The Development of an Automated System of Deep Learning Computer Vision Models for Dental Anomaly and Disease Diagnosis on Panoramic Radiographs

James Faure - 20844700

April 2021

1 Problem background and origin

Analysis of the orthodontic region can be a complicated and time-consuming process for any dentist or radiologist. A common procedure used to complement the clinical examination of a patient is to take a panoramic radiograph. A panoramic radiograph is a two-dimensional X-Ray that examines the full mouth in a single image. Features including teeth, upper and lower jaws and tissues can be studied by observe a panoramic radiograph. The examiner must carefully check each area of the radiograph in search of tooth anomalies, diseases and infections.

The problem arose from the current status of radiographic interpretation in large scale orthodontic radiology screening centres across South Africa. Due to the time-consuming nature of reporting and analysis of panoramic radiographs, and the shortage of qualified dental radiologists, anomalies in many of these radiographs are left undiagnosed. According to Dr Shoayeb Shaik of the University of Pretoria, on average, a screening centre in South Africa has on average 120 patients per day with only one or two orthodontic radiologists available for diagnosis. The under staffing in the screening centres results in only around 30 patients getting adequate reporting. The radiographs are screened, and if there are no significant findings, no report is compiled. There are also medico-legal implications if a disease or lesions are not reported. The solution is to build a model that allows automatic reporting on radiographs during the screening process.

The task is to develop a deep learning algorithm to identify and characterize objects in a panoramic radiograph. The objects range from the scientific naming convention of teeth, to various diseases or anomalies that are present in the mouth. The area of study, within deep learning, is called computer vision. Computer vision is an artificial intelligence process that trains computer to be able to interpret visual objects by observing the pixelation in images.

In order to develop a well-working computer vision predictive model and system, the following high-level steps must be taken:

1. Obtain access to raw image data.
2. Pre-process the raw data.
3. Annotate the data by adding classifications or feature labels.
4. Split the labelled data into three subsets, namely, a training set, a validation set and a testing set.
5. Use the training set to develop a predictive model.
6. Use the validation set to evaluate the generalization performance of the prediction model.
7. Repeat the training process several times to test for model stability.
8. Test the model on new data.
9. Deploy the models in an automated pipeline within a system architecture.

The following research proposal will give an in-depth view as to how the system can be developed.

2 Preliminary literature review

An in-depth literature review must be conducted over three distinct fields of research. As the dataset consists of images the first section will study the formation of digital images. Next, research will be conducted into the fields of deep learning and computer vision to better understand the technologies available for image analysis. Further research into the field of dentistry will help better understand the domain of image dataset. Finally, research must be conducted to understand the most efficient way to integrate all the models into an automated pipeline.

2.1 Images and image preprocessing

Before discussing the different areas of the project, it is necessary to understand the various aspects of images.

An image is a two-dimensional visual representation of a physical array of smaller two-dimensional squares, which picture elements (pixels) [4]. The number of pixels an image has correlates to the resolution. The resolution is an averaging of the signal across a finite area [8]. A level of greyness is assigned to each pixel in the image, ranging from black to white, where black is zero and white is 255.

There are many different types of image files. In this project, the original images are in Tagged Image File Format (TIFF) but must be converted to Portable Network Graphics (PNG) for processing. TIFF images are of large file sizes as the image data contains vast details. PNG images are compressed files which still allow for a full range of colour.

2.1.1 Data preprocessing

Image processing is the improvement in the appearance of images, and the preparation for measurement of the features and structures which they reveal [8]. Revision of the appearance involves cleaning methodologies, such as correcting image defects or image enhancement. Preparing the images includes labelling techniques of features that are apparent. Once the images have been processed, they are ready for modelling.

2.1.2 Annotating the data

As much as 85% of the data in the world exists in an unstructured format [3]. Unstructured data comes in the form of e-mails, contracts, documents, memos, social media feeds, and images, among others. In the new age, data has been seen as the new oil due to it's value in applications such as artificial intelligence. The supervised machine learning process requires this data to exists in a structured format. The process of transforming unstructured data to structured data is called annotating.

A human labeller uses a segmentation algorithm to identify a foreground region in the image [2]. The human demarcates the region of interest with a boundary in the form of hand-drawn lines. The algorithm then assigns pixels to objects based on low-level priorities. Various annotating techniques can be used for labelling. The best option depends on the nature of what needs to be labelled on the image.

Two of the most common approaches to labelling annotations include bounding boxes and polygonal segmentation. A bounding box is a rectangular box that is used to define the location of the targeted object. Polygonal segmentation is used for objects that are not rectangular. Using complex polygons to label an image results in a much more precise location but is more time consuming for the annotator.

A deep learning model then learns the pixelation for each annotation so that predictions can be made on novel data. The annotations are stored in various file formats, such a java script object notation (JSON), which can be used during the modelling process. The structure of the file format will include the image file name, the prediction label, and the geometrical co-ordinates of the demarcated area.

