Applying the FAIR Principles to Accelerate Health Research in Europe in the Post COVID-19 Era
J. Delgado et al. (Eds.)

doi:10.3233/SHTI210801

A Learning Framework for Medical Image-Based Intelligent Diagnosis from Imbalanced Datasets

Tetiana BILOBORODOVA[a,1], Inna SKARGA-BANDUROVA[b], Mark KOVERHA[c], Illia SKARHA-BANDUROV[d] and Yelyzaveta YEVSIEIEVA[e]
[a]*G.E. Pukhov Institute for Modelling in Energy Engineering, Ukraine*
[b]*Oxford Brookes University, United Kingdom*
[c]*Volodymyr Dahl East Ukrainian National University, Ukraine*
[d]*Luhansk State Medical University, Ukraine*
[e]*School of Medicine, V. N. Karazin Kharkiv National University, Ukraine*

Abstract. Medical image classification and diagnosis based on machine learning has made significant achievements and gradually penetrated the healthcare industry. However, medical data characteristics such as relatively small datasets for rare diseases or imbalance in class distribution for rare conditions significantly restrains their adoption and reuse. Imbalanced datasets lead to difficulties in learning and obtaining accurate predictive models. This paper follows the FAIR paradigm and proposes a technique for the alignment of class distribution, which enables improving image classification performance in imbalanced data and ensuring data reuse. The experiments on the acne disease dataset support that the proposed framework outperforms the baselines and enable to achieve up to 5% improvement in image classification.

Keywords. Medical image classification, imbalanced data, machine learning oversampling

1. Introduction

The recent success of machine learning (ML) and computer vision allows elevating medical diagnostics to a new level, particularly in the classification of visual data enabling to solve problems of medical analytics and clinical decision making more rapid and accurate. Being a key component of intelligent diagnosis, medical images classification includes identifying the features in an image and predicting the class of a specific object in an image. In this context, data quality has a significant impact on the success of ML algorithms and is a core of scientific knowledge. As mentioned in [1], the ideal medical image dataset for the ML application is Findable, Accessible, Interoperable, and Reusable (FAIR) [2]. This basically means that image datasets should have adequate volume and class distribution, be well-annotated, verifiable, ground-truth, and reusable. However, in the wild, due to their nature, medical image datasets are being the long way

[1] Corresponding Author, Tetiana Biloborodova, G.E. Pukhov Institute for Modelling in Energy Engineering, Kyiv, Ukraine; E–mail: beloborodova.t@gmail.com.

of the FAIR principles, and in many cases, they are closed, limited distributed, relatively few annotated and highly imbalanced.

Meanwhile, the ML algorithms require a large number of annotations, which is a time-consuming and labour-intensive process and often, the class distribution inside datasets is not equal. This is because of identifying and predicting process often includes rare events [3]. Medical data demonstrate an uneven distribution of classes in rare clinical cases or diseases, which makes it difficult to form a balanced dataset for training which in turn leads to poor reproducibility of the ML algorithms. Rare cases result in data imbalance, namely the imbalance in the number of objects in different classes. Imbalanced data refers to a dataset where the class distribution is not uniform among the classes. The prevailing class is called the majority class, and the smallest class in terms of objects is the minority class [4]. Imbalanced data can negatively affect the accuracy of the models and lead to incorrect or erroneous classification results.

To following the FAIR principles, this study proposes the technique for dealing with heavily imbalanced datasets and introduces the concept of a machine-readable data preprocessing and resampling for model learning. We aim to extend previous research in imbalanced data classification, improve quality of computer vision-based disease diagnostic and provide usability and reusability of medical image datasets. Inoculation of the FAIR principles as a new data management strategy results in significant improvements in automation of medical image diagnostic through machine readability and enable reuse data and improve their scalability.

2. Methods

Methods to handle imbalanced data can be divided into three large categories: data-layer methods, algorithm-layer methods, and cost-sensitive learning methods [5]. Data layer methods include resampling (oversampling, undersampling and hybrid) techniques. This is the most straightforward and widely adopted approach for dealing with highly imbalanced datasets. All of these techniques follow FAIR principles in part of being machine-readable for reusable. Algorithm-level methods include the use of essemble methods based on machine learning algorithms [6]. Cost-sensitive learning methods target the problem of imbalanced learning by using other evaluation metrics and different cost matrices that describe the costs for misclassifying any particular data example [7].

