

HOSTED BY



Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Full Length Article

Towards automated eye cancer classification via VGG and ResNet networks using transfer learning

Daniel Fernando Santos-Bustos^{a,b}, Binh Minh Nguyen^b, Helbert Eduardo Espitia^{a,*}^a Universidad Distrital Francisco José de Caldas, Bogotá, Colombia^b École Nationale Supérieure Mines-Télécom Atlantique Bretagne-Pays de la Loire, Nantes, France

ARTICLE INFO

Article history:

Received 14 July 2021

Revised 22 May 2022

Accepted 28 June 2022

Available online 22 July 2022

Keywords:

Computer vision

Convolutional neural networks

Medical image analysis

Uveal melanoma

ABSTRACT

Complex tasks such as disease diagnosis or semantic segmentation are now becoming easier to tackle in part due to increasing advances in computing and storage. This study provides an exploratory approach utilising convolutional neural networks (CNNs) to detect ocular abnormalities with an illustrative case of uveal melanoma (UM), a type of ocular cancer. In previous studies UM has been researched employing different computational techniques focusing on discriminative features using fuzzy systems, neural networks, and adaptive neuro-fuzzy systems. However, given the inheritable nature of the problem, it was decided to use CNNs with transfer learning as a promising alternative to improve the accuracy of the results. As for the main contributions, the results outperforms different state-of-the-art computational algorithms studied to detect UM, in particular improvements in sensitivity, precision and accuracy, achieving 99%, 98% and 99%, respectively. Besides, two algorithms were implemented to reduce the bias of the dataset: a data augmentation algorithm using the Gabor filter, and an algorithm to remove light spots using Navier–Stokes approach.

© 2022 Karabuk University. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Numerous cancer patients are diagnosed symptomatically or during routine medical appointments. For instance, breast cancer, one of the most common types in women [1], and uveal melanoma (UM) intraocular cancer arising in the melanocytes of the iris, ciliary body or choroid which is a common malignant tumor in adults [2,3]. Evidence suggests that smaller tumors are associated with longer survival than larger tumors [4]; therefore early diagnosis and local treatment are crucial as tumors grow as the cancer progresses. For UM it is estimated that approximately 50% of patients will develop metastases with a survival of 6–12 months from diagnosis [5,3]. According to [3], approximately 40% of UMs have metastatic disease, with the liver being the most affected, causing a high mortality rate as liver metastases are particularly refractory to treatment [6]. When the cancer is confined to the eye, the 5-year relative survival rate is higher about 80%. This is due, in part, to the fact that most treatments have been developed for non-metastatic Uveal Melanoma [6]. For individuals with ocular melanomas that

have spread to distant parts of the body, the 5-year relative survival rate is about 15% [7].

In recent years, the use of computational algorithms in medicine has transformed and revamped the way diseases are diagnosed, enabling the development of computer-aided detection (CAD) [8] to boost disease detection; consequently, these algorithms are presented as tools for prevention. Such prevention, together with early detection and treatment are vital to improve the rates in cancer control [9,10]. As an example in [11] is presented the development of a system for providing aided clinical diagnosis service via a deep learning framework. In this regard, early detection of chronic diseases becomes relevant since such diseases limit daily life activities. For example in [12] is illustrated how deep learning models can be used to detect diseases like chronic kidney disease and type-2 diabetes; in addition, this learning scheme can be employed to recognize other diseases. Also, numerous investigations have been carried out to detect a user's mental state based on her/his posting information [13], and works to track the level of consciousness and delirium through recurrent neural networks employing frontal electroencephalography (EEG) signals [14,15].

Connected works to eye focus on pattern examination, colors and other characteristics of the iris to determine different factors like the patient's health [16–18]. For instance, in [19], a method for detecting diseases in the iris using the Gabor filter is shown

* Corresponding author.

E-mail addresses: dfsantosb@correo.udistrital.edu.co (D.F. Santos-Bustos), binh-minh.nguyen@imt-atlantique.net (B.M. Nguyen), heespitiac@udistrital.edu.co (H.E. Espitia).

allowing the diagnosis of Corneal Edema, Iridotomies, and Conjunctivitis.

On the other hand, regarding deep learning health applications, in [20] are analyzed different models for distinguishing COVID-19 from other chest related infections in X-ray images. Some architectures considered are AlexNet [21], GoogleLeNet [22], SqueezeNet, DenseNet201, MobileNetV2, and InceptionV3 using different configurations and training approaches. In regards to transfer learning used for cancer detection, in [23] is proposed a CNN based transfer learning method with structured filter pruning approach for improved breast cancer classification, in addition, this strategy seeks to decrease the run time to training the deep learning models. An additional article to consider is [24] where it is made a review of artificial neural networks, multi-layer perceptron neural network and convolutional neural network to detect breast malignancies focused on early diagnosis of breast cancer, in this work it is considered the accuracy metric to identify the suitable method for the diagnosis of breast cancer.

1.1. Proposal approach and document organization

Although UM is a comparatively rare disease it is also the most common primary intraocular tumour in adults [25]. In an attempt to address this concern, this paper presents different implementations of convolutional neural networks (CNNs) for UM diagnosis and identification. On top of that it is also given a comparison with previous implemented techniques as fuzzy systems (FSs), neural networks (NNs), and adaptive neuro-fuzzy systems (ANFIS). In this order, 1 describes the main aspects addressed in this work.

This paper is an extension of a previous (published) work for the diagnosis of UM using artificial intelligence (AI) techniques. The main contributions of this extended version are listed below:

- It is shown that the study results using CNNs with transfer learning and data augmentation to detect UM surpasses the previous works in all the performance metrics considered.
- It is introduced a data augmentation technique based on a specific application of texture kernels to improve the performance and reduce the generalized error.
- It is shown that the implementation of the algorithm to remove light spots in the eyes with Navier–Stokes acts as regularization method.

- A comparative framework is given with previous implemented techniques as FSs, NNs, and ANFIS.

A noteworthy aspect to keep in mind in the use of computational techniques like in this paper is the “No Free Lunch Theorem”, which states that considering the set of all possible problems, each computational technique will perform as well as any other on average. In other words, There is no single technique that solves all problems with the best performance [26]. In particular, for this problem, it was not intended that the developed algorithm would achieve high interpretability or be able to detect all types of irregularities in the eyes; rather the intention of this research is to obtain high success rates in the detection of cancer-like abnormalities in the eye.

The paper is organized as follows. In 2, the previously methods used to tackle UM are briefly mentioned. The technique of image processing and data augmentation used are shown in 3, while 4 contains the architectures employed. In 5 it is shown preliminary data analyses based on statistical tests. 6 describes the dataset considered and the experimental results. The discussion are then stated in 7 followed by the conclusions in 8.

2. Background

Techniques of AI have a wide number of applications, for instance, motion tracking [27], disease diagnoses [28], and image segmentation [29], among others; thus these techniques leverage and extend traditional disease diagnosis. As an example, in [30] it is presented an approach to diagnose diseases by radiotherapy, the algorithm uses machine learning and specific features to distinguish closed and open eyes; alternatives to these works include advanced techniques with particles such as [2] which employs nanoparticles manipulated to identify cancerous sites.

This paper compares the work shown in [31,32] on a method that integrates algorithms related to iris segmentation and proposes algorithm using fuzzy logic, neural networks and adaptive neuro-fuzzy networks (ANFIS); obtaining 76% of correct classification in the fuzzy logic system, 96.04% for the neural network system, and 91.68% for ANFIS. The input to these classifiers is a dataset whose columns are the seven moments of Hu $h_0, h_1, h_2, h_3, h_4, h_5, h_6$ along with the classes “healthy” and “un-

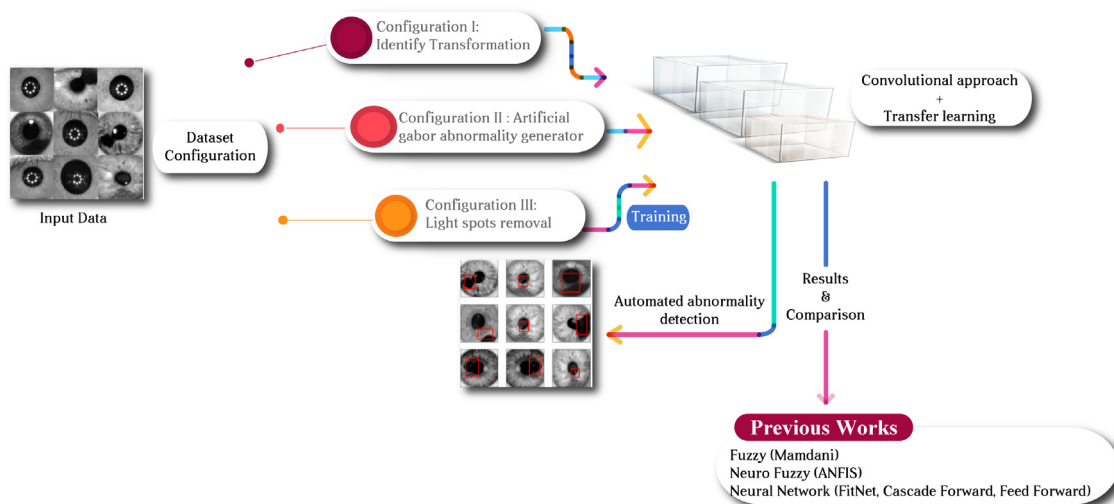


Fig. 1. Visual abstract of the study case: I. Sample data input (iris images); II. Proposed configurations, a) with noise b) abnormality generator c) without light spots); III. CNNs with transfer learning, IV. Compared techniques and V. System detection..

healthy”; nevertheless, these approaches are considered to have the following deficiencies:

1. When images are transformed from polar to cartesian, an interpolation occurs thereby losing information and adding uncertainty to the measurements.
2. It is considered the imbalance between the number of class elements in the training dataset given the low prevalence of the disease, which increases the generalized error.
3. The previous models only employ the seven Hu moments for training, which is suitable as the moments are powerful descriptors; however, other features may influence the detection of this disease.

