UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DOCTORADO EN CIENCIAS DE LA INGENIERÍA CON MENCIÓN EN INGENIERÍA ELÉCTRICA

Deep learning en la detección de cáncer de piel utilizando termografía activa

---

Profesor supervisor: PhD. Sebastián E. Godoy Medel
Departamento de Ingeniería Eléctrica
Facultad de Ingeniería
Universidad de Concepción

Tesis para optar al grado académico de Doctor en Ciencias de la Ingeniería con mención en Ingeniería Eléctrica

RICARDO FELIPE ANTONIO SOTO CATALÁN

CONCEPCIÓN - CHILE

2024

Prefacio

Esta tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias de la Ingeniería con mención en Ingeniería Eléctrica, de la Universidad de Concepción, Chile, y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Ingeniería Eléctrica, durante el período comprendido entre el año 2019 y 2024, bajo la dirección del Doctor Sebastián E. Godoy Medel.

Ricardo Felipe Antonio Soto Catalán
ricsoto@udec.cl
Departamento de Ingeniería Eléctrica
Facultad de Ingeniería
Universidad de Concepción
Concepción, 2024

A mis amorosos abuelitos Ruth y Mario...
A Mariana, mi madre y a mi hermanita Carla por su apoyo incondicional...
A Macarena, por darme su amor y paciencia en todo momento...

Agradecimientos

La realización de esta tesis implicó a muchas personas y entidades que aportaron desde diferentes aristas. Agradezco por su valiosa contribución a:

Agencia Nacional de Investigación y Desarrollo (ANID), que me apoyó financieramente con la Beca Doctorado Nacional Folio: 2019-21191485.

Sebastián Godoy, profesor supervisor de esta tesis, por la oportunidad y la motivación para trabajar en el área de procesamiento digital de imágenes. Por todos los consejos académicos, laborales y personales.

Profesores del Departamento de Ingeniería Eléctrica de la Universidad de Concepción por participar en mi formación.

Don Luis Arteaga, por su buena disposición a ayudar en todas las actividades en que necesité su apoyo. Por su amabilidad y profesionalismo con que se desempeña.

Mis compañeros de laboratorio: Yasmany, Laura, Christian y Felipe por acogerme al llegar a mi entonces nuevo lugar de trabajo.

Mi familia, en especial a mi mamá Mariana, por apoyarme siempre, desde el primer día en que decidí continuar estudiando. De igual forma a mi hermana Carla, que siempre me hace saber lo orgullosa que está de mi.

Macarena, mi otra mitad, que conocí gracias a estar en Laboratorio de Transmisión. Gracias por animarme y apoyarme en todo momento, comprenderme y amarme.

Contents

List of Figures

List of Tables

Abstract

Skin cancer is the most common form of cancer worldwide, with early detection being critical to reducing mortality and minimizing the need for invasive therapies. This thesis proposes an automatic and low-cost skin cancer detection system using active infrared thermography (IRT), machine learning and deep learning. Active IRT leverages controlled thermal stimuli to identify malignant lesions based on unique thermal regulation curves (TRCs) observed in tumor-affected areas.

The study introduces a novel detection scheme that integrates spatio-temporal features extracted from TRCs, visible imagery, and machine learning classifiers. Along with this, the use of a U-Net convolutional neural network for automatic lesion segmentation, trained on high-quality datasets and tested in real clinical conditions, is proposed. Additionally, the research evaluates a range of low-cost infrared cameras and develops a video degradation model to simulate their performance. This evaluation showed that the proposed detection algorithm using expert segmentation allows working with low-cost cameras with an accuracy of approximately 85% with mid-range infrared cameras and 83% with a low-end camera.

Finally, the application of transfer learning to TRCs was explored, with a proposed strategy to generate images from the curves. Using a ResNet152V2 network, the curves were classified as benign or malignant, and the average malignancy of the analyzed TRC set was calculated. Incorporating this new feature, the detection system achieved an accuracy of 90.86% with expert segmentation and 89.20% with automatic segmentation. These results suggest that incorporating this feature into the detection algorithm for lower-quality cameras could enhance its performance, potentially matching the accuracy achieved with high-quality video processing.

Resumen

El cáncer de piel es la forma más común de cáncer a nivel mundial, su detección temprana es crucial para reducir la mortalidad y minimizar la necesidad de terapias invasivas. Esta tesis propone un sistema automático y de bajo costo para la detección de cáncer de piel utilizando termografía infrarroja activa (IRT), machine learning y deep learning. La IRT activa permite diferenciar lesiones malignas de benignas, dado que los tejidos cancerígenos presentan un comportamiento de recuperación térmica diferente al de tejido sano.

El estudio introduce un novedoso esquema de detección que integra características espacio-temporales extraídas de las curvas de recuperación térmica (TRC), imágenes visibles y clasificadores de machine learning. Junto con esto, se propone el uso de una red neuronal convolucional U-Net para la segmentación automática de lesiones, entrenada en conjuntos de datos de alta calidad y probada en condiciones clínicas reales. Además, la investigación evalúa una variedad de cámaras infrarrojas de bajo costo y desarrolla un modelo de degradación de video para simular su desempéño. Esta evaluación demostró que el algoritmo de detección propuesto, utilizando segmentación experta, permite trabajar con cámaras de bajo costo con una precisión de aproximadamente el 85% para cámaras infrarrojas de gama media y el 83% para una cámara de gama baja.

Finalmente, se exploró la aplicación de transfer learning a las TRCs, proponiendo una estrategia para generar imágenes a partir de las curvas. Utilizando una red ResNet152V2, las curvas se clasificaron como benignas o malignas, y se calculó la malignidad promedio del conjunto de TRCs analizadas. Al incorporar esta nueva característica, el sistema de detección logró una precisión del 90.86% con segmentación experta y del 89.20% con segmentación automática. Estos resultados sugieren que incluir esta característica en el algoritmo de detección al trabajar con cámaras de menor calidad, su rendimiento podría alcanzar poten-

cialmente el el que se obtiene al procesar videos de alta calidad.

1 Introduction

Skin cancer is considered the most common type of cancer, and its early detection is quite relevant, since it has a direct impact on the decrease in mortality rates [2]. Moreover, early detection provides the additional benefit of requiring less aggressive and lower-cost therapies [3]. It is estimated that globally in 2018, 1.28 million new cases of skin cancer were diagnosed, with 1 million corresponding to non-melanoma skin cancer and 280,000 to malignant melanoma (MM), which are associated with 65,000 and 60,000 deaths, respectively [4].

Currently, skin cancer is diagnosed by anatomical pathology analysis of the biopsy of each suspicious lesion, normally indicated by a dermatologist. Typically, a suspicious lesion is analyzed by a dermatologist by performing visual inspection based on Asymmetry, Border, Color, Diameter and Evolution over time using the ABCDE rule [5]. In addition, the specialist considered the patient's medical history and other factors to determine if the lesion was probably cancerous and referred the patient for a biopsy.

A biopsy is a surgical intervention; therefore, it is an invasive and expensive procedure. Normally, there is a waiting list for patients requiring biopsy because of the limited availability of both specialized professionals and resources. Moreover, several patients are referred for biopsy solely because the practitioner may want to avoid false-negative identification of some lesions [6]. Several technological approaches have been developed to optimize resources and minimize the number of patients for whom biopsies can be avoided. These technologies include dermoscopy [7], developments based on thermography [8, 9, 10, 11, 12, 13, 14], visible image analysis systems [15, 16, 17, 18, 19, 20], ultrasound-based tools [21, 22, 23], and multispectral digital skin lesion analysis [24].

Dermoscopy is commonly performed by dermatologists. This technique enables better visualization of the spatial patterns and structures of lesions. When compared with naked-eye evaluation, it increases the sensitivity from 71% to 90%

and the specificity from 81% to 90%. Moreover, dermoscopy decreases the biopsy ratio from 15:1 (1 of 15 exams corresponds to MM, while the others are benign lesions) to 5:1 [4]. A key drawback is that specialists require years of training to achieve this performance. Motivated by the importance of improving diagnosis, several tools have been developed based on visible image analysis and infrared thermography (IRT).

Technologies that analyze visible images have been of great interest to the scientific community in recent years. Various systems have been developed to assist in the diagnosis of skin cancer using machine learning techniques, such as support vector machines, artificial neural networks, and K-nearest neighbors [25]. Currently, deep learning techniques are the most commonly used because of the successful processing of visible-range images. There are different deep learning approaches that involve pre-processing stages, such as artifact removal, contrast enhancement, and the use of different color spaces [26]. Deep learning approaches have proven to be effective in two main areas: lesion segmentation and classification of skin cancer lesions [27, 28, 29, 26]. First, the results reported in the literature demonstrate that deep learning can be effectively used to detect the contour of lesions versus the surrounding skin. The similarity reaches a Jaccard index of up to 0.9 [30].

Second, outstanding results have been obtained by classifying skin lesions, particularly when using strategies that involve pre-trained networks, exceeding 90% accuracy [17, 18, 31, 32, 26, 33, 34]. However, most of these studies focused on the problem of detecting melanoma lesions, unlike this study, which sought to detect any type of malignant lesions.

Another approach explored for skin cancer detection is the use of IRT, which can be either passive or active [35]. The main difference between these techniques is that active thermography requires a thermal stimulus that produces a thermal contrast between the area of interest and the background. Passive thermography is applied to problems where the area of interest is naturally at a higher or lower temperature with respect to the background [36]. Passive IRT has not been successful in detecting skin cancer lesions [37], and no developments related to skin cancer have been observed using this technology until a few years ago. Through digital image processing techniques and statistical analysis, it was determined that passive thermography images can distinguish cancerous from healthy skin areas [38, 39]. For example, Magalhaes et al. [14] proposed a deep learning configuration to process passive IRT. This network could differentiate melanoma from nevi and melanoma from non-melanoma skin cancer lesions with precisions of 96.65% and 92.59%, respectively. However, the performance was inferior when working on the problem of distinguishing benign from malignant lesions, with a sensitivity of 62.81% and specificity of 57.58%. In contrast, active IRT has been successful in the detection of skin cancer [8, 9, 10, 11, 12, 13], with sensitivity and specificity

of up to 99% in sensitivity and specificity [12]. Another advantage of active IRT is that it provides useful information for lesion modeling, which provides more insight into lesion morphology. Verstockt et al. [40] proposed lesion modeling using the finite element method (FEM), a technique that has been successful in modeling different structures present in the human body [41, 42, 43]. Skin cancer detection can be considered solved based on the results reported by Godoy et al. [12]. However, this method and other methods based on active thermography require expert knowledge of the selection of lesion and non-lesion areas, as these systems are sensitive to the operator. Moreover, machine-learning techniques have not been well explored in the detection of skin cancer using active thermography, probably owing to the key aspects of data pre-processing for successful detection.

In spite of decades of research and development of medical applications based on IRT, its massification has been limited mainly due to the high cost of IR cameras. According to Narayanamurthy et al. [44] IRT needs optimum instrumentation for the recording purpose, given that it is widely affected by the external noise. Thanks to the continuous development of electronic technology, new structures of arrays of detectors and new semiconductor alloys are available. This has reduced the costs and size, increase the resolution and the precision of the IR devices [45, 46]. The development of longwave infrared (LWIR) microbolometer technology stands out of which multipurpose low-cost cameras have been developed, which work along with smartphones [47] and in several other applications.

Recently, there has been a high interest in the development of medical applications based on low-cost IR cameras, with tests being conducted mainly to support the evaluation of diabetic foot [48, 49, 50, 51, 52, 53, 54, 55]. Studies have also been conducted that aim to evaluate skin burns [56] and to assess the healing progress of thoracic surgical incisions [57]. Villa et al. [52] characterizes and compares the low-cost cameras ST and Thermal Expert TE-Q1 Plus (TE-Q1), using as a reference the high-end camera INO IRXCAM-640. This study considers that camera TE-Q1 is suitable for e-health applications, particularly in the assessment of diabetic foot ulcers. Due to the quality of the measured noise equivalent temperature difference (NETD) and the residual non-uniformity (RNU) validated at 25 °C, it is comparable to the IRXCAM-640 camera.

In this study, an automatic image-based skin cancer detection scheme is proposed for a real clinical environment, where lower-quality images, partially blurred areas owing to patient movement, poor illumination, and similar factors are common. This scheme provides a key pre-processing stage for the extraction of discriminative features, along with the proposal of a pool of features that can be used with any machine learning technique, such as a support vector machine (SVM) classifier. Another novel point addressed in this work is the use of a convolutional neural network (CNN) U-Net to segment the lesion area, which was trained to work with low-quality imagery. This was performed to assess the performance of

the algorithm in the worst-case scenario. In another aspect of the research, the noise present in the microbolometer IR cameras is analyzed and from these observations an IR video degradation model is proposed. Using the degradation model, the feasibility of using three IR cameras on the skin cancer detection problem is evaluated by adaptation of a dataset to the characteristics of the different cameras and processed by the proposed detection scheme. Finally, a classifier for TRCs is proposed, which discriminates between malignant and benign classes. Using this classifier, a metric is introduced to describe the malignancy of the IR video. This new feature is evaluated both individually and in combination with other previously selected features.

1.1 Hypothesis

It is possible to implement a low-cost skin cancer detection system using active thermography with sensitivity and specificity higher than 95%, utilizing a spatio-temporal scheme based on deep learning.

1.2 Objectives

General objective:

To develop a spatio-temporal scheme integrating the information provided by thermoregulation curves and a visible image of the suspicious lesion using deep learning to demonstrate the feasibility of using low-cost IR cameras for skin cancer detection.

Specific objectives:

i) To evaluate different skin cancer detection algorithms based on deep learning.

ii) To develop and implement an automatic skin lesion segmentation algorithm using CNNs on visible spectrum images.

iii) To develop an algorithm for classifying thermoregulation curves extracted from a sequence of IR images using CNN.

iv) To develop a skin cancer detection algorithm based on the analysis of thermoregulation curves using CNNs and spatio-temporal analysis.

v) To evaluate the performance of the various developed algorithms on IR videos captured with cameras of different ranges.

1.3 Contributions of this dissertation

This thesis presents 7 contributions related to the detection of skin cancer using infrared thermography and the use of low-cost IR technology.

The first contribution is an algorithm that performs the automatic selection of the lesion area by using a CNN U-Net. This algorithm segments the lesion area from the surrounding tissue in low-quality visible images, as shown in Fig. 1.1.c. As a consequence, a previously-published skin cancer detection algorithm was improved, such that the resulting algorithm has no human bias.

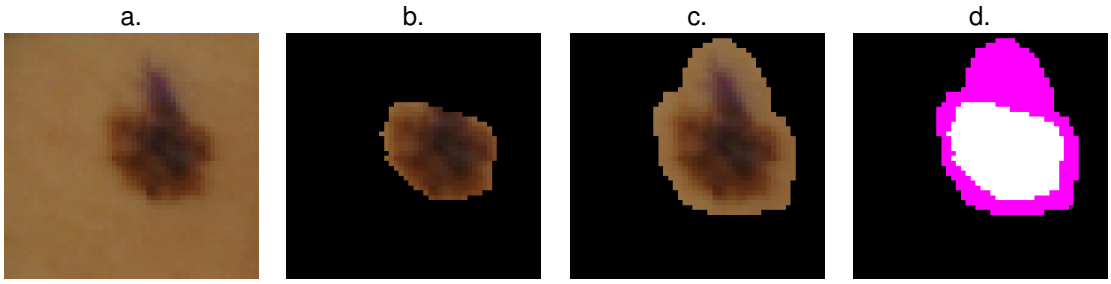

Figure 1.1: Sample of automatic lesion selection using the CNN U-Net. a. Image with the lesion to be segmented, b. lesion manually selected based on expert criteria, c. lesion automatically selected with CNN U-Net, and d. comparison between expert and automatic segmentation.

The second contribution is the development and evaluation of a pool of spatiotemporal features extracted from TRCs, which are related to the energy, thermal evolution, and similarities of the TRCs. These features are calculated either on representative TRCs from each video or on a set of TRCs. When applied to a set of TRCs, features based on statistical measures such as mean and standard deviation were proposed.

The third contribution is the evaluation of different machine learning techniques together with the proposed feature pool. Among the classification techniques evaluated are: K-Nearest Neighbors, decision trees, random forests, support vector machines and extreme gradient boosting. Table 1.1 demonstrates that incorporating machine learning techniques along with feature selection enhances the specificity of the Euclidean Detection (ED) algorithm by approximately 45%, without compromising its sensitivity.

Table 1.1: Performance of the detection algorithm when working with a set of proposed features and different classification techniques.

Classifier	**TPR (%)**	**TNR (%)**
ED	86.15	42.72
KNN	70.87	83.44
SVM	82.09	87.12
XGBoost	85.18	86.05
Ramndom Forest	87.26	87.39

The fourth contribution corresponds to a methodology to extract features from thermoregulation curves using deep learning. Since deep learning techniques are more developed for image classification, a strategy was developed to generate an image from a thermoregulation curve, which is subsequently processed by a ResNet152V2 network, which classifies an image as benign or malignant TRC. The proposed feature measures the average malignancy of the set of TRCs contained in an IR video. Figure 1.2(a,b) presents a sample of representative images of benign and malignant TRCs, respectively.

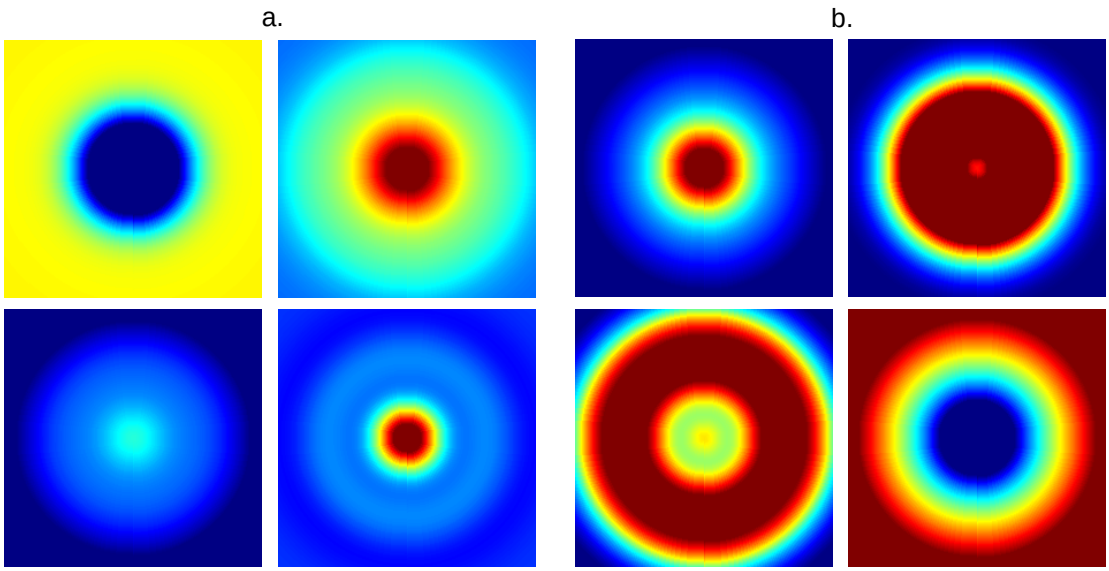

Figure 1.2: Sample circular images generated from TRCs. a. Representative images of benign TRCs, b. representative images of malignant TRCs.

The fifth contribution corresponds to the characterization of spatial and temporal noise present in microbolometer cameras. The identification of the low frequency temporal noise that affects the cameras, which had not been identified

in the literature and is of great importance in applications that use active thermography, stands out. Figure 1.3(a-c) shows the low-frequency noise present in the Xenics, Opgal, and Seel cameras, respectively. It is observed that this noise is correlated with ambient temperature fluctuations.

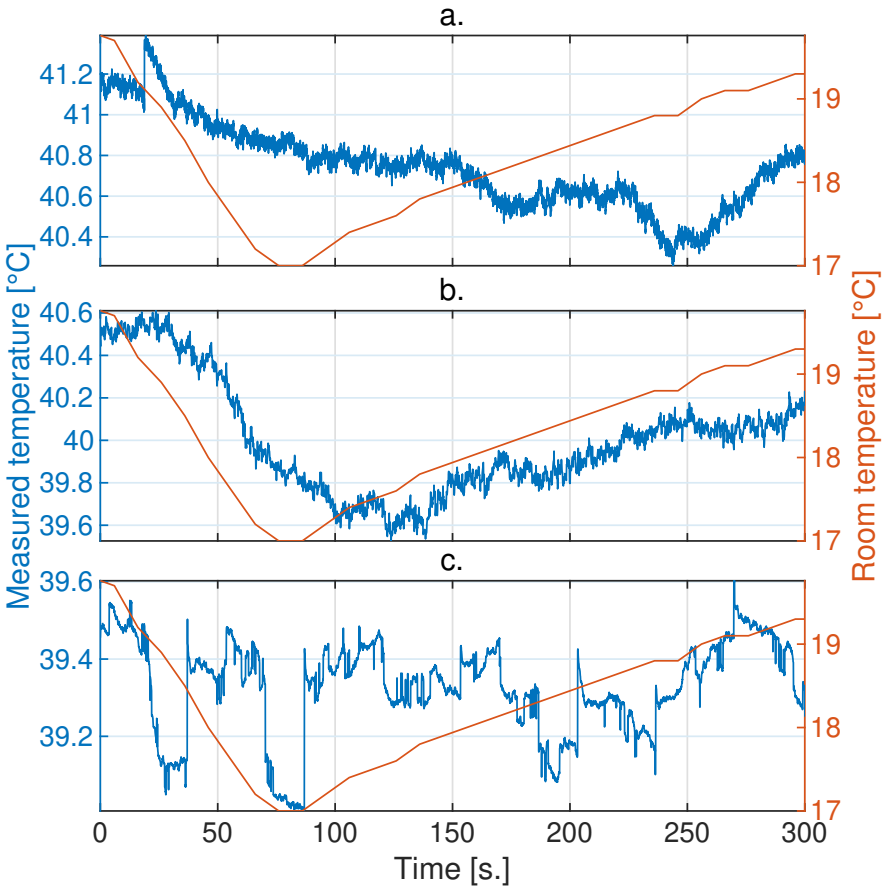

Figure 1.3: Samples of behavior of measurements from different IR cameras on a blackbody stabilized at $40°C$ in a room with controlled ambient temperature at $20°C$ using an AC unit. The order of measurements taken with the different cameras is as follows: a. Xenics Gobi-640, b. Opgal Therm-App and c. Seek Thermal CompactPRO.

The sixth contribution is the development of an IR video degradation model, which considers spatial and temporal noise, together with the temporal resolution of the camera to be modeled. Figure 1.4 shows an example of the degradation resulting from adapting an IR video captured with a high-quality IR camera (QmagiQ) to the characteristics of lower-quality cameras (Xenics Gobi-640, Opgal Therm-App, and Seek Thermal CompactPRO).

