Probabilistic combination of non-linear eigenprojections for ensemble classification

Juan E Arco[a,b,c,*], Andrés Ortiz[b,c], Javier Ramírez[a,c], Francisco J. Martínez-Murcia[a,c], Yu-Dong Zhang[d], Jordi Broncano[e], Álvaro Berbís[e], Javier Royuela-del-Val[e], Antonio Luna[f], Juan M. Górriz[a,c]

[a]*Department of Signal Theory, Networking and Communications, University of Granada, 18010 Granada, Spain*
[b]*Department of Communications Engineering, University of Malaga, 29010 Malaga, Spain*
[c]*DaSCI Institute, University of Granada, Spain*
[d]*School of Informatics, University of Leicester, Leicester, LE1 7RH, UK*
[e]*Department of Radiology, Hospital San Juan de Dios, HT Médica, 14012 Spain*
[f]*Department of Radiology, Clínica Las Nieves, HT Médica, 23007 Spain*

Abstract

The emergence of new technologies has changed the way clinicians perform diagnosis. Medical imaging play a crucial role in this process, given the amount of information that they usually provide as non-invasive techniques. Despite the high quality offered by these images and the expertise of clinicians, the diagnostic process is not a straightforward task since different pathologies can have similar signs and symptoms. For this reason, it is extremely useful to assist this process with the inclusion of an automatic tool that reduces the bias when analyzing this kind of images. In this work, we propose an ensemble classifier based on probabilistic Support Vector Machine (SVM) in order to identify relevant patterns while providing information about the reliability of the classification. Specifically, each image is divided into patches and features contained in each one of them are extracted by applying kernel PCA. The use of base classifiers within an ensemble allows our system to identify the informative patterns regardless of their size or location. Decisions of each individual patch are then combined according to the reliability of each individual classification: the lower the uncertainty, the higher the contribution. Performance is evaluated in a real

*Corresponding author

scenario where distinguishing between pneumonia patients and controls from chest Computed Tomography (CCT) images, yielding an accuracy of 97.86%. The large performance obtained and the simplicity of the system (use of deep learning in CCT images would highly increase the computational cost) evidence the applicability of our proposal in a real-world environment.
Keywords: Computer-aided diagnosis, Medical imaging, Probabilistic machine learning, uncertainty, Pneumonia.

---

1. Introduction

Medical imaging play a crucial role in the clinical practice worldwide. As a non-invasive technique, they provide crucial information for the diagnosis of a wide range of diseases. Despite the high spatial resolution that current medical imaging provide, performing a correct diagnosis is not a straightforward task. Success depends on factors such as the expertise of the doctors or the overlapping between symptoms associated with similar pathologies. This leads to a manual, time-consuming process that may delay diagnosis and the election of a correct treatment. Previous studies have employed machine learning methods for the automatic detection of diseases. Some of them have been used for delimitating the regions of interest as an initial step of the classification pipeline [1, 2, 3], whereas others have been successfully employed in the classification stage [4, 5, 6]. The emergence of deep neural networks has revolutionized the automatic classification of medical images. These alternatives provide an excellent performance both as a feature extractor and in the classification stage of a wide range of pathologies [7, 8, 9]. However, the use of too much complex alternatives can degrade performance in contexts with a high amount of information. This complexity is partially alleviated in scenarios where two-dimensional images are employed, but when using three-dimensional ones, the computational burden highly increases.

One solution is to select only one slice, constraining the images to have two dimensions. [10] successfully employed this alternative within a transfer learning

setting and discriminant correlation analysis. [11] also used it in combination with DenseNet, leading to a high classification performance. However, this extreme dimensionality reduction leads to a vast loss in the spatial information that this kind of images provide. In fact, this is essentially the advantage of 3D images over 2D ones: higher resolution and volumetric data. Thus, eliminating the third dimension can drastically mitigate the potential of three-dimensional images. Another important issue is the way the target slice is selected from the volumetric image, since it is not a trivial task. One possibility is to choose the slice displaying the largest number of lesions. However, this process requires clinicians to do it manually, which is not ideal when trying to automatize the diagnosis. It seems clear that the simplest solution is to employ the whole 3D image and not selecting only one slice, but this entails a high computational cost. This workload would be even worse when using deep learning and three-dimensional convolutions for feature extraction. Besides, training a deep learning model requires a considerable amount of data, especially when the number of features is high like in 3D images. Thus, it is necessary to find a solution that leads to a high classification performance while mitigating the computational load associated with the processing of volumetric data.

In this work, we employ an ensemble classifier based on probabilistic SVM in order to identify relevant patterns while providing information about the reliability of the classification. In particular, each three-dimensional image is divided into a number of cubic patches. Features contained in each one of them are extracted by applying kernel PCA, and the most informative components are then entered into an RBF-SVM classifier. The use of base classifiers within an ensemble allows our system to identify the informative patterns regardless of their size or location. Decisions of each individual patch are then combined according to the reliability of each individual classification: the lower the uncertainty, the higher the contribution, and vice versa. Performance is evaluated in a real scenario where distinguishing between controls *vs* pneumonia patients.

The method presented in this work successfully employs machine learning from a probabilistic perspective, guiding the ensemble classification according

to the uncertainty of individual predictions. This allows to strike a balance between performance and the computational burden, which is especially relevant when processing three-dimensional images. The rest of the paper is organized as follows. Section 2 summarizes similar works developed for image classification. Section 3 provides a complete explanation of the method presented in this work. First, the kernel PCA is described, whereas the probabilistic SVM is also detailed. After that, it is explained how the different base classifiers are fused within the ensemble framework. Afterwards, the applicability of the proposed method is assessed in the detection of pneumonia associated with COVID-19. The database employed for that purpose is explained in Section 4, in addition to the preprocessing pipeline and the three experiments conducted. Results are described in Section 5 and discussed in Section 6. Finally, conclusions and future lines of research are contained in Section 7.

2. Related works

The creation of intelligent systems for the automation of image classification is commonly used in a wide range of scenarios. Algorithms based on machine learning have been successfully used in the automotive industry in order to ensure the integrity and quality of different components [12]. These approaches have also been employed in other applications such as the automatic evaluation of the porosity of different materials [13], in addition to the prediction of the production capacity of hydropower industries [14]. Other alternatives in the field of computer vision have focused on the prediction and management of traffic flow from cameras located in roads [15]. This work combines a mixture of Gaussian (MOG) modeling for background removal with a transfer learning approach to detect the different objects in the video images. This allows tracking the position of each pedestrian and bicycle, predicting their trajectory and the risk of collision with another vehicle. Another work focused on the analysis of satellite images for improving the agricultural production [16]. This alternative employed an expanded version of Randomized Quasi-Exhaustive (RQE) to extract a set

of interdependent features from integral images. Then, a classifier based on random forest was used to compute the posterior probability that each pixel belongs to each agricultural field. [17] presented a framework for predicting a tennis game outcome by analyzing the players performance. They employed a deep neural network for the extraction of high-level statistics, in addition to recognize fine-grained tennis actions based on an Inception architecture [18].

