



Data Augmentation of Minority Class with Transfer Learning for Classification of Imbalanced Breast Cancer Dataset Using Inception-V3

Manisha Saini^(✉) and Seba Susan

Delhi Technological University, Delhi, India
manisha.saini44@gmail.com

Abstract. In this paper, deep learning based experiments are conducted to investigate the effect of data augmentation on the minority class for the imbalanced breast cancer histopathology dataset (BREAKHIS). Two different pre-trained networks are fine-tuned with the minority-augmented dataset. The pre-trained networks were already trained on the well-known ImageNet dataset comprising of millions of high resolution images belonging to multiple object categories. The model so trained is further subjected to transfer learning, to correctly classify cancerous pattern from non-cancerous conditions, in a supervised manner. Our experiments were carried out in two phases. Phase-I investigates the effect of data augmentation applied on minority class for the Inception-v3 and ResNet-50 pre-trained networks. Results of phase-I are further enhanced in phase-II by the transfer learning approach in which features extracted from all layers of Inception-v3 are learnt by the SVM and weighted SVM classifiers. From experimental results, it was found that the pre-trained Inception-v3 model with data augmentation on minority class outperforms other network types. Results also indicate that Inception-v3 with data augmentation of minority class and transfer learning with weighted SVM gives the highest classification accuracies.

Keywords: Data augmentation · Imbalanced dataset · ResNet-50 · Inception-v3 · SVM

1 Introduction

Breast Cancer is a commonly found cancer related disease in women which leads to a high increase in the death rate every year [1]. So it is essential to detect and diagnose cancer at an early stage to begin with the right treatment. Even so, it was found that breast cancer histopathology dataset (BREAKHIS) is an imbalanced image dataset. Hence, this imbalanced breast cancer dataset is challenging to deal with because of the disproportionate number of samples in each class respectively [2]. It is important to recognize cancer (Malignant) class from non-cancer (Benign) class correctly so that an appropriate diagnosis of cancer is done at the right stage. It is often observed that most real-world problems involve imbalanced datasets. The imbalanced nature of such a

dataset yields results that are biased towards the majority class and the minority class is left undetected. An investigation is conducted in our work using the breast cancer dataset, to effectively discriminate minority class samples from the majority class through deep learning based experiments. A data augmentation based approach using pre-trained networks and transfer learning is used to correctly classify cancerous pattern from non-cancerous conditions. The transfer learning approach is used in the experiment to transfer the knowledge from one domain to another domain for performing a two-class classification task to segregate Benign from Malignant class. Transfer learning is a way of transferring knowledge from one specific domain or task to another task [3]. There are multiple advantages of using transfer learning approach (1) Firstly, model is already trained on a large scale dataset, and we are using the already existing model along with predefined weights for our own classification task, which saves a lot of computational processing time. (2) Secondly, we can transfer the knowledge from large scale dataset and perform classification well even with the small dataset. Pre-trained networks are trained on a large scale dataset such as ImageNet, which is composed of millions of high-resolution images belonging to multiple categories or classes (approximately 1.4 million images and 1000 classes).

Data augmentation [4] is the process of synthetically increasing the number of samples in the dataset, which helps to resolve the overfitting problem that arises due to inadequate number of samples in the dataset. There are several transform operations which can be applied for data augmentation such as scaling, rotation, horizontal and vertical flip, etc. This transformation help in increasing the number of samples of minority class as well as majority class so that we are able to correctly detect the minority class while performing the classification task. This paper is ordered into following subsequent sections. Section 2, discusses related work. In Sect. 3, the methodology for extracting features using pre-trained networks along with the data augmentation approach is described. Sections 4 and 5 discuss about the dataset and the experimental work conducted respectively. Finally, the conclusion is inferred in Sect. 5.