A study by Jain and Grauman was conducted in 2013 at the University of Texas on predicting sufficient annotation strength. The experiment focused on three annotation techniques to determine the best approach based on the difficulty of labelling features in images. The approach of the experiment was to first define the interactive annotation techniques, to define features that reveal image difficulty, then to predict annotations, and finally to determine the cost implications. The techniques observed were a bounding box, where the annotator provides a tight box around the foreground object, a tight polygonal where a tight polygon is drawn on the exact boundary of the feature, and a "sloppy contour" where a rough outline is drawn around the objects (not to precision) [2]. The performance of the segmentation was directly related to the degree of separation between the foreground and background regions. Four features were used to measure this degree, namely colour separability, edge complexity, label uncertainty, and boundary alignment and object coherence [2]. A support vector machine[1] (SVM) model was deployed to perform predictions on the new images. The algorithm predicted where the object might be located. The results of the experiment showed that the tight polygons yielded better accuracy than the bounding boxes, but was a longer, and hence more expensive process to execute.

Semantic segmentation is a less common but more precise technique. The process involves the annotation of features on a pixel-level. Every pixel is assigned a class that carries semantic meaning. Semantic segmentation relies heavily on pixel-level annotations and requires costly labelling efforts [9].

The annotation techniques discussed have become easily accessible as many different labelling software options have been developed. The labelling software can be accessed via cloud-based computing or downloaded as an open-source program. A commonality between the different options is the way the human operator interacts with the program. The interface allows a human to apply the annotation technique of choice to the feature of interest. Some of the common programs used for annotating are; CVAT[2], LabelMe[3], Roboflow[4], LabelBox[5], VGG image annotator[6], and LabelImg[7], among others.

Once the labelling process is complete for a given image, the annotation must be compiled into an easily interpretable file format. A JavaScript Object Notation (JSON) file is one of the preferred formats. The file creates a dictionary with the location of the file and the coordinates of the annotation. The JSON file is easily interpretable and manipulated using a Jupyter notebook or other programming software.

2.2 Deep learning and Computer vision

Deep learning is a subfield of artificial intelligence that "allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction" [6]. A common deep learning algorithm, used for image data, is called a convolutional neural network (CNN).

According to [5] a convolutional neural network is a specialized type of neural network that process data in a grid with typology. Time series data would be an example of data that exists in a one dimensional grid, where the pixelation of image data is an example of two dimensional grid typology.

2.2.1 Convolution neural network

The following section on convolutional neural networks (CNN) gives a high-level description of the technology. During the actual report, the technology will be discussed in much more depth.

A CNN is a class of deep neural networks that is applied to analysing images. The algorithm takes an input image and assigns importance to the pixelation of the features and objects. The primary purpose is to find features, put them in a feature map, while still preserving the pixelation. A CNN has four layers, namely, convolutional layers, ReLU layers, pooling layers, and a fully connected layer [1]. This section of the research proposal will give a high-level overview of the different layers, further research and understanding will be conducted in the later research.

Figure 1 shows the process of the convolutional layer. Features are extracted from an input image and create a feature map by applying many different filters. The feature map is made up of an array of numbers that represent the depth of a filter. The filter slides across that receptive field (image), and in each position, the filter multiplies the value of the filter by the original values of the pixels. The multiplications are summed up, creating a single number, which represents the area just filtered. The CNN learns the values of the filters during training. A feature map is smaller than the original image and therefore easier and faster to process. Information is lost during this process, but features are still able to be detected. The model will not detect

[1] A machine learning algorithm use for binary classification [7].
[2] https://cvat.org/
[3] http://labelme.csail.mit.edu/Release3.0/
[4] https://roboflow.com/
[5] https://labelbox.com/
[6] http://www.robots.ox.ac.uk/ vgg/software/via/
[7] https://github.com/tzutalin/labelImg

every pixel, but will instead detect features made up of a variety of pixels. This allows for the detection of many different features which the model can use to learn.

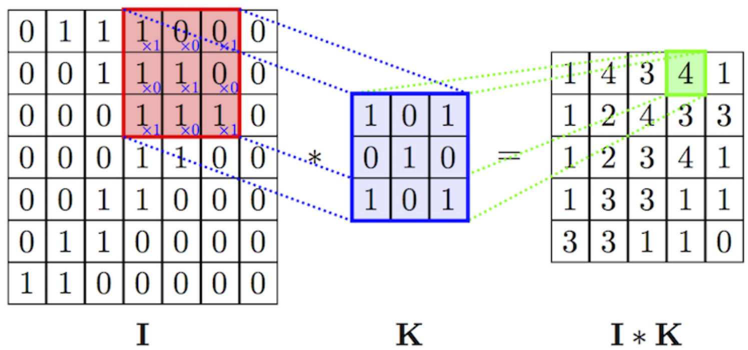


Figure 1: Filter K is an array of 6 elements, which slides over each array of 6 elements in the Image I and creates feature map I * K.

The rectified linear unit (ReLU) layer applies an ReLU activation function to the feature maps. This increases the non-linearity of the network. The function removes all negative pixel values from the activation map by setting them to zero. This changes the negative pixel values to black. The ReLU layer trains the network faster, without hindering the generalization accuracy.