Ensemble classification has also been used for image analysis, especially in contexts where its performance is larger than the one obtained by individual classifiers. Numerous works in literature have combined the decision of a number of base classifiers in order to improve the modeling of some data. [19] proposed a variant of ensemble architecture based on random forest in order to detect and classify epileptic seizure. The EEG data was entered into all the classifiers within the ensemble, obtaining the final decision as a combination of individual outputs through a majority voting. Similarly, [20] presented a two-stage ensemble of deep architectures for melanoma classification. Skin lesions were previously segmented by an U-sharped architecture. Afterwards, the preprocessed images were entered into five state-of-the-art networks. The final decision was computed by combining each individual network through its output probability, so that a network with a high output probability had more influence than one with a lower probability. [20] developed a robust quality estimation for ensemble architectures in segmentation contexts. In this case, they employed the SIMPLE [21] method in order to combine the outputs of the individual classifiers within the ensemble. This approach is based on an iterative procedure that combines atlas selection and the performance of propagated segmentation as a weight in the fusion process. Results showed a better performance than the classical majority voting.

With reference to probabilistic approaches, previous studies have used them for different purposes. [22] employed these alternatives as a feature transformation method. Specifically, they entered features in their original space into a probabilistic Support Vector Machines (SVM) classifier, leading to new features in a transformed space. These outputs were then combined by using an

adaptive similarity fusion, yielding a high performance in the classification of medical images. [23] proposed a Bayesian classification framework based on a local probabilistic model for image classification. The idea behind Bayesian classifiers is that they offer a measure of the risk of each classification decision, which is considerably more informative than providing the class label itself. The main finding in [23] is that probability distribution for each class is simpler and more accurate when computed in a local sample space than when models are optimized in the whole sample space. Following a similar approach, [24] proposed a probabilistic model for simultaneously classifying and annotating images from different sports, whereas [25] presented a method based on the expectation–maximization (EM) algorithm to refine the final annotation of the images.

3. Methodology

3.1. Kernel Principal Component Analysis

One of the main challenging problems in classification is related to the small sample size problem [26], which occurs when datasets are formed by high-dimensional data but with a small number of samples. Unfortunately, it is not uncommon to find large differences between the number of features and samples, so that finding a solution that alleviates this issue is crucial. Principal Component Analysis (PCA) is a multivariate approach that has been widely used to reduce the dimensionality of the data [27, 28, 29]. This method attempts to find a linear subspace with a lower dimensionality than the original space. Given a set of N samples $\mathbf{x}_k$, $\mathbf{x}_k = [\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kn}] \in \mathbb{R}^n$, the aim of PCA is to find the projection directions that maximize the variance of a subspace [30]. This is equivalent to compute the eigenvalues from the covariance matrix. There are some occasions in which features can not be linearly extracted. In kernel PCA [31, 32], vector $\mathbf{x}$ is projected from the input space, $\mathbb{R}^n$, to a high-dimensional space, $\mathbb{R}^f$, by applying a non-linear mapping function $\mathbf{\Phi} : \mathbb{R}^n \rightarrow \mathbb{R}^f, f > n$. In

the new feature space, $\mathbb{R}^f$, the eigenvalue problem can be described as follows:

$$C^{\Phi}\mathbf{w}^{\Phi} = \lambda\mathbf{w}^{\Phi} \tag{1}$$

where C^{Φ} is a covariance matrix. All the solutions $\mathbf{w}^{\Phi}$ with $\lambda \neq 0$ are in the transformed space $\Phi(\mathbf{x}_1, \ldots, \Phi(\mathbf{x}_N))$, and there exist coefficients α_i such that:

$$\mathbf{w}^{\Phi} = \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}_i) \tag{2}$$

Defining an NxN matrix K by

$$K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \tag{3}$$

the PCA problem becomes:

$$N\lambda K\boldsymbol{\alpha} = K^2\boldsymbol{\alpha} \equiv N\lambda\boldsymbol{\alpha} = K\boldsymbol{\alpha} \tag{4}$$

where $\boldsymbol{\alpha}$ denotes a column vector with entries $\alpha_1 \ldots \alpha_N$ [32].

A nonlinear version of PCA is obtained when using a nonlinear kernel such as the radial basis function (RBF), defined as follows:

$$k(\mathbf{x}_i, \mathbf{x}_j) = exp\left(\frac{-0.5\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{\sigma^2}\right) \tag{5}$$

Finally, vectors in the high-dimensional feature space are projected into a lower dimensional spanned by the eigenvectors $\mathbf{w}^{\Phi}$. Given a sample $\mathbf{x}$ whose projection is $\Phi(\mathbf{x})$ in $\mathbb{R}^f$, the projection of $\Phi(\mathbf{x})$ onto the eigenvectors $\mathbf{w}^{\Phi}$ is the nonlinear principal components corresponding to Φ, as follows:

$$\mathbf{w}^{\phi} \cdot \Phi(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i(\Phi(\mathbf{x}_i)\Phi(\mathbf{x})) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}) \tag{6}$$

The dominant eigenvectors of the covariance matrix described in Equation 1 span a new subspace. Then, the number of eigenvectors that explained 90% of the total variance are selected for the subsequent classification step.

3.2. Classifier

The resulting eigenvectors were then entered into the classification stage, which was based on an SVM classifier with RBF kernel [33]. This alternative can be mathematically described as follows:

$$r_i = sign\Big(\sum_{i=1}^{N_{sv}} \alpha_i y_i K(\mathbf{x}_i, \mathbf{z}_j) + b\Big) \tag{7}$$

where r_i is the classification response for sample i; N_{sv} is the number of support vectors; α_i is the Lagrange multiplier; y_i is the class membership of sample i; $K(\mathbf{x}_i, \mathbf{z}_j)$ is the kernel function and b is the bias parameter [34]. In order to mitigate the effect of class imbalance, we incorporated the weights of the classes into the cost function of the SVM in order to assign to each sample a different relevance in the classification decision [35, 36]. This means that samples from the majority class had a lower influence in the penalty term than the ones from the minority class.

The output of the classifier informs us about the class each sample belongs to. When classifying medical images it is particularly convenient to know not only if a patient suffers or not a disease but a degree of certainty of the prediction. However, standard SVMs do not provide any additional information of the predictions. [37] proposed a method for mapping the outputs of SVMs to probabilities. This was based on the decomposition of the feature space into a direction orthogonal to the separating hyperplane and the rest of dimensions of the feature space. Despite its good performance, this approach requires a linear solution for every evaluation of the SVM. [38] suggested an alternative based on training a logistic regression model on the classifier outputs in order to transform them into a probability distribution. The posterior probability can be defined as follows:

$$P(y = 1|f) = \frac{1}{1 + exp(Af + B)} \tag{8}$$

The parameters A and B are fit using maximum likelihood estimation from a

training set (f_i, y_i). Let (f_i, t_i) a new training set, where t_i are target probabilities defined as:

$$t_i = \frac{y_i + 1}{2} \tag{9}$$

Parameters can be estimated by minimizing the negative log likelihood of the training data, which is a cross-entropy error function:

$$min\Big[-\sum_i t_i \, log(p_i) + (1 - t_i) \, log(1 - p_i) \Big] \tag{10}$$

where

$$p_i = \frac{1}{1 + exp(Af_i + B)} \tag{11}$$

We employed Equation 8 for computing the probability that a test sample belonged to a specific class. However, this measure itself doest not quantify the uncertainty of the prediction. To do so, we used a Bootstrap method following a random sampling with replacement scheme [39, 40]. This process consists on selecting part of the training set (80% in our case), training the SVM classifier and computing the posterior probability of each test sample. Then, a new subset is randomly picked up from the training images and computed again the posterior probability of the test sample. This operation was repeated 500 times in order to build a distribution of probabilities. Finally, the uncertainty for a specific test sample k was computed as the variance of the posterior probabilities, as follows:

$$u_i = \frac{\sum_{i=1}^{K} (x_i - \mu)^2}{K} \tag{12}$$

where x_i is the i-th element of the probabilities distribution x, μ is the mean of the distribution and K is the number of times that the Boostrapping process is repeated. Section 3.3 provides a detailed explanation about the use of uncertainty in our classification framework.