2 Literature Survey

BREAKHIS involves the imbalanced dataset having an unequal class distribution with the number of Benign samples significantly lesser than Malignant class. A large flux is observed in these images. Several studies have been conducted in order either to resolve the issue or study the effect of the imbalance. Gardner & Nicholas (2017), experiment on satellite images by implementing a Convolution Neural Network (CNN) model to perform multi-label classification through a comparison between VGG-16, Inception-v3, ResNet-50 and ResNet-50 with data augmentation. Best performance is achieved with ResNet-50 model after combination with data augmentation and ensemble technique by using various evaluation parameters such as F1-score, precision and recall [5]. Wang et al. (2017) proposed a method called neural augmentation which allows a neural net to learn augmentations that improve the classifier performance and comparison was carried out with traditional approaches used for data augmentation in the image classification task. Results state that a combination of

traditional augmentation followed by neural augmentation further improves the classification strength [6].

In data science, sampling is the most popular remedy for the class imbalance problem. Chawla (2009), described in detail about changes that can be performed at data level or algorithmic level to overcome the class imbalance issue, such as sampling techniques at data level i.e., oversampling, undersampling, SMOTE and even variants of SMOTE such as SMOTEBOOST etc [7]. Hybrid strategies that combine oversampling of the minority class with the undersampling of the majority class have been proposed [21, 22]. In the field of computer vision, data augmentation has proved to mitigate the imbalance problem to a certain extent, and hence it is the approach used in our deep learning experiment to counter the class imbalance.

Chen, et al. (2012) conducted an experiment by applying data augmentation on various pre-trained networks such as VGG16, ResNet-50, and DenseNet for satellite images. From the experimental results, it was found that the best performance is achieved in case of the ensemble model which is formed after a combination of the results obtained from three DenseNet121 models with VGG-16 model [8].

3 Proposed Methodology and Techniques Used

Following the cue from the existing work, our work induces data augmentation [4], however on the minority class only, similar to the oversampling step in [21]. The minority augmented dataset, which is in essence balanced, is then applied for transfer learning through pre-trained networks and eventually classified using Weighted Support Vector Machine (SVM).

3.1 Data Augmentation of the Minority Class

The following data augmentation operations were applied on the minority class for the complete dataset: shear with range of 10, upper and lower zoom with the range of 20 percent and horizontal flip, along with resizing and pre-processing operations. Instances from the augmented minority set are shown in Fig. 1.

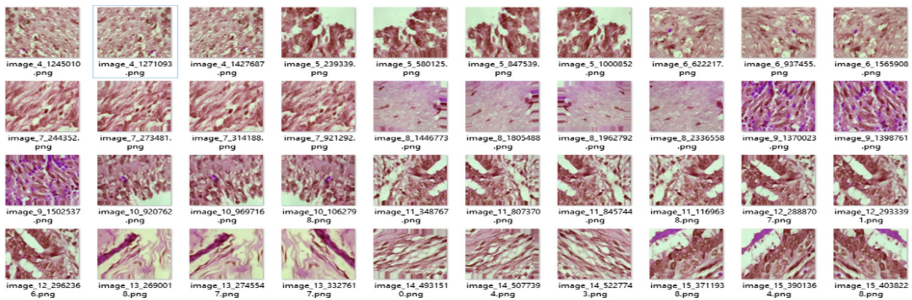


Fig. 1. Illustration of data augmentation applied on the minority class (Benign)

3.2 Phase I: Classification Using Pre-trained Networks

The pre-trained model was used for performing the classification experiment [19]. The pre-trained networks used for our experiment are described below.

- (i) ResNet-50:- Architecture of ResNet-50 [5] consists of various residual blocks. Each residual block consists of a set of repeated layers. ResNet-50 architecture basically consists of 50 layers comprising of repeated convolution and pooling layers along with fully connected layers.
- (ii) Inception-v3:- GoogleNet introduced the first Inception model. With time span, various other versions of Inception were introduced such as Inception-v2, Inception-v3, and Inception-v4. For our experiment, we have used Inception-v3, which has factorization as the main idea along with batch Normalization [8]. Inception-v3 architecture basically consist of 42-layer architecture, having fewer parameters, and more computational efficiency in comparison to previous Inception architectures.

