

# Convolutional Neural Networks on Assembling Classification Models to Detect Melanoma Skin Cancer

<https://doi.org/10.3991/ijoe.v18i14.34435>

Hugo Vega-Huerta<sup>1</sup>(✉), Renzo Villanueva-Alarcón<sup>1</sup>, David Mauricio<sup>1</sup>,  
Juan Gamarra-Moreno<sup>1</sup>, Hugo D. Calderon-Vilca<sup>1</sup>, Diego Rodriguez<sup>2</sup>, Ciro Rodriguez<sup>1</sup>

<sup>1</sup>Department of Computer Science, Faculty of System Engineering, Universidad Nacional  
Mayor de San Marcos (UNMSM), Lima, Perú

<sup>2</sup>Medicine Program, Faculty of Health Sciences, Universidad Peruana de Ciencias Aplicadas  
(UPC), Lima, Perú  
hvegah@unmsm.edu.pe

**Abstract**—In 2020, there were more than 1.2 million new skin cancer diagnoses, and melanoma was the most recurrent type of cancer. On the other hand, melanoma is the least common but most serious form of skin cancer affecting both men and women. This work aims to assemble classification models to detect a case of melanoma with high accuracy based on a Convolutional Neural Networks system. The methodology considers training 21 models for image classification, with the best assembly performance of EfficientNet and VGG-19 architectures, the data augmentation technique was used to the images to improve its performance. The results show 92.85% of accuracy, 71.50% of sensitivity, and 94.89% of specificity, with an improvement of 0.06% in accuracy and specificity. The assembly of the classification models achieved higher accuracy in melanoma skin cancer image classification.

**Keywords**—melanoma, skin cancer, convolutional neural networks, classification model, deep learning

## 1 Introduction

A prevalent type of cancer is skin cancer; within this group is melanoma, which is the least common of all cases and yet is the most dangerous, as it causes 75% of most skin cancer deaths by spreading to other regions of the body, affecting treatment efficacy and patient survival, considering the research of [1] and what is happening in the worldwide, where 132,000 people are diagnosed with melanoma-type skin cancer [2,3], according to statistical data from the World Health Organization [4], with exposure to ultraviolet rays being one of the leading causes of this. According to [5], the dermoscopy technique was developed to improve the diagnostic yield of skin cancer; in it, the dermatologist uses various noninvasive techniques of image analysis, such as focal microscopy, optical coherence tomography, and pattern analysis that allow him to classify and detect skin lesions according to their type [6–8]. This article presents the procedure used as a solution to classify images of patient skin lesions and identify

which ones represent melanoma. In addition, contextual image information (metadata) is used to evaluate whether its inclusion improves image classification performance.

The first step in diagnosing a possible malignant lesion or melanoma of the suspicious skin region is a visual inspection by the dermatologist [9]. However, automatic identification of skin cancer by dermatologists remains complicated as it faces many challenges [10]. The incidence of various skin lesions has a high degree of visual similarity, and variations in skin conditions, such as color and natural hair, can make detection of these lesions difficult [11–13]. Diagnosis from visual examination of images by a dermatologist range from 75% to 84% accuracy in detecting melanomas, so it is essential to include computer-assisted platforms that can increase the accurate detection of diagnosis of skin lesions using imaging [14–16].

The objective is to assemble classification models using image analysis to accurately detect skin lesions and detect if it's a melanoma problem along with patient input.

## **2 Background**

Lately, deep learning models, specifically convolutional neural networks, evidence medical image analysis results that allow an approach to detect skin cancer problems accurately [17–19]. Therefore, this type of neural network has been widely used to overcome this image classification problem [20–22]. Table 1 lists some of the articles related to the skin cancer classification problem, highlighting the use of the following convolutional neural network architectures: MobileNet [23], DenseNet [24], ResNet [25], Inception [26], VGG [27], PNASNet [28], AlexNet [29], Xception [30], EfficientNet [31], as well as ensemble methods, the application of the “Data Augmentation” technique, K-Fold and the inclusion of MetaMetrics [32]. Inclusion of K-Fold and MetaMetrics.

**Table 1.** Summary of state of the art regarding skin cancer classification

Article	Data	Architectures	Methods Used	ACC %	PREC %	SEN %	SPC %
[33]	ISIC 016	-MobileNet -DenseNet	-Data Augmentation -Assembly	–	84.5	–	–
[34]	ISIC 017	-ResNet-50 -InceptionV3	-Assembly	–	81.6	–	–
[35]	ISIC 017	-InceptionV4	-Data Augmentation	89.0	–	55.6	97.1
[36]	ISIC 017	-AlexNet -VGG-19 -ResNet-18 -ResNet-101	-Data Augmentation -Inclusion of metadata -Assembly	87.7	–	85.0	73.3
[37]	ISIC 018	-Resnet-50	-Data Augmentation	92.1	–	–	92.5
[38]	ISIC 019	-AlexNet	-Data Augmentation -K-Fold	92.9	62.8	70.4	96.0
[39]	ISIC 019	-PNASNet -VGG-19	-Inclusion of metadata -Assembly	90.1	–	–	–
[40]	ISIC 019	-EfficientNet-B3 -EfficientNet-B4	-Data Augmentation -Inclusion of metadata -Assembly	91.5	–	66.2	95.2
[41]	ISIC 2019	-Resnet-50 -DenseNet -Xception	-Data Augmentation -Assembly	88.0	87.0	–	94.0

Notes: The table shows the main state of the art reviewed in the case study of this article. The table shows a dashboard of performance metrics of the models proposed. The metrics used are ACC (Accuracy), PREC (Precision), SEN (Sensitivity), and SPC (Specificity).

### 3 Proposal design

#### 3.1 SIIM-ISIC dataset

The present article uses the SIIM-ISIC dataset (from the year 2020) and the 2019 dataset, which consolidates lesion images from previous years [42]. The 2020 ISIC repository contains 33,126 dermatoscopic images of skin lesions classified as benign (without Melanoma) and malignant (with the presence of Melanoma). At the same time, the 2019 dataset contains 25,331 images.

Each image is associated with an individual by a unique patient identifier. According to [43], all diagnoses of melanoma-type skin cancer have been confirmed by histopathology, and diagnoses of benign skin lesions have been expertly verified by longitudinal follow-up examinations or histopathology.