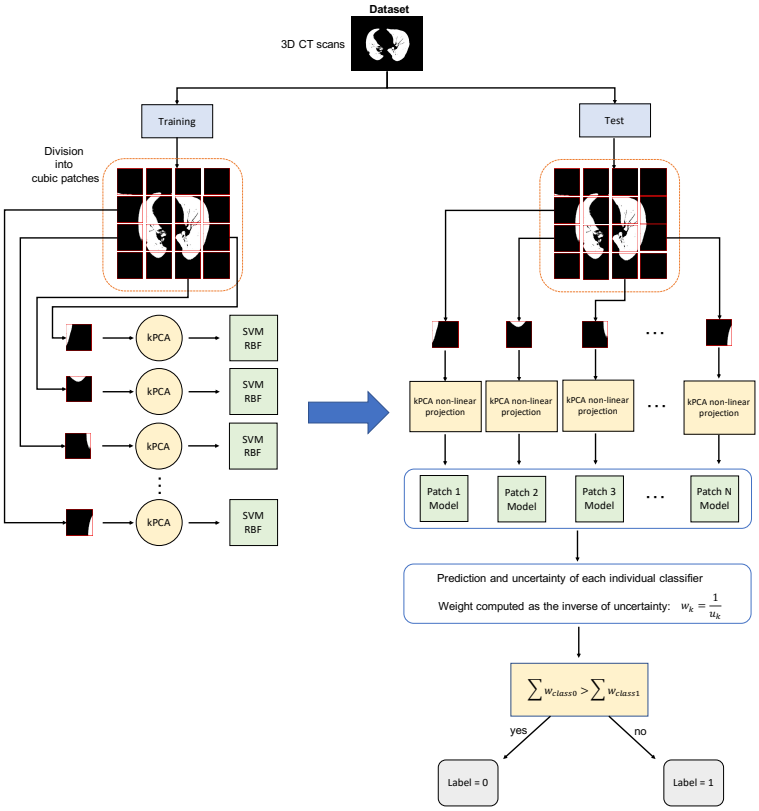

Figure 1: Schema of the classification system proposed in this work.

3.3. Ensemble Classification

The development of a CAD (Computer-aided diagnosis) system for the detection of a certain disease relies on the assumption that patterns associated with this pathology are similar among patients. However, the signs and symptoms may differ depending on the virulence of the disease. The presence of other abnormal findings that are not related to the disorder can also affect the reliability of the classification system, in addition to the artifacts during the acquisition of the images. To overcome these potential pitfalls, we propose an ensemble framework in which different classifiers analyze local regions of the images. Afterwards, the decisions of each classifier are combined according to the reliability of each classifier's prediction. This means that images are first divided into different regions. When using three-dimensional images, they are divided into cubic patches [41]. For each individual patch, kernel PCA is applied, entering the resulting components into the classifier. The number (and hence, size) of the patches were selected in order to match the potential size of the informative patterns, guaranteeing to strike a balance between performance and computational cost. Finally, each individual classifier was then fused into a global one following a specific procedure.

Majority voting has been widely used as a way of combining the output of individual classifiers into a global decision [42, 43]. However, this is not the optimum choice to combine different classifiers decisions because some of them can be more reliable than others. In this work, we compute the weight associated with each patch according to the uncertainty derived from the posterior probabilities obtained by the SVM classifier. If the variance of the probabilities of a classifier in a specific prediction is high, its contribution to the final ensemble would be low, and viceversa [44]. Defining $u_l^k(\mathbf{y})$ as the uncertainty of the test sample $\mathbf{y}$ obtained from the k-th classifier corresponding to the l-th class, the empirical average of the l-th weights (inverse of uncertainties) over the K

classifiers can be calculated as follows:

$$E_l(\mathbf{y}) = \frac{\sum_{k=1}^{K} \frac{1}{u_l^k(\mathbf{y})}}{K} \tag{13}$$

The class label of the test sample $\mathbf{y}$ is then assigned to the class with the maximum average weight as:

$$Label(\mathbf{y}) = \arg\max_{l} E_l(\mathbf{y}) \tag{14}$$

Figure 1 shows a scheme of the ensemble classification framework proposed in this work.

4. Application to pneumonia detection

The following sections evaluate the applicability of the proposed method as a tool for identifying pneumonia from CT images. Specifically, we assessed the performance of our approach in order to detect the patterns associated with this pathology, distinguishing between controls and COVID-19 patients.

4.1. Database description

The dataset employed in this work was provided by HT Médica, a company specialized in radiology that offers innovative solutions for image diagnosis. The dataset comprises 513 CCT images, including 100 control patients and 413 characterized as depicting pneumonia associated with COVID-19. Controls are formed by healthy subjects and patients who have been diagnosed from atelectasis, bronchopneumonia, chronic obstructive pulmonary disease, emphysema and pneumothorax . All images were obtained as part of patient's routines clinical care during the first wave of the COVID-19 pandemic in Spain (March to June 2020). Data were anonymized before being used in this study following the requirements stated by medical ethics committees. Figure 2 shows a slice of a CCT scan from a control (CL) and a patient suffering from pneumonia (PNEU).

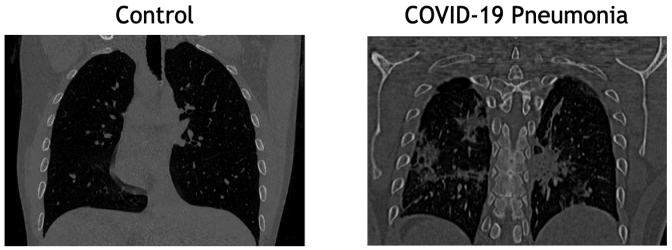

Figure 2: Slices of two CCT images of a control (left) and a pneumonia caused by COVID-19 (right). Note some clear artifacts in COVID-19 image.

4.2. Image preprocessing

When working with medical images, it is crucial to apply preprocessing in order to improve the subsequent classification performance. Therefore, this operation must adapt images to the requirements of the classification framework. Given the high computational and memory requirements of CCT volumes, we downsampled the input images to obtain a final map of 128x128x128. We also performed an automatic lung segmentation in order to separate voxels corresponding to lung tissues from those of the surrounding anatomy. To do so, we applied a histogram equalization to improve the contrast of the images by modifying the intensity distribution of the voxels in the image. After that, the Otsu's method [45] was employed for computing the threshold that separated target and non-target voxels. This alternative relies on the maximization of the between-class variance to separate the voxels belonging to the different classes.