2.1. Imbalanced datasets

We define imbalanced dataset S with m objects, $|S| = m$, as $S = \{(x_i, y_i), i = 1, ..., m$, where $x_i \in X$ is an object in n-dimension space of input features $X=\{f_1, f_2, ..., f_n\}$, and $y_i \in Y = \{1, \ldots, C\}$ is the label, associated with object x_i. At it simplest, C = 2 means the binary classification task where two subsets are defined as $S_{min} \subset S$ the minor class subset S_{min} in S and $S_{maj} \subset S$ is a major class subset such as $S_{min} \cap S_{maj} = \{\Phi\}$ and $S_{min} \cup S_{maj} = \{S\}$.

The objects generated from the dataset S are defined as E, with disjoint subsets E_{min} and E_{maj} that represent the minority and majority classes of E, respectively, each time they are used.

2.2. Proposed approach

The basic structure of proposed approach to obtaining an accurate model for classification of medical images in the imbalanced datasets is shown in Fig. 1.

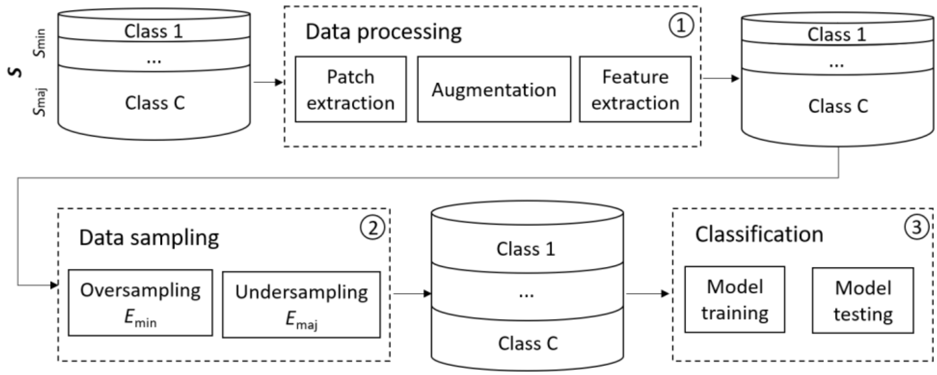

Figure 1. The structure of proposed approach

The learning framework for medical image-based diagnosis from imbalanced datasets incorporates data processing, data sampling, and classification. Since we are dealing with imbalanced datasets, collected data are being annotated and analysed in terms of minority and majority classes. The data processing phase includes automatic patch extraction, data augmentation and feature extraction. The output data at this phase are the extracted features. For phase 2, class distribution is evaluated, and the resampling technique is selected depending on the size and type (minority or majority) of the imbalance in different classes. It can be done by removing samples from the majority class (under-sampling) and/or adding more examples from the minority class (over-sampling). Oversampling is used for sampling minority class objects, while undersampling is used for sampling majority class objects. From this phase, we obtain a quasi balanced machine-readable dataset ready for model training. Finally, the model training and validation are performed.

3. Results

To evaluate the performance of the proposed technique, the experiments with an open medical image dataset ACNE04 provided by Wu et al. [8] were conducted. The dataset includes 1457 face images and expert annotations according to the Japanese acne grading scale. There are four acne severity classes, namely 0 Mild equal to 410 samples, 1 Moderate equal to 506 samples, 2 Severe equal to 146 samples, and 3 Very severe equal to 103 samples. In general, we used 1165 images to train the model and 291 images to test the model, which corresponds to a distribution of 80% for training and 20% for testing. Proposed approach runs on an NVIDIA GeForce GTX 1060 with 3 GB VRAM and is implemented based on the PyTorch framework.

The data processing phase included patch extraction, data augmentation and feature extraction. At the patch extraction stage, we utilised two pre-trained models: (1) shape_predictor_68_face_landmarks model [9] and (2) the One Eye model [10]. In case when any of these models could not process the image, the entire original image was used for the next phase. Each patch inherited the label of the original image and had a binding to it, which was used later to get a general estimate of the degree severity of acne for a photo. For augmentation, a sliding translation of patches was used. Further, the feature extraction for each patch was carried out using the ResNet-152 model [11]. Data distribution by classes after augmentation is presented as follow. 0 Mild: 3556 samples, 1 Moderate: 4333 samples, 2 Severe: 1843 samples, and 3 Very severe: 1514 samples. Each patch is bound to the original image, and thus the extracted features inherit the dependencies of the patches. Then, the classes were revised following acne severity grades, and extracted features were used for sampling. Sampling and minority class generation were done via Synthetic Minority Oversampling Technique (SMOTE) [12]. The total number of samples for each class was fitted to the most numerous class after feature extraction and equal to 4333 for each class.

