for orthodontic extraction/non-extraction decision in a racially and ethnically diverse patient population," *Int Orthod*, vol. 21, no. 3, p. 100759, Sep. 2023, doi: 10.1016/j.ortho.2023.100759.

[84] N. Chaiprasittikul, B. Thanathornwong, S. Pornprasertsuk-Damrongsri, S. Raocharernporn, S. Maponthong, and S. Manopatanakul, "Application of a Multi-Layer Perceptron in Preoperative Screening for Orthognathic Surgery," *Healthc Inform Res*, vol. 29, no. 1, pp. 16–22, Jan. 2023, doi: 10.4258/hir.2023.29.1.16.

[85] S. Caruso, S. Caruso, M. Pellegrino, R. Skafi, A. Nota, and S. Tecco, "A Knowledge-Based Algorithm for Automatic Monitoring of Orthodontic Treatment: The Dental Monitoring System. Two Cases," *Sensors*, vol. 21, no. 5, p. 1856, Mar. 2021, doi: 10.3390/s21051856.

[86] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif Intell Rev*, vol. 57, no. 4, p. 99, Mar. 2024, doi: 10.1007/s10462-024-10721-6.

[87] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Med Imaging*, vol. 22, no. 1, p. 69, Dec. 2022, doi: 10.1186/s12880-022-00793-7.

[88] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[89] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.

[90] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[91] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[92] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[93] A. Howard *et al.*, "Searching for MobileNetV3," May 2019.

[94] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019.

[95] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020.

[96] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," Jan. 2022.

[97] Z. Dai, H. Liu, Q. V. Le, and M. Tan, “CoAtNet: Marrying Convolution and Attention for All Data Sizes,” Jun. 2021.

[98] Y. Yue and Z. Li, “MedMamba: Vision Mamba for Medical Image Classification,” Mar. 2024.

[99] C. Ma, J. Donnelly, W. Liu, S. Vosoughi, C. Rudin, and C. Chen, “Interpretable Image Classification with Adaptive Prototype-based Vision Transformers,” Oct. 2024.

[100] D. Wang, Q. Meng, D. Chen, H. Zhang, and L. Xu, “Automatic Detection of Arrhythmia Based on Multi-Resolution Representation of ECG Signal,” *Sensors*, vol. 20, no. 6, p. 1579, Mar. 2020, doi: 10.3390/s20061579.

[101] F. Ramzan *et al.*, “A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer’s Disease Stages Using Resting-State fMRI and Residual Neural Networks,” *J Med Syst*, vol. 44, no. 2, p. 37, Feb. 2020, doi: 10.1007/s10916-019-1475-2.

[102] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” Sep. 2016.

[103] Thomas Kipf, “Graph Convolutional Networks,” https://tkipf.github.io/graph-convolutional-networks/.

[104] X. Li, Y. Jiang, Y. Liu, J. Zhang, S. Yin, and H. Luo, “RAGCN: Region Aggregation Graph Convolutional Network for Bone Age Assessment From X-Ray Images,” *IEEE Trans Instrum Meas*, vol. 71, pp. 1–12, 2022, doi: 10.1109/TIM.2022.3190025.

[105] C. Tang, C. Hu, J. Sun, S.-H. Wang, and Y.-D. Zhang, “NSCGCN: A novel deep GCN model to diagnosis COVID-19,” *Comput Biol Med*, vol. 150, p. 106151, Nov. 2022, doi: 10.1016/j.compbiomed.2022.106151.

[106] J. Chen *et al.*, “A deep learning-based multimodal medical imaging model for breast cancer screening,” *Sci Rep*, vol. 15, no. 1, p. 14696, Apr. 2025, doi: 10.1038/s41598-025-99535-2.

[107] F. Hinterwimmer *et al.*, “Impact of metadata in multimodal classification of bone tumours,” *BMC Musculoskelet Disord*, vol. 25, no. 1, p. 822, Oct. 2024, doi: 10.1186/s12891-024-07934-9.

[108] K. Kita *et al.*, “Bimodal artificial intelligence using TabNet for differentiating spinal cord tumors-Integration of patient background information and images.,” *iScience*, vol. 26, no. 10, p. 107900, Oct. 2023, doi: 10.1016/j.isci.2023.107900.