The pooling layer adds flexibility to the CNN. It does this by giving the model spatial variance. This trains the model to recognize the same features in many different forms of the images, hence dealing with overfitting. Data Augmentation is when images are manipulated in one way or another to increase the size of the dataset. The pooling layer will deal with difference in data so that no matter the orientation or contrast, the features are always recognized.

Finally, the fully connected layer adds on an artificial neural network. The primary purpose of this is to combine features into more attributes, which will predict classes and objects with high accuracy. This is done by back propagating the calculated error so that the weight of the model can be adjusted. This is repeated until the model is fully optimized.

2.3 The Field of dentistry

Understanding the nature of the problem requires an insight into the field of dentistry. As mentioned in section 2.1.2., the labelling process will consist of annotating features that are apparent on a panoramic radiograph. The information in the following section on dentistry derives from an interview conducted with Dr Shoayeb Shaik. Dr Shaik graduated with a Bachelors degree in dentistry from UWC in 2003, completed a postgraduate diploma in oral and maxillofacial radiology in 2010 and was awarded a masters degree in 2013. He joined the department of radiology at UWC as a sessional staff member in 2005, followed by a stint in private practice as a dentist from 2006 to 2009. Dr Shaik has been in academia since then, sharing his knowledge and experience with aspiring dentists. Dr Shaik is currently at the university of Pretoria and is also a co-supervisor on this project.

Figure 2 shows the human mouth, which in most cases has 32 teeth, 16 at the top and 16 at the bottom. The top and bottom are divided symmetrically so that the mouth is displayed in quadrants, where each quadrant has eight teeth. The anatomy of each tooth can be further broken down into two key elements, as seen in Figure 3, namely the crown and the root. The crown is above the gum, whereas the root is below. When seen on a radiograph, the enamel and dentine appear most prominent as both are made of a dense, bone-like substance, therefore absorbing the most radiation.

As part of the Data Science process, features within the dental area have to be labelled. This will include labelling teeth according to the ISO dental notation, as well as the presence of elements that suggest anomalies, lesion, disease or manifestation of systemic conditions. The anomalies are related to size, position,

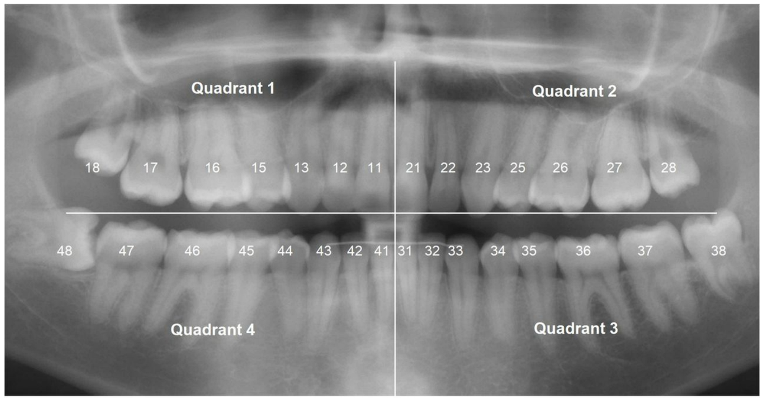


Figure 2: The human mouth is divided into four quadrants, each of which having eight teeth.

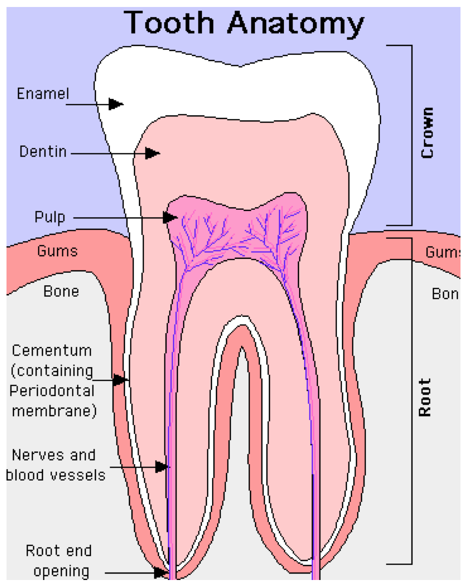


Figure 3: The anatomy of a tooth. (Source: `https://www.enchantedlearning.com/subjects/anatomy/teeth/toothanatomy.shtml`).

number and structure of the teeth. Teeth may also be infected, such as having cavities or caries. Figure 4 shows a image from the dataset that will to be labelled.

3 Problem statement and research questions

After conducting a review of various literature sources and consulting professionals in both dentistry and data science, the problem statement is as follows:

A series convolutional neural networks must be trained in an effort to build a highly accurate computer vision models. The models will be able to detect anomalies and diseases, such as impactions and diabetes, that exist by observing panoramic radiographs. The series of detection models will then be integrated into an automated system that can instantaneously provide the user with a diagnostics report.