For the aforementioned reasons, it was decided to implement a new configuration of models which addresses these problems using CNNs. However, collecting data from this particular disease to effectively train a CNN is difficult. In consequence, data augmentation techniques are selected to generate input data, thus reducing the reliance on training data and reducing the generalized error (operating as a regularization method) [33]. For example, applying noise as data augmentation technique involves altering some pixels of the image, while keeping the actual meaning as a whole. Works have proposed different methods to augment the data, for example, in [29] authors training a U-net with different training set size and employing different artificial data augmentation strategies such as horizontal flip, affine transform, Gaussian random field, Gaussian white noise, rotation and dilation in which the accuracy improved significantly when performing horizontal flips. Meanwhile, in [34] for an application in voice command recognition, the error rate reduced when the original dataset was combined with augmentation methods involving speed perturbation and room impulse response reverberation, improving the generalization capability of convolutional neural network when used for voice command recognition. Room impulse response reverberation produces voice command variations caused by reflected sound paths while speed perturbation generates voice command variations caused by shorter or longer time duration. In [35], it is designed a 13-layer CNN, using three types of data augmentation methods: Gamma correction, noise injection, and image rotation, these approaches are used for image-based fruit classification; the results show that data augmentation can increase the overall accuracy. Finally, in [36] it is displayed a patch shuffle stochastic pooling neural network for an explainable diagnosis (PSSPNN), a model used for classification among COVID-19, healthy subjects, secondary pulmonary tuberculosis, and community-captured pneumonia. In this application, an improved multiple-way data augmentation helps to avoid overfitting.

Considering the above, it is proposed a simple and suitable algorithm for creating eye abnormalities as an augmentation technique, which is convenient for the present case by using a particular application of the Gabor filter for texture generation instead of texture recognition [37]. It is noteworthy that these filters are most effective when used as preprocessors and feature extractors in image processing applications such as pattern recognition and motion analysis or stereoscopy [38].

3. Data augmentation and image processing

For the present study case, the iris is the region of interest (ROI) which will be segmented to perform the classification. Segmentation in CAD is critical as incorrect segmentation can lead to false positives (FP) and false negatives (FN), therefore, the database is critical, particularly, when the samples are unbalanced. This section first presents the background on data augmentation and the

Gabor filter application since the proposed approach employs these techniques. Subsequently, the algorithm to perform the balance of the dataset is also shown.

3.1. Data augmentation

Data augmentation techniques (AT) generate input data for machine learning models to increase performance, reduce both the reliance on training data and overfitting [33,39]. Thus, as the occurrence of UM is rare, it is then suggested to use AT. In particular, generic input transformations are useful for this task as CNNs can robustly classify objects, even if they are positioned in different orientations, as they are invariant to translation, viewpoint, size, and illumination (or a combination of these). Moreover, rotation AT is potentially useful for the iris dataset as the iris can be seen as circle. According to [40], data augmentation techniques are proposed being the most popular:

- Flip: The image can be flipped horizontally and vertically depending on the object. For example, it is not possible to flip an image of a car vertically, as it becomes a different object (e.g., a horizontal flipped car is still car but a flipped “3” will not be a number any more).
- Zoom: This technique creates new versions of an image with different zooming views. The resulting images are zoomed in or out according to a predefined range. In [41], the range can be specified by a value or a list of two threshold values (e.g, objects in a scene can be zoomed into obtain greater detail while losing contextual details).
- Rotation: This technique rotates the image with a predefined or random angle. Each rotated image is unique in the dataset. Typically, there are 359 degrees of rotation (e.g, it is still possible to read the characters in the book by turning the book 2 degrees, but it will be almost impossible to do so by turning it 180 degrees).
- Noise addition: Applying noise involves altering some pixels of the image, while keeping the actual meaning of the image as a whole (e.g, adding white noise evenly to an image or a song in a light way will not affect the meaning).
- Channel shift: For each channel R, G, B of an image, channel shift randomly move the pixel values in each channel by a predefined range. As a result, multiple color versions were produced by this technique. This augmentation technique increases the model robustness to color variation. As the color is influenced by light, it is particularly useful to apply it when the detection system is used in a different light context (e.g, self-driving cars).

The correct zoom ranges are defined for each dataset since given that the location and zoom of the region of interest vary in each set of images. In particular, very large zoom values can cut off the region of interest and very small values can blur the region due to loss of information. For this case, the values were calculated experimentally; a similar procedure can be employed to find the ranges of rotation and noise, since for each case study the valid zoom, rotation ranges, and noise values are different.

In a formal way, given a dataset $X \in \mathbb{R}^{n \times m}$ where n is the number of rows and m the number of features and Y the classes (with/without UM), so that the classifier $f: X \rightarrow Y$ with performance $P(f)$ which is not accurate enough, thus the n can be extended by means of using AT, so that $Z = [X', X]$ where X' is the augmented data and $f': Z \rightarrow Y$ resulting in $P(f') \geq P(f)$.

The mutual information inequality states that $I(Z; Y) \leq I(X; Y)$, however, as this study helps to generate new data with an external method like the Gabor kernels, it is therefore possible to say that

$I(Z;Y) \geq I(X;Y)$ which provides theoretical bases to the employed method.

3.2. Gabor filter

Gabor filter is fundamentally a Gaussian function modulated by a complex sinusoidal of frequency and orientation. It has the ability to operate in both the spatial and frequency domains and can be used in any number of dimensions [42].

Gabor filters have been used in a wide range of image processing tasks, in particular those related to texture information extraction [43]. Nevertheless, in this study this filter was used as the foundation of a data augmentation method that contains a Gabor generator kernel for generating tumour-like textures. 3 describes the procedure to generate an augmented image $A(x,y)$ from an image $I(x,y)$ by using a Gabor filter $K(f,\theta,\sigma)$. These kernels can be used for features extraction which consists of a set of filters with different frequencies f and orientations θ [44,45]. For two-dimensional the Gabor filters are:

$$G_c(x,y) = B \exp(-\psi) \cos(\rho) \quad (1)$$

$$G_s(x,y) = C \exp(-\psi) \sin(\rho) \quad (2)$$

$$\rho = 2\pi f(x \cos(\theta) + y \sin(\theta)) \quad (3)$$

$$\psi = \frac{(x^2 + y^2)}{2\sigma^2} \quad (4)$$

In Eqs. (1), (2) B and C are the normalization factors, f determines the frequency to be pursued in the texture. Changing θ defines the texture oriented in a particular direction; and varying σ results in changing the size of the region of the image being analyzed. In 2 it is shown an example of a Gabor filter contour, given by Eq. (5).

$$G(x,y) = e^{-(x^2+y^2)/10} \cos(2\pi x) \quad (5)$$

Considering the above, a schema to create tumour-like spots to augment dataset is presented in 3.

3.3. Light reflect compensation

In the dataset is observed some spot lights in the pupil, potentially caused by the device that captured the photo. Especially in healthy class dataset, as this may be a distinguishing mark, it is therefore necessary to remove them in order to minimize the bias. In this way, it is proposed a configuration III (see 6.5) with an algorithm to compensate the light reflection in the eyes, as described in 4. This process is divided in the following steps:

- Step (a): Build a mask using the threshold and the circular Hough transform. As the light reflection has a strong grey value, it is fairly simple to use a threshold to obtain the desired

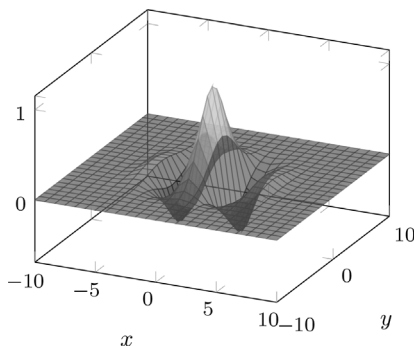


Fig. 2. Contour for a Gabor filter given by Equation (5).

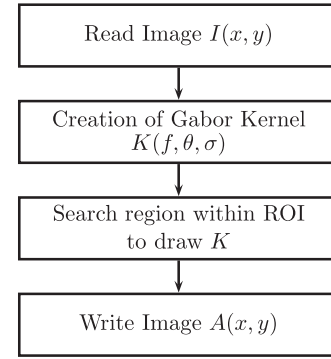


Fig. 3. Proposed schema to create tumour-like spots to produce an augmented dataset.

regions. However, in some images of the healthy class, the parts outside the iris may contain high grey level pixels. As it is only needed to cover the light pattern in the pupil for the healthy class, it is therefore necessary to locate the region where the pupil is located. Hence, from the input image, the Hough's circle transform is applied to find the region of the pupil circle and mask that region. After an element-wise multiplication is performed, the final mask containing the region of light pattern is obtained.

- Step (b): Image inpainting technique to fill in the region and make it smooth. The chosen algorithm is Navier–Stokes image inpainting. Navier–Stokes method based on fluid mechanics (Navier–Stokes equation), defined by partial differential equation (PDE). It is essential to use higher order PDEs for the modeling of incompressible fluids, and thus to produce the continuation of isophotes (the fluids in the images) resulting in smooth interpolation[46]. The idea here consists of considering the image as fluid, and let this flow to fill the holes, which are the regions indicated in the mask. Before applying the Navier–Stokes inpainting, it is applied morphological operation dilation. The aim of this action is to ensure the holes and the dark part of pupil are connected directly. As a result, the dark pixels can flow and fill the holes.

