A shallow learning investigation for COVID-19 classification

Luca Zedda[1], Andrea Loddo[1], and Cecilia Di Ruberto[1]

Department of Mathematics and Computer Science, University of Cagliari,
l.zedda12@studenti.unica.it
{andrea.loddo, dirubert}@unica.it

Abstract. COVID-19, an infectious coronavirus disease, triggered a pandemic that resulted in countless deaths. Since its inception, clinical institutions have used computed tomography as a supplemental screening method to reverse transcription-polymerase chain reaction. Deep learning approaches have shown promising results in addressing the problem; however, less computationally expensive techniques, such as those based on handcrafted descriptors and shallow classifiers, may be equally capable of detecting COVID-19 based on medical images of patients. This work proposes an initial investigation of several handcrafted descriptors well known in the computer vision literature already been exploited for similar tasks. The goal is to discriminate tomographic images belonging to three classes, COVID-19, pneumonia, and normal conditions, and present in a large public dataset. The results show that kNN and ensembles trained with texture descriptors achieve outstanding accuracy in this task, reaching accuracy and F-measure of 93.05% and 89.63%, respectively. Although it did not exceed state of the art, it achieved satisfactory performance with only 36 features, enabling the potential to achieve remarkable improvements from a computational complexity perspective.

Keywords: Computer Vision · Shallow Learning · Image Processing · COVID-19 detection · Texture features · CT scan Images

1 Introduction

COVID-19 is a disease caused by the SARS-CoV-2 virus, declared a pandemic by the World Health Organisation on 11 March 2020. At the time of writing, COVID-19 has more than 458 million confirmed cases and has caused more than six million deaths [24]. The infection starts in the throat's mucous membranes and spreads to the lungs through the respiratory tract. COVID-19 is a highly contagious disease; therefore, rapid screening is essential for timely diagnosis and treatment. Diagnosis of COVID-19 infection by imaging-based methods has been reported to give accurate results, both for screening and quantifying the amount of damage [7]. At the same time, attempts are being made to develop rapid diagnostic techniques for detecting COVID-19 using chest X-ray (CXR)

and chest computed tomography (CT) images that radiologists frequently analyze. Like many other types, manual diagnosis of COVID-19 is time-consuming, prone to human error, and requires the help of a competent radiologist. The presence of an experienced radiologist is also necessary because abnormalities in the early stages of COVID-19 may resemble other lung diseases such as severe acute respiratory syndrome (SARS) or viral pneumonia (VP), which may delay the diagnosis and treatment of COVID.

In particular, CXR is the most easily accessible and fastest form of imaging with the fewest side effects on the human body. CXR imaging has traditionally been used for the detection of pneumonia and cancer. Although it can detect COVID-19 infection, it fails to provide fine-order details of infected lungs. CT is a more sophisticated technique for assessing is used to evaluate the level of infection. CXR imaging can be used to detect COVID-19; however, to assess the level of severity of infection, a CT scan is mandatory [1].

For these reasons, any automated solution designed to diagnose COVID-19 should also consider other respiratory disorders to develop a more comprehensive and robust diagnostic system [28].

In this manuscript, our focus is on computed tomography from a machine and deep learning point of view. CT is a widely explored medical imaging technique that allows non-invasive visualisation of the interior of an object [8,2,33,32,15] and is widely used in many applications, such as medical imaging for clinical purposes [18,30,14,23].

Two CT scans of COVID-19 and non-COVID-19 are shown in fig. 1.