The information is available in DICOM standard format, which is a special image data format commonly used in the medical field. In the data set to be analyzed, each record contains the image and the image contextual information (also known as metadata). For each record, the following information is available:

- **image:** Dermatoscopic image of the skin lesion.
- **image\_array:** Unidimensional image array (grayscale).
- **image\_name:** Identifier of the image corresponding DICOM standard.
- **patient\_id:** Unique patient identifier.
- **sex:** Sex of the patient.
- **age\_approx:** Age of the patient at image capture moment.
- **anatom\_site\_general\_challenge:** Anatomical location of the skin lesion.
- **target:** Target variable (No Melanoma:0; Melanoma:1).
- **benign\_malignant:** Qualitative indicator (No: “Benign”; Yes: “Malignant”).

### 3.2 Training and evaluation strategy

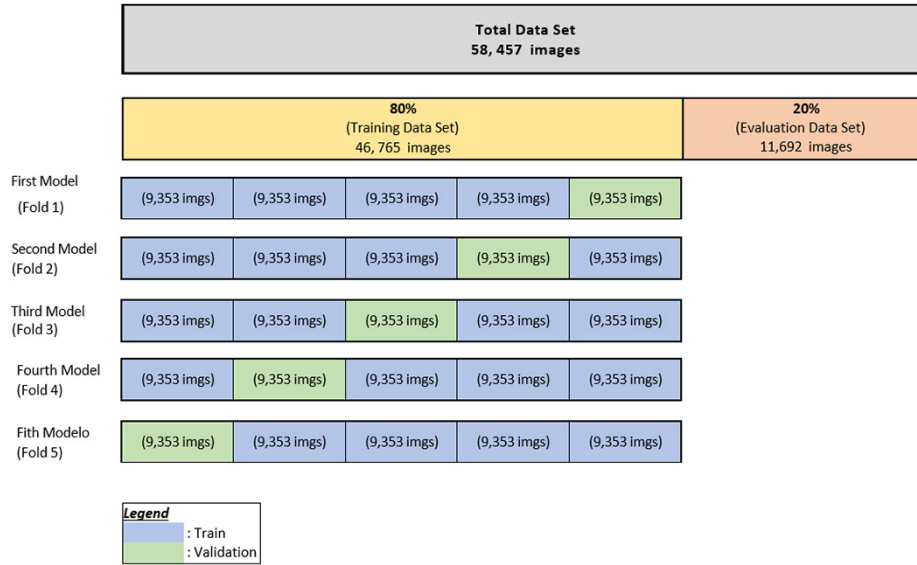
Based on the statistical principle “Pareto’s Law,” a proposed training and evaluation split of 80 and 20. This law describes the statistical phenomenon whereby a small proportion of any population contributing to a common effect [44] contributes to most of the impact. The 80/20 rule suggests that, in any given example, the few (20%) are vital and the many (80%) are trivial.

As explained in the previous paragraph and taking as a reference the article [45], was considered 80% images for the training process, leaving 20% images for the evaluation process of the trained modules of classification. Both partitions are duly represented by benign and malignant types of melanoma. Table 2 shows the proposed 80/20 split for the entire data set.

**Table 2.** Number of images per process (training and evaluation)

Process	Number of Images
Training (80%)	46, 765
Evaluation (20%)	11, 692
Total (100%)	58, 457

As a methodology of data division for training, the “K fold” cross-validation technique has been applied, where the parameter used was: K=5, i.e., 5 folds of the dataset where a convolutional neural network model is trained on each fold, which when evaluating a record has a probability of melanoma presence for each fold. Finally, the 5 probabilities are averaged with this technique to obtain a final stable probability. The reason for the application is to use the entire data set for training and, at the same time to reserve a data set for model validation; this validation data set is also used in the loss function. See Figure 1.



**Fig. 1.** Distribution of the dataset (training and evaluation) [44]

Note: Division of training and evaluation (80/20), followed by the application of K-Fold.

## 4 Classification modules

In this chapter, the development and validation of the performance of the classification modules are addressed. As their name indicates, these modules focus on the classification task through neural networks to provide a probability as output in each module. In designing these modules, the following assumptions are made:

- The objective of each module is to provide a classification through a probability of the presence of Melanoma, based on their input data.
- The image classification module receives as input data the preprocessed images in 3-dimensional array format.
- The contextual information classification module receives as input data the categorical and discrete variables in a dataframe.
- Both modules are subjected to 5-fold K-Fold cross-validation. So, each module has 5 sub-models. The final probability per module is the average of the 5 trained sub-models.

### 4.1 Image classification module

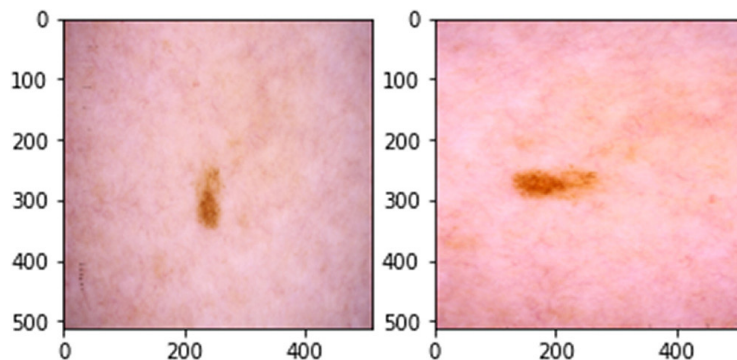
For the training of the image classification module, a model was built consisting of a 3-dimensional input layer, a base architecture, a Pooling layer, and a Dense layer. As part of the training process and to evaluate and select the best model, it was proposed to train on 7 different convolutional neural network architectures: EfficientNet B4, EfficientNet B5, EfficientNet B6, InceptionV3, ResNet50, ResNet110, VGG19. The main parameter settings are shown in Table 3.

**Table 3.** Parameter settings for the image classification module

Input Layer	(Height, Width, 3)
Pooling layer	GlobalAveragePooling2d
Dense layer	Sigmoid
Training scheme: Epochs	12
Training scheme: Batch Size	36
Loss function	Binary Cross-Entropy Loss
Learning Rate	0.001
Optimizer	Adam

*Note:* The values of the parameters used in the neural network for the classification module.

