

# Comparison of machine learning strategies for infrared thermography of skin cancer

Carolina Magalhaes<sup>a</sup>, João Manuel R. S. Tavares<sup>b, \*</sup>, Joaquim Mendes<sup>b</sup>, Ricardo Vardasca<sup>a, c</sup>

<sup>a</sup> *Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal*

<sup>b</sup> *Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal*

<sup>c</sup> *ISLA Santarem, Santarem, Portugal; Faculty of Computing, Engineering and Science, University of South Wales, UK*

**\* corresponding author:**

João Manuel R. S. Tavares – tavares@fe.up.pt

## **ORCID:**

Carolina Magalhaes: 0000-0001-5602-718X

João Manuel R. S. Tavares: 0000-0001-7603-6526

Joaquim Mendes: 0000-0003-4254-1879

Ricardo Vardasca: 0000-0003-4217-2882

**Funding:** This research was funded by Project LAETA [grant numbers UIDB/50022/2020, UIDP/50022/2020]; and the PhD Scholarship supported by FCT (national funds through Ministério da Ciência, Tecnologia e Ensino Superior (MCTES)) and co-funded by ESF through the Programa Operacional Regional do Norte (NORTE 2020) (EU funds) [grant number SFRH/BD/144906/2019].

## **ABSTRACT**

**Objective:** The aim of this work was to explore the potential of infrared thermal imaging as an aiding tool for the diagnosis of skin cancer lesions, using artificial intelligence methods.

**Methods:** Thermal parameters of skin tumours were retrieved from thermograms and used as input features for two machine learning based strategies: ensemble learning and deep learning.

**Results:** The deep learning strategy outperformed the ensemble learning one, showing good predictive performance for the differentiation of melanoma and nevi (Precision=0.9665, Recall=0.9411, f1-score=0.9536, ROC(AUC)=0.9185) and melanoma and non-melanoma skin cancer (Precision=0.9259, Recall=0.8852, f1-score=0.9051, ROC(AUC)=0.901).

**Conclusion:** IRT imaging combined with deep learning techniques is promising for simplifying and accelerating the diagnosis of skin cancer.

**Significance:** Despite ongoing awareness campaigns for skin cancer' risk factors, its incidence rate has continuously been growing worldwide, becoming a major public health issue. The standard first detection method – dermoscopy –, is largely experience-dependent and mostly used to assess melanocytic lesions. As infrared thermal imaging is an innocuous imaging technique that maps skin surface temperature, which may be associated to pathological states, e.g., tumorous lesions, it could be a potential aiding tool for all skin cancer conditions. The application of artificial intelligence methods to process the collected temperature data can save time and assist health care professionals with low experience levels in the diagnosis task. To the best of our knowledge, this is the first study where a data set of skin cancer thermograms is expanded and used for skin lesion differentiation with a deep learning approach.

**Keywords:** Biomedical; deep learning; ensemble learning; infrared thermal imaging.

## 1 Introduction

The human skin is the largest organ of the human body, influenced by internal and external factors and, consequently, highly subjected to lesions [1]. Some of them are of neoplastic origin, developing upon metabolic defects that lead cells to divide without control [2]. Malignant neoplasms, as melanoma, squamous cell carcinoma (SCC) and basal cell carcinoma (BCC), require special attention, due to their ability to rapidly grow and possibly result in major disfiguration or even death [3].

Even though continuous awareness campaigns for skin cancer risk factors, people demonstrate taste for tanning beds and poor sunscreen habits. The incidence rates have continued to grow worldwide, resulting in major public health threat [4,5]. The standard diagnosis is based on dermoscopy assessment, using qualitative guidelines (e.g., ABCDE criteria) that can result in a misdiagnosis, if carried out by an inexperienced physician, and the performance of unnecessary histopathological tests [6,7]. Several researchers are investigating the use of alternative techniques to allow an innocuous in-vivo diagnosis that ultimately lessens skin cancer mortality and its associated costs [8].

Infrared thermography (IRT) is an imaging method that measures the electromagnetic radiation emitted by an object surface in the IR range, converting this in temperature records. In the case of the human body, it can identify skin temperature shifts triggered by physiological alterations, that may be connect to possible pathological states [9,10]. Hence, a different thermal signature is expected from healthy and tumorous tissue, due to alterations on angiogenesis and metabolic rate [11]. Apart from potentially given a quantitative diagnosis of skin neoplasms based on functional parameters, IRT holds the advantage of being contactless and non-ionizing [12]. Its recent theoretical application for skin cancer assessment, presents mathematical models for the construction of melanoma lesions [13–16]. The results bear for the occurrence of a hyperthermic gradient across the skin tumour, with a maximum temperature difference between the tumour' centre and the surrounding healthy tissue. This behaviour is emphasized if the tumour' radii, depth and blood perfusion rate is increased. The practical implementation of IRT imaging in skin cancer diagnosis is the most explored topic. Still, studies are centred on the assessment of overall average lesion temperature and statistical attributes, or in a simple visual analysis of the thermal image, being the primary focus on the differentiation of melanoma and nevi and benign and malignant lesions [17–20].

As in other imaging modalities, the process of the retrieved IRT data can be performed with artificial intelligence (AI) approaches to save time and assist health care professionals in the diagnosis task. The use of machine learning (ML) algorithms for skin cancer diagnosis is widely studied when using dermoscopy images with a range of detection strategies, from simple algorithms to deep learning-based approaches [21,22]. As this image technique is predominantly implemented to evaluate melanocytic lesions, research with dermoscopy images usually targets the differentiation of melanoma and nevi. Most recent studies includes datasets with 200 to 23906 images, and all with good classification results [23,24]. Oliveira et al. tested several classifiers to perform this task, with a perfectly balance dataset of 1104 lesions [25]. A model based on optimum-path forest delivered top performance with accuracy (ACC), sensitivity (SN) and specificity (SP) values of 91.6%, 87% and 96.2%. Learners based on support vector machines (SVM) are commonly found, with Alamri and Alsaeed [26] showing a SVM learner capable of outperforming an ensemble model (151 lesions, ACC= 92.6%, SN=90%, SP= 96%). Another example, is Kalwa et al. [23] that combined SVM with Synthetic Minority Over-sampling

Technique (SMOTE) to improve performance metrics (200 lesions, ACC=88%, SN=80%, SP=90% and ROC(AUC)=0.85). Other implementations of single learners are also found in prior years, but with far less incidence than SVM and artificial neural network-based strategies (ANN) [27]. In fact, implementation of ANN has shifted from basic networks to deep learning (DL) models. Nowadays, several authors aim to construct complex approaches and reach higher diagnostic performance, without performing elaborate feature extraction and/or dealing with complex data during classification. Despite its inherent complexity, new optimization ideas and applications are constantly being studied. Hosny et al. [28] applied theory of transfer learning to a convolutional neural network and replaced the classification layer with a softmax layer (9350 images, ACC=97.7%, SN=97.34%, SP=97.34%), while Zhang et al. [29] developed a meta-heuristic optimization algorithm to deal with biases and distribution of weights during training of a Convolutional Neural Network (CNN) for diagnosis of melanocytic lesions (>20000 images, ACC=91%, SN=95%, SP=92). It is worth mentioning that the vast majority of authors chooses to test its AI strategies with dermoscopy images available from public databases, e.g., PH2, ISBI, DermIS, Dermquest, to ease the comparison of achieved results with those of other authors.

For IRT imaging, almost any author encompasses a ML stage for skin cancer classification. The ones that do prefer single learners, with a small sample of thermograms to classify, e.g., Stringasci et al. that used a SVM model to report a SN and SP of 82.9 and 85.7% for differentiation of SCC and actinic keratosis (AK) [30]. In a previous study, SVM, ANN and k-NN classifiers were tested to differentiate melanoma (16 samples) and melanocytic nevi (30 samples) lesions based on thermal profiles. The SVM learner achieved an accuracy of 84.2% and a SN of 91.3%, but a poor SP (11.1%) [31]. Thus, further expansion is needed. Other approaches, non-based on AI can be found, as the one of Godoy et al. that compared the thermal recovery curves of lesion and surrounding healthy tissue, after the application of a thermal stress. The authors differentiated benign (59 samples) and malignant (43 samples) lesions based on the temperature difference between the two areas. The temperature threshold defined by the authors that achieved top differentiation (SN=95, SP=83%) was 0.238 °C [32]. The same authors retrieved 10 Karhunen-Loève coefficients from the thermal recovery curves, a few years later, and applied it in a statistical algorithm. Almost 80% of instances were used to train the classifier 200 times and create a different classification threshold each time. Eleven samples were used as a validation set, being successfully identified by 36% of threshold classifiers [33].