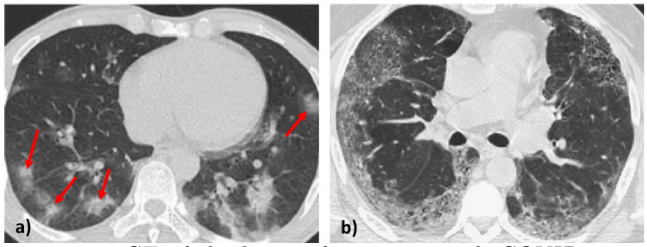

Fig. 1: *(a) represents a CT of the lungs of a patient with COVID-19, in which there are traces of ground glass opacity indicated by red arrows. (b) shows a CT of the lungs of a non-COVID-19 patient with diffuse opacity in the outer parts of the lungs. These images are courtesy of [35].*

The main target of this study is an early investigation regarding the extent to which two types of handcrafted (HC) features can be beneficial for the classification of the CTs of patients. We focus our work on invariant moments and texture features, which are widely employed in the context of MR and CT imaging[26], to train several shallow classifiers.

We performed a three-class classification on the public COVIDx CT-2A dataset, divided explicitly into COVID-19, pneumonia, and healthy cases; we investigated the robustness of the descriptors analyzed through five different

machine learning classifiers; we aim to demonstrate how this task can be faced employing HC features, even with low-end devices, for a real-time diagnosis.

Moreover, several works in the context of COVID-19 diagnostics have considered small or private datasets or lacked rigorous experimental methods, potentially leading to overfitting and overestimation of performance [29,28]. For this reason, we carefully selected an extended dataset composed of multi-source CT images on which to conduct the experiments described. Roberts et al. [28] have recently shown that most of the datasets used in the literature for the diagnosis or prognosis of COVID-19 suffer from duplication and quality problems. It has already been provided with train, validation, and testing splits.

Our proposed approach achieves promising results on COVID-19 identification, although it is only a preliminary analysis that does not consider deep learning methods, for example.

The rest of the manuscript is organized as follows. section 2 illustrates the dataset, our approach, and the setup. In section 3 the experimental results are given and discussed. Finally, in section 4 we draw the findings and directions for future works.

2 Materials and Methods

A publicly available dataset has been used in this study, as detailed in section 2.1, while the evaluation metrics are described in section 2.4.

2.1 Dataset: COVIDx CT-2A

COVIDx CT-2A [9] is an open-access dataset. At the time of writing, it consisted of 194,922 CT images of 3745 patients from 15 different countries, aged 0-93 years (median age 51), belonging to a particular class clinically verified by experienced pathologists.

Specifically, the classes are COVID-19, indicating CT images of COVID-19 positive patients, pneumonia, and patients in a normal condition.

The countries involved are part of a multinational cohort that consists of patient cases collected by the following organizations and initiatives from around the world:

1. China National Center for Bioinformation (CNCB) [34] (China);
2. National Institutes of Health Intramural Targeted Anti-COVID-19 (ITAC) Program (hosted by TCIA [12], countries unknown);
3. Negin Radiology Medical Center [27] (Iran);
4. Union Hospital and Liyuan Hospital of the Huazhong University of Science and Technology [20] (China);
5. COVID-19 CT Lung and Infection Segmentation initiative annotated and verified by Nanjing Drum Tower Hospital [17] (Iran, Italy, Turkey, Ukraine, Belgium, some countries unknown);
6. Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) [4] (countries unknown);

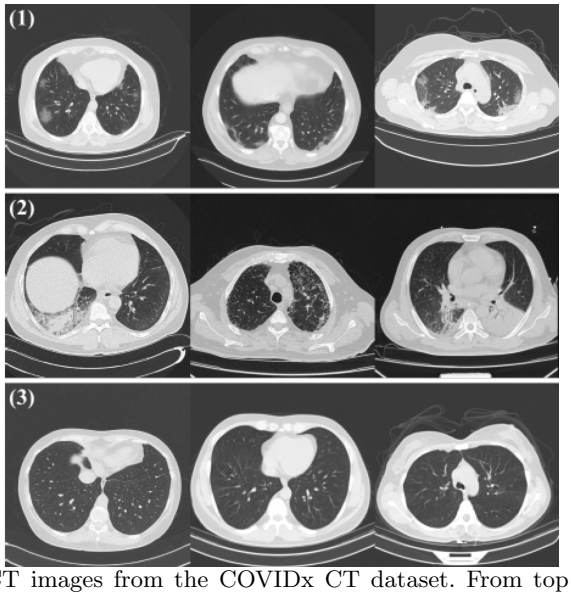

Fig. 2: Sample CT images from the COVIDx CT dataset. From top to bottom, the images in (**1**) represent coronavirus pneumonia due to SARS-CoV-2 infection (NCP), (**2**) are common pneumonia (CP), and (**3**) are healthy lungs. Images are courtesy of [10].

