Effects of Class Imbalance Problem in Convolutional Neural Network Based Image Classification

Arghasree Banerjee, Kushankur Ghosh, Arya Sarkar, Mayukh Bhattacharjee, and Sankhadeep Chatterjee

Abstract Convolutional Neural Networks (CNNs) have revolutionized the image classification tasks to a greater extent. Various research works reported the high accuracy of CNNs in image classification problems related to science and engineering. It has been found that many image classification problems carry an additional challenge of imbalanced classes in the dataset. In absence of a suitable analysis of the effect of imbalanced image data on the performance of CNNs, the current study thoroughly analyzed the same. CNN models are evaluated in terms of Sensitivity, Specificity, G-Mean and Bookmaker Informedness with varying degrees of imbalance ratio. The experimental results have revealed that the performance of CNN is severely affected by increased imbalance ratio.

Keywords CNN · Imbalance · Image classification

1 Introduction

Almost every modern individual utilizes the facilities of the current technologies for the creation and sharing of images of any kind. This sharing of images has eased the mode of visual communication over the years. However, the quality of this communication mostly depends on the processing [1] than the quality of image. An image can be of any type starting from biomedical images [2, 3], document images [4], satellite images [1] and PolSAR images [5]. An intelligent system is very much necessary to identify the category of an image from a thick chunk of wide variety of images. The application of machine learning in real life issues has been studied and experimented over a wide range of time [6–11]. With the expansion of the field, the application of Neural Networks (NN) is gradually suppressing the implementations of other machine learning algorithms [12]. A NN based image classification [13] model can

A. Banerjee · K. Ghosh · A. Sarkar · M. Bhattacharjee · S. Chatterjee (✉)
Department of Computer Science & Engineering, University of Engineering & Management,
Kolkata, India
e-mail: chatterjeesankhadeep.cu@gmail.com

generate better results compared to any other traditional image classification technique due to the robustness of the prior model. A popular way of classifying an image is by using Convolutional Neural Networks (CNN) [14]. Over the years CNN based models have gained a wide popularity and has been implemented to resolve various real-life issues starting from face detection [15] to language matching [16]. However, in the recent years CNN has become a very concrete tool for image classification and segmentation. The classification of images incorporates the task of putting any selected image of a kind to its relative category based on intelligent computations. Various studies over the years successfully developed concrete models delivering effective classifications on chunks of images both offline and online. However, in the recent years the techniques of image classifications using CNN are documented to suffer from the imbalanced classification problem. The Class-Imbalanced problem was coined in the early 1990s and have been studied for more than two decades. A imbalanced dataset is mostly a common situation and can be very much prominent in any day to day practical scenarios. Apart from image classification, the problem has also become very popular in sentiment analysis [10], sarcasm detection [17], botnet detection [18] and twitter-spam detection [19]. In this paper a detailed documentation of class-imbalance problem in CNN based image classification is presented. The classification challenges and the future research scope are also discussed.

2   Image Classification Using CNN

A Convolutional Neural Network (Conv-Net/CNN) is a Deep Learning Based algorithm which is basically applied to analyse image data. A CNN consists of an input layer, an output layer and multiple hidden layers. Input is a visual imagery tensor with shape (number of images) * (image width) * (image height) * (image depth). After passing through the convolutions, the image gets absorbed to a feature map. In [20], the authors make an account for the early applications of CNN in speech recognition, text recognition, handwriting recognition and later in natural image recognition. In [21] the authors documented the behaviour of CNN on a set of 1.2 million high resolution images consisting of 1000 different variety of classes in ILSVRC-2010 competition. The authors found that CNN gives us a remarkable performance but can undergo some complications if a layer is removed. In the ILSVRC-2012 competition, the authors averaged the predictions of two CNNs that were pre-trained on the entire Fall 2011 release with the five CNNs which gave an error rate of 15.3%. The application of CNNs to multi-categorical classifications applied in the earlier paper, attracted a lot of attention in the computer vision research community. In [22] a pre-trained CNN model generated at the full-connected layers extracts features better than conventional hand-crafted features on all datasets in image retrieval application. In a comparative study done by Yi Hou et al. in [23], we get to know how CNN-based image descriptors are better than the state-of-the-art hand-crafted image descriptors in the visual loop closure detection application. In [24], the authors introduced 3D CNN models for action recognition. From adjacent input frames the authors generate