Global Average Pooling 2D is added along with the dense layers [11], and Sigmoid and RELU activation functions are used.

3.3 Phase II: Deep Feature Extraction and Transfer Learning

Phase II, comprises of deep feature extraction from the pre-trained model fine-tuned with the minority augmented dataset. The deep features are extracted from all the layers of the Inception-v3 pre-trained model taken in the right order. The classifiers used for our experiment involving transfer learning based on the deep feature extracted are explained below:-

SVM. SVM classifier can be used for linearly separable and non-separable problems based upon the type of kernel used. The kernel is the parameters set to change/transform the dimension of input feature space from low to high. The kernel can be selected as 'Linear', 'Polynomial', 'rbf' etc., based upon the task performed. SVM [9, 10] is commonly used as a classifier to solve numerous tasks by selecting optimal decision boundary or hyperplane. Selected optimal decision boundary helps to segregate the samples from the two classes. For the experimental task, we have used 'rbf' kernel.

Weighted SVM. Weighted SVM [11] is also used for the evaluation in Phase II. The class weight is assigned the balanced kernel value in order to deal with the imbalanced dataset. The kernel used while performing the classification task is 'linear' along with other values that were set such as $C = 1.0$ (C is penalty parameter). When the value of C is set to 1, it indicates that the classifier is easily able to tolerate the wrong classification data items/points. Other parameters which was considered in training the classifier was a random state which was set equal to zero and probability is set as true while performing the experiment.

4 Experimental Results and Discussions

For the purpose of our deep learning experiments, the popular python framework Keras v2.1.6 has been used. It uses a Tensorflow backend to perform all its internal computations. The model will be trained across 3 epochs with 29 steps in each epoch. While training the pre-trained network, sparse categorical cross-entropy is used as the loss function for each fold along with Adam optimizer (Adaptive Moment Estimation), batch size set as 32 in each fold, with RELU activation function [12] in dense layer, is used for the experiment.

4.1 Dataset

Images present in the BREAKHIS dataset belong to of four different magnification factors i.e. 40X, 100X, 200X and 400X [18, 20]. Experimental evaluation was performed using 400X magnification factor images. Dataset comprises of 5,88 Benign and 1,232 Malignant images as shown in Table 1. Random sampling is applied in order to segregate testing images from training images, such that 100 samples each from Malignant and Benign class will be used for testing. Whereas the remaining number of the sample is kept for training purpose. Each input image was resized into 224×224 size before performing the experiment. For the experiment, five-fold cross-validation technique has been used in Phase I.

Table 1. Imbalanced image dataset of a patient having cancer and those not suffering from cancer obtained from 400 X magnification factor in training and testing dataset.

Class	Dataset		Train dataset	Test dataset
	Magnification factor	No of images		
Benign	400 X	588	488	100
Malignant	400 X	1232	1132	100

4.2 Performance Evaluation

Several studies show that accuracy is not always a reliable parameter to be used to determine the model performance while dealing with the imbalanced dataset. Accuracy illustrates how proficiently classifier is able to classify the samples in the different classes. Precision denotes the fraction of relevant samples which are correctly classified from the total number of samples. Recall denotes the fraction of samples which are correctly classified from the correct samples. F-Measure/F1-score is obtained from the combination of both recall and precision as shown in Eq. 1. Whereas the ROC curve is the graph plotted between the true positive rate and false negative rate.

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (1)$$

Classification report has been generated using the scikit-learn package in python for analyzing the performance of the model with precision, recall, F1-score [16] and ROC curve (Receiver Operating curve).