4. Convolutional neural networks

A special type of deep learning model for processing data that has a grid pattern [47], designed to recognise visual patterns from images, CNNs have shown high accuracy in recognizing abstract patterns and features that a human would not easily obtain [48], perhaps due to the reduction in the number of parameters with respect to NNs [49]. Accuracy is a significant indicator of whether a CNN-based classification method is excellent [50]; the deeper the network the greater the feature abstraction. Perhaps some of the remarkable aspects about this type of network is the suppression of dependency on feature maps in the spacial dimension, in other words, deep learning models can detect an object independently of its location [51], and the deeper data is in the network, the greater is the abstract characteristic that these data represent [49]. For the aforementioned reasons, and looking at the current state of the art for the diagnosis of diseases, VGG-16 [52] and ResNet-18 with pre-trained CNN with transfer learning are used.

4.1. Transfer learning

Transfer learning is a technique that uses a model that was trained on one task and is transferred to a new related task. For the case considered in this work, an image classification model

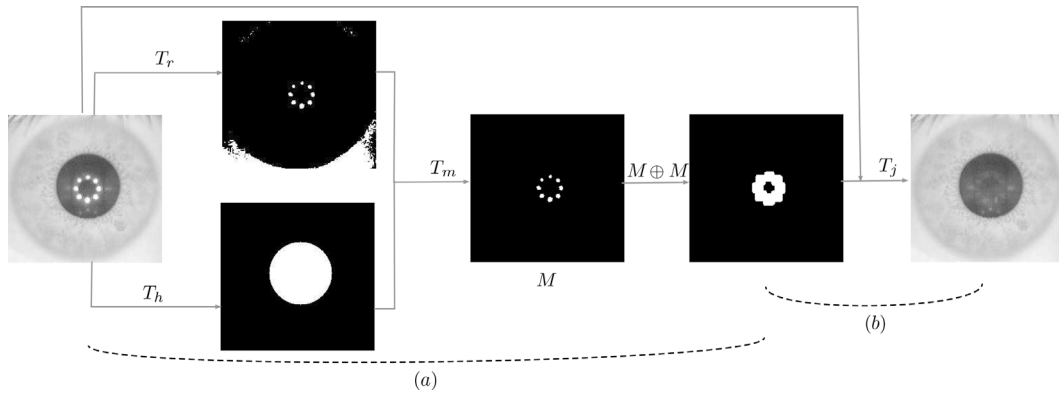


Fig. 4. Proposed schema to compensate light reflect, where (a) represents the process to create the mask and (b) stands for the image inpainting, T_r is a thresholding operation, T_h is a Hough circle transform, T_m is the element wise multiplication and \oplus is dilation process.

trained on ImageNet (a huge dataset containing 1.2 million labelled training images) can be reused for classification and detection of ocular anomalies. For this reason, transfer learning is useful when the size of the dataset does not allow deep neural networks to be trained from scratch. Using transfer learning, neural networks can have adequate initial parameters and require only minor modifications to adapt to new tasks [53].

When using transfer learning, there are two main approaches. In the first one, the pre-trained model weights are considered as the initial weights for the model in a new job, and these weights are trained and updated during the training process. This approach is called *fine-tuning transfer learning*. On the other hand, in the second approach, the pre-trained model is used only to extract features, which implies that the pre-trained weights are not affected during the training progress of the new task. In this approach, only one classifier (usually containing fully connected layers) is trained on the pre-trained model to provide classification results [53]. The name of this approach is feature-extracting transfer learning (FETL). Fine-tuning transfer learning method is used in this paper, as the performance of this approach is surpassing that of feature extraction transfer learning.

Transfer learning is particularly useful when training data are scarce, obtaining comparable performance as to health experts in disease diagnosis as observed in [54,55], and can be used for the case of UM. Within applications of TL in disease diagnosis research, it is proposed to use networks for the identification of Alzheimer's disease [56], in particular a VGG (VIN)-inspired network is considered as a backbone network (backbone network: a network that takes the image as input and extracts the feature map). In addition, the use of attention mechanisms is addressed; attention is a mechanism to improve the performance of the network inspired in human-eye way to process images, other methods to improve a network are depth, width and cardinality. In this order, the authors integrate convolutional block attention modules into a VIN backbone. Data augmentation in 18-way is also proposed to avoid overfitting. Experimental results demonstrate the effectiveness of VGG and 18-way data augmentation.

4.2. Deep convolutional neural network models

This work evaluates the performance of two well-known pre-trained models, VGG-16 and ResNet-18, the figure describes the scheme used with ResNet for the present work. In both figures the convolutional layer parameters are denoted with the following notation: *receptive field size, conv, number of channels*.

In this order, 5 presents the ResNet-18 architecture, where the solid lines represent identity shortcuts when the dimensions of

input and output are the same. The dotted lines represent the input and output shortcuts with different dimensions. In that case, there are two options: the shortcuts are still considered as identity shortcuts with extra zero entries padded to increase input's dimensions, or 1×1 convolutions used to project input's dimensions to output's. In second option, the shortcuts are called projection shortcuts (PS) [57].

VGG net is a convolutional neural network proposed by the Visual Geometry Group and DeepMind of Oxford University [58] that has been widely applied in many fields as observed in [59–64] for places image recognition, fruit detection, thyroid nodules identification and objects tracking, respectively. The structure of VGG-16 can be observed in 6.

The input to "Conv 1" layer is of fixed size 224×224 RGB image. The image is passed through a stack of convolutional layers where the filters were used with a very small receptive field 3×3 . In one configuration, it uses 1×1 convolution filters operating as a linear transformation of the input channels. The convolution stride is 1 pixel; the spatial padding of convolutional layer input is in order to spatial resolution preserved after convolution. Spatial pooling is carried out by five max-pooling layers that follow some of the convolutional layers. Max-pooling is performed over a 2×2 pixel window, with step of 2. Three fully-connected layers follow a stack of convolutional layers and the final is the soft-max layer. All hidden layers use rectification (ReLU) non-linearity activation function.

ResNet is a well know convolutional neural network architecture employed to train hundreds or thousands of layers achieving outstanding performance. The biggest advantage of ResNet is the ability to reduce the vanishing gradient problem. Before ResNet, deep networks were difficult to train since the gradients need to propagate backwardly through an enormous number of layers in a deep network, which made the gradient to be infinitively small. To solve this problem, ResNet proposed the novel idea of using the bypass pathway concept as a way to skip the backward connections between layers and create identity shortcuts in the gradients' path, which allows the gradients flow faster to the initial layer [65].

ResNet also has been used in medical image classification and diagnosis. In [53], S. Minaee et al. used ResNet-18 and ResNet-50 with transfer learning for predicting COVID-19 from chest X-ray images. These two models were pre-trained on ImageNet dataset. Their classification results were 89.58% with ResNet-18 and 92.29% with ResNet-50. In another research, S. Ayyachamy et al. [66] applied ResNet-18 into their content-based image retrieval framework to retrieve the images which have similar pathology of the one under examination. Their pre-trained ResNet-18 provided 92% classification accuracy with 0.9 mean average of retrieval precision.

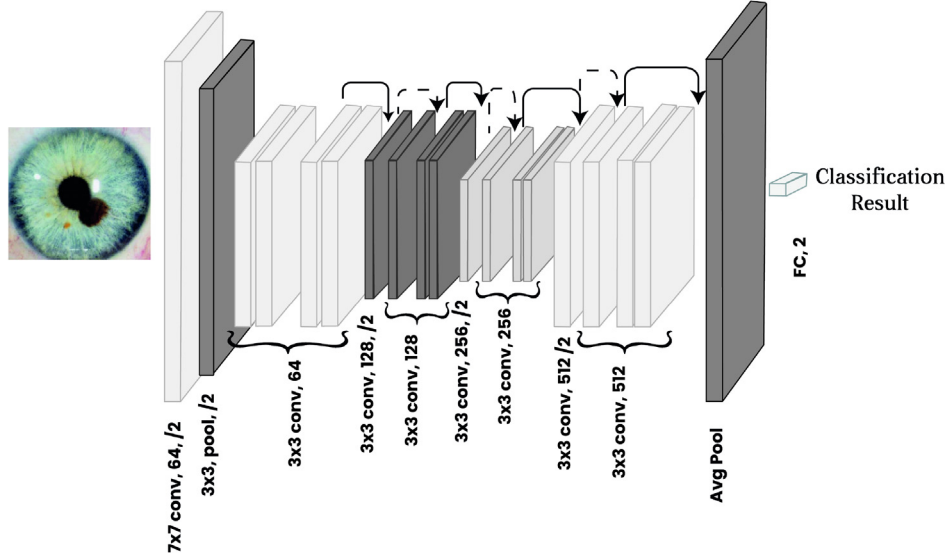


Fig. 5. Example of network architecture for ResNet-18 applied for the case study of UM diagnosis. The dotted shortcuts increase dimensions. Conv blocks consist of a convolution, batch normalization and max pooling operation.