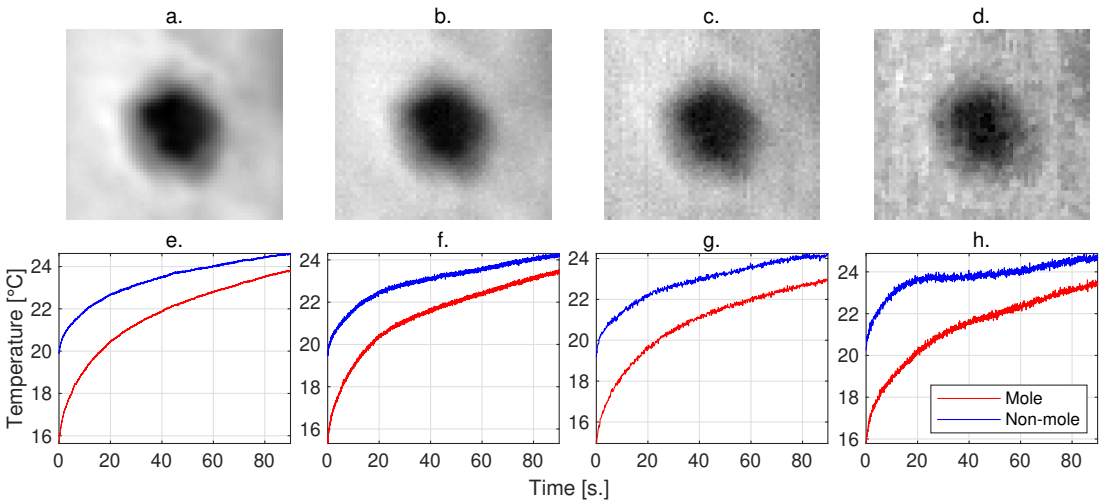

Figure 1.4: Sample result of the degradation of an IR video to the characteristics of different simulated cameras in a case of a malignant lesion, diagnosed as SCC. The top row shows an image from the video and the bottom row shows a characteristic TRC of the mole and non-mole areas. a. and e. Correspond to high-quality data captured with the QmagiQ camera; b. and f. data adapted to Xenics Gobi-640; c. and g. data adapted to Opgal Therm-App; d. and h. data adapted to Seek Thermal CompactPRO.

The seventh contribution is the evaluation of the feasibility of using low-cost IR cameras in skin cancer detection. Using the proposed IR video degradation algorithm, a dataset saturated with a high-quality camera was adapted to the characteristics of 3 lower-quality IR cameras, the performance of the proposed machine learning detection algorithm was evaluated and based on their performance indices, the feasibility of using each camera in skin cancer detection was analyzed. Table 1.2 presents the main detection algorithm's performance indices when working with each camera.

Table 1.2: Performance of the proposed detection algorithm using IR videos captured with the high-quality QmagiQ camera and its adaptations to features of the Xenics Gobi-640, Opgal Therm-App and Seek Thermal CompactPRO cameras.

Camera	**TPR (%)**	**TNR (%)**
QmagiQ	87.26	87.39
Xenics Gobi-640	83.03	85.28
Opgal Therm-App	83.23	84.94
Seek Thermal CompactPRO	79.77	83.74

1.4 Products of this dissertation

The following journal papers and conference presentations were obtained as direct results of this thesis, and they allowed the dissemination of this research work:

- **Soto, R. F.**, & Godoy, S. E. (2023, October). A novel feature extraction approach for skin cancer screening using active thermography. In 2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (pp. 1-6). IEEE.
- **Soto, R. F.**, & Godoy, S. E. (2024). Feasibility Study on the Use of Infrared Cameras for Skin Cancer Detection under a Proposed Data Degradation Model. Sensors, 24(16), 5152.
- **Soto, R. F.**, & Godoy, S. E. (2024). An automatic approach to detect skin cancer utilizing active infrared thermography. Heliyon, 10(23).

The following journal papers were obtained as indirect results of this thesis:

- Viafora, L. A., Torres, S. N., Machuca, G., Gutierrez, P., Jara, A., Coelho, P., & **Soto, R. F.** (2024). Infrared imaging technique for weightlifting exercise assessment. Applied Optics, 63(28), 7529-7539.
- Montes, C. S., **Soto, R. F.**, Sanhueza, M. I., Sanhueza, I., Luarte, D., Godoy, S. E., Torres, S. N., Castillo, R., Escribano, R., & Urbina, M. THE CHALLENGE OF BEING INVISIBLE IN OPEN SHALLOW WATERS: REFLECTION MECHANISMS IN THE SILVERY COATING OF CHILEAN PELAGIC FISH. Article under review in Scientic Reports.

1.5 Thesis Manuscript Structure

The thesis manuscript presented here is organized as follows.

In Chapter 2, the materials used in the development of the thesis are presented, including the corresponding datasets and hardware. Chapter 3 introduces an automatic detection algorithm, which improves upon a classical detection algorithm by automating the selection of the lesion area, proposing new feature extraction techniques, and utilizing machine learning classifiers. In Chapter 4, an IR video degradation model is proposed. Using this model, IR videos captured with a high-quality camera are adapted to the characteristics of three different IR cameras of varying ranges, and the feasibility of using each camera is evaluated by studying the performance of the proposed detection algorithm. In Chapter 5, a

classifier for TRCs using deep learning is proposed. Using this classifier, a metric is introduced to measure the malignancy of the IR video. This new feature is evaluated both individually and as part of the feature selection process described in Chapter 3.

Finally, Chapter 6 presents the conclusions and outlines the future steps required to enhance the outcomes of this research.

2 Materials

This chapter presents the materials used to develop this thesis, including the datasets, the data acquisition protocol, the IR video registration algorithm, and the hardware utilized.

2.1 Datasets

The development of this work considered the use of two datasets; first, the HAM10000 [58] set, which was used for training and validating the CNN. The HAM10000 set contains 10,015 dermoscopic images, of which 3,327 were selected as belonging to common skin cancer categories, namely lesions declared as MM, basal cell carcinoma (BCC) and benign. Second, a dataset called the UdeC set was used for testing. This dataset is private, it was acquired by one of the authors in his doctoral research with the support of the University of New Mexico Dermatology Clinic staff. The data was captured with informed consent, which does not allow its public release. As explained below, the UdeC set contains lower-quality, poor illuminated imagery acquired in clinical trials our team conducted. The UdeC set contains 144 image cubes of size $M \times N \times K$, where K is the number of images contained in the cube, M the number of rows, and N the number of columns of each picture. $K-1$ of K images are IR images corresponding to the thermal recovery of the skin after cold excitation. The remaining image corresponds to a visible image, which, unlike the HAM10000 set, was captured with a smartphone. The spatial dimension of the data is variable because the size of the lesions is variable. A lesion typically involves an area of 5x5 cms., which correspond to images of 100x100 pixels, i.e., $M = N = 100$. The IR data was captured for 1.5 minutes at a rate of 60 frames per second, so each cube contains $K - 1 = 5400$ IR images. This set is composed of MM, BCC, SCC and benign lesions. The visible images of the set were used to evaluate the performance of

Table 2.1: Distribution of lesions captures from the HAM10000 and UdeC datasets according to type of lesion: Malignant Melanoma (MM), Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC) and benign (BN).

Dataset	MM	BCC	SCC	BN	Total
HAM10000	1113	514	0	1700	**3327**
UdeC	7	41	9	87	**144**

the segmentation algorithm, contrasting the segmentation with the manually generated masks. Table 2.1 shows the distribution of the datasets by type of lesion.

It is important to note that for all images of both sets there is a mask that defines the lesion area. For the HAM10000 set the mask was faithfully defined as the pigmented area of the skin, i.e., the area whose color stands out from the healthy skin. For the UdeC set, the lesion area may coincide with the pigmented area, however, the expert who made the masks considered the nature of the lesions, so that sometimes he defined that the suspicious parts of the lesion encompasses more or less area than solely the pigmented area.

2.1.1 Data acquisition process

The data acquisition was carried out using active thermography in a controlled environment, with room temperature in range of 20° C to 22° C. The acquisition procedure consists of selecting a region of interest (ROI) with a plastic marker, as shown in the Figure 2.1. Then, take a visible image and a 15 seconds IR sequence of the ROI, which is used as a reference. Subsequently, using an air conditioning unit, the skin registered in the ROI is cooled for 30 seconds. After the cooling stage, the skin is heated by thermoregulation to room temperature. During the cooling and thermal recovery stages, a 1.5 minutes IR sequence is captured.

2.1.2 Image registration algorithm

For the correct analysis of the acquired images, it is necessary to apply an image registration algorithm, which is responsible for correcting involuntary movements of the patient, aligning the data sequence. Along with this, it is useful to differentiate the areas of lesion (pigmented area) and healthy skin on IR images.

The registration algorithm used corresponds to the one presented in Díaz et al. [59], which consists of the following stages:

1. Manually select the corners of the plastic marker in the visible image and in the first IR image. In the case of subsequent images of the IR sequence, the corners are detected automatically using as reference the selected in the previous image.

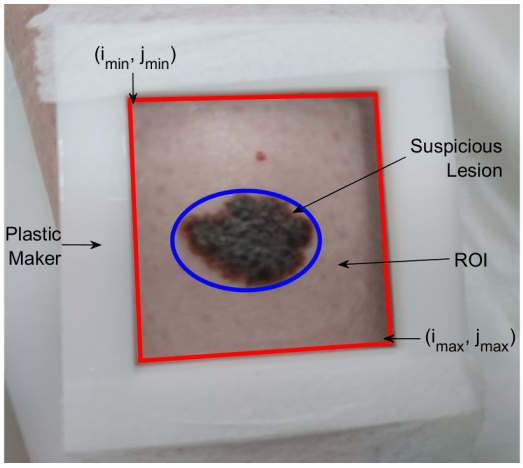

Figure 2.1: Example of a plastic marker with which the region of interest is selected. The region of interest is indicated in red and the suspicious lesion in blue.

2. Estimate an affine transformation matrix that maps motion between consecutive images (one matrix is estimated for each pair of images).

3. Apply the inverse transformation to each image to align the image sequence relative to the first IR image.

Once the image registration process is finished, a 3D array $\mathbf{u} \in \mathbb{R}^{I \times J \times K}$ is obtained, where I and J are the amounts of horizontal and vertical pixels, respectively, within the ROI. K is the number of images that compose the array (1 visible and K-1 IR images). In what follows we utilize the term $\boldsymbol{u_k}$ to name the k-th frame within the video sequence, i.e., $\boldsymbol{u_k} = u(\cdot, \cdot, k) \in \mathbb{R}^{I \times J}$. Additionally, we refer to the (i, j)-th pixel within the k-th frame as $u_{i,j,k}$.

This registration algorithm provides a maximum image shift of 3 pixels. Because of this, once the data cube is registered, a 3-pixel border is removed to avoid processing TRCs with temperature measurements of the plastic marker. A set of nearby TRCs are highly correlated, so in general this does not affect the data processing, thanks to the TRC selection stage of the detection algorithm explained in Section 3.2. It is possible that with poorer quality images, the image motion may be greater. However, this can be corrected with digital image processing techniques such as filter smoothing and contrast adjustment.

2.1.3 Hardware

This study involves data captured by a high-quality IR camera, brand QmagiQ [60], whose technology is quantum dots in a well (DWELL). These data were

Table 2.2: Technical features of the involved cameras.

Characteristic \ Camera	QmagiQ	Xenics Gobi-640	Opgal Therm-App	Seek Thermal CompactPRO
Focal distance [mm]	50	25	6.8	12.5
Spectral range [μm]	8-14	8-14	7.5-14	7.5-14
FOV [°]	5.5 × 4.4	25.84 × 19.38	55.56 × 41.67	33.40 × 24.81
Temporal resolution [fps]	60	50	8.7	15
FPA size [pixels]	320×256	640×480	384×288	320×240
NETD [mK]	20	50	70	70

Figure 2.2: Materiales utilizados en el desarrollo de la investigación. a. Xenics Gobi-640 camera, b. Opgal Therm-App camera, c. Seek Thermal CompactPro camera, d. modified portable air conditioning unit, e. Mikron M345X12 black-body radiator and f. temperature and humidity datalogger Elitech RC-4HC.

adapted in terms of thermal, spatial and temporal resolution to the characteristics of the microbolometer cameras Xenics Gobi-640 [61], Opgal Therm-App [62], and Seek Thermal CompactPRO [63]. Table 2.2 summarizes the technical features of these cameras. All cameras have a similar spectral range; the camera QmagiQ presents the best features in terms of field of view (FOV), temporal resolution and NETD, followed by the camera Xenics Gobi-640, Seek Thermal CompactPRO and Opgal Therm-App. Focal plane array (FPA) size is a unique feature where the camera QmagiQ is inferior to the others.

Figure 2.2 shows an image of the equipment used for the development of the research.

3

An automatic approach to detect skin cancer utilizing active infrared thermography

3.1 Introduction

The skin is a large organ, covering approximately 16% of the body mass. It is organized into two primary layers, epidermis and dermis. The epidermis is considered the body's armor, it is the peripheral layer of skin that contacts with the environment, working as a physiochemical barrier against environmental stressors such as pathogens, chemicals and ultraviolet (UV) radiation. The dermis, originated from the mesoderm, underlies the epidermis and anchorages cutaneous structures such as hair follicles, nerves, sebaceous glands and sweat glands. The dermis also contains abundant immune cells and fibroblasts, which actively participate in many physiological responses in the skin [64].

Thermal regulation is a relevant skin function to maintain the core temperatures within a small range, between 36 and 38°C. Vasodilation and vasoconstriction are among the main factors that control thermal regulation. Vasodilation and vasoconstriction refer to changes in blood vessel diameter, which affect skin temperature by changing the rate of blood exchange with the interior. If the ambient temperature rises, then increased conductance below the skin surface (due to increased blood flow) facilitates heat transfer from body interior to the skin. In the cold, muscle tensing and shivering increase heat production and body temperature. As blood flow decreases, the thermal conductance also decreases, and such decreasing heat loss from the body to the environment [65].

Skin, as any other organ, may be affected by solid malignant tumors. Solid malignant tumors, require a blood supply in order to grow larger than a few millimeters in diameter. Tumors induce the growth of new capillary blood vessels, a process called angiogenesis, by producing specific angiogenesis-promoting growth factors [66].

With the presence of these new blood vessels and the increase of blood supply, the thermal response of an area with a tumor changes with respect to the thermal response without a tumor. By thermal response, we mean the recovery of the skin temperature when a stimulus is applied. Under this fundament, many studies have highlighted the potential of analyzing thermal recovery curves obtained from dynamic thermography in order to distinguish malignant from benign lesions [11, 12, 8, 9, 67].

3.1.1 Skin cancer detection using active thermography

IRT is a fast, non-contact and non-invasive technique to analyze the temperature of an area of interest. It has been used for the detection of arthritis, allergies, breast cancer, burns, among others. It allows a high resolution view of the temperature of the lesion, from the epidermis to the depth of the tissue, in addition to measuring the variation in body temperature [35, 68].

There are two types of IRT: passive and active. The passive IRT is used to evaluate the temperature of an object in steady state, while the active IRT is used to evaluate the transient state after a thermal stimulus [69]. Active IRT is considered more powerful than passive IRT, because it allows to obtain quantitative information related to the thermal properties of the sample [67]. When active IRT is used, one can plot the temperature measured by each pixel with respect to time obtaining one TRC per pixel.

Among the dermatological applications where the use of active IRT stands out is the detection of skin cancer. Buzug et al. [8] have provided evidence that active thermography can differentiate between malignant and benign skin areas by observing distinct behavior in the thermal recovery of the skin, achieving results consistent with biopsies. However, the data lacked representativeness due to the limited number of infrared images. Çetingül and C. Herman [9] proposed a methodology to quantify the differences between the thermal responses of healthy skin and melanoma. Di Carlo et al. [10] demonstrated that, unlike dermoscopy, active thermography allows for the identification of a clear pattern to differentiate BCC from actinic keratosis (AK). In Godoy et al. (2015) [11], a standardized analysis of dynamic thermography was proposed and tested on more than 100 patients with benign lesions, BCC, SCC, or MM. Their algorithm differentiates between malignant and benign lesions, achieving a sensitivity of 95% and a specificity of 83%.

Godoy et al. (2017) [12] provided a more comprehensive analysis of TRCs, including modeling these curves. Combined with a proposed detection theory framework, this achieved a sensitivity and specificity of 99%. Bu et al. [70] explored the use of thermal excitation, modeling the lesion in 3D from the captured data. Simulations revealed a dependency between tumor thickness and the maxi-

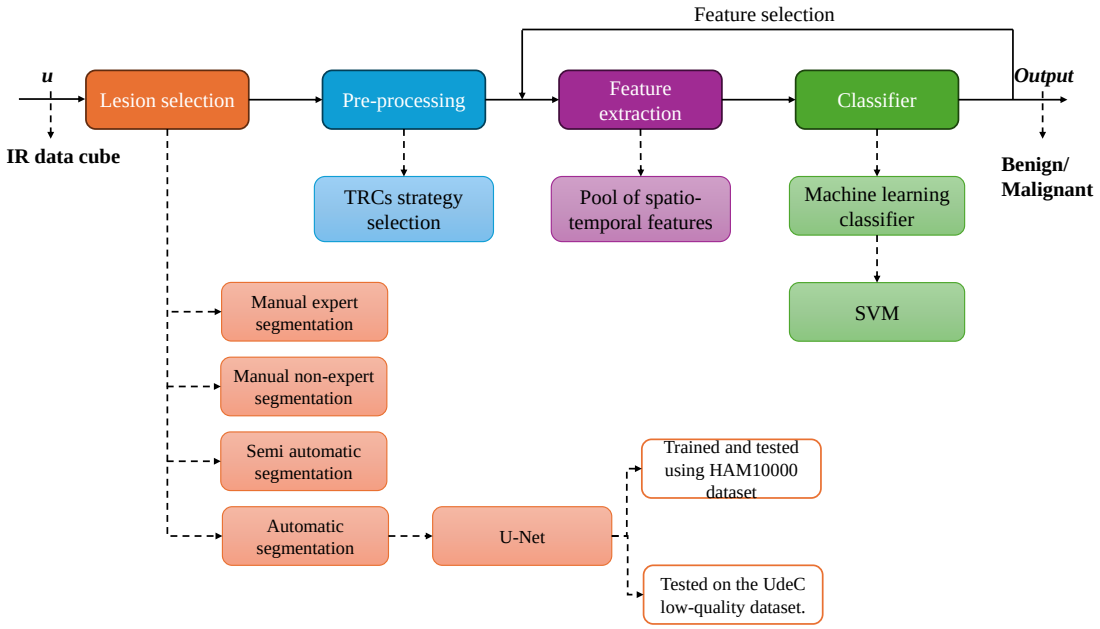

Figure 3.1: Diagram of the methodology used in the study.

mum contrast parameter, making it a discriminative factor in tumor classification.

Stringasci et al. (2018) [71] conducted a binary classification of clinically similar tumors using passive IRT, identifying relevant variables for differentiation. Magalhaes et al. (2021) [14] proposed a deep learning framework applied to passive IRT, using 40 temperature measurements from a 40×40 IR image, corresponding to the image diagonal. Several two-class classifiers were developed to differentiate lesion types (e.g., SCC vs. AK, MM vs. nevi, MM vs. non-melanoma, and benign vs. malignant). The classifier for MM vs. nevi achieved notable results with 96.91% accuracy, 94.12% sensitivity, and 98.41% specificity. In contrast, the benign vs. malignant classifier showed lower sensitivity (62.81%) and specificity (57.58%). Given these lower results compared to Godoy et al. (2015) [11], the authors emphasize the importance of using active thermography for this purpose.

3.1.2 Discussion

Skin cancer detection systems that process active IRT present two problems in the area of study: the systems are not automatic, so the results vary depending on who operates them, and there is a lack of exploration of machine-learning techniques. In this chapter, an automatic system for skin cancer detection is proposed, which is a base scheme for exploring machine learning techniques. A diagram of the methodology used is presented in Fig. 3.1, which is composed of four blocks: Lesion selection, Pre-processing, Feature extraction, and Classifier.

Lesion selection addresses the automatic detection of the lesion area using a U-Net segmentation CNN, which was trained and tested with the high-quality dermoscopic dataset HAM10000 and was then tested with the low-quality visible image set UdeC, which is the target dataset. The segmentation quality was measured with respect to expert manual segmentation, non-expert manual segmentation, and the semi-automatic segmentation method proposed by Diaz et al. [59]. The next block corresponds to Pre-processing, which corresponds to the selection considerations of features to be processed, based on the work proposed by Godoy et al. [11]. Feature extraction corresponds to the computation of the parameters that are then processed by the classifier. In this study, a large set of features is proposed, which are described in Sec. 3.3. Finally, there is a classifier in which any machine-learning technique can be used. In this study, an SVM classifier was used because in a previous study [72], the classifier that reported the best performance. In this scheme, a feedback loop was observed because, in this study, we searched for the best combination of up to five features.

3.2 Proposed thermography-based detection scheme and its foundations

It has been shown that in the detection of skin cancer using active thermography, the comparison between a descriptive TRC from a healthy skin area with the descriptive TRC contained in the lesion area is key to obtaining good classification results. The algorithm proposed by Godoy et al. [11] provides a solid foundation for feature extraction and subsequent classification, with the key stage being the selection of TRCs to obtain features that facilitate discrimination between benign and malignant lesions. For this reason, it is proposed to use the pre-processing stage of this algorithm as a basis for detection (Sec. 3.2.1), incorporating new techniques for feature extraction and classification described in Sec. 3.2.2.

3.2.1 Euclidean distance detection algorithm

The Euclidean distance detection algorithm (ED) proposed by Godoy et al. [11] comprises the following five stages, as shown in Fig. 3.2.

1. **Lesion selection.** A mask was generated over the visible image, which defined the lesion area. The original algorithm proposes creating a mask that manually encloses the pigmented area to generate the set L which contains the list of pixel locations of the lesion area (mask). With the remaining pixel locations of the visible image, another set, named N is created. It contains a list of pixel locations that are not within the lesion, but the surrounding skin.

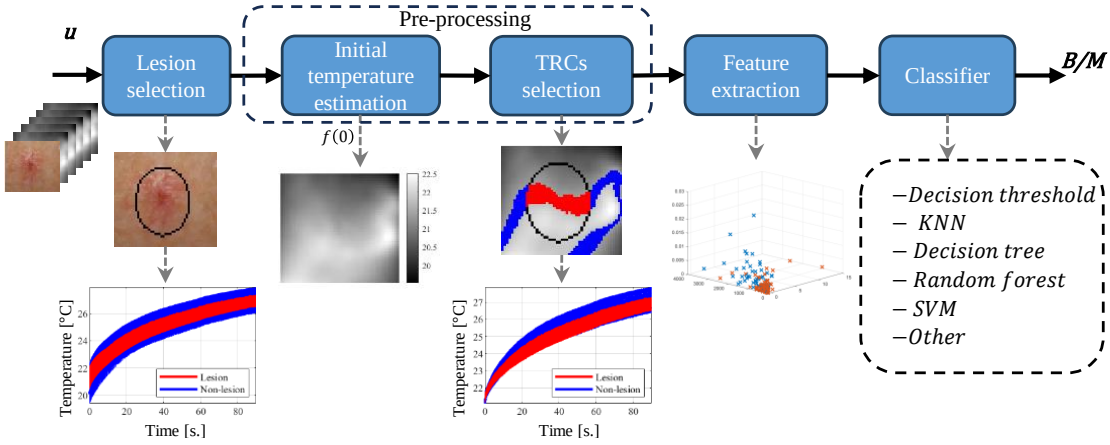

Figure 3.2: Base detection scheme for skin cancer using active thermography and machine learning.