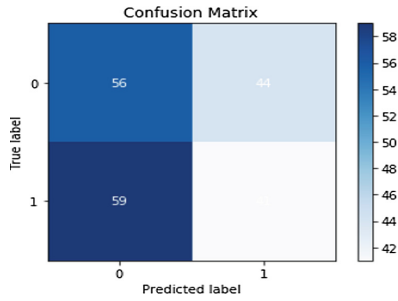
4.3 Results and Discussions

4.3.1 Phase I Experiments

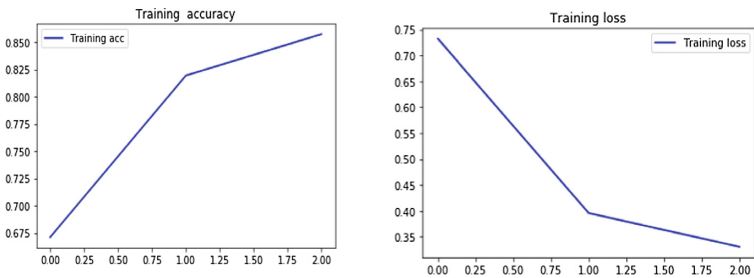
From the experimental analysis, it was found that there is a huge difference between test and train accuracy. As average train accuracy in case of Inception-v3 is approximately between 70 to 80% range with and without data augmentation. Whereas, average test accuracy has improved from 47 to 50% as shown in Table 3. Experiments performed using Inception-v3 pre-trained networks with and without data augmentation are shown in Tables 2 and 3. From the experimental evaluation, it was found that Inception-v3 is unable to detect minority class. Whereas experiments conducted by applying data augmentation technique on minority class separately are effectively able to detect the minority class, as illustrated in Table 2 and Fig. 2. According to the literature, it was found that while dealing with the imbalanced dataset, results are biased towards the majority class. Figure 3 depicts the test samples tested on Inception-v3 pre-trained network after applying data augmentation on minority class using one of the best fold out of the average (Avg) five-fold. Whereas, data augmentation applied separately on the minority is able to detect minority class efficiently in comparison to Inception-v3 pre-trained network after data augmentation is applied on both classes. Figure 4 represents sample test images without applying data augmentation. Similar experiment was conducted with ResNet-50 architecture to measure the performance of data augmentation technique as illustrated in Tables 4 and 5. From experiment results, it was found that Inception-v3 outperforms ResNet-50 in all the cases.

Table 2. Comparison of Performance evaluation on the basis of Precision, Recall, and F1-score using cancer dataset in case of Inception-v3 pre-trained network.

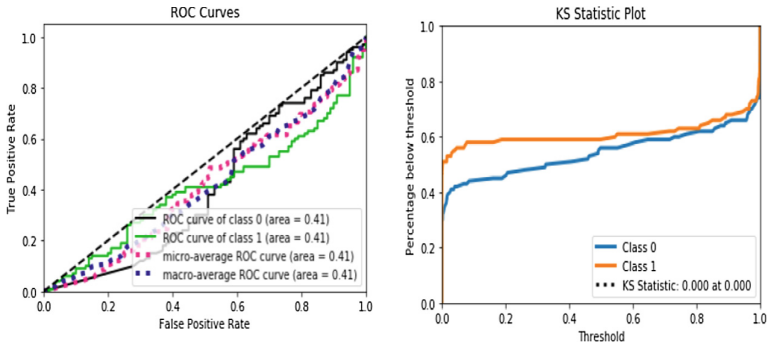
Classes	Inception-v3 [14]		Inception-v3 (with data augmentation on both class) [15]		Proposed Inception-v3 (with data augmentation on minority class)	
	Benign	Malignant	Benign	Malignant	Benign	Malignant
Avg precision (five-fold)	0.076	0.49	0.05	0.49	0.46	0.44
Avg recall (five-fold)	0.006	0.48	0.002	0.99	0.56	0.35
Avg F1-score (five-fold)	0.012	0.66	0.004	0.66	0.50	0.39
Support	100	100	100	100	100	100



(i) Confusion Matrix in case of Inception-v3



(ii) Accuracy and loss obtained while training the Inception-v3 pre-trained network



(iii) ROC Curve (iv) KS statistic Plot in case of Inception-v3 pre-trained Network

Fig. 2. Various visualization Curves demonstrates from (i) to (iv) in case of Inception-v3 pre-trained networks after applying data augmentation on Minority class with 4 fold validation

Table 3. Comparison of performance evaluation of train and test accuracy using cancer dataset in case of Inception-v3 pre-trained network.