The following research questions have been formulated:

1. Within the scope of computer vision there are a variety of different convolutional neural network

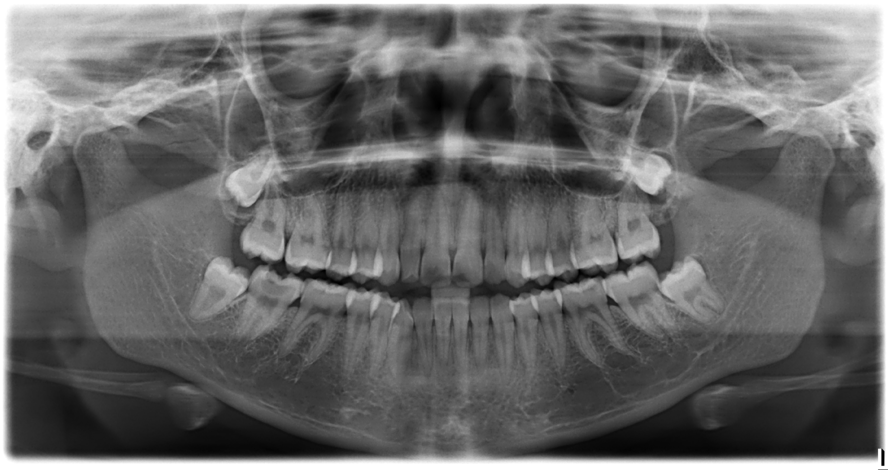


Figure 4: A panoramic radiograph from the dataset that will need to be annotated.

architectures that can be trained. What is the most effective and efficient convolutional neural network that can be used to detect features on a grey-scale panoramic radiographs.

2. After deciding on an appropriate CNN, a variety of different models will have to be trained on a number of different datasets. What are the most appropriate evaluation metrics to assess model performance, and what is a realistic benchmark to base successful generalization performance on?

3. Once the models have been trained, they will be ready for deployment. What will be the most streamlined software architecture that can be built to automatically deploy the models for inference on a commercial and scalable scale.

4 Research objectives

After extensive research into the project, the following objectives have been defined:

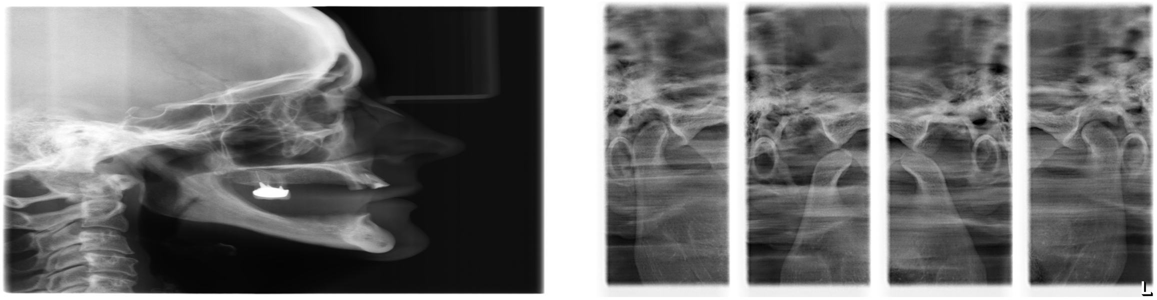
First and foremost an in-depth literature review will cover the range of topics that must be understood in the scope of the project. The literature review will discuss the formation of digital images, deep learning and computer vision, an insight into the field of dentistry, system architecture, and commercial viability.

An essential part of any data science project is data cleaning and manipulation. In the case of this project, the images need to be manipulated before the labelling can begin. Research has found that in order to process the images through a CNN, they must be a png or jpeg file instead of a TIFF file. A python script should therefore be developed to assist in converting images from TIFF format to png.

The radiographs must go through two separate labelling processes. The first process is an image classification where the class of an image will be identified as either “workable” or “unworkable”. An X-Ray is classified as “workable” if the teeth and jaw area are clearly seen from a front view of the mouth. Figure 4 above shows examples of workable X-Rays. An X-Ray is classified as “unworkable” if the image is not a panoramic radiograph, a blurry or inconclusive view of the mouth, or when there is no clear visual at all. Figure 5 shows a subset of “unworkable” images. An image classification model must then be built to train a convolutional neural network. The model will be able to detect whether an image is workable or unworkable.

The “workable” images then will go through a second labelling procedure. A rectangular box must be manually drawn, by a qualified dentist, around each tooth or feature of interest and labelled accordingly. These labels will be used to identify:

- Teeth - labelled according to ISO dental notation.



(a) "Unworkable" because it is not a panoramic view (b) "Unworkable" because it is not a panoramic view

Figure 5: "Unworkable" radiographs that must be removed from the dataset.

- Cavities and caries can be show areas of disease.
- Anomalies related to size, position, number, shape and structure of teeth.
- Areas of the lesion.
- Areas of infection.