2. **Initial temperature estimation.** The next stage corresponds to the selection of the TRCs, which depends on the initial temperature of each TRC. To select curves whose initial temperature is less affected by non-uniformities in the cooling process, the TRC of the (i, j)-th pixel is modeled as a double exponential function by

$$f_{i,j}(t) = \theta_{i,j}^{(1)} + \theta_{i,j}^{(2)} \exp\left[\theta_{i,j}^{(3)} t\right] + \theta_{i,j}^{(4)} \exp\left[\theta_{i,j}^{(5)} t\right], \tag{3.1}$$

whose parameters $\boldsymbol{\theta_{i,j}} = \begin{bmatrix} \theta_{i,j}^{(1)} & \cdots & \theta_{i,j}^{(5)} \end{bmatrix}$ are computed using a non-linear least-squares fitting approach. With this model, the initial temperature is estimated by setting $t = 0$, i.e., $f_{(i,j)}(0) = \theta_{i,j}^{(1)} + \theta_{i,j}^{(2)} + \theta_{i,j}^{(4)}$.

3. **TRCs selection.** The reference temperature is calculated as $T_{ref} = E\left[\boldsymbol{f}_L(0)\right]$, where E is the expectation operator, and $\boldsymbol{f}_L(0)$ is the vector function of the initial temperature of the TRCs of the modeled lesion area within set L. Simply put, pixels within set L are selected, and the average initial temperature is calculated to determine T_{ref}.

   The selection of points for processing by the algorithm considers a margin of error of $p \cdot 100\%$ with respect to T_{ref}. Thus, the set of points of interest for the algorithm is defined as

$$S = \{(i, j) : |f_{i,j}(0) - T_{ref}| \leq p \cdot T_{ref}\} \ . \tag{3.2}$$

   Experimentally $p = 0.01$ was defined as a good value. Using the set S of all the locations within a $p\%$ of T_{ref}, the sets of TRCs with similar initial temperature from the lesion and non-lesion areas are defined as $L^* = L \cap S$ and $N^* = N \cap S$, respectively.

4. **Feature extraction.** From the representative TRCs, a combination of features was extracted (feature vector). The detection algorithm proposed by Godoy et al. [11] uses only Euclidean distance as a feature to feed the next step. The formula used to compute this feature is given in Eq. 3.3.

5. **Classifier.** In the final step, the feature vector is processed by a classifier that defines whether the lesion is suspicious. The original detection algorithm used a simple threshold to discriminate between malignant and benign lesions. This threshold was defined by the training set.

3.2.2 Proposed detection algorithm

The proposed detection algorithm considers the following 3 modifications to the algorithm described above:

1. **Lesion selection.** The main drawback of the base algorithm is that when a non-expert selects the lesion the performance drops from 86% to 75%, as shown in Sec. 3.5.2. As such, the first proposed modification is to use the CNN U-Net described in Sec. 3.4 to automatically select the lesion by the segmentation of the pigmented area.

2. **Feature extraction.** As already mentioned, the base algorithm uses only one feature that accounts for the difference between the thermal responses. In Sec. 3.3, 32 feature extraction techniques are described, of which 31 were proposed in this study. Using the complete detection scheme, the feature vector that provides the highest accuracy was selected, considering a maximum of five features out of the 32 available.

3. **Classifier.** Based on prior work, the SVM classifier with a radial basis function (RBF) kernel provides the best performance compared to techniques such as XGBoost and Random Forest for skin cancer detection using active thermography [72]. SVM works by finding a hyperplane that best separates the data points of different classes in a way that maximizes the margin, i.e., the distance between the hyperplane and the closest data points from both classes. The closest data points are called support vectors. The hyperplane's position and orientation are determined by the data distribution, and the SVM aims to find the hyperplane that minimizes the classification error and generalizes well to new, unseen data [73]. SVM classifiers are known for their ability to handle complex decision boundaries, their resistance to overfitting, and their effectiveness in high-dimensional spaces. Owing to the property of projecting the features onto a kernel such as RBF, it is possible to map the characteristics into a higher-dimensional space where the data might

become separable or enhance their separability. As such, whereas the base algorithm only uses thresholding of one feature, here the best combination of up to five features selected by an SVM classifier was used.

3.3 Descriptive feature extracted from TRCs

Initially, Euclidean distance was a good discriminating parameter between benign and malignant cases of TRCs, as demonstrated by Godoy et al. [11]. However, when considering new data acquired after the publication of the article, this performance decreased from an accuracy of 89% to 60%. Therefore, new features are proposed in this study to improve the differentiation between malignant and benign cases in a more general manner.

From the representative TRC of the lesion ($\overline{TRC_{L^*}}$) area and non-lesion area ($\overline{TRC_{N^*}}$), the following features were extracted:

1. **Euclidean distance (d).** This feature is calculated as the norm of the difference $\overline{TRC_{L^*}} - \overline{TRC_{N^*}}$, normalized by the number of points, i.e., as shown in the following Eq.:

$$d = \frac{1}{K-1}\|\overline{TRC_{L^*}} - \overline{TRC_{N^*}}\|. \tag{3.3}$$

   The idea behind this feature is that if the Euclidean distance is small, the curves are similar, and it is very likely that the lesion is benign (*i.e.*, the thermal recovery of the lesion behaves like normal tissue). Conversely, if the Euclidean distance is large, the curves show differences in thermal recovery, and the lesion is likely to be malignant because the lesion thermal recovery does not behave like a normal tissue [11].

2. **Energy difference (E_d).** Let $\overline{\overline{TRC_{X^*}}} = \overline{TRC_{X^*}} - min(\overline{TRC_{X^*}})$, *i.e.*, the unbiased TRC of the X^* area. The energy difference E_d is calculated as follows:

$$E_d = \left| \|\overline{\overline{TRC_{L^*}}}\|^2 - \|\overline{\overline{TRC_{N^*}}}\|^2 \right|. \tag{3.4}$$

   This feature is closely related to d. However, because of the triangular inequality $E_d \leq d$, thus E_d quantifies smaller differences than d.

3. **Rise time difference (RT_d).** To calculate this characteristic, unbiased characteristic TRCs of each area were used. Over $\overline{\overline{TRC_{L^*}}}$ and $\overline{\overline{TRC_{N^*}}}$ the time at which 63.2% of the maximum amplitude is reached is calculated,

which is referred to as $T63_L$ and $T63_N$, respectively. The rise time difference was then calculated as $RT_d = |T63_L - T63_N|$.

4. **Area difference (A_d).** Let $\overline{\overline{TRC_{L^*}}}$ and $\overline{\overline{TRC_{N^*}}}$ be the unbiased characteristic TRCs of lesion and non-lesion areas, respectively. The normalized area of a $\overline{\overline{TRC_{X^*}}}$ is calculated as:

$$A_X = \frac{\sum_{k=1}^{K-1} \overline{\overline{TRC_{X^*}}}(k)}{(K-1) \cdot \max(\overline{\overline{TRC_{X^*}}})}. \tag{3.5}$$

Then, A_d was calculated as $A_d = |A_L - A_N|$.

5. **Statistical similitude features.** Here, six base features were defined to measure the similarity of a set of TRCs with a normalized modeled TRC using the dual exponential model shown in Eq. 3.1. It is assumed here that the modeled TRC has K sample points, and thus, $f_m \in \mathbb{R}^K$. With this, its inner product is computed by $< f, g >= f^T g$, where $f, g \in \mathbb{R}^K$ and ^{T} is the column-vector transpose. f_m was obtained by computing the five model parameters $\boldsymbol{\theta}_{i,j}$ using a non-linear least-squares fitting approach. These parameters were averaged to obtain the *descriptive TRC* f_M for each class (cancerous and non-cancerous TRCs). Then f_m is forced to have a unit norm, *i.e.*, $\|f_m\| \triangleq \sqrt{< f_m, f_m >} = 1$.

Let f_m and f_n be a model TRC and an arbitrary TRC, respectively, both modeled and normalized. (Here, it was assumed that $\|f_m\| = 1$.) The characteristics of projection, correlation, and Euclidean distance are described below.

   a. **Projection.** The projection of f_n onto f_m is calculated as $proj_n =< f_n, f_m >$. This operation is performed for a set of at least 10 TRCs, and the mean projection, $\overline{proj}$, and the standard deviation of the projection, σ_{proj}, are then calculated over this set of projections.

   b. **Correlation.** The correlation between f_n and f_m was calculated using the following expression:

$$\rho = \frac{cov(f_m, f_n)}{\sigma_{f_m} \sigma_{f_n}}, \tag{3.6}$$

   where $cov(f_m, f_n)$ is the covariance between f_m and f_n, and σ_{f_m} and σ_{f_n} are the standard deviations (in the time samples) of f_m and f_n, respectively. This operation was performed for a set of at least 10 TRCs,

and then the $\overline{\rho}$ (the mean correlation) and σ_ρ (the standard deviation of the correlation) were calculated.

c. **Distance.** This parameter was calculated according to Eq. 3.3 but using f_m as a reference and a family of at least 10 TRCs, f_n. The mean value $\overline{d}$ and standard deviation σ_d were calculated for this set of distance values.

Considering a model of malignant and benign TRC, f_M and f_B, respectively, for each data cube, the projection characteristics $\overline{proj}$ and σ_{proj}, correlation $\overline{\rho}$ and σ_ρ, and distance $\overline{d}$ and σ_d were calculated for the sets of TRCs L^* and N^*. In this manner, 24 features were extracted and grouped into four categories based on the origin of the model curve and the analysis group. For example, the malignant model f_M over a non-lesion area N^* allows the extraction of $\overline{proj_{M-N}}$, $\sigma_{proj_{M-N}}$, $\overline{\rho_{M-N}}$, $\rho_{proj_{M-N}}$, $\overline{d_{M-N}}$ and $\sigma_{d_{M-N}}$ features.

6. **Difference in temperature at 20 seconds (ΔT_{20}).** Let $\overline{TRC_{L^*}}$ and $\overline{TRC_{N^*}}$ be the characteristic TRCs of lesion and non-lesion areas, respectively. Each TRC is normalized and modeled as a simple exponential function $g(t) = 1 - e^{\theta t}$, which is adjusted by the least-squares method. The benign, $g_b(t)$, and malignant, $g_m(t)$, models are evaluated at $t = 20$ to obtain the temperature value for each case. The difference in temperature is defined by

$$\Delta T_{20} = |g_b(20) - g_m(20)| . \tag{3.7}$$

Experimentally, $t = 20$ was chosen because after this time, the slope of change between malignant and benign TRCs starts to look alike. Thus, the difference in the time evolution of the thermal recovery was more easily appreciated at the chosen time.

7. **Principal Component Analysis (PCA) of the difference of the exponential model parameters.** By using the double-exponential model shown in Eq. 3.1, the TRCs representative of the lesion and non-lesion areas were modeled to obtain sets of parameters $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_N$, respectively. The difference between these parameters is calculated as $\Delta\theta = |\theta_L - \theta_N|$, generating a vector of five components.

Considering the training set, the dimension reduction technique PCA was applied to the five difference parameters $\Delta\theta$. The covariance matrix was analyzed to reduce the dimensions of these five characteristics to three principal components: $PCA_{\Delta\theta 1}$, $PCA_{\Delta\theta 2}$ and $PCA_{\Delta\theta 3}$.

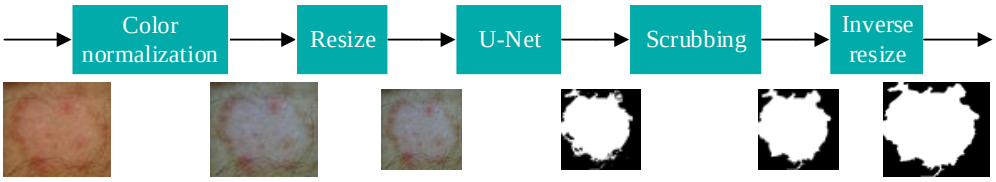

Figure 3.3: Proposed segmentation scheme.

3.4 Skin lesion segmentation algorithm

As mentioned previously, the main drawback of the ED algorithm [11] is that it requires good manual selection of the lesion to perform well. The algorithm breaks down by wrongfully selecting the lesion tissue. As such, this work focuses on automatically identifying the lesions with respect to the surrounding skin to eliminate this weakness of the ED algorithm.

This is not a trivial problem to address because the segmentation algorithm to do so should work with low resolution visible-range images acquired in a clinical environment. Namely, the patient may move, and the angle between the lesion and the camera may not be the best. Moreover, the main problem with the proposed application is that the images are not acquired with controlled light, as they are commonly imaged in the literature.

In recent years, several algorithms have been reported in which a U-Net convolutional network or its variations are used to perform this task [74, 75, 30]. U-Net was specially designed to solve biomedical image-segmentation problems [76]. The network merges a convolutional network architecture with a deconvolutional architecture to generate a semantic segmentation. As described by Ashraf et al. [30]: *It is a fully connected network that adds network layers to a traditional contracting network, replacing pooling operators with up-sampling operators to improve the resolution of the predicted image. A subsequent convolution layer assembles a more specific output for localization and backward propagation by concatenating high resolution features from up-sampled output and contraction path. The network can spread context information to higher resolution layers with up-sampling components, resulting in a U-shaped architecture.* The architecture utilized in this study was a U-Net model with modifications proposed by Bisla et al. [1]. The main difference from the original U-Net is the consideration of up-sample stages and the use of dilatation kernels of size 2×2 and 4×4. This allowed us to obtain an output with the same spatial resolution, without being clipped. To reiterate, these algorithms work well, but are trained and tested over high-quality imagery, which may not always be available in a clinical trial. Here, the U-Net is pushed to the limit so that it can operate under these extreme circumstances.

Figure 3.3 shows the proposed segmentation scheme, which comprises the fol-

lowing five stages:

1. **Color normalization.** Standardize the color conditions of images, which can be affected by illumination conditions and the use of different setups to acquire images. This study considers the use of the algorithm Shades of Gray [77], which has been shown to be effective in obtaining greater performance in the analysis of dermoscopic images [78].

2. **Resize.** The size of the image was adjusted to dimensions 380×380, to meet the input conditions considered by the implemented U-Net network. As previously mentioned, images may have been acquired with different setups; therefore, image size standardization is necessary.

3. **U-Net.** Segmentation using a variant of the U-Net network proposed by Bisla et al. [1]. This network requires as input an RGB image concatenated with the transformation of the same image to HSV and L*a*b format, generating a cube of size $380 \times 380 \times 7$. This size was considered in the design of the network because it is a representative size of the images of the UdeC set, which is the objective set to be segmented. This model has nine blocks and 21 convolutional layers, as illustrated in Fig. 3.4.

4. **Scrubbing.** Morphological dilation and erosion filters were applied to smooth the segmented area and fill the holes. An analysis by connected components was also carried out; in the event that there was more than one segmentation, the segment with the largest area was selected.

5. **Inverse resize.** The output mask is adjusted to the original dimensions of the input image.

The segmentation algorithm was trained using the HAM10000 dataset [58], which was split by allocating 70% to train and to validate the algorithm, and the remaining 30% was used for testing. This proportion was selected for being a typical division ratio used in machine learning, which according to Nguyen et al. [79], reports better performance. As a final test, the algorithm was evaluated over the visible images of the UdeC dataset.

The training parameters of the U-Net network were summarized in Table 3.1. To obtain an algorithm that is robust to variations in lighting conditions and the use of different cameras, the following data augmentation techniques were used in the training stage: vertical and horizontal flip with a 50% probability of being flipped, adjustments of ±50% shuffled brightness, ±20% shuffled saturation, and ±50% shuffled contrast, and blur was added using a Gaussian kernel filter of size 5×5 and a variable standard deviation randomly selected between 0.1 and 20.

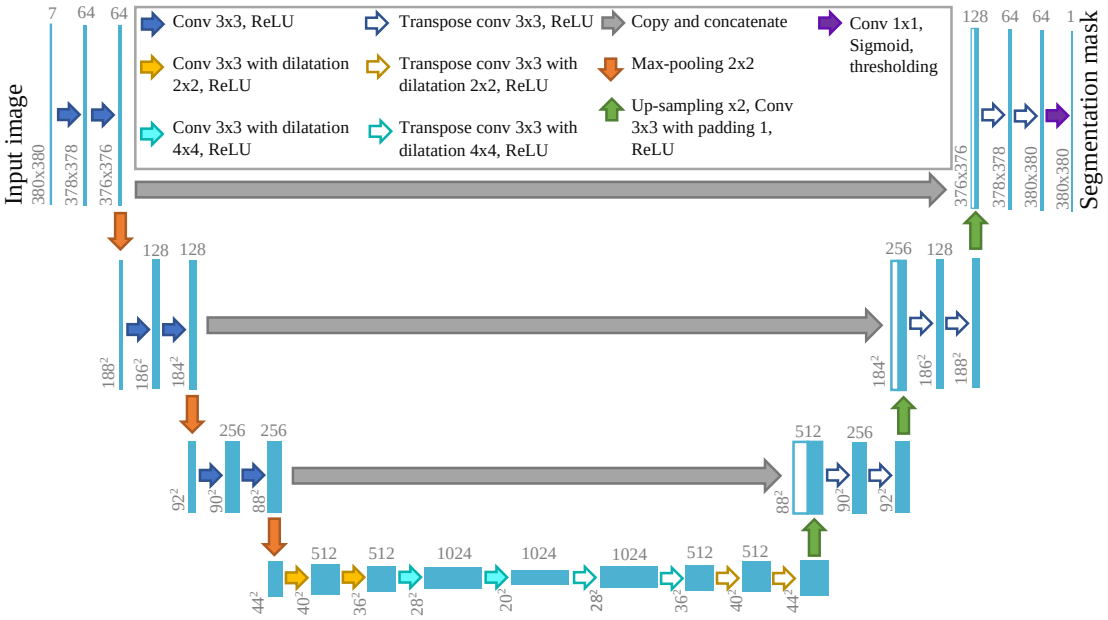

Figure 3.4: Scheme of U-Net architecture proposed by Bisla et al. [1].

In the training process, different learning rates were evaluated, whose value 10^{-r} varied by modifying r between 3 and 5 in steps of 1. The best performance was achieved by training the network with a learning rate of 0.0001, Adam's optimizing function [80], and the BCE loss function with sigmoid [81], adjusting the network for 33 epochs. The training and validation curves obtained over 100 epochs are shown in Fig. 3.5. Because 33 the result did not improve from epoch 33, it was decided not to continue with the training. It can be seen that after 33^{th} epoch, the loss in the training set was reduced. However, the validation loss increased, indicating that the network did not learn, thus overfitting the training set.

3.5 Results and discussion

3.5.1 Segmentation performance

As mentioned above, the segmentation algorithm was evaluated on two datasets: with the data partition of the HAM10000 set intended for testing and with the UdeC set. The well-known metrics of accuracy [82], sensitivity (also known as true positive rate (TPR)) [83], specificity (also known as true negative rate (TNR)) [83], precision (also known as positive predictive value (PPV)) [83] and Jaccard index [84] were used as performance measures.

Table 3.1: Detail of U-Net training parameters.

Parameter	**Value**
Learning rate	0.0001
Optimizer	Adam
Function loss	BCE
Data augmentation	Vertical and horizontal flip, brightness, saturation, contrast and blurring adjustments
Maximum training iterations	100
Best performace iteration	33
Training/validation samples (HAM10000 dataset)	2340
Testing samples 1 (HAM10000 dataset)	987
Testing samples 2 (UdeC dataset)	144

Table 3.2 shows the performance of the algorithm on the HAM10000 and UdeC sets, reporting the aforementioned metrics. A considerable reduction in performance was observed when evaluating the segmentation of images of the UdeC set, decreasing the Jaccard index by 0.19 points and the precision by 15%. This is because the images of the UdeC set were captured under different lighting conditions, with a low-cost visible camera that, unlike a dermatoscope, is not made to provide the highest quality of clinical conditions.

Moreover, the HAM10000 set contained segmentations faithfully based on discriminating the pigmented area from the background, while the masks of the UdeC set were created considering expert knowledge, adjusting the masks to the criteria of a dermatologist, who determined that part of the lesion was malignant or benign. Because the segmentation algorithm was trained with the HAM10000 set, its performance was expected to be lower when evaluating masks created with expert criteria. To measure the impact of the proposed segmentation algorithm, using the UdeC set, the segmentation performance was contrasted with manual masks (non-experts) adjusted solely to the pigmented areas and masks created with the semi-automatic segmentation algorithm proposed by Díaz et al. [59], hereinafter referred to as the Díaz algorithm.

Figure 3.6 shows the cross-performance indices resulting from different segmentation techniques. In contrast to expert segmentation, it is observed that the highest Jaccard index is obtained by non-expert segmentation, with a value of 0.67, followed by U-Net segmentation with 0.60, with the worst performance obtained with the Díaz algorithm. Similarly, the best accuracy performance was obtained by non-expert segmentation with 98.26%, followed by the U-Net and Díaz segmentation algorithms with 98.00% and 97.37%, respectively. Regarding

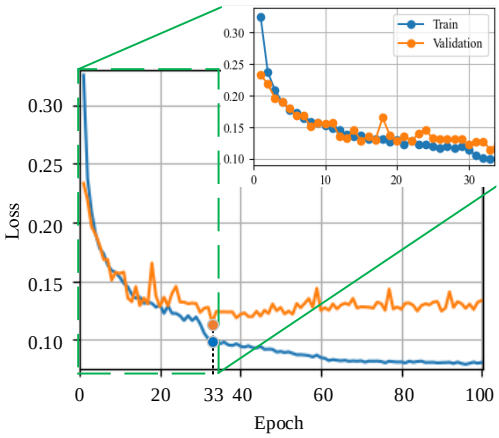

Figure 3.5: Training and validation curves of U-Net network.

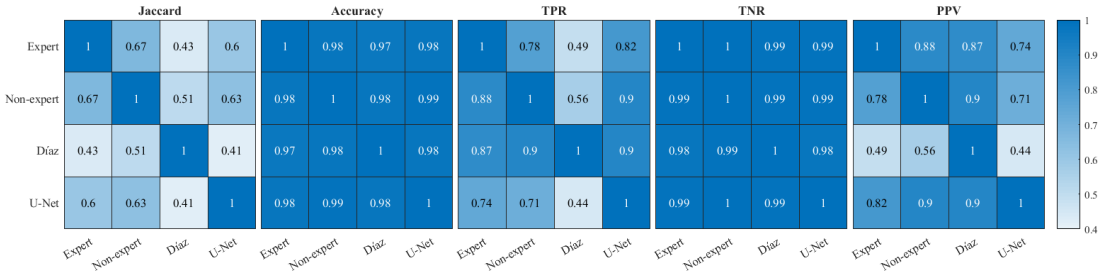

Figure 3.6: Comparison of segmentation performance indices using different benchmarks. The rows represents the benchmarks and the columns are the evaluated masks.

TPR, the best performance was obtained by U-Net segmentation with 82.02%, followed by inexperienced segmentation and Díaz segmentation with 77.71% and 48.58%, respectively. Whereas in specificity the worst result was obtained by the U-Net segmentation with 98.92%. This last result is consistent with the precision, in which the worst result was obtained by the U-Net segmentation. Thus, despite the fact that the U-Net segmentation presents a high specificity, the difference that exists with the other techniques generates a lower precision of approximately 15%, since most of the images present a much larger zone of no lesion than of injury, which means that generating a mask slightly larger than the reference produces a large drop in precision performance.

As expected, the indices indicated that the U-Net segmentation algorithm is similar to non-expert segmentation, with the advantage of being an approach with standardized criteria, unlike manual segmentation, which depends on the criteria of each user, whose decision can be affected by conditions such as stress and fatigue, among others.