The present work focusses on the collection of IRT parameters for different skin tumour types, followed by its machine classification using a learning ensemble method and a DL model. The aim was to assess the potential of this imaging modality along with AI techniques to function as a decision-support system for the diagnosis of skin cancer lesions.

This document encompasses five sections, the first introduces the problem, related background and research aim, it is followed by the methodology, which describes how the images were processed, the retrieved dataset and the data classification. At the third section the results of the data extraction and processing are presented, this is followed by a discussion relating the current work with previous research and is remarked by the conclusions and further work suggestions.

## 2 Methodology

The proposed approach focusses on the correct identification of skin neoplasms using AI strategies with input vectors composed by thermal features. The defined classification tasks concern the differentiation of: (1) melanoma (malignant) and melanocytic nevi (benign); (2) Squamous cell carcinoma (SCC) (malignant) and Actinic keratosis (AK) (Pre-cancerous lesion); (3) Melanoma (malignant) vs Non-melanoma skin cancer (NMSC) (SCC and Basal cell carcinoma (BCC)) (malignant); (4) Benign (AK, melanocytic nevi, seborrheic keratosis, cyst, other benign conditions) and Malignant lesions (melanoma, SCC, BCC). These tasks were chosen considering the current trends for detection challenges [34].

### 2.1 Image Processing and feature collection

Normally, the first step to assess lesion malignancy with dermoscopy images involves a pre-processing stage to eliminate possible artefacts or correct poor illumination and noise removal of the collected images. In the case of thermograms, the captured temperature values are most likely affected by heat dissipated from the body and environmental factors. Thus, if the international guidelines for correct thermal data collection are followed, this aspect should not be of concern [35,36]. As it was the case for the image data set used in this study, the pre-processing stage was simply composed by a histogram equalization process to adjust image contrast and ease lesion visualization.

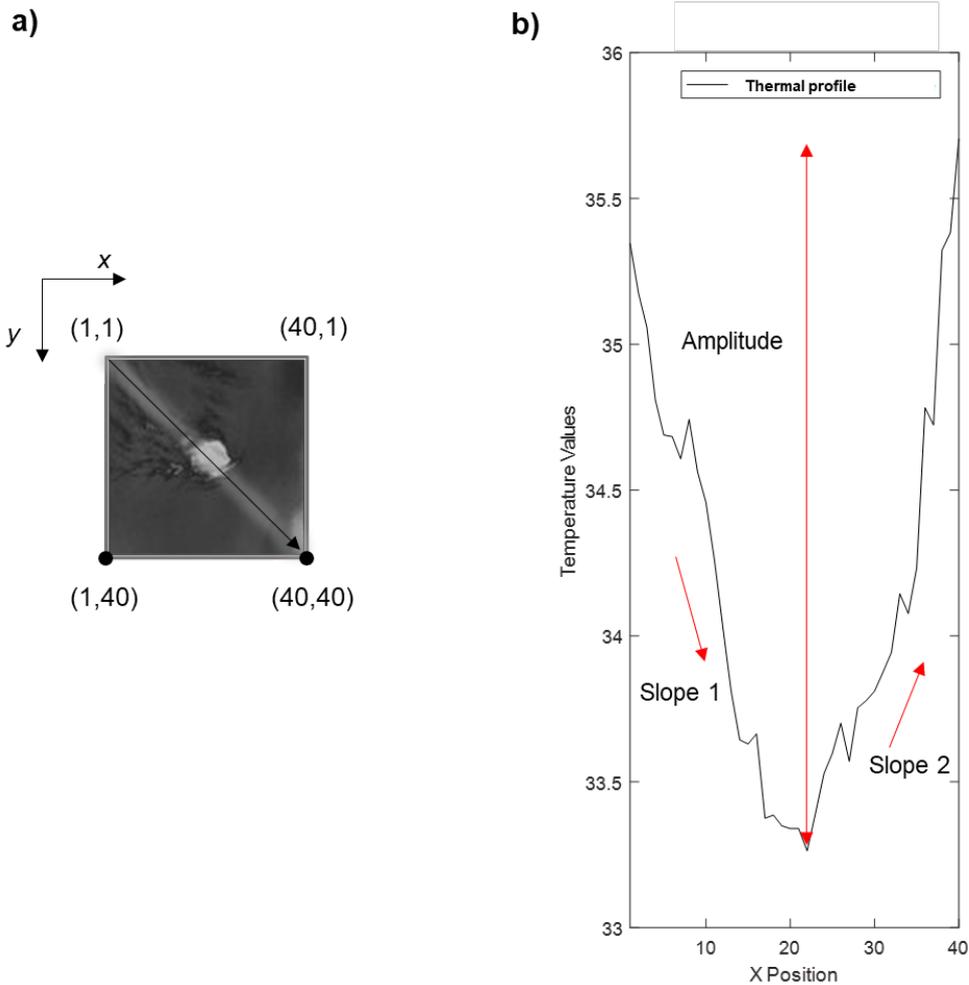
The presence of an unclear boundary between the pathological area and its background is common in infrared images, producing poor segmentation results when automatic strategies are used. Thus, the selection of an area including the skin neoplasm under evaluation was initialized with the selection of lesion centre by a human operator. Following, a region of 40 by 40 pixels, centred on the coordinates of the selected point, was automatically cropped including the entire lesion. To note that the 40x40 dimension was chosen, since it encompassed the leased area and surrounding healthy tissue of all skin neoplasms included in the image data set.

To construct a feature vector for classification, a diagonal was collected across the tumour area composed by 40 temperature measurements ( $T_{(1,1)}, \dots, T_{(40,40)}$ ), Figure 1a. The temperature vector described the thermal gradient across the neoplasm and the surrounding healthy tissue, Figure 1b. Normally, this gradient decreases/increases as it approaches lesion centre, increasing/decreasing again when moving away from it. So, feature reduction was performed through the estimation of two temperature slopes ( $S_1, S_2$ ) and central temperature amplitude ( $A$ ), using the collected 40 temperature measurements (Figure 1b):

$$S_1 = (T_{(20,20)} - T_{(1,1)}) / (20 - 1), \quad (1)$$

$$S_2 = (T_{(40,40)} - T_{(20,20)}) / (40 - 20), \quad (2)$$

$$A = |T_{(1,1)} - T_{(20,20)}|. \quad (3)$$



**Figure 1** – Collection of the thermal gradient across the neoplasm and surrounding healthy tissue for the selected 40x40 region (a) and correspondent computed thermal profile, thermal slopes and central temperature amplitude (b).

Hence, two final input vectors for comparison during the classification stage were built:  $T_v$  – with forty temperature values, and  $S_v$  – with two slope values and amplitude value, Table 1.

Input vector	Input features
$T_v$	$T_{(1,1)}, T_{(2,2)}, \dots, T_{(39,39)}, T_{(40,40)}$
$S_v$	$S_1, S_2, A$

**Table 1** – Computed input vectors for the classification stage and respective features.

## 2.2. Used image dataset

The used image dataset includes: 16 melanomas, 30 melanocytic nevi, 51 SCC, 29 AK, 118 BCC, 14 SK, 11 Cyst and 29 other benign conditions. The number of images included in most classes is considered low to perform classification with AI models. Thus, an attempt to expand the available dataset was made by the collection of temperature vectors in alternative directions ( $T_A$ ,  $T_B$ ,  $T_C$  and  $T_D$ ), Figure 2a:

- $T_A$ : composed by the temperature values of the diagonal from the top left corner (1,1) to the bottom right corner (40,40) of the selected area;