Likewise, for the present solution, according to the performance seen of the “Data Augmentation” technique in the reviewed state of the art, it is proposed to apply pre-processing functions: (saturation, contrast, brightness) and transformation functions: (rotation, zoom, flip), randomly and sequentially. The application procedure of these functions depends on the result of a random operation between 0 and 1 as shown in Figure 2; in the case of the random part of 1, the process is applied; otherwise, continue with the following.



**Fig. 2.** Comparison of the application of Data Augmentation on an example image

*Notes:* Left: Normal test image. Right: Test image with saturation at 110%, contrast at 120%, brightness at 120%, rotation at 90 degrees, zoom to 6, and mirror rotation. Own elaboration.

Considering the characteristics mentioned in Figure 2, it was proposed to validate in 3 scenarios the performance of the application of the “Data Augmentation” and “K-Fold” techniques with five folds. The following 3 scenarios were built considering by default image size of  $384 \times 384$  and EfficientNet-b6 as convolutional neural network architecture:

- First Scenario: Model without applying Data Augmentation and the 5-fold K-Fold technique.

- Second Scenario: Model with the application of Data Augmentation and the 5-fold K-Fold technique.
- Third Scenario: Model with the application of Data Augmentation and the 10-fold K-Fold technique.

Table 4 shows that of the scenarios proposed for validation, the one that had the best performance under the AUC metric with 94.83% was the Data Augmentation technique and considering 5 folds for the Cross Validation technique “K-Fold”; for this reason, the application of both methods is regarded for training and evaluation.

**Table 4.** Comparison of the performance of the 3 proposed scenarios

Scenario	Data Augmentation	Number of Folds	Area Under the Curve
1	No	5	76.84%
2	Yes	5	<b>94.83%</b>
3	Yes	10	93.62%

*Note:* This table shows the results of the proposed scenarios, measured in the AUC metric.

#### 4.2 Metadata classification module

For the training process, we proposed the construction of the neural network model, a neural network consisting of 3 layers is proposed, considering the following variables:

- Independent variables:  
“mean\_color”, “sex”, “age\_approx”, “anatom\_site\_general”
- Depending variables:  
“Target”

A creation of a new variable was proposed, “mean\_color,” which contains the average contrast of the image. This is obtained by averaging all the array values, thus getting a value between 0 to 255, representing the overall contrast of the image.

The validation process consists of comparing the performance under the AUC metric with the final prediction of each of the following scenarios; the final prediction is the average of the 5 sub-models, changing the values of the following parameters:

- Batch size: 20,40,80,160
- Epochs: 20,40,80,160
- Learning Rate: 0.0001, 0.001, 0.01
- Activation: Softmax, Relu, Sigmoid
- Optimizer: SGD, RMSprop, Adam
- Neurons: 5,10,15,20

Once the validation process was completed, it was concluded that the best parameter configuration among the combinations is the one shown in Table 5, obtaining a performance value of 82.23% in AUC, evaluated in validation data.

**Table 5.** Parameter configuration for the metadata classification module

Training scheme: Epochs	20
Training scheme: Batch Size	40
Activation	Softmax
Learning Rate	0.001
Optimizer	SGD
Neurons	5

*Note:* This table shows the parameter values used in the neural network for the classification module.

## 5 Results

This section evaluates each of the classification modules developed and the union of the predictions of both modules. As described in the design section, 11,692 images were used, representing 20% of the total dataset. Likewise, for the evaluation process of each of the two modules (images and metadata), the performance of the predictions of its 5 submodels (Folds) and the final model was validated.

### 5.1 Evaluation of the final proposal

Table 6 shows the performance of the 21 proposed models (7 convolutional neural network architectures iterated by 3 image sizes). Note that the average gives each final model for each evaluation of the 5 predictions of the submodels (5 folds for each model); also, the Data Augmentation technique was applied for each model.

**Definition of validation metrics.** Metrics of accuracy, sensitivity, and specificity are the most essential variables when designing diagnostic tests, these variables help to determine how reliable are the tests and the results obtained.

Accuracy is the fraction of predictions that our assembling model got right, as shown in equation (1).

$$ACC = \frac{\text{Number of classifications a model correctly predicts}}{\text{Total number of predictions made}} \quad (1)$$

Sensitivity is the percentage of true positives, it evaluates the ability of a model to predict the true positives of each available category, it is also known as the True Positive Rate as shown in equation (2).

$$SEN = \frac{\text{Number of True positive test}}{\text{Total number of individuals with the disease in a population}} \quad (2)$$

Specificity is the percentage of true negatives, it evaluates the ability of a model to predict the true negatives of each available category, it is also known as the True Positive Rate as shown in equation (3).



$$SPC = \frac{\text{Number of True negative test}}{\text{Total number of healthy individuals in a population}} \quad (3)$$

**Table 6.** Model performance to evaluate the image classification module [1]

Model	Architecture	Image Size	AUC	ACC	PREC	SEN	SPC	F SCORE
1	EfficientNet B4	256 × 256	92.38%	91.35%	50.41%	60.63%	94.29%	55.05%
2		384 × 384	92.93%	91.91%	52.94%	66.11%	94.38%	58.80%
3		512 × 512	93.06%	92.37%	55.13%	67.87%	94.71%	60.84%
4	EfficientNet B5	256 × 256	92.01%	90.14%	45.59%	66.80%	92.37%	54.19%
5		384 × 384	92.98%	91.50%	51.00%	67.68%	93.78%	58.16%
6		512 × 512	92.16%	89.54%	42.99%	60.72%	92.31%	50.30%
7	EfficientNet B6	256 × 256	92.67%	90.80%	48.11%	68.66%	92.92%	56.78%
8		384 × 384	93.16%	91.48%	50.88%	70.52%	93.49%	59.11%
9		512 × 512	93.09%	91.94%	52.99%	68.46%	94.19%	59.74%
10	INCEPTIONV3	256 × 256	91.25%	90.25%	46.00%	67.09%	92.47%	54.58%
11		384 × 384	90.00%	90.54%	46.48%	54.95%	93.95%	50.36%
12		512 × 512	91.88%	91.70%	52.06%	63.08%	94.44%	57.04%
13	RESNET101	256 × 256	87.98%	89.96%	39.09%	26.84%	96.00%	31.82%
14		384 × 384	88.05%	90.33%	45.81%	58.96%	93.33%	51.56%
15		512 × 512	89.27%	89.32%	41.59%	55.24%	92.58%	47.45%
16	RESNET50	256 × 256	88.59%	89.78%	15.75%	3.92%	97.99%	6.27%
17		384 × 384	88.97%	89.89%	43.63%	53.97%	93.33%	48.25%
18		512 × 512	89.30%	88.08%	38.02%	57.79%	90.98%	45.86%
19	VGG19	256 × 256	84.57%	79.67%	25.42%	68.66%	80.72%	37.10%
20		384 × 384	85.20%	80.74%	26.55%	68.27%	81.93%	38.23%
21		512 × 512	84.34%	81.29%	26.57%	64.74%	82.88%	37.67%

Note: The table shows a dashboard of performance metrics of the proposed models of the imaging module.