multiple channels of information. Convolution and subsampling is performed separately in each channel after that. The CNN based image classification is thus now used in many medical examinations as it significantly improved the best performance for many image databases [25]. CNN based image classification is still an interesting topic among researchers. In [26] a recently published paper presents a method to optimize input for CNN-based HSI classification. Liu et al. [27] proposed a detailed study by investigating the middle layers of the CNN and conducted experiments on different image datasets. Their approach fused the latent features extracted in the middle layers of the CNN which effectively improves the classifier's performance. Another remarkable application of CNN was demonstrated by Peng et al. [28] to detect fruit flies by using the FB-CNN technique. In [29], the authors utilized the texture filters to classify between histopathological images.

3 Class-Imbalance Problem in Image Classification

The imbalanced classification problem can occur in both binary or multi-class dataset. The problem occurs when the count of a particular class in a dataset becomes more dominant compared to the other classes. At this scenario, the classifying algorithm will have a tendency to generate decisions in favour of the class with the majority count. This mostly results in a wide deflection in performance of a machine learning model. The most common approaches to mitigate the issue are oversampling of minority instances or the under-sampling of majority instances. However, the most popular way to handle this problem is by making the minority class denser as under-sampling can result in a loss of valuable information from the dataset. Some of the popular techniques developed for oversampling over the years are SMOTE [30], Borderline-SMOTE [31], ADASYN [32] and Safe-Level-SMOTE [33] and. However, in [34, 35] the authors described the consequences of oversampling and also presented a remarkable majority focussed approach to overcome that. The earliest reference of the effects of class imbalance in image classification is found in [36]. However, in the recent years it has become a significant problem in the domain of image classification [37–41]. The problem mostly arise when the model is made to classify images in any practical domain. Classification with CNNs also undergo the same problem, like any other techniques of image classification. In [42] Lee et al. encountered with the class imbalance problem in Plankton image classifications. The authors proposed a novel approach based on CNN and avoided the problem by training the CNN with normalized data. The principle idea behind their approach depends on the value of threshold μ which is selected to construct the normalized data by sampling the original data. The normalized data is constructed in a way such that no class is larger than μ. The primary problem behind this approach is the loss of data. In the following year Wang et al. [43] developed a Generative Adversarial Network (GAN) based Plankton classification model. The primary concern of their technique was to perform an effective image classification by resolving the class imbalance problem. The proposed architecture CGAN-Plankton generates the minority class

data using the Generative model which is discriminated using the Discriminative model. The approach widely outperforms other CNN based models. Mahmood et al. [44] also faced the class imbalance problem while proposing a technique to classify corals. Their proposed approach was also based on CNN and used the MLC dataset for the experiment which is prone to the class imbalance problem. A simple down-sampling of data was done to mitigate the problem which also resulted in the loss of valuable data. Rahman and Wang [45] used a Fully-Convolutional Neural Network (FCNN) for an augmentation in the IoU score. The main objective of the paper is to mitigate the object category segmentation problem and in their series of experiments they encountered the class-imbalance problem. In [46] the authors developed an efficient model by hybridizing CNN with uneven-margin based SVM to conduct an effective performance on imbalanced visual problems. Uneven margins were introduced to SVM to handle imbalanced data and was constructed by implying a margin parameter to the standard SVM model which is a ratio of the positive margin to the negative margin. In the proposed model of [46], the rearmost layer of the CNN is replaced by the uneven SVM which produced better performance than the standard model of SVM. In 2017 [47] Yue discussed the effects of imbalanced classification on CNN based malware image classification and also introduced the weighted softmax loss approach to mitigate the issue. The weight ω_k is calculated by,