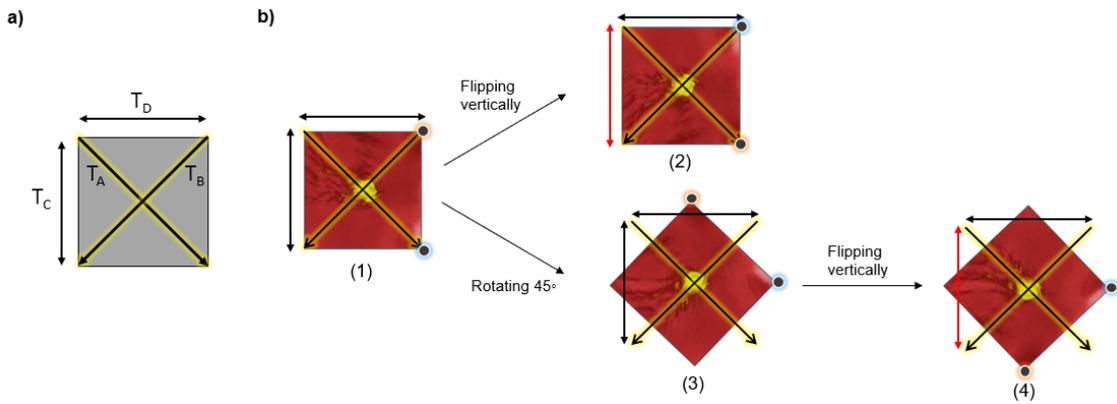
- $T_B$ : composed by the temperature values of the diagonal from the top right corner (40,1) to the bottom left corner of the selected area (1,40);
- $T_C$ : composed by the multiplication of  $T_A$  and  $T_B$  values, followed by its square root, starting on  $T_{A(1,1)} \times T_{B(1,40)}$ ,  $T_{A(2,2)} \times T_{B(2,39)}$ , ... and ending on ...,  $T_{A(39,39)} \times T_{B(39,2)}$ ,  $T_{A(40,40)} \times T_{B(40,1)}$ ;
- $T_D$ : composed by the multiplication of  $T_A$  and  $T_B$  values, followed by its square root, starting on  $T_{A(1,1)} \times T_{B(40,1)}$ ,  $T_{A(2,2)} \times T_{B(39,2)}$ , ... and ending on ...,  $T_{A(39,39)} \times T_{B(2,39)}$ ,  $T_{A(40,40)} \times T_{B(1,40)}$ ;

An example of the calculation of the first value for vectors  $T_C$  and  $T_D$  follows:

$$T_C = \sqrt{T_{A(1,1)} \times T_{B(1,40)}}, \quad (4)$$

$$T_D = \sqrt{T_{A(1,1)} \times T_{B(40,1)}}. \quad (5)$$

In addition, four different images of the same skin neoplasm case were considered: (1) original, (2) original vertically flipped, (3) original rotated 45°, (4) original rotated 45° and vertically flipped. For each image, the referred temperature vectors were computed, Figure 2b. Each temperature vector was defined as  $T_{z k}$ , being  $z$  the image number (1,2,3 or 4) and  $k$  the vector index (A, B, C or D). The input vectors  $T_{2c}$  and  $T_{4c}$  were equal to vectors  $T_{1c}$  and  $T_{3c}$ , respectively, and so were excluded from the expanded classification dataset.



**Figure 2** – Used data expansion strategy: temperature vectors collected for each image ( $T_A$ ,  $T_B$ ,  $T_C$  and  $T_D$ ) (a); thermal images considered for data collection, emphasizing in red thermal profiles excluded from expanded dataset (b).

All in all, 14 input vectors were able to be extracted from each input image, expanding the original experimental dataset fourteen times. Table 2 presents the number of instances considered for each classification task, before and after dataset expansion.

		<i>without DE</i>	<i>with DE</i>
<b>Classification task</b>	Melanoma vs Nevi	16 vs 30	224 vs 420
	SCC vs AK	51 vs 29	714 vs 406
	Benign vs Malignant	102 vs 185	1428 vs 2590
	Melanoma vs NMSC	16 vs 169	224 vs 2366

**Table 2** – Number of instances considered for the classification stage, without and with the dataset expansion (DE).

### 2.3 Skin neoplasm classification

The skin neoplasm classification stage was performed with two different strategies: a learning ensemble method and a DL network. An Intel® Core™ i7-6700HQ Quad Core 2.6GHz with Turbo Boost up to 3.5GHz (8GB DDR4 RAM) personal computer with a NVIDIA GeForce® GTX 950M graphic card was used for the implementation of both strategies, with the Anaconda 3 – Spyder 4.0 (Python 3.7 programming language) software.

As previously mentioned, each instance was characterized by the input vectors described in section 2.1:  $T_v$ , composed by 40 temperature values ( $T_{(1,1)}, T_{(2,2)}, \dots, T_{(39,39)}, T_{(40,40)}$ ), and  $S_v$  (feature reduced input vector), composed only by 3 values: two temperature slopes and central amplitude ( $S_1, S_2, A$ ). Both vectors were tested separately for each classification strategy to attest performance.

#### 2.3.1 Training and test sets

After the image processing step, stratified samples were collected to be divided for the classification instances into training and test sets with a 70:30 ratio. Table 3 presents the number of instances included in each training and test set, with and without data expansion.

	<i>without DE</i>		<i>with DE</i>	
	<i>Training set</i>	<i>Test set</i>	<i>Training set</i>	<i>Test set</i>
Melanoma vs Nevi	11 vs 21	5 vs 9	156 vs 294	68 vs 126
SCC vs AK	35 vs 20	16 vs 9	499 vs 284	215 vs 122
Benign vs Malignant	71 vs 129	31 vs 56	999 vs 1813	429 vs 777
Melanoma vs NMSC	11 vs 118	5 vs 51	156 vs 1656	68 vs 710

**Table 3** – Number of instances considered for the training and test sets, without and with the dataset expansion (DE).

The available image dataset presented a clear unbalance between classes. To correct this, the oversampling method Synthetic Minority Over-sampling Technique (SMOTE) was implemented in the training set to generate synthetic samples of the minority class [37].

### 2.3.2 Learning ensemble method

The first ML based strategy consisted of the construction of an ensemble model composed by two classifiers: random forest (RF) and support vector machine (SVM) (Figure 3). The preference for an ensemble as oppose to a single learner to solve the current prediction problems, is based on two motives. This type of models reduces the spread of outcome classes and consequently, attained performance metrics, increasing the robustness of the results [38]. Also, the combination of several learners tends to lead to an overall better predictive performance, if the base models are distinct and independent [38,39]. At the end, majority voting was used to combine the estimated class probabilities and reach a final prediction (Fig. 3). The classification results were compared with and without the use of SMOTE and with and without the use of the expanded dataset. The model performance was assessed using the metrics of Precision, Recall, F1-score, Area Under the Receiver Operating Characteristic Curve (ROC AUC).

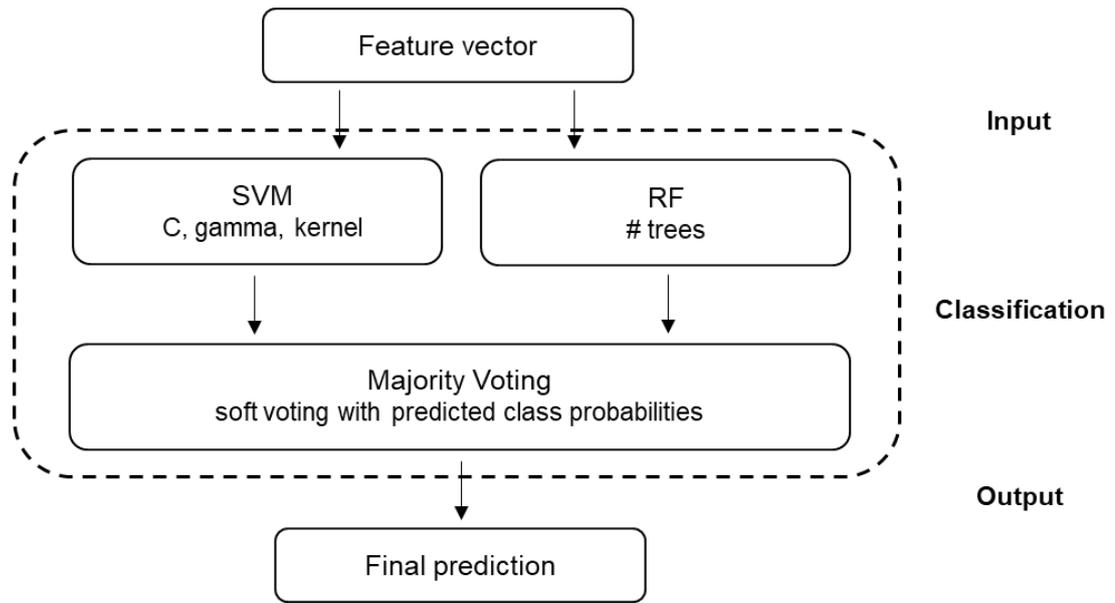
Random Forest is considered itself to be an ensemble of decision trees (DT) [40]. The algorithm is composed by a set number of individual DTs that deliver a class prediction. After running the entire “forest”, the results are averaged and the class with the highest vote is considered the final output [41]. Its working is summarized in the following procedures [42]:

1. Random selection of  $n$  samples from the training set (with replacement);
2. Create a decision tree from the constructed set of  $n$  sample. For each tree node:
  - a. Arbitrarily choose  $x$  features (without replacement);
  - b. Choose optimal features among  $x$  and split the node;
3. Repeat the first two steps for the desired number of nodes among the tree;
4. Repeat the first tree steps  $y$  times to create a forest with  $y$  number of trees;
5. Combine all tree results with majority voting to reach a final class prediction.