An assembly of the 3 best and the 5 best rated models according to their Sensitivity is evaluated to achieve better performance. According to Table 7, it is observed that the 5 best models have been using the EfficientNetB6 architecture in all image sizes and the VGG19 architecture in the size of 256 × 256 and 384 × 384.

**Table 7.** Performance of the 5 best models [1]

Top Sensitivity	Architecture	Image Size	AUC	ACC	PREC	SEN	SPC	F SCORE
1	EfficientNet B6	384 × 384	93.16%	91.48%	50.88%	70.52%	93.49%	59.11%
2	EfficientNet B6	256 × 256	92.67%	90.80%	48.11%	68.66%	92.92%	56.58%
3	VGG19	256 × 256	84.57%	79.67%	25.42%	68.66%	80.72%	37.10%
4	EfficientNet B6	512 × 512	93.09%	91.94%	52.99%	68.46%	94.19%	59.74%
5	VGG19	384 × 384	85.20%	80.74%	26.55%	68.27%	81.93%	38.23%

*Note:* The table shows a dashboard of performance metrics of the 5 best-proposed models of the imaging module according to their Sensitivity.

The next step is the application of the assembly methods with the 5 best models, using Soft Voting and Hard Voting techniques. For this purpose, it was proposed to evaluate 4 scenarios:

- First Scenario: The 3 best models according to Sensitivity, applying Soft Voting.
- Second Scenario: The 3 best models according to Sensitivity, applying Hard Voting.
- Third Scenario: The 5 best models according to Sensitivity, applying Soft Voting.
- Fourth Scenario: The 5 best models according to Sensitivity, applying Hard Voting.

Table 8 shows that the Hard Voting technique improves performance considering the 5 best models.

**Table 8.** Performance of imaging module assemblies [1]

Scenario	VOTING	P R E C	A C C	S P C	S E N	F SCORE
1	SOFT – TOP 3	57.92%	92.95%	95.10%	70.52%	63.60%
2	HARD – TOP 3	57.36%	92.85%	94.98%	70.62%	63.30%
3	SOFT – TOP 5	56.97%	92.78%	94.88%	70.81%	63.14%
4	HARD – TOP 5	<b>56.94%</b>	<b>92.79%</b>	<b>94.83%</b>	<b>71.50%</b>	<b>63.40%</b>

*Note:* The table shows a performance metrics dashboard of the Soft Voting and Hard Voting assemblies considering the 3 and the 5 best models in Sensitivity evaluated.

Once the evaluation of the classification module was finished, having as a winner the Hard Voting Model Assembly with the 5 best models in Sensitivity; it was evaluated how much the performance of the image module improved with the incorporation of the predictions of the metadata module. For this purpose, it was proposed to evaluate the assembly in 5 scenarios, where a weight is assigned to each module in each evaluation:

- First Scenario: 100% of the prediction of the Images module (the prediction of the Metadata module is not considered).
- Second Scenario: 90% of the prediction of the Images module and 10% of the prediction of the Metadata module.

- Third Scenario: 80% of the prediction of the Images module and 20% of the prediction of the Metadata module.
- Fourth Scenario: 70% of the prediction of the Images module and 30% of the prediction of the Metadata module.
- Fifth Scenario: 50% of the prediction of the Images module and 50% of the prediction of the Metadata module.

**Table 9.** Performance of the assemblies of the images module with the metadata module

Scenario	Image Module Weight	Metadata Module Weight	ACC	PREC	SEN	SPC	F SCORE
1	1	0	92.79%	56.94%	71.50%	94.83%	63.40%
2	0.9	0.1	<b>92.85%</b>	<b>57.25%</b>	<b>71.50%</b>	<b>94.89%</b>	<b>63.59%</b>
3	0.8	0.2	94.23%	67.47%	65.43%	96.98%	66.43%
4	0.7	0.3	94.65%	73.97%	59.84%	97.99%	66.16%
5	0.5	0.5	94.70%	82.50%	49.85%	98.99%	62.15%

*Note:* Comparison of the performance of the weighted assemblies of the images module with the metadata module.

As shown in Table 9, the best performance is observed in the weighted assembly considering 90% of the Images module and 10% of the Metadata module, with 92.85% Accuracy, 71.50% Sensitivity, 94.89% Specificity. It shows an improvement of 0.06% in Accuracy and 0.06% in Specificity compared to the performance of only considering the predictions of the Images module, so it is decided to consider as the final solution the Assembly of the two modules with the explained weighting. [46×] also presents important results about of predict university student’s dropout based on accuracy and sensitivity metrics.