Table 3.2: Performance of the segmentation algorithm in terms of accuracy, sensitivity (TPR), specificity (TNR), precision (PPV) and Jaccard index.

Dataset	**Acc.**	**TPR**	**TNR**	**PPV**	**Jaccard**
HAM10000	0.93	0.85	0.96	0.89	0.79
UdeC	0.98	0.82	0.99	0.74	0.60

In relation to the multiple image processing stages that include the proposed segmentation scheme, the best result is obtained considering all stages, with a Jaccard index value of 0.5963. By deleting the scrubbing step, the index is reduced to 0.5894, while removing the color normalization and scrubbing steps, the index is reduced to 0.5721. The segmentation results are reflected in the detection performance, such that the complete scheme provides an Area Under Curve (AUC) of 0.792, whereas deleting the scrubbing step reduces the AUC to 0.778, and removing the color normalization and the scrubbing steps reduces the AUC to 0.759.

Figure 3.7(a) shows an image of a lesion of the UdeC set, together with manual segmentation of the lesion using expert knowledge (Fig. 3.7(b)), manual non-expert segmentation (Fig. 3.7(c)), and the proposed automatic segmentation (Fig. 3.7(d)). It was observed that all the segmentation algorithms selected the lesion. It should be noted that the expert segmentation selects the inner area of the lesion, whereas the other segmentation output contains the lesion with a border of non-pigmented skin. However, the lesion area was small with respect to the total area of the image, which implies that small differences in the selection of the lesion negatively affect the TPR, PPV and Jaccard indices. This situation is characteristic in the UdeC set, where the lesion area covers an average of $3.34 \pm 7.29\%$ of the image. While with the HAM10000 set it addresses $27.82 \pm 19.41\%$. Despite the fact that to the human eye the segmentation results are quite similar, the performance index used is highly reduced if the lesion is composed by few pixels. Considering an extreme case where lesion is composed by 5 pixels, not detecting one implies reducing the Jaccard index from 1 to 0.8. It could be worse if the segmentation algorithm generates an over or under segmentation with a 1-pixel border. Therefore, the fewer the pixels the lesion has, the worse the index when the algorithm mis-segments one pixel. This problem will always be present when the suspicious lesion is imaged with low-quality sensors.

3.5.2 Detection performance

The detection algorithms described in Sec. 3.2 were evaluated using the four lesion segmentation techniques previously described: (i) Expert, (ii) Non-expert, (iii) Díaz, and (iv) U-Net. Both expert and non-expert segmentations are manually

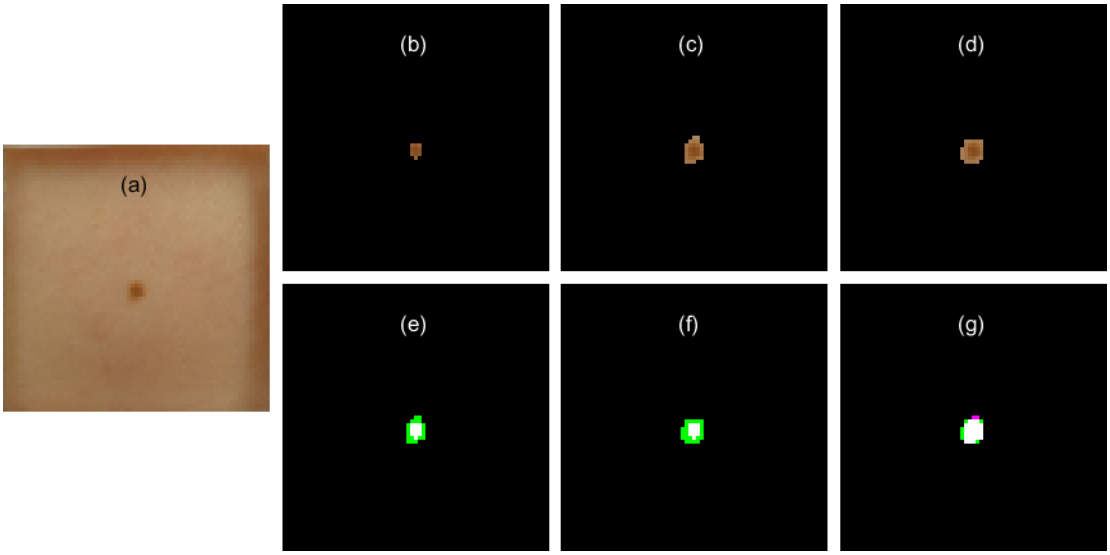

Figure 3.7: Segmentation contrast. (a) Image with lesion to be segmented, (b) segmentation with expert knowledge, (c) manual non-expert segmentation, (d) automatic segmentation using U-Net network, (e) contrast of expert segmentation with non-expert segmentation, (f) expert segmentation contrast with automatic segmentation, (g) non-expert segmentation contrast with automatic segmentation.

performed, whereas Díaz is semi-automatic in the sense that it requires interaction with the user. In contrast, the proposed approach (U-Net) is fully automatic in the sense that it does not require any input from the user.

The proposed detection algorithm considers a novel feature-extraction stage that feeds an SVM classifier. From the set of features described in Sec. 3.3, a combination of up to five characteristics that provided the greatest accuracy was sought. When working with expert segmentation, the best feature combinations were $[A_d, RT_d, \Delta T_{20}, PCA_{\Delta\theta 3}, \sigma_{\rho_{B-L}}]$. Whereas, with the other types of segmentations, the following feature vectors were selected: $[E_d, \sigma_{\rho_{B-L}}, \sigma_{proj_{B-N}}, \sigma_{d_{M-N}}]$.

The performance and variability of the ED algorithm and proposed detection algorithm based on the new features were evaluated using the bootstrap method. This technique entails drawing multiple samples with replacements from the original dataset, creating a new dataset of the same size. The modified dataset was then divided by allocating 80% of the data for training and 20% for testing. The entire procedure was repeated 2,000 times such that the mean results were presented, as well as the minimum and maximum results for each approach.

Table 3.3 lists the performance of the ED detection algorithm on the active thermography dataset. Considering a sensitivity of 85% as a reference point, the best result was achieved with expert segmentation, followed by the U-Net, non-

expert, and Díaz segmentation algorithms, with an average specificity of 42.72%, 32.39%, 28.90%, and 17.03%, respectively. Similarly, Table 3.4 shows the performance of the detection algorithm using an SVM detection algorithm; it is observed that the best result is obtained with the expert segmentation, followed by the non-expert, U-Net, and Díaz segmentation algorithms, with average specificities of 87.62%, 69.63%, 68.50%, and 64.45%, respectively.

Both detection algorithms exhibited their best performance using expert segmentation as input, followed by detection when non-expert and U-Net segmentation algorithms were used. The worst result was obtained using Díaz segmentation as input. Analyzing these results, it is evident that the selection of TRCs is crucial for the performance of skin cancer detection algorithms that process active thermography. These results indicate that the detection algorithm used, which is based on active thermography analysis, requires expert selection of the lesion area to obtain the best performance. Thus, expert knowledge is relevant for determining the key pigmented area to determine whether a lesion is benign or malignant.

By linking the performance of the detection algorithm using different lesion segmentation techniques with the performance in segmentation quality, it is observed that these are highly correlated. The proposed detection algorithm, when working with U-Net segmentation, exhibited superior performance compared to the Díaz segmentation algorithm, and similar performance to that of non-expert segmentation. Thus, it was demonstrated that accurate selection of the lesion area is relevant for successful detection. Thus, it is important to automate these types of algorithms so that they are not susceptible to the operator. Or, in contrast, utilize algorithms that do not use the time difference of the TRCs (or forms of it) as their discriminant factor.

Regarding the proposed modifications to feature extraction techniques and the detection algorithm, it is highlighted that the combination of the proposed features along with the SVM classification technique enables an increase in accuracy of 26.58%, 24.14%, 21.63%, and 28.16% when selecting the lesion area with expert, non-expert, U-Net, and Díaz segmentations, respectively. Thus, these modifications make the proposed detection algorithm less sensitive to segmentation differences than the ED algorithm. However, its performance suffers if the input segmentation is incorrect.

Table 3.3: Detection performance using Euclidean classifier and different lesion area selection techniques.

Lesion selection	**Expert**				**Non-expert**				**U-Net**				**Díaz**			
Index	**Min.**	**Max.**	**AVG**	**SD**	**Min.**	**Max.**	**AVG**	**SD**	**Min.**	**Max.**	**AVG**	**SD**	**Min.**	**Max.**	**AVG**	**SD**
Accuracy	31.03	86.21	60.03	8.91	20.69	82.76	51.63	9.01	20.69	82.76	53.66	9.14	10.34	79.31	44.51	9.15
TPR	33.33	100.00	86.15	10.16	30.00	100.00	85.98	10.35	37.501	100.00	85.73	10.63	42.86	100.00	86.10	10.39
TNR	5.26	80.00	42.72	11.90	0.00	73.33	28.90	10.68	0.00	70.00	32.39	11.02	0.00	57.14	17.03	9.13
PPV	6.67	84.00	49.91	10.99	8.70	84.00	44.46	10.34	8.70	84.00	45.63	10.59	7.41	83.33	40.73	9.86

Table 3.4: Detection performance using SVM classifier and different lesion area selection techniques.

Lesion selection	**Expert**				**Non-expert**				**U-Net**				**Díaz**			
Index	**Min.**	**Max.**	**AVG**	**SD**	**Min.**	**Max.**	**AVG**	**SD**	**Min.**	**Max.**	**AVG**	**SD**	**Min.**	**Max.**	**AVG**	**SD**
Accuracy	65.52	100.00	86.61	6.65	27.59	96.55	75.77	9.26	24.14	100.00	75.29	9.99	17.24	96.55	72.67	10.97
TPR	33.33	100.00	85.02	12.02	25.00	100.00	85.04	11.97	25.00	100.00	85.10	11.80	30.00	100.00	85.05	12.44
TNR	40.00	100.00	87.62	8.80	0.00	100.00	69.63	15.48	0.00	100.00	68.80	16.30	0.00	100.00	64.45	18.90
PPV	20.00	100.00	82.37	11.62	20.00	100.00	65.95	13.53	18.18	100.00	65.41	13.94	14.29	100.00	62.82	14.24

The presented results seem to be lacking in impact when compared to others that process dermatoscopic images, which achieved an accuracy higher than 95%, as shown in Table 3.5. However, they mainly addressed the classification of MM vs. non-MM or MM vs. benign. It is important to mention that the proposed approach tackles the problem of detecting benign lesions vs. malignant lesions, *i.e.*, all types of cancer are malignant cases for our approach. In this scenario, the detection of skin cancer is not trivial owing to the small variability between the two classes.

This was reflected in the work of Magalhaes et al. [14], who obtained 62.81% sensitivity and 57.58% specificity. Their study is one of the most recent approaches in the literature that uses thermography. The performance reported in this manuscript surpasses that of Magalhaes, achieving an 85% sensitivity and 69% specificity when working with automatic segmentation and an SVM classifier, respectively.

By implementing our own classifier based on a ResNet152V2 network, which was trained using the HAM10000 dataset in conjunction with the U-Net network, the ResNet152V2 achieved an accuracy of 90.63% on the HAM10000 test set and 72.49% on the UdeC visible image test set. This result demonstrates the great potential of transfer learning techniques. However, their performance declines by almost 20% when processing poorer quality data, obtained in a clinical environment with limited resources. This result is concordant with the study by Akilandasowmya et al. [34], which showed high variability in the performance of systems processing transfer learning-based dermoscopy images by varying the training set sizes.

3.6 Conclusions

This study presents two improvements to a previously published detection algorithm, resulting in the first automatic skin cancer detection algorithm that processes active thermography. The first improvement involved automating the lesion area selection stage using U-Net CNN. The second improvement involves a new selection of features to be classified, which are processed using an SVM-based algorithm.

Regarding the lesion segmentation algorithm, training was performed using a set of dermoscopic images with the aim of segmenting lower-quality images, which were acquired with standard smartphone cameras under varying degrading conditions. The algorithm presented a Jaccard index of 0.79 using dermoscopic images, a value similar to that reported by other authors. As expected, when evaluating a set generated with a smartphone camera under varying acquisition conditions, the performance dropped to 0.60. The performance on the target image set is

considered satisfactory because it achieved 0.07 Jaccard points lower than that obtained using non-expert segmentation, which is based on pigmentation. It is consistent with the fact that the training dataset (HAM10000) considers pigmented pixels as lesions. This drop was consistent with the degraded resolution, lower illumination, and blurred areas of the test dataset. Even though this drop is significant, the segmentation result that can be performed without the training of a dermatologist.

The performances of both detection algorithms were evaluated using expert, non-expert, and Díaz (semi-automatic algorithm) segmentation. Considering a sensitivity of 85%, both detection algorithms showed their best performance with expert segmentation and the worst with Díaz segmentation, whereas U-Net segmentation provided a detection performance similar to that of non-expert segmentation.

The use of machine learning in classification increased the detection accuracy, reaching values of 86.61%, 75.77%, 75.29%, and 72.67% when using expert, non-expert, U-Net, and Díaz lesion selection, respectively. All of these results outperformed the original detection algorithm, which achieved an accuracy of 60.03%.

When comparing the performance of the automatic detection system with that of the ResNet152V2 network on the UdeC visible image set, the thermography-based approach demonstrated a 3% increase in the accuracy. However, when using expert segmentation, the thermography system outperformed it by nearly 14 percentage points in accuracy. These results indicate that the proposed automatic detection system has significant potential for further improvement.

This leads to the conclusion that the main problem with the proposed automatic algorithm is that the automatic segmentation fits the pigmented area of the skin very well, which is correct according to the HAM10000 set. However, there are differences with the expert criterion, which considers knowledge of the nature of the lesion, such that sometimes the lesion area does not necessarily match only the pigmented area, meaning that the actual lesion may be larger or smaller than the pigmented area. Thus, the pigmented area is only a projection of the lesion over the skin surface, and its malignancy may spread below the surface in a random fashion.

The difference between selecting only the pigmented area with respect to the expert criterion implies a low accuracy of the detection algorithm of approximately 11%, owing to the manner in which the classifiers were designed. This implies the detection of more benign and malignant cases, that is, when using the automatic segmentation, more biopsies would be requested than when using the segmentation performed by an expert. To improve this aspect, it is necessary to teach the segmentation network about segmentations performed using expert criteria, which is the main limitation of the proposed algorithm. As there is no set of hundreds of

lesions labeled with expert criteria, it is not yet possible to teach this knowledge to the network, but we are currently working on it.

We consider that skin cancer detection using active thermography is still relatively unexplored, possibly due to the challenges involved in addressing the detection problem with this technology, such as the high cost that this technology implied some years ago, and the knowledge required to successfully process active thermography data. Regarding processing, we consider the judicious selection of TRCs crucial for more successful discrimination between benign and malignant cases. Therefore, this study proposes a generic framework for skin cancer detection using active thermography. Owing to the high cost of IR technology, in the last decade, low-cost microbolometer technology has emerged, which may have lower image quality, but makes it feasible to transfer the technology developed for skin cancer detection.

It is expected that by using more robust detection algorithms, such as deep-learning algorithms over dynamic infrared imagery, their performance will be less affected by the lesion area selection stage. This is an ongoing research by our group, and we expect to report a novel detection algorithm that uses only deep-learning techniques within a few months. The purpose of this novel algorithm is to work with lower-quality IR images acquired using low-cost equipment.

Table 3.5: Comparison of related articles for detecting skin cancer using different types of imaging, differentiating the classification problems, where MM: Malignant melanoma; Malignant includes MM, Basal-cell carcinomas and Squamous-cell carcinomas.

Reference	**Methodology**	**Classification problem**	**Dataset**	**Technology**	**Performance**
[85]	Nasnet Mobile with transfer learning	MM vs. non-MM	HAM10000	Dermoscopy	97.90% accuracy
[86]	MobileNetV2-based transfer learning	MM vs. benign	ISIC2020	Dermoscopy	98.20% accuracy
[87]	CNN DenseNet-121 with multi-layer perceptron (MLP)	MM vs. non-MM	ISIC 2016, ISIC 2017, PH2 y HAM10000	Dermoscopy	Accuracy of 98.33%, 80.47%, 81.16% and 81.00% on PH2 , ISIC 2016, ISIC 2017 and HAM10000 datasets.
Own implementation	ResNet152V2-based transfer learning	MM vs. benign	HAM10000	Dermoscopy	90.63% accuracy
Own implementation	ResNet152V2-based transfer learning	MM vs. benign	UdeC	Low-quality visible images	72.49% accuracy
[14]	Deep learning	MM vs.benign		Passive IRT	96.91% accuracy
[14]	Deep learning	Malignant vs.benign		Passive IRT	57.58% accuracy
Proposed method using automatic U-Net segmentation	Machine learning	Malignant vs.benign	UdeC	Active IRT	75.29% accuracy
Proposed method using manual expert segmentation	Machine learning	Malignant vs.benign	UdeC	Active IRT	86.61% accuracy

4 Feasibility study of the use of infrared cameras for skin cancer detection under a proposed data degradation model

4.1 Microbolometer technology

IR technology is divided according to wavelength response range: 0.75-1.4 μm Near IR (NIR), 1.4-3 μm Short Wave IR (SWIR), 3-8 μm Medium Wave IR (MWIR), 8-15 μm LWIR and 15-1000 μm Far IR (FIR) [88].

Due to the skin emissivity is close to 1, it can be considered a black body. Based on Wien's Displacement Law, it is possible to define the maximum radiation wavelength of the skin λ_{max} [67]. Typically, the superficial skin temperature is $27°C$, so λ_{max} is approximately $10\mu m$. Then, LWIR technology is the most appropriate for skin screening, being the most widely used [35].

The most popular LWIR commercial detectors are made from an alloy of $Hg_{1\text{-}x}Cd_xTe$, corresponding to Quantum well and not refrigerated Bolometer technologies [89]. The quantum well infrared photodetector (QWIP) technology is preferred by the research and development area for presenting high thermal sensitivity (approximately 30 mK), higher temporal resolution (100 to 200 Hz) and higher spatial resolution (up to 1280×1024 pixels) [67]. However, cameras based on QWIP technology are refrigerated, which implies a high cost. As an alternative, non-refrigerated microbolometer technology has emerged, which today is the most popular, especially the low-cost commercial versions [47].

Recently, low-cost microbolometers have entered the market have lower sensitivity (about 70 mK), spatial resolution (about 320×240 pixels), and temporal (8 to 25 Hz). To tackle the lower sensitivity of these cameras, Bonmarin et al. [90] proposed the use of dynamic thermography, specifically the lock-in thermal imaging technique. The lower spatial resolution should not be a problem because

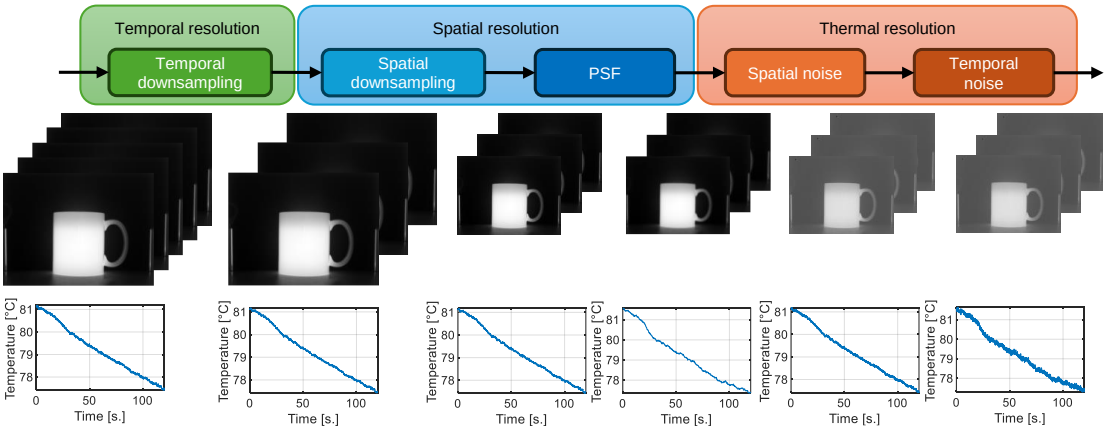

Figure 4.1: Schematic of the proposed degradation model, which addresses 3 areas: temporal, spatial and thermal resolution. Giving rise to a process composed of 5 stages.

it can be improved by choosing a suitable optical set-up. In relation to temporal resolution, according to [67], dermatological applications do not require high temporal resolution, because the conduction of heat through the human skin is relatively slow (with respect to metals, for example). Based on these fundamentals, we propose the use of low-cost microbolometer cameras for medical applications, specifically related to skin cancer.

4.2 Proposed data adaptation process

The proposed data adaptation scheme was designed to adapt IR videos acquired with a high-quality imager, e.g., the dataset described in Sec. 2.1, to a lower-quality camera characteristics. In this study, we intend to simulate the data acquisition with the cameras Xenics Gobi-640, Opgal Therm-App and Seek CompactPRO, but the proposed approach can be utilized in any pair of cameras. In order to define the scheme, the temporal, spatial and temperature resolution of the cameras were analyzed.

In Fig. 4.1, the proposed model for IR video degradation is presented, which as mentioned earlier, consists of the following 3 edges involving 5 stages:

1. Temporal resolution. This edge only contemplates the temporary downsampling stage. Due to the temporal resolution of the cameras to adapt is lower than the camera that captured the data, the sequence of IR images is downsampled following the proportion $fps_{HQ} : fps_{cam}$, where fps_{HQ} and fps_{cam} are the sample rates of the high-quality camera and the camera to adapt respectively.

2. Spatial resolution. This aspect addresses two areas: spatial resolution and optical distortion.

   (a) Spatial downsampling. The camera QmagiQ has a high instantaneous field of view (IFOV); however, the size of its FPA is smaller than the FPA of the cameras to model, in terms of the amount of pixels. In order to not add artificial elements of unknown effect, the spatial dimension of the images was not modified.

   (b) Point spread function (PSF). Images captured by each camera are usually affected by blurring due to optical distortions of the lens of each camera. The point spread function (PSF) applied corresponds to a 2D model, which was calculated as described by Jara et al. [91]. This model considers both optical and electronic aberrations.

   In order to transfer the optical response of the camera to be simulated to the high-quality videos, the PSF obtained from the lower-quality camera is applied to each frame of the higher-quality camera. Thus, the modified frame is the result of the spatial convolution of the image $\boldsymbol{u_k}$ with the PSF of the camera to simulate $\boldsymbol{PSF_{cam}}$:

$$\tilde{\boldsymbol{u}}_{\boldsymbol{k}} = \boldsymbol{u_k} * \boldsymbol{PSF_{cam}}. \tag{4.1}$$

   It is important to mention that usually the PSF is used to correct the optical distortions. In this study it is used to degrade the images, such a worst-case scenario is to be evaluated. So it is expected that using the real camera can improve the performance of the detection algorithm by correcting the optical distortions.

3. Thermal resolution. In order to simulate the thermal sensitivity of a camera, it is proposed to add characteristic noise of the camera to be modeled.

   The IR images contain spatial and temporal noise. According to Feng et al. [92], the noise affecting images captured by microbolometer IR cameras contains low, medium, and high-frequency spatial noise, along with horizontal and vertical component noise and non-uniformities.

   For applications utilizing active thermography, temporal noise is equally or more important than spatial noise affecting the cameras.

