



Ensemble method using real images, metadata and synthetic images for control of class imbalance in classification

Rogers Aloo¹ · Atsuko Mutoh¹ · Koichi Moriyama¹ · Tohgoroh Matsui² · Nobuhiro Inuzuka¹

Received: 24 March 2022 / Accepted: 5 July 2022
© International Society of Artificial Life and Robotics (ISAROB) 2022

Abstract

Binary classification and anomaly detection face the problem of class imbalance in data sets. The contribution of this paper is to provide an ensemble model that improves image binary classification by reducing the class imbalance between the minority and majority classes in a data set. The ensemble model is a classifier of real images, synthetic images, and metadata associated with the real images. First, we apply a generative model to synthesize images of the minority class from the real image data set. Secondly, we train the ensemble model jointly with synthesized images of the minority class, real images, and metadata. Finally, we evaluate the model performance using a sensitivity metric to observe the difference in classification resulting from the adjustment of class imbalance. Improving the imbalance of the minority class by adding half the size of the majority class we observe an improvement in the classifier's sensitivity by 12% and 24% for the benchmark pre-trained models of RESNET50 and DENSENet121 respectively.

Keywords Image classification · Patient metadata · Chest X-rays · Pneumonia detection · Imbalance data · Image synthesis

1 Introduction

General classification and anomaly detection of images is a major problem in a wide variety of domains. Some research have provided solutions for image detection in fields such as finance, manufacturing, healthcare, and security [1]. Pneumonia is an acute pulmonary infection that can be caused by bacteria, viruses, or fungi and infects the lungs [2].

In normal cases, pneumonia detection is identified through instances of increased opacity's on the X-ray images by radiologists [3]. However, detection of these opacity's may not be easy in some cases depending on a few factors

that determine the interpretation of these images. Additionally, there is a number of research studies in this area of chest X-ray images classification with some good results accurately detecting Pneumonia cases in chest X-ray images [4, 5].

Deep convolution neural networks (DCNN) have been at the forefront of the development of the classification models [6]. For these networks to provide a guarantee of high performance, there is need of a large availability of data sets on which to train the models [7]. Due to the sensitive nature of data sets, especially relating to medical data of patients, the availability of large data sets is a key setback [8]. Additionally, a low number of disease positive cases compared to normal chest X-ray cases exist, since it is infeasible to acquire large amounts of positive data [9]. The setbacks have resulted in innovative ways to be able to increase data set samples for training through image augmentation, sample-pairing, and cut-out methods [10, 11]. The key idea is to help generalize the classification models better through an increase of sample sizes. Arguably the method has provided successful classification outputs [12].

Class-imbalance also proves a major hindrance in obtaining a good generalization models [13]. An ideal situation would be a detection model that has an equal number of normal and abnormal cases on the chest X-rays. However,

This work was presented in part at the joint symposium of the 27th International Symposium on Artificial Life and Robotics, the 7th International Symposium on BioComplexity, and the 5th International Symposium on Swarm Behavior and Bio-Inspired Robotics (Online, January 25–27, 2022)

✉ Rogers Aloo
r.owit.965@stn.nitech.ac.jp
Nobuhiro Inuzuka
inuzuka@nittech.ac.jp

¹ Nagoya Institute of Technology, Nagoya, Japan

² Chubu University, Kasugai, Japan

in most data sets, the number of abnormal observations is the minority despite being the focal point. As a solution, metadata relating to images has been jointly used with the real X-ray images for classification [14]. Alternatively, synthetic image generation of highly realistic images based on mapping the distribution of the existing real images has been applied in controlling class imbalance [15]. The use of ensemble techniques based on combining the synthesized images with real images or combining the metadata with real images has also been explored to provide better classification outputs [16].

The paper's contribution is to provide an ensemble model that improves image binary classification by reducing the class imbalance between the minority and majority classes in a data set. The ensemble model contains a combination of synthetic images, real images and metadata associated with the real images. The ensemble method applied in this research aims to reduce the high imbalance in pneumonia images classes and hence provide a better classifier for pneumonia detection from chest X-ray images (CXR images). To the best of our knowledge, an ensemble classifier technique that utilizes the three different approaches has not been conducted before.

The rest of the paper is organized as follows. In Sect. 2, we provide an overview of the related work concerning class imbalance, synthetic image generation, ensemble image classification using metadata and synthetic images. In Sect. 3, we present a methodology for image generation and our ensemble classifier model. In Sect. 4 we describe the pre-processing of the data sets. In Sects. 5 and 6, we present experimental setup, outcomes and results.