Data generated at the oversampling phase was used to train a convolutional neural network (CNN) model and estimate the severity of the acne from the face image. Model training was run for 17985.2 seconds. The classification problem was transformed into a regression task at this phase by defining acne severity grades as integer equivalents. It was done to reduce possible subjectivity in the expert's annotation of the acne severity. The inverse transformation was done using [0.5, 1.5, 2.5] as the edge list. Model evaluation using the trained CNN is implemented on test data. Since the problem was reduced to a regression problem, the corresponding criteria for assessing the regression quality were used. As a result, the following values were obtained for the ACNE04: RMSE = 0.397419, EV = 0.826736, MAPE = 0.199264, R2 = 0.826682.

4. Discussion

In order to test effectiveness of the proposed approach, we compared our results with outcomes without oversampling (Table 1). Classification accuracy was calculated after converting the results back from continuous to discrete scale using [0.5, 1.5, 2.5] as a list of edges. Delta was calculated as the difference in the results between these two approaches in percentage.

Table 1. The results of experiments with the basic and proposed approach

Metrics	Without oversampling	Proposed approach	Delta, %
RMSE	0.422356	0.397419	5.904261
EV	0.826736	0.873874	5.7017
MAPE	0.199264	0.171855	13,75512
R^2	0.826682	0.873646	5.68102
Accuracy	80 %	85 %	5

Both RMSE and MAPE show a smaller error, while the EV and R^2 criteria show higher values for proposed approach, which indicates a higher quality of the model. Comparison of the obtained results with two benchmark models is presented in Table 2.

Table 2. Comparison of acne classification research

Approach for acne classification	Accuracy (%)	ER (%)
Wu et al. [8]	84.11	15.89
Lim et al. [13]	67	33
Ours	85	15

As can be seen from the table, the proposed approach showed the highest accuracy and the lowest error rate in comparison with studies with imbalanced data, however, it should be mentioned that it did not outperform the results for study [14] where data was initially balanced (stated accuracy 99.44%), which sounds natural but needs further investigation.

5. Conclusion

Accurate classification of medical images is one of the first steps towards the wide adoption of computer vision into healthcare industry. In this paper, we propose a complex approach for imbalanced medical image classification. It is grounded on FAIR prinsiple where medical image datasets remain useful even in high inbalance and can be reused to train, test, validate, verify, and regulate ML products. Experiments on acne image dataset showed that the proposed learning framework able to improve the classification performance metrics and proved their advantages in comparison with basic approach without oversampling.

References

[1] Kohli MD, Summers RM, Geis JR. Medical image data and datasets in the era of machine learning—whitepaper from the 2016 C-MIMI meeting dataset session. Journal of digital imaging. 2017 Aug;30(4):392-9.
[2] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J. Andra Waagmeester. Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016.
[3] Weiss GM, Hirsh H. Learning to predict extremely rare events. InAAAI workshop on learning from imbalanced data sets 2000 Jul (pp. 64-68). Austin: AAAI Press.
[4] Yijing L, Haixiang G, Xiao L, Yanan L, Jinling L. Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. Knowledge-Based Systems. 2016 Feb 15;94:88-104.
[5] Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications. 2017 May 1;73:220-39..
[6] *Imbalanced learning: Foundations, Algorithms, Applications* (Eds.: Haibo He, Yunqian Ma) Wiley-IEEE Press, 2013.
[7] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48 (2019).
[8] Wu X, Wen N, Liang J, Lai YK, She D, Cheng MM, Yang J. Joint acne image grading and counting via label distribution learning. InProceedings of the IEEE/CVF International Conference on Computer Vision 2019 (pp. 10642-10651)..
[9] Davisking/dlib-models. Github.com. [Online] https://github.com/davisking/dlib-models
[10] OpenCV. Github.com. [Online] https://github.com/opencv/opencv/blob/master/data/haarcascades/haarcascade_eye.xml.
[11] The Microsoft Cognitive Toolkit. Cntk.ai. [Online] https://www.cntk.ai/Models/Caffe_Converted/ResNet152_ImageNet_Caffe.model
[12] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002 Jun 1;16:321-57.
[13] Lim ZV, Akram F, Ngo CP, Winarto AA, Lee WQ, Liang K, Oon HH, Thng ST, Lee HK. Automated grading of acne vulgaris by deep learning with convolutional neural networks. Skin Research and Technology. 2020 Mar;26(2):187-92.
[14] Junayed MS, Jeny AA, Atik ST, Neehal N, Karim A, Azam S, Shanmugam B. Acnenet-a deep cnn based classification approach for acne classes. In2019 12th International Conference on Information & Communication Technology and System (ICTS) 2019 Jul 18 (pp. 203-208). IEEE.