An Approach of Dempster-Shafer Decision Fusion to Diagnose COVID-19 in Chest X-ray Imagery by Using Controlled Asymmetric Training of the two CNNs Ensemble

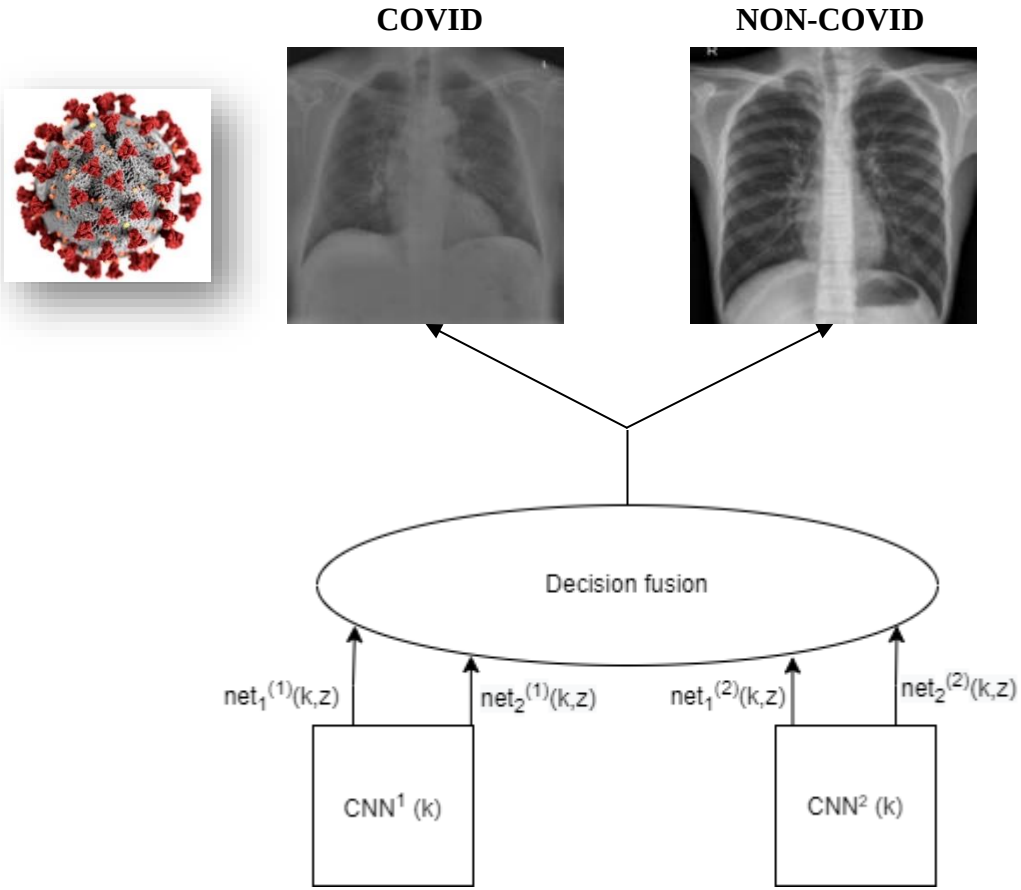
Victor-Emil Neagoe[1], and Gabriela-Loredana Ghenea[2]

[1,2] Faculty of Electronics, Telecomm, & Inform. Technol.,
Polytechnic University of Bucharest, Bucharest, Romania

Overview

- **This paper proposes a model of Dempster-Shafer decision fusion based on controlled training of the ensemble of two Convolutional Neural Networks (CNNs) by the asymmetry parameter k, defined as the ratio of the numbers of training data per class assigned to each CNN module.**
- **The proposed model is dedicated to COVID-19 diagnose in chest X-ray imagery. We have considered two CNN modules with identical architectures. First CNN module has been trained with 2837 COVID-19 labeled images and (2837/k) NON-COVID images. Second CNN module has been trained with (2837/k) COVID-19 labeled images and 2837 NON-COVID images. We have evaluated the influence of control parameter k on the diagnosis performances.**
- **As a result of Dempster-Shafer fusion, for k=2.1, one obtains a maximum Overall Accuracy (OA) of 95.18% The above performance is clearly better than the corresponding OA obtained by a single CNN (92.26%) for the same k, and at the same time it is better than OA obtained by any single CNN module for any considered k. Moreover, one can remark, that by controlled training, for k=20, a CNN module can lead to an incredible low Missing Alarm Rate (MAR) of only 0.63%!**