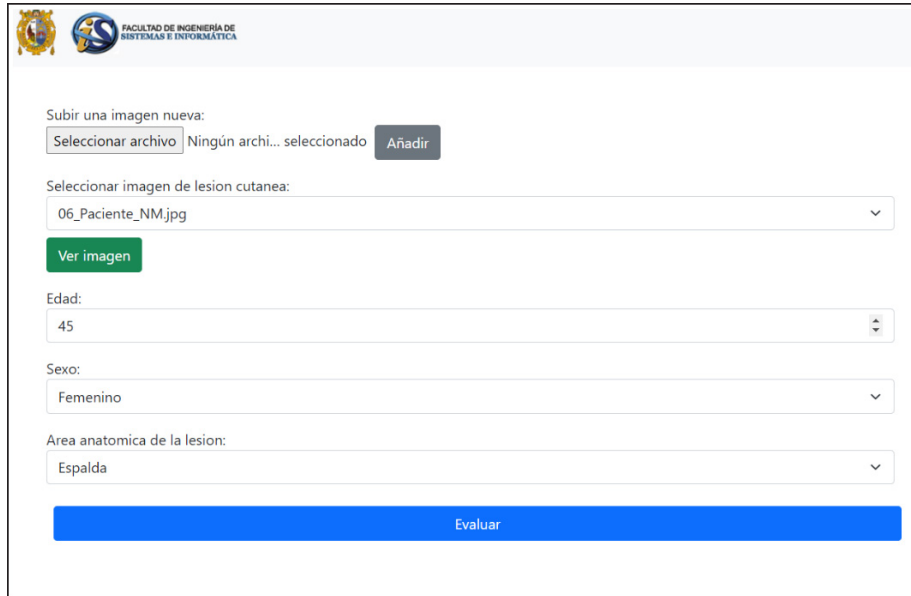
## 5.2 System interfaces

This section shows the web interface of the solution, explaining each of the functionalities. As preliminary considerations, the implementation is developed in Python language using the free software Google Colab. The 5 submodels of each module and the static files (HTML) and images are hosted in a Google Drive repository. Screenshots of the developed web system are shown below.

Figure 3 shows Variables and image form: The interface includes a form for entering the variables: age, sex, and anatomical area of the lesion. The cast also includes a list of all the images of skin lesions in the repository. There is also an option to upload a new photo to the image repository.

Figure 4 presents screenshot of the View Image Option, below the list of images to select, the “View Image” button is included where the variables: Height, Width, and Average Contrast are displayed, as well as the selected image.

Figure 5 shows screenshot of the View Prediction option, once all the variable fields have been filled in and the skin lesion image has been selected, clicking on the “Evaluate” button processes the image and metadata, evaluating the record for each module.



The screenshot shows a web interface for a melanoma detection application. At the top left, there are logos for the 'FACULTAD DE INGENIERIA DE SISTEMAS E INFORMÁTICA'. The main form contains the following elements:

- Subir una imagen nueva:** A button labeled 'Seleccionar archivo' followed by the text 'Ningún archi... seleccionado' and an 'Añadir' button.
- Seleccionar imagen de lesion cutanea:** A dropdown menu with '06\_Paciente\_NM.jpg' selected and a 'Ver imagen' button below it.
- Edad:** A text input field containing the number '45'.
- Sexo:** A dropdown menu with 'Femenino' selected.
- Area anatomica de la lesion:** A dropdown menu with 'Espalda' selected.
- Evaluar:** A large blue button at the bottom of the form.

Fig. 3. Screenshot of the variables and image form

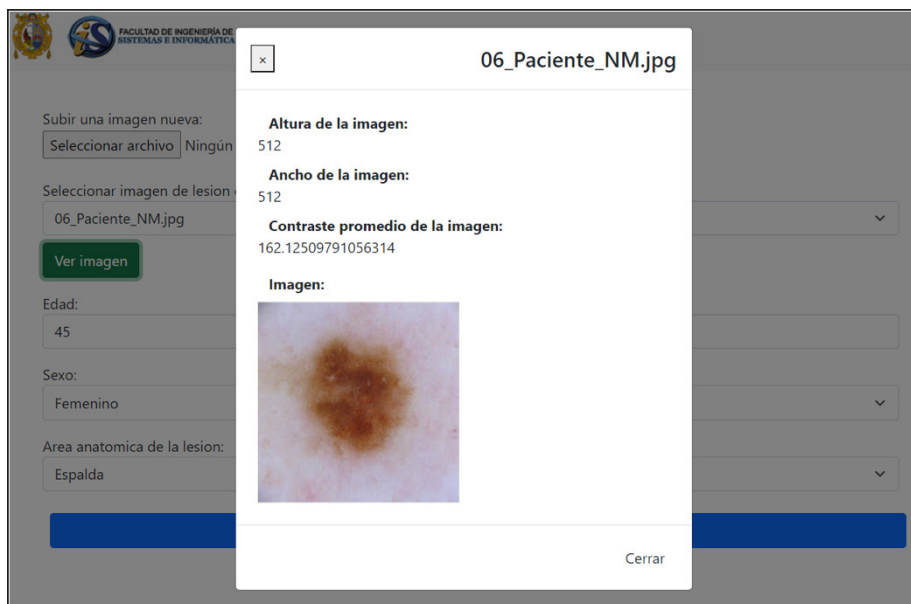


Fig. 4. Screenshot of the view image option

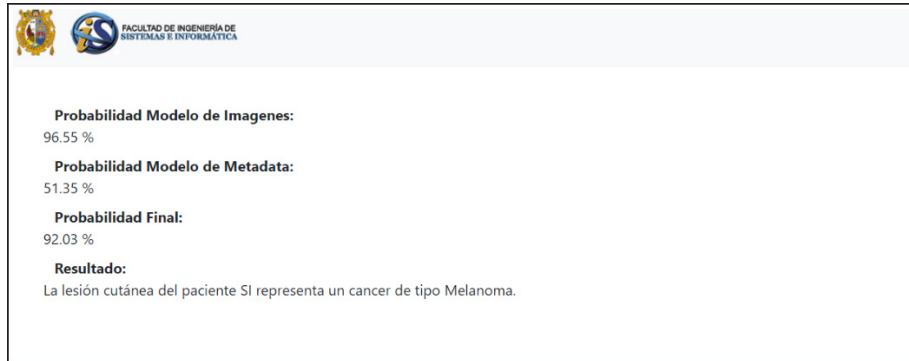


Fig. 5. Screenshot of the view prediction option

## 6 Discussion and conclusions

Comparing the results of previous investigations with the results of the study, we appreciate that in Table 1, [35] presents us with results ACC 87.7%, SEN 85.0%, SPC 73.3%, while the values reached in this work, shown in scenario 2 of Table 9, are ACC 92.85, SEN 71.5%, SPC 94.89%. However, there is a lower SEN of 14.5%, a higher ACC of 5.2% and a higher SPC of 11.6% have been achieved. The value obtained for the SEN of 71.5%, despite being lower than that achieved by [35], is higher than the 51.6% achieved by [34], the 70.4% achieved by [39], and the 66.2% achieved [39], for which we can affirm that the results achieved in this investigation have been superior in some indicators achieved in other similar studies on the same subject.