[109] R. Panneerselvam, S. Balasubramaniam, V. Sureshkumar, V. Ravi, and S. S. Maidin, “Multimodal Skin Cancer Prediction: Integrating Dermoscopic Images and Clinical Metadata with Transfer Learning,” *Open Bioinforma J*, vol. 18, no. 1, Jan. 2025, doi: 10.2174/0118750362358444250120070327.

[110] F. Khader *et al.*, “Multimodal Deep Learning for Integrating Chest Radiographs and Clinical Parameters: A Case for Transformers,” *Radiology*, vol. 309, no. 1, Oct. 2023, doi: 10.1148/radiol.230806.

[111] M. A. Saghiri, P. Freag, A. Fakhrzadeh, A. M. Saghiri, and J. Eid, “Current technology for identifying dental implants: a narrative review,” *Bull Natl Res Cent*, vol. 45, no. 1, p. 7, Dec. 2021, doi: 10.1186/s42269-020-00471-0.

[112] J.-H. Lee and S.-N. Jeong, “Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs,” *Medicine*, vol. 99, no. 26, p. e20787, Jun. 2020, doi: 10.1097/MD.0000000000020787.

[113] J.-E. Kim, N.-E. Nam, J.-S. Shim, Y.-H. Jung, B.-H. Cho, and J. J. Hwang, “Transfer Learning via Deep Neural Networks for Implant Fixture System Classification Using Periapical Radiographs,” *J Clin Med*, vol. 9, no. 4, p. 1117, Apr. 2020, doi: 10.3390/jcm9041117.

[114] R. da Mata Santos *et al.*, “Automated Identification of Dental Implants Using Artificial Intelligence,” *Int J Oral Maxillofac Implants*, vol. 36, no. 5, pp. 918–923, Sep. 2021, doi: 10.11607/jomi.8684.

[115] B. Tiryaki, A. Ozdogan, M. T. Guller, O. Miloglu, E. A. Oral, and I. Y. Ozbek, “Dental implant brand and angle identification using deep neural networks,” *J Prosthet Dent*, Sep. 2023, doi: 10.1016/j.prosdent.2023.07.022.

[116] W. Park, J.-K. Huh, and J.-H. Lee, “Automated deep learning for classification of dental implant radiographs using a large multi-center dataset,” *Sci Rep*, vol. 13, no. 1, p. 4862, Mar. 2023, doi: 10.1038/s41598-023-32118-1.

[117] J.-H. Lee, Y.-T. Kim, and J.-B. Lee, “Identification of dental implant systems from low-quality and distorted dental radiographs using AI trained on a large multi-center dataset,” *Sci Rep*, vol. 14, no. 1, p. 12606, Jun. 2024, doi: 10.1038/s41598-024-63422-z.

[118] V. Benakatti, R. P. Nayakar, M. Anandhalli, and R. sukhasare, “Advanced deep learning techniques for recognition of dental implants,” *J Oral Biol Craniofac Res*, vol. 15, no. 2, pp. 215–220, Mar. 2025, doi: 10.1016/j.jobcr.2025.01.016.

[119] J.-H. Park, H. S. Moon, H.-I. Jung, J. Hwang, Y.-H. Choi, and J.-E. Kim, “Deep learning and clustering approaches for dental implant size classification based on periapical radiographs,” *Sci Rep*, vol. 13, no. 1, p. 16856, Oct. 2023, doi: 10.1038/s41598-023-42385-7.

[120] InsightAce Analytics, “AI in Dentistry Market – Share & Trends Analysis Report 2025-2034 (Report ID: 3004) – Report Summary [Press release],” 2025.