7. Radiopaedia collection [5] (Iran, Italy, Australia, Afghanistan, Scotland, Lebanon, England, Algeria, Peru, Azerbaijan, some countries unknown).

Figure 2 shows some sample images taken from the dataset.

2.2 Feature extraction

We evaluated two different feature sets: invariant moments and texture features.

The invariant moments computed were the **Hu**[13] and **Zernike** moments. The former are invariant to changes in scale, translation, and rotation, while the latter are orthogonal and represent image properties without redundancy[22]. The order of the Zernike moments is 6, as higher orders would reduce the system's performance by adding features representing irrelevant details or noise[6].

As texture features, the rotationally invariant Gray Level Co-occurrence Matrix (GLCM), as proposed in [25], and the rotationally invariant **LBP** features[21] have been computed. In both cases, we focused on the finest textures, so we computed four GLCMs with $d = 1$ and $\theta = [0°, 45°, 90°, 135°]$ and the LBP map in the neighborhood identified by r and n equal to 1 and 8, respectively. From the GLCMs we extracted 26 features [11] and converted them to the rotationally invariant features Har_{ri}[25]). The LBP map is converted to a rotationally invariant one, and its histogram is used as the feature vector LBP_{ri} [21].

2.3 Shallow classifiers

The classification algorithms considered in this study were as follows: **kNN** (k-Nearest Neighbors) (kNN), **Decision Tree** (DT), **Naive Bayes** (NB), **Ensemble** (Ens), and **Support Vector Machine** (SVM).

To ensure the heterogeneity of the training set, we trained each classifier with stratified 10-fold cross-validation to ensure that the proportion of positive and negative examples is respected in all folds. For each case, we selected the model with the largest area under the ROC curve (AUC).

The hyperparameters characterizing each classifier were not fine-tuned. Our goal for this study was not to obtain the best absolute performance but only to understand the extent to which the descriptors used were feasible for analysis. Furthermore, several authors have empirically observed that in many cases, the use of tuned parameters cannot significantly exceed the default values of a classifier[3,31].

However, to make the results reproducible, we specify the values of the hyperparameters chosen. For the kNN classifier, the distance metric adopted was the euclidean, and the number of nearest neighbors is 3, with an inverse squared distance weighting function. For the Decision Tree classifier, the maximum number of splits to control the depth of the trees is 50. The distribution chosen to model the data is normal for the Naive Bayes classifier with a normal kernel. The ensemble classifier was Random Undersampling Boosting (RUSBoost), the students were decision trees. Finally, for the SVM classifier, we used a linear kernel.

2.4 Metrics

The performance measures used to quantify the performance are the accuracy (A), precision (P), specificity (SP), sensitivity (SE), F1-score (F1) as following defined:

$$accuracy = \frac{TP + TN}{TP + TF + FP + FN}, \tag{1}$$

$$precision = \frac{TP}{TP + FP}, \tag{2}$$

$$specificity = \frac{TN}{FP + TN}, \tag{3}$$

$$recall = \frac{TP}{TP + FN}, \tag{4}$$

$$F1 = \frac{2 * precision * recall}{precision + recall} = \frac{2 * TP}{2 * TP + FP + FN}. \tag{5}$$

Accuracy measures the number of correctly labeled items belonging to the positive class divided by the items correctly or incorrectly labeled as belonging to the same class. Specificity measures the proportion of correctly identified

negatives (the true negative rate), while sensitivity measures the proportion of correctly identified positives (the true positive rate). The fourth measure is accuracy, defined as the ratio of correctly labeled instances to the entire pool of cases. The last is the F1 score, which conveys the balance between precision and recall.