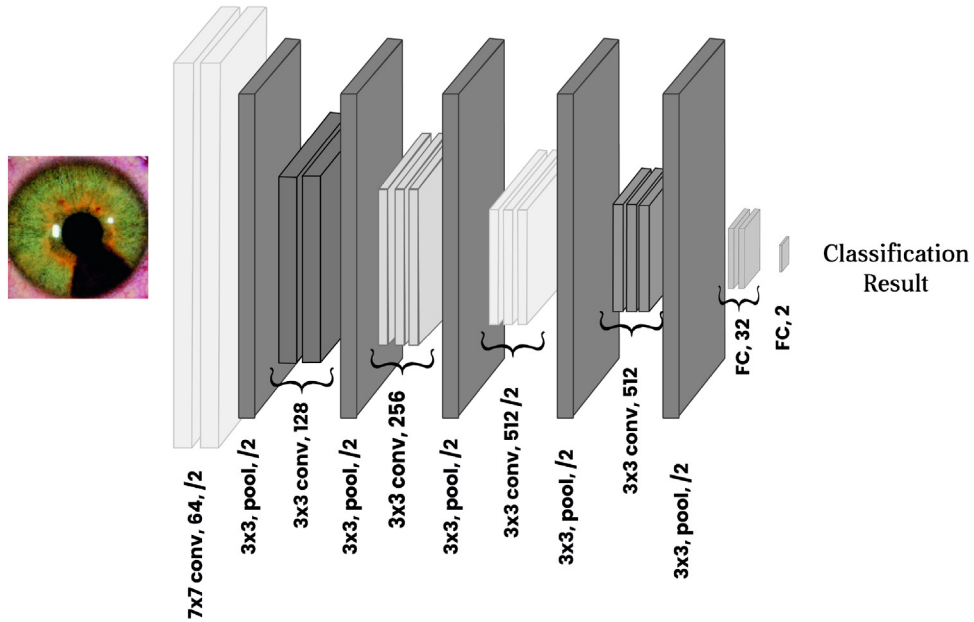


Fig. 6. Example network architecture for VGG-16 for the case study of UM diagnosis. Conv block consisting on a convolution + batch normalization + max pooling operation.

Both models are trained with cross-entropy loss function aiming to minimize the distance between predicted probabilities and ground-truth probabilities. It is defined as in Eq. (6), with p_i and q_i as the ground-truth and predicted probability, respectively.

$$\text{Cross - entropy loss} = -\frac{1}{N} \sum_{i=1}^N p_i \log q_i \quad (6)$$

To update parameters, it is used Adam (Adaptive Moment Estimation) optimizer. This optimizer takes advantages of momentum and adaptive learning rate to converge faster by considering past gradients and squared gradients. Adam optimizer combines the benefits of both AdaGrad (Adaptive Gradient) optimizer and

RMSProp (Root Mean Square Propagation) optimizer, which means that Adam works well with sparse gradients and on-line non-stationary settings [67]. The update rule of Adam optimizer is described as:

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{s}_t + \epsilon}} \quad (7)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (8)$$

$$\hat{s}_t = \frac{s_t}{1 - \beta_2^t} \quad (9)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} J(\theta) \quad (10)$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2) (\nabla_{\theta} J(\theta))^2 \quad (11)$$

In this set of equations θ represents the network's parameters, J is the loss function to optimize, m_t and s_t are the first and second moments at time t . In addition, ∇ represents the gradient operator, α is learning rate and ϵ is smoothing term, which prevents division by zero. The hyperparameters β_1 and β_2 are exponential decay rates for moment estimation.

5. Preliminary data analysis

For classification process, it is especially important to check the differentiation between classes in the dataset, which allows the classifiers to learn and distinguish between different types of images to save time and resources in the design and formation of configurations. Clearly from the figures it could be seen that there is a difference between the feature groups, this is an evidence of such differentiation in the sample, but is not enough in a statistical way. Therefore, it is necessary to establish a statistical test that provides stronger evidence of whether the two samples come from two different populations.

For this procedure, it is used the moments of an image, which are measures that provide a generic representation of an object with simple or complex figures [68]. In this case, Hu moments are used as measures in the preliminary analysis.

In order to make the statistical test, Shapiro–Wilk was performed to confirm the non-Gaussian nature of the data [69], but this test is very sensitive under certain conditions; therefore, the D'Agostino–Pearson test was also performed since this considers the asymmetry (skewness) and flattening (kurtosis [70]) of the distribution.

Considering the Shapiro–Wilk and D'Agostino–Pearson results shown in 1 for normality tests, it can be seen that for the Hu descriptors h_i the $p_{value} < 0.05$. In conclusion, the data sets do not come from a normal distribution; therefore, it is used the Kruskal–Wallis test, since ANOVA is implemented only when the data follows a normal distribution. Kruskal–Wallis test is a non-parametric “analysis of variance” used to compare several independent samples [71].

In this order, to determine the difference between groups, the null and alternative hypothesis are the following:

Null hypothesis H_0 : the mean of the values of Hu for the images without UM is equal to the values of Hu for the images with UM.

$$\mathcal{H}_0 : \forall \mu_i | \mu_i = \mu_j, 1 \leq i, j \leq 7 \quad (12)$$

Alternative hypothesis H_1 : The mean Hu values for the images without UM are different from the Hu values for the images with UM.

$$\mathcal{H}_1 : \exists \mu_i | \mu_i \neq \mu_j, 1 \leq i, j \leq 7 \quad (13)$$

After carrying out the Kruskal–Wallis test, the results are shown in the last two columns in 1. Considering the probability of type I error, H_0 is rejected concluding that a differentiation can be established between the two groups of characteristics as they come from different populations, thus it is possible to continue to the stage where the classifiers and the parameters will be selected, and different configurations will be defined so that in a later stage the training and generation of models can be carried out.

6. Experimental results

For generating the results it was used Colab of Google that provides GPU computing and enough memory to train the classifiers. Regarding the dataset, there are 150 healthy and 33 unhealthy images, some of them are shown in 7. Two experimental configurations were designed, *Configuration I* and *Configuration II* to train VGG-16 and ResNet-18, to reduce bias and to increment the robustness of the proposed method. Additionally, *Configuration III* is proposed to compensate the light reflect in the eyes.

Also to be considered in these results in certain images within the healthy class, the parts outside the iris may contain high grey level pixels, this is due to the conditions in which the pictures were taken; nevertheless, as the conditions were in a dark environment, this caused parts such as the sclera to have a grey tone; fortunately, this also increased the tone of the iris. In fact, this is not a problem as it can be interpreted as a magnification technique.

In this order, 2 shows the results obtained in this work and others presented in previous implementations using other approaches. In 2, the implementation of convolutional networks surpasses the previous works in all the performance metrics considered, showing that CNN can be a suitable option to detect uveal melanoma.

6.1. Parameters, metrics and tests

The context in which the investigation is situated is a machine learning task of a binary classification problem. Healthy individuals here are labeled as positives (*without UM*), while the negative label is attributed on images with anomalies (*with UM*). From this binary classification, it is obtained the tags T_P , T_N , F_P and F_N that will facilitate the computation of the following metrics: *accuracy* (how often the model is right), *error rate* (how often the model is wrong), *sensitivity* (how often the model classifies the positive instances right), *specificity* (how often the model classifies the negative instances right), *precision* (how often the model classifies right when it detects a positive), F_{score} (harmony mean between precision and sensitivity) and Jaccard similarity coefficient (JSC) corresponding to the similarity between the predicted and the

Table 1
Normality test results Shapiro–Wilk and D'Agostino–Pearson on the classes with and without UM, and results of Kruskal–Wallis test.

Variable	Without UM				With UM				Kruskal–Wallis	
	Shapiro–Wilk		D'Agostino–Pearson		Shapiro–Wilk		D'Agostino–Pearson			
	Statistic	p_{value}	Statistic	p_{value}	Statistic	p_{value}	Statistic	p_{value}	Statistic	p_{value}
h_1	0.77	2.86e-16	110.38	1.07e-24	0.63	0	1312.81	8.42e-286	100.73	1.05e-23
h_2	0.49	1.42e-24	256.97	1.58e-56	0.31	0	1922.19	0	83.52	6.29e-20
h_3	0.46	3.82e-24	179.93	8.46e-40	0.12	0	2545.11	0	107.01	4.43e-25
h_4	0.44	1.58e-24	192.42	1.67e-42	0.12	0	2545.23	0	75.57	3.53e-18
h_5	0.23	6.49e-28	319.40	4.39e-70	0.04	0	3131.43	0	56.54	5.49e-14
h_6	0.27	2.22e-27	303.13	1.49e-66	0.08	0	2612.77	0	47.54	5.39e-12
h_7	0.21	4.04e-28	258.06	9.17e-57	0.16	0	655.27	5.11e-143	4.62	0.031

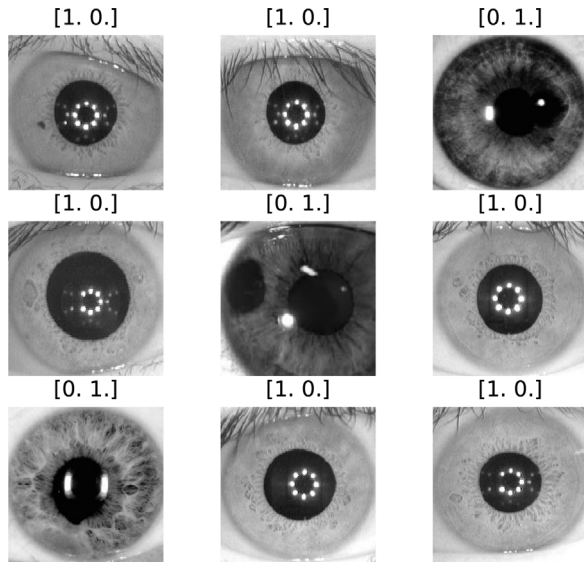


Fig. 7. Some images of the dataset where [1. 0.] indicates that the image belongs to the first class *healthy* and [0. 1.] to the second class *unhealthy*.

actual class, this coefficient is linear dependant with the F score; the equations are shown below.

$$\text{Specificity} = \frac{T_N}{F_P + T_N} \quad (14)$$

$$\text{Sensitivity} = \frac{T_P}{T_P + F_N} \quad (15)$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \quad (16)$$

$$F_{\text{score}} = 2 * \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (17)$$

$$\text{JSC} = \frac{T_P}{T_P + F_N + F_P} \quad (18)$$

Additionally, it is also computed the metric Matthews correlation coefficient (MCC) a trustworthy measure when it comes to binary classification more than accuracy and F score, as accuracy can present misleading results when a certain label performs much better than the other [72], unlike the other measures, this measures ranges from -1 (worst) to 1 (best), the equation of MCC is as follows.

$$\text{MCC} = \frac{T_P \cdot T_N - F_P \cdot F_N}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}} \quad (19)$$

Table 2
Comparative results of techniques to classify UM.