After lung segmentation, the resulting images were registered employing the Elastix software [46]. This process consists on finding a coordinate transformation $T(\mathbf{x})$ that modifies a moving image $\mathbf{I}_M(\mathbf{x})$ to be aligned with a fixed image $\mathbf{I}_F(\mathbf{x})$. A transformation model $T_\mu(\mathbf{x})$, with parameters $\boldsymbol{\mu}$, can be formulated as an optimization problem in which a cost function C is minimised with respect to $\boldsymbol{\mu}$, as follows:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} C\left(T_\mu; I_F, I_M\right) \tag{15}$$

We employed the average of all the images in the database as the fixed image

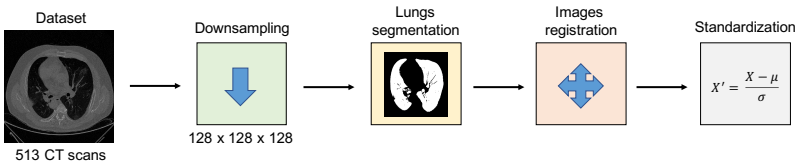

Figure 3: Diagram of the preprocessing steps.

, whereas each individual image was iteratively selected as the moving one. We used the Mean Square Difference (MSD) as the cost function in order to evaluate the similarity between the fixed and the moving image. This function is defined as follows:

$$MSD(\mathbf{T}_{\boldsymbol{\mu}}; I_F, I_M) = \frac{1}{N} \sum_{\mathbf{x} \in \Omega_F} (I_F(\mathbf{x}) - I_M(\mathbf{T}_{\boldsymbol{\mu}}(\mathbf{x})))^2 \tag{16}$$

where Ω_F denotes the fixed image domain and N the number of voxels $\mathbf{x}$ sampled from the domain of the fixed image. See [47] for a more detailed explanation. Finally, we performed an intensity normalization procedure for each individual image based on standardization. Each image was transformed such the resulting distribution had a zero mean and unit variance, as follows:

$$I' = \frac{I - \mu}{\sigma} \tag{17}$$

where I is the original image and I' is the resulting one. Figure 3 shows a scheme of all the stages of the preprocessing.

4.3. Performance Evaluation

We employed a Leave-One-Out cross-validation scheme to estimate the generalization ability of our method [48]. From the 513 CCT scans, 512 were employed to train the model, whereas the remaining one was used for testing. This was performed within an iterative process in which all scans were used at some point as the test sample. The performance of the classification framework was evaluated in terms of different metrics derived from the confusion matrix:

balanced accuracy, sensitivity, specificity, precision and F1-score, computed as follows:

$$
\begin{aligned}
Bal\ Acc = \frac{1}{2}\Big(\frac{TP}{P} + \frac{TN}{N}\Big) \quad Sens = \frac{T_P}{T_P + F_N} \\
Spec = \frac{T_N}{T_N + F_P} \quad Prec = \frac{T_P}{T_P + F_P} \\
F1 - score = \frac{2 \times Prec \times Sens}{Prec + Sens}
\end{aligned}
\tag{18}
$$

where T_P is the number of pneumonia patients correctly classified (true positives), P is the number of pneumonia patients, T_N is the number of controls correctly classified (true negatives), N is the number of controls, F_P is the number of controls classified as pneumonia (false positives) and F_N is the number of pneumonia patients classified as controls (false negatives). The area under the ROC curve was also employed as an additional measure of the classification performance [49, 50].

One crucial aspect is to evaluate the level of agreement of the different classifiers within the ensemble. To do so, we used a kappa-uncertainty diagram [51, 52]. This measure relies on Cohen's kappa coefficient [53], a statistic that compares an observed accuracy with an accuracy obtained by chance, providing a measure of how closely instances classified by a classifier match the ground truth [54]. Cohen's kappa can be mathematically defined as:

$$
k = \frac{p_A - p_E}{1 - p_E} \tag{19}
$$

where p_A is the observed relative agreeement between two annotators, and p_E is the probability of agreement by chance. Although acceptable kappa statistic values vary on the context, the closer to 1, the better the classification. Section 5 summarizes the kappa scores obtained by the different members of the ensemble, in addition to how accuracies of individual classifiers and kappa values are related.

Table 1: Summary of previous works focused on the automatic identification of pneumonia in addition to the best results obtained by our method.

Research work	Dataset	Method	Classification context	Results (%)
[55]	772 CT scans	MSANet	Normal *vs* Bacterial *vs* Viral *vs* COVID	Acc = 97.46
[56]	1000 CT scans	GAN model	Normal *vs* COVID	Acc = 99.95
[57]	1397 CT scans	DenseNet-121	Normal *vs* COVID	Acc = 90.80
[56]	4356 CT scans	COVNet	Normal *vs* COVID	AUC = 0.96
[58]	852 CT scans	ResNet-50	Normal *vs* COVID	Acc = 93.01
[59]	137 CT scans	3D-Resnet-10	Severe *vs* Critical COVID	AUC = 0.91
[60]	400 CT scans	VGG16	Normal *vs* COVID	Acc = 99.00
[61]	436 CT scans	ResNet-50	Normal *vs* COVID Other pathologies	Acc = 97.10
[62]	234 CT scans	DenseNet-121	Normal *vs* COVID	Acc = 99.00
[63]	542 CT scans	3D CNN	Normal *vs* COVID	Acc = 90.80
[64]	1110 CT scans	COV-CAF	Normal *vs* COVID	Acc = 97.76
[65]	1164 CT scans	CCSHNet	Normal *vs* COVID *vs* Pneumonia *vs* Tuberculosis	Acc = 96.46
[66]	4154 CT scans	ResNet-50	Normal *vs* COVID *vs* Other pathologies	Acc = 95.00
[67]	63849 CT scans	ResNet-50V2	Normal *vs* COVID	Acc = 99.49
Our method	**513 CCT scans**	**Probabilistic Ensemble**	**Normal *vs* COVID**	**Acc = 97.86**

4.4. Experimental setup

Performance of the classification framework developed in this work was evaluated according to three different experiments:

- **Experiment 1: Classification between controls and COVID-19 patients**. The aim is to detect the presence of pneumonia patterns in the different patches each CCT is divided into. The classification system was based on an ensemble of RBF-SVM classifiers, whose parameters γ and C were optimized by using a grid search within a 5-Fold Cross-Validation scheme. The final values used were $\gamma = 3$ and $C = 1$. Besides, patches were only evaluated if 20% of their voxels contain lung regions.

- **Experiment 2: Evaluation of the effect of the patch size** in the ensemble classification performance. A crucial aspect in the system proposed is the size of the cubic patches used for each individual classifier. A wide range of values were employed (from 24 to 56) in order to check the existence of an optimum size. This would be clearly related to the size of the patterns associated with COVID-19. Moreover, we study the relationship between the patch size and the uncertainty measures derived from kappa scores in order to guide the election of a proper cubic region as a member of the ensemble.

