AI-based solutions for the analysis of biomedical images and signals

Fabiola De Marco[1,*], Alessia Auriemma Citarella[1], Luigi Di Biasi[1], Lorenzo D'Errico[3], Rita Francese[1], Giovanni Mettivier[3], Mariacarla Staffa[2] and Genoveffa Tortora[1]

[1]*University of Salerno, Department of Computer Science, Italy*
[2]*Parthenope University of Naples, Italy*
[2]*University Federico II, Italy*

Abstract
Artificial intelligence (AI) has emerged as a disruptive technology that is transforming the medical field. The use of AI algorithms, machine learning, and neural networks has revealed a great potential in enhancing healthcare delivery, including early detection and diagnosis of diseases, improving patient outcomes, and reducing healthcare costs. The application of AI in medical imaging analysis is going to obtaining remarkable success in identifying tumors and other anomalies, leading to earlier diagnoses and better treatment outcomes. In addition, AI has also been applied to medical data analysis, drug discovery, and personalized medicine. In this review article we present an overview of the research work conducted by our team in the field of medical AI. Our primary focus has been on the development and validation of AI-based approaches for the early detection and accurate diagnosis of breast cancer, skin melanoma, and heart disease. Our research has demonstrated the potential of AI in improving the accuracy and efficiency of medical diagnosis and treatment. With the growing availability of medical data and advances in AI technology, we anticipate that the future of medical AI will bring about a paradigm shift in the way healthcare is delivered.

Keywords
Artificial Intelligence, neural network, classification, data augmentation, digital breast tomosynthesis, Generative Adversarial Network, Deep Convolutional Neural Network, ECG, pattern detection, skin melanoma

1. Introduction

Artificial Intelligence (AI) is emerging as a game-changing technology in the field of medicine, capable of ushering in a new era of patient-centered care. By leveraging machine learning (ML), deep learning (DL), natural language processing (NLP), and other AI-based technologies, medical practitioners can make more informed and accurate decisions, leading to improved healthcare outcomes. AI-based algorithms can mine vast amounts of data from electronic health records, medical images, genomic data, and other sources, providing clinicians with critical insights and enabling them to make more personalized treatment decisions. Moreover, the potential of AI extends beyond clinical settings, with applications in telemedicine, remote patient monitoring, and public health management. AI-powered devices such as wearables and sensors can help monitor patient health and alert clinicians to potential health risks before they escalate. The development of innovative solutions for early disease diagnosis is a critical objective in the biomedical field. In this context, the project utilizing AI-based techniques to analyze biomedical images and signals is well-aligned. By leveraging advanced algorithms and machine learning models, this project aims to provide accurate and efficient diagnosis results to healthcare professionals. With the potential to detect diseases at their earliest stages, these solutions have the potential to revolutionize the healthcare industry and improve patient outcomes. Overall, our project represents a significant step forward in the quest to develop innovative solutions for the early diagnosis of diseases.

2. Research Fields

The application of cutting-edge AI methods to the task of supporting medical care is one of the primary focuses of our research group. We concentrate our research activity mainly in medical image analysis, which involves the development of AI-based algorithms to analyze medical images and supporting the detection and diagnosis of diseases.

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
fdemarco@unisa.it (F. D. Marco); aauriemmacitarella@unisa.it (A. A. Citarella); ldibiasi@unisa.it (L. D. Biasi); lorenzo.derrico@unina.it (L. D'Errico); francese@unisa.it (R. Francese); giovanni.mettivier@unina.it (G. Mettivier); mariacarla.staffa@uniparthenope.it (M. Staffa); tortora@unisa.it (G. Tortora)
0000-0003-4285-9502 (F. D. Marco); 0000-0002-6525-0217 (A. A. Citarella); 0000-0002-9583-6681 (L. D. Biasi); 0000-0001-8044-8224 (L. D'Errico); 0000-0002-6929-0056 (R. Francese); 0000-0001-6606-4304 (G. Mettivier); 0000-0001-7656-8370 (M. Staffa); 0000-0003-4765-8371 (G. Tortora)

CEUR Workshop Proceedings (CEUR-WS.org)

Currently, we are investigating a variety of medical imaging techniques, including tomosynthesis, electrocardiogram (ECG), dermoscopy, and clinical. By leveraging the power of AI, we aim at developing more accurate and efficient methods for identifying and diagnosing a wide range of conditions, from cardiovascular disease to skin cancer.