$$\omega_k = (\Delta - \emptyset / C \times \Delta) + 1$$

where Δ is the dimension of the majority class, $\emptyset$ is dimension the selected class and C is the constant controlling the scaling of the weighted loss. Sahu et al. [48] performed a CNN based multi-label classification to detect a surgical tool. Their approach also faced a class-imbalance problem while training the CNN. The authors performed their experiments on M2CAI tool detection dataset by considering the co-occurrences of the tools. In [49] Li et al. proposed an imbalanced-aware CNN to mitigate the problem in vehicle classifications from aerial images. The authors extended the CNN model with a cost-effective imbalance aware feature map. In [50] the authors presented a very novel technique detect lithographic hotspots in electronic circuits based on CNN. The authors encountered a the imbalanced classification problem while performing the experiments and applied mirror-flipping and up-sampling to mitigate the issue, which resulted a significant improvement in the performance. Pouyanfar et al. in [51] performs a dynamic on the samples of each class of a dataset based on the F1-score to mitigate the imbalance data. A data augmentation module is present in the model is used to generate real time image data. In recent years some of the best improvements of CNN in imbalanced datasets can be found in mammal detection from UAV dataset [52], Marine image classification [53], sewer damage detection from CCTV images [54] and automatic bird detection [55].

4 Experimental Analysis

The most commonly known metric to evaluate a model's performance is Accuracy. However, highly imbalanced training or testing data provides a biasness to the accuracy [56] and deflects its values. This problem has been coined in various experiments [57, 58] suggesting it to be a very inferior metric to test on imbalanced data. However, some of the most commonly used metrics to evaluated these datasets are Area Under Curve (AUC) and Receiver Operating Characteristic curve (ROC). In [58] the authors suggested that Sensitivity (S_s) and Specificity (S_p) are undoubtedly an improvement over the Accuracy metric to evaluate the performance classes individually. The metrices can be defined as,

$$S_s = \frac{\epsilon_1}{\epsilon_1 + \epsilon_3} \tag{1}$$

$$S_p = \frac{\epsilon_4}{\epsilon_4 + \epsilon_2} \tag{2}$$

where, ϵ_1 represents True Positive values, ϵ_2 represents False Positive values, ϵ_3 represents False Negative values and ϵ_4 represents True Negative values. However, the authors constructed the G-Mean metric G, for their experiments. The metric can be written as,

$$G = \sqrt[2]{S_p \times S_s} \tag{3}$$

Johnson and Khoshgoftaar [59] also mentioned the accuracy-based evaluation anomalies in their comprehensive study. The authors mentioned that Precision values alone is critically affected by imbalanced distribution. However, Sensitivity values always remains unaffected as it only deals with the positive class. Luque et al. [60] stated the G-Mean metric and the Bookmaker Informedness B_{If}, as the best metrics which provides unbiased performance values. The B_{If} values can be calculated as,

$$B_{If} = S_p + S_s - 1 \tag{4}$$

We are going to concern ourselves with the particular case of Imbalanced Image Classification and compare these metrics that are a good fit for the problem while using CNN as a solution to our classification problem [47]. To document the performance of CNN, in our experiment we concentrated on using CIFAR-10 [61] and Fashion-MNIST [62] datasets. We eliminated data from 4 individual classes in Cifar-10 and 5 individual classes in Fashion-MNIST at similar rates to deliberately achieve imbalance and document the performance deflection. The CNN is trained at 100 epochs.

From Fig. 1a we see that the sensitivity deflection of the CNN for Cifar-10 indicates a downward trend in the metric score hence suggesting an overall reduction of the score with increase in Imbalanced data. However, initially the curve followed a
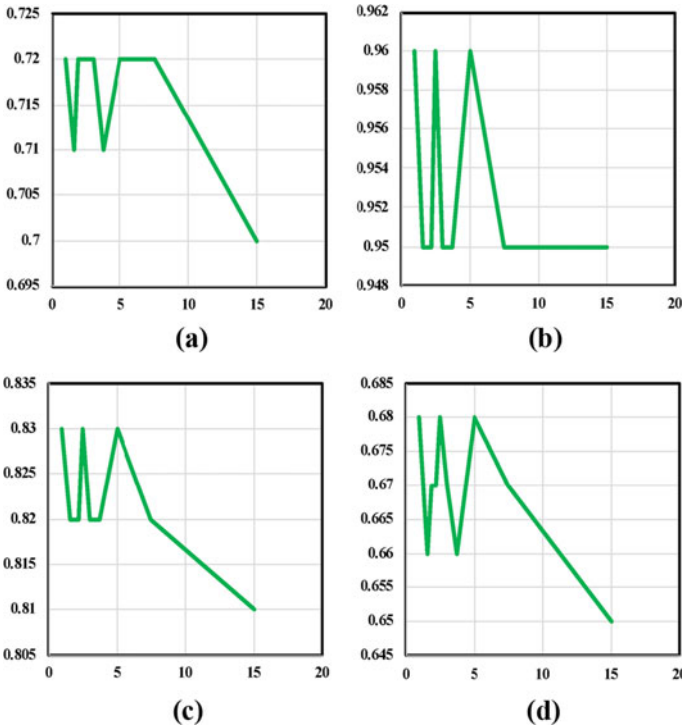
Fig. 1 Performance deflection of CNN for Cifar-10 dataset with the increase in imbalance ratio: **a** sensitivity, **b** specificity, **c** G-Mean and, **d** bookmaker informedness