- **Experiment 3: Evaluation of the level of agreement of the different classifiers within the ensemble**. It is of great relevance to measure

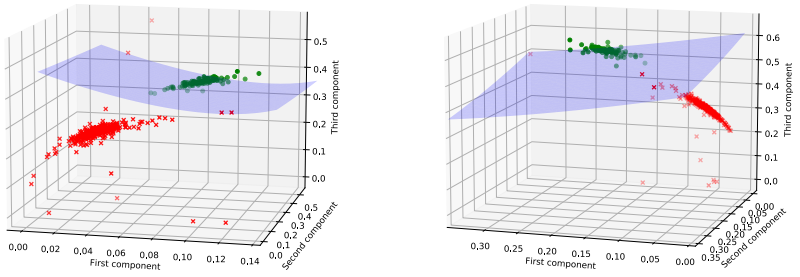

Figure 4: Data image coefficients after polynomial kernel PCA projection and decision surfaces from the RBF-SVM classifier for two different patches. The number of eigenvectors needed for retaining 90% of the variance ranges from 20 to 30. The hyperplane allows the separation between Controls and COVID-19 patients.

the performance of each individual classifier and compare it with the one obtained by the ensemble. Besides, we study how this relationship varies for different patch sizes.

Table 1 summarizes recent works focused on the automatic detection of pneumonia. Besides, it includes information about the methodology employed as well as a comparison with the results obtained by the method proposed in this work.

Table 2: Performance of the ensemble classification proposed in this work for the different patch sizes evaluated.

Patch size	Bal Acc (%)	Sens (%)	Spec (%)	Prec (%)	AUC	F1-score (%)
Baseline approach: Voxels as Features						
28 x 28 x 28	58.73 ±2.43	67.2 ±1.92	59.14 ±1.46	60.02 ±2.04	0.59 ±0.14	61.12 ±1.86
Ensemble approach: kernel PCA						
24 x 24 x 24	96.89 ±1.43	100	84.45 ±0.87	96.42 ±1.01	0.91 ±0.76	98.08 ±0.55
28 x 28 x 28	**97.86 ±0.76**	**100**	**90.18 ±0.86**	**96.98 ±1.02**	**0.95 ±0.01**	**98.75 ±0.45**
32 x 32 x 32	97.27 ±0.98	100	86.98 ±1.24	96.72 ±1.12	0.93 ±0.13	98.33 ±0.52
42 x 42 x 42	89.68 ±1.64	100	81.04 ±1.09	88.82 ±1.49	0.89 ±0.15	94.08 ±0.87
48 x 48 x 48	85.50 ±1.47	100	71.35 ±2.01	93.44 ±0.76	0.86 ±0.20	96.61 ±0.79
56 x 56 x 56	84.00 ±1.52	100	68.19 ±1.87	92.81 ±0.56	0.84 ±0.18	96.27 ±0.73

5. Results

We first explore the performance of the ensemble classifier in terms of different measures, as summarized in Table 2. We can see that the maximum accuracy obtained is 97.86% when a cubic of 28x28x28 voxels is employed in each individual classifier. This manifests the high discrimination ability of the ensemble

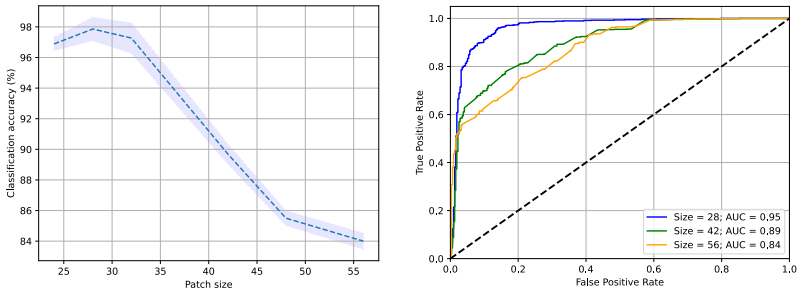

Figure 5: Left: Classification accuracies and their confidence bands obtained by the ensemble classifier for different patch sizes. The size is given in terms of the side of the cubic, so that a patch size = 24 refers to a cubic region of 24x24x24 voxels. Right: ROC curves obtained by the ensemble approach for different patch sizes.

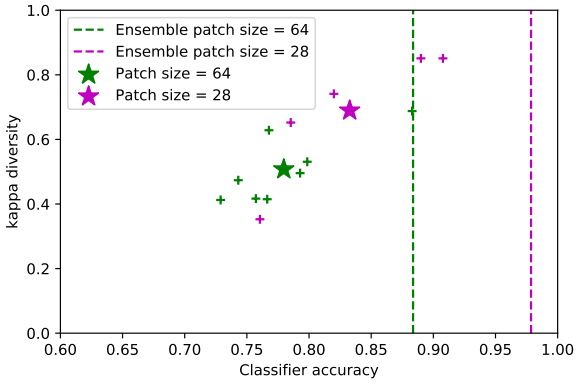

Figure 6: Diversity-accuracy diagrams of the ensemble classifier for two different patch sizes. The x-axis represents the balanced accuracy obtained by each individual classifier and the resulting ensemble. The y-axis represents diversity of the classifiers evaluated by the kappa measure. Each marker represents the kappa-accuracy score obtained by by each individual classifier, whereas large stars represent the centroid of the resulting distribution. The dashed vertical lines represent the ensemble accuracies for the different patch sizes.

system. We also compared our system with a voxel-as-features (VAF) baseline method, in which all voxels of the patch are used as a feature vector. Results show that our approach highly outperformed the baseline, evidencing the need of applying feature extraction before classification. In fact, the projection provided by the eigenvectors allows the classifier to find an optimum solution for separating the two classes (see Figure 4). We also evaluate how performance differs according to the patch sizes of the base classifiers within the ensemble. Figure 5 provides a visual representation of the relationship between accuracy and patch size. Despite accuracies are lower than the one obtained by the cubic of 28x28x28 voxels, other patch sizes also lead to high performance, confirming that the system proposed can detect the presence of pneumonia even when the size of the patches is not ideal. However, results evidence that accuracy starts decreasing when too large patches are employed. A similar behavior occurs when referring to the area under the ROC curve, as Figure 5 shows. Patches that cover too wide regions can add confounds due to other anatomical structures unrelated to COVID-19 but with a similar appearance. Besides, the use of an exceeding size for the region covered by each base classifier can be detrimental for identifying the location of pneumonia. This can be especially relevant when the pulmonary affection of the patient is not severe.