The advantage associated with RF is its diversity, as it can be modelled for continuous or categorical values, it handles missing data values, deals with regression and classification problems and reduces the chances of overfitting the data if enough trees are considered [42].

On the other hand, SVM creates a decision boundary (hyperplane) in an  $N$ -dimensional space (being  $N$  the number of features) to separate data from different classes [43]. The main goal is to maximize the distance between this hyperplane and the data examples that are closest to it (support vectors) [44]. This decision boundary is created using a kernel function, e.g., linear, polynomial, gaussian. When dealing with non-linear problems, the kernel function is essential to transform the original data and map it into a new hyperdimensional space, making it separable [42]. One of the major advantages of using SVM is its ability to function well with large input vectors.

The performance of the learning ensemble model was compared for the use of input vector  $T_v$  or input vector  $S_v$ , implementation of SMOTE and its absence during the training stage and the use of the expanded dataset versus the original one.

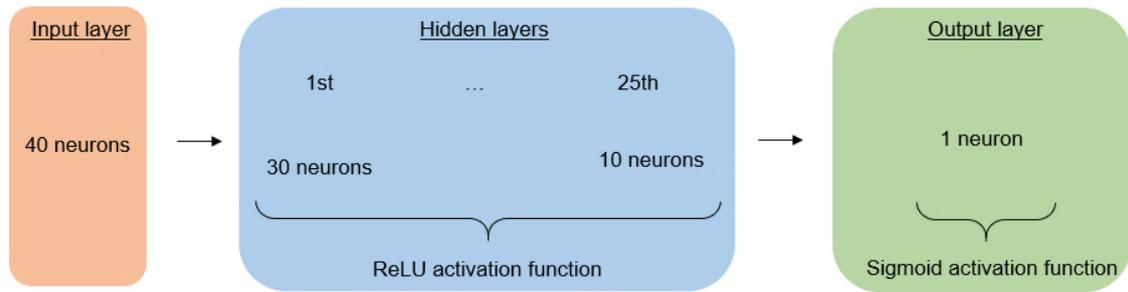


**Figure 3** – Scheme of the used learning ensemble method.

### 2.3.3 Deep learning network

Deep learning pertains to a ML subfield that uses artificial neural networks with multiple layers to imitate the workings of the human brain [45]. DL models can also be referred to as deep neural networks or deep neural learning. A scheme of the DL network built for this study is presented in Figure 4. An input layer is defined with 40 input neurons, due to the number of input features (40 temperature values of lesion' thermal profile). The input vector  $T_v$  was used to assess the classification performance of this model, instead of the feature reduced one, since DL strategies are usually capable of dealing with larger input vectors. Twenty-five hidden layers followed, being the first 5 layers composed by 30 hidden neurons and the remaining layers by 10 nodes. The output layer presents a single neuron, giving the final result for the proposed binary classification problems. The rectified linear unit (ReLU) activation function was implemented on the hidden layers, while a sigmoid activation function was preferred for the output layer to guarantee a final prediction between 0 (zero) and 1 (one) that can be easily transformed to an integer value [45,46]. The optimization algorithm Adam was used with a cross-entropy loss function and a batch size of 30 [47,48].

Hence, the performance of the DL approach was evaluated with the input vector  $T_v$  and the expanded dataset, instead of the original one, since DL strategies are used to deal with large amounts of data and even needed it to be effectively trained. The influence of SMOTE during the training stage was also tested. The model performance was assessed using the metrics of Precision, Recall, f1-score and Area Under the Receiver Operating Characteristic Curve (ROC AUC).



**Figure 4** – Scheme of the used deep learning network.

#### 2.3.4 Hyperparameter tuning

To select a set of optimal parameters for the built classification models, hyperparameter tuning was performed during the training stages of classification tasks for both AI approaches, using grid search [49].

In the ensemble model, different C, gamma and kernel options were tested for SVM. The regularization parameter C varied in {0.1, 1, 10, 100}, gamma in {0.1, 1, 2, 3} and kernel in {Linear, Radial Basis Function (RBF), 3<sup>rd</sup> grade polynomial (Poly-3<sup>rd</sup>)}. For RF, the only parameter assessed with grid search strategy was the number of trees included in the model, varying from 10 to 150 in increments of ten. Since the model deals quite well with noise, normally there is no increased need to search for other optimal parameters [42].

For the DL model, the number of hidden layers varied in {5, 15, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55} and the number of epochs in {10, 50, 100, 150, 200, 250, 300} were tested. Overall, the best balance between increased model performance and reduced computational time was achieved with the use of 25 hidden layers and 100 epochs, as previously mentioned, for all classification tasks, so that structure was preferred for the building of the used network.

### 3 Results

Four classification problems were addressed using the built learning ensemble model and DL network. The classification results obtained using the learning ensemble method with and without the entire feature set are presented in Table 4 and 5, respectively, and using the DL model in Table 6, based on the evaluation metrics of precision, recall, f1-score and ROC(AUC).

		without DE		with DE	
		without SMOTE	with SMOTE	without SMOTE	with SMOTE
SCC vs AK	Precision	0.7667	0.6625	<b>0.7535</b>	0.76
	Recall	0.6333	0.7	<b>0.7069</b>	0.698
	F1-score	0.6936	0.6807	<b>0.729</b>	0.7277
	ROC(AUC)	0.6444	0.6812	<b>0.7151</b>	0.6547
Melanoma vs Nevi	Precision	0.8	1	0.8718	1
	Recall	0.8	0.7778	0.8571	0.7403
	F1-score	0.8	0.875	0.8644	0.8508
	ROC(AUC)	0.8444	0.8889	0.9028	0.8701
Melanoma vs NMSC	Precision	0.7221	0.7324	0.8776	0.8718
	Recall	0.7455	0.8705	0.8431	0.7701
	F1-score	0.7336	0.7955	0.86	0.8178
	ROC(AUC)	0.7865	0.7764	0.8627	0.8888
Benign vs Malignant	Precision	0.5652	0.6724	0.7319	0.6275
	Recall	0.4727	0.6964	0.5349	0.73
	F1-score	0.5148	0.6842	0.6181	0.6749
	ROC(AUC)	0.5578	0.5418	0.3305	0.3667

**Table 4** – Results of precision, recall, f1-score and ROC(AUC) for classification tasks with and without dataset expansion and with or without SMOTE with entire feature set, using the learning ensemble model.

		without DE		with DE	
		without SMOTE	with SMOTE	without SMOTE	with SMOTE
SCC vs AK	Precision	0.6842	0.8182	0.7139	0.6725
	Recall	0.8667	0.5625	0.6735	0.7041
	F1-score	0.7647	0.6667	0.6931	0.6879
	ROC(AUC)	0.6	0.7146	0.6673	0.7066

Melanoma vs Nevi	Precision	0.79	0.8	0.8429	<b>0.8861</b>
	Recall	0.8	0.8778	0.8194	<b>0.8118</b>
	F1-score	0.7950	0.8371	0.8310	<b>0.8473</b>
	ROC(AUC)	0.7778	0.7222	0.8333	<b>0.8333</b>

Melanoma vs NMSC	Precision	0.8524	0.86	0.8555	<b>0.8461</b>
	Recall	0.7195	0.8412	0.7561	<b>0.8382</b>
	F1-score	0.7803	0.8505	0.8027	<b>0.8421</b>
	ROC(AUC)	0.7585	0.851	0.7651	<b>0.8107</b>