The number of dataset that can be labelled are dependent on the time availability of the labeller. Within the medical domain, it is essentially that the data is labelled by a qualified dentist to avoid inaccuracies. It would be ideal to have a portion of the data labelled so that the testing phase can begin. The development of several models will mean that the system architecture can start to be developed. Future labelled dataset can be trained and easily deployed into the ready built system.

The newly labelled dataset will be used for an object detection algorithm. Each model should be trained several times to ensure that the evaluation metrics are stable and accurate. Transfer learning can be applied to improve a model's accuracy. Transfer learning is the process of using an existing pre-trained model, along with augmented data, to train a new model. The pre-trained model will make it easier for the new model to learn shapes and specific objects.

Finally, once we are satisfied with the model's performance, we can deploy the model for inference. A user-friendly program must be developed which either take in a batch of images or images in real-time. Figure 6 shows a diagram of the automated pipeline that will be used to integrate the different models. An image will enter in an image collection porthole and saved to a database with an unique id. The image will then be converted from the original format to a png file and saved to a new database. Next, the image will be checked for quality to ensure that it is good enough for further diagnostics. An image that is good enough to continue down the pipeline can be seen in Figure 4. If an image is unworkable, the user will be prompted to take a new one. The prediction label will be saved with the image id to a new database. Images that are predicted as "workable" will progress to the pipeline of anomaly and disease detection models. The models will run simultaneous and all predictions will be saved to a new database. The predictions for each image will be compiled into a user report that will be displayed to the radiographer. In a system that has autonomous entries of multiple radiograph each second, radiographs that have severely rated predictions will be flagged for further diagnostics.

5 Proposed research approach, design and methodology

There are two aspects to the project. The first is the data science perspective where image data will need to be cleaned, labelled and trained using a CNN.

Figure 7 displays the labelling diagram that is going to be used for the project. The diagram is divided into three levels, each representing a different labelling job. Level 1 is an image classification job where an

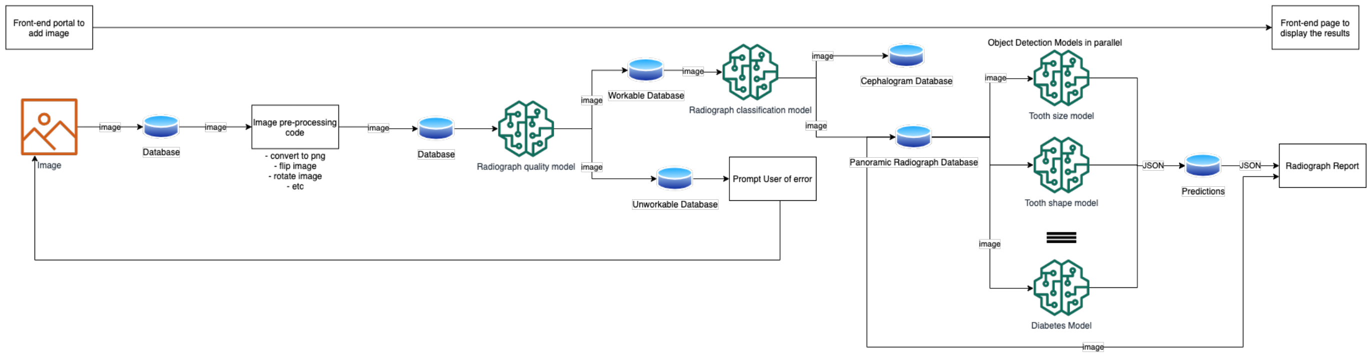


Figure 6: A diagram showing a pipeline of the integrated models.

entire image will be labelled according to a "workable" or "unworkable" class. Level 2 is where objects are labelled. A bounding box (rectangular or polygonal) will be used to mark out the features of interest. The annotating process can be see as the process of transforming unstructured data into structured data. The data labelling is expected to be the most time consuming part of the project, and will need to be done by a professional dentist to avoid inconsistencies in the data.

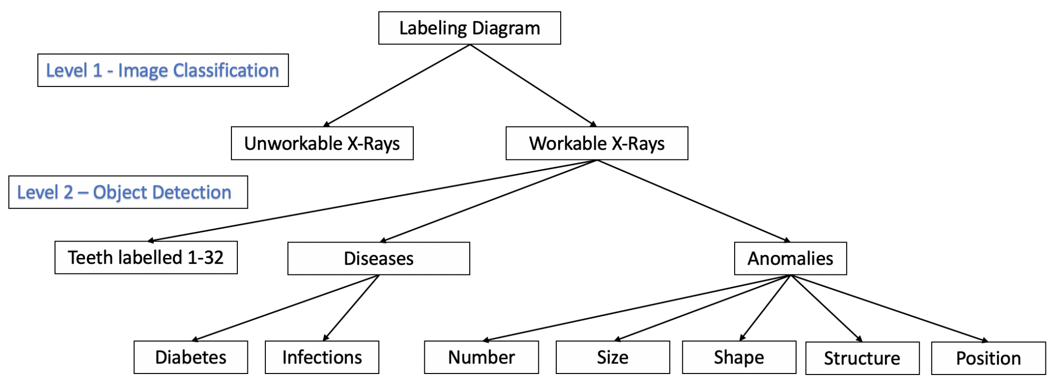


Figure 7: A diagram showing the labelling technique that will be deployed to annotate the X-Rays.