2.5 Experimental Setup

The images were not preprocessed or augmented because this work is used as a basis for further investigation. We retained the dataset splits provided by the authors and did not use any randomization approach to make the studies repeatable.

In addition, COVIDx CT-2A has already been divided by the authors according to the following percentages: 70%, 20%, and 10% for training, validation, and testing, respectively. For the sake of reproducibility, we left this division unchanged.

Finally, all the experiments have been conducted on a single machine with the following configuration: Intel(R) Core(TM) i9-8950HK @ 2.90GHz CPU with 32 GB RAM and NVIDIA GTX1050 Ti 4GB GPU.

3 Experimental Results and Discussion

Three main results obtained from the experiments are reported. In particular, fig. 3 and fig. 4 show the general behavior of the four sets of descriptors with two different metrics, accuracy, and F-measure, respectively. We give a general indication of the descriptors' effectiveness for the task.

Next, we analyze the performance class by class, remembering that the problem is divided into three classes. fig. 5 show the performance of the accuracy and F-measure computed for the best combination of descriptor and classifier.

Figure 4 shows that the configuration based on the kNN classifier trained with LBP features can achieve the highest performance, with 93.05% accuracy and 89.63% F-measure. Overall, it seems clear that texture outperformed the moment features in this scenario, as both HAR and LBP achieved the highest performance with every classifier except SVM.

Although the DT and Ensemble strategies achieved interesting results, kNN seems the most suitable for this task, especially when trained with LBP descriptors, being the only one above 89% on both metrics.

In general, looking at accuracy, there are no distinct performance differences between the texture descriptors to justify one over the other. In the case of SVM, the accuracy achieved with HAR as a descriptor is even higher than LBP.

However, the scenario changes when looking at the results of multiclass classification performance evaluated using both metrics computed for all classes. In particular, fig. 4 shows that the F-measure values are generally lower than the accuracy. However, the texture descriptors helped keep the performance high, especially LBP, which reached 89.63% with the kNN classifier.

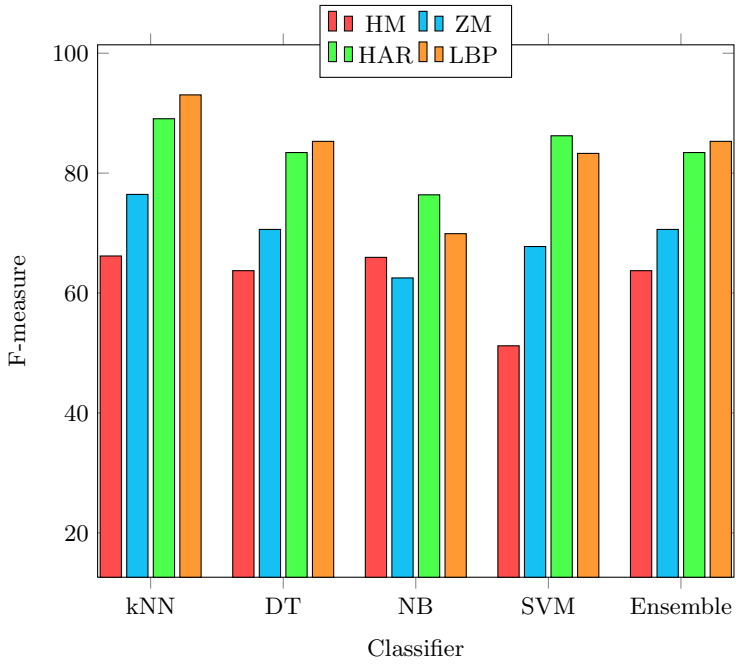

Fig. 3: Accuracy trends with the different classifiers adopted.

As a general rule, on the one hand, the results provided by the numerous experiments conducted seem to show that the kNN trained with LBP features can perform the multiclass classification task with performance that outperforms any other combination of descriptors and classifier analyzed. Moreover, as represented by fig. 5, this solution seems to be the most robust, achieving the highest results in each comparative test.