Benign vs Malignant	Precision	0.5962	0.6463	0.6782	<b>0.6981</b>
	Recall	0.5636	0.7464	0.4628	<b>0.7352</b>
	F1-score	0.5794	0.6928	0.5502	<b>0.7161</b>
	ROC(AUC)	0.5943	0.3055	0.3784	<b>0.3878</b>

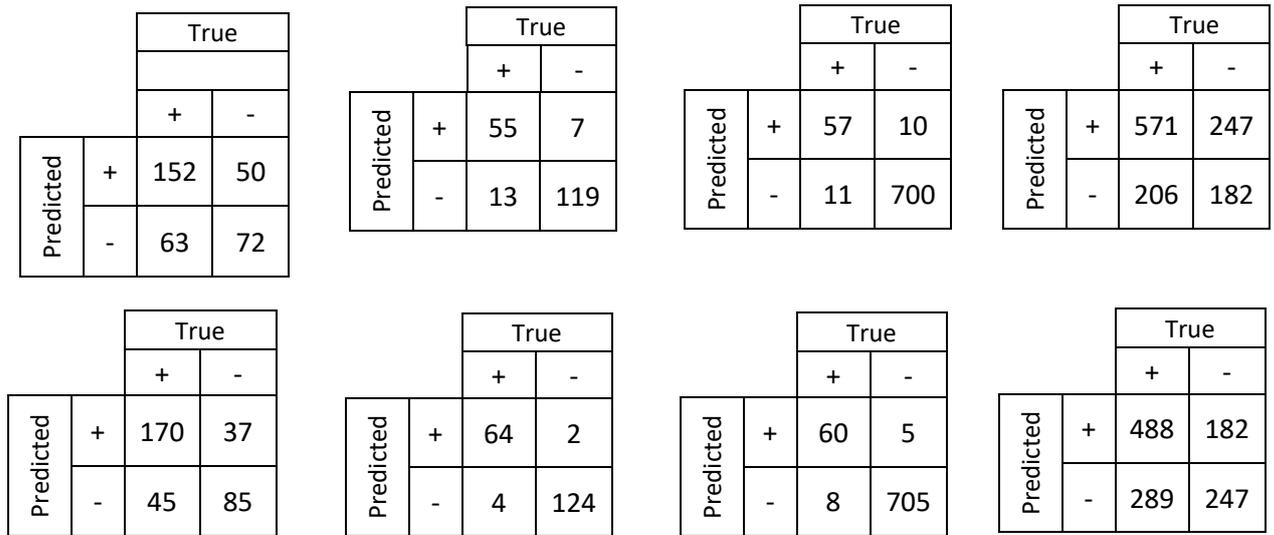
**Table 5** – Results of precision, recall, f1-score and ROC(AUC) for classification tasks with and without dataset expansion and with or without smote with feature reduced dataset, using the learning ensemble model.

	SCC vs AK		Melanoma vs Nevi		Melanoma vs NMSC		Benign vs Malignant	
	without SMOTE	with SMOTE	without SMOTE	with SMOTE	without SMOTE	with SMOTE	without SMOTE	with SMOTE
Precision	0.789	<b>0.8217</b>	0.9225	<b>0.9665</b>	0.9105	<b>0.9259</b>	0.7123	<b>0.7286</b>
Recall	0.7926	<b>0.7904</b>	0.9506	<b>0.9411</b>	0.8752	<b>0.8852</b>	0.6105	<b>0.6281</b>
F1-score	0.7908	<b>0.8057</b>	0.9363	<b>0.9536</b>	0.8925	<b>0.9051</b>	0.6575	<b>0.6747</b>
ROC(AUC)	0.7098	<b>0.7094</b>	0.882	<b>0.9185</b>	0.9265	<b>0.901</b>	0.4986	<b>0.4847</b>

**Table 6** – Results of precision, recall, f1-score and ROC(AUC) for classification tasks with entire dataset, with and without SMOTE, using the deep learning based strategy.

For the learning ensemble method, the best results were attained with the use of SMOTE during the training stages, as well as the use of the expanded dataset with feature reduced input vector SV, except for the SCC vs AK classification task. With the DL based approach, the use of SMOTE was also beneficial, with the overall classification results exceeding the ones attained with the ensemble model. Both models performed best for the classification tasks: Melanoma vs NMSC and Melanoma vs Nevi, followed by SCC vs AK. The distinction of benign and malignant neoplasms was tricky for both approaches.

The confusion matrixes respective to each top classification result are shown in Figure 5.



**Figure 5** – Confusion matrixes obtained for the defined classification tasks according to the best results found using the learning ensemble and deep learning models.

The accuracy, sensitivity and specificity values of the best classification results for learning ensemble and DL method were also calculated, as a means of comparing with the ones of other research works focused on the diagnosis of skin cancer with imaging modalities (Table 7).

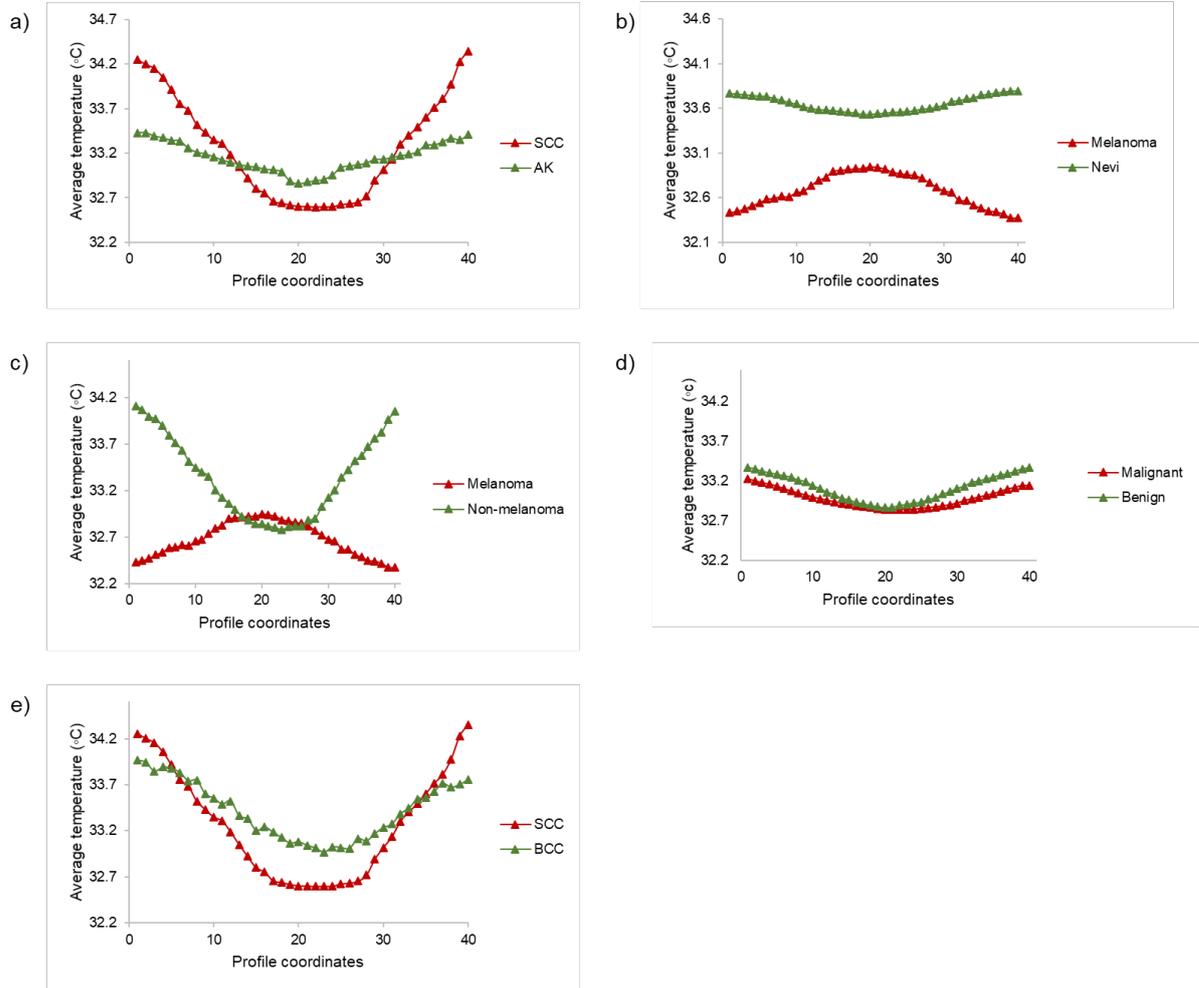
		ACC	SN	SP
Ensemble method	SCC vs AK	0.6647	0.7069	0.5902
	Melanoma vs Nevi	0.8969	0.8088	0.9444
	Mel. Vs Non-mel.	0.9730	0.8382	0.9859
	Benign vs Malignant	0.6244	0.7349	0.4242
Deep learning	SCC vs AK	0.7567	0.7907	0.6967
	Melanoma vs Nevi	0.9691	0.9412	0.9841
	Mel. Vs Non-mel.	0.9586	0.8824	0.9658
	Benign vs Malignant	0.6095	0.6281	0.5758

**Table 7** – Results of accuracy (ACC), sensitivity (SN) and specificity (SP) for top classification results of established tasks for the learning ensemble model and deep learning model.