2 Related work

One of the main reasons for generating data sets is to counter the inadequate training data and protect the confidentiality of data, highly observable within the healthcare domain. The generation of synthetic data has been researched for a long period with the usage of different techniques [17]. In recent years, with the advent of Generative Adversarial Networks (GAN's), a move towards more realistic synthesized images is almost achievable. Various methodologies associated with GAN's and the synthesis of chest X-ray images have raised interest [12]. Particularly, the ability to generate images from unpaired labels is a big leap in the synthesis of images used for classification [15]. Zunair et al. [18] uses pneumonia images in the generation of scarce data on COVID-19 chest X-rays for classification. The study uses the synthesized images in order to leverage the class imbalance where there are few instances of positive COVID-19 CXR images.

The issue of class imbalance is also an eminent factor that undermines classification [19]. In most instances, the

inter-class variation has caused over-fitting due to the high effect on the model weighting by the majority class [20]. Transactions on medical imaging [21] indicates an improvement in the performance of classification models by implementing synthetic generated images to an imbalanced data set as an augmentation technique for the minority class. Qasim et al. [22] addresses the problem of class imbalance through the conditional synthesis of medical images that are applied for the classification of brain tumors.

Further, other techniques of addressing class imbalance have been depicted through ensemble models of convolution neural networks. A case in this respect has been ensemble methods consisting of metadata associated with images, metadata extracted from images, and the images [16]. The study implements classification tasks on skin lesions based on metadata associated with the patient and images of skin lesions taken from the patient. A comparison of the two methods implemented indicates better performance in classification made using an ensemble of images and metadata.

However, ensemble models have also been proven not to be the ideal models in some instances. Calderisi et al. [14] implements an ensemble of image-based metadata and a simple convolution neural network architecture on the classification of severe defects. The study utilizes the principle component analysis (PCA) and Q residuals on the metadata and structures a combination of the network for two data sets. The study notes that the results on feature selection and dimensionality reduction offer better results than other more sophisticated ensemble methods.

From these studies on the use of real images and metadata towards the improvement of classification, our research aims to extend the concepts. We propose a new addition to the techniques by adding synthetic image data to the minority class of the images used for classification with the aim of determining whether an increase in the number of the minority classes which implies a decrease in class imbalance, improves classification. Addition of the new approach provides three avenues for classification hence forming a basis for the ensemble classification method.

3 Methodology

We highlight the building blocks geared towards our ensemble method. the methodology encompasses a flow from synthetic image generation, binary classification of images, binary classification of metadata, and later an ensemble classifier based on real images, metadata associated with the images and synthetic images.

3.1 Synthetic image generation

The goal of a basic GAN, in theory, is to learn the mapping functions of a domain (X) distribution and reproduce the mappings on another domain (Y) [23]. This is achieved using two types of models: a generator and a discriminator which are based upon neural networks due to the Universal Approximation theory. The generator (G) enables the model to learn the joint distribution of the input variable and the target variable, while the Discriminator (D) learns a target variable given an input variable. The mapping function of a generator $G: X \rightarrow Y$ and the discriminator D_y which are expressed by an objective as:

$$\mathbb{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}} [\log(1 - D_Y(G(x)))] \tag{1}$$

The generator G is responsible for generating images $G(x)$ that imitate the distribution of images in a domain Y . The discriminator D_y determines the difference between the generated images $G(x)$ and the real images Y . In doing so the objective function tries to optimize the adversarial loss by improving the generator to produce more realistic images through reducing the mapping spaces of the two domains. In this paper, we employ the same principles of a basic GAN but using a cycle consistent GAN.

The cycle GAN introduces a cycle consistency loss that claims to guarantee that the mapping of functions between domains to cycle consistent [15]. The cycle GAN enforces a cycle consistency that is able to ensure that for each image y in domain Y , the resultant generated image should be able to cycle back to the original image and vice versa. That implies that the real domain and the generated domain distribution are almost similar. Such that: $y \rightarrow G(y) \rightarrow F(G(y))y$. The cycle GAN combines two objective functions of a GAN and introduces a parameter lambda that provides relative importance to the two GAN models.

$$\mathbb{L}_{GAN}(G, D_X, X, Y) = \mathbb{L}_{GAN}(G, D_Y, X, Y) + \mathbb{L}_{GAN}(F, D_X, Y, X) + \lambda \mathbb{L}_{Cyc}(G, F) \tag{2}$$

3.2 Classification

To attain an ensemble of the three data sets real images, synthetic images and metadata, we train a joint classifiers for a deep convolution neural network (DCNN) and a Multi-Layer Perceptron (MLP). The model consists of two main underlying networks. We train a convolution network on the images of the balanced data set. The balanced dataset contains real images of the minority class and the synthesized images. Secondly, we train a normal Multi-Layer Perceptron(MLP)

on the patients’ metadata associated with the images and merge the outputs of the networks to produce an ensemble.