Dempster-Shafer Decision Fusion to Diagnose COVID-19



I. Introduction

- The SARS-CoV-2 virus has suffered genetic mutations so that now there are multiple variants of this virus. This complicates all the studies in progress since a treatment developed for a specific variant of virus may not be as effective for another mutation. According to [1] there are more than 6 million deaths caused by COVID-19

- In [2] multiple classifiers were used to differentiate COVID-19 from other lung diseases. In the above paper, E. J. Mortani Barbosa Jr et al have reported a classification performance of 94%.

- The variations of this disease come with multiple types of symptoms, including malfunction of respiratory system, which is being analyzed in this paper by exploring the X-Ray COVID-19 diagnosis.

- As presented in [3] and [4], to overcome the problem of limited datasets, we have used the idea to maximize the decision accuracy by combining two classifiers that have been independently trained. Keidar et. al [5] have applied an ensemble of neural networks to diagnose COVID-19 based on X-Ray images. This method led to an accuracy of 90.3%.

I. Introduction

- **We propose a deep learning technique based on asymmetric training of an ensemble of two Convolutional Neural Networks (CNNs) of identical architectures but with opposite and asymmetric kind of training. The proposed classifier is applied for COVID-19 diagnosis based on X-ray imagery.**

- **We have used the Dempster-Shafer theory to calculate belief functions and thus have deduced the final decision. Being a generalization of Bayesian theory, Dempster-Shafer theory of evidence allows us to combine degrees of belief based on independent evidence [6], [7].**

- **We have defined the asymmetry control parameter k as the ratio between the volumes of labeled training data of the two classes for a given CNN module. We have evaluated the performances of proposed COVID-19 diagnose classifier as a function of parameter k, considered in the interval [1].**

II. Proposed decision fusion classifier with asymmetric training using the ensemble of two CNN modules

A. Asymmetrical Controlled Training

- The concept of the project is to implement two identical CNN modules that are asymmetrically trained. We have used a dataset that is available on the Mendeley public website [9]. An example of images extracted from the database is illustrated in Fig. 2. The considered Mendeley dataset images have been normalized and triplicated to the sizes of 224x224x3 pixels, corresponding to the input specifications of VGG. We have applied transfer learning, so that the convolutional layers have been previously trained on the ImageNet dataset. In this paper, the term "training" refers only to the fully-connected (dense) layers of CNN.
- We have chosen an equal number of images of each class and split the data into 70% images for training and 30% images for testing. This way, the training set is composed by 2837 images of healthy patients and 2837 images of patients with COVID-19 disease.
- We have defined the asymmetry parameter k to give us control on the training process as follows: the first CNN classifier is trained with 2837 images from COVID class and 2837/k images from NON-COVID class; the second CNN is trained with 2837/k images with COVID and 2837 NON-COVID images. Choosing this procedure, we obtain two CNN units trained for different tasks: first module has better performance on detecting COVID-19 in X-ray images and the second module is better specialized on detecting the NON-COVID images (without COVID-19 pathologies).

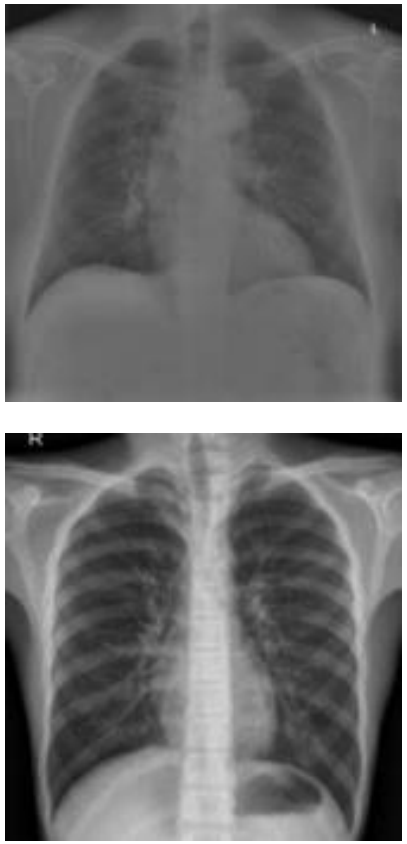


Fig. 2. Examples of chest X-ray images from the database: COVID class (up) and NON-COVID class (down)