Method	Accuracy	Precision	Sensitivity	Error Rate	F1-Score	MCC	JSC
Mamdani Fuzzy System [31]	0.7663	0.8392	0.8885	0.2337	0.8614	–	0.7565
Feed Forward [31]	0.9604	–	–	0.0396	–	–	–
Cascade Forward [31]	0.9569	–	–	0.0431	–	–	–
FitNet [31]	0.9554	–	–	0.0446	–	–	–
Neuro-fuzzy configuration I [32]	0.9020	0.7671	0.2828	0.0980	0.4133	0.4117	0.2605
Neuro-fuzzy configuration II [32]	0.8958	0.7671	0.4848	0.1042	0.5942	0.4767	0.4227
Neuro-fuzzy configuration III [32]	0.9168	0.7368	0.4949	0.0832	0.5921	0.5615	0.4206
VGG-16 configuration I	0.9760	0.9520	0.9917	0.0240	0.9714	0.9448	0.9444
ResNet-18 configuration I	0.9800	0.9600	1.0	0.0200	0.9796	0.9608	0.9600
VGG-16 configuration II	0.9620	0.9680	0.9565	0.0380	0.9624	0.9241	0.9275
ResNet-18 configuration II	0.9640	0.9440	0.9833	0.0360	0.9632	0.9287	0.9290
VGG-16 configuration III	0.9900	0.9840	0.9880	0.0010	0.9860	0.9720	0.9724
ResNet-18 configuration III	0.9680	0.9800	0.9570	0.0320	0.9694	0.9363	0.9406

Regarding the hyper-parameters that will model the behavior of the neural networks for VGG-16 are as follows: batch size to 128 and learning rate to 0.0001. On the other hand, ResNet-18, the batch size, and learning rate were set to 20 and 0.0003 respectively. All the input images are resized to $(224 \times 224 \times 3)$ to match the input's requirement of the network.

6.2. Dataset

The original database is taken from [73]. After the process of augmentation, a dataset of 2048 images was obtained, 1024 healthy and 1024 unhealthy. A dataset of 500 images was used for testing, some of the results are shown in 7.

6.3. Configuration I

For this dataset configuration, ResNet-18 and VGG achieve 98% and 97.6% testing accuracy. However, it is considered that this high accuracy is not due to the tumour, but to other differentiating features, such as the light pattern in the pupil and the scale of the iris. Precisely, almost all images of the healthy class contain a pattern of reflected light in their pupils, hence a biased dataset, which makes these networks primarily trained based on this noise. The confusion matrixes are shown in 8 and 9. Additionally, the metrics are shown in 2.

Even though the performance of CNN is heavily impacted by the number of images, it was created a novel way to create abnormalities as described above. On the other hand, this can cause the neural network to be focused on learning on this artificial-created tumors whose pattern is intended to be as similar as the one presented in the unhealthy images. Nevertheless, by employing this system the abnormality will be identified (even if it is artificially created), this is not a major concern since after detecting the abnormality the doctor (medic) assigns additional studies to verify the CAD results.

6.4. Configuration II

This section displays the results using the augmented dataset with the artificial texture-like abnormalities, some of them can be seen in 10.

The classifiers were again trained and tested with 1024 and 250 images respectively, this time, half of them were with images generated simulating abnormalities to reduce the bias caused by illumination and brightness; a sample of this dataset can be seen in 11.

After training and testing, both models VGG-16 and ResNet-18 provide suitable results with 96% testing accuracy. In 12 and 13 it

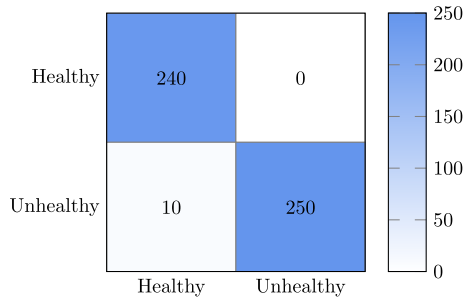


Fig. 8. Confusion Matrix ResNet-18 – Configuration I.

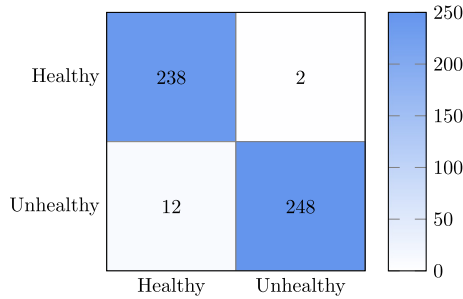


Fig. 9. Confusion Matrix VGG-16 – Configuration I.

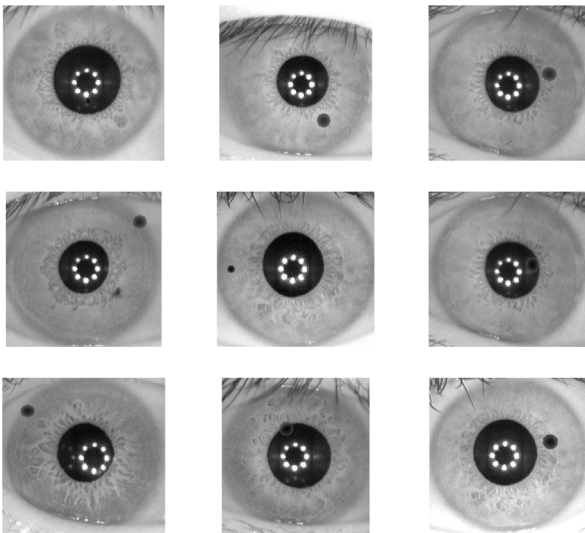


Fig. 10. Healthy images after process to add abnormalities tumour-like.

is possible to see the results of *configuration II*, the metrics are shown in 2.

6.5. Configuration III

In this part, the model is trained and tested using the dataset, in which the light reflection is compensated in healthy and unhealthy class.

The classifiers were still trained and tested with 1024 and 250 images for each class. The arrangement of dataset was the same as configuration I (which means there is no mixture in unhealthy class like configuration II).

After training and testing, both models VGG-16 and ResNet-18, provide remarkable results with 99% and 97% and testing accuracy respectively, in 14 and 15.

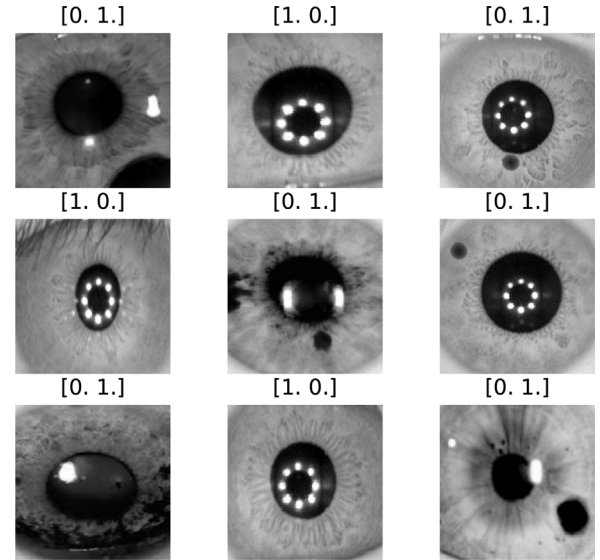


Fig. 11. Dataset unhealthy with images after process to add abnormalities rounded-like, where [1. 0.] indicates that the image belongs to the first class (healthy) and [0. 1.] to the second class unhealthy.

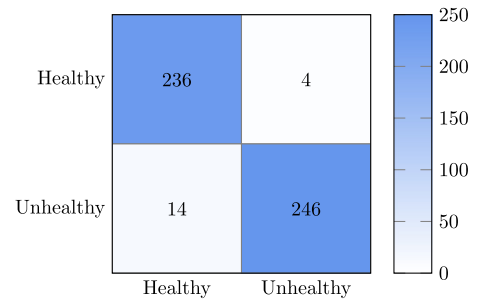


Fig. 12. Confusion Matrix ResNet-18 – Configuration II.

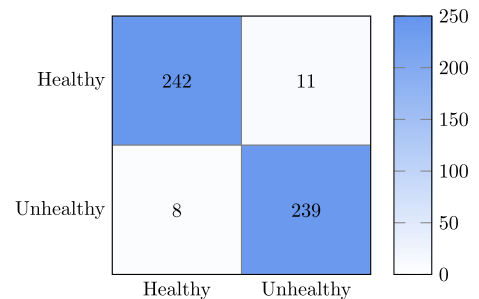


Fig. 13. Confusion Matrix of VGG-16 – Configuration II.