The DCNN model is composed of architectures based on three benchmark image classification models, the RESNET50 [27], VGG16 [28], and DENSENet121 [29]. First, we train a classifier of the real images only and call it (*DCNN-real*). For the metadata associated with the real images we use a MLP that consists of a Neural network with three fully connected layers (*MLP-meta*). Finally, for the target ensemble method we train the DCNN model and the MLP model jointly to obtain an ensemble classifier.

The data set used for the ensemble is achieved by adding the generated images to the minority classes to balance the minority class(pneumonia cases) of the real data set, integrate it to the corresponding metadata, and create an ensemble DCNN (*DCNN-real-synth-meta*). The ensemble classification flow is highlighted in Fig. 1.

4 Data

4.1 Sample selection

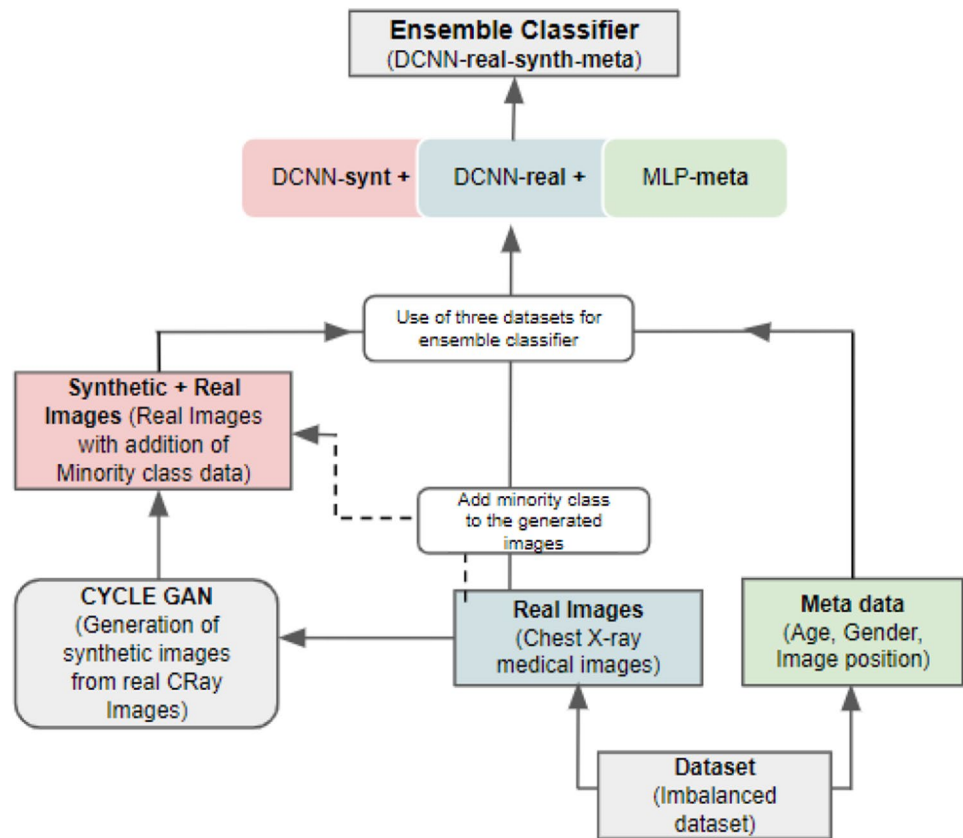
We obtained chest X-ray (CXR) images data set from the NIH website (<https://nihcc.app.box.com/v/ChestXray-NIHCC>). The data consists of 112,120 frontal view x-ray images with only 30,805 unique patients. Each of the patient undergoes a number of rounds for testing of the chest X-ray(CXR) images.