A final statement concerns the number of features extracted from each descriptor. More specifically, 36 features were extracted from the LBP descriptor, while HAR, ZM, and HM had 26, 10, and 7, respectively. If we consider the number of discriminating features, the results obtained with LBP features can be considered satisfactory in a low-resource environment, even compared to the computational cost of using a CNN, and open the way for possible combinations of heterogeneous features.

The behavior expressed by texture features and LBP, in particular, is due to their high representation power in MR and CT imaging, even for fine-grained analysis[19]. However, the accuracy metric does not represent the detail of the multiclass problem addressed in this work. For this reason, we adopted the F-measure to have a more unambiguous indication of which features are best suited for the task. Figure 4 confirms that, in terms of F-measure, the kNN and ensemble classifiers benefit from texture features in general and LBP in particular, outperforming all the rest.

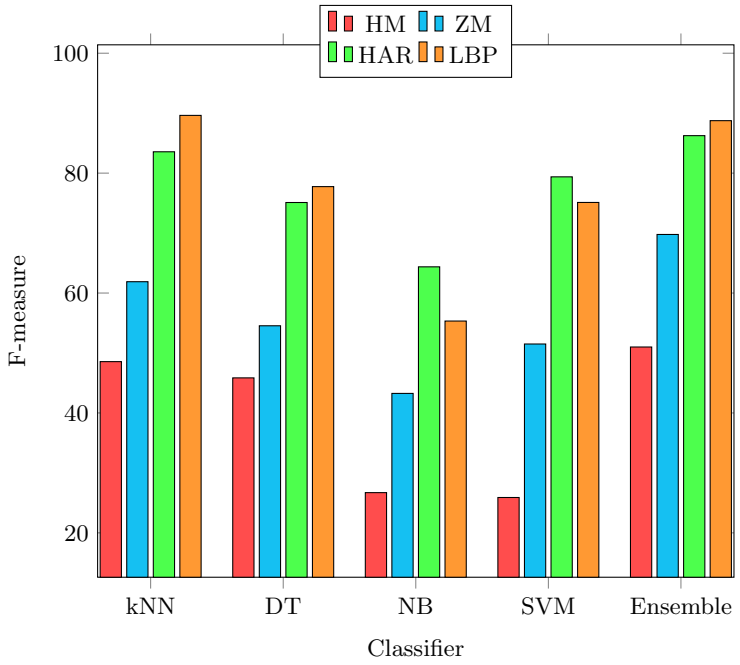

Fig. 4: F-measure trends with the different classifiers adopted.

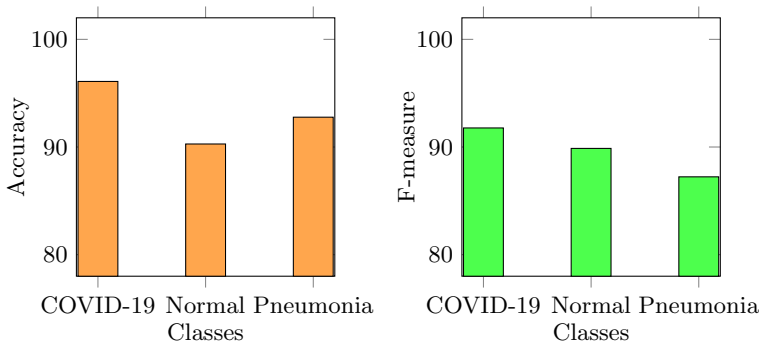

Fig. 5: Classwise F-measure computed from the best model: kNN with LBP features.

Furthermore, table 1 shows the metrics computed with all the classifiers and descriptors. As can be seen, the LBP descriptors are the best and second-best performing, with kNN and ensemble, respectively. It is also notable the result obtained with the same classifiers trained with *HARri* features, clearly being the second-best descriptor. It must be noted that with only 36 and 26 features, the mentioned classifiers can achieve more than 90% accuracy.