   (a) Spatial noise. The spatial noise characteristic of the camera to be simulated is added according to the following expression: $\tilde{\boldsymbol{u}}_{\boldsymbol{k}} = \tilde{\boldsymbol{u}}_{\boldsymbol{k}} + \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the spatial noise resulting from blackbody radiator measurements as described in Sec. 4.3.1. This spatial noise already contains

the low, medium and high-frequency components proposed by Feng et al. [92] since we have extracted it from actual measurements.

(b) Temporal noise. The temporal noise is composed of 2 components: low frequency ($\boldsymbol{N}^{LF}$) and high frequency ($\boldsymbol{N}^{HF}$). Where the low-frequency noise is associated with ambient temperature fluctuations, while the high-frequency noise is related to chemical and electrical characteristics of each camera component.

Therefore, the response of each detector over time is degraded according to the following expression:

$$\tilde{u}^{cam}_{i,j,k} = \tilde{u}_{i,j,k} + N^{LF}_{i,j,k} + N^{HF}_{i,j,k}. \tag{4.2}$$

To reiterate, the subscripts i, j, k mean that we are evaluating the noise component at the (i, j)-th pixel, at the k-th frame.

4.3 Noise characterization of imagers

4.3.1 Spatial noise characterization

As described earlier, the spatial noise of microbolometer IR cameras consists of low, medium, and high-frequency spatial noise, horizontal and vertical component noise, and non-uniformities. To transfer these noise features, a characteristic noise $\boldsymbol{\beta}$ of the camera to be modeled was extracted from measurements on a blackbody radiator.

To extract spatial noise, a sequence of 180 frames was captured. In our study, we considered setting the room temperature to $20°C$ because the clinical data acquisition protocol stipulates that the ambient temperature should be in the range of $20°C$ to $22°C$ [11]. Each camera was positioned in front of the blackbody, ensuring that the entire width of it was visible so that all cameras covered the same field of view at the same time.

Using the cube $\boldsymbol{u}$, an image $\bar{\boldsymbol{u}}$ was generated by averaging all the 180 frames. This effectively removes the temporal noise present in each frame, leaving only the spatial-noise components. Then, the spatial noise of a camera, $\boldsymbol{\beta}$, is defined as $\bar{\boldsymbol{u}}$ minus its two-dimensional mean value. Thus, $\boldsymbol{\beta}$ contains the horizontal, vertical, high, medium, and low-frequency noise, as well as the non-uniformities of each camera to be simulated, as we previously discussed.

In the study, for each camera, $\boldsymbol{\beta}$ was calculated for 6 measurements on the blackbody, i.e., between $15°C$ and $40°C$ with a step of $5°C$. The temperature range was selected to model the camera noise within the body temperature ranges (recall that we want to assesses the feasibility of these cameras in medical applications).

4.3.2 Temporal noise characterization

Usually, applications utilizing active thermography require a controlled temperature environment. In our skin cancer detection application, according to the data acquisition protocol, the room is set to a temperature between $20°C$ and $22°C$ using an air conditioning (AC) unit [11]. In our laboratory experiments with these cameras, we have observed that due to the on-off control performed by the AC, fluctuations in ambient temperature affect the camera measurements.

With the aim of understanding and modeling the ambient temperature fluctuations generated by the AC affecting the cameras to be simulated, measurements were taken using a Mikron-345 blackbody with the Xenics Gobi-640, Opgal Therm-App, and Seek CompactPRO microbolometer cameras. All these elements were placed in a 5 m^2 room, which is conditioned with an AC unit of 10000 BTU. The ambient temperature fluctuations were measured with the Elitech RC-4HC thermometer.

Similar to the spatial noise characterization, measurements were conducted by stabilizing the blackbody between $15°C$ and $40°C$ with a $5°C$ step. These measurements were performed by setting the ambient temperature of the AC to $15°C$, $20°C$, and $25°C$.

In Figs. 4.2.a, 4.2.b, and 4.2.c, a sample of measurements over 5 minutes on the blackbody with the Xenics, Opgal, and Seek cameras, respectively, is presented. It can be observed that the Xenics and Seek cameras exhibit jumps, which are effects of the Non-Uniformity Correction (NUC). It is also noticeable that all cameras exhibit low-frequency oscillatory behavior related to ambient temperature fluctuations. In the case of the Opgal camera, a highly correlated behavior to ambient temperature fluctuations is observed, with a linear relationship in the states of increasing or decreasing ambient temperature. Instances of changes in ambient temperature behavior are reflected with a variable delay, and the camera response is attenuated.

On the other hand, the Xenics and Seek cameras exhibit fluctuations in their measurements over time, whose behaviors are not easy to comprehend because the NUC modifies the gain of the measurements to correct errors originating from changes in ambient temperature. Applications utilizing active thermography ideally require smoothness over time in the temperature measurements of each detector in a camera. Therefore, jump correction produced by the NUC was applied to the Xenics and Seek cameras.

According to these observations, it is not possible to accurately model low-frequency temporal noise as a function dependent on ambient temperature variations. This is because the NUC makes the actions the camera will take to correct the measurements unpredictable. However, low-frequency oscillations are observed, which we propose to model using a Fourier series for each detector,

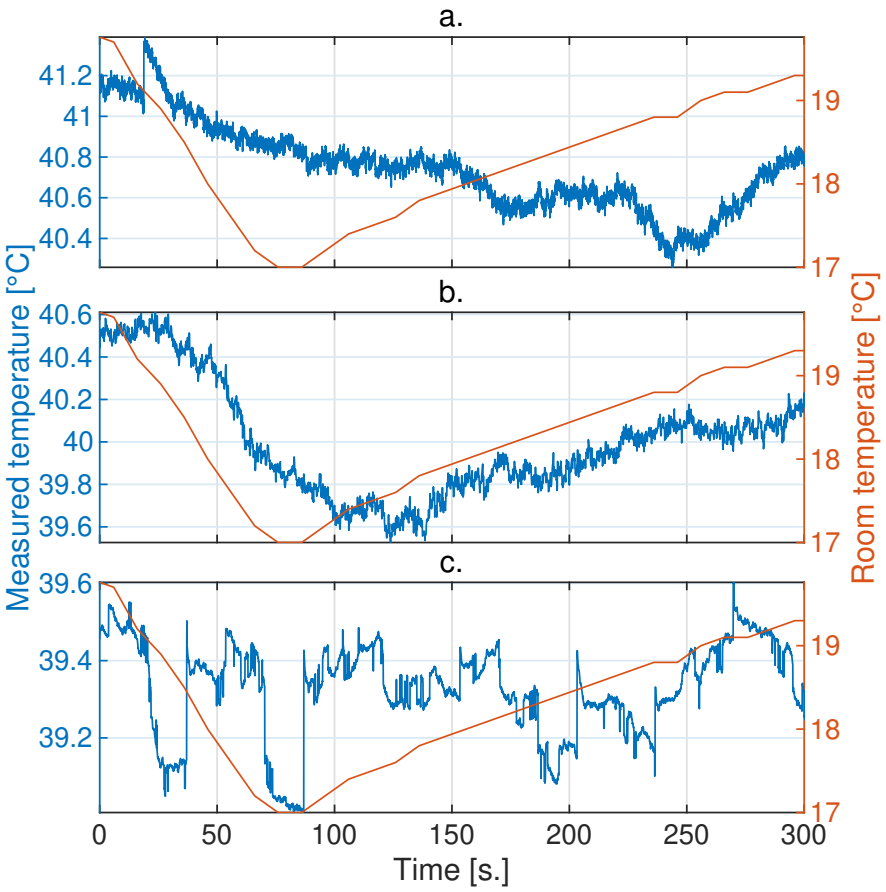

Figure 4.2: Samples of behavior of measurements from different IR cameras on a blackbody stabilized at $40°C$ in a room with controlled ambient temperature at $20°C$ using an AC unit. The order of measurements taken with the different cameras is as follows: a. Xenics Gobi-640, b. Opgal Therm-App and c. Seek CompactPRO.

according to the following Eq.:

$$N_{i,j,k}^{LF} = \sum_{z=0}^{Z} a_{i,j}^{z} cos\left(z\omega_{i,j} \cdot \frac{k}{f_s}\right) + b_{i,j}^{z} sin\left(z\omega_{i,j} \cdot \frac{k}{f_s}\right). \tag{4.3}$$

Then, the high-frequency component $N_{i,j}^{HF}$ corresponds to white Gaussian noise with standard deviation $\sigma_{i,j}$, whose value is calculated over the difference between the measurement over time of the detector $u(i,j)$ and its low-frequency component $N_{i,j}^{LF}$.

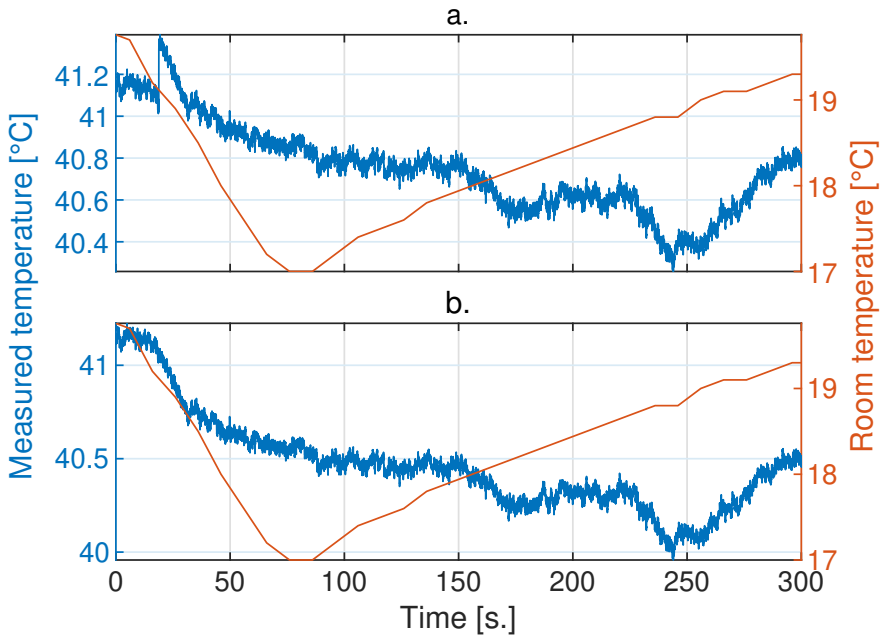

Figure 4.3: Sample of the result of jump correction produced by the NUC in the Xenics Gobi-640 camera. a. Uncorrected measurements; b. Corrected measurements.

4.4 Results

Two main experiments were conducted and shown here: The first aims to validate the proposed degradation model, and the second evaluates the feasibility of using low-cost IR cameras in skin cancer detection using active thermography.

4.4.1 Validation of the degradation model

To validate the degradation model, a simultaneous video was acquired with the 3 cameras positioned over a lesion on a volunteer. The video was captured according to the protocol described in Sec. 2.1.1, and then each video was registered using the registration algorithm described in Sec. 2.1.2.

For the purpose of validating the model, the video acquired with the Xenics Gobi-640 camera was considered as the high-quality video. Thus, the videos captured with the Opgal Therm-App and Seek CompactPRO cameras, which are of lower quality and cost, will be mimicked. Following the methodology described in Sec. 4.2, the video captured with the Xenics camera was adapted to the noise characteristics of the other cameras. First, the frame rate and image dimensions were reduced to match those of the camera being modeled. Second, the char-

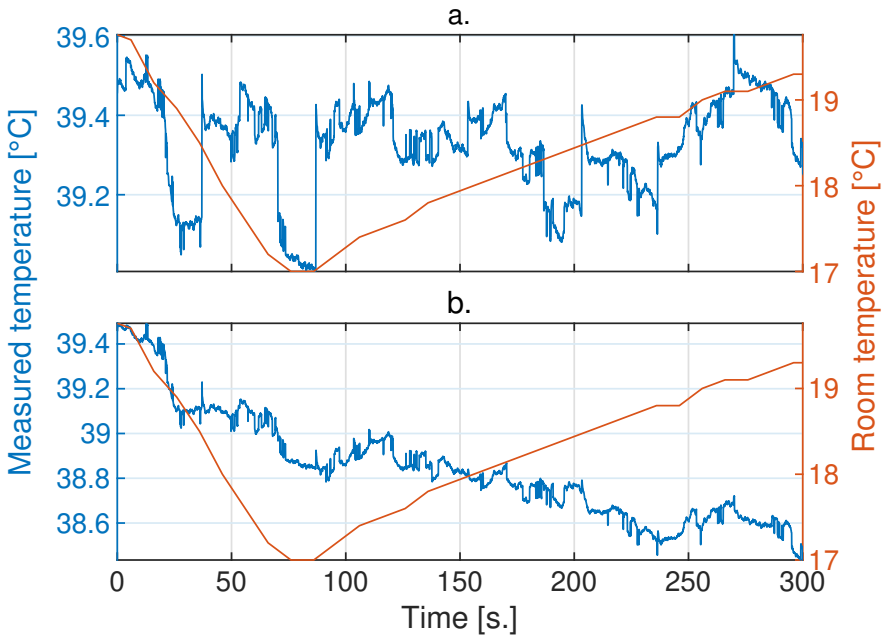

Figure 4.4: Sample of the result of jump correction produced by the NUC in the Seek CompactPRO camera. a. Uncorrected measurements; b. Corrected measurements.

acteristic PSF of each camera was applied to the reduced-sized imagery; third, the characteristic spatial noise was added to the degraded imagery, and, fourth, temporal noise specific to the camera being modeled is also added.

Thus, let $u_{i,j,\cdot}$ be a TRC captured at position (i, j); the $\cdot$ indicates we are considering all the time-samples of the video. It is considered that the temporal noise present in $u_{i,j,\cdot}$ is $NT_{i,j} = u_{i,j} - f_{i,j}$, where $f_{i,j}$ corresponds to the curve $u_{i,j}$ modeled as a double exponential, as defined in (3.1). Then, the low-frequency component is obtained by modeling $NT_{i,j}$ as a Fourier series, as defined in Sec. 4.3.2, to obtain $N^{LF}_{i,j}$. Meanwhile, the high-frequency component corresponds to white Gaussian noise, whose standard deviation is calculated over the residual noise, as described above.

Now, having the data-cube $\boldsymbol{u^c}$ acquired with camera c and the data-cube $\tilde{\boldsymbol{u}}^{\boldsymbol{c}}$ of a high-quality camera degraded to mimic the camera c, the Pearson correlation coefficient $\rho(u^c_{i,j}, \tilde{u}^c_{i,j})$ was calculated to estimate the level of similarity between the actual and simulated TRCs. It was obtained that the mimicking of the Opgal Therm-App camera, which includes 6889 TRCs, achieved an average correlation of 0.9756 ± 0.0104. Meanwhile, the mimicking of the Seek CompactPRO camera, which includes 7031 TRCs, achieved an average correlation of 0.9271 ± 0.0861.

Given that on average the Pearson correlation coefficient is greater than 0.9,

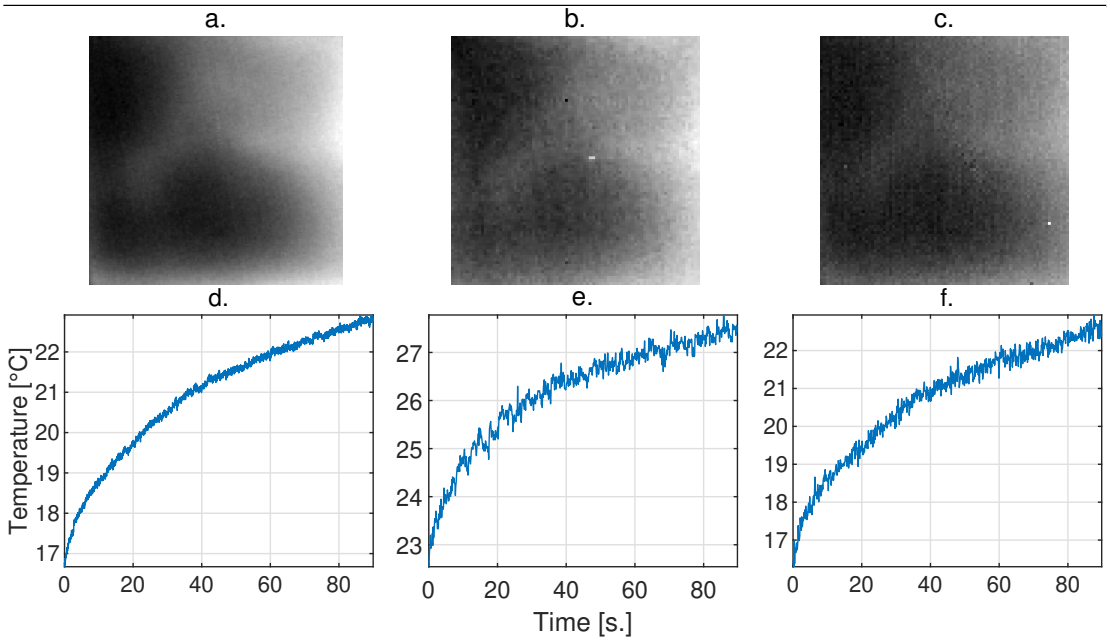

Figure 4.5: Sample of the degradation performed to a high quality video to Opgal Therm-App camera features. a. Image captured with the Xenics Gobi-640 camera, b. image captured at the same instant of time and same area with the Opgal Therm-App camera, c. Xenics image adapted to Opgal camera features, d. and e. correspond to representative TRCs of video acquired with the Xenics and Opgal cameras, respectively. f. TRC resulting from the adaptation.

we consider that the correlation between the real and simulated curves is very strong [93], i.e., the similarity between the data collected with the actual camera versus the data that mimic it is very high. Therefore, we consider the proposed degradation model to be valid within the scope of the TRC measurements. Clearly, more evidence must be collected to validate our model to mimic more cameras and on different cases of study.

A sample of the degradation model applied to the data-cube acquired with the Xenics camera to Opgal camera characteristics is presented in Fig. 4.5. Where Fig. 4.5.a corresponds to an image of the video IR acquired with the Xenics camera, and Fig. 4.5.d to a TRC of the cube; Fig. 4.5.b corresponds to an image from the video acquired with the Opgal camera, and Fig. 4.5.e corresponds to an image of the cube; Fig. 4.5.c corresponds to an image of the video acquired with the Xenics camera that was modified to features of the Opgal camera according to the proposed model, and Fig. 4.5.f corresponds to an image of the degraded cube.

Similarly, a sample of the degradation model applied to the same camera but

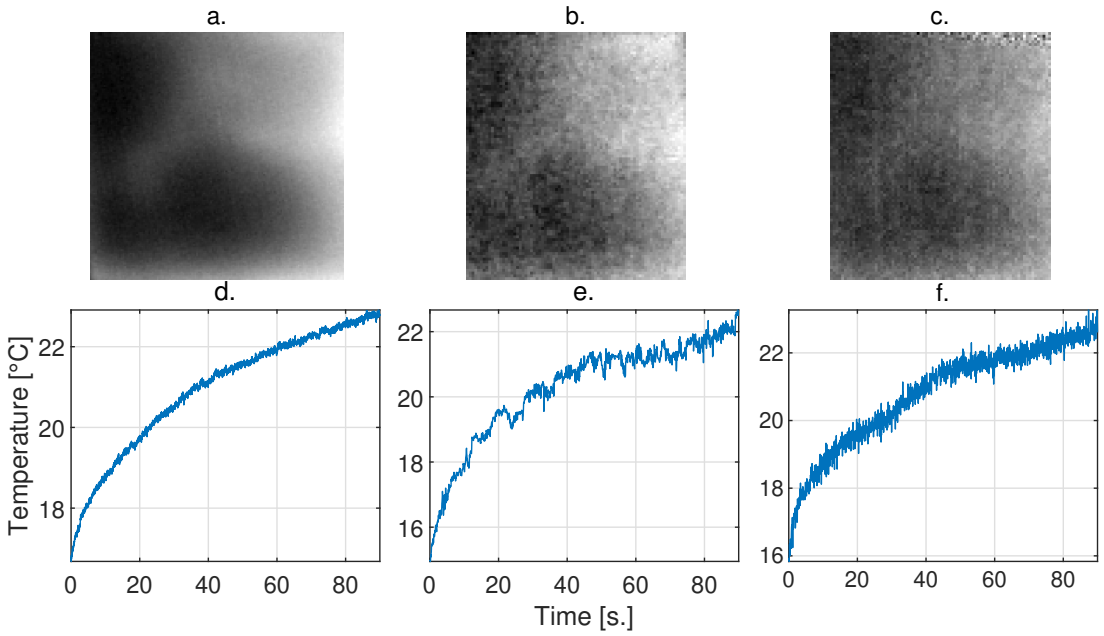

Figure 4.6: Sample of the degradation performed to a high quality video to Seek CompactPRO camera features. a. Image captured with the Xenics Gobi-640 camera, b. image captured at the same instant of time and same area with the Seek CompactPRO camera, c. Xenics image adapted to Seek camera features, d. and e. correspond to representative TRCs of video acquired with the Xenics and Seek cameras, respectively. f. TRC resulting from the adaptation.

to mimic the Seek camera features is presented in Fig. 4.6. Here, Fig. 4.6.a corresponds to an image of the video IR acquired with the Xenics camera, and Fig. 4.6.d to a TRC of the cube; Fig. 4.6.b corresponds to an image from the video acquired with the Seek camera, and Fig. 4.6.e corresponds to an image of the cube; Fig. 4.6.c corresponds to an image of the video acquired with the Xenics camera that was modified to features of the Seek camera according to the proposed model, and Fig. 4.6.f corresponds to an image of the degraded cube.

The proposed method does not rely on mimicking the TRCs exactly at every sample point, but on transferring all the noise characteristics from one camera to the other. At such, we observe that even though the TRCs are not visually identical, the low-frequency oscillations and the high-frequency noise from the low-cost cameras are actually included in the modeled data (see Figs. 4.6.e and 4.6.f, for example). Some of the differences that are not transferred are due to the NUC correction shutter, the cameras have constantly operating.

A larger graphic sample is presented in Appendix A. Where the result of mimicking each camera on the 4 types of lesions studied in this work is shown.

4.4.2 Feasibility study of using low-cost IR cameras in skin cancer detection

We now utilize the degraded data-cube to assess the performance of the skin-cancer detection algorithm we described in Sec. 3.2.2. As such, we can evaluate the performance one may get when low-cost infrared imagers can achieve in this task.

To understand the variability of the classifier performance, each one of the modeled datasets was trained using the bootstrap method, which involves taking multiple samples with replacement from the original dataset, generating a new dataset of the same size. The modified dataset was then divided, allocating 80% of the data for training and 20% for testing. This process was repeated 2000 times. The results previously reported by our group.

The algorithm performance with the original dataset acquired with the QmagiQ camera is shown in the second column of Table 4.1. The results obtained by degrading these videos to match the characteristics of the Xenics, Opgal, and Seek cameras are presented in the following columns of the same table. The reported indices include accuracy, TPR, TNR, and PPV, indicating the minimum, maximum, average (AVG), and standard deviation (SD) values for evaluation.