Concerning tomosynthesis analysis, we focused on Digital Breast Tomosynthesis (DBT) images facing two crucial and strongly correlated aspects of deploying Computer Aided Diagnosis (CAD) systems: classification and data augmentation. The classification task was carried out using a Deep Convolutional Neural Network (DCNN) for mass- and micro-calcification detection, a state-of-the-art image analysis model that was contextually adjusted. The need for huge volume of data for training such models represents a crucial problem due to privacy restrictions, great effort for labelling images, costs of the medical exams and so on. Thus, there is the need of investigating affordable data augmentation technique. A novel approach to generative models, namely Evolutionary Generative Adversarial Network (E-GAN) has been used to perform data augmentation on the same kind of images.

Similarly, in the case of ECG, we are exploring how AI can be used to analyze cardiac signals and identify patterns that may indicate the presence of underlying conditions. By developing algorithms that can detect these patterns quickly and accurately, we hope to improve the accuracy of diagnoses and ultimately improve patient outcomes.

Finally, in the case of dermoscopic and clinical images, we are exploring how AI can be used to analyze images of the skin and detect early signs of skin cancer. By identifying suspicious lesions and providing early intervention, we hope to improve the prognosis for patients with this deadly disease.

These activities involve the University of Salerno, Parthenope University of Naples, and the University of Naples Federico II. The CAISLab laboratory (Computer Science Department of the University of Salerno) provides support for the development of time and cost-consuming algorithms.

3. AI Applications on biomedical images and signals

3.1. Classification of melanoma images

Melanoma is a severe form of skin cancer that accounts for approximately 99,780 new malignant diagnoses each year [1].It come in a variety of shapes, sizes, and colors, and affect people of all skin types. In this scenario, early detection is critical to ensure patient survival [1].

In the following we describe the main results of our research related to the melanoma detection.

3.1.1. Minimization of False Negative Rate

It is important to provide diagnostic support tools able to achieve both high accuracy and minimize type 2 errors, related to the *False Negative Rate* (FNR). Consequently, we explored the behavior of nine types of CNNs (including Alexnet, DenseNet, GoogleNet Inception V3, GoogleNet, MobileNet, ShuffleNet, SqueezeNet, and VGG16) on MED-NODE, a dataset of 170 clinical images [2]. To investigate the impact of data augmentation and image pre-processing on the final classification performance, four datasets were generated from the original (INA, NIA, IA, NINA), with pre-processing quality step (IIQ) and a simple segmentation process (OTSU). INA contains MED-NODE original images by using quality improved and data augmentation techniques; NIA is composed by MED-NODE original images not quality improved but using data augmentation techniques; IA includes the NSA images by using quality improved and data augmentation techniques while NINA is the original MED-NODE dataset not improved and not data augmented. All tested neural networks perform better without data augmentation, with AlexNet and SqueezeNet achieving a maximum accuracy of 78%. Without pre-processing and data augmentation, AlexNet performed best with 89%, 75% and 82% of accuracy, sensitivity, and specificity, respectively, as shwon in Fig. 1. VGG can guarantee the lowest FNR at the expense of global accuracy, whereas AlexNet can guarantee comparable FNR to VGG but with the highest global accuracy.

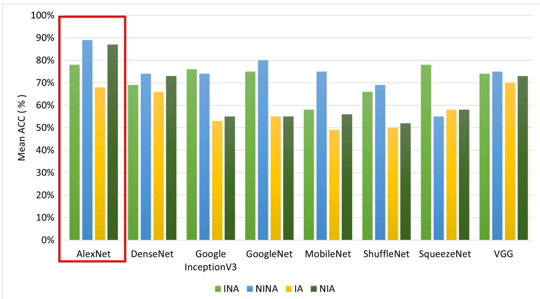

Figure 1: The global performances of the CNN on the four datasets

3.1.2. CNNs design employing genetic algorithms

In [3] we adopted genetic algorithms to designing an architecture for a convolutional neural network (Fig. 2).