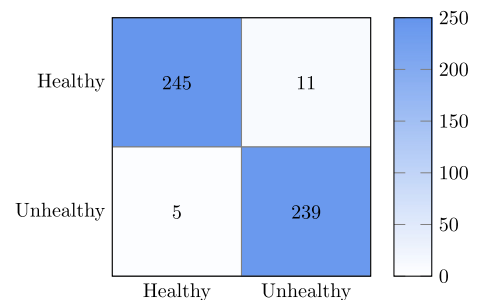


Fig. 14. Confusion Matrix ResNet-18 – Configuration III.

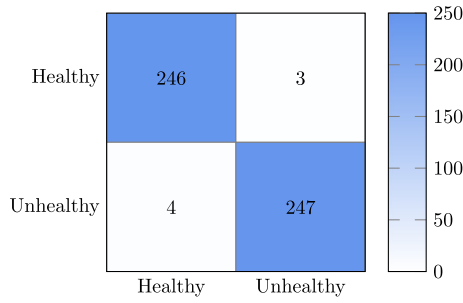


Fig. 15. Confusion Matrix of VGG-16 – Configuration III.

As the results shown in the confusion matrices, after light reflect compensation, the dataset turned to be less bias. Especially with VGG-16, the classifier provided an excellent result.

6.6. Uveal melanoma detection

After producing the classifiers for uveal melanoma diagnosis, an object detection system was trained to identify the region where this abnormality is located. Then it is reused the pre-trained ResNet-18 model and put on top of it a fully connected layer that has 4 neurons, whose outputs represent 2-dimensional coordinates of two points that define a rectangle box: the top left point and the right bottom. The network is trained with a dataset of 900 images labeled with bounding boxes, the training and testing ratios are 0.8 and 0.2 respectively. During the training process, it is applied L_1 loss function to measure the improvement of the model and Adam optimizer to update parameters. The L_1 loss function is defined as in Eq. (21), where Y_{True} and $Y_{Predict}$ are the ground-truth coordinates and predicted coordinates respectively. The mean square error (MSE) in Eq. (20) is also studied when performing the training process, however, the results were not significantly different from the model using L_1 loss given by Eq. (21).

$$MSE = \frac{1}{N} \sum_{i=1}^N [Y_{True}(i) - Y_{Predict}(i)]^2 \quad (20)$$

$$L_1 \text{ loss} = \frac{1}{N} \sum_{i=1}^N |Y_{True}(i) - Y_{Predict}(i)| \quad (21)$$

The convolutional neural network was trained using 100 epochs, learning rate of 0.008, and batch size equal to 20. The training and validation loss are shown in 16. As observed, in the final epochs the validation loss is lower than the training loss.

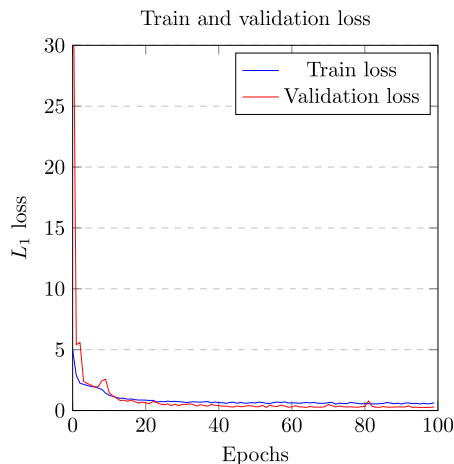


Fig. 16. Training loss and validation loss through the training process.

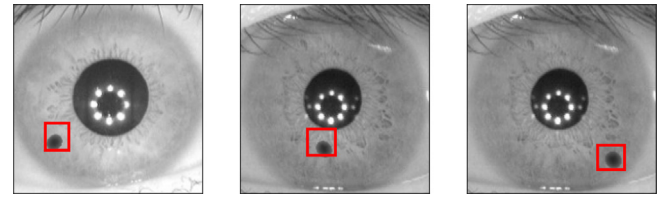


Fig. 17. Abnormality detected using ResNet-18 over dataset with abnormalities generated with Gabor filter.

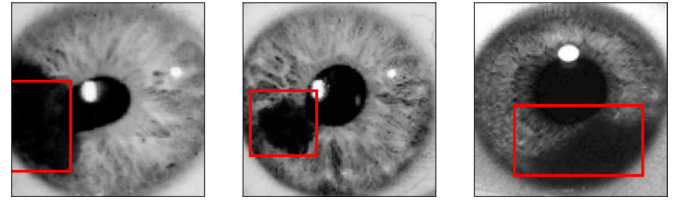


Fig. 18. Uveal melanoma detected using convolutional neural network ResNet-18.

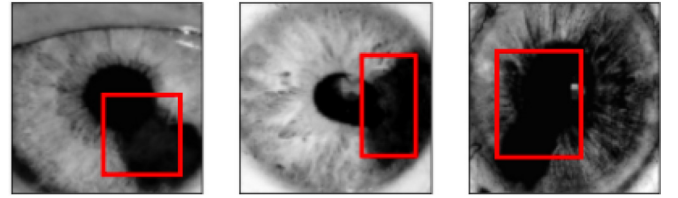


Fig. 19. Uveal melanoma detected employing ResNet-18 with dataset without light spot on iris using process in 4.

After carrying out this process, some outputs generated by the developed system are shown below, where the red rectangle corresponds to the abnormality detected by the system. 17 displays the abnormality generated using Gabor filter, and the detection employing ResNet-18. In this way, an example of uveal melanoma detected with ResNet-18 is shown in 18. Meanwhile, 19 exhibits the results of uveal melanoma detected employing ResNet-18 with dataset without light spot on iris using process in 4.

7. Discussion

In the present study two convolutional neural network architectures are used, according to the results, it is observed that in this particular case the deeper the better is not a case as the VGG16 models outperformed Resnet18 albeit it required more time to train. On the other hand, it is important to consider the overfitting effect of the healthy images; for overcoming this problem, a different data augmentation technique using the Gabor filter has been suggested to create a texture similar to the ocular anomalies, which led to a reduction of the generalised error, making this augmentation technique applicable for related works.

The proposed CNN models were evaluated by means of resampling methods, in particular with k-fold cross-validation, from which the mean scores of the metrics are calculated and directly compared in each training period; therefore; additional statistical methods on the results were not applied since the probability due to the number of training epochs reduced the possibility of a statistical fluke. However, the possibility is not discarded and it is foreseen for future work to implement statistical significance tests.

Additionally, an aspect to consider is the possibility of integrating interpretability into the CNN design to achieve a better config-

uration of the CNN. This can be achieved by incorporating concepts and operators of digital signal processing in the design of the convolutional layers.

As shown in the results, the use of AT is a suitable option for this application, in order to improve the results, it can be used the generative adversarial network (GAN) technique proposed by Goodfellow [74], whose approach is to create a generative network model that can produce realistic data, in this case, tumours images or any other kind of eye abnormalities.

Finally, the framework employed might be limited by the low number of data available, and also by the classification method adopted since the techniques used are not based in feature extraction.

8. Conclusions

The main advantage of deep learning over the algorithms of machine learning compared is the ability to obtain high-dimensional features, which may not be interpretable by humans, but have again shown to be able to solve a rather complex problem with a high level of accuracy and thus make the creation of an automatic diagnostic system for UM viable.

As a future work in the configuration, the geometric Hu's invariant moments can be used as an experimental method to perform a configuration and pre-initialization of the convolutional neural network, as well as the use of different architectures under this scheme as an alternative of transfer learning. On the other hand, the theoretical basis and the proof of concept to perform the detection of cancerous regions is shown in this paper; therefore, future engineering work can evaluate the computational complexity of the implemented strategies so that an efficient implementation can be established in the system.

Regarding technologies, it is recommended the use of protocol buffers (Protobuf) to transfer the images from client to server, the use of GPU with NVIDIA Triton Server which helps to standardize the deployment of the model and automatically optimizes the kernels and converts them into code for execution on NVIDIA GPUs, the use of Fast Data Preprocessing with NVIDIA DALI which provides a set of highly optimized building blocks and a runtime engine to accelerate the input data preprocessing for deep learning applications like the one of this study. Besides, it is suggested to use dynamic batching, which combines multiple requests in the same execution so that the model can provide larger performance.

This study has shown two architectures with three configurations, each for a total of six models tested under cross-validation and configurations specially designed to address the problems seen in the literature. In particular, the notable results to highlight are ResNet-18 with accuracy 98% in configuration I, VGG-16 96.2% and 99% in configuration II and III. This study may serve to students and researchers as reference to design diagnosis systems of anomalies as ocular cancer.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We wish to thank Dr. Paul T. Finger, MD, the New York Eye Cancer and the Institute of Automation of the Chinese Academy of Sciences for providing the data required to conduct this study. We would like to thank Universidad Distrital Francisco José de Caldas, École Nationale Supérieure Mines-Télécom Atlantique

Bretagne-Pays and University of Nantes for encouraging research. The study was self-funded.