Likewise, we highlight that in the research [47] a hybrid model of CNN and SVM is proposed to classify carcinoma, in their training and testing they used the HAM10000 dataset with 19,267 dermoscopic images comparing the proposed model to detect carcinoma with assembly model also. We use several CNN models, the dataset we use is from SIIM-ISIC.

The experiments performed are consistent, since they were performed under the same conditions taking as a basis the initial value of  $k = 5$  and comparing our proposed assembled model with the other models evaluated under the same conditions. [48] suggesting experiments with  $k = 10$  folds validate the consideration of the third scenario of the assembling model with the application of Data Augmentation and the  $k = 10$  folds. The experiments performed are also congruent with [49] to consider  $k = 10$  to obtain the best error estimate, which is also congruent with the third scenario. The results show that not necessarily the best results are given with the highest value of  $k$ , in this sense the results obtained are congruent with [50] in his article on the automated classification of liver disorders using ultrasound images shows that his best result is given with  $k = 5$  and not with  $k = 10$  folds, which is also concordant with our results.

With the results, we conclude that it is possible to detect in the images with skin lesions if there is a case of Melanoma skin cancer or not by assembling classification models.

The proposed solution allows for accurately identifying cases of Melanoma in images of skin lesions, using binary classification, of the presence or absence of Melanoma.

For its realization, the “ISIC SIIM” dataset has been used, a repository composed of many images of skin lesions in DICOM format. Each of the models that make up the proposed solution was trained.

Specifically, the implementation of this research project has been designed based on 3 main modules to organize the development and achieve the proposed objectives more efficiently. These are the preprocessing module, the image classification module, and the metadata classification module.

In the preprocessing module, the use of the “Data Augmentation” technique, increasing the variability between images, was a critical factor to improve the classification model. Applying the “K-Fold” technique was necessary to use the totality of data to optimize the model in the training process with the validation data. Likewise, it is observed that with this technique, the solution generalizes better, having a stable performance in evaluation.

By adding metadata to the model, the performance improves slightly compared to only using the image module. This indicates that metadata does help models that do not take advantage of the complete information available only in the images.

## 7 References

- [1] Villanueva, R. (2021). Sistema inteligente basado en redes neuronales para la identificación de cáncer de piel de tipo melanoma en imágenes de lesiones cutáneas. Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM. [https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/17574/Villanueva\\_ar.pdf?sequence=1&isAllowed=y](https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/17574/Villanueva_ar.pdf?sequence=1&isAllowed=y)
- [2] American Cancer Society. (2021). “Estadísticas sobre el cáncer de piel tipo melanoma,” Junio 2021. [En línea]. Available: <https://www.cancer.org/es/cancer/cancer-de-piel-tipo-melanoma/acerca/estadisticas-clave.html>
- [3] Shweikeh, Lu, E. J. and M. Al-Rajab. (2021). “Detection of cancer in medical images using deep learning,” *International Journal of Online and Biomedical Engineering*, vol. 17, no. 14, pp. 164–171. <https://doi.org/10.3991/ijoe.v17i14.27349>
- [4] World Health Organization. (2021). “How common is skin cancer?”. [En línea]. Available: <https://www.who.int/uv/resources/FAQ/skincancer/en/index1.html>
- [5] Binder, M., M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, y H. Pehamberger, “Epiluminescence microscopy. A useful tool for diagnosing pigmented skin lesions for formally trained dermatologists,” *Jama Dermatology*, 1995. <https://doi.org/10.1001/archderm.1995.01690150050011>
- [6] Carli, P., E. Quercioli, S. Sestini, M. Stante, L. Ricci, G. Brunasso, and V. DE Giorgi. (2003). “Pattern analysis, not simplified algorithms, is the most reliable method for dermoscopy for melanoma diagnosis to residents in dermatology,” *Br J Dermatol*. <https://doi.org/10.1046/j.1365-2133.2003.05023.x>
- [7] Afza, F., M. A. Khan, M. Sharif, y A. Rehman. (2019). “Microscopic skin laceration segmentation and classification: A framework of statistical normal distribution and optimal feature selection,” *Microscopy Research and Technique*, vol. 82, no. 9, pp. 1471–1488. <https://doi.org/10.1002/jemt.23301>