As presented in table 2, these results cannot outperform the 98.87% accuracy obtained in a previous work that employed only CNNs for the same task[16], and

Classifier	Descriptor	Size	A	P	SP	SE	F1
DT	HM	7	63.73	52.56	48.97	74.55	45.85
	HARri	26	83.43	76.06	74.44	86.52	75.10
	LBP	36	85.30	77.48	78.70	88.42	77.75
	ZM	10	70.61	56.50	56.09	78.54	54.54
Naive Bayes	HM	7	65.95	44.52	35.47	68.24	26.71
	HARri	26	76.38	65.39	65.13	81.27	64.37
	LBP	36	69.90	54.80	57.84	77.14	55.33
	ZM	10	62.54	51.44	46.03	73.85	43.26
SVM	HM	7	51.20	27.88	25.63	62.25	25.90
	HARri	26	86.23	79.41	79.80	88.85	79.38
	LBP	36	83.29	74.81	76.14	86.70	75.11
	ZM	10	67.76	53.95	54.33	76.76	51.50
Ensemble	HM	7	67.77	56.66	53.43	77.20	51.00
	HARri	26	90.75	86.15	86.37	92.52	86.25
	LBP	36	**92.43**	**88.57**	**88.97**	**93.89**	**88.75**
	ZM	10	81.30	71.10	70.83	86.58	69.77
kNN	HM	7	66.18	52.55	50.45	75.81	48.56
	HARri	26	89.07	83.50	83.66	91.17	83.56
	LBP	36	**93.05**	**89.66**	**89.63**	**94.33**	**89.63**
	ZM	10	76.45	63.87	63.39	82.97	61.89

Table 1: Weighted average performance computed using the five different shallow classifiers trained with each of the four descriptors. The best results is in black bold, while red indicates the second-best.

Method	Accuracy	N. features
COVID-Net CT-1 [10]	94.5%	$> 1,000$
COVID-Net CT-2 L [10]	98.1%	$> 1,000$
COVID-Net CT-2 S [10]	97.9%	$> 1,000$
Bit-M [36]	99.2%	$> 1,000$
Loddo [16]	98.87%	1,024
This work	93.05%	36

Table 2: Comparison of this work with the state of the art on COVIDx CT-2A.

others using deep learning strategies. However, a satisfactory accuracy has been obtained with an enormously smaller quantity of features (only 36 for LBP) compared to that works. It opens the field to further improvements, e.g., , with features combination or by using this indicator to realize attention mechanisms to drive CNNs towards the right portions of the images needed to perform a correct diagnosis.

4 Conclusions

In this study, we have investigated the performance of some handcrafted features, invariant moments, and texture for a three-class classification of CT images. In particular, we have performed an early investigation of the feature feasibility for the classification task that involves COVID-19 and pneumonia classes.

This investigation can be considered a novel, rapid and lightweight approach for the diagnosis of COVID-19 to address the problem of faster diagnosis and patient prioritization. It can potentially also confirm the severity. The task is faced with five different shallow classification methods, each one trained with four different categories of handcrafted descriptors.

We worked on a dataset composed of 194,922 images, of which about 19,000 have been exploited in testing the mentioned descriptors. Among all the experiments, the kNN classifier trained with LBP features reached a weighted accuracy of 93.05% and an F-measure of 89.63%, being the best performer. In any case, the texture features generally demonstrated high representative power with both kNN and ensemble strategies.

These results show significant and promising results regarding the state of the art, even though we planned several further investigations. First of all, we aim to propose a system able to combine the potential of CNNs, expressed in[16], with the features presented in this work. To cite some examples, they can be combined, even with the ones extracted from CNNs, or their indicators can even be used to realize attention-based mechanisms to drive CNNs towards the right portions of the images to perform a correct diagnosis. Secondly, we plan to extend this work to generalize it to different diseases and conditions. Finally, we aim to further study these methods to test their robustness across multiple datasets and assess their feasibility in clinical practice.