9 Appendix

9.1 Appendix A: EfficientNetB2 + GNN Training Figures

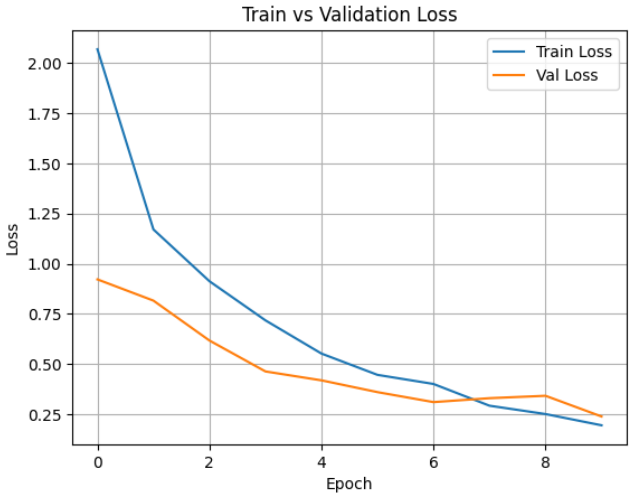


Figure A1. *Training vs Validation Loss*

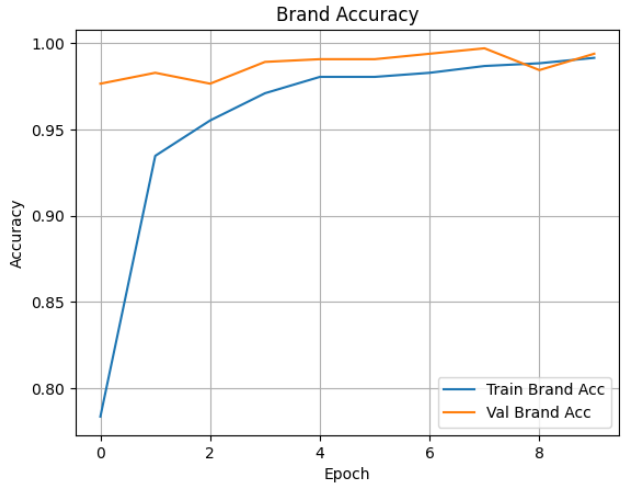


Figure A2. *Train vs Validation accuracy for model classification*

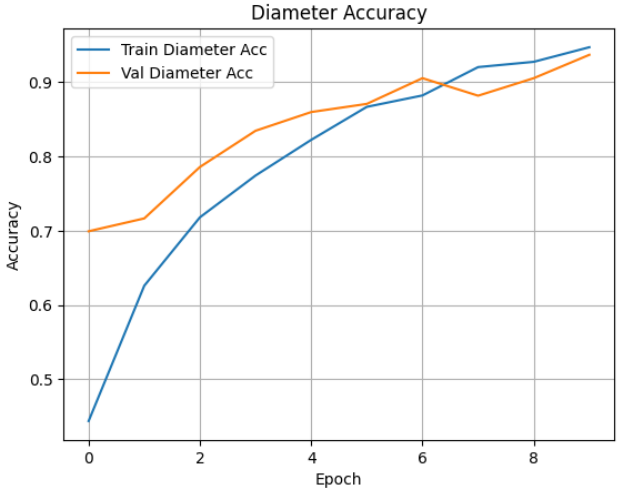
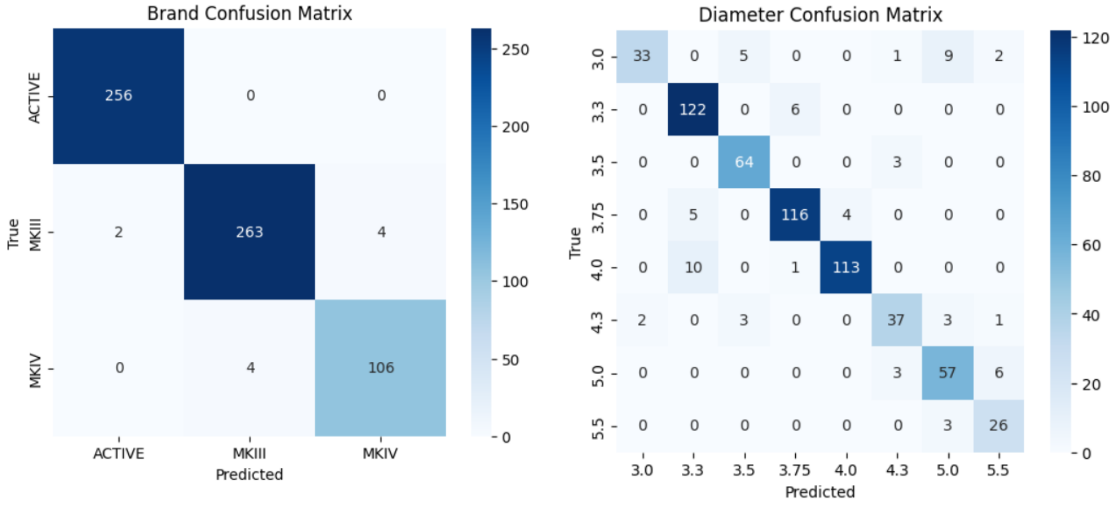