- [8] Javed, R., M. Shafry, M. Rahim, T. Saba, y A. Rehman. (2020). “A comparative study of features selection for skin lesion detection from dermoscopic images,” *Network Modeling and Analysis in Health Informatics and Bioinformatics*, vol. 1, pp. 1–13. <https://doi.org/10.1007/s13721-019-0209-1>
- [9] Kassem, M., A. K. M. Hosny, R. Damaševičius, y M. M. Eltoukhy. (2021). “Machine learning and deep learning methods for skin lesion classification and diagnosis: A systematic review,” *Diagnostics*, vol. 11, no. 8. <https://doi.org/10.3390/diagnostics11081390>
- [10] Alpana, K. G., B. Mausumi, and M. Ravi. (2016). “Skin cancer concerns in people of color: Risk factors and prevention,” *Skin Cancer Concerns in People of Color*, pp. 5257–5264.
- [11] Fabbrocini, G., M. Triassi, M. C. Mauriello, G. Torre, M. C. Annunziata, V. a. P. F. a. D. V. De Vita, y G. Monfrecola. (2010). “Epidemiology of skin cancer: Role of some environmental factors,” *Cancers*, vol. 2, no. 4, pp. 1980–1989. <https://doi.org/10.3390/cancers2041980>
- [12] Haenssle, H., C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. Hassen, L. Thomas, A. Enk, y +L. Uhlmann. (2018). “An against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists,” *Annals of Oncology*, vol. 29. <https://doi.org/10.1093/annonc/mdy166>
- [13] Ferrante di Ruffano, L., Y. Takwoingi, J. a. S. N. Dinnes, S. Bayliss, C. Davenport, R. Matin, K. Godfrey, C. O’Sullivan, A. Gulati, S. a. D. A. Chan, S. O’Connell, and M. Gardiner. (2018). “Computer assisted diagnosis techniques (dermoscopy and spectroscopy-based) for the diagnosis of skin cancer in adults,” *Cochrane Database of Systematic Reviews*, 2018. <https://doi.org/10.1002/14651858.CD013186>
- [14] Argenziano, G., G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, y M. Delfino. (1998). “Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis,” *Archives of Dermatology*. <https://doi.org/10.1001/archderm.134.12.1563>
- [15] Ali A.-R. A, and T. M. Deserno. (2012). “A systematic review of automated melanoma detection in dermoscopic images and its ground truth data,” *Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment*, vol. 8318, pp. 421–431. <https://doi.org/10.1117/12.912389>
- [16] Fabbrocini, G., V. Vita, F. Pastore, V. D’Arco, C. Mazzella, M. C. Annunziata, S. M. Cacciapuoti, C. Mauriello, y A. Monfrecola. (2011). “Teledermatology: From prevention to diagnosis of nonmelanoma and melanoma skin cancer,” *International Journal of Telemedicine and Applications*. <https://doi.org/10.1155/2011/125762>
- [17] Litjens, G., T. Kooi, B. E. Bejnordi, A. A. Adiyoso Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, y C. I. Sánchez. (2017). “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, pp. 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- [18] Stanganelli, I., A. Brucale, L. Calori, R. A. a. M. S. Gori, B. Lovato, R. Kopf, V. Bacchilega, Rapisarda, Testori, A. Ascierio, P. E. Simeone, y M. Ferri. (2005). “Computer-aided diagnosis of melanocytic lesions,” *Anticancer Research*, vol. 25, pp. 77–82.
- [19] Jaworek-Korjakowska, J. (2018). “A deep learning approach to vascular structure segmentation in dermoscopy colour images,” *BioMed Research International*, pp. 8. <https://doi.org/10.1155/2018/5049390>
- [20] Sun, X., J. Yang, M. Sun, and K. Wang. (2016). “A benchmark for automatic visual classification of clinical skin disease images,” vol. 9910, pp. 206–222. [https://doi.org/10.1007/978-3-319-46466-4\\_13](https://doi.org/10.1007/978-3-319-46466-4_13)
- [21] Hoque, M. E. and K. Kipli. (2021). “Deep learning in retinal image segmentation and feature extraction: A review,” *International Journal of Online and Biomedical Engineering*, vol. 17, no. 14, pp. 103–118. <https://doi.org/10.3991/ijoe.v17i14.24819>



- [22] S. Hadiyoso, F. Fahrozi, Y. S. Hariyani, and M. D. Sulistiyo. (2022). “Image based ECG signal classification using convolutional neural network,” *International Journal of Online and Biomedical Engineering*, vol. 18, no. 14, pp. 64–78. <https://doi.org/10.3991/ijoe.v18i04.27923>
- [23] Tajbakhsh, N., J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. (2016). “Convolutional neural networks for medical image analysis: Full training or fine tuning?,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312. <https://doi.org/10.1109/TMI.2016.2535302>
- [24] Howard, A. G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Wey, y H. Adam. (2017). “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *Computer Vision and Pattern Recognition*.
- [25] Huang, G., L. Zhuang, L. Van der Maaten, y K. Q. Weinberger. (2016). “Densely connected convolutional networks,” *Cornell University*. <https://doi.org/10.1109/CVPR.2017.243>
- [26] He, K., X. Zhang, S. Ren, y J. Sun, “Deep residual learning for image recognition,” *Computer Vision and Pattern Recognition (CVPR)*, 2015. <https://doi.org/10.1109/CVPR.2016.90>
- [27] Szegedy, C., W. Liu, J. Yangqing, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, y A. Rabinovich, “Going deeper with convolutions,” *Computer Vision and Pattern Recognition*, 2015. <https://doi.org/10.1109/CVPR.2015.7298594>
- [28] Parraga, F. T., C. Rodriguez, Y. Pomachagua, and D. Rodriguez. (2021). “A review of image-based deep learning algorithms for cervical cancer screening,” *13th International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 155–160. <https://doi.org/10.1109/CICN51697.2021.9574680>
- [29] Liu, C., B. Zoph, M. Neumann, J. Shlens, W. Hua, L.-J. Li, L. Fei-Fei, A. Yuille, J. Huang, y K. Murphy. (2017). “Progressive neural architecture search,” *Computer Vision and Pattern Recognition*. [https://doi.org/10.1007/978-3-030-01246-5\\_2](https://doi.org/10.1007/978-3-030-01246-5_2)
- [30] Krizhevsky, A., I. Sutskever, y G. E. Hinton. (2017). “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, pp. 84–90. <https://doi.org/10.1145/3065386>
- [31] Chollet, F. (2017). “Xception: Deep learning with depthwise separable convolutions,” *Chollet, Francois*, pp. 1800–1807. <https://doi.org/10.1109/CVPR.2017.195>
- [32] Tan, M. and Q. V. Le. (2019). “EfficientNet: Rethinking model scaling for convolutional neural networks,” *International Conference on Machine Learning*.
- [33] Wei, L., K. Ding, and H. Hu. (2020). “Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network,” *IEEE Access*, vol. 8, pp. 99633–99647. <https://doi.org/10.1109/ACCESS.2020.2997710>
- [34] Vanheule, R., Artist. (2018). *Deep Learning in Skin Lesion Classification Tasks*. [Art]. Ghent University – Faculty of Engineering and Architecture.
- [35] Pham, T.-C., C. Luong, M. Visani, y H. V. Dung. (2018). “Deep CNN and data augmentation for skin lesion classification,” *Intelligent Information and Database Systems*. [https://doi.org/10.1007/978-3-319-75420-8\\_54](https://doi.org/10.1007/978-3-319-75420-8_54)
- [36] Mahbod, A., G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang. (2019). “Fusing fine-tuned deep features for skin lesion classification,” *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29. <https://doi.org/10.1016/j.compmedimag.2018.10.007>
- [37] Kassani, S. and K. P. Hosseinzadeh. (2019). “A comparative study of deep learning architectures on melanoma detection,” *Tissue and Cell*.
- [38] Kassem, M., K. Hosny, y M. Fouad. (2020). “Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning,” *Cairo International Biomedical Engineering Conference*, pp.90–93. <https://doi.org/10.1109/ACCESS.2020.3003890>