References

1. Aggarwal, P., Mishra, N.K., Fatimah, B., Singh, P., Gupta, A., Joshi, S.D.: Covid-19 image classification using deep learning: Advances, challenges and opportunities. Computers in Biology and Medicine p. 105350 (2022)
2. Ahmad, M., Ai, D., Xie, G., Qadri, S.F., Song, H., Huang, Y., Wang, Y., Yang, J.: Deep belief network modeling for automatic liver segmentation. IEEE Access **7**, 20585–20595 (2019)
3. Arcuri, A., Fraser, G.: Parameter tuning or default values? an empirical investigation in search-based software engineering. Empir. Softw. Eng. **18**(3), 594–623 (2013)
4. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. Medical Physics **38**(2), 915–931 (2011)
5. Bell, D.J.: Covid-19. https://radiopaedia.org/articles/covid-19-4 (2021), online; accessed 9 August 2021

6. Di Ruberto, C., Putzu, L., Rodriguez, G.: Fast and accurate computation of orthogonal moments for texture analysis. Pattern Recognition **83**, 498–510 (2018)
7. Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., Ji, W.: Sensitivity of chest ct for covid-19: comparison to rt-pcr. Radiology **296**(2), E115–E117 (2020)
8. Furqan Qadri, S., Ai, D., Hu, G., Ahmad, M., Huang, Y., Wang, Y., Yang, J.: Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in ct images. Applied Sciences **9**(1), 69 (2019)
9. Gunraj, H.: Covid-net open source initiative - covidx ct-2 dataset. https://www.kaggle.com/hgunraj/covidxct (2020), online; accessed 30 June 2021
10. Gunraj, H., Sabri, A., Koff, D., Wong, A.: Covid-net ct-2: Enhanced deep neural networks for detection of covid-19 from chest ct images through bigger, more diverse learning (2021)
11. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **SMC-3**(6), 610–621 (1973)
12. Harmon, S.A., Sanford, T.H., Xu, S., Turkbey, E.B., Roth, H., Xu, Z., Yang, D., Myronenko, A., Anderson, V., Amalou, A., et al.: Artificial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets. Nature Communications **11**(1), 1–7 (2020)
13. Hu, M.K.: Visual pattern recognition by moment invariants. IRE Trans. Inf. Theory **8**(2), 179–187 (1962)
14. Isaac, A., Nehemiah, H.K., Isaac, A., Kannan, A.: Computer-aided diagnosis system for diagnosis of pulmonary emphysema using bio-inspired algorithms. Computers in Biology and Medicine **124**, 103940 (2020)
15. Liu, M., Dong, J., Dong, X., Yu, H., Qi, L.: Segmentation of lung nodule in ct images based on mask r-cnn. In: 9th International Conference on Awareness Science and Technology, iCAST 2018, Fukuoka, Japan, September 19-21, 2018. pp. 1–6. IEEE (2018)
16. Loddo, A., Pili, F., Di Ruberto, C.: Deep learning for covid-19 diagnosis from ct images. Applied Sciences **11**(17) (2021)
17. Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., et al.: Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation. arXiv e-prints pp. arXiv–2004 (2020)
18. Ma, L., Liu, X., Gao, Y., Zhao, Y., Zhao, X., Zhou, C.: A new method of content based medical image retrieval and its applications to ct imaging sign retrieval. Journal of Biomedical Informatics **66**, 148–158 (2017)
19. Maheshwari, S., Sharma, R.R., Kumar, M.: Lbp-based information assisted intelligent system for covid-19 identification. Computers in Biology and Medicine **134**, 104453 (2021)
20. Ning, W., Lei, S., Yang, J., Cao, Y., Jiang, P., Yang, Q., Zhang, J., Wang, X., Chen, F., Geng, Z., et al.: Open resource of clinical data from patients with pneumonia for the prediction of covid-19 outcomes via deep learning. Nature Biomedical Engineering **4**(12), 1197–1207 (2020)
21. Ojala, T., Pietikäinen, M., T., M.: Multiresolution gray-scale and rotation invariant texture classification with local binary pattern. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
22. Oujaoura, M., Minaoui, B., Fakir, M.: Image annotation by moments. In: Moments and Moment Invariants - Theory and Applications. vol. 1, pp. 227–252 (2014)
23. Oulefki, A., Agaian, S., Trongtirakul, T., Laouar, A.K.: Automatic covid-19 lung infected region segmentation and measurement using ct-scans images. Pattern Recognition **114**, 107747 (2021)