We only sample the pneumonia cases and images taken from the first round of the CXR image scanning. After sampling, we obtain a total of 212 pneumonia-positive cases and a random sample of 700 normal CXR images. We use this sample images in the GAN for synthetic image generation. Additionally, we obtain metadata associated with each of the sampled patients’ images. The metadata consists of patient age, gender, image orientation (including frontal/rear view images).

4.2 Data preprocessing

The selected images for the sample are obtained in the original size and transformed for the training set to 256x256 pixels. The training and test samples are split in the ratio of 7:3 on the sampled data set. The patient’s metadata only consists of the data observations matched to the sampled images. We add a new target feature on the metadata that corresponds with the image status ‘ones’ being pneumonia positive and ‘zeros’ being pneumonia negative. Further, we process the metadata with all the categorical variables being subjected to a one-hot encoding, and the numerical variables to a Min-Maxscaler to regularize the scale range. The images from

Fig. 1 Schematic illustration of the flow of image and metadata. Right indicated an MLP classified on the metadata (MLP-meta). Middle indicates a DCNN-real based on the real Images (DCNN-real). Left indicates a GAN generated images and applying them jointly with real images and metadata for an ensemble classifier (DCNN-real-synth-meta)



preprocessing are used in the generation of the synthetic image set as highlighted in Sect. 5.2.

5 Experimental setup

We train and test our ensemble on pre-trained DCNN architectures due to the nature of our research hypothesis. In this case, we select the pre-trained models of the Resnet50, Densenet121, and VGG16 to evaluate the ensemble's performance on each of the models. We also change the input layer of the models to a single channel for the images with a final binary output subjected to dropout during training. Both the real and synthetic images are trained based on these models.

5.1 Metric evaluation

We are focused on evaluating the hypothesis that the addition of the synthesized data on the minority class reduces class imbalance, hence having an impact on the performance of a classifier. To achieve this goal, the choice of evaluation metric for our model is important. A popular metric for evaluation in the measure for classification is the basic accuracy method. However, since we are aware that our data set displays a high level of class imbalance we cannot apply

the accuracy metric because it may lead to misleading results [24].

An ideal outcome for pneumonia classification is a scenario where all the positive cases are detected by the classifier. Since we are interested in the evaluation of all the positive cases which is the optimal case for our model, we use the recall metric to evaluate the performance of our model. Recall also called sensitivity indicates how well our model represented the minority cases which are our main interest in the study. The recall is the ratio of the true positive instances against the total positive instances and the False Negative's in the classifier defined as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (3)$$

5.2 Synthetic image generation

We apply cycle GAN for generation of synthetic images by using sampled images highlighted in Sect. 4.1. After synthesis of the real images, we generate a total of 350 for the normal images(majority class) and 350 pneumonia images(minority class). Later we add the 350 minority synthetic images to the real image data set to improve the class imbalance. Under ideal circumstances a point of convergence is achieved when the discriminator loss is 0.5

(ability of the discriminator to differentiate between real and synthetic images). To evaluate the realness of the images generated by the GAN, we do not rely on the discriminator loss of 0.6213 obtained from our GAN. Rather we use the difference exhibited by the classifier as a measure of impact for balancing the class (see Fig. 2).

5.3 Classification

Since the aim is to distinguish if balancing the minority class sample and employing an ensemble classifier based on different data sets results in an improvement in the classification. We set up different classifiers as an ablation to the target ensemble.

- (i) *Classifier of real images only* First we carry out a classification on the real images data set only across the three pre-trained classification benchmark models and observe the change in recall during the training.
- (ii) *Ensemble classifier on real images and metadata associated with the real images* Secondly, we evalu-