II. Proposed decision fusion classifier with asymmetric training using the ensemble of two CNN modules

B. Dempster-Shafer Theory

- The proposed architecture is composed by two CNN classifiers that generate decision probabilities from two different points of view, so the aim is to combine these partial decisions into a final decision. To achieve this, we have chosen the Dempster-Shafer theory of evidence [6], [7]. The architecture of the corresponding system is described in Fig. 3.
- The set of concurrent classifiers consists in CNN^1 (k) and CNN^2 (k), where k is the asymmetry control parameter. The output NET function of each unit is $net_p^{(n)}(k, z)$, where z is the test image, p is the class index and n is the CNN module index.

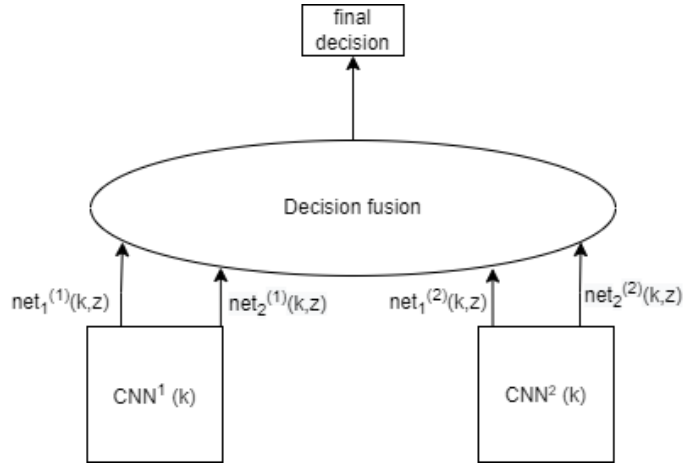


Fig. 3. Proposed decision fusion system

II. Proposed decision fusion classifier with asymmetric training using the ensemble of two CNN modules

C. Proposed Architecture

- As we have mentioned before, in this paper we use two CNN units with identical architectures. We have chosen the VGG-Net architecture [8]. Multiple experiments have been conducted on non-medical images and VGG-Net has proven to be the most performant. The model architecture for any CNN unit is presented in Fig.1.
- VGG-Net is composed of multiple convolutional layers followed by Max-Pooling layers. The kernel size used in this architecture is 3x3. We trained each model for a number of epochs equal with 20. This number has been chosen as a result of multiple experiments. We observed that the performance was not increasing by changing the number of training epochs. The training has been performed on a batch size equal with 64. The optimizer is Adam with a learning rate of 0.0001 and the computed loss is Cross Entropy.

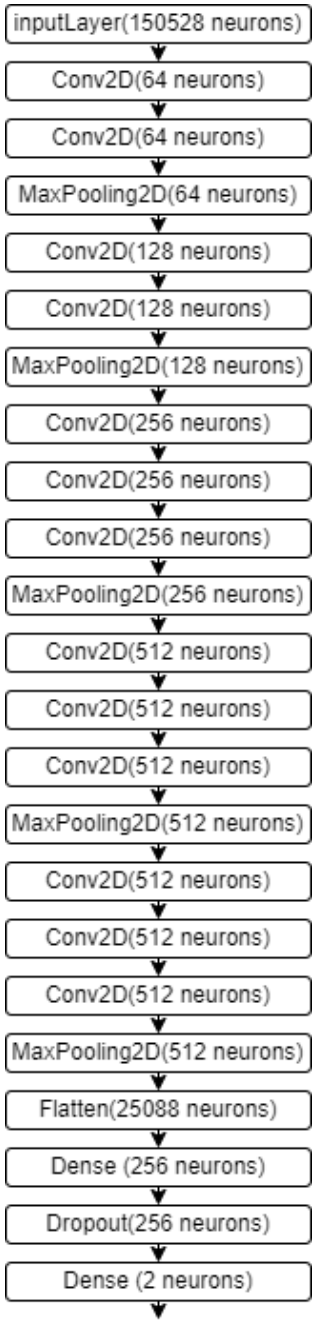


Fig. 1. VGG architecture of each CNN module of the ensemble.