We also use the kappa-accuracy diagram to evaluate the level of agreement between the classifier outputs. Figure 6 shows these diagrams for two different patch sizes. The cloud points represent the kappa score-accuracy obtained by each individual classifier, whereas large stars represent the centroid of the resulting distribution. The most interesting aspects to highlight are the relationship between kappa score and accuracy of individual classifiers, in addition to the differential performance between each base classifier and the resulting ensemble. We can see that kappa diversity and accuracy are linearly dependent: a low kappa score leads to a low classification accuracy, and vice versa. Moreover, the accuracy of the ensemble classifier is much higher than the one obtained by individual members, supporting the suitability of this approach in this context. Although predictions of all base classifiers are used for taking the final decision,

it is worth remembering that their contribution are weighted by the uncertainty of their predictions. If a classifier is much better than the others (in terms of accuracy and reliability), its decision will contribute much more than the rest. Figure 6 shows an extreme case in which one base classifier overcomes by far the rest of the members of the ensemble. When the patch size was 64 (green line and markers), performance of the ensemble was exactly the same as performance of the best individual classifier, which is due to the large weight of this classifier compared to the rest. With reference to the patch size of 28, it shows a more desirable behavior: the final result is given by the combination of more than one member of the ensemble.

6. Discussion

In this study, an image classification framework based on a probabilistic combination of classifiers is proposed. This approach relies on a scheme in which each image is divided into different patches. From each one of them, features are extracted by using kernel PCA, and classification is then performed through a probabilistic RBF-SVM classifier. The outputs of individual classifiers are combined in an ensemble according to the reliability of their predictions. We evaluated the performance of this approach in terms of different measures and studied the influence of the patches size in the global performance and the relationship between individual and global decisions.

The high performance shown by the classification system proposed in this work led to an accurate tool for detecting the presence of pneumonia in CCT scans, in addition to spatially identify where this affection is located. It is extremely important that a simple method like kernel PCA was able to extract the relevant information from each patch. There is a current tendency to use deep learning for the analysis of medical images, especially due to the large performance that these alternatives provide. However, there are some scenarios in which deep learning is not always the best choice for two main reasons. First, the use of convolutional blocks in 3D images requires a high amount of

mathematical operations, leading to an excessive increase in the computational cost. Besides, the convergence and generalization ability of a convolutional neural network are highly influenced by the size of the dataset. Previous studies have performed dimensionality reduction of the input data by selecting only one slice from the 3D CCT. However, this is not the optimal choice for several reasons. It is likely that using only one slice discards information that could be relevant for the classification process. Moreover, one of the main advantages of CCT over CXR is its high resolution and the three-dimensional volumes that it provides, so that reducing the images in order to have two dimensions can be detrimental. Another important issue is that the process for selecting the slice that contains the pulmonary affection is not trivial. This is especially challenging when patients show an incipient pneumonia that is only located in a few lung regions. In this scenario, the selection of the slice would require the help of clinicians, eliminating the desired automaticity that our approach provides.

Another crucial aspect of our method is the way different members of the ensemble are combined. Unfortunately, COVID-19 often causes severe bilateral pneumonia, which means that the pathology is spread across large regions of the lungs. However, in first or intermediate stages of the disease only small pulmonary regions are damaged. This means that when a CCT image is divided into patches, most of them would be classified as controls since no pneumonia patterns would be found. When employing majority voting for the combination of individual classifiers within an ensemble, the final decision only depends on the number of patches that votes for a certain class. In the described scenario, all images from the first stages of the disease (where lung damage is scarce) would be labelled as control patients, invalidating its use as an accurate tool for the detection of pneumonia. However, since we weighted each member of the ensemble according to their uncertainty, the decisions of some members are more important than others. Patches that contain features similar to controls will lead to a high uncertainty. On the other hand, patches with lung lesions will be easily distinguished from controls, resulting in a low uncertainty and therefore, a high

weight in the final decision. Thus, our system detects pneumonia for different grades of severity, from early stages to the hyperinflammation phase. This is extremely useful for an early diagnosis of the pathology and can help doctors to select the proper treatment that speeds up the recovery of the patient.

We have developed a complete system that is able to identify the patterns associated with pneumonia caused by COVID-19. It is worth highlighting the high performance obtained by our proposal: the accuracy and the AUC obtained were 97.86% and 95.31%, respectively. These results overcome other similar techniques in previous studies [8, 68, 43, 69, 70]. There are some relevant aspects regarding our system to be mentioned. First, our system obtained excellent results while keeping a simple solution for the classification of three-dimensional images. Second, the probabilistic nature of the classification scheme provides extremely useful information for clinicians. Our approach detects the presence (or not) of pneumonia and a measure of the uncertainty of its prediction, which can be converted in visual maps (patches with highest accuracy and lowest uncertainty) that help doctors to identify the pulmonary affection associated with COVID-19.

7. Conclusions

The emergence of new technologies has manifested their high usefulness in the diagnosis of a wide range of diseases. This utility is even higher when applying to medical imaging, given the amount of information that they usually provide. In this paper, we propose a method to process images and identify relevant features by using an ensemble probabilistic framework. This is addressed by dividing the images into small regions, applying kernel PCA and performing classification for each individual region. The outputs of each individual classifier are then combined into a global one according to the reliability of each individual prediction: the lower the uncertainty, the higher the contribution. Performance is evaluated in a real scenario where distinguishing between pneumonia patients and controls from chest Computed Tomography (CCT) images,

yielding an accuracy of 97.86%. The combination of individual classifiers provides an automatic tool that detects the presence of the pathology, identifies its location and quantifies the reliability of its prediction. The large performance obtained and the simplicity of the system (use of deep learning in CCT images would highly increase the computational cost) evidence the applicability of our proposal to assist clinicians in a real-world environment. Future versions could optimize the system to adapt to the idiosyncrasy of different types of medical imaging, as well as exploring more complex frameworks in contexts where the computational cost is not too problematic (e.g. two-dimensional images).

Funding

This work was supported by projects PGC2018-098813-B-C32 and RTI2018-098913-B100 (Spanish "Ministerio de Ciencia, Innovación y Universidades"), UMA20-FEDERJA-086, A-TIC-080-UGR18 and P20 00525 (Consejería de economía y conocimiento, Junta de Andalucía) and by European Regional Development Funds (ERDF); and by Spanish "Ministerio de Universidades" through Margarita-Salas grant to J.E. Arco.

References

[1] Q. Guan, Y. Huang, Multi-label chest X-ray image classification via category-wise residual attention learning, Pattern Recognit. Lett. 130 (2020) 259–266.

[2] Z. Barzegar, M. Jamzad, Wlfs: Weighted label fusion learning framework for glioma tumor segmentation in brain MRI, Biomedical Signal Processing and Control 68 (2021) 102617. doi:https://doi.org/10.1016/j.bspc.2021.102617.

[3] Y. Xu, Y. Wang, J. Yuan, Q. Cheng, X. Wang, P. L. Carson, Medical breast ultrasound image segmentation by machine learning, Ultrasonics 91 (2019) 1–9. doi:https://doi.org/10.1016/j.ultras.2018.07.006.

[4] E. H. Houssein, M. M. Emam, A. A. Ali, P. N. Suganthan, Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, Expert Systems with Applications 167 (2021) 114161. doi:https://doi.org/10.1016/j.eswa.2020.114161.

[5] M. R. Islam, M. Nahiduzzaman, Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach, Expert Systems with Applications 195 (2022) 116554. doi:https://doi.org/10.1016/j.eswa.2022.116554.

[6] J. E. Arco, J. Ramírez, J. M. Górriz, M. Ruz, Data fusion based on Searchlight analysis for the prediction of Alzheimer's disease, Expert Systems with Applications 185 (2021) 115549. doi:https://doi.org/10.1016/j.eswa.2021.115549.