[1] https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html

The goal is to determine the optimal neural network structure for melanoma classification. In our previous work, we used the same approach on clinical dataset [4]. In this work, a revised subset of images from ISIC, one of the most widely used datasets for melanoma classification, were employed in an experimental study to test the proposed methodology. The convolutional neural network architecture uses a genetic algorithm which enables the population to evolve over successive generations in order to obtain the best fitness. The initial generation of neural network (NN) is stochastic. As a result, the initial accuracy is remarkably low, but as the experiment progresses, the error attenuates due to a higher level of fitness in the NN population. In the most recent evolutionary iteration, a set of equivalent NN with high classification performance was available. Our hybrid approach to melanoma detection CNN design achieves 94% accuracy, 90% sensitivity, 97% specificity, and 98% precision. According to preliminary results, the proposed method could improve melanoma classification by eliminating the need for user interaction and avoiding a priori network architecture selection.

Figure 2: Structure of the best network

3.1.3. A cloud approach for melanoma detection

In [5], we investigated two aspects of melanoma detection problem, compared the performance of AlexNet, Google InceptionV3 and GoogleNet. The first issue is related to the Transfer Learning adoption because we examined how even minor changes to the parameters in the dataset (such as data augmentation and segmentation) affect the accuracy of classifiers. For this purpose, we constructed four datasets from origianal clinical MED-NODE dataset: MD1 (MED-NODE original images), MD2 (MED-NODE images segmented with Otsu), MD3 (MED-NODE images and augmented images without segmentation) and MD4 (MED-NODE images and augmented images segmented with Otsu). Based on the findings of the first analysis, we propose that continuous training-test iterations are required to produce robust prediction models (see Table 1). The second point is the requirement for a more adaptable system architecture that can deal with changes in training datasets.

Net	Measure	MD1	MD2	MD3	MD4
AlexNet	Best	0.97	0.91	0.97	0.89
	Average	0.81	0.72	0.81	0.73
	Drop	-19.75	-26.38	-19.75	-21.91
Google InceptionV3	Best	0.91	0.88	0.90	0.89
	Average	0.75	0.72	0.75	0.74
	Drop	-21.33	-22.22	-2.0	-20.27
GoogleNet	Best	0.94	0.93	0.91	0.89
	Average	0.81	0.77	0.75	0.74
	Drop	-16.04	-20.77	-21.33	-20.27

Table 1
Performance drop after 100 training steps

In this context, we designed and deployed a hybrid architecture based on Cloud, Fog, and Edge Computing to provide a Melanoma Detection service based on clinical and dermoscopic images (see Fig. 3). Data buckets are kept in the cloud, and system training is executed. After each formation in the Fog area, where services are executed, the orchestrator is in charge of distributing the optimized services. Local calculations are performed in the Edge area on IoMT devices (for example, smartphones). HiC-Otsu is a software component of the Fog system on the IoMT device that performs preliminary data analysis. To improve system performance, the QoS moderator annotates content. The generic user uses the output of services, but by loading data, he contributes to the system's knowledge base.

3.1.4. Real Time melanoma detection support

In [6] we proposed an augmented reality smartphone application for supporting the dermatologist in the real-time analysis of a skin lesion. The app augments the camera view with information related to the lesion features generally measured by the dermatologist for formulating the diagnosis. The lesion is also classified by a deep learning approach for identifying melanoma. Results revealed that the real-time process may be entirely executed on the smartphone and that the support provided is well judged by the target users.

3.2. Classification of ECG images

3.2.1. Classification of Premature Ventricular Contractions

Cardiovascular disease (CVD) is still the leading cause of death worldwide. According to the World Health Organization (WHO), in 2017 more than 17.9 million people died from cardiovascular disease (31% of all deaths worldwide). Premature Ventricular Contraction (PVC) is an additional heartbeat that occurs in one of two heart ventricles that delays the normal pumping order, first the atria, then the ventricles. PVC is of crucial importance in the cardiology field, not only to improve the health system but also to reduce the workload of experts who analyze ECG manually.