The average  $T_{1A}$  profiles for each skin neoplasm group are presented in Figure 6, to ease the interpretation of the classification results.

Some neoplasm groups showed distinct average profiles, as melanoma and nevi lesions, Figure 6b, or melanoma and NMSC, Figure 6c. For SCC and AK, both neoplasms types displayed

a decrease in temperature when approaching lesion centre, Figure 6a. The same hypothermic tendency was verified for malignant and benign skin tumour groups, with similar average profiles Figure 6d.



**Figure 6** – Average  $T_{1A}$  of skin neoplasm types included in the classification tasks; a) SCC vs AK, b) Melanoma vs Nevi, c) Melanoma vs NMSC, d) Benign vs Malignant, and e) average  $T_{1A}$  for non-melanoma skin cancer types (BCC and SCC).

The computation times recorded for the best classification results are indicated in Table 8. There was an increase in running time when a higher number of instances was included, with the DL model surpassing the ensemble model for all tasks.

	Running time (s)	
	EM	DL
SCC vs AK	48.0432	72.4117
Melanoma vs Nevi	39.2710	45.4879
Melanoma vs NMSC	49.0215	152.4029
Benign vs Malignant	68.9087	194.4475

**Table 8** – Computation time needed to solve the proposed classification problems with top performance.

To facilitate the comparison of the attained classification results with the ones of similar methods, Table 9 summarizes the performance metrics that have been reported for related methods and the ones obtained by the proposed method.

	Authors [Ref.]	Imaging method	Best overall classifier	ACC (%)	SN (%)	SP (%)
<b>Melanoma vs Nevi</b>	Oliveira et al. [25]	Dermoscopy	optimum-path forest	91.6	87	96.2
	Alamri and Alsaeed [26]	Dermoscopy	SVM	92.6	90	96
	Kalwa et al. [23]	Dermoscopy	SVM	88	80	90
	Hosny et al. [28]	Dermoscopy	DL	97.7	97.34	97.34
	Zhang et al. [29]	Dermoscopy	CNN	91	95	92
	Magalhaes et al. [31]	IRT imaging	SVM	84.2	91.3	11.1
	<b>Proposed method</b>	<b>IRT imaging</b>	<b>DL</b>	<b>96.91</b>	<b>94.12</b>	<b>98.41</b>
<b>SCC vs AK</b>	Stringasci et al. [30]	IRT imaging	SVM	-	82.9	85.7
	<b>Proposed method</b>	<b>IRT imaging</b>	<b>DL</b>	<b>75.67</b>	<b>79.07</b>	<b>69.67</b>
<b>Benign vs Malignant</b>	Godoy et al. [32]	IRT imaging	Statistical classifier	-	95	83
	<b>Proposed method</b>	<b>IRT imaging</b>	<b>DL</b>	<b>60.95</b>	<b>62.81</b>	<b>57.58</b>

**Table 9** – Comparison of results for published literature and proposed method, for different classification tasks.

## 4 Discussion

The ML ensemble model showed good performance when dealing with the distinction of melanoma and nevi lesions and the differentiation of melanoma and NMSC (Table 4). The SCC vs AK task followed, with the classifier performing poorer when it concerned the Benign vs Malignant task. The same tendency was found for the DL model results; yet, this strategy seemed to yield higher performance metrics with a more balanced distribution of precision, recall, f1-score and ROC(AUC) (Table 5). The superior predictive ability of the deep models over ensemble ones has been previously suggested, especially when a larger dataset is used, at the cost of extra computational time (Table 8) [51–53]. To emphasize that all top results were attained when the expanded dataset was used, improving the quality of the used classifiers, in comparison to the implementation of the original dataset. The use of SMOTE also boosted the learning stage, an event previously documented in the literature [23], being equally preferred for the achievement of best precision, recall, f1-score and ROC(AUC) results by both approaches. This finding was somewhat anticipated, since the original dataset presented a significant class imbalance and could be prone to result in erroneous and biased classifications [37]. Another interesting finding was the achievement of best lesion differentiation with the feature reduced input vector. The slopes and amplitude values retrieved from the thermal gradient values proved to have higher discriminative power, reducing the computational cost.

Looking at the thermal curves included in Figure 6, it is easier to understand the classification results for the defined tasks. The similarities/dissimilarities between the average thermal curves of some skin tumours types result in resemblances/unlikeness of the attained slopes and amplitude values. Hence, the good differentiation of melanoma and nevi and melanoma and non-melanoma cases was justified by the clear differences between the thermal gradient of this types, Figure 6b and 6c. The melanoma lesions behaved like “hot spots” as opposed to the hypothermic tendency of nevi and non-melanoma tumours. This is the main reason why both classifiers performed poorly when it came to the distinction of benign and malignant cases. Because the non-melanoma tumours are included in the malignant group, its thermal curve acquires a concave shape very similar to the one of benign lesion, Figure 6d. A physiological explanation supports these claims, with benign lesions presenting far less vascularization and, therefore, decreased blood supply and heat losses, than malignant ones. Nonetheless, it is important to emphasize the practical validation of the thermal profiles documented in theoretical studies for melanoma tumours [13–15]. For SCC and BCC (NMSC), it is suspected that the increased blood flow is masked by anatomical elements, since these skin cancer types can often develop a crust on top of the main lesion, functioning as an insulating structure [3], Figure 6e. Thus, an improved image analysis approach is needed to retrieve features that better characterize and differentiate benign and malignant skin tumours. Anatomical trades can also justify the difficulty in distinguishing AK and SCC lesions. The first is considered to be a pre-cancerous lesion of SCC; thus, it displays some of the characteristics of SCC, as the “crusty” appearance [54]. Yet, in a mitigated way, making its hypothermic concavity similar, but not as accentuated, Figure 6a.

When comparing the best classification results with previous research, the developed methodology performs close to that of studies involving dermoscopy images for melanoma and nevi diagnosis. For traditional ML, the suggested ensemble showed an ACC and SP of 89.69% and 94.44%, close to [25] and [26]. Still, sensitivity is a metric to improve, as it is the most important indicator of the capability of an AI system to detect those ill. For DL (ACC=96.91%,

SN=94.12%, SP=98.41%), Hosny et al. presented a more balanced distribution of metrics (ACC=97.7%, SN=97.34%, SP=97.34%) using a transfer learning approach, while the optimization strategy of Zhang et al. [29] yield them higher sensitivity (95%) with a lower accuracy and specificity (91% and 92%) (Table 9). It is important to refer most of these studies possessed an elevated number of samples when compared to the ones included here. Focusing on IRT studies, as in this work, Stringasci et al. [30] performed a manual selection of tumorous area, collecting then statistical attributes for classification with SVM. A SN and SP value of 82.9% and 85.7% for SCC and AK distinction were found, with a balanced data set composed by 35 images for each type. Both values surpass the numbers reached by our work (SN=79.07%, SP=69.67%) (Table 9), hinting the need of a different approach to improve the encountered metrics. Godoy et al. [32] found positive results for the distinction of benign and malignant lesions (SN=95, SP=83%), as oppose to the ones reported here (Table 9). Still, the metrics reported were achieved through the selection of a temperature threshold based on temperature differences of lesion and surrounding skin, during thermal recovery after lesion cooling. In other words, a dynamic approach of IRT imaging was used, different from the one presented here, i.e., steady state (no stimulus before image acquisition). The lesions were also assessed as a hole, being no training/test set division. Nonetheless, the approach of the authors [32] is of extreme value and could indicate that dynamic IRT imaging is the way forward when distinction of benign and malignant lesions is intended.