It is observed that, as expected, the best performance is achieved with the QmagiQ camera, with an average accuracy of 87.29%, sensitivity (TPR) of 87.26%, specificity of 87.15%, and precision (PPV) of 87.39%. It is noted that the adaptations to the Xenics and Opgal cameras offer a similar performance to that of the QmagiQ camera. The Xenics camera reached an accuracy of 84.33% and a sensitivity of 83.03%, while the Opgal camera achieved an accuracy of 84.20% and a sensitivity of 83.23%. Based on these results, we consider that both the Xenics and Opgal cameras are suitable for skin cancer detection, as they achieve similar levels of performance to the QmagiQ camera, with approximately 3% difference in accuracy.

Table 4.1: Performance of the skin cancer detection algorithm with a Random Forest classifier processing data captured with the high-quality QmagiQ camera and its adaptation to features of the Xenics, Opgal and Seek cameras.

Camera adaptation	**QmagiQ (original)**			**Xenics**			**Opgal**			**Seek**		
Index	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**
Accuracy (%)	62.07	100.00	87.29 ± 6.60	51.72	100.00	84.33 ± 7.11	55.17	100.00	84.20 ± 7.09	51.72	100.00	82.13 ± 7.43
TPR (%)	33.33	100.00	87.26 ± 11.49	28.57	100.00	83.03 ± 12.62	27.27	100.00	83.23 ± 12.58	28.57	100.00	79.77 ± 13.79
TPR MM (%)	0.00	100.00	76.42 ± 36.29	0.00	100.00	66.91 ± 40.11	0.00	100.00	71.68 ± 38.14	0.00	100.00	74.98 ± 37.30
TPR BCC (%)	20.00	100.00	87.23 ± 13.60	20.00	100.00	83.18 ± 14.96	0.00	100.00	83.60 ± 14.63	16.67	100.00	78.68 ± 16.40
TPR SCC (%)	0.00	100.00	95.94 ± 15.81	0.00	100.00	94.74 ± 18.37	0.00	100.00	90.92 ± 24.09	0.00	100.00	87.66 ± 26.62
TNR (%)	47.62	100.00	87.39 ± 8.98	47.62	100.00	85.28 ± 9.50	44.44	100.00	84.94 ± 9.75	36.84	100.00	83.74 ± 10.23
PPV (%)	28.57	100.00	82.45 ± 11.94	28.57	100.00	79.11 ± 12.90	25.00	100.00	78.91 ± 12.74	26.67	100.00	76.96 ± 13.25

Table 4.2: Performance of classical skin cancer detection methods and automatic passive and active thermography methods.

Methodology	**Detection problem**	**TPR (%)**	**TNR (%)**
Naked eye evaluation [7]	MM vs. benign	71.00	81.00
Dermoscopy evaluation [7]	MM vs. benign	90.00	90.00
Passive thermography + deep learning[14]	MM vs. benign	94.12	98.41
Passive thermography + deep learning[14]	Malignant vs. benign	62.81	57.58
Active thermography - QmagiQ camera	Malignant vs. benign	87.26	87.39
Active thermography - Xenics camera	Malignant vs. benign	83.03	85.28
Active thermography - Opgal camera	Malignant vs. benign	83.23	84.94
Active thermography - Seek camera	Malignant vs. benign	79.77	83.74

The worst performance was obtained by adapting to the characteristics of the Seek camera, with an average accuracy of 82.13%, sensitivity of 79.77%, and specificity (TNR) of 83.74%. The sensitivity is approximately 8% lower than the QmagiQ camera. Consequently, this camera is not really suitable for the skin cancer application we are investigating.

Regarding the system's performance in detecting the different types of skin cancer, the highest sensitivity was obtained in the detection of SCC lesions, with values of 87.66%, 90.92%, 94.74% and 95.94%, with the Seek, Opgal, Xenics and QmagiQ cameras, respectively. In the case of BCC lesions, a sensitivity of 78.68%, 83.60%, 83.18% and 87.23% was achieved with the Seek, Opgal, Xenics and QmagiQ cameras, respectively. The worst performance was obtained in the detection of MM lesions, with sensitivity values of 74.98%, 71.68%, 66.91%, and 76.42%, with the Seek, Opgal, Xenics, and QmagiQ cameras, respectively.

Table 4.2 presents the performance in terms of sensitivity and specificity of the proposed method using different cameras, along with the performance of highly trained dermatologists using naked eye evaluation and dermoscopy. The results obtained by Magalhaes et al. [14] using passive thermography and deep learning are also presented. Dermoscopy significantly outperformed naked-eye evaluation, demonstrating very high performance. However, it is important to note that this level of performance is achieved by dermatologists with years of clinical training. Along with this, the reported performance corresponds to a different detection problem. The passive thermography analysis method using deep learning proposed by Magalhaes et al. [14] outperforms dermatologists in this detection problem. However, when addressing the problem of distinguishing between malignant and benign lesions (with the malignant class including MM, BCC, and SCC), the algorithm's performance drops considerably. This is where the relevance of active thermography analysis becomes evident, achieving over 80% sensitivity and 85% specificity. This work demonstrates that it is possible to significantly reduce equipment costs while maintaining similar performance.

4.5 Discussion and Conclusions

In this work, a degradation model of IR videos was proposed and evaluated with the aim to mimicking the performance of different cameras in medical applications under laboratory conditions. Based on this model, videos captured with a high-quality camera were degraded to the characteristics of three low-cost imagers, namely, the Xenics Gobi-640, Opgal Therm-App and Seek CompactPRO cameras. These synthetic datasets were then used to evaluate the feasibility of using the modeled cameras over the skin cancer detection algorithm proposed by Soto and Godoy [72].

The proposed degradation model considers 3 key areas to mimic the performance of any camera: temporal, spatial and thermal resolution. It was demonstrated that the model achieves a high level of similarity at the level of TRC, which is the most important aspect in applications that use active thermography. Moreover, qualitatively, it is appreciated that spatially the model manages to transmit the texture of the images captured by the modeled cameras.

We consider that the proposed model may not fully transfer the characteristics of a lower-quality camera to images captured with a higher-quality camera. The main characteristics that the model does not consider are the following:

1. The adjustment of the size of the high-quality images to match the size of images captured with lower-quality cameras when the FPA of the higher-quality camera is larger. This approach was not considered because it requires an interpolation process, which may introduce noise that is not characteristic of the camera being simulated.

2. Shape and size of the detector. This characteristic is important for determining the minimum size of detectable objects. However, since the size of the skin lesions is much larger than the detector size, this characteristic was not considered. The size of one detector of microbolometer technology is approximately 20 μm, with the right optics it would be possible to detect a 40×40 μm lesion.

3. Temporal noise introduced by the shutter. When analyzing the temporal variations in the cameras due to changes in ambient temperature, the camera adjusts the offset and modifies the gain. However, incorporating this feature is problematic because the camera's logic for applying these adjustments is unknown and could not be determined.

Despite not considering the points mentioned above, the proposed model provides a high similarity between the degraded videos and the original ones. We expect that when taking these cameras to a clinical environment, the behavior

will be similar to the one described in this study. This is because the cameras were evaluated in one of the worst scenarios in terms of ambient temperature fluctuations. The air-conditioning unit aggressively controls the room temperature, and when activated, it decreases the temperature by approximately 5°C, which translates into a drift in the temperature measurements. However, we do not rule out that the measurements by the cameras are affected by other possible problems we may have not considered.

We believe that this model can be useful to evaluate the use of an IR camera in different applications, without the need to spend time, space and other resources generating a dataset with a camera that does not fit the requirements of the application. However, as we have shown, IR cameras are highly affected by the ambient temperature, so it is important to characterize the behavior of the camera in the environment in which it will be used, so that the simulation is as close as possible to its real performance. Nevertheless, some nuances may not be captured by our model, and that is exactly what we are currently doing in our current research.

We have shown that the model manages to simulate the behavior of the 3 cameras studied, achieving a high similarity of TRCs, with a Pearson correlation coefficient higher than 0.9. We believe that this model can be applied to most microbolometer technology chambers. However, we cannot be sure, especially in cameras with worse characteristics than those studied.

For our case study, we have demonstrated within certain boundaries, that the Xenics Gobi-640 and Opgal Therm-App are the most suitable IR cameras for skin cancer detection using active thermography. This is because their characteristics allow the evaluated algorithm to achieve a similar performance of the high-quality camera. Moreover, the NUC approach in these two cameras is less aggressive, allowing discontinuities to be easily corrected over time. Meanwhile, the Seek camera presents an average decay of 5% in accuracy and 7% in TPR, so we do not consider it suitable for skin cancer detection using active thermography.

It is relevant to analyze the performance of the tool in the detection of different types of skin cancer to observe its robustness to different scenarios. The performance of the system is outstanding when processing SCC-type lesions, reaching an average sensitivity between 87.66% and 95.94%, and with BCC-type lesions, an average sensitivity between 78.68% and 87.23% is achieved. Whereas, when analyzing MM lesions, the average sensitivity is in the range of 66.91% to 76.42%. It is important to note that the data set used is small, with only 7 cases of MM, which implies that the tests performed with the bootstrap technique contain at most 2 IR cubes of MM lesions. This is a disadvantage at the time of training the detection models because it does not manage to represent the generality of the behavior of this type of lesion; and at the time of evaluation, it implies that when a case is missed, the sensitivity is reduced to 50%. Because of this it is very

important to increase the dataset to make the system more robust.

This study was conducted on a dataset captured in the New Mexican population, i.e., primarily Caucasian or Hispanic individuals. We consider the data set to be small, but representative of this population. We strongly believe that the malignancy of the lesions is encrypted in the thermal recovery of the skin, so we consider it important to increase the dataset, considering that the sample is representative of all types of population. We are currently working to conduct a similar study in the Chilean population of the Biobío region. Conducting a clinical study involves many challenges, starting with the authorization of the clinical field, which must ensure the confidentiality of the patient's data and that the patient is not exposed to any treatment harmful to his or her physical and mental health. For this study we have informed consent of the subjects, and the clinical staff was in charge of blanking the data, so we never can correlate the patient identity with the corresponding data. Another challenge is adjusting to the limited space and time to capture the data in a real public health clinical setting. This has been the main reason that has led us to develop this video degradation model, which allowed us to select the right equipment for the research, reducing the amount of time involved in using many devices simultaneously and providing greater comfort when working in a small space.

Most of the algorithms try to find a good approach to detect skin cancer with the best-available sensor, our research tries to change the paradigm by evaluating the feasibility of these algorithms when applied to data acquired with low-cost sensors. The similarity of the results between QmagiQ, Xenics and Opgal cameras is due to the robustness of the detection algorithm. As evidenced by the results presented in Sec. 4.4.2 and Appendix B, more advanced classification techniques such as Random Forest, SVM and eXtreme Gradient Boosting (XGBoost) allow for similar performance in skin cancer detection when using high- or low-quality technology, as opposed to simpler algorithms such as K-nearest neighbors (KNN). Thus, we can conclude that thanks to advances in machine learning and feature extraction techniques, it is possible to use worse quality technology. We hope to report better results soon, we are developing new statistical tools to extract the spatial thermal information of a process that is hidden within noisy TRCs. For this, the degradation model will play a key role to understand the way the actual data is acquired with low-cost imagers.

As in any other type of cancer, early-detection is key for skin cancer patients. The tool presented in this work uses non-invasive, non-contact and low-cost technology that in a few minutes provides a screening of a suspicious lesion. The portability and the low-cost of our tool allow its rapid massification, being possible to be accessible in primary care centers, even in villages of difficult access, in which it is complex for a patient to be evaluated by a trained dermatologist. Our tool also aims to support specialists in the decision of whether to perform a

biopsy or not; with this, we hope to contribute to the optimization of resources, avoiding unnecessary biopsies. In this way our tool contributes to public health in the early detection of skin cancer, being a first filter to refer a patient to a dermatologist, support in the decision to perform a biopsy and reduce costs of more serious diseases caused by late detection of cancerous lesions.

5

Skin cancer detection by extracting information from thermoregulation curves using Transfer learning

5.1 Introduction

CNNs have become a cornerstone of modern artificial intelligence, particularly in fields like image recognition, object detection, and medical imaging. CNNs are uniquely capable of learning hierarchical features from raw data, such as edges, textures, and shapes, which enables them to achieve high accuracy in complex tasks. However, training CNNs from scratch often requires massive labeled datasets and significant computational resources, making this approach impractical for many specialized or resource-limited applications.

Transfer learning addresses these challenges by leveraging pre-trained CNN models, such as ResNet or VGG, which have been trained on large, diverse datasets like ImageNet [94]. This process involves fine-tuning the pre-trained model for a new, often smaller, dataset, enabling faster training with fewer resources. Transfer learning offers high performance, reduction of training time and computational requirements; these advantages are highly valued in problems where labeled data is scarce and high accuracy is critical, such as medical image analysis [95, 96].

Skin cancer detection through dermoscopy image processing is not the exception, there are developments that present accuracy higher than 90% using transfer learning models[17, 19, 20, 97].

Dorj et al. [17] proposes a skin lesion classification scheme, which uses AlexNet pre-trained CNN as a feature extraction stage, which are used by an ECOC SVM classifier for lesion classification. The classifier differentiates between 4 classes: AK, BCC, SCC and MM. It presents accuracy between 92.3% and 95.1%, sen-

sitivity between 96.9% and 98.9%, specificity between 86.73% and 94.17%. Experiments are performed with 3753 images collected from Google, Bing and other search engines.

Rezvantalab et al. [19] performed several classification systems for 8 types of skin cancer, using the DenseNet 201, ResNet-152, Inception v3 and InceptionResNet v2 architectures. The results obtained with the ResNet-152 architecture in the detection of MM with presenting 94.4% ROC AUC stand out. This ResNet network was retrained by adjusting all its layers to the HAM10000 set. The final layers were replaced, so that the network is able to classify 8 different diagnoses.

Acosta et al. [20] proposed a melanoma detector. The scheme consists of two stages: selection of the region of interest and classification using a ResNet-152 network, which discriminates between two classes, malignant (MM) and benign (nevus and AK). This scheme reports an accuracy of 90.4%, sensitivity of 82.0% and specificity of 92.5%.

Çakmak and Tenekeci [85] utilized transfer learning with the Nasnet Mobile architecture and the HAM10000 dataset to classify melanoma with an accuracy of 97.90% after data augmentation, demonstrating the effectiveness of deep learning methods for early melanoma detection.

Gajera et al. [87] proposed an automated framework using deep features extracted from eight CNN models, with DenseNet-121 combined with an MLP achieving superior melanoma detection accuracy of up to 98.33% on the PH dataset, highlighting the impact of boundary localization and normalization techniques.

Active thermography has also yielded outstanding results. Typically the videos that result from capturing thermal recovery are analyzed at the level of TRCs, i.e., time series. A strategy to classify TRCs using transfer learning, a lesion malignancy metric and its impact when working with other feature extraction techniques and machine learning is proposed below.

5.2 Study of lesion malignancy using Transfer learning

Given how successful Deep Learning, specifically CNNs has proven to be in image processing, it is logical to consider processing TRCs using CNNs for time series data. However, the success of Deep Learning relies on large datasets or the use of pre-trained networks (transfer learning). Since the thermography dataset is small, it was proposed to generate an image from a TRC, which would then be processed by a ResNet152V2 network. Each IR video contains hundreds or thousands of TRCs, so a feature representing the malignancy of most TRCs is proposed. This feature, called Average Malignancy (AM), is subsequently pro-

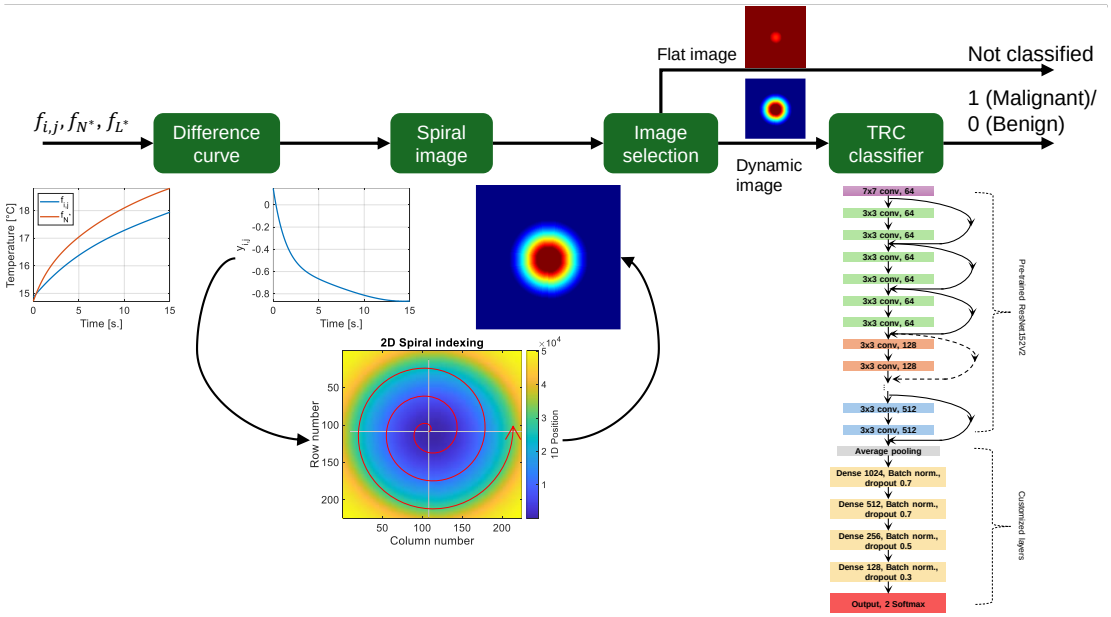

Figure 5.1: Schematic detector of malignant TRCs using transfer learning on images generated from the curves.

cessed by a classifier either alone or alongside other features to classify the lesion as benign or malignant.

Similar to other features described in Sec. 3.3, the calculation of the AM feature is performed on the set of TRCs L^*, i.e., TRCs with a similar initial temperature according to Eq. 3.2 that belong to the lesion area. Each TRC is processed according to the scheme presented in Fig. 5.1, which is composed of the following four stages:

1. **Difference curve.** The malignancy of a lesion is encoded in the evolution of the thermal recovery of the skin. This behavior varies according to the individual and the area of the body where the lesion is located. Therefore, it is necessary to analyze the behavior of the thermal recovery with respect to healthy tissue surrounding the lesion. Therefore, it is proposed to evaluate the malignancy of a TRC by means of the difference curve $y_{i,j} = f_{i,j} - f_{N^*}$, resulting from the subtraction of the TRC modeled by the double exponential (Eq. 3.1) $f_{i,j}$ with the characteristic TRC of the non-lesion area $\overline{TRC_{N^*}}$, also modeled as double exponential f_{N^*}. The modeling of the curves was performed in the time range of 0 - 15 seconds, considering 50176 points. The time range was selected because there is more dynamism in the thermal evolution, and the number of points due to the image size required to be generated.

2. **Spiral image.** From the curve $y_{i,j}$ the image I is generated, whose size is 224×224 pixels. I is formed by inserting the samples from the center outwards in a spiral fashion in increasing order of time, as shown in Fig. 5.1. Then, the image is adjusted with the following contrast formula: $I = \frac{I - min(y_{MN})}{max(y_{MN}) - min(y_{MN})}$, where $y_{MN} = f_{L^*} - f_{N^*}$, which represents the difference between the modeled feature TRCs for the lesion area (f_{L^*}) and the non-lesion area f_{N^*}. The I image is a single-channel (grayscale) image. Most pre-trained networks work with RGB images, so using the colormap jet, we generated 3-channel images, where the blue and red colors represent smaller and larger magnitudes, respectively.

3. **Image selection.** The contrast adjustment performed generates many flat images, i.e., images of mainly one color. So, images whose area of red, green or blue is greater than 60% are not classified.

4. **TRC classifier.** Using 2500 TRC images, of which 1250 correspond to TRCs from benign cubes and the rest to TRCs from malignant cubes, a ResNet-152V2 network was trained to discriminate between benign and malignant TRCs. The network was trained with 5-fold cross-validation, a learning rate of 0.0001 and early stopping with 50 epochs of patience. On average it obtained an accuracy of 92.3%.

Then, using the classification of the TRCs of the set L^* whose image turned out to be non-flat (dynamic), the average malignancy characteristic is calculated, which corresponds to the average value of the classification of the different TRCs.

5.3 Results

Figure 5.2.a shows the boxplot of the AM feature by class, extracted using expert segmentation. It is observed that in the case of benign lesions more than 75% of the cases present a value of $AM < 0.6$, and all lesions of type MM and SCC have a value of $AM < 0.6$. While BCC type lesions present more than 75% of the cases with values higher than 0.4. Therefore, this feature presents a low overlap of MM and SCC classes with BN lesions and a higher confusion of BCC and BN type lesions, as seen in the histograms shown in Fig. 5.2.b.

Figure 5.3.a shows the boxplot of the AM feature by class, extracted by using automatic segmentation using the U-Net network presented in Chapter 3. It is observed that in the case of benign lesions more than 75% of the cases present a value of $AM < 0.5$; all lesions of the type MM have a value of $AM \geq 0.6$, and most of SCCs present $AM \geq 0.4$. While BCC type lesions present more than 75% of the cases with values higher than 0.5. As when extracting the feature using

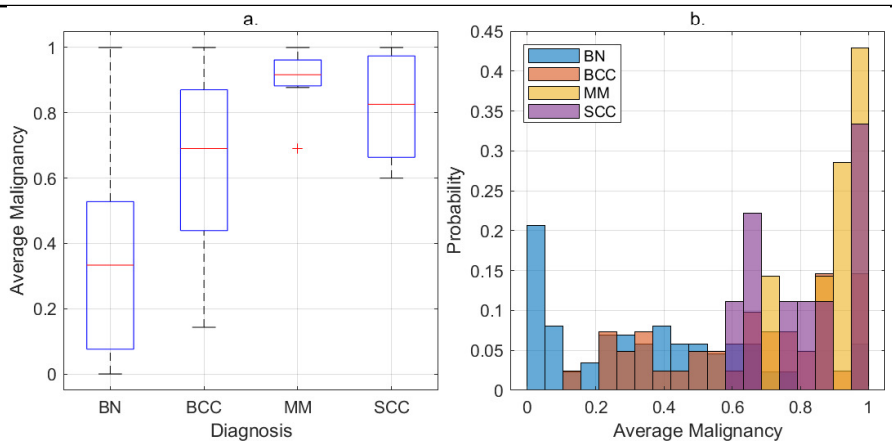

Figure 5.2: AM characteristic extracted using expert segmentation. a. Boxplot of AM by class. b. AM histograms by class.

expert segmentation, this feature presents a low overlap of MM and SCC classes with BN lesions and a higher confusion of BCC and BN type lesions, as seen in the histograms shown in Fig. 5.3.b.

In order to measure the contribution in the detection of the AM feature, detectors were trained using the detection algorithm proposed in Chapter 3, using expert and automatic segmentation (U-Net). Initially, the detectors were trained with the best feature vector that had been previously selected: $[E_d, \sigma_{rho_{B-L}}, \sigma_{proj_{B-N}}, \sigma_{d_{M-N}}]$. Then, the AM feature was added. The Random Forest classification technique was used in all these detectors.

Table 5.1 shows the results obtained with the 4 detectors. By using expert segmentation and adding the AM feature, the average detector accuracy increased by 4.66%, mainly attributable to better detection of MM lesions whose TPR increased by almost 10% and also to better classification of BN cases, implying an increase of TNR by 7.19%. Similarly, by using U-Net segmentation and adding the AM feature, the average detector accuracy increased by 6.20%, mainly influenced by better detection of MM lesions whose TPR increased by 8.62% and also to better classification of BN cases, implying an increase of TNR by 9.41%.