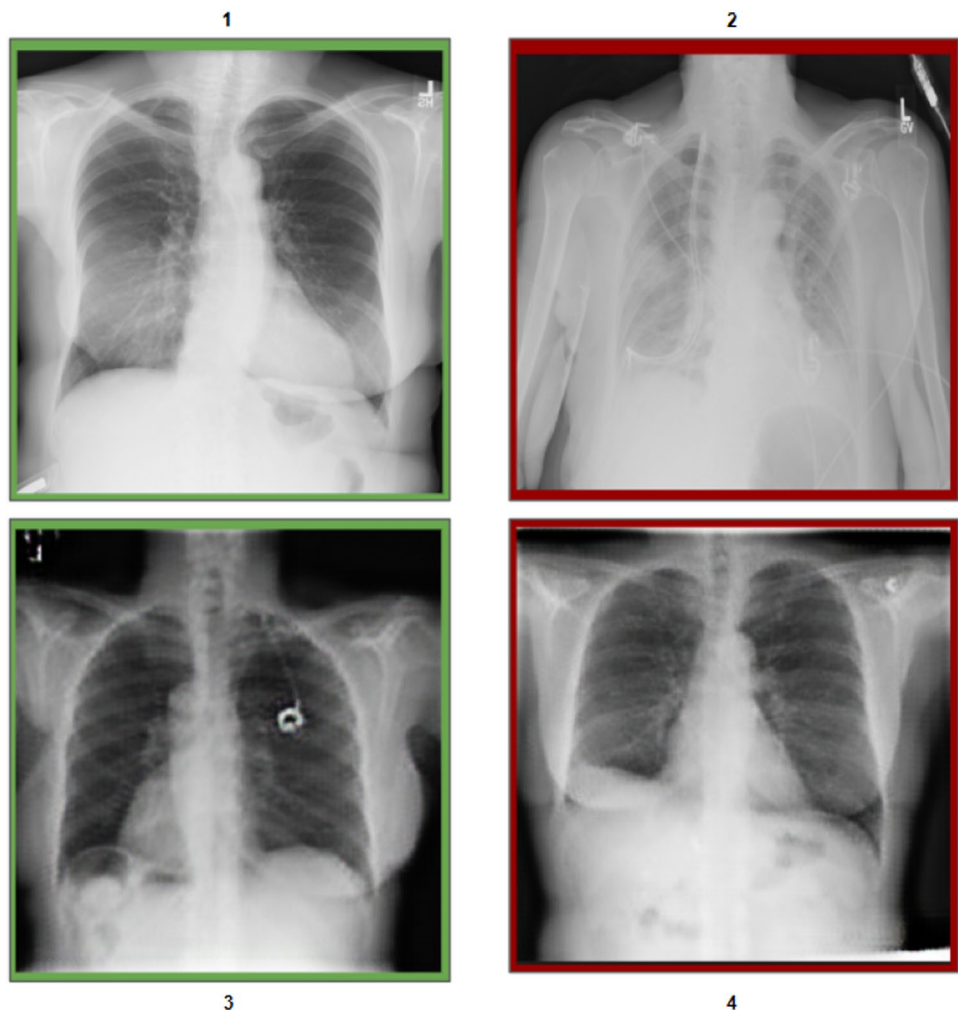
ate the performance of an ensemble classifier based on real image data set and the metadata. The metadata consists of patient age, gender and image orientation features.

- (iii) *Ensemble classifier on real images with metadata and synthetic balance on the minority class* Finally, we evaluate the proposed ensemble classifier which contains the data set of the real images, synthetic images added to the minority class and metadata associated with the real images. The three different pre-trained classifiers are used because we note that the result from a single model may experience randomness in some instances.

6 Results

- (i) *Ensemble classifier on real images* First, we demonstrate the performance of the real image data set on a classification model. We observe that the Recall acts as a great point of measure due to the steady increase

Fig. 2 Image 1 and 2 represent real normal and pneumonia positive images respectively. Image 3 and 4 represent synthetic normal and pneumonia positive images respectively



train_loss	valid_loss	accuracy	precision	recall	rocauc
0.561679	0.527434	0.805970	0.941176	0.571429	0.844017
0.470754	0.383602	0.855721	0.966102	0.678571	0.911783
0.422518	0.351939	0.855721	0.910448	0.726190	0.925824
0.393544	0.389078	0.815920	0.752688	0.833333	0.912291
0.361102	0.373931	0.825871	0.845070	0.714286	0.902829
0.321716	0.355270	0.835821	0.814815	0.785714	0.917074
0.289945	0.359746	0.830846	0.857143	0.714286	0.914733
0.259145	0.374910	0.810945	0.750000	0.821429	0.916158
0.241794	0.368823	0.850746	0.846154	0.785714	0.913207