References

- [1] L. Abujamous, A. Tbakhi, M. Odeh, O. Alsmadi, F.F. Kharbat, H. Abdel-Razeq, Towards digital cancer genetic counseling, in: 2018 1st International Conference on Cancer Care Informatics (CCI), 2018, pp. 188–194. URL: <https://doi.org/10.1109/CANCERCARE.2018.8618229>.
- [2] S. Zhang, J. Zhou, Z. Hu, A. Nair, L. Tang, Nanoparticles for uveal melanoma treatment, in: 2008 8th IEEE Conference on Nanotechnology, 2008, pp. 822–825. URL: doi: 10.1109/NANO.2008.245.
- [3] H. Luo, C. Ma, A novel ferroptosis-associated gene signature to predict prognosis in patients with uveal melanoma, *Diagnostics* 11 (2). URL: doi: 10.3390/diagnostics11020219.
- [4] D. Ball, A. Mitchell, D. Giroux, R. Rami-Porta, Effect of tumor size on prognosis in patients treated with radical radiotherapy or chemoradiotherapy for non-small cell lung cancer: An analysis of the staging project database of the international association for the study of lung cancer, *J. Thoracic Oncol.* 8 (3) (2013) 315–321, <https://doi.org/10.1097/jto.0b013e31827dc74d>.
- [5] B.A. Krantz, N. Dave, K.M. Komatsubara, B.P. Marr, R.D. Carvajal, Uveal melanoma: epidemiology, etiology, and treatment of primary disease, *Clin. Ophthalmol.* 11 (2017) 279–289. URL: <https://doi.org/10.2147/OPTH.S89591>.
- [6] A.M. Lane, I.K. Kim, E.S. Gragoudas, Survival rates in patients after treatment for metastasis from uveal melanoma, *JAMA Ophthalmol.*
- [7] American Cancer Society, Eye cancer (ocular melanoma), <https://www.cancer.org/cancer/eyecancer/detection-diagnosis-staging/survival-rates.html#references>, (Accessed on: 01-02-2021).
- [8] H.R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, R.M. Summers, Improving computer-aided detection using convolutional neural networks and random view aggregation, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1170–1181. URL: doi: 10.1109/TMI.2015.2482920.
- [9] P.C. Pearlman, R. Divi, M. Gwede, P. Tandon, B.S. Sorg, M.R. Ossandon, L. Agrawal, V. Pai, H. Baker, T.B. Lash, The national institutes of health affordable cancer technologies program: Improving access to resource-appropriate technologies for cancer detection, diagnosis, monitoring, and treatment in low- and middle-income countries, *IEEE J. Transl. Eng. Health Med.* 4 (2016) 1–8. URL: <https://doi.org/10.1109/JTEHM.2016.2604485>.
- [10] N. Khosravan, U. Bagci, Semi-supervised multi-task learning for lung cancer diagnosis, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 710–713. URL: <https://doi.org/10.1109/EMBC.2018.8512294>.
- [11] J. Wu, X. Liu, X. Zhang, Z. He, P. Lv, Master clinical medical knowledge at certificated-doctor-level with deep learning model, *Nature Communications* 9 (1). URL: doi: 10.1038/s41467-018-06799-6.
- [12] K. Zhang, X. Liu, J. Xu, J. Yuan, W. Cai, T. Chen, K. Wang, Y. Gao, S. Nie, X. Xu, X. Qin, Y. Su, W. Xu, A. Olvera, K. Xue, Z. Li, M. Zhang, X. Zeng, C.L. Zhang, O. Li, E.E. Zhang, J. Zhu, Y. Xu, D. Kermany, K. Zhou, Y. Pan, S. Li, I.F. Lai, Y. Chi, C. Wang, M. Pei, G. Zang, Q. Zhang, J. Lau, D. Lam, X. Zou, A. Wumaier, J. Wang, Y. Shen, F. Hou, P. Zhang, T. Xu, Y. Zhou, G. Wang, Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images, *Nature Biomedical Engineering* 5 (6) (2021) 533–545. URL: doi: 10.1038/s41551-021-00745-6.
- [13] J. Kim, J. Lee, E. Park, J. Han, A deep learning model for detecting mental illness from user content on social media, *Sci. Rep.* 10 (1). URL: doi: 10.1038/s41598-020-68764-y.
- [14] H. Sun, E. Kimchi, O. Akeju, S.B. Nagaraj, L.M. McClain, D.W. Zhou, E. Boyle, W.L. Zheng, W. Ge, M.B. Westover, Automated tracking of level of consciousness and delirium in critical illness using deep learning, *npj Digital Med.* 2 (1). URL: doi: 10.1038/s41746-019-0167-0.
- [15] M.C. Guerrero, J.S. Parada, H.E. Espitia, EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks, *Heliyon* 7 (6) (2021), <https://doi.org/10.1016/j.heliyon.2021.e07258> e07258.
- [16] P. Perner, Iris acquisition and detection for computer-assisted iridology, in: Signal Processing and Communications Applications Conference (SIU), in: 2014 22nd, 2014, pp. 2291–2295. URL: <https://doi.org/10.1109/SIU.2014.6830722>.
- [17] S. Amerifar, A.T. Targhi, M.M. Dehshibi, Iris the picture of health: Towards medical diagnosis of diseases based on iris pattern, in: 2015 Tenth International Conference on Digital Information Management (ICDIM), 2015, pp. 120–123. URL: <https://doi.org/10.1109/ICDIM.2015.7381861>.
- [18] D.F. Santos, Chaos-based digital image encryption using unique iris features, *Int. J. Appl. Eng. Res.* 15 (4) (2020) 358, <https://doi.org/10.37622/ijaer/15.4.2020.358-363>.
- [19] D.P.G. Durga-Devi, Disease identification in iris using gabor filter, *Int. J. Eng. Comput. Sci.* 3 (2014) 5369–5399.
- [20] F. Ahmad, M.U. Ghani Khan, K. Javed, Deep learning model for distinguishing novel coronavirus from other chest related infections in x-ray images, *Comput. Biol. Med.* 134 (2021). URL: <https://doi.org/10.1016/j.compbiomed.2021.104401> 104401.
- [21] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM* 60 (6) (2017) 84–90.
- [22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE

- Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9. URL: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [23] T. Choudhary, V. Mishra, A. Goswami, J. Sarangapani, A transfer learning with structured filter pruning approach for improved breast cancer classification on point-of-care devices, *Comput. Biol. Med.* 134 (2021). URL: <https://doi.org/10.1016/j.combiomed.2021.104432>.
- [24] M. Desai, M. Shah, An automatization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (mlp) and convolutional neural network (cnn), *Clinical eHealth* 4 (2021) 1–11. URL: <https://doi.org/10.1016/j.ceh.2020.11.002>.
- [25] S. Kaliki, C.L. Shields, Uveal melanoma: relatively rare but deadly cancer, *Eye* 31 (2) (2016) 241–257. URL: [10.1038/eye.2016.275](https://doi.org/10.1038/eye.2016.275).
- [26] H. Migallón, A. Belazi, J.L. Sánchez-Romero, H. Rico, A. Jimeno- Morenilla, Settings-free hybrid metaheuristic general optimization methods, *Mathematics* 8 (7). URL: [10.3390/math8071092](https://doi.org/10.3390/math8071092).
- [27] D. Santos, L. Dallos, P.A. Gaona-García, Algoritmos de rastreo de movimiento utilizando técnicas de inteligencia artificial y machine learning, *Información tecnológica* 31 (3) (2020) 23–38. URL: [10.4067/s0718-07642020000300023](https://doi.org/10.4067/s0718-07642020000300023).
- [28] D.F. Santos-Bustos, H.E. Espitia-Cuchango, Hepatitis diagnosis using optimized KD-trees and neural networks, *Int. J. Eng. Res. Technol.* 13 (9) (2020) 2269. URL: <https://doi.org/10.37624/ijert/13.9.2020.2269-2274>.
- [29] M. Martin, B. Sciolla, M. Sdika, P. Quélin, P. Delachartre, Segmentation of neonates cerebral ventricles with 2d cnn in 3d us data: suitable training-set size and data augmentation strategies, in: 2019 IEEE International Ultrasonics Symposium (IUS), 2019, pp. 2122–2125. URL: <https://doi.org/10.1109/ULTSYM.2019.8925799>.
- [30] M. Cavusculu, I.S. Yetik, M. Yeginer, Fully automated blink detection for uveal melanoma radiotherapy, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017. URL: <https://doi.org/10.1109/embc.2017.8036903>.
- [31] D.F. Santos, H.E. Espitia, Detection of uveal melanoma using fuzzy and neural networks classifiers, *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 18 (4) (2020) 2213. URL: <https://doi.org/10.12928/telkomnika.v18i4.14228>.
- [32] D.F. Santos, H.E. Espitia, Proposal for a neuro-fuzzy system for uveal melanoma detection, *ARPN J. Eng. Appl. Sci. (JEAS)* 16 (4). URL: http://www.arpnjournal.org/jeas/research_papers/rp_2021/jeas_0221_8519.pdf.
- [33] K. Maharana, S. Mondal, B. Nemade, A review: Data preprocessing and data augmentation techniques, *Globel Transitions Proceedings*. doi: 10.1016/j.gltp.2022.04.020. URL: <https://doi.org/10.1016/j.gltp.2022.04.020>.
- [34] A. Azarang, J. Hansen, N. Kehtarnavaz, Combining data augmentations for cnn-based voice command recognition, in: 2019 12th International Conference on Human System Interaction (HSI), 2019, pp. 17–21. URL: <https://doi.org/10.1109/HSI47298.2019.8942638>.
- [35] Y.D. Zhang, Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, S.H. Wang, Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation, *Multimed. Tools Appl.* 78 (2022) 3613–3632. URL: <https://doi.org/10.1007/s11042-017-5243-3>.
- [36] S.H. Wang, Y. Zhang, X. Cheng, X. Zhang, Y.D. Zhang, Psspn: Patchshuffle stochastic pooling neural network for an explainable diagnosis of covid-19 with multiple-way data augmentation, *Comput. Math. Methods Med.* (2021). URL: <https://doi.org/10.1155/2021/6633755>.
- [37] E. David, P. Ungureanu, L. Goras, On the feature extraction performances of cnn gabor-type filters in texture recognition applications, in: 2006 10th International Workshop on Cellular Neural Networks and Their Applications, 2006, pp. 1–6. URL: <https://doi.org/10.1109/CNNA.2006.341621>.
- [38] B.S. Manjunath, W.Y. Ma, Texture features for browsing and retrieval of image data, *IEEE Trans. Pattern Anal. Mach. Intell.* 18 (8) (1996) 837–842. URL: <https://doi.org/10.1109/34.531803>.
- [39] N. Dvornik, J. Mairal, C. Schmid, On the importance of visual context for data augmentation in scene understanding, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (6) (2021) 2014–2028. URL: <https://doi.org/10.1109/tpami.2019.2961896>.
- [40] A. Mikolajczyk, M. Grochowski, Data augmentation for improving deep learning in image classification problem, in: 2018 International Interdisciplinary PhD Workshop (IIPHDW), 2018, pp. 117–122. URL: <https://doi.org/10.1109/IIPHDW.2018.8388338>.
- [41] Keras: the python deep learning api, <https://keras.io/>, (Accessed on: 01-03-2021).
- [42] R. Roslan, N. Jamil, Texture feature extraction using 2-d gabor filters, in: 2012 International Symposium on Computer Applications and Industrial Electronics (ISCAIE), IEEE, 2012. doi: 10.1109/iscaie.2012.6482091. URL: <https://doi.org/10.1109/iscaie.2012.6482091>.
- [43] T.P. Weldon, Gabor filter design for multiple texture segmentation, *Opt. Eng.* 35 (10) (1996) 2852. URL: <https://doi.org/10.1117/1.600971>.
- [44] A. Ramakrishnan, S. Raja, H. Ram, Neural network-based segmentation of textures using gabor features, in: 12th IEEE Workshop on Neural Networks for Signal Processing, 2002, pp. 365–374. URL: <https://doi.org/10.1109/NNSP.2002.1030048>.
- [45] M. Haghghat, S. Zonouz, A.M.M., Identification Using Encrypted Biometrics, Wilson R., Hancock E., Bors A., Smith W. (eds) *Computer Analysis of Images and Patterns*. CAIP 2013. Lecture Notes in Computer Science, vol 8048. Springer, Berlin, Heidelberg, 2013. URL: [10.1007/978-3-642-40246-3_55](https://doi.org/10.1007/978-3-642-40246-3_55).
- [46] M. Bertalmio, A. Bertozzi, G. Sapiro, Navier-stokes, fluid dynamics, and image and video inpainting, in: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001, IEEE Comput. Soc. doi: 10.1109/cvpr.2001.990497. URL: <https://doi.org/10.1109/cvpr.2001.990497>.
- [47] R. Yamashita, M. Nishio, R. Do, K. Togashi, Convolutional neural networks: an overview and application in radiology, *Insights into Imaging* 9. doi: 10.1007/s13244-018-0639-9.
- [48] P. Vecchiotti, F. Vesperini, E. Principi, S. Squartini, F. Piazza, Convolutional neural networks with 3-d kernels for voice activity detection in a multiroom environment, in: *Multidisciplinary Approaches to Neural Computing*, Springer International Publishing, 2017, pp. 161–170. URL: https://doi.org/10.1007/978-3-319-56904-8_16.
- [49] S. Albawi, T. Abed Mohammed, S. ALZAWI, Understanding of a convolutional neural network, in: 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1–6. URL: <https://doi.org/10.1109/ICEngTechnol.2017.8308186>.
- [50] C. Luo, X. Li, L. Wang, J. He, D. Li, J. Zhou, How does the data set affect cnn-based image classification performance?, in: 2018 5th International Conference on Systems and Informatics (ICSAI), 2018, pp. 361–366. URL: <https://doi.org/10.1109/ICSAI.2018.8599448>.
- [51] G. Li, Z. Liu, F. Li, J. Cheng, Block convolution: Toward memoryefficient inference of large-scale CNNs on FPGA, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 41 (5) (2022) 1436–1447. URL: <https://doi.org/10.1109/tcad.2021.3082868>.
- [52] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Y. Bengio, Y. LeCun (Eds.), *International Conference on Learning Representations*, 2015. URL: <http://arxiv.org/abs/1409.1556>.
- [53] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G. Jamalipour Soufi, Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning, *Med. Image Anal.* 65 (2020). URL: <https://doi.org/10.1016/j.media.2020.101794>.
- [54] H. Malik, M.S. Farooq, A. Khelifi, A. Abid, J.N. Qureshi, M. Hussain, A comparison of transfer learning performance versus health experts in disease diagnosis from medical imaging, *IEEE Access* 8 (2020) 139367–139386. URL: <https://doi.org/10.1109/access.2020.3004766>.
- [55] A. Abbas, M.M. Abdelsamea, M. Medhat Gaber, Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network, *medRxiv*. URL: <https://doi.org/10.1101/2020.03.30.20047456>.
- [56] S.H. Wang, Q. Zhou, M. Yang, Y.D. Zhang, Advian: Alzheimer's disease vgg-inspired attention network based on convolutional block attention module and multiple way data augmentation, *Front. Aging Neurosci.* 13. doi: 10.3389/fnagi.2021.687456. URL: <https://www.frontiersin.org/article/10.3389/fnagi.2021.687456>.
- [57] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. URL: <https://doi.org/10.1109/CVPR.2016.90>.
- [58] M. Robnik-Sikonja, I. Kononenko, Theoretical and empirical analysis of relieff and relieff, *Mach. Learn.* 53 (1/2) (2003) 23–69. URL: [10.1023/a:1025667309714](https://doi.org/10.1023/a:1025667309714).
- [59] H. Qassim, A. Verma, D. Feinzimer, Compressed residual-vgg16 cnn model for big data places image recognition, in: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 169–175. URL: <https://doi.org/10.1109/CCWC.2018.8301729>.
- [60] Z. Song, L. Fu, J. Wu, Z. Liu, R. Li, Y. Cui, Kiwifruit detection in field images using faster r-cnn with vgg16, *IFAC-PapersOnLine* 52 (30) (2019) 76–81, 6th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2019. URL: [10.1016/j.ifacol.2019.12.500](https://doi.org/10.1016/j.ifacol.2019.12.500).
- [61] Y. Chen, X. Zhang, D. Li, J. Jin, Y. Shen, Classification of a small data-set set thyroid nodules based on supplementary feature layer improved vgg16, in: 2020 39th Chinese Control Conference (CCC), 2020, pp. 7316–7321. URL: <https://doi.org/10.23919/CCC50068.2020.9188671>.
- [62] K.S. Chandrasekar, P. Geetha, Multiple objects tracking by a highly decisive three-frame differencing-combined-background subtraction method with gmpfm-gmpfd filters and vgg16- lstm classifier, *J. Visual Commun. Image Represent.* 72 (2020). URL: <https://doi.org/10.1016/j.jvcir.2020.102905>.
- [63] Z. Khan, F.G. Khan, A. Khan, Z.U. Rehman, S. Shah, S. Qummar, F. Ali, S. Pack, Diabetic retinopathy detection using VGG-NIN a deep learning architecture, *IEEE Access* 9 (2021) 61408–61416. URL: <https://doi.org/10.1109/access.2021.3074422>.
- [64] A.K. Rangarajan, R. Purushothaman, Disease classification in eggplant using pre-trained VGG16 and MSVM, *Sci. Rep.* 10 (1). doi: 10.1038/s41598-020-59108-x. URL: <https://doi.org/10.1038/s41598-020-59108-x>.
- [65] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaria, M.A. Fadel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, *J. Big Data* 8 (1). doi: 10.1186/s40537-021-00444-8. URL: <https://doi.org/10.1186/s40537-021-00444-8>.
- [66] S. Ayyachamy, V. Alex, M. Khened, G. Krishnamurthi, Medical image retrieval using Resnet-18, in: P.H. Chen, P.R. Bak (Eds.), *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, vol. 10954, International Society for Optics and Photonics, SPIE, 2019, pp. 233–241. URL: <https://doi.org/10.1117/12.2515588>.
- [67] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: *International Conference for Learning Representations*, 2014. URL: <http://arxiv.org/abs/1412.6980>.

- [68] F. Chaumette, Image moments: A general and useful set of features for visual servoing, *IEEE Trans. Robot.* 20 (4) (2004) 713–723, <https://doi.org/10.1109/tro.2004.829463>.
- [69] B. Guner, M. Frankford, J.T. Johnson, On the shapiro-wilk test for the detection of pulsed sinusoidal radio frequency interference, in: *IGARSS 2008–2008 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2008, <https://doi.org/10.1109/igarss.2008.4778951>.
- [70] R. D'Agostino, E.S. Pearson, Tests for departure from normality. empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika* 60 (3) (1973) 613, <https://doi.org/10.2307/2335012>.
- [71] E. Theodorsson-Norheim, Kruskal-wallis test: BASIC computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples, *Comput. Methods Programs Biomed.* 23 (1) (1986) 57–62, [https://doi.org/10.1016/0169-2607\(86\)90081-7](https://doi.org/10.1016/0169-2607(86)90081-7).
- [72] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation, *BMC Genom.* 21(1). doi: 10.1186/s12864-019- 6413-7. URL: <https://doi.org/10.1186/s12864-019-6413-7>.
- [73] New York Eye Cancer Center, Iris tumors, <https://eyecancer.com/eye-cancer/image-galleries/iris-tumors/>, (Accessed on: 01-02- 2021).
- [74] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde- Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, *Adv. Neural Inform. Process. Syst.* 3. doi: 10.1145/3422622.