Despite the attainment of promising results, our study presents methodological limitations that could be hampering the metrics achieved for some classification tasks. The tumorous area selected during image processing was set to a 40 by 40 pixels in dimension. This size was chosen considering the present dataset, being sufficient to include the lesions in its entirety. However, this is not completely guaranteed for other skin neoplasms, as the use of a temperature curve with a fixed direction and length of 40 values influences the amount of lesion and surrounding healthy tissue that is assessed. The present data set is also considered limited, as there is a clear class imbalance between the different skin cancer types. It is also clear that the comparison of results attained with dermoscopy images and IRT images is not ideal,

Hence, for future work, it is indicated the expansion of the data set with the assessment of different skin tumour types. Ultimately, the same number of lesions should be addressed for each category, allowing the performance of training and test stages without the need of artificial samples generated by the original cases. It is also suggested an image processing strategy that allows the selection of a preferred temperature curve by the user, as per example, a line with any direction that passes over the two points on the periphery of the lesion that are furthest away. Thus, guaranteeing a more extensive analysis. The implementation of IRT imaging involving a thermal stress before image collection might be an interesting option, in the hopes of enhancing tumour contour and metabolic and vascular capability. The detection of different skin tumour types could benefit from the inclusion of data regarding anatomical traits in the classification input vector, as asymmetry, since that type of information is not perceptible with IRT imaging and could aid in the differentiation of benign and malignant lesions. So, it is suggested the building of a database and an image analysis strategy with IRT and dermoscopy images to extract and combine physiological and anatomical data and ultimately improve the predictive performance of the proposed models. The constructed database would also facilitate scientific advances in the area of IRT imaging for skin cancer assess, as it would allow the comparison of results of different image analysis and classification methods applied to the same dataset, as it is usually the case with dermoscopy studies.

## 5 Conclusion

The learning ensemble model and a deep learning network were implemented and assessed to differentiate skin tumour types, using thermal values as input features. Promising classification results were found for the detection of melanoma lesions among nevi and non-melanoma skin cancer cases, showing the potential of IRT parameters for its implementation in a decision-support system for skin cancer diagnosis. Though, the differentiation of benign and malignant classes has fallen short.

Therefore, as to future work, it is suggested the use of input parameters collected from IRT and dermoscopy images, to provide complementary information for the classification model and facilitate and improve its accuracy.

**Acknowledgments:** The authors appreciate the availability of the patients at “Clínica da Pele” (IPO-Porto) that generously agreed to take part in this study.

## 6 REFERENCES

- [1] A.A. Romanovsky, Skin temperature: its role in thermoregulation, *Acta Physiol.* 210 (2014) 498–507. <https://doi.org/10.1111/apha.12231>.
- [2] R.A. Cairns, I.S. Harris, T.W. Mak, Regulation of cancer cell metabolism, *Nat. Rev. Cancer.* 11 (2011) 85–95. <https://doi.org/10.1038/nrc2981>.
- [3] C. Ricotti, N. Bouzari, A. Agadi, C.J. Cockerell, Malignant Skin Neoplasms, *Med. Clin. North Am.* 93 (2009) 1241–1264. <https://doi.org/10.1016/j.mcna.2009.08.011>.
- [4] S.M. Mahon, Skin cancer prevention: Education and public health issues, *Semin. Oncol. Nurs.* 19 (2003) 52–61. <https://doi.org/10.1053/sonu.2003.50005>.
- [5] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA. Cancer J. Clin.* 68 (2018) 394–424. <https://doi.org/10.3322/caac.21492>.
- [6] J. Kato, K. Horimoto, S. Sato, T. Minowa, H. Uhara, Dermoscopy of Melanoma and Non-melanoma Skin Cancers, *Front. Med.* 6 (2019) 1–7. <https://doi.org/10.3389/fmed.2019.00180>.
- [7] C. Conforti, R. Giuffrida, R. Vezzoni, F.S.S. Resende, N. Meo, I. Zalaudek, Dermoscopy and the experienced clinicians, *Int. J. Dermatol.* 59 (2020) 16–22. <https://doi.org/10.1111/ijd.14512>.
- [8] C. Fink, H.A. Haenssle, Non-invasive tools for the diagnosis of cutaneous melanoma, *Ski. Res. Technol.* (2017) 1–11. <https://doi.org/10.1111/srt.12350>.
- [9] S. U, P.T. K., S. K, Computer aided diagnosis of obesity based on thermal imaging using various convolutional neural networks, *Biomed. Signal Process. Control.* 63 (2021) 102233. <https://doi.org/10.1016/j.bspc.2020.102233>.
- [10] A. Di Carlo, Thermography and the Possibilities for Its Applications in Clinical and Experimental Dermatology, *Clin. Dermatol.* 13 (1995) 329–336.
- [11] M. Pirtini Çetingül, C. Herman, Quantification of the thermal signature of a melanoma lesion, *Int. J. Therm. Sci.* 50 (2011) 421–431. <https://doi.org/10.1016/j.ijthermalsci.2010.10.019>.
- [12] S. Hossain, F.A. Mohammadi, M. Abdelaal, Localization and parameter estimation of tumor by thermography, in: *Can. Conf. Electr. Comput. Eng.*, 2014. <https://doi.org/10.1109/CCECE.2014.6901120>.
- [13] A. Bhowmik, R. Repaka, R. Mulaveesala, S.C. Mishra, Suitability of frequency modulated thermal wave imaging for skin cancer detection-A theoretical prediction, *J. Therm. Biol.* 51 (2015) 65–82. <https://doi.org/10.1016/j.jtherbio.2015.03.007>.
- [14] M. Bonmarin, F.A. Le Gal, Lock-in thermal imaging for the early-stage detection of cutaneous melanoma: A feasibility study, *Comput. Biol. Med.* 47 (2014) 36–43. <https://doi.org/10.1016/j.combiomed.2014.01.008>.
- [15] E. Agyingi, T. Wiandt, S. Maggelakis, A Quantitative Model of Cutaneous Melanoma Diagnosis Using Thermography, in: *Math. Comput. Approaches Adv. Mod. Sci. Eng.*, Springer International Publishing, Cham, 2016: pp. 167–175. [https://doi.org/10.1007/978-3-319-30379-6\\_16](https://doi.org/10.1007/978-3-319-30379-6_16).