Figure 8 shows the scope of the entire project. Similarly to the labelling process in Figure 7, there are two stages to the process diagram. The first being image classification and the second being object detection. The process starts with raw, unstructured data in the form of radiographs, which will need to be labelled according to Figure 7. A CNN can only process a PNG or JPEG image as an input, therefore images must be processed and manipulated in order to be in the correct format. Next the images are labelled for image classification. By referring back to Figure 7, the two broad categories are workable or unworkable. After labelling each image in dataset, the labels are evaluated and any data augmentation techniques are applied to increase and enhance the training set. Augmentation techniques help reduce overfitting when the model is being trained. Once the training set is satisfactory, an image classification model must be trained using a CNN. After training, the model is evaluated. A satisfactory model will have an accuracy above 90%. If the accuracy is less than 90%, the labels and training set can be further enhance with additional augmentation and image preprocessing. If the accuracy is above 90%, the model can be deployed and the workable image progress to the next labelling process. The next stage of object detection is similar to the image classification process, except now features of interest are labelled using rectangular bounding boxes. The box is manually demarcated around an object and the coordinates of the box are saved along with the annotation label. The images then progress to augmentation and training. Finally, the model is evaluated to ensure satisfactory validation results. The model is then saved if it is of satisfactory standard, and can be deployed into the system architecture.

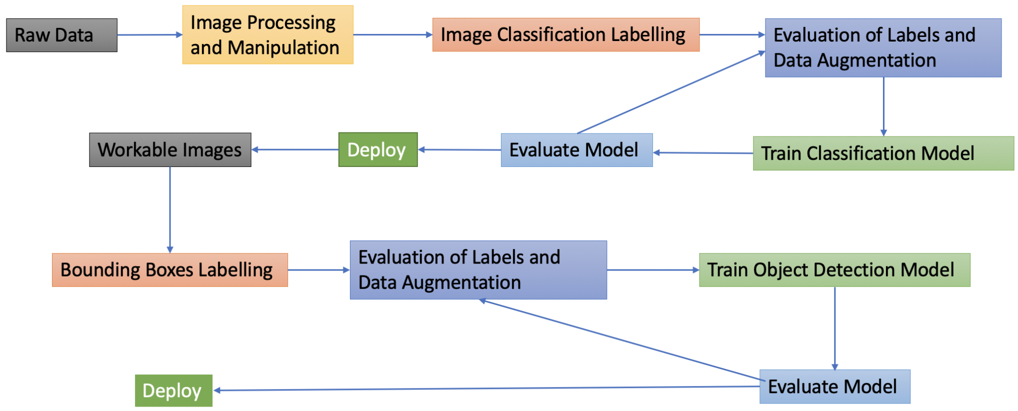


Figure 8: The computer vision process within the scope of this project.

Referring back to Figure 6 shows a design of the system that is to be developed. The system will integrate all the models in a seamless pipeline. The final program will take a unlabelled input radiograph and simultaneously run all the developed models to perform a diagnostics. The diagnostics report will be displayed to the user on screen. Depending on the diseases and anomalies detected, the input image will be assigned a severity score. A radiologist will be notified of patients with high-severity for further diagnosis and analysis.

There are many methods of application deployment that can be researched. One particular method is to use Amazon Web Services (AWS) severless to quickly develop and deploy the application. AWS severless allows a developer to use the AWS services to manage the flow of data within the backend of the application. Figure 9 shows the integration of different AWS services that can be utilized to deploy the application. AWS amplfy is a service that is used to integrate the frontend code with a backend service. The frontend code will be hosted of a versioning control service, such as Github, and called directly through amplify. An image collection bucket, deployed on the frontend, will be use to pull images into an S3 database. S3 is a cloud-based storage service that is offered by AWS. The raw image files will remain saved in an S3 bucket. Three lambda functions will be required for the project. Essentially a lambda function acts as the glue between different AWS services where code can be hosted. The first lambda function needed will be used for pre-processing the images. Each image must be converted from its medical form, such as a TIFF image, to a png file format. The png files must be saved to a new S3 bucket. The other two lambda functions will be used for the predictions. The png image file will be pulled into the lambda function and the model will make predictions. DynamoDB is the final AWS service in this architecture. DynamoDB is a NoSQL database for saving the prediction data. Each image prediction will be saved to this database system. Predictions can then easily be displayed on the fronted using the amplify function.

6 Scope, limitations, assumptions

The project will begin with an in-depth literature review of various sources within the scope of the project. First, the structure of images will be explored in order to learn more about the data that will be used throughout the project. Next, the deep learning technologies available to create a learning model must be researched in order to know what the most efficient experimenting method will be. Finally, research must be conducted into the field of dentistry. A data scientist must always seek to understand the domain in which they are operating in. Additional research can be conducted to the development of the automated a system. This will involve researching the best data stores, system architecture and integration of multiple models.