II. Proposed decision fusion classifier with asymmetric training using the ensemble of two CNN modules

A. Asymmetrical Controlled Training

Taking these into consideration, we propose the evidence of belief in the decision made by classifier n regarding class p with reference to input image z to be

$$m_p^{(n)}(k,z) = \frac{1}{1+e^{-net_p^{(n)}(k,z)}} \qquad (1)$$

The decision fusion is further presented.

We compute $m_p^{(n)}(k,z)$ **for p = 1, 2 and n = 1, 2 then we compute the belief functions**

$$m_1(k,z) = \frac{m_1^{(1)}(k,z)\cdot m_1^{(2)}(k,z)}{m_1^{(1)}(k,z)\cdot m_1^{(2)}(k,z)+m_2^{(1)}(k,z)\cdot m_2^{(2)}(k,z)} \qquad (2)$$

and

$$m_2(k,z) = \frac{m_2^{(1)}(k,z)\cdot m_2^{(2)}(k,z)}{m_1^{(1)}(k,z)\cdot m_1^{(2)}(k,z)+m_2^{(1)}(k,z)\cdot m_2^{(2)}(k,z)} \qquad (3)$$

Consider the competition rule $m_j(k,z) = MAX\{m_1(k,z), m_2(k,z)\}$**, and thus, the final decision is that the image z is assigned to class j.**

III. Experiments and results

▶To analyze the performance of the system, we further evaluate the results, by considering the following performance parameters [10], [4]

- **Overall Accuracy**

$$OA = 100 \times \frac{TP+TN}{TP+TN+FP+FN} \qquad (4)$$

- **False Alarm Rate**

$$FAR = 100 \times \frac{FP}{TN+FP} \qquad (5)$$

- **Miss Alarm Rate**

$$MAR = 100 \times \frac{FN}{TP+FN} \qquad (6)$$

where

▶**TP (*true positives)* = number of COVID images that have been correctly diagnosed as being COVID**

▶**TN (*true negatives)* = number of images belonging to patients without COVID that have been correctly classified as NON-COVID**

▶**FP (*false positives)* = number of NON-COVID images that have been erroneously predicted as being COVID**

▶**FN (*false negatives)* = number of -COVID images that have been erroneously classified as being NON-COVID**

III. Experiments and results

k	OA_CNN1 [%]	MAR_CNN1	FAR_CNN1
1	93.87	7.95	4.12
1.1	93.70	8.18	4.22
1.2	94.69	4.51	6.06
1.3	94.77	4.12	6.26
1.4	94.53	3.83	6.98
1.5	94.44	5.80	5.29
1.6	94.20	6.40	5.17
1.7	94.12	3.94	7.65
1.8	94.57	5.35	5.50
1.9	93.95	6.22	5.86
2	94.2	5.17	6.40
2.1	92.22	2.23	12.15
2.2	93.37	8.23	4.87
2.3	94.57	6.14	4.68
2.4	92.26	10.64	4.35
2.5	92.76	3.90	10.12
3	92.1	3.11	11.79
5	88.03	3.14	17.99
7	85.85	1.55	21.54
10	85.27	1.58	22.25
15	80.63	0.66	27.78
20	75.45	0.63	32.85

TABLE 1. EXPERIMENTAL CLASSIFICATION PERFORMANCES OF FIRST CNN MODULE AS A FUNCTION OF ASYMMETRY CONTROL PARAMETER K

- **Table 1 shows the results that have been obtained for the first CNN module. Acting individually, this module has an Overall Accuracy of 94.77%. This result has been obtained for an asymmetry parameter k = 1.3. An important remark is that once the asymmetry parameter increases, the MAR rate decreases, reaching a minimum of 0.63. This means that the module has learned not to miss almost any COVID test image**