Figure A3. *Train vs validation accuracy for diameter classification*

9.2 Appendix B: Confusion Matrixes

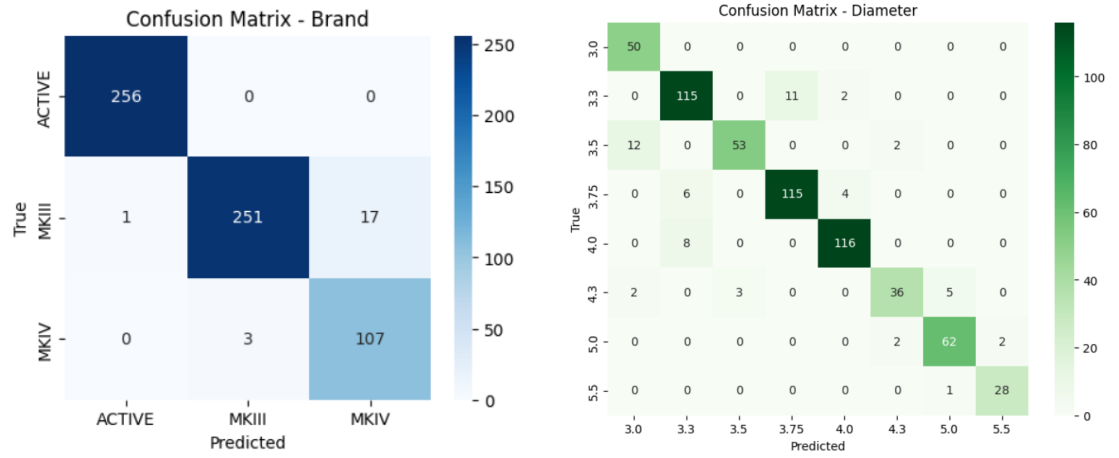
9.2.1 GNN + ResNet-18



a b

Figure B1. *GNN + ResNet-18 confusion matrixes. a, model confusion matrix. b, diameter confusion matrix*

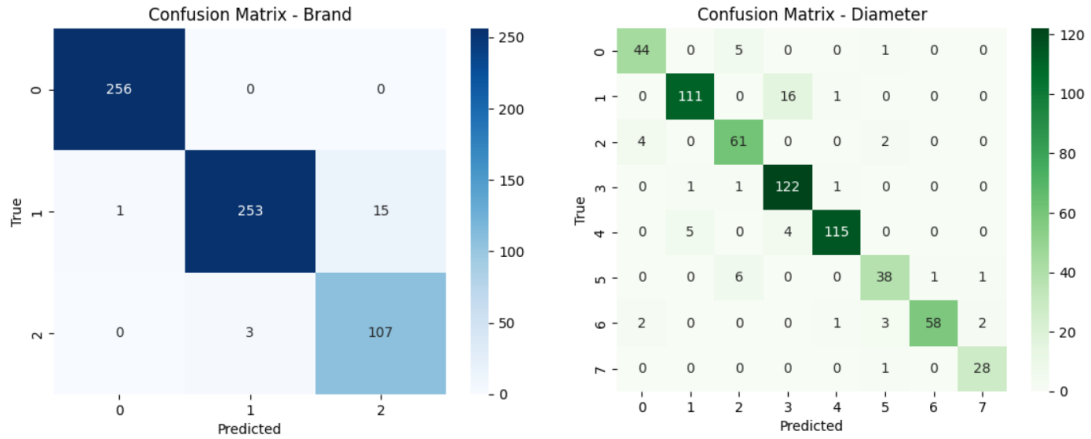
9.2.2 TabNet + EfficientNetB2



a b

Figure B2. *TabNet + EfficientNetB2 confusion matrixes. a, model confusion matrix. b, diameter confusion matrix*

9.2.3 EfficientNetB2



a b

Figure B3. *EfficientNetB2 confusion matrixes. a, model confusion matrix. b, diameter confusion matrix*