[7] A. Lozano, J. S. Suárez, C. Soto-Sánchez, F. J. Garrigós, J. J. Martínez-Álvarez, J. M. Ferrández, E. Fernández, Neurolight: A deep learning neural interface for cortical visual prostheses, International journal of neural systems (2020) 2050045.

[8] Y. D. Zhang, S. C. Satapathy, L. Y. Zhu, J. M. Górriz, S. H. Wang, A seven-layer convolutional neural network for chest CT based COVID-19 diagnosis using stochastic pooling, IEEE Sensors Journal (2020) 1–1doi:10.1109/JSEN.2020.3025855.

[9] Álvaro S. Hervella, J. Rouco, J. Novo, M. Ortega, End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images, Applied Soft Computing 116 (2022) 108347. doi:https://doi.org/10.1016/j.asoc.2021.108347.

[10] S.-H. Wang, D. R. Nayak, D. S. Guttery, X. Zhang, Y.-D. Zhang, Covid-19 classification by ccshnet with deep fusion using transfer learning and discriminant correlation analysis, Information Fusion 68 (2021) 131–148. doi:https://doi.org/10.1016/j.inffus.2020.11.005.

[11] Y. Zhang, S. Satapathy, X. Zhang, S. Wang, COVID-19 diagnosis via densenet and optimization of transfer learning setting, Cognitive Computation (2021) 1 – 17.

[12] A. Theissler, J. Pérez-Velázquez, M. Kettelgerdes, G. Elger, Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry, Reliability Engineering & System Safety 215 (2021) 107864.

[13] A. Haagsma, M. Scharenberg, L. Keister, J. Schuetter, N. Gupta, Secondary porosity prediction in complex carbonate reefs using 3D CT scan image analysis and machine learning, Journal of Petroleum Science and Engineering 207 (2021) 109087. doi:https://doi.org/10.1016/j.petrol.2021.109087.

[14] Y. Wang, J. Liu, Y. Han, Production capacity prediction of hydropower industries for energy optimization: Evidence based on novel extreme learning machine integrating monte carlo, Journal of Cleaner Production 272 (2020) 122824. doi:https://doi.org/10.1016/j.jclepro.2020.122824.

[15] H. Wang, H. Owens, J. Smith, W. P. Chernicoff, M. Mazari, M. Pourhomayoun, An end-to-end traffic vision and counting system using computer vision and machine learning: The challenges in real-time processing, in: SIGNAL 2018 : The Third International Conference on Advances in Signal, Image and Video Processing, 2018.

[16] S. Karlsen, Automated front detection - using computer vision and machine learning to explore a new direction in automated weather forecasting (2017).

[17] S. V. Mora, Computer vision and machine learning for in-play tennis analysis: framework, algorithms and implementation, Ph.D. thesis, University of London (2018).

[18] 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. doi:10.1109/CVPR.2015.7298594.

[19] J. Shanmugasundaram, G. Raichal, G. Dency Flora, P. Rajasekaran, V. Jeevanantham, Classification of epileptic seizure using rotation forest ensemble method with 1D-LBP feature extraction, Materials Today: Proceedings (2021). doi:https://doi.org/10.1016/j.matpr.2021.12.258.

[20] J. Ding, J. Song, J. Li, J. Tang, F. Guo, Two-stage deep neural network via ensemble learning for melanoma classification, Frontiers in Bioengineering and Biotechnology 9 (2022) 758495. doi:10.3389/fbioe.2021.758495.

[21] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. van Vulpen, J. P. W. Pluim, Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple), IEEE Transactions on Medical Imaging 29 (12) (2010) 2000–2008. doi:10.1109/TMI.2010.2057442.

[22] M. M. Rahman, B. C. Desai, P. Bhattacharya, Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion, Computerized Medical Imaging and Graphics 32 (2) (2008) 95–108. doi:https://doi.org/10.1016/j.compmedimag.2007.10.001.

[23] C. Mao, L. Lu, B. Hu, Local probabilistic model for Bayesian classification: A generalized local classification model, Applied Soft Computing 93 (2020) 106379. doi:https://doi.org/10.1016/j.asoc.2020.106379.

[24] S. N. M. Foumani, A. Nickabadi, A probabilistic topic model using deep visual word representation for simultaneous image classification and annotation, Journal of Visual Communication and Image Representation 59 (2019) 195–203. doi:https://doi.org/10.1016/j.jvcir.2019.01.009.

[25] L. Laib, M. S. Allili, S. Ait-Aoudia, A probabilistic topic model for event-based image classification and multi-label annotation, Signal Processing: Image Communication 76 (2019) 283–294. doi:https://doi.org/10.1016/j.image.2019.05.012.

[26] S. Raudys, A. Jain, Small sample size effects in statistical pattern recognition: Recommendations for practitioners, Pattern Analysis and Machine Intelligence, IEEE Transactions on 13 (1991) 252–264. doi:10.1109/34.75512.

[27] I. T. Jolliffe, Principal Component Analysis and Factor Analysis, Springer New York, 1986. doi:10.1007/978-1-4757-1904-8_7.

[28] L. Khedher, J. Ramírez, J. Górriz, A. Brahim, F. Segovia, Early diagnosis of alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images, Neurocomputing 151 (2015) 139 – 150. doi:https://doi.org/10.1016/j.neucom.2014.09.072.

[29] M. López, J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, I. Álvarez, F. Segovia, C. Puntonet, Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules, Electronics Letters 45 (2009) 389–391.

[30] M. López, J. Ramïrez, J. Gorriz, I. Illan, D. Salas-Gonzalez, F. Segovia, R. Chaves, SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA, Neuroscience letters 464 (2009) 233–8. doi:10.1016/j.neulet.2009.08.061.

[31] Q. Wang, Kernel principal component analysis and its applications in face recognition and active shape models, ArXiv abs/1207.3538 (2012).

[32] B. Schölkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (5) (1998) 1299–1319. doi:10.1162/089976698300017467.

[33] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1023/A:1022627411411.

[34] C. L. Morais, K. M. Lima, F. L. Martin, Uncertainty estimation and misclassification probability for classification models based on discriminant analysis and support vector machines, Analytica Chimica Acta 1063 (2019) 40 – 46. doi:https://doi.org/10.1016/j.aca.2018.09.022.

[35] Xulei Yang, Qing Song, A. Cao, Weighted support vector machine for data classification, in: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., Vol. 2, 2005, pp. 859–864 vol. 2. doi:10.1109/IJCNN.2005.1555965.

[36] G. King, L. Zeng, Logistic regression in rare events data, Political Analysis 9 (2001) 137–163.

[37] V. N. Vapnik, The nature of statistical learning theory (1999).

[38] J. C. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, in: ADVANCES IN LARGE MARGIN CLASSIFIERS, MIT Press, 1999, pp. 61–74.

[39] B. Efron, Bootstrap methods: another look at the jackknife, The Annals of Statistics 7 (1) (1979) 1–26.