Following the research process, development will begin. The project will require two types of models to be developed, namely, image classification and object detection. In both cases, image data will need to be labelled. The labelling process could potentially be the first limitation of the project. Although the scope

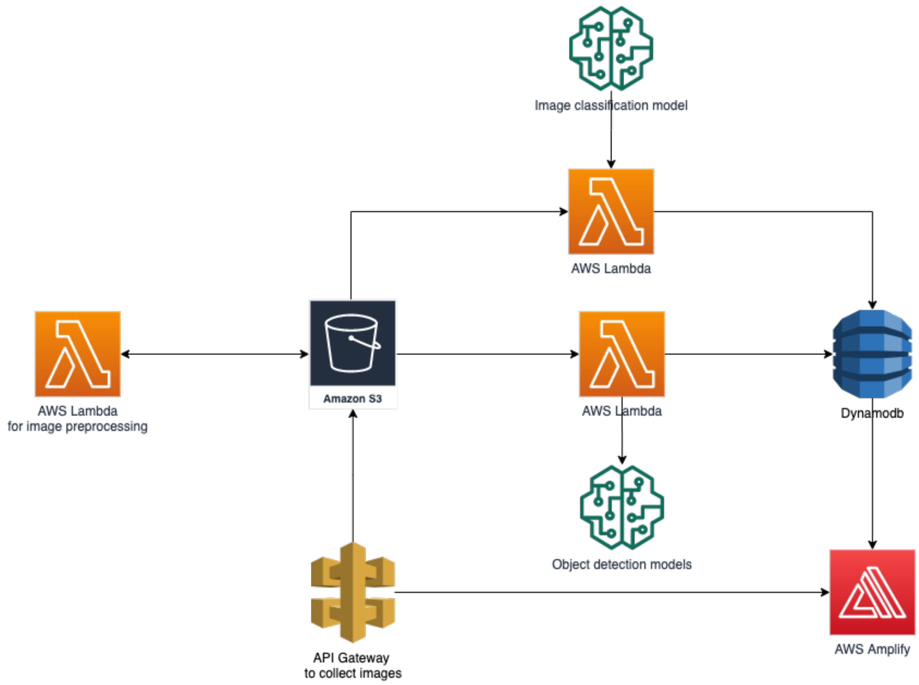


Figure 9: The AWS serverless architecture.

project falls within the deep learning domain, the data is specific to the domain of dentistry. The extent of labelled features is vast, meaning that many different models will be built. Although this will increase the density of feature learning, the time availability of qualified dentists to label the data may be limiting. With over 30 possible anomalies, and even more diseases, an image could take between two and twenty minutes to label. The best labelling strategy may be to label images within their anomaly or disease class, and test models every 400 labelled images. This will allow for model development to happen at a quicker rate, hence not limiting the development of the project. Once a model has been built, it can be saved and stored until the automated system is developed. And every time more data is labelled and another model is trained, it too can easily be integrated into the system's architecture. This process will be an iterative process of development. It will allow for multiple parts of the project to simultaneously taken place, hence never limiting the progress of the project.

7 Ethical implications

Due to dataset of the project using medical data, the project will need an ethical clearance through the HREC committee at Stellenbosch University. The X-Rays can be seen as identifiable human data, which raises ethical concerns. All data must be anonymized in order to eliminate any direct association with a human. In addition, it must be ensured that all data is kept secure on a database that cannot be accessed by the public. Any labelling tools or machine learning software must use login credentials for anyone assisting with the project.

8 Expected contributions

From the perspective of Dr Shaik, the intentions of this project are, firstly, to inform the field through publication and, secondly, to develop a working program for institutions and the country to adapt to current programs.

The project is expected to make contribution is five key areas.

The first area is the development of several computer vision models that are able to detect anomalies and diseases in the mouth and jaw. The number of models depends on the grouping of different anomalies and diseases. A model will be developed for each group and then deployed in an automated pipeline.

The second contribution will be the integration of all developed models into an automated pipeline. Software architecture will be built to house the models, so that a novel image can seamlessly enter the system and have a diagnosis performed. The results of the diagnosis will be neatly presented to the radiographer.

The third contribution that this project is expected to make is to provide a fully labelled dataset for any future projects that involve dentistry radiographs. The labelled dataset will contain a variety of features that could be processed to fit other projects of a similar nature.

The fourth contribution will be a number of conference and article publications that can form the basis of future research in the field of both computer vision and dentistry. These publications will inform people of the best methods for radiograph analysis as well as optimizing performance of convolutional neural networks on image datasets.

Finally, such a program can be readily available in rural areas across Africa, where there is a need for professional dentistry analysis. A national government controlled PACS (picture archiving and communication system) system can be used to access radiographs in remote areas and diagnose them with this product. The program can be accessed via an app linked to provincial government sites that house PACS and HIS/RIS (hospital information system/radiograph information system) systems. Radiographers and referring dentists in the rural areas can upload the images into the app and then basic diagnosis can be provided with the click of a button. Any additional features or diagnosis required can be done on a referral system thus streamlining the workflow, improving patient outflow and throughput.