III. Experiments and results

k	OA_CNN2 [%]	MAR_CNN2	FAR_CNN2
1	93.99	8.20	3.55
1.1	94.16	7.24	4.33
1.2	94.40	6.23	4.92
1.3	94.07	6.84	4.95
1.4	92.92	10.09	3.55
1.5	92.72	10.19	3.90
1.6	94.65	6.69	3.90
1.7	92.68	10.50	3.57
1.8	93.50	6.28	6.71
1.9	92.84	10.16	3.64
2	93.25	9.72	3.28
2.1	92.26	11.00	3.85
2.2	92.55	10.76	3.50
2.3	90.78	14.00	2.94
2.4	92.51	9.92	4.72
2.5	91.57	11.44	4.90
3	92.92	9.59	4.21
5	85.3	22.17	1.35
7	82.68	25.23	1.93
10	85.56	20.59	4.99
15	69.77	37.55	1.80
20	64.51	41.41	2.93

TABLE 2. EXPERIMENTAL CLASSIFICATION PERFORMANCES OF SECOND CNN MODULE AS A FUNCTION OF ASYMMETRY CONTROL PARAMETER K

- **Table 2 shows the performance of the second CNN module. One can remark that the overall performance of this module is lower than the first CNN. A reason for this behavior is that CNN2 has been trained with a bigger number of NON-COVID images than that of COVID images. However, in this manner, CNN2 leads to a minimum FAR of 1.35 for k=5.**

III. Experiments and results

k	OA_FUSION [%]	MAR_FUSION	FAR_FUSION
1	94.65	7.31	3.18
1.1	95.14	6.42	3.15
1.2	94.81	5.40	4.95
1.3	94.81	5.25	5.10
1.4	94.81	6.46	3.81
1.5	94.07	7.39	4.34
1.6	94.90	6.45	3.65
1.7	94.86	5.97	4.27
1.8	94.94	4.94	5.16
1.9	95.02	6.71	3.08
2	94.86	6.04	4.19
2.1	95.18	4.54	5.06
2.2	93.79	8.68	3.41
2.3	93.87	8.48	3.48
2.4	93.33	9.27	3.69
2.5	94.07	6.06	5.77
3	94.28	5.01	6.39
5	93.5	5.69	7.26
7	93.21	5.42	8.06
10	92.31	5.91	9.32
15	92.64	7.43	6.87
20	87.95	5.41	16.96

TABLE 3. EXPERIMENTAL CLASSIFICATION PERFORMANCES OF PROPOSED DECISION FUSION MODEL AS A FUNCTION OF ASYMMETRY CONTROL PARAMETER K

- **In Table 3 there are shown the results obtained for the fusion algorithm using Dempster-Shafer theory of evidence. One can observe that the fusion algorithm leads to a higher performance than a single CNN module.**

III. Experiments and results

k	OA_CNN1 [%]	OA_CNN2 [%]	OA_FUSION [%]
1	93.87	93.99	94.65
1.1	93.70	94.16	95.14
1.2	94.69	94.40	94.81
1.3	94.77	94.07	94.81
1.4	94.53	92.92	94.81
1.5	94.44	92.72	94.07
1.6	94.20	94.65	94.90
1.7	94.12	92.68	94.86
1.8	94.57	93.50	94.94
1.9	93.95	92.84	95.02
2	94.2	93.25	94.86
2.1	92.22	92.26	95.18
2.2	93.37	92.55	93.79
2.3	94.57	90.78	93.87
2.4	92.26	92.51	93.33
2.5	92.76	91.57	94.07
3	92.1	92.92	94.28
5	88.03	85.3	93.5
7	85.85	82.68	93.21
10	85.27	85.56	92.31
15	80.63	69.77	92.64
20	75.45	64.51	87.95

TABLE 4. OVERALL ACCURACY (OA) FOR EACH CNN MODULE AND FOR DECISION FUSION AS A FUNCTION OF ASYMMETRY CONTROL PARAMETER K

- **As highlighted in Tables 3 and 4, the best results have been obtained for k = 2.1. For this specific value of asymmetry, we have obtained the maximum OA-FUSION (95.18%) and the minimum MAR_FUSION (4.54). The main purpose of the diagnose (here based on fusion using Dempster-Shafer theory) is to maximize the Overall Accuracy (OA) and to minimize the Miss Alarm Rate (MAR).**