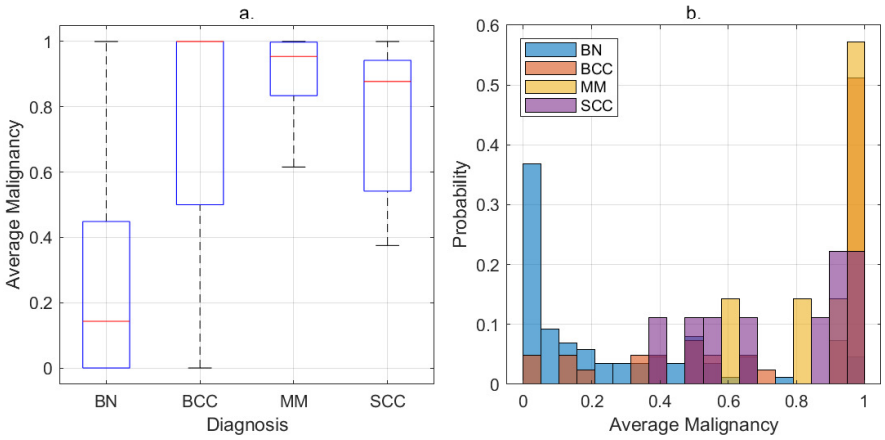

Figure 5.3: AM characteristic extracted using U-Net segmentation. a. Boxplot of AM by class. b. AM histograms by class.

Table 5.1: Performance of the skin cancer detection algorithm with expert and automatic U-Net lesion segmentation, processing the best feature combination with and without the AM feature.

Camera adaptation	**Expert**			**Expert+AM**			**U-Net**			**U-Net+AM**		
Index	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**
Accuracy (%)	17.24	100.00	86.20 ± 11.02	24.14	100.00	90.86 ± 7.87	17.24	100.00	83.00 ± 11.28	17.24	100.00	89.20 ± 8.89
TPR (%)	85.00	100.00	92.85 ± 4.94	85.00	100.00	93.73 ± 5.07	85.00	100.00	92.19 ± 4.71	85.71	100.00	93.59 ± 5.03
TPR MM (%)	0.00	100.00	86.43 ± 29.03	0.00	100.00	97.33 ± 13.52	0.00	100.00	90.25 ± 25.21	0.00	100.00	98.87 ± 8.24
TPR BCC (%)	66.67	100.00	92.63 ± 7.36	66.67	100.00	91.65 ± 7.49	66.67	100.00	91.49 ± 7.36	66.67	100.00	91.80 ± 7.37
TPR SCC (%)	0.00	100.00	97.57 ± 12.30	0.00	100.00	99.82 ± 3.41	0.00	100.00	95.86 ± 15.82	0.00	100.00	97.37 ± 12.10
TNR (%)	0.00	100.00	81.93 ± 18.21	0.00	100.00	89.12 ± 12.78	0.00	100.00	77.07 ± 18.64	0.00	100.00	86.48 ± 14.49
PPV (%)	17.24	100.00	79.63 ± 15.13	21.43	100.00	86.23 ± 12.60	17.24	100.00	74.79 ± 15.19	17.24	100.00	83.62 ± 13.75

5.4 Discussion and Conclusions

In this Chapter, a methodology to analyze TRCs using transfer learning techniques that process images was proposed. This is relevant, since CNNs present a low development for time series processing, unlike images.

A strategy was proposed to generate images from TRCs, which were termed spiral images, and thus be able to use transfer learning techniques to analyze the TRCs. These images make it possible to visualize a clear difference between malignant and benign TRCs. Transfer learning and in particular the ResNet152V2 network proved to be highly effective, providing better than 90% accuracy at the level of TRCs discrimination with the small training set.

Due to the variability of data, product of the non-uniformity in the skin cooling process and thermal noise, it is possible to find TRCs of a benign lesion with malignant behavior and vice versa. Because of this drawback it is necessary to propose a metric such as AM, which provides a result on the set of TRCs that compose the lesion.

By individually visualizing the AM feature, its potential to discriminate mainly MM and SCC lesions from benign cases was observed. By integrating AM into skin cancer detectors with other features, it was observed to increase the accuracy of the detectors by approximately 10%. Along with this, AM reduces the difference in accuracy of the detector using expert lesion selection with the one using automatic segmentation from 3.20% to 1.66%. With both detectors having an accuracy of about 90%.

6

Conclusions and Future Work

This thesis presents a comprehensive approach to the development of an automated and low-cost skin cancer detection system using active IRT, machine learning and deep learning methodologies. The conclusions of the study are as follows:

In this study, a skin cancer detection scheme that processes active thermography videos using machine learning was proposed. The scheme has several stages: lesion area selection, TRCs selection, feature extraction and classification.

The development of a set of 32 novel spatio-temporal features extracted from TRCs significantly enhanced detection performance when combined with advanced machine learning classifiers, such as SVMs, Random Forest, and XGBoost. Compared to the Euclidean detection system, this approach increased accuracy from 60.03% to 86.61%.

One of the objectives pursued in this thesis was to develop an automated skin cancer detection system, which was achieved through the implementation of a CNN U-Net. This automated approach addressed a critical limitation of previous systems, which relied heavily on manual lesion selection, thereby reducing variability in results due to operator expertise. The implementation of the automatic segmentation was trained and validated the system under diverse conditions, including low-quality clinical images with noise and non-uniform illumination. The proposed system demonstrated robustness and reliability in real-world clinical environments. This ensures its applicability in resource-limited settings where optimal imaging conditions may not be available.

In relation to the feasibility of using low-cost IR cameras for skin cancer detection, the characteristics and performance of the following three microbolometer cameras were studied: Xenics Gobi-640, Opgal Therm-App and Seek Thermal CompactPRO. From the observations made of the spatial and temporal noise characterization of each camera, a high-quality video degradation model was pro-

posed and validated, with the aim to mimic the performance of different cameras in medical applications under laboratory conditions. Using the proposed model, we mimic that the patient data was acquired with three different lower-cost cameras, namely Xenics Gobi-640, Opgal Therm-App, and Seek CompactPRO. The degraded data is used to evaluate the performance of a skin-cancer detection algorithm. The Xenics and Opgal cameras achieved an accuracy of 84.33% and 84.20%, respectively, and a sensitivity of 83.03% and 83.23%, respectively. These values closely match those from the undegraded data, indicating that employing these lower-cost cameras is appropriate for skin cancer detection. The Seek camera achieved an accuracy of 82.13% and a sensitivity of 79.77%. Based on these results, we conclude that this camera is appropriate for less critical applications.

With the aim of extracting more information from TRCs, a malignancy classifier for TRCs was developed by fine-tuning a ResNet152V2 CNN, which processes images generated from TRCs. This methodology is the result of a process that initially explored CNNs for time series, but yielded no positive results. CNNs for images are more advanced than those for time series, offering pre-trained networks such as ResNet, which can be fine-tuned for other purposes using small datasets. For this reason, I believe the classifier achieved an accuracy exceeding 90% in the classification of TRCs, based on the set of 2500 TRCs selected to train the network.

The number of TRCs analyzed in each video is variable, so the AM metric was proposed to measure the malignancy of the video. This metric corresponds to the average result of the TRCs contained within the lesion area. Analyzing the distribution of this metric revealed a high capacity to distinguish MM and SCC lesions from benign cases, with greater confusion observed between BCC cases and benign ones. By incorporating this feature into the previously selected set of four, the detection algorithm's performance improved: with expert segmentation, sensitivity increased from 92.85% to 93.73%, and specificity improved from 81.93% to 89.12%. Similarly, when working with automatic segmentation, sensitivity increased from 92.10% to 93.59%, and specificity improved from 77.07% to 86.48%.

Thus, a spatiotemporal framework was developed, which uses a CNN to process a visible image and automatically select the lesion area, a key step for calculating the classification parameters. The classification framework shows significant performance improvements compared to a classical detector. Furthermore, when working with data adapted to the characteristics of lower-cost and lower-quality cameras, similar performance was achieved. This makes the use of microbolometer cameras such as Xenics Gobi-640 and Opgal Therm-App suitable for skin cancer detection.

This study demonstrates that it is possible to implement a low-cost skin cancer detection system using active thermography with sensitivity and specificity higher than 90%, utilizing a spatio-temporal scheme based on deep learning.

The use of active IRT for early skin cancer detection is debatable, given the excellent results reported by studies processing dermatoscopic images, which exceed 95% accuracy. However, most of these studies focus on detecting MM. Additionally, they are conducted using high-quality images under controlled lighting conditions and without blurring caused by involuntary movements. In this work, it was demonstrated that a ResNet152V2 CNN can achieve over 90% accuracy on the HAM10000 dermatoscopy dataset. However, when applied to visible images without controlled lighting and, in some cases, blurred, its performance dropped to 72.49%. Therefore, the development of a low-cost system based on active IRT processing is considered relevant, as it is a technology that enables the correlation of thermal evolution differences between healthy and malignant tissue with the physiological properties of cancerous tissue.

IR technology continues to improve in quality and becomes more accessible each year. Furthermore, this thesis evaluated the performance of IR cameras under the worst-case scenario. It is expected that the performance of detectors will improve by enhancing image quality, eliminating NUC, and correcting optical distortions.

The *AM* feature demonstrates high potential for analysis at the TRC level using deep learning. There are many options to consider when processing the individual decisions of the TRCs, such as a voting scheme, analyzing a fixed number of TRCs, employing a hierarchical classification framework, among others. Therefore, I consider it possible to exceed the 95% sensitivity and specificity outlined in the hypothesis of this thesis.

In this manner, this investigation contributes to the ongoing efforts to make cancer diagnostics more accessible and less invasive. By leveraging affordable technology and advanced computational methods, the system minimizes the need for surgical biopsies, reduces healthcare costs, and improves patient outcomes.

6.1 Future Work

As future work, a new feasibility assessment will be conducted for the use of low-cost IR cameras, which will include a phase for improving image quality and a more extensive review of features extracted using the malignancy classifier for TRCs.

Along with this, there are plans to undertake a project for validation using real data acquired with low-cost IR cameras. This project will enable the characterization of skin in at least one region of Chile, validate the developed detection models, and further improve them.

Finally, it is important to encapsulate the developed software technology into a device that integrates both data acquisition and processing, enabling the transfer

of the technology to the population. This would serve as a tool to support skin cancer detection for medical professionals.

A

IR Video Degradation Sample

In this Section, supplementary material is presented, corresponding to a graphical sample of the result of degrading an active thermography IR video for a case of benign, malignant lesion of type MM, BCC and SCC, corresponding to Figures A.1, A.2, A.3, A.4, respectively.

It is observed that the curves of the QmagicQ camera have a low noise level, similar to the image. Regarding the degraded images, the best quality is achieved with the Xenics camera, followed by the Opgal and Seek cameras. It is evident that the model successfully transfers non-uniformities, vertical and horizontal noise, transferring the characteristic texture of each camera. Regarding TRCs, it is observed that the model successfully transfers the low-frequency noise associated with ambient temperature fluctuations, with variations in the noise affecting each detector.

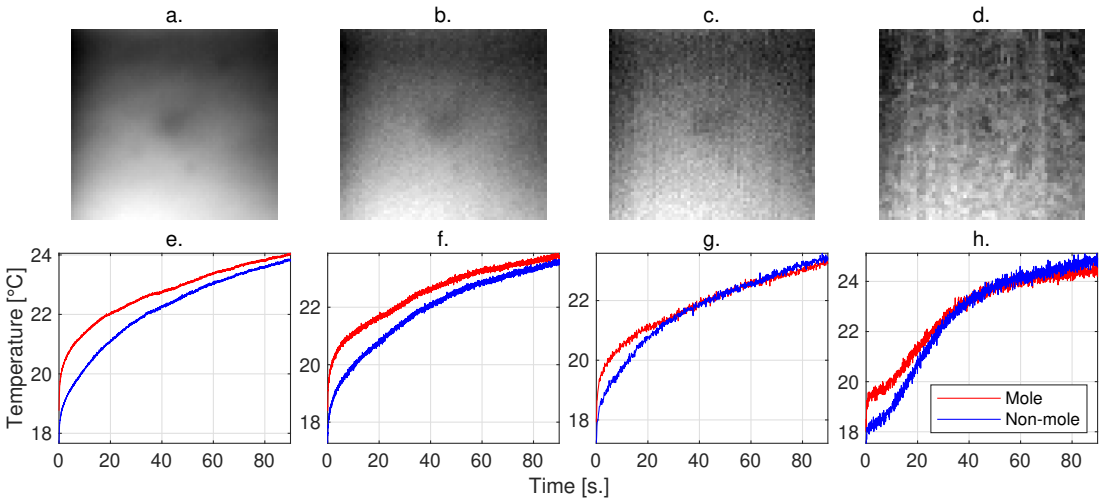

Figure A.1: Sample result of the degradation of an IR video to the characteristics of different simulated cameras in a case of a benign lesion. The top row shows an image from the video and the bottom row shows a characteristic TRC of the mole and non-mole areas. a. and e. Corresponds to high quality data captured with the QmagiQ camera; b. and f. data adapted to Xenics Gobi-640; c. and g. data adapted to Opgal Therm-App; d. and h. data adapted to Seek CompactPRO.

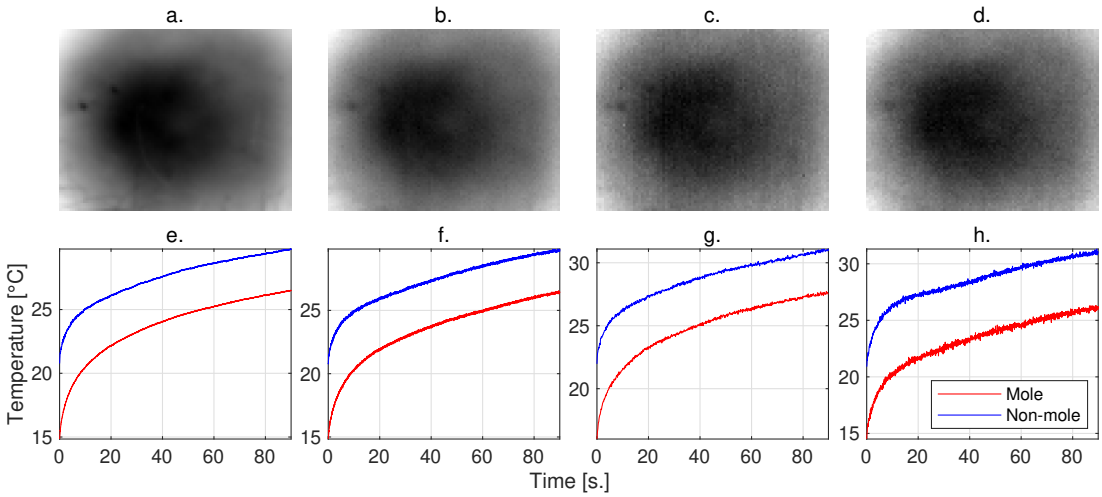

Figure A.2: Sample result of the degradation of an IR video to the characteristics of different simulated cameras in a case of a malignant lesion, diagnosed as MM. The top row shows an image from the video and the bottom row shows a characteristic TRC of the mole and non-mole areas. a. and e. Corresponds to high quality data captured with the QmagiQ camera; b. and f. data adapted to Xenics Gobi-640; c. and g. data adapted to Opgal Therm-App; d. and h. data adapted to Seek CompactPRO.

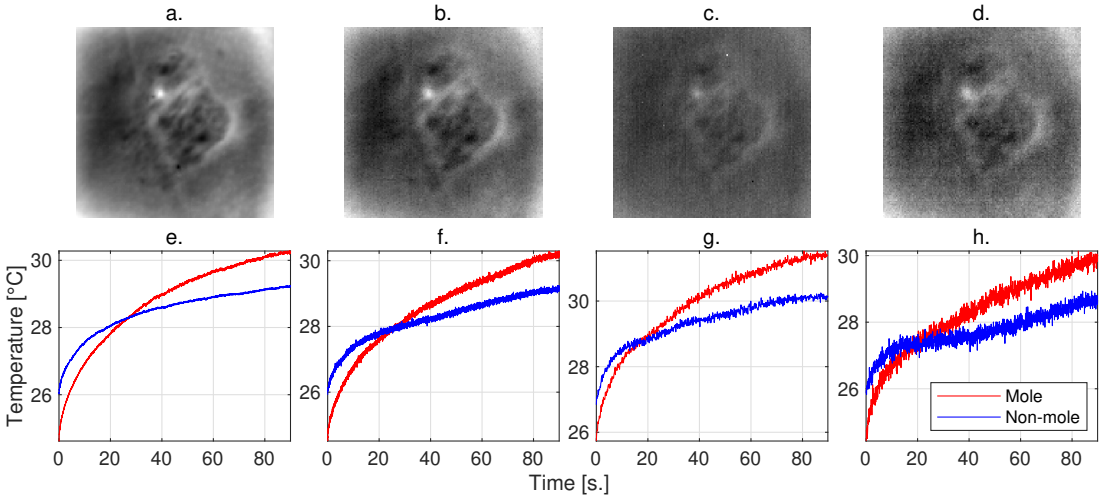

Figure A.3: Sample result of the degradation of an IR video to the characteristics of different simulated cameras in a case of a malignant lesion, diagnosed as BCC. The top row shows an image from the video and the bottom row shows a characteristic TRC of the mole and non-mole areas. a. and e. Corresponds to high quality data captured with the QmagiQ camera; b. and f. data adapted to Xenics Gobi-640; c. and g. data adapted to Opgal Therm-App; d. and h. data adapted to Seek CompactPRO.

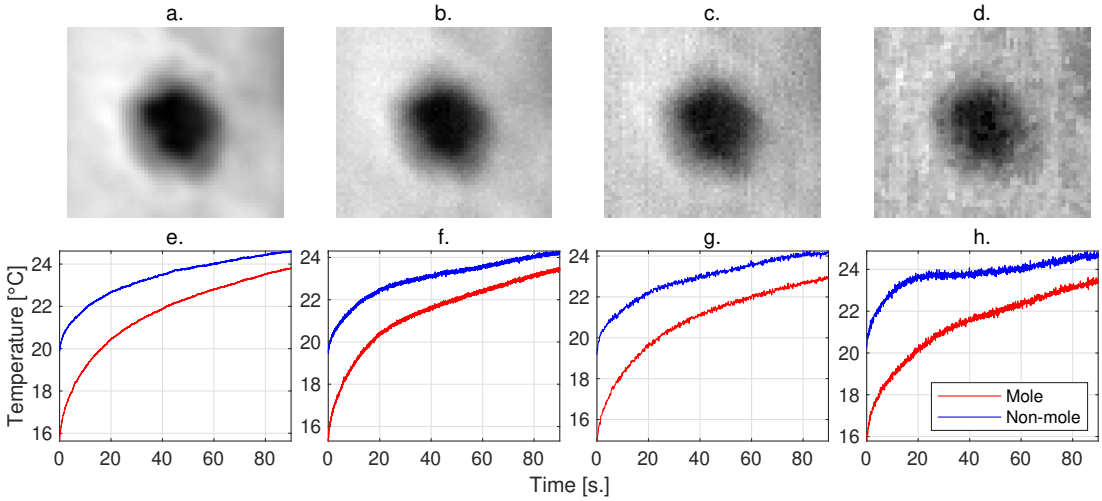

Figure A.4: Sample result of the degradation of an IR video to the characteristics of different simulated cameras in a case of a malignant lesion, diagnosed as SCC. The top row shows an image from the video and the bottom row shows a characteristic TRC of the mole and non-mole areas. a. and e. Corresponds to high quality data captured with the QmagiQ camera; b. and f. data adapted to Xenics Gobi-640; c. and g. data adapted to Opgal Therm-App; d. and h. data adapted to Seek CompactPRO.

B

Performance of the Detection Algorithm Using Different Machine Learning Techniques

This section presents the results of the detection system on the different types of cameras analyzed using a KNN, SVM and XGBoost classifier, corresponding to Tables B.1, B.2 and B.3, respectively. It is observed that the worst performance is achieved with the KNN technique, followed by XGBoost and SVM. It is important to note that when using the XGBoost and SVM techniques, a slightly lower performance is achieved than that reported in Sec. 4.4.2 with the Random Forest technique.

Table B.1: Performance of the skin cancer detection algorithm with a KNN classifier processing data captured with the high-quality QmagiQ camera and its adaptation to features of the Xenics, Opgal and Seek cameras.

Camera adaptation	**QmagiQ (original)**			**Xenics**			**Opgal**			**Seek**		
Index	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**
Accuracy (%)	48.28	100.00	78.39 ± 8.13	44.83	100.00	76.59 ± 8.15	41.38	96.55	75.37 ± 8.40	41.38	96.55	70.85 ± 8.83
TPR (%)	11.11	100.00	70.87 ± 15.39	9.09	100.00	66.57 ± 15.92	5.88	100.00	65.80 ± 16.19	7.69	100.00	58.97 ± 16.33
TPR MM (%)	0.00	100.00	53.41 ± 42.83	0.00	100.00	45.45 ± 42.45	0.00	100.00	52.73 ± 42.87	0.00	100.00	58.18 ± 42.16
TPR BCC (%)	0.00	100.00	71.44 ± 17.98	0.00	100.00	66.77 ± 19.28	0.00	100.00	65.87 ± 19.11	0.00	100.00	59.03 ± 19.56
TPR SCC (%)	0.00	100.00	82.25 ± 31.04	0.00	100.00	82.26 ± 31.26	0.00	100.00	75.85 ± 35.20	0.00	100.00	60.65 ± 39.37
TNR (%)	38.89	100.00	83.44 ± 9.97	41.18	100.00	83.25 ± 10.18	40.00	100.00	81.72 ± 10.48	26.67	100.00	78.75 ± 11.41
PPV (%)	11.11	100.00	74.26 ± 14.24	16.67	100.00	73.00 ± 15.00	14.29	100.00	70.83 ± 14.79	14.29	100.00	65.21 ± 15.92

Table B.2: Performance of the skin cancer detection algorithm with a SVM classifier processing data captured with the high-quality QmagiQ camera and its adaptation to features of the Xenics, Opgal and Seek cameras.

Camera adaptation	**QmagiQ (original)**			**Xenics**			**Opgal**			**Seek**		
Index	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**
Accuracy (%)	51.72	100.00	85.03 ± 7.31	51.72	100.00	84.12 ± 7.28	58.62	100.00	84.55 ± 7.33	48.28	100.00	79.73 ± 7.99
TPR (%)	11.11	100.00	82.09 ± 15.60	12.50	100.00	80.33 ± 15.45	22.22	100.00	80.25 ± 13.86	15.38	100.00	73.15 ± 15.12
TPR MM (%)	0.00	100.00	73.57 ± 38.00	0.00	100.00	69.63 ± 39.45	0.00	100.00	72.96 ± 38.23	0.00	100.00	68.69 ± 39.78
TPR BCC (%)	0.00	100.00	81.77 ± 17.07	0.00	100.00	81.26 ± 17.11	0.00	100.00	81.23 ± 15.71	11.11	100.00	74.02 ± 17.31
TPR SCC (%)	0.00	100.00	90.72 ± 25.04	0.00	100.00	85.83 ± 28.51	0.00	100.00	82.03 ± 31.62	0.00	100.00	73.18 ± 36.24
TNR (%)	43.75	100.00	87.12 ± 9.47	44.44	100.00	86.65 ± 9.43	52.94	100.00	87.42 ± 8.77	29.41	100.00	84.20 ± 10.04
PPV (%)	14.29	100.00	81.47 ± 12.70	25.00	100.00	80.53 ± 12.89	20.00	100.00	81.15 ± 12.48	25.00	100.00	75.86 ± 14.15

Table B.3: Performance of the skin cancer detection algorithm with a XGBoost classifier processing data captured with the high-quality QmagiQ camera and its adaptation to features of the Xenics, Opgal and Seek cameras.