24. University of Oxford: Coronavirus pandemic (covid-19) – the data. https://ourworldindata.org/coronavirus-data (2021), online; accessed 30 June 2021
25. Putzu, L., Di Ruberto, C.: Rotation invariant co-occurrence matrix features. In: 19th International Conference ICIAP on Image Analysis and Processing, vol. 10484, pp. 391–401. Springer (2017)
26. Putzu, L., Loddo, A., Ruberto, C.D.: Invariant moments, textural and deep features for diagnostic MR and CT image retrieval. In: Tsapatsoulis, N., Panayides, A., Theocharides, T., Lanitis, A., Pattichis, C.S., Vento, M. (eds.) Computer Analysis of Images and Patterns - 19th International Conference, CAIP 2021, Virtual Event, September 28-30, 2021, Proceedings, Part I. vol. 13052, pp. 287–297 (2021)
27. Rahimzadeh, M., Attar, A., Sakhaei, S.M.: 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**, 102588 (2021)
28. Roberts, M., Driggs, D., Thorpe, M.e.a.: Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans. Nature Machine Intelligence **3**, 199–217 (2021)
29. Signoroni, A., Savardi, M., Benini, S., Adami, N., Leonardi, R., Gibellini, P., Vaccher, F., Ravanelli, M., Borghesi, A., Maroldi, R., Farina, D.: Bs-net: Learning covid-19 pneumonia severity on a large chest x-ray dataset. Medical Image Analysis **71**, 102046 (2021)
30. Sivaranjini, S., Sujatha, C.: Deep learning based diagnosis of parkinson's disease using convolutional neural network. Multimedia Tools and Applications **79**(21), 15467–15479 (2020)
31. Soda, P., D'Amico, N.C., Tessadori, J., Valbusa, G., Guarrasi, V., Bortolotto, C., Akbar, M.U., Sicilia, R., Cordelli, E., Fazzini, D., Cellina, M., Oliva, G., Callea, G., Panella, S., Cariati, M., Cozzi, D., Miele, V., Stellato, E., Carrafiello, G., Castorani, G., Simeone, A., Preda, L., Iannello, G., Bue, A.D., Tedoldi, F., Alì, M., Sona, D., Papa, S.: Aiforcovid: Predicting the clinical outcomes in patients with COVID-19 applying AI to chest-x-rays. an italian multicentre study. Medical Image Anal. **74**, 102216 (2021)
32. Tu, X., Xie, M., Gao, J., Ma, Z., Chen, D., Wang, Q., Finlayson, S.G., Ou, Y., Cheng, J.Z.: Automatic categorization and scoring of solid, part-solid and non-solid pulmonary nodules in ct images with convolutional neural network. Scientific Reports **7**(1), 1–10 (2017)
33. Zhang, B., Qi, S., Monkam, P., Li, C., Yang, F., Yao, Y.D., Qian, W.: Ensemble learners of multiple deep cnns for pulmonary nodules classification using ct images. IEEE Access **7**, 110358–110371 (2019)
34. Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., Zha, Y., Liang, W., Wang, C., Wang, K., et al.: Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. Cell **181**(6), 1423–1433 (2020)
35. Zhao, J., Zhang, Y., He, X., Xie, P.: Covid-ct-dataset: a ct scan dataset about covid-19. arXiv preprint arXiv:2003.13865 (2020)
36. Zhao, W., Jiang, W., Qiu, X.: Deep learning for covid-19 detection based on ct images. Scientific Reports **11**(1), 1–12 (2021)