III. Experiments and results

- **The control parameter k has an important influence on the results, not only on the final decision, but also on the individual CNN components. For example, for k = 1.3, we can see that the OA_CNN1 and MAR_CNN1 performance indices are significantly improved by comparison to the ones obtained for k = 1 (without asymmetry). To better understand the influence of parameter k, we can look at Fig. 4**

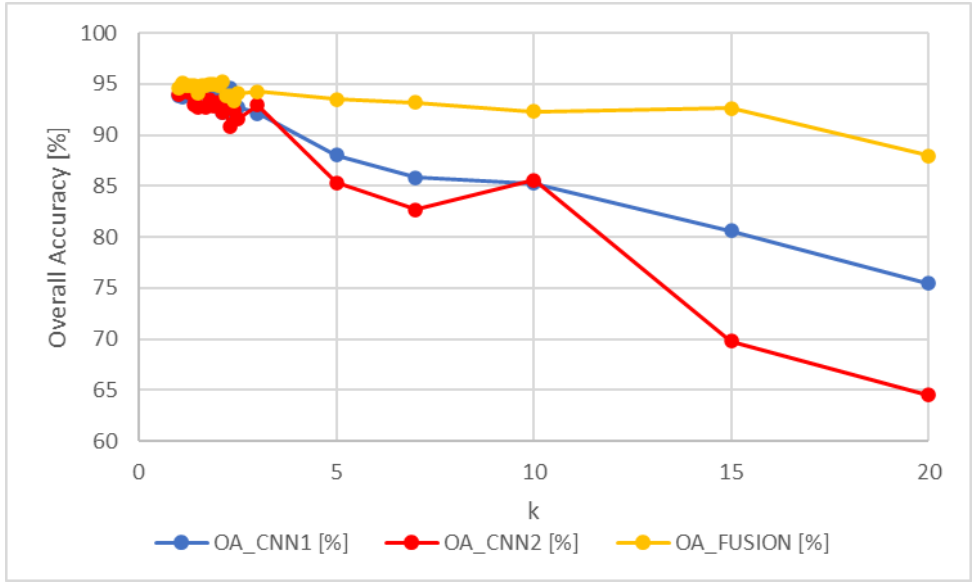


Fig. 4. Overall Accuracy (OA) as a function of asymmetry control parameter k

III. Experiments and results

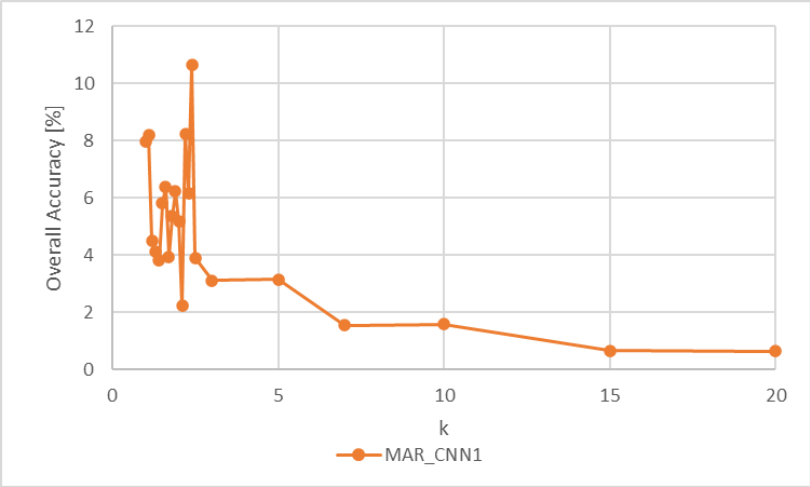


Fig. 5. Miss Alarm Rate (MAR_CNN1) as a function of control parameter k

- In Fig. 5 there is shown the Miss Alarm Rate (MAR) as a function of asymmetry parameter k for the first CNN module and respectively for decision fusion. As discussed, the first CNN module has been trained with 2837 images of patients who suffer of COVID-19 disease and 2837/k images with NON-COVID-19 patients. Fig. 5 shows that the MAR value is minimized with the increase of k, which demonstrates that CNN1 has been designed to correctly classify images from COVID class.