zigzag trajectory but eventually ended up with a score of 0.7 from an initial recorded score of 0.72 during balanced state. A sharp downward curve of specificity from a score of 0.96–0.95 is also visible from Fig. 1b after crossing an IR value of 5 and an immediate stabilization of deflection indicates a unique characteristic of the data in the prediction of negative cases. From Fig. 1c we see the gradual drop in the G-mean score on crossing an IR value of 5 and eventually recording a deflection in score from 0.83 to 0.81. The deflection curve of the Bookmarker Informedness in Fig. 1d exhibits a trajectory similar to that of the G-Mean. A similar sharp drop in scores are visible after crossing an IR value of 5 recording a deflection from 0.68 to 0.65.

In Fig. 2a the Sensitivity value has a highest count of 0.92 when imbalance ratio is lowest. When the IR is at approximately 2.5, the sensitivity value decreases to 0.91. When the ratio is around 6 it further decreases to 0.89. Finally, when the ratio is approximately 12.5 the count reduces to 0.86. From here we can conclude that the overall sensitivity value decreases in a very rapid pace. Figure 2b suggests that the Specificity value has a highest count of 0.99 when IR is lowest. The specificity count drastically decreases when the IR count increases very little. After this decrease, the specificity count remains constant throughout the increase in imbalance. The G-Mean values exhibits abrupt rise and fall in Fig. 2c. During initial state, the G-Mean
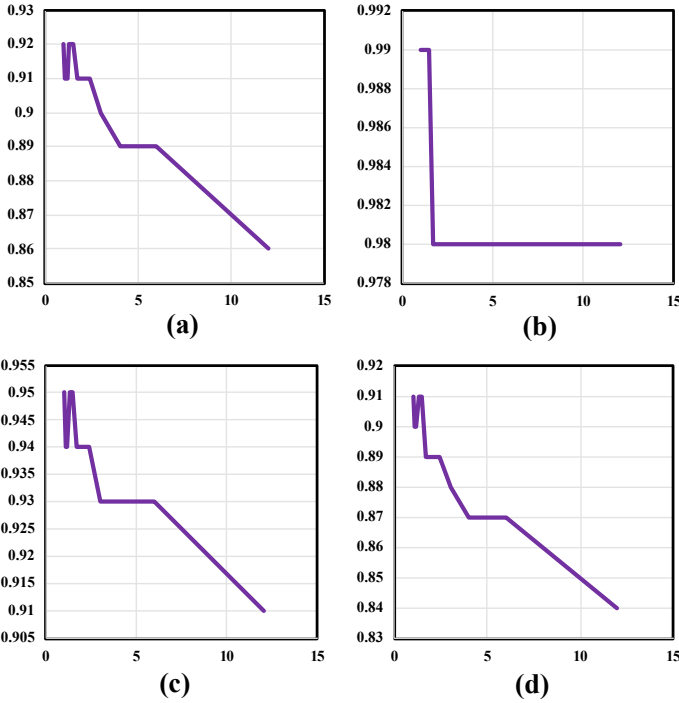
Fig. 2 Performance deflection of CNN for fashion MNIST dataset with the increase in imbalance ratio: **a** sensitivity, **b** specificity, **c** G-Mean and, **d** bookmaker informedness

count is highest, that is 0.95. With the gradual increase in the IR value, the G-Mean value decreases and then again rises. When the IR count is near to 6, the value is constant at 0.93. Finally, the G-Mean count decreases to 0.91 when the IR value is extreme. From Fig. 2d we see that the Bookmaker Informedness value is 0.91 during the balanced condition. It reduces and again increases within a very short interval. Till the imbalance ratio count is near to 6 the value keeps on decreasing after which we can see the value decreasing in a steep way. For the extreme value of IR, the Bookmaker Informedness value is as low as 0.84.