- [16] J. Iljaž, L.C. Wrobel, T. Gomboc, M. Hriberšek, J. Marn, Solving inverse bioheat problems of skin tumour identification by dynamic thermography, *Inverse Probl.* 36 (2020). <https://doi.org/10.1088/1361-6420/ab2923>.
- [17] C. Laurino, B. Palmieri, Wide instrumental screening in monitoring early melanoma, *Eur. J. Oncol.* 20 (2015) 41–52.
- [18] F.M. Solivetti, F. Desiderio, A. Guerrisi, A. Bonadies, C.L. Maini, S. Di Filippo, V. D’Orazi, I. Sperduti, A. Di Carlo, HF ultrasound vs PET-CT and telethermography in the diagnosis of In-transit metastases from melanoma: a prospective study and review of the literature, *J. Exp. Clin. Cancer Res.* 33 (2014) 96. <https://doi.org/10.1186/s13046-014-0096-3>.
- [19] M.D. Stringasci, L.T. Moriyama, A.G. Salvio, V.S. Bagnato, C. Kurachi, Thermographic diagnostics to discriminate skin lesions: a clinical study, in: C. Kurachi, K. Svanberg, B.J. Tromberg, V.S. Bagnato (Eds.), *Biophotonics South Am. Proc. SPIE Proc. SPIE*, 2015. <https://doi.org/10.1117/12.2180967>.
- [20] A. Di Carlo, F. Elia, F. Desiderio, C. Catricalà, F.M. Solivetti, L. Laino, Can video thermography improve differential diagnosis and therapy between basal cell carcinoma and actinic keratosis?, *Dermatol. Ther.* 27 (2014) 290–297. <https://doi.org/10.1111/dth.12141>.
- [21] P.M.M. Pereira, R. Fonseca-Pinto, R.P. Paiva, P.A.A. Assuncao, L.M.N. Tavora, L.A. Thomaz, S.M.M. Faria, Skin lesion classification enhancement using border-line features – The melanoma vs nevus problem, *Biomed. Signal Process. Control.* 57 (2020) 101765. <https://doi.org/10.1016/j.bspc.2019.101765>.
- [22] S. Pathan, K.G. Prabhu, P.C. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review, *Biomed. Signal Process. Control.* 39 (2018) 237–262. <https://doi.org/10.1016/j.bspc.2017.07.010>.
- [23] U. Kalwa, C. Legner, T. Kong, S. Pandey, Skin cancer diagnostics with an all-inclusive smartphone application, *Symmetry (Basel)*. 11 (2019). <https://doi.org/10.3390/sym11060790>.
- [24] M.A. Albahar, Skin Lesion Classification Using Convolutional Neural Network with Novel Regularizer, *IEEE Access.* 7 (2019) 38306–38313. <https://doi.org/10.1109/ACCESS.2019.2906241>.
- [25] R.B. Oliveira, A.S. Pereira, J.M.R.S. Tavares, Skin lesion computational diagnosis of dermoscopic images: Ensemble models based on input feature manipulation, *Comput. Methods Programs Biomed.* 149 (2017) 43–53. <https://doi.org/10.1016/j.cmpb.2017.07.009>.
- [26] A. Alamri, D. Alsaeed, On the development of a skin cancer computer aided diagnosis system using support vector machine, *Biosci. Biotechnol. Res. Commun.* 12 (2019) 297–308. <https://doi.org/10.21786/bbrc/12.2/12>.
- [27] C. Magalhaes, J. Mendes, R. Vardasca, The role of AI classifiers in skin cancer images, *Ski. Res. Technol.* 25 (2019) 750–757. <https://doi.org/10.1111/srt.12713>.
- [28] K.M. Hosny, M.A. Kassem, M.M. Foad, Classification of skin lesions using transfer learning and augmentation with Alex-net, *PLoS One.* 14 (2019) 1–17. <https://doi.org/10.1371/journal.pone.0217293>.
- [29] L. Zhang, H.J. Gao, J. Zhang, B. Badami, Optimization of the Convolutional Neural

Networks for Automatic Detection of Skin Cancer, *Open Med.* 15 (2020) 27–37.  
<https://doi.org/10.1515/med-2020-0006>.

- [30] M.D. Stringasci, A.G. Salvio, D. Sbrissa Neto, J.D. Vollet-Filho, V.S. Bagnato, C. Kurachi, Discrimination of benign- versus -malignant skin lesions by thermographic images using support vector machine classifier, *J. Appl. Phys.* 124 (2018) 044701.  
<https://doi.org/10.1063/1.5036640>.
- [31] C. Magalhaes, R. Vardasca, M. Rebelo, R. Valenca-Filipe, M. Ribeiro, J. Mendes, Distinguishing melanocytic nevi from melanomas using static and dynamic infrared thermal imaging, *J. Eur. Acad. Dermatology Venereol.* 33 (2019).
- [32] S.E. Godoy, D.A. Ramirez, S.A. Myers, G. von Winckel, S. Krishna, M. Berwick, R.S. Padilla, P. Sen, S. Krishna, Dynamic infrared imaging for skin cancer screening, *Infrared Phys. Technol.* 70 (2015) 147–152. <https://doi.org/10.1016/j.infrared.2014.09.017>.
- [33] S.E. Godoy, M.M. Hayat, D.A. Ramirez, S.A. Myers, R.S. Padilla, S. Krishna, Detection theory for accurate and non-invasive skin cancer diagnosis using dynamic thermal imaging, *Biomed. Opt. Express.* 8 (2017) 2301. <https://doi.org/10.1364/boe.8.002301>.
- [34] M. Goyal, T. Knackstedt, S. Yan, S. Hassanpour, Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities, *Comput. Biol. Med.* 127 (2020) 104065.  
<https://doi.org/10.1016/j.combiomed.2020.104065>.
- [35] K. Ammer, The Glamorgan Protocol for recording and evaluation of thermal images of the human body, *Thermol. Int.* 18 (2008) 125–129.
- [36] E. Ring, K. Ammer, The technique of infrared imaging in medicine, in: *Infrared Imaging*, IOP Publishing, 2015: pp. 1–10. <https://doi.org/10.1088/978-0-7503-1143-4ch1>.
- [37] N. V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, *J. Artif. Intell. Res.* 16 (2002) 321–357.  
<https://doi.org/10.1613/jair.953>.
- [38] R. Polikar, Ensemble Learning, in: Y. Zhang, Cha, Ma (Ed.), *Ensemble Mach. Learn.*, 2012: pp. 1–34.
- [39] V. Kotu, B. Deshpande, Data Mining Process, in: *Predict. Anal. Data Min.*, Elsevier, 2015: pp. 17–36. <https://doi.org/10.1016/B978-0-12-801460-8.00002-1>.
- [40] A. Murugan, S.A.H. Nair, K.P.S. Kumar, Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers, *J. Med. Syst.* 43 (2019). <https://doi.org/10.1007/s10916-019-1400-8>.
- [41] A. Cutler, R.D. Cutler, J.R. Stevens, Random Forest, in: Y. Zhang, Cha, Ma (Ed.), *Ensemble Mach. Learn.*, 2012: pp. 157–175.
- [42] S. Raschka, V. Mirjalili, A Tour of Machine Learning Classifiers Using scikit-learn, in: J. Malysiak, S. Jain, J. Lovell, C. Nelson, S. D’silva, R. Atitkar (Eds.), *Python Mach. Learn.*, Third, Packt Publishing, 2019: pp. 111–181.
- [43] E. La Torre, B. Caputo, T. Tommasi, Learning methods for melanoma recognition, *Int. J. Imaging Syst. Technol.* 20 (2010) 316–322. <https://doi.org/10.1002/ima.20261>.
- [44] J. Miguel, [Supervised Learning] SVM – Support Vector Machine explained with examples, *Lipman’s Artif. Intell. Dir.* (2015). <http://laid.delanover.com/supervised-learning-svm-support-vector-machine-explained-with-examples/> (accessed January 8,

2018).

- [45] S. Raschka, V. Mirjalili, Implementing a Multilayer Artificial Neural Network from Scratch, in: J. Malysiak, S. Jain, J. Lovell, C. Nelson, S. D'silva, R. Atitkar (Eds.), Python Mach. Learn., Third, Packt Publishing, 2019: pp. 523–574.
- [46] J.C. Platt, Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, (1999).
- [47] I. Goodfellow, Y. Bengio, A. Courville, Optimization for training deep models, in: Deep Learn., 2016: pp. 267–320.
- [48] S. Raschka, V. Mirjalili, Loss functions for classification, in: J. Malysiak, S. Jain, J. Lovell, C. Nelson, S. D'silva, R. Atitkar (Eds.), Python Mach. Learn., Third, Packt Publishing, 2019: pp. 724–727.
- [49] S. Raschka, V. Mirjalili, Learning Best Practices for Model Evaluation and Hyperparameter Tuning, in: J. Malysiak, S. Jain, J. Lovell, C. Nelson, S. D'silva, R. Atitkar (Eds.), Python Mach. Learn., Third, Packt Publishing, 2019: pp. 280–321.
- [50] R.B. Oliveira, A.S. Pereira, J.M.R.S. Tavares, Computational diagnosis of skin lesions from dermoscopic images using combined features, *Neural Comput. Appl.* 31 (2019) 6091–6111. <https://doi.org/10.1007/s00521-018-3439-8>.
- [51] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, J. Xiao, LSUN: Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop, (2016).
- [52] C.-Y. Hung, W.-C. Chen, P.-T. Lai, C.-H. Lin, C.-C. Lee, Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database, in: 2017 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE, 2017: pp. 3110–3113. <https://doi.org/10.1109/EMBC.2017.8037515>.
- [53] D.T. Bui, P. Tsangaratos, V.-T. Nguyen, N. Van Liem, P.T. Trinh, Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment, *CATENA*. 188 (2020) 104426. <https://doi.org/10.1016/j.catena.2019.104426>.
- [54] R.A. Schwartz, Actinic keratosis, in: *Ski. Cancer Facts Stat.*, 2nd ed., Blackwell Publishing, Newark, New Jersey, 2008: pp. 5–15.