Pre-trained networks	Fold	Train accuracy (%)	Train loss	Test accuracy (%)	Avg test accuracy (five-fold)	Avg train accuracy (five-fold)
Inception-v3 [14]	1	0.80	0.46	0.50	0.49	0.78
	2	0.77	0.46	0.50		
	3	0.78	0.47	0.49		
	4	0.88	0.44	0.49		
	5	0.78	0.47	0.48		
Inception-v3 with data augmentation on both the classes [15]	1	0.74	0.51	0.50	0.49	0.77
	2	0.80	0.44	0.50		
	3	0.76	0.49	0.50		
	4	0.76	0.48	0.50		
	5	0.80	0.42	0.50		
Inception-v3 with data augmentation on minority class	1	0.83	0.38	0.46	0.47	0.83
	2	0.84	0.34	0.48		
	3	0.82	0.4	0.47		
	4	0.85	0.33	0.48		
	5	0.83	0.36	0.46		

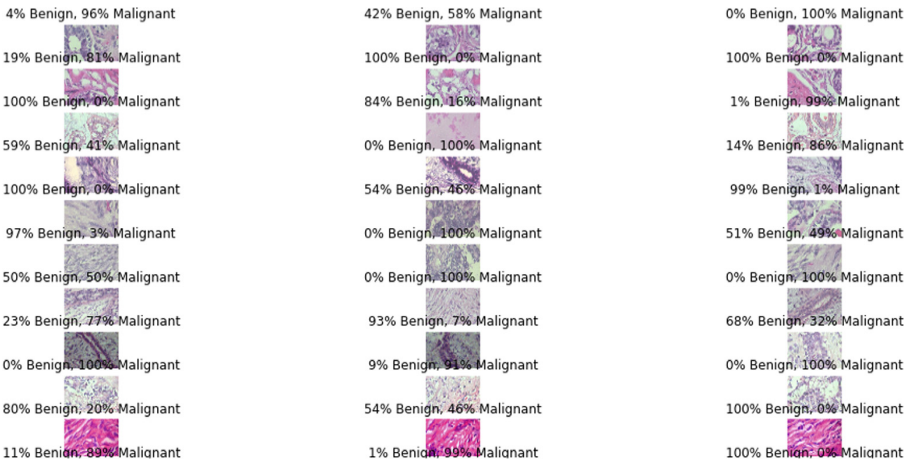


Fig. 3. Samples tested on Inception-v3 pre-trained network after applying data augmentation on minority class

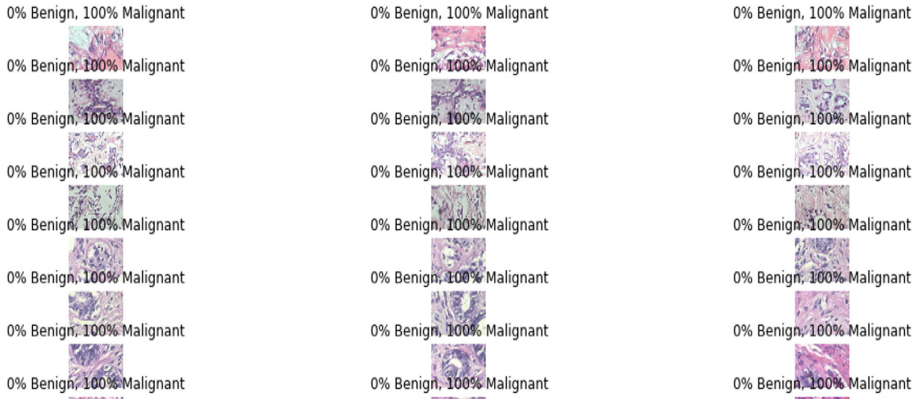


Fig. 4. Samples tested on Inception-v3 pre-trained network without applying data augmentation

Table 4. Comparison of performance evaluation on the basis of precision, recall, and F1-score using cancer dataset in case of ResNet-50 pre-trained network.