III. Experiments and results

- **On the other side, the CNN2 has been trained with 2837/k images from COVID class and 2837 images from NON-COVID class. We can see in Fig. 6 that by increasing control parameter k, we can significantly reduce the False Alarm Rate, which is in a good agreement with the fact that CNN2 is experienced in detecting images that are part of NON-COVID class. One can obtain the minimum FAR_CNN2 of 1.35 for k=5.**

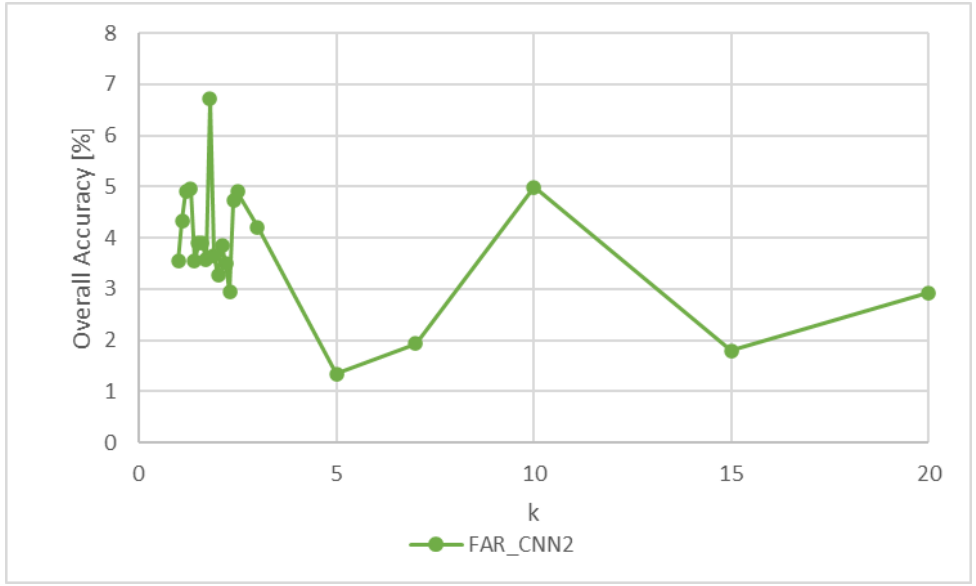


Fig. 6. False Alarm Rate (FAR_CNN2) as a function of control parameter k

III. Experiments and results

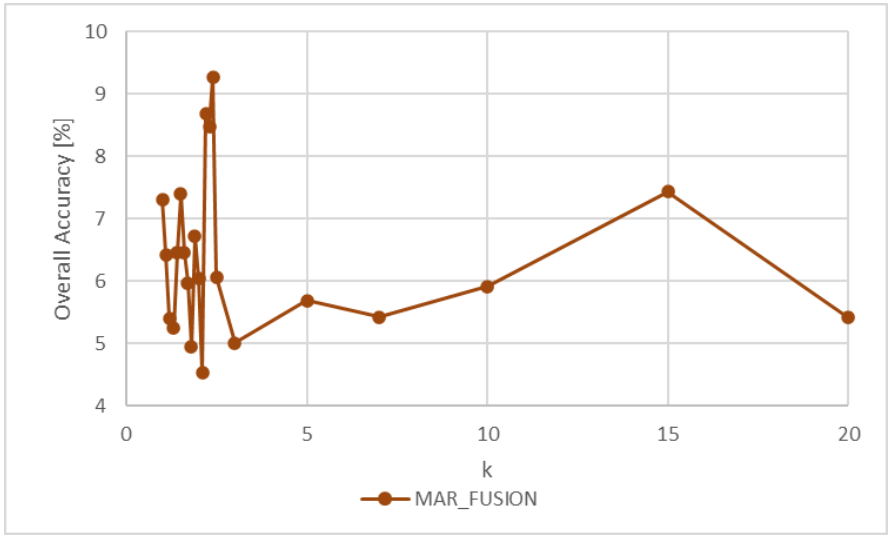


Fig. 7. Miss Alarm Rate after the Dempster-Shafer fusion (MAR_FUSION) as a function of control parameter k

- **The aim of the diagnose (performed by decision fusion algorithm) is to minimize the MAR rate (as well as to maximize OA). As expected, it can be concluded from Fig. 7 that our proposed model gives satisfactory results for the purpose of minimizing MAR (as 4.65 for k=2.1).**