5 Conclusion

The current study evaluated CNNs in presence of imbalance classes. The study involved two artificially modified well known dataset CIFAR-10 and Fashion-MNIST. Multiple classes are perturbed by gradually removing samples and thereby making the dataset more imbalanced every time. The CNN is trained and tested

in terms of well known performance metrics namely Sensitivity, Specificity, G-Mean and Bookmaker Informedness which are specifically used to measure classifier performance in presence of imbalanced classes. The results revealed that the performance of CNN is severely affected when the imbalance ratio is increased. In case of CIFAR-10 the performance monotonically decreased after imbalance ratio is increased above 5 and a similar trend is observed for fashion-MNIST above 2.5. Future studies can be focused on developing algorithms to mitigate the performance issue of CNNs in presence of imbalanced classes.

References

1. Längkvist, M., Kiselev, A., Alirezaie, M., Loutfi, A.: Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sens. **8**(4), 329 (2016)
2. Roy, M., Chakraborty, S., Mali, K., Swarnakar, R., Ghosh, K., Banerjee, A., Chatterjee, S.: Data security techniques based on DNA encryption. In: International Ethical Hacking Conference, pp. 239–249. Springer, Singapore (2019)
3. Xu, H., Murphy, B., Fyshe, A.: Brainbench: a brain-image test suite for distributional semantic models. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2017–2021 (2016)
4. Kang, L., Kumar, J., Ye, P., Li, Y., Doermann, D.: Convolutional neural networks for document image classification. In: 2014 22nd International Conference on Pattern Recognition, pp. 3168–3172. IEEE (2014)
5. Yang, C., Hou, B., Ren, B., Hu, Y., Jiao, L.: CNN-based polarimetric decomposition feature selection for PolSAR image classification. IEEE Trans. Geosci. Remote Sens. **57**(11), 8796–8812 (2019)
6. Choi, J., Hwang, S.J., Sigal, L., Davis, L.S.: Knowledge transfer with interactive learning of semantic relationships. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
7. Ahmed, M.O., Vaswani, S., Schmidt, M.: Combining bayesian optimization and lipschitz optimization. Mach. Learn. 1–24 (2019)
8. Joty, S., Carenini, G., Ng, R., Murray, G.: Discourse processing and its applications in text mining. In: 2018 IEEE International Conference on Data Mining (ICDM), pp. 7–7. IEEE (2018)
9. Fatemi, B., Kazemi, S.M., Poole, D.: Record Linkage to Match Customer Names: A Probabilistic Approach. arXiv preprint arXiv:1806.10928 (2018).
10. Ghosh, K., Banerjee, A., Chatterjee, S., Sen, S.: Imbalanced twitter sentiment analysis using minority oversampling. In: 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), pp. 1–5. IEEE (2019)
11. Liu, X., Mou, L., Cui, H., Zhengdong, Lu., Song, S.: Finding decision jumps in text classification. Neurocomputing **371**, 177–187 (2020)
12. Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A.S., Balas, V.E.: Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Comput. Appl. **28**(8), 2005–2016 (2017)
13. Zhu, F., Ma, Z., Li, X., Chen, G., Chien, J.T., Xue, J.H., Guo, J.: Image-text dual neural network with decision strategy for small-sample image classification. Neurocomputing **328**, 182–188 (2019)
14. Dey, S., Singh, A.K., Prasad, D.K., Mcdonald-Maier, K.D.: SoCodeCNN: program source code for visual CNN classification using computer vision methodology. IEEE Access **7**, 157158–157172 (2019)

15. Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G.: A convolutional neural network cascade for face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325–5334 (2015)