9 Project timeline

The project began in February 2019 as a final year industrial engineering project. The project achieved great success with a cum laude and recognition as the "most innovative and creative" industrial engineering as a nominee for the Jac van der Merwe competition. Since then the research has been extended into a master in engineering thesis. Up to this point models were developed for both image classification and object detection, hence providing evidence that the algorithms will work on the radiograph dataset.

The research conducted going forward will be conducted into three main areas.

Firstly, a variety of different types of CNNs must be tested on a dataset to determine which will be best for the remainder of the project. The models must be evaluated based on accuracy, computation expense, and robustness to overfitting. This research should be carried out first. A list of different models, as well as evaluation metrics, must be compiled in order to understand how to execute effectively.

Secondly, a larger dataset must be labelled for model development. The labels must include all types of anomalies and diseases that can be detected on a panoramic radiograph. The labelling of the data is expected to take the longest time. The ideal strategy for is to outline what needs to be labelled and compile an instruction document. The labelling job must be spilt into categories of anomalies and diseases so that once a category is finished, it can be developed and tested as a model.

Finally, a model must be trained for each category of anomaly and disease. These models must then be integrated into an automated pipeline as seen in Figure 6. The code for each training instance will remain the same for each model. Therefore, this set just requires the same iterative process across different datasets. The model will be saved an deployed in a pipeline within the system architecture.

10 Proposed table of contents or chapter outline

The sections to be covered are as follows:

I. Introduction

II. Images

III. Deep Learning and Computer Vision

IV. The Field of Dentistry

V. Dataset Preprocessing

VI. System Design

VII. Radiograph Quality Classification Model

VIII. Object Detection for Anomalies

IX. Object Detection for Diseases and Infections

X. Integration of Models into Automated System

XI. Economic Feasibility of the System

XII. Conclusion

XIII. References

XIV. Appendices

Chapter I will be the introduction and will follow a fairly similar layout to the content in this document. Chapter II, III and IV will cover form the basis of the literature review with further literature discussed in chapter VI and XI when applicable. The literature in chapter II will aim to introduce the concept of digital images, how they are developed and the different type of digital images. The chapter will further discuss radiographs and image processing techniques. Chapter III will then discuss deep learning and more specifically computer vision. This will include an introduction to deep learning and how the technologies progressed until convolutional neural network were discovered and developed. Convolutional neural networks will then be discussed in more depth as the backbone for computer vision. Furthermore, the chapter will discuss real-world applications of computer vision and how it can be implemented on different datasets. Chapter IV will give a brief introduction to the field of dentistry as well as the identifying the various anomalies and diseases that can be detected in the mouth. This chapter will be vital as it will put the project in the perspective of the dataset.

Chapter V, VI, VII, VIII, IX, X and XI will cover the development part of the project. Chapter V will showcase different image manipulation and preprocessing techniques that can be applied to the dataset. Chapter VI will discuss the system architecture as an automated pipeline of computer vision models. The content of the chapter will be off the diagram seen in Figure 6. The chapter will also include a further literature review into software development, databases, cloud computing and integrating deep learning models using application programming interfaces (APIs). Chapter VII will cover the first computer vision models that will be developed. These models will be classification models that are able to classify images until in order to find the perfect image that can progress to the object detection stage of the pipeline. Chapter VIII and IX will then cover the object detection models of the pipeline. This will be the panicle of the project. The development of highly robust and efficient object detection models for a variety of anomalies, diseases and infections. Chapter X will follow suit by showing how the models are integrated into a pipeline that instantaneously diagnoses novel data that enters the system. Chapter XI will then finally discuss the economic feasibility of the commercial product.

References

[1] Charu C. Aggarwal. *Neural Networks and Deep Learning*. Springer International Publishing, 2018, pp. 315–371. DOI: 10.1007/978-3-319-94463-0. URL: https://doi.org/10.1007%2F978-3-319-94463-0.

[2] Suyog Dutt Jain and Kristen Grauman. "Predicting Sufficient Annotation Strength for Interactive Foreground Segmentation". In: (Dec. 2013), pp. 1–8.

[3] Adanma Eberendu. "Unstructured Data: an overview of the data of Big Data". In: *International Journal of Computer Trends and Technology* 38 (Aug. 2016), pp. 46–50. DOI: 10.14445/22312803/IJCTT-V38P109.

[4] Andrew J Gerber and Bradley S Peterson. "What is an image?" In: *Journal of the American Academy of Child and Adolescent Psychiatry* 47.3 (2008), p. 245.

[5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[6] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), pp. 436–444.

[7] William S Noble. "What is a support vector machine?" In: *Nature Biotechnology* 24.12 (2006), pp. 1565–1567.

[8] John C. Russ. *The Image Processing Handbook*. 6th ed. New York: CRC Press, 2011.

[9] Yiheng Zhang et al. "Fully Convolutional Adaptation Networks for Semantic Segmentation". In: (June 2018).