IV. Conclusions

- **In this paper we have proposed, implemented, and evaluated an exciting model of chest X-ray imagery classification to diagnose COVID-19 by using the combination of following techniques: asymmetric controlled training of the ensemble of two CNNs, decision fusion according to Dempster-Shafer theory of evidence, and optimization of training control parameter k.**

- **By choosing the control parameter k=2.1, one obtains by fusion a maximum Overall Accuracy (OA_FUSION) of 95.18%; this performance is better than the corresponding OA obtained for the same k by any of the two CNNs (92.26%), and at the same time it is better than OA obtained by any single CNN module, for any k in the interval [1, 20].**

- **Moreover, by controlling the training with the asymmetry parameter k, one can improve the performances obtained by a single CNN classifier (without fusion). For example, if we choose k=20, the first CNN module can lead to an extremely low MAR of only 0.63%.**

- **At the same time, by choosing k=1.3, we have obtained OA_CNN1 of 94.77% and MAR_CNN1 of 4.12%, performances that are better than those obtained for k=1 (without asymmetry of class training data volumes), namely OA_CNN1 of 93.87% and MAR_CNN1 of 7.95%. This means the asymmetry parameter k (defined as the ratio between the volumes of labeled training data of the two classes) can be useful to optimize the CNN classification performance even in the case without decision fusion.**

References

[1] WHO COVID-19 Dashboard. World Health Organization. Available online: https://covid19.who.int/ .

[2] E. J. Mortani Barbosa Jr, B. Georgescu, S. Chaganti, G. B. Aleman, J. Broncano Cabrero, G. Chabin, T. Flohr, P. Grenier, S. Grbic, N. Gupta, F. Mellot, S. Nicolaou, T. Re, P. Sanelli, A. W. Sauter, Y. Yoo, V. Ziebandt, D. Comaniciu, "Machine learning automatically detects COVID-19 using chest CTs in a large multicenter cohort," European Radiology, vol. 31, pp. 8775-8785, May 2021, Springer.

[3] V. E. Neagoe, P. Diaconescu, "An ensemble of deep convolutional neural networks for drunkenness detection using thermal infrared facial imagery," Proc. 13th International Conference on Communications, June 18-20, 2020, Bucharest, Romania, pp 147-150.

[4] L. G. Ghenea, V. E. Neagoe, "Concurrent convolutional neural networks with decision fusion to diagnose COVID-19 using chest X-ray imagery," Proc. 13 th International Conference on Electronics, Computers, and Artificial Intelligence (ECAI), July 1-3, 2021, Pitesti, Romania, pp. 1-4.

[5] D. Keider, D. Yaron, E. Goldstein, A. Blass, L. Charbinsky, I. Aharony, L. Lifshitz, D. Lumelsky, Z. Neeman, M. Mizrachi, M. Hajouj, N. Eizenbach, E. Sela, C. S. Weiss, P. Levin, O. Benjaminov, G. N. Bachar, S. Tamir, Y. Rapson, D. Suhami, E. Atar, A. Dror, N. Bogot, A. Grubstein, N. Shabshin, Y. Elyada, Y. Eldar, "COVID-19 classification of X-ray images using deep learning networks," European Radiology, vol. 31, pp. 9654-9663, May 2021, Springer.

[6] G. Shafer, A Mathematical Theory of Evidence, vol. 42. Princeton, NJ, USA : Princeton Univ. Press, 1976.

[7] V. E. Neagoe, A. Ropot, and A. Mugioiu, "Real time face recognition using decision fusion of neural classifiers in the visible and thermal infrared spectrum," Proc. of the 2007 IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS 2007), London (United Kingdom), 5-7 September 2007, pp.1-6.

[13] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Proc. International Conference on Learning Representation (ICLR-2015), May 7-9, 2015, San Diego, California, USA, pp. 1-14.

[14] W. El-Shafai, F. Abd El-Samie, "Extensive COVID-19 X-Ray and CT chest images dataset," (Mendeley Data), version 3, DOI: 10.17632/8h65ywd2jr.2, June 2020.

[15] V. Neagoe, Neural Networks for Data Mining (in Romanian), MatrixRom, Bucharest, 2018.

THANK YOU!