Classes	ResNet-50 [5]		ResNet-50 (with data augmentation on both class) [16]		ResNet-50 (with data augmentation on minority class)	
	Benign	Malignant	Benign	Malignant	Benign	Malignant
Avg precision (five-fold)	0.20	1.00	0.20	0.50	0.00	0.50
Avg recall (five-fold)	0.002	1.00	0.002	1.00	0.00	1.00
Avg F1-score (five-fold)	0.004	0.67	0.008	0.67	0.00	0.67
Support	100	100	100	100	100	100

Table 5. Comparison of performance evaluation of train and test accuracy using cancer dataset in case of ResNet-50 pre-trained network.

Pre-trained networks	Fold	Train accuracy (%)	Train loss	Test accuracy (%)	Avg test accuracy (%) (five-fold)	Avg train accuracy (%) (five-fold)
ResNet-50 [5]	1	0.93	0.16	0.50	0.50	0.92
	2	0.92	0.19	0.50		
	3	0.93	0.18	0.50		
	4	0.91	0.21	0.50		
	5	0.92	0.17	0.50		
ResNet-50 with data augmentation on both the classes [16]	1	0.79	0.71	0.50	0.50	0.82
	2	0.86	0.31	0.50		
	3	0.89	0.26	0.50		
	4	0.69	0.49	0.50		
	5	0.88	0.27	0.50		

(continued)

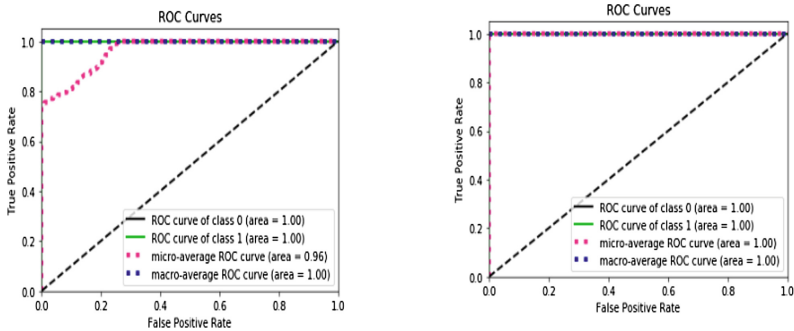


Fig. 5. (i) ROC curve in case of Inception-v3 with weighted SVM (ii) ROC Curve in case of Inception-v3 with weighted SVM and data augmentation applied on the minority class

5 Conclusion

A solution for countering class imbalance in deep learning is proposed in this work specifically for BREAKHIS breast cancer dataset. It is proved that we cannot only rely on accuracy for the identification of imbalanced dataset. Various other parameters need to be considered such as F1-score, Precision, Recall, and ROC. We have conducted experiments in two phases. Phase-I investigates the effect of data augmentation technique when applied on minority class only for the pre-trained networks Inception-v3 and ResNet-50. Results obtained are found better with the application of data augmentation on the minority class in Inception-v3 pre-trained model. In Phase-II, features were extracted from all layers of Inception-v3 and learned by the SVM and weighted SVM classifiers. Results show that Inception-v3 with data augmentation on minority class works best with transfer learning using weighted SVM as compared to other networks. However, it was also observed that despite having high test accuracy, if the model is unable to give correct predictions and the results are biased towards the majority class then the minority class is left undetected and the model will evaluate every sample image from test data as non-cancerous even though the person is having cancer. So, it is important to consider model performance using all the described parameters instead of using only accuracy as the evaluation criteria. Future work involves using ensemble deep feature approach to extract a new set of features to effectively distinguish the minority class from the majority class using pre-trained models.