16. Hu, B., Lu, Z., Li, H., Chen, Q.: Convolutional neural network architectures for matching natural language sentences. In: Advances in Neural Information Processing Systems, pp. 2042–2050 (2014)

17. Liu, P., Chen, W., Ou, G., Wang, T., Yang, D., Lei, K.: Sarcasm detection in social media based on imbalanced classification. In: International Conference on Web-Age Information Management, pp. 459–471. Springer, Cham (2014)

18. Khanchi, S., Vahdat, A., Heywood, M.I., Nur Zincir- Heywood, A.: On botnet detection with genetic programming under streaming data label budgets and class imbalance. Swarm Evol Comput **39**, 123–140 (2018).

19. Li, C., Liu, S.: A comparative study of the class imbalance problem in Twitter spam detection. Concur. Comput.: Pract. Exp. **30**(5), e4281 (2018)

20. Liang, Z., Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., Huang, J.X.: CNN-based image analysis for malaria diagnosis. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 493–496. IEEE (2016)

21. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

22. Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y., Li, J.: Deep learning for content-based image retrieval: a comprehensive study. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 157–166. ACM (2014)

23. Hou, Y., Zhang, H., Zhou, S.: Convolutional neural network-based image representation for visual loop closure detection. In: 2015 IEEE International Conference on Information and Automation, pp. 2238–2245. IEEE (2015)

24. Ji, S., Xu, W., Yang, M., Yu, K.: 3D convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 221–231 (2013)

25. Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M.: Medical image classification with convolutional neural network. In: 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 844–848. IEEE (2014)

26. He, X., Chen, Y.: Optimized input for CNN-based hyperspectral image classification using spatial transformer network. IEEE Geosci. Remote Sens. Lett. (2019)

27. Liu, X., Zhang, R., Meng, Z., Hong, R., Liu, G.: On fusing the latent deep CNN feature for image classification. World Wide Web **22**(2), 423–436 (2019)

28. Peng, Y.Q., Liao, M.X., Song, Y.X., Liu, Z.C., He, H.J., Deng, H., Wang, Y.L.: FB-CNN: feature fusion based bilinear CNN for classification of fruit fly image. IEEE Access (2019)

29. de Matos, J., de Souza Britto, A., de Oliveira, L.E.S., Koerich, A.L.: Texture CNN for Histopathological Image Classification. arXiv preprint arXiv:1905.12005 (2019).

30. Chawla, N.V., Bowyer, K.W., Hall, L.O., Philip Kegelmeyer, W.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**: 321–357 (2002)

31. Han, H., Wang, W.-Y., Mao, B.- H.: Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: International Conference on Intelligent Computing, pp. 878–887. Springer, Berlin, Heidelberg (2005)

32. He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322–1328. IEEE (2008)

33. Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C.: Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 475–482. Springer, Berlin, Heidelberg (2009)

34. Sharma, S., Bellinger, C., Krawczyk, B., Zaiane, O., Japkowicz, N.: Synthetic oversampling with the majority class: a new perspective on handling extreme imbalance. In: 2018 IEEE International Conference on Data Mining (ICDM), Singapore, pp. 447–456 (2018)