- [39] Pacheco, A. G. C., A.-R. Ali, and T. Trappenberg. (2019). “Skin cancer detection based on deep learning and entropy to detect outlier samples,” *Mathematics, Computer Science*.
- [40] Sun, Q., C. Huang, M. Chen, H. Xu, y Y. Yang. (2021). “Skin lesion classification using additional patient information,” *BioMed Research International*. <https://doi.org/10.1155/2021/6673852>
- [41] Zillur, R., H. Md, I. Md, H. Md, y H. Rubaiyat. (2021). “An approach for multiclass skin lesion classification based on ensemble learning,” *Informatics in Medicine Unlocked*, vol. 25, pp. 1–9. <https://doi.org/10.1016/j.imu.2021.100659>
- [42] Codella, N., V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, y A. Halpern. (2019). “Skin lesion analysis toward melanoma detection: A challenge hosted by the international skin imaging collaboration (ISIC),” *Computer Vision and Pattern Recognition arXiv*.
- [43] Society for imaging informatics in medicine (SIIM). (2021). “Melanoma classification challenge,”. Available: [https://siim.org/page/melanoma\\_classification\\_challenge](https://siim.org/page/melanoma_classification_challenge)
- [44] Bustamante, J., C. Rodriguez, and D. Esenarro. (2019). Real-time facial expression recognition system based on deep learning. *International Journal of Recent Technology and Engineering*, vol. 8, no: 2 Special Issue 11, pp. 4047–4051. <https://doi.org/10.35940/ijrte.B1591.0982S1119>
- [45] Alofi, N., W. Alonezi, & W. Alawad. (2021). WBC-CNN: Efficient CNN-based models to classify white blood cells subtypes. *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 13, pp. 135–150. <https://doi.org/10.3991/ijoe.v17i13.27373>
- [46] Vega, H., E. Sanéz, P. De la Cruz, S. Moquillaza, and J. Pretell. “Intelligent system to predict university students dropout,” vol. 18, no. 07, pp. 27–43, 2022. <https://doi.org/10.3991/ijoe.v18i07.30195>
- [47] Ángeles Rojas, J. A., H. D. Calderon Vilca, E. N. Tumi Figueroa, K. J. Cuadros Ramos, S. S. Matos Manguinuri, and E. F. Calderón Vilca. (2021). “Hybrid model of convolutional neural network and support vector machine to classify basal cell carcinoma,” *CyS*, vol. 25, no. 1. <https://doi.org/10.13053/cys-25-1-3431>
- [48] Kohavi, Ron. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, 1995.
- [49] Witten, I. H. and E. Frank. *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, San Francisco, CA, USA, 2nd edition, 2005.
- [50] Minhas Fu, D. Sabih, and M. Hussain. Automated classification of liver disorders using ultrasound images. *J Med Syst*. vol. 36, no. 5, pp. 3163–72, 2012. Epub 2011 Nov 10. PMID: 22072280. <https://doi.org/10.1007/s10916-011-9803-1>

## 8 Authors

**Hugo Vega-Huerta.** Doctor in Systems Engineering at UNFV, Master in Administration Science at UNMSM. Is a principal professor at UNMSM and a researcher specializing in Artificial Intelligence. He was Academic Vice Dean at the Faculty of Systems Engineering at UNMSM, Director of the Software Engineering Erogam at URP. Is responsible for the YACHAY Research Group at UNMSM. (email: [hvegah@unmsm.edu.pe](mailto:hvegah@unmsm.edu.pe))

**Renzo Villanueva-Alarcón.** Systems engineer at UNMSM with experience in the analysis, design, and implementation of Business Intelligence solutions in the financial sector. I am characterized by responsibility, a sense of teamwork, and constantly learn-

ing new technologies related to advanced analytics. (email: [renzo.villanueva@unmsm.edu.pe](mailto:renzo.villanueva@unmsm.edu.pe)).

**David Mauricio.** Doctor of Science in Systems Engineering and Computing, and Master of Science in Applied Mathematics from the Federal University of Rio de Janeiro, Brazil. He has been a professor at the North Fluminense State University of Brazil (1994–1998), and since 1998 he has been a professor at the National University of San Marcos. Areas of interest: Mathematical Programming, Artificial Intelligence, Software Engineering, Entrepreneurship (email: [dmauricios@unmsm.edu.pe](mailto:dmauricios@unmsm.edu.pe)).

**Juan Gamarra-Moreno.** Master of Science in Systems Engineering, member of the “Internet of Things” Research Group, undergraduate professor of Artificial Intelligence and graduate professor of Machine Learning and Big Data in the Facultad de Ingeniería de Sistemas de la Universidad Nacional Mayor de San Marcos – Perú (email: [juan.gamarra@unmsm.edu.pe](mailto:juan.gamarra@unmsm.edu.pe)).

**Hugo D. Calderon-Vilca.** PhD in Computer Science, research professor of the “Artificial Intelligence” Group of the Universidad Nacional Mayor de San Marcos – Peru, advisor of undergraduate and graduate thesis projects related to Neural Networks, Machine Learning and Natural Language Processing. Professor of doctoral programs in other universities. (email: [hcalderonv@unmsm.edu.pe](mailto:hcalderonv@unmsm.edu.pe)).

**Diego Rodriguez.** Peruvian researcher member of the research group in medicine and health sciences at Universidad Peruana de Ciencias Aplicadas (UPC), he has collaborated on IA projects related to biomedicine, and he has published scientific articles indexed on Scopus and IEEE (email: [U202115933@upc.edu.pe](mailto:U202115933@upc.edu.pe)).

**Ciro Rodriguez.** Peruvian researcher, and engineering educator at National University Mayor de San Marcos (UNMSM), in Lima Peru. He is member of the International Alliance for Sensing and IoT Collaboration (IASIC), and he belongs to the research group of Information Technologies Applied to Data Science (ITDATA), He is reviewer at Multimedia Tools and Applications and Applied Sciences MDPI (email: [crodriguezro@unmsm.edu.pe](mailto:crodriguezro@unmsm.edu.pe)).

Article submitted 2022-08-03. Resubmitted 2022-09-01. Final acceptance 2022-09-02. Final version published as submitted by the authors.