Camera adaptation	**QmagiQ (original)**			**Xenics**			**Opgal**			**Seek**		
Index	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**	**Min.**	**Max.**	**AVG ± SD**
Accuracy (%)	55.17	100.00	85.69 ± 6.76	51.72	100.00	82.66 ± 7.41	55.17	100.00	82.64 ± 7.36	48.28	100.00	80.25 ± 7.75
TPR (%)	25.00	100.00	85.18 ± 12.31	28.57	100.00	81.26 ± 13.17	21.43	100.00	80.19 ± 13.59	14.29	100.00	79.19 ± 14.02
TPR MM (%)	0.00	100.00	76.21 ± 36.24	0.00	100.00	64.17 ± 40.81	0.00	100.00	67.43 ± 39.53	0.00	100.00	71.17 ± 38.85
TPR BCC (%)	16.67	100.00	84.39 ± 14.50	14.29	100.00	81.38 ± 15.42	0.00	100.00	80.01 ± 16.06	11.11	100.00	78.19 ± 16.61
TPR SCC (%)	0.00	100.00	96.36 ± 15.48	0.00	100.00	94.48 ± 18.60	0.00	100.00	91.22 ± 23.38	0.00	100.00	88.92 ± 25.86
TNR (%)	35.00	100.00	86.05 ± 9.18	50.00	100.00	83.67 ± 9.95	47.37	100.00	84.34 ± 9.91	38.46	100.00	80.93 ± 10.92
PPV (%)	22.22	100.00	80.45 ± 12.24	25.00	100.00	77.05 ± 13.21	20.00	100.00	77.62 ± 13.11	22.22	100.00	73.85 ± 13.40

Bibliography

[1] Devansh Bisla, Anna Choromanska, Jennifer A Stein, David Polsky, and Russell Berman. Skin lesion segmentation and classification with deep learning system. *arXiv preprint arXiv:1902.06061*, pages 1–6, 2019.

[2] Christopher Wild, Elisabete Weiderpass, and Bernard W Stewart. *World cancer report: cancer research for cancer prevention*. IARC Press, 2020.

[3] Natalia Sabatini-Ugarte, Montserrat Molgó, and Gustavo Vial. Melanoma en Chile ¿Cuál es nuestra realidad? *Revista Médica Clínica Las Condes*, 29(4):468–476, 2018.

[4] Deborah N. Dorrell and Lindsay C. Strowd. Skin cancer detection technology. *Dermatologic clinics*, 37(4):527–536, 2019.

[5] Darrell S. Rigel, Robert J. Friedman, Alfred W. Kopf, and David Polsky. Abcde—an evolving concept in the early detection of melanoma. *Archives of dermatology*, 141(8):1032–1034, 2005.

[6] Kirsten Bibbins-Domingo, David C. Grossman, Susan J. Curry, Karina W. Davidson, Mark Ebell, John W. Epling, Francisco AR García, Matthew W. Gillman, Alex R. Kemper, Alex H. Krist, et al. Screening for skin cancer: Us preventive services task force recommendation statement. *Jama*, 316(4):429–435, 2016.

[7] Oriol Yélamos, Ralph P. Braun, Konstantinos Liopyris, Zachary J. Wolner, Katrin Kerl, Pedram Gerami, and Ashfaq A. Marghoob. Usefulness of dermoscopy to improve the clinical and histopathologic diagnosis of skin cancers. *Journal of the American Academy of Dermatology*, 80(2):365–377, 2019.

[8] Thorsten M. Buzug, Steffen Schumann, Lucas Pfaffmann, Uwe Reinhold, and Jurgen Ruhlmann. Functional infrared imaging for skin-cancer screening. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2766–2769. IEEE, 2006.

[9] M. Pirtini Çetingül and C. Herman. Quantification of the thermal signature of a melanoma lesion. *International Journal of Thermal Sciences*, 50(4):421–431, 2011.

[10] Aldo Di Carlo, Fulvia Elia, Flora Desiderio, Caterina Catricalà, Francesco M Solivetti, and Luigi Laino. Can video thermography improve differential diagnosis and therapy between basal cell carcinoma and actinic keratosis? *Dermatologic therapy*, 27(5):290–297, 2014.

[11] Sebastián E. Godoy, David A. Ramirez, Stephen A. Myers, Greg von Winckel, Sanchita Krishna, Marianne Berwick, R. Steven Padilla, Pradeep Sen, and Sanjay Krishna. Dynamic infrared imaging for skin cancer screening. *Infrared Physics & Technology*, 70:147–152, 2015.

[12] Sebastián E. Godoy, Majeed M. Hayat, David A. Ramirez, Stephen A. Myers, R. Steven Padilla, and Sanjay Krishna. Detection theory for accurate and non-invasive skin cancer diagnosis using dynamic thermal imaging. *Biomedical optics express*, 8(4):2301–2323, 2017.

[13] C. Magalhaes, Ricardo Vardasca, and J. Mendes. Recent use of medical infrared thermography in skin neoplasms. *Skin Research and Technology*, 24(4):587–591, 2018.

[14] Carolina Magalhaes, João Manuel RS Tavares, Joaquim Mendes, and Ricardo Vardasca. Comparison of machine learning strategies for infrared thermography of skin cancer. *Biomedical Signal Processing and Control*, 69:102872, 2021.

[15] Ebrahim Nasr-Esfahani, Shadrokh Samavi, Nader Karimi, S. Mohamad R. Soroushmehr, Mohammad H Jafari, Kevin Ward, and Kayvan Najarian. Melanoma detection by analysis of clinical images using convolutional neural network. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1373–1376. IEEE, 2016.

[16] Jeremy Kawahara, Aicha BenTaieb, and Ghassan Hamarneh. Deep features to classify skin lesions. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, pages 1397–1400. IEEE, 2016.

[17] Ulzii-Orshikh Dorj, Keun-Kwang Lee, Jae-Young Choi, and Malrey Lee. The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications*, 77(8):9909–9924, 2018.

[18] KS Rajasekhar and T Ranga Babu. Skin lesion classification using convolution neural networks. *Indian Journal of Public Health Research & Development*, 10(12), 2019.

[19] Amirreza Rezvantalab, Habib Safigholi, and Somayeh Karimijeshni. Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. *arXiv preprint arXiv:1810.10348*, 2018.

[20] Mario Fernando Jojoa Acosta, Liesle Yail Caballero Tovar, Maria Begonya Garcia-Zapirain, and Winston Spencer Percybrooks. Melanoma diagnosis using deep learning techniques on dermatoscopic images. *BMC Medical Imaging*, 21(1):1–11, 2021.

[21] Ximena Wortsman and Jacobo Wortsman. Clinical usefulness of variable-frequency ultrasound in localized lesions of the skin. *Journal of the American Academy of Dermatology*, 62(2):247–256, 2010.

[22] N. Meyer, V. Lauwers-Cances, S. Lourari, J. Laurent, M. P. Konstantinou, J. M. Lagarde, B. Krief, Hadj Batatia, L. Lamant, and C. Paul. High-frequency ultrasonography but not 930-nm optical coherence tomography reliably evaluates melanoma thickness in vivo: a prospective validation study. *British Journal of Dermatology*, 171(4):799–805, 2014.

[23] Andrea Kozárová, Marek Kozar, Ingrid Tonhajzerová, Tana Pappova, and Eva Minariková. The value of high-frequency 20 mhz ultrasonography for preoperative measurement of malignant melanoma thickness. *Acta Dermatovenerologica Croatica*, 26(1):15–15, 2018.

[24] Darrell S. Rigel, Mrinalini Roy, Jane Yoo, Clay J. Cockerell, June K. Robinson, and Richard White. Impact of guidance from a computer-aided multispectral digital skin lesion analysis device on decision to biopsy lesions clinically suggestive of melanoma. *Archives of dermatology*, 148(4):541–543, 2012.

[25] Nazia Hameed, Anglia Ruskin, Kamal Abu Hassan, and M Alamgir Hossain. A comprehensive survey on image-based computer aided diagnosis systems for skin cancer. In *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*, pages 205–214. IEEE, 2016.

[26] Mehwish Zafar, Muhammad Imran Sharif, Muhammad Irfan Sharif, Seifedine Kadry, Syed Ahmad Chan Bukhari, and Hafiz Tayyab Rauf. Skin lesion analysis and cancer detection based on machine/deep learning techniques: A comprehensive survey. *Life*, 13(1), 2023.

[27] Adekanmi Adegun and Serestina Viriri. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. *Artificial Intelligence Review*, 54(2):811–841, 2021.

[28] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, Abdur Rehman Mahmood, Soliman Ayed Alsaiari, Abdul Hakeem M Saeed, Mohammed Olaythah Alraddadi, and Mater Hussen Mahnashi. Skin cancer detection: A review using deep learning techniques. *International journal of environmental research and public health*, 18(10):5479, 2021.

[29] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8(1):1–74, 2021.

[30] Hassan Ashraf, Asim Waris, Muhammad Fazeel Ghafoor, Syed Omer Gilani, and Imran Khan Niazi. Melanoma segmentation using deep learning with test-time augmentations and conditional random fields. *Scientific Reports*, 12(1):1–16, 2022.

[31] Manoj Kumar, Mohammed Alshehri, Rayed AlGhamdi, Purushottam Sharma, and Vikas Deep. A de-ann inspired skin cancer detection approach using fuzzy c-means clustering. *Mobile Networks and Applications*, 25:1319–1329, 2020.

[32] Anupama Damarla and Sumathi Doraikannan. Optimized one-shot neural architecture search for skin cancer classification. *Journal of Electronic Imaging*, 31(6):063053, 2022.

[33] Sanjay Singh, Urmila Pilania, Manoj Kumar, and Satya Prakash Awasthi. Image processing based skin cancer recognition using machine learning. In *2023 3rd International Conference on Intelligent Technologies (CONIT)*, pages 1–7. IEEE, 2023.

[34] G. Akilandasowmya, G. Nirmaladevi, SU. Suganthi, and A. Aishwariya. Skin cancer diagnosis: Leveraging deep hidden features and ensemble classifiers for

early detection and classification. *Biomedical Signal Processing and Control*, 88:105306, 2024.

[35] B. B. Lahiri, S. Bagavathiappan, T. Jayakumar, and John Philip. Medical applications of infrared thermography: a review. *Infrared Physics & Technology*, 55(4):221–235, 2012.

[36] Jan Verstockt, Simon Verspeek, Filip Thiessen, Wiebren A. Tjalma, Lieve Brochez, and Gunther Steenackers. Skin cancer detection using infrared thermography: Measurement setup, procedure and equipment. *Sensors*, 22(9), 2022.

[37] Mario Cristofolini, Bruna Perani, Francesco Piscioli, Giovanni Recchia, and Giuseppe Zumiani. Uselessnes of thermography for diagnosis and follow-up of cutaneous malignant melanoma. *Tumori Journal*, 67(2):141–143, 1981.

[38] Shazia Shaikh, Nazneen Akhter, and R Manza. Application of image processing techniques for characterization of skin cancer lesions using thermal images. *Indian Journal of Science and Technology*, 9(15):1–7, 2016.

[39] E. Benjumea, Y. Morales, C. Torres, and J. Vilardy. Characterization of thermographic images of skin cancer lesions using digital image processing. In *Journal of Physics: Conference Series*, volume 1221, page 012076. IOP Publishing, 2019.

[40] Jan Verstockt, Ruben Somers, Filip Thiessen, Isabelle Hoorens, Lieve Brochez, and Gunther Steenackers. Finite element skin models as additional data for dynamic infrared thermography on skin lesions. *Quantitative InfraRed Thermography Journal*, 0(0):1–20, 2023.

[41] Yılmaz Güvercin, Murat Yaylacı, Ayberk Dizdar, Ayhan Kanat, Ecren Uzun Yaylacı, Sevil Ay, Ahmet Atilla Abdioğlu, and Ahmet Şen. Biomechanical analysis of odontoid and transverse atlantal ligament in humans with ponticulus posticus variation under different loading conditions: Finite element study. *Injury*, 53(12):3879–3886, 2022.

[42] Mohammed El Sallah Zagane, Moulgada Abdelmadjid, Murat Yaylacı, Sahli Abderahmen, and Ecren Uzun Yaylacı. Finite element analysis of the femur fracture for a different total hip prosthesis (charnley, osteal, and thompson). *Structural Engineering and Mechanics, An Int'l Journal*, 88(6):583–588, 2023.

[43] Ayça Kurt, Murat Yaylacı, Ayberk Dizdar, Muhammed Enes Naralan, Ecren Uzun Yaylacı, Şevval Öztürk, and Binali Çakır. Evaluation of the

effect on the permanent tooth germ and the adjacent teeth by finite element impact analysis in the traumatized primary tooth. *International Journal of Paediatric Dentistry*, n/a(n/a), 2024.

[44] Vigneswaran Narayanamurthy, P Padmapriya, A Noorasafrin, B Pooja, K Hema, K Nithyakalyani, Fahmi Samsuri, et al. Skin cancer detection using non-invasive techniques. *RSC advances*, 8(49):28095–28130, 2018.

[45] Nibir K. Dhar, Ravi Dat, and Ashok K. Sood. Advances in infrared detector array technology. *Optoelectronics-Advanced Materials and Devices*, 7:149–188, 2013.

[46] P. Martyniuk, Jarek Antoszewski, Mariusz Martyniuk, Lorenzo Faraone, and A. Rogalski. New concepts in infrared photodetector designs. *Applied Physics Reviews*, 1(4):041102, 2014.

[47] Fariba Khodayar, Saeed Sojasi, and Xavier Maldague. Infrared thermography and ndt: 2050 horizon. *Quantitative InfraRed Thermography Journal*, 13(2):210–231, 2016.

[48] Luay Fraiwan, Mohanad AlKhodari, Jolu Ninan, Basil Mustafa, Adel Saleh, and Mohammed Ghazal. Diabetic foot ulcer mobile detection system using smart phone thermal camera: a feasibility study. *Biomedical engineering online*, 16(1):117, 2017.

[49] Luay Fraiwan, Jolu Ninan, and Mohanad Al-Khodari. Mobile application for ulcer detection. *The open biomedical engineering journal*, 12:16, 2018.

[50] RFM Van Doremalen, JJ Van Netten, JG Van Baal, MMR Vollenbroek-Hutten, and F Van der Heijden. Validation of low-cost smartphone-based thermal camera for diabetic foot assessment. *diabetes research and clinical practice*, 149:132–139, 2019.

[51] Rania Niri, Yves Lucas, Sylvie Treuillet, and Hassan Douzi. Smartphone-based thermal imaging system for diabetic foot ulcer assessment. 2019.

[52] Enrique Villa, Natalia Arteaga-Marrero, and Juan Ruiz-Alzola. Performance assessment of low-cost thermal cameras for medical applications. *Sensors*, 20(5):1321, 2020.

[53] Ayca Kirimtat, Ondrej Krejcar, Ali Selamat, and Enrique Herrera-Viedma. Flir vs seek thermal cameras in biomedicine: comparative diagnosis through infrared thermography. *BMC bioinformatics*, 21(2):1–10, 2020.

[54] Natalia Arteaga-Marrero, Lucas Christian Bodson, Abián Hernández, Enrique Villa, and Juan Ruiz-Alzola. Morphological Foot Model for Temperature Pattern Analysis Proposed for Diabetic Foot Disorders. *Applied Sciences*, 11(16), 2021.

[55] Ajat Shatru Arora et al. Automated prediction of diabetes mellitus using infrared thermal foot images: recurrent neural network approach. *Biomedical Physics & Engineering Express*, 10(2):025025, 2024.

[56] Erica Y Xue, Laurel K Chandler, Stephen L Viviano, and Jonathan D Keith. Use of flir one smartphone thermography in burn wound assessment. *Annals of plastic surgery*, 80(4):S236–S238, 2018.

[57] Fanfan Li, Min Wang, Ting Wang, Xiaolan Wang, Xiaoli Ma, Hua He, Guojing Ma, Dan Zhao, Qin Yue, Panpan Wang, and Minjie Ma. Smartphone-based infrared thermography to assess progress in thoracic surgical incision healing: A preliminary study. *International Wound Journal*, 20(6):2000–2009, 2023.

[58] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.

[59] Silvana Díaz, Thomas Krohmer, Álvaro Moreira, Sebastián E. Godoy, and Miguel Figueroa. An instrument for accurate and non-invasive screening of skin cancer based on multimodal imaging. *IEEE Access*, 7:176646–176657, 2019.

[60] QmagicQ-LLC. Datasheet QmagicQ Falcon, 2024. Accessed: 25/05/2024.

[61] Xenics. Datasheet Xenics Gobi-640, 2024. Accessed: 25/05/2024.

[62] Nax-Instruments. Datasheet Opgal Therm-app, 2024. Accessed: 25/05/2024.

[63] Seek Thermal Inc. Datasheet Seek CompactPRO, 2024. Accessed: 25/05/2024.

[64] MC Filipe Simões, JJ Simões Sousa, and AAC Canelas Pais. Skin cancer and new treatment perspectives: A review. *Cancer letters*, 357(1):8–42, 2015.

[65] Edward A. Arens and H. Zhang. The skin's role in human thermoregulation and comfort. 2006.

[66] Naoyo Nishida, Hirohisa Yano, Takashi Nishida, Toshiharu Kamura, and Masamichi Kojiro. Angiogenesis in cancer. *Vascular health and risk management*, 2(3):213, 2006.

[67] Mathias Bonmarin and Frédérique-Anne Le Gal. Thermal imaging in dermatology. In *Imaging in Dermatology*, pages 437–454. Elsevier, 2016.

[68] Abhijit Achyut Gurjarpadhye, Mansi Bharat Parekh, Arita Dubnika, Jayakumar Rajadas, and Mohammed Inayathullah. Infrared imaging tools for diagnostic applications in dermatology. *SM journal of clinical and medical imaging*, 1(1):1, 2015.

[69] Subramaniam Bagavathiappan, B. B. Lahiri, T. Saravanan, John Philip, and T. Jayakumar. Infrared thermography for condition monitoring–a review. *Infrared Physics & Technology*, 60:35–55, 2013.

[70] Chiwen Bu, Haijiang Xu, Zeqing Mao, Dong Zhang, and Chibin Pu. Non-destructive testing theoretical study on skin tumor detection using long-pulsed infrared thermal wave testing technology. *Thermal Science*, 23(3 Part A):1401–1408, 2019.

[71] Mirian Denise Stringasci, Ana Gabriela Salvio, David Sbrissa Neto, José Dirceu Vollet-Filho, Vanderlei Salvador Bagnato, and Cristina Kurachi. Discrimination of benign-versus-malignant skin lesions by thermographic images using support vector machine classifier. *Journal of Applied Physics*, 124(4):044701, 2018.

[72] Ricardo F. Soto and Sebastián E. Godoy. A novel feature extraction approach for skin cancer screening using active thermography. In *2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6. IEEE, 2023.

[73] Corinna Cortes and Vladimir Vapnik. Support vector machine. *Machine learning*, 20(3):273–297, 1995.

[74] RD Seeja and A Suresh. Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (svm). *Asian Pacific journal of cancer prevention: APJCP*, 20(5):1555, 2019.

[75] Sabari Nathan and Priya Kansal. Lesion net–skin lesion segmentation using coordinate convolution and deep residual units. *arXiv preprint arXiv:2012.14249*, 2020.

[76] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[77] Graham D. Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[78] Catarina Barata, M. Emre Celebi, and Jorge S. Marques. Improving dermoscopy image classification using color constancy. *IEEE Journal of Biomedical and Health Informatics*, 19(3):1146–1152, 2015.

[79] Quang Hung Nguyen, Hai-Bang Ly, Lanh Si Ho, Nadhir Al-Ansari, Hiep Van Le, Van Quan Tran, Indra Prakash, and Binh Thai Pham. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021, 2021.

[80] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[81] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, pages 1–7. IEEE, 2020.

[82] David L. Olson and Dursun Delen. Performance evaluation for predictive modeling. In *Advanced data mining techniques*, pages 137–147. Springer, 2008.

[83] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[84] John M Hancock. Jaccard distance (jaccard index, jaccard similarity coefficient). *Dictionary of Bioinformatics and Computational Biology*, 2004.

[85] Mustafa Çakmak and Mehmet Emin Tenekecı. Melanoma detection from dermoscopy images using nasnet mobile with transfer learning. In *2021 29th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2021.

[86] Javed Rashid, Maryam Ishfaq, Ghulam Ali, Muhammad R Saeed, Mubasher Hussain, Tamim Alkhalifah, Fahad Alturise, and Noor Samand. Skin cancer disease detection using transfer learning technique. *Applied Sciences*, 12(11):5714, 2022.

[87] Himanshu K. Gajera, Deepak Ranjan Nayak, and Mukesh A. Zaveri. A comprehensive analysis of dermoscopy images for melanoma detection via deep cnn features. *Biomedical Signal Processing and Control*, 79:104186, 2023.

[88] Abbas AlZubaidi, Yahya Ethawi, Georg Schmölzer, Sherif Sherif, Michael Narvey, and Molly Seshia. Review of biomedical applications of contactless imaging of neonates using infrared thermography and beyond. *Methods and protocols*, 1(4):39, 2018.

[89] Mary Diakides, Joseph D. Bronzino, and Donald R. Peterson. *Medical infrared imaging: principles and practices*. CRC press, 2013.

[90] Mathias Bonmarin and Frédérique-Anne Le Gal. Lock-in thermal imaging for the early-stage detection of cutaneous melanoma: a feasibility study. *Computers in biology and medicine*, 47:36–43, 2014.

[91] Anselmo Jara, Sergio N. Torres, Guillermo Machuca, Pablo Coelho, and Laura A. Viafora. Three-dimensional point spread function estimation method for mid-wave infrared microscope imaging. *Applied Optics*, 61(28):8467–8474, 2022.

[92] Tao Feng, Weiqi Jin, and Junjie Si. Spatial-noise subdivision evaluation model of uncooled infrared detector. *Infrared Physics & Technology*, 119:103954, 2021.

[93] Patrick Schober, Christa Boer, and Lothar A. Schwarte. Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, 126(5):1763–1768, 2018.

[94] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

[95] Ahmad Waleed Salehi, Shakir Khan, Gaurav Gupta, Bayan Ibrahimm Alabduallah, Abrar Almjally, Hadeel Alsolai, Tamanna Siddiqui, and Adel Mellit. A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, 15(7), 2023.

[96] Srinivasa Rao Burri, Sachin Ahuja, Abhishek Kumar, and Anupam Baliyan. Exploring the effectiveness of optimized convolutional neural network in transfer learning for image classification: A practical approach. In *2023 International Conference on Advancement in Computation Computer Technologies (InCACCT)*, pages 598–602, 2023.

[97] Abdelkader Alrabai, Amira Echtioui, and Fathi Kallel. Skin cancer detection based on transfer learning techniques. In *2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP)*, volume 1, pages 8–13, 2024.