35. Bellinger, C., Sharma, S., Japkowicz, N., Zaıane, O.: Framework for Extreme Imbalance Classification SWIM: Sampling With the Majority Class
36. Mari Antonie, M.-L., Zaiane, O.R., Coman, A.: Application of data mining techniques for medical image classification. In: Proceedings of the Second International Conference on Multimedia Data Mining, pp. 94–101. Springer (2001)
37. Ryan, C., Fitzgerald, J., Krawiec, K., Medernach, D.: Image classification with genetic programming: building a stage 1 computer aided detector for breast cancer. In: Handbook of Genetic Programming Applications, pp. 245–287. Springer, Cham (2015)
38. Su, F., Xue, L.: Graph learning on K nearest neighbours for automatic image annotation. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval–ICMR '15 (2015)
39. Schaefer, G., Nakashima, T.: Strategies for addressing class imbalance in ensemble classification of thermography breast cancer features. In: 2015 IEEE Congress on Evolutionary Computation (CEC), pp. 2362–2367. IEEE (2015)
40. Pérez-Ortiz, M., Sáez, A., Sánchez-Monedero, J., Gutiérrez, P.A., Hervás-Martínez, C.: Tackling the ordinal and imbalance nature of a melanoma image classification problem. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 2156–2163. IEEE (2016)
41. Dai, B., Xiangqian, Wu., Wei, Bu.: Retinal microaneurysms detection using gradient vector analysis and class imbalance classification. PLoS ONE **11**(8), e0161556 (2016)
42. Lee, H., Park, M., Kim, J.: Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In: 2016 IEEE International Conference on Image Processing (ICIP) (2016)
43. Wang, C., Yu, Z., Zheng, H., Wang, N., Zheng, B.: CGAN-plankton: towards large-scale imbalanced class generation and fine-grained classification. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 855–859. IEEE (2017)
44. Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., Fisher, R.B.: Coral classification with hybrid feature representations. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 519–523. IEEE (2016)
45. Rahman, Md.A., Wang, Y.: Optimizing intersection-over-union in deep neural networks for image segmentation. In: International Symposium on Visual Computing, pp. 234–244. Springer, Cham (2016)
46. Geng, M., Wang, Y., Tian, Y., Huang, T.: CNUSVM: hybrid CNN-uneven SVM model for imbalanced visual learning. In: 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), pp. 186–193. IEEE (2016)
47. Yue, S.: Imbalanced malware images classification: a CNN based approach. arXiv preprint arXiv:1708.08042 (2017).
48. Sahu, M., Mukhopadhyay, A., Szengel, A., Zachow, S.: Addressing multi-label imbalance problem of surgical tool detection using CNN. Int. J. Comput. Assist. Radiol. Surg. **12**(6), 1013–1020 (2017)
49. Li, F., Li, S., Zhu, C., Lan, X., Chang, H.: Cost-effective class-imbalance aware CNN for vehicle localization and categorization in high resolution aerial images. Remote Sens. **9**(5), 494 (2017)
50. Yang, H., Luo, L., Jing, Su., Lin, C., Bei, Yu.: Imbalance aware lithography hotspot detection: a deep learning approach. J. Micro/Nanolith. MEMS MOEMS **16**(3), 033504 (2017). https://doi.org/10.1117/1.JMM.16.3.033504https://doi.org/10.1117/1.JMM.16.3.033504
51. Pouyanfar, S., Tao, Y., Mohan, A., Tian, H., Kaseb, A.S., Gauen, K., Dailey, R. et al.: Dynamic sampling in convolutional neural networks for imbalanced data classification. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 112–117. IEEE (2018)
52. Kellenberger, B., Marcos, D., Tuia, D.: Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. **216**, 139–153 (2018)
53. Langenkämper, D., van Kevelaer, R., Nattkemper, T.W.: Strategies for tackling the class imbalance problem in marine image classification. In: International Conference on Pattern Recognition, pp. 26–36. Springer, Cham (2018)

54. Li, D., Cong, A., Guo, S.: Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification. Autom. Constr. **101**, 199–208 (2019)
55. Niemi, J., Tanttu, J.T.: Deep learning case study on imbalanced training data for automatic bird identification. In: Deep Learning: Algorithms and Applications, pp. 231–262. Springer, Cham (2020)
56. Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw **106**, 249–259 (2018)
57. Sarkar, S., Khatedi, N., Pramanik, A., Maiti, J.: An ensemble learning-based undersampling technique for handling class-imbalance problem. In: Proceedings of ICETIT, pp. 586–595. Springer, Cham (2020)
58. Tang, Y., Zhang, Y.-Q., Chawla, N.V., Krasser, S.: SVMs modeling for highly imbalanced classification. IEEE Trans. Syst. Man, Cybern. Part B (Cybernetics) **39**(1): 281–288 (2008)
59. Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. J. Big Data **6**(1), 27 (2019)
60. Luque, A., Carrasco, A., Martín, A., de las Heras, A.: The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognit. **91**, 216–231 (2019)
61. Yang, L., Bankman, D., Moons, B., Verhelst, M., Murmann, B.: Bit error tolerance of a CIFAR-10 binarized convolutional neural network processor. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5. IEEE (2018)
62. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)