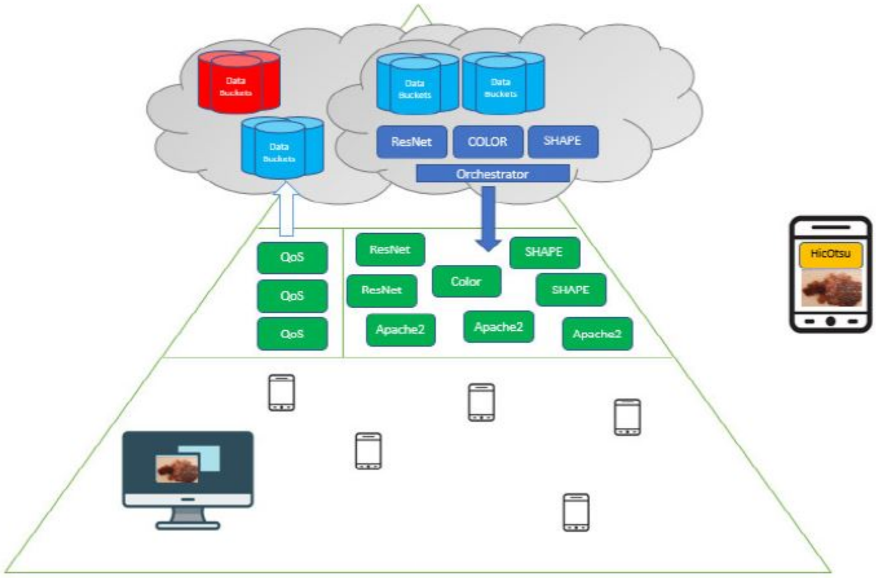

Figure 3: General operation of the three layers architecture for melanoma detection

PVC is a non-harmful common occurrence represented by extra heartbeats, whose diagnosis is not always easily identifiable, especially when done by long-term manual ECG analysis. In some cases, it may lead to disastrous consequences when associated with other pathologies. This work introduces an approach to classify PVCs using machine learning techniques [7]. In particular, a group of six classifiers is used: Decision Tree, Random Forest, Long-Short Term Memory (LSTM), Bidirectional LSTM, ResNet-18, MobileNetv2, and ShuffleNet. Two types of experiments are performed on data extracted from the MIT-BIH Arrhythmia database: *(i)* the original dataset and *(ii)* the balanced dataset. MobileNetv2 came in first in both experiments with high performance and promising results for PVCs' final diagnosis. The final results showed 99.90% of accuracy in the first experiment and 99.00% in the second one, despite no feature detection techniques were used. The approach we used, which was focused on classification without feature extraction, allowed us to dramatically reduce costs and computational times while providing excellent performance and obtaining better results. Finally, this research defines as a first step toward understanding the explanations for deep learning models' incorrect classifications.

3.2.2. Identification of a Pattern in Premature Ventricular Contractions

The primary goal of this study is to find PVCs-related patterns in ECG signals [8]. Specifically, our focus was to highlight the latent patterns of non-PVC signal data that indicate belonging to different arrhythmias. We first pre-processed the ECG signals with noise remove technique, and then we created a matrix based on the wave distances of each pair of analyzed images. This distance matrix as used for the next clustering analysis. One of the most important goal of a clustering algorithm is to classify the data into a set of clusters in order to group similar elements. K-means is a method based on the centroids, the points belonging to the space of the features, that mediates the distances between all the items belonging to the identified cluster. This algorithm chooses the k value in an arbitrary way, without knowing the classes in the input dataset. The Elbow method was used to objectively choose the best value for k. As a result, we were able to calculate the optimal number of clusters for the explored data. In particular, this optimization on the k-means analyzes the intrinsic pathological meaning of non-PVC signals. Although these signals are not adequately labeled in the dataset, we could identify possible common patterns connected to various arrhythmias in the non-PVC class. At present, our work is still in the experimental phase. The identification of well-defined clusters allows us to hypothesize the presence of morphological patterns common to different arrhythmias, in particular between PVC and non-PVC as shown in Fig. 4. Future studies will be able to help us better describe the clusters that exist in non-PVCs, making it easier to identify and label them based on their ECG signal feature space.