Fig. 3 Ensemble classifier performance metrics on DENSENet121

train_loss	valid_loss	accuracy	precision	recall	rocauc
0.629704	0.617497	0.776119	0.854545	0.559524	0.781848
0.509213	0.465126	0.820895	0.944444	0.607143	0.884717
0.444378	0.408893	0.820895	0.842857	0.702381	0.897639
0.411609	0.378080	0.845771	0.844156	0.773810	0.906695
0.379731	0.379098	0.805970	0.747253	0.809524	0.898352
0.355034	0.374072	0.820895	0.792683	0.773810	0.908628
0.334861	0.365480	0.825871	0.775281	0.821429	0.915140
0.314928	0.368222	0.825871	0.781609	0.809524	0.915649
0.295105	0.351258	0.825871	0.795181	0.785714	0.921042
0.275160	0.360081	0.830846	0.784091	0.821429	0.915954

Fig. 4 Ensemble classifier performance metrics on RESNet50

- over a range of training steps. Loss on the other side indicates fluctuation, evidence of overfitting in some instance.
- (ii) *Ensemble classifier on real images and metadata associated with the real images* We also evaluate the performance of the joint classifier of the real images data set and the metadata associated with the real images. This model indicates improvement in the recall metric compared to real images only. We associate the improvement to the addition of the metadata providing a more enriched classifier that is able to learn more features. Even though, we note that the train and validation loss is still high, it still displays a lower representation to that of the model trained on image sets only. The ensemble of the two classifiers indicates a steadier increase in the recall as opposed to the images only which exhibit a larger plateau phase.
- (iii) *Ensemble classifier on real images with metadata and synthetic balance on the minority class* Finally, we evaluate an ensemble classifier on the real images, the metadata related to patient images, and the synthetic images of the minority class to balance the class. We observe an improved performance in the recall after training and testing the model on the data set. We maintain hyper-parameters on the ensemble model similar to the first and second experiments (i) and (ii).

We note that the results indicates slight improvements in sensitivity based on the three benchmark. In Table 1 we showcase the performance metrics on the three models. In Figs. 3, 4 and 5 we provide performance metric of the ensemble classifier based on the three benchmark models DENSENet121, RESNET50 and VGG16 respectively.

train_loss	valid_loss	accuracy	precision	recall	rocauc
0.606720	0.512667	0.810945	1.000000	0.547619	0.876374
0.494671	0.413479	0.830846	1.000000	0.595238	0.898555
0.441924	0.370346	0.850746	0.935484	0.690476	0.894892
0.403527	0.397359	0.796020	0.772152	0.726190	0.895910
0.377023	0.357103	0.825871	0.845070	0.714286	0.908730
0.344674	0.358684	0.840796	0.833333	0.773810	0.908934
0.312429	0.364681	0.830846	0.820513	0.761905	0.907916
0.287421	0.367752	0.840796	0.825000	0.785714	0.909035
0.268602	0.379566	0.845771	0.844156	0.773810	0.905678
0.255029	0.376581	0.840796	0.825000	0.785714	0.906695

Fig. 5 Ensemble classifier performance metrics on VGG16

Table 1 Sensitivity scores on test set over three models and three data sets

Model	Real	Real-Meta	Ensemble
RESNET50	0.5667	0.7333	0.8214
DENSENet121	0.400	0.6333	0.7857
VGG16	0.6000	0.8667	0.7857

The bold values indicate the most significant results for the experiment as explained in the results

Evaluating the models on the test set we observe a set of performance difference in the recall of the models. The RESNet50 and DENSENet121 architecture represent a higher recall value on the ensemble method. However, the VGG16 architecture does not favour a better sensitivity on the ensemble model.

By improving the minority class imbalance by half the size of the majority class we observe an improvement

in the sensitivity by 12% and 24% (The percentages are calculated based on the rate of sensitivity for (ii) against (iii) for the classifiers of RESNET50 and DENSENet121 respectively).

6.1 Discussion

While improvements can be observed in our ensemble model as opposed to the other two classifiers, we are not claiming state-of-the-art results. Rather, we showcase the contribution of the synthetic image on imbalanced data towards improving generalization in a classification task.

Additionally, though we extract results from GAN through images that indicate an effect in the classification, we experience two drawbacks. First, the nature of GAN generation is generally highly unstable and therefore there is need for evaluation of the synthetic images through GAN evaluation methods such as Inception score and the Frechet Inception Distance. Secondly, further improvement on the current sensitivity results is achievable based on the work conducted for improvement of the data set through balancing the minority class.

7 Conclusion

We propose an ensemble classifier using real chest x-ray images, the patient metadata associated with images, and synthetic images. We generate the synthetic images using a cycle consistent GAN. The ensemble classifier shows an improvement in the image classification based on the sensitivity metric. By improving the minority class imbalance by half the size of the majority class we observe an increase in the sensitivity by 12% and 24% for the classifiers of RESNET50 and DENSENet121 respectively. We associate the difference to improvement of class imbalance and using the synthetic images jointly with metadata related to real images in an ensemble classifier.