References

1. Parkin, D.M., Bray, F., Ferlay, J., Pisani, P.: Global cancer statistics, 2002. *CA Cancer J. Clin.* **55**(2), 74–108 (2005)
2. Estabrooks, A., Jo, T., Japkowicz, N.: A multiple resampling method for learning from imbalanced data sets. *Comput. Intell.* **20**(1), 18–36 (2004)

3. Ferlaino, M., et al: Towards deep cellular phenotyping in placental histology. arXiv preprint. [arXiv:1804.03270](https://arxiv.org/abs/1804.03270) (2018)
4. Kumar, A., Kim, J., Lyndon, D., Fulham, M., Feng, D.: An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J. Biomed. Health Inf.* **21**(1), 31–40 (2017)
5. Gardner, D., Nichols, D.: Multi-label Classification of Satellite Images with Deep Learning, p. 2017 (2017)
6. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. *Neurocomputing* **234**, 11–26 (2017)
7. Chawla, N.V.: Data mining for imbalanced datasets: an overview. In: *Data Mining and Knowledge Discovery Handbook*, pp. 875–886. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-09823-4_45
8. Chen, Y., Dong, F., Ruan, C.: Understanding the Amazon from Space, pp. 1–9 (2012)
9. Vapnik, V., Isabel, G., Hastie, T.: Support vector machines. *Mach. Learn.* **20**(3), 273–297 (1995)
10. Camps-Valls, G., Gomez-Chova, L., Muñoz-Mari, J., Vila-Francis, J., Calpe-Maravilla, J.: Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **3**(1), 93–97 (2006)
11. https://chrisalbon.com/machine_learning/support_vector_machines/imbalanced_classes_in_svm/
12. Kim, M., Zuallaert, J., De Neve, W.: Towards novel methods for effective transfer learning and unsupervised deep learning for medical image analysis. In: *Doctoral Consortium (DCBIOSTEC 2017)*, pp. 32–39 (2017)
13. Niu, X.-X., Suen, C.Y.: A novel hybrid CNN–SVM classifier for recognizing handwritten digits. *Pattern Recogn.* **45**(4), 1318–1325 (2012)
14. Chang, C., Lin, C.: LIBSVM : A Library for Support Vector Machines, pp. 1–39 (2013)
15. Nguyen, G.H., Bouzerdoum, A., Phung, S.L.: Learning pattern classification tasks with imbalanced data sets. In: *Pattern Recognition*. IntechOpen (2009)
16. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.: Using transfer learning for image-based cassava disease detection. arXiv preprint. [arXiv:1707.03717](https://arxiv.org/abs/1707.03717) (2017)
17. Barai, M., Heikkinen, A.: Impact of data augmentations when training the inception model for image classification (2017)
18. Xie, J., Liu, R., Joseph Luttrell, I.V., Zhang, C.: Deep learning based analysis of histopathological images of breast cancer. *Front. Genet.* **10**, 80 (2019)
19. Kieffer, B., Babaie, M., Kalra, S., Tizhoosh, H.R.: Convolutional neural networks for histopathology image classification: training vs. using pre-trained networks. In: *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6. IEEE (2017)
20. <https://web.inf.ufpr.br/vri/databases/>. Accessed 10 May 2019
21. Susan, S., Kumar, A.: Hybrid of intelligent minority oversampling and PSO-based intelligent majority undersampling for learning from imbalanced datasets. In: Abraham, A., Cherukuri, A.K., Melin, P., Gandhi, N. (eds.) *ISDA 2018*. AISC, vol. 941, pp. 760–769. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-16660-1_74
22. Susan, S., Kumar, A.: SSOMaj-SMOTE-SSOMin: three-step intelligent pruning of majority and minority samples for learning from imbalanced datasets. *Appl. Soft Comput.* **78**, 141–149 (2019)