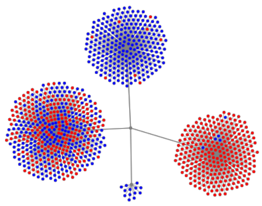

Figure 4: Cluster with PVC (red) and non-PVC (blue)

3.3. Deep learning for breast cancer detection

3.3.1. DCNN for cancer detection in DBT images

Breast cancer is the most common type of cancer in women, and mammography is the most effective method for early detection. However, mammography has limitations in terms of sensitivity and specificity, especially for dense breasts. To overcome these limitations, new imaging technologies such as digital breast tomosynthesis (DBT) and breast computed tomography have been developed. DBT, in particular, offers a (pseudo) three-dimensional representation of the breast tissue and a clearer localization of possible lesions (masses and microcalcifications). However, interpreting a DBT exam requires the analysis of tens of image slices, adding complexity to the radiological clinical workflow. To improve experts' performance in DBT exam analysis, computer-aided detection (CAD) systems have been developed. These systems help in managing the complexity of DBT lesion search space and potentially improve diagnostic accuracy. One specific goal of the CAD system is to achieve an acceptable trade-off between the computational costs arising from automatic analysis and classification performance. Previous studies have focused on developing CAD systems using hand-crafted features. In this study [9], a deep convolutional neural network for DBT images (DBT-DCNN) was developed to automatically classify the presence or absence of mass lesions in DBT exams. The performance of this DCNN was compared to that of popular architectures (AlexNet and VGG19) (see Tab.2).

Table 2
Evaluation in terms of classification absolute numbers, accuracy (acc) and sensitivity (s), of the DCNN architectures.

	TP	TN	FP	FN	acc	s
	(#)	(#)	(#)	(#)	(%)	(%)
TL-AlexNet	940	249	208	9	84.6	99.0
TL-VGG19	832	215	242	17	74.5	87.7
DBT-DCNN	948	374	83	1	94.0	99.0

The classification performance of the three neural networks was comparatively assessed on different datasets provided by two hospitals. This process allowed the evaluation of the robustness of the architecture versus the influence of different hardware instrumentation or acquisition protocols. The study concluded that the DCNN architecture performed better in terms of sensitivity and specificity and had the potential to reduce false positives. Additionally, a technique called Grad-CAM was implemented to highlight pixels in all DBT slices of a given exam that were more relevant for the final classification task performed by the network. This technique could provide an indication of the position of the mass inside the slice(s) classified as abnormal and show the location of possible network activation in zones less relevant for the diagnostic task (see Fig.5).

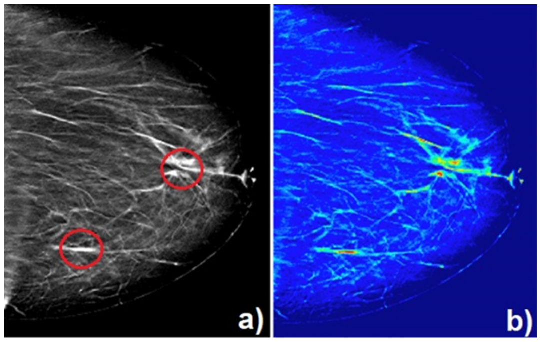

Figure 5: Comparison between the input image (a) and the image to which the Grad-Cam technique has been applied to display the saliency maps of the neural activation gradients (b) belonging to the same patient suffering from infiltrating ductal carcinoma.

3.3.2. EGAN for DBT data augmentation

As discussed in Sec.3.3.1 DBT exam consists of tens of image slices, adding complexity to the radiological clinical workflow and, at the same time, representing a critical problem when approaching to DL solutions: building a dataset of medical images is a complex task due to privacy restrictions, the need for expert clinicians to report the images, and the costs and manual efforts required to process them. An additional challenge is the non-balancing of datasets, especially when a particular class is more abundant than others. To address these challenges, new techniques to augment and balance existing DBT datasets with realistic synthetic samples become necessary. The use of data augmentation techniques, and in particular of generative models, represents a possible solution to the problem. In this work, we investigated an innovative solution in the field of generative models, in particular, Generative Adversarial Network (GAN) models. However, GANs often suffer from training difficulties such as the problem of the gradient vanishing and mode collapse. To cope with these problems, a new GAN architecture, known as Evolutionary GAN (E-GAN), has been designed that uses different metrics together to optimize the generator through an evolutionary approach. In this work [10] such a model has been applied to the problem of increasing the data of a DBT image dataset to generate a more significant number of samples of "sick" slices to balance the starting dataset. In the proposed E-GAN approach, a discriminator acts as an environment while a population of generators mutates to produce offspring that adapt to the environment. Furthermore, once the current best discriminator is fixed, the quality and diversity of the samples generated are assessed, and only the best one for future training steps is kept. A first training session was performed on the entire dataset (results are shown in Fig.6) clearly showing some artifacts: non-continuity in borders, shadows, incoherence of the nipple etc. Although not definitive and usable for classification, the results represent a starting point for the development of future architectures in charge of 2.5D or even 3D.