References

- Pang G, Shen C, Cao L, Van Den Hengel A (2021) Deep learning for anomaly detection: a review. *ACM Comput Surv (CSUR)* 54(2):1–38
- Pneumonia. <https://www.who.int/news-room/fact-sheets/detail/pneumonia>. Accessed 5 Dec 2021
- Goldenberg JM, Cárdenas-Rodríguez J, Pagel MD (2018) Preliminary results that assess metformin treatment in a preclinical model of pancreatic cancer using simultaneous [18F]FDG PET and acidoCEST MRI. *Mol Imaging Biol* 20(4):575–583
- Noor NM, Rijal OM, Yunus A, Abu-Bakar SAR (2010) A discrimination method for the detection of pneumonia using chest radiograph. *Comput Med Imaging Graph* 34(2):160–166
- Oliveira LL, Silva SA, Ribeiro LH, de Oliveira RM, Coelho CJ, Andrade ALS (2008) Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *Int J Med Inform* 77(8):555–564
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP, Ng AY (2017) Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning
- Parveen NRS, Sathik MM (2011) Detection of pneumonia in chest x-ray images. *J X-Ray Sci Technol* 19(4):423–428
- Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK (2018) Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med* 15(11):e1002683
- Doshi K (2019) Synthetic image augmentation for improved classification using generative adversarial networks. [arXiv:1907.13576](https://arxiv.org/abs/1907.13576)
- DeVries T, Taylor GW (2017) Improved regularization of convolutional neural networks with cutout. [arXiv:1708.04552](https://arxiv.org/abs/1708.04552)
- May P (2019) Improved image augmentation for convolutional neural networks by copyout and CopyPairing. [arXiv:1909.00390](https://arxiv.org/abs/1909.00390)
- Huang CC, Wu YL, Tang CY (2019) Human face sentiment classification using synthetic sentiment images with deep convolutional neural networks. In: 2019 International Conference on Machine Learning and Cybernetics (ICMLC). IEEE, pp 1–5
- Ha Q, Liu B, Liu F (2020) Identifying melanoma images using EfficientNet ensemble: winning solution to the SIIM-ISIC melanoma classification challenge. [arXiv:2010.05351](https://arxiv.org/abs/2010.05351)
- Calderisi M, Galatolo G, Ceppia I, Motta T, Vergentini F (2019) Improve image classification tasks using simple convolutional architectures with processed metadata injection. In: 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). IEEE, pp 223–230
- Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2223–2232
- Ningrum DNA, Yuan S-P, Kung W-M, Wu C-C, Tzeng I-S, Huang C-Y, Yu-Chuan Li J, Wang Y-C (2021) Deep learning classifier with patient's metadata of dermoscopic images in malignant melanoma detection. *J Multidiscip Healthc* 14:877–885
- Phua C, Lee V, Smith K, Gayler R (2010) A comprehensive survey of data mining-based fraud detection research. [arXiv:1009.6119](https://arxiv.org/abs/1009.6119)
- Zunair H, Hamza AB (2021) Synthesis of COVID-19 chest x-rays using unpaired image-to-image translation. *Soc Netw Anal Min* 11(1):23
- Lequan Yu, Chen H, Dou Q, Qin J, Heng P-A (2017) Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans Med Imaging* 36(4):994–1004
- Shie CK, Chuang CH, Chou CN, Wu MH, Chang EY (2015) Transfer representation learning for medical image analysis. In: 2015 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp 711–714
- Shin H-C et al (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 35(5):1285–1298. <https://doi.org/10.1109/TMI.2016.2528162>
- Qasim AB, Ezhov I, Shit S, Schoppe O, Paetzold JC, Sekuboyina A et al (2021) Red-GAN: attacking class imbalance via conditioned generation. Yet another perspective on medical image synthesis for skin lesion dermoscopy and brain tumor MRI. *Proc Mach Learn Res* 1:14
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S et al (2014) Generative adversarial nets. *Adv Neural Inf Process Syst* 27
- Menon AK, Williamson RC (2018) The cost of fairness in binary classification. In: Conference on Fairness, accountability and transparency, pp 107–118. PMLR

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article

is solely governed by the terms of such publishing agreement and applicable law.