[40] R. Wehrens, H. Putter, L. M. Buydens, The bootstrap: a tutorial, Chemometrics and Intelligent Laboratory Systems 54 (1) (2000) 35 – 52. doi:https://doi.org/10.1016/S0169-7439(00)00102-7.

[41] A. Ortiz, J. Munilla, J. M. Górriz, J. Ramírez, Ensembles of Deep Learning architectures for the early diagnosis of the Alzheimer's disease, International Journal of Neural Systems 26 (07) (2016) 1650025. doi:10.1142/S0129065716500258.

[42] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, S. S. Netam, Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting

based classifier ensemble, Expert Systems with Applications 165 (2021) 113909. doi:https://doi.org/10.1016/j.eswa.2020.113909.

[43] T. Zhou, H. ling Lu, Z. Yang, S. Qiu, B. qiang Huo, Y. Dong, The ensemble deep learning model for novel covid-19 on CT images, Applied Soft Computing (2020) 106885doi:https://doi.org/10.1016/j.asoc.2020.106885.

[44] M. Liu, D. Zhang, D. Shen, Ensemble sparse classification of Alzheimer's disease, NeuroImage 60 (2) (2012) 1106 – 1116. doi:https://doi.org/10.1016/j.neuroimage.2012.01.055.

[45] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66. doi:10.1109/TSMC.1979.4310076.

[46] K. Marstal, F. Berendsen, M. Staring, S. Klein, Simpleelastix: A user-friendly, multi-lingual library for medical image registration, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, pp. 574–582. doi:10.1109/CVPRW.2016.78.

[47] S. Klein, M. Staring, K. Murphy, M. Viergever, J. Pluim, Elastix: A toolbox for intensity-based medical image registration, IEEE transactions on medical imaging 29 (2009) 196–205. doi:10.1109/TMI.2009.2035616.

[48] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, p. 1137–1143.

[49] J. N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment, Journal of Thoracic Oncology 5 (9) (2010) 1315 – 1316. doi:https://doi.org/10.1097/JTO.0b013e3181ec173d.

[50] K. Hajian-Tilaki, Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation, Caspian journal of internal medicine 4 (2013) 627–635.

[51] J. J. Rodriguez, L. I. Kuncheva, C. J. Alonso, Rotation forest: A new classifier ensemble method, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (10) (2006) 1619–1630. doi:10.1109/TPAMI.2006.211.

[52] J. Wang, Y. Yang, B. Xia, A simplified cohen's kappa for use in binary classification data annotation tasks, IEEE Access 7 (2019) 164386–164397. doi:10.1109/ACCESS.2019.2953104.

[53] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1960) 37 – 46.

[54] B. Di Eugenio, M. Glass, The kappa statistic: A second look, Comput. Linguist. 30 (1) (2004) 95–101. doi:10.1162/089120104773633402.

[55] P. K. Wong, T. Yan, H. Wang, I. N. Chan, J. Wang, Y. Li, H. Ren, C. H. Wong, Automatic detection of multiple types of pneumonia: Open dataset and a multi-scale attention network, Biomedical Signal Processing and Control 73 (2022) 103415. doi:https://doi.org/10.1016/j.bspc.2021.103415.

[56] R. Alizadehsani, D. Sharifrazi, N. H. Izadi, J. H. Joloudari, A. Shoeibi, J. M. Gorriz, S. Hussain, J. E. Arco, Z. A. Sani, F. Khozeimeh, A. Khosravi, S. Nahavandi, S. M. S. Islam, U. R. Acharya, Uncertainty-aware semi-supervised method using large unlabeled and limited labeled COVID-19 data, ACM Transactions on Multimedia Computing, Communications, and Applications 17 (3s) (2021). doi:10.1145/3462635.

[57] S. Harmon, T. Sanford, S. Xu, E. Turkbey, H. Roth, Z. Xu, D. Yang, A. Myronenko, V. Anderson, A. Amalou, M. Blain, M. Kassin, D. Long, N. Varble, S. Walker, A. Ierardi, E. Stellato, G. Plensich, B. Turkbey,

Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets, Nature Communications 11 (2020) 4080. doi:10.1038/s41467-020-17971-2.

[58] Y. Pathak, P. Shukla, A. Tiwari, S. Stalin, S. Singh, P. Shukla, Deep transfer learning based classification model for COVID-19 disease, IRBM (2020). doi:https://doi.org/10.1016/j.irbm.2020.05.003.

[59] C. Li, D. Dong, L. Li, W. Gong, X. Li, Y. Bai, M. Wang, Z. Hu, Y. Zha, J. Tian, Classification of severe and critical COVID-19 using deep learning and radiomics, IEEE Journal of Biomedical and Health Informatics 24 (12) (2020) 3585–3594. doi:10.1109/JBHI.2020.3036722.

[60] N. K. Mishra, P. Singh, S. D. Joshi, Automated detection of COVID-19 from CT scan using convolutional neural network, Biocybernetics and Biomedical Engineering 41 (2) (2021) 572–588. doi:https://doi.org/10.1016/j.bbe.2021.04.006.

[61] Q. Qi, S. Qi, Y. Wu, C. Li, B. Tian, S. Xia, J. Ren, L. Yang, H. Wang, H. Yu, Fully automatic pipeline of convolutional neural networks and capsule networks to distinguish COVID-19 from community-acquired pneumonia via CT images, Computers in Biology and Medicine 141 (2022) 105182. doi:https://doi.org/10.1016/j.compbiomed.2021.105182.

[62] S. H. Kassania, P. H. Kassanib, M. J. Wesolowskic, K. A. Schneidera, R. Detersa, Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: A machine learning based approach, Biocybernetics and Biomedical Engineering 41 (3) (2021) 867–879. doi:https://doi.org/10.1016/j.bbe.2021.05.013.

[63] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for COVID-19 from chest CT using weak label, medRxiv (2020). doi:10.1101/2020.03.12.20027185.

[64] M. R. Ibrahim, S. Youssef, K. M. Fathalla, Abnormality detection and intelligent severity assessment of human chest computed tomography scans using deep learning: a case study on SARS-COV-2 assessment, Journal of Ambient Intelligence and Humanized Computing (2021) 1 – 24.

[65] S.-H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, Y.-D. Zhang, Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network, Information Fusion 67 (2021) 208 – 229.
URL http://www.sciencedirect.com/science/article/pii/S1566253520303705

[66] H. Kaheel, A. Hussein, A. Chehab, AI-based image processing for COVID-19 detection in chest CT scan images, Frontiers in Communications and Networks 2 (2021). doi:10.3389/frcmn.2021.645040.

[67] M. Rahimzadeh, A. Attar, S. M. Sakhaei, A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset, Biomedical Signal Processing and Control 68 (2021) 102588. doi:https://doi.org/10.1016/j.bspc.2021.102588.

[68] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, B. Xu, A deep learning algorithm using CT images to screen for corona virus disease (covid-19), medRxiv (2020). doi:10.1101/2020.02.14.20023028.

[69] E. E.-D. Hemdan, M. A. Shouman, M. E. Karar, COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images (2020). arXiv:2003.11055.

[70] I. Apostolopoulos, M. Tzani, Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, Australasian physical & engineering sciences in medicine 43 (03 2020). doi:10.1007/s13246-020-00865-4.