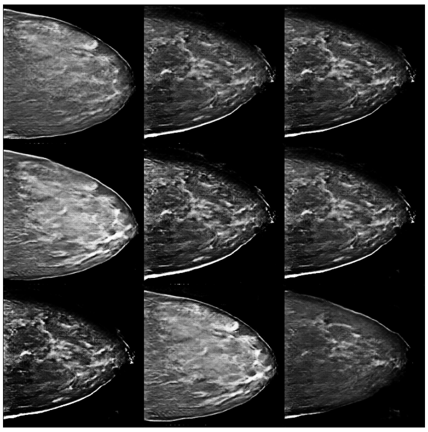

Figure 6: Randomly selected samples obtained after 90 training epochs.

4. Project

We proposed a PRIN on the thematic of reconstruction algorithms for 3D breast computed tomography; Some of the works presented are part of the Artificial Intelligence in Medicine (AIM) project, funded by INFN.

5. Acknowledgements

We acknowledge financial support from the project PNRR MUR project PE0000013-FAIR.

References

[1] N. R. Abbasi, H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. McCarthy, I. Osman, A. W. Kopf, D. Polsky, Early diagnosis of cutaneous melanoma: revisiting the abcd criteria, Jama 292 (2004) 2771–2776.

[2] L. Di Biasi, F. De Marco, A. A. Citarella, M. Castrillón-Santana, P. Barra, G. Tortora, Refactoring and performance analysis of the main cnn architectures: using false negative rate minimization to solve the clinical images melanoma detection problem, BMC Bioinformatics (2022).

[3] L. Di Biasi, F. De Marco, A. Auriemma Citarella, P. Barra, , S. Piotto Piotto, G. Tortora, Hybrid approach for the design of cnns using genetic algorithms for melanoma classification, in: 2nd International Workshop on Artificial Intelligence for Healthcare Applications (AIHA 2022), 2022.

[4] L. Di Biasi, A. Auriemma Citarella, F. De Marco, M. Risi, G. Tortora, S. Piotto, Exploration of genetic algorithms and cnn for melanoma classification, in: J. J. Schneider, M. S. Weyland, D. Flumini, R. M. Füchslin (Eds.), Artificial Life and Evolutionary Computation, Springer Nature Switzerland, Cham, 2022, pp. 135–138.

[5] L. Di Biasi, A. A. Citarella, M. Risi, G. Tortora, A cloud approach for melanoma detection based on deep learning networks, IEEE Journal of Biomedical and Health Informatics 26 (2021) 962–972.

[6] R. Francese, M. Frasca, M. Risi, G. Tortora, A mobile augmented reality application for supporting real-time skin lesion analysis based on deep learning, J. Real Time Image Process. 18 (2021) 1247–1259.

[7] F. De Marco, F. Ferrucci, M. Risi, G. Tortora, Classification of qrs complexes to detect premature ventricular contraction using machine learning techniques, Plos one 17 (2022) e0268555.

[8] F. De Marco, L. Di Biasi, A. A. Citarella, M. Tucci, G. Tortora, Identification of morphological patterns for the detection of premature ventricular contractions, in: 2022 26th International Conference Information Visualisation (IV), IEEE, 2022, pp. 393–398.

[9] R. Ricciardi, G. Mettivier, M. Staffa, A. Sarno, G. Acampora, S. Minelli, A. Santoro, E. Antignani, A. Orientale, I. Pilotti, V. Santangelo, P. D'Andria, P. Russo, A deep learning classifier for digital breast tomosynthesis, Physica Medica 83 (2021) 184–193.

[10] M. Staffa, L. D'Errico, R. Ricciardi, P. Barra, E. Antignani, S. Minelli, G. Mettivier, How to increase and balance current dbt datasets via an evolutionary gan: preliminary results, in: 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), 2022, pp. 913–920.