On the Impact of Imbalanced Data in Convolutional Neural Networks Performance

Francisco J. Pulgar(✉), Antonio J. Rivera, Francisco Charte, and María J. del Jesus

Depart of Computer Science, University of Jaén, Jaén, Spain
{fpulgar,arivera,fcharte,mjjesus}@ujaen.es
http://simidat.ujaen.es

Abstract. In recent years, new proposals have emerged for tackling the classification problem based on Deep Learning (DL) techniques. These proposals have shown good results in certain fields, such as image recognition. However, there are factors that must be analyzed to determine how they influence the results obtained by these new algorithms. In this paper, the classification of imbalanced data with convolutional neural networks (CNNs) is analyzed. To do this, a series of tests will be performed in which the classification of real images of traffic signals by CNNs will be performed based on data with different imbalance levels.

Keywords: Deep learning · Convolutional neural network · Image recognition · Imbalanced dataset

1 Introduction

Classification is one of the most widely studied tasks within automatic learning, the fundamental reason being its application to real cases. The objective is to obtain a model that allows the classification of new examples, starting from a set of examples that are correctly labeled [1]. Image classification is a problem that can be solved by applying a classifier.

In recent years, there has been a rise in the use of techniques based on Deep Learning to tackle the classification problem. This heyday was mainly due to two reasons: the large amount of data available and the increase in processing capacity. These DL techniques have provided good results in classification, especially in fields such as image and sound recognition [15,16].

One of the techniques that have obtained better results in the task of image recognition are convolutional neural networks (CNNs) [20]. Given the nature of the convolutions, these networks are able to learn to classify all the types of data where the attributes are distributed continuously along the entrance, as it is given in the images.

Despite the good results obtained with CNN, the fact that these techniques are very recent causes new factors to come into play. One of these is the need to analyze how data imbalance affects the classification performance of this type of

F.J. Martínez de Pisón et al. (Eds.): HAIS 2017, LNAI 10334, pp. 220–232, 2017.
DOI: 10.1007/978-3-319-59650-1_19

tool. Most real data sets show some imbalance degree, thus the importance of studying how this aspect influences classification performance.

Therefore, this study focuses on analyzing the effect of the imbalance of data in the classification of real images of traffic signals using CNN. Our starting hypothesis is that classification performance will improve as the imbalance level decreases. The reason for this hypothesis is a question of similarity with other techniques used in classification, since if the imbalance of data influences the quality of classification using techniques such as traditional neural networks, it could also influence the classification obtained through CNNs. Similarly, the nature of CNNs can be influenced by the imbalance, since the adjustment of the parameters can depend on the different number of examples of the classes.

This paper is structured as follows: Sect. 2 explains how imbalanced datasets influence classification. In Sect. 3 the DL concept is introduced and the CNNs which will be used to classify the images in the experimentation are analyzed in more detail. Section 4 exposes the hypothesis raised in this study. In Sect. 5 we attempt to verify the hypothesis established by CNN classification, starting from data with different degree of imbalance. Lastly, in Sect. 6 some conclusions are drawn.

2 The Imbalance Problem in Classification

Classification is a predictive task that usually uses supervised learning methods [2]. Its purpose is to learn, based on previously labeled data, patterns in order to predict the class to assign to future examples which are not labeled. In traditional classification, datasets are composed of a set of input features and a unique value in the output attribute, the class or label.

In many applications aimed at solving the classification problem there is a significant difference between the number of elements of the different classes, so the probability that an example belongs to each of those classes will also be different. This situation is known as the imbalance problem [3–5]. In many cases, the minority class is the one that has the most interest in the classification and has a greater cost in the case of not doing well.

Most of the standard classification algorithms obtain a good coverage when classifying the elements of the majority class, but the minority class is misclassified frequently. Therefore, these classification algorithms, which obtain good results for a traditional framework, will not necessarily work well with imbalanced data. There are several reasons for this:

- Many of the performance measures used to guide the overall procedure, such as the accuracy rate, disadvantage the minority class.
- The rules that predict the minority class are very specialized and their coverage very low, and therefore are usually discarded in favor of more general rules, that is, those that predict the majority class.
- The treatment of noise can affect the classification of minority classes since, such classes could be treated as noise and discarded erroneously or the actual noise may degrade the classification of the minority classes.

The main obstacle caused by this type of problem is that the classification algorithms are biased toward the majority class and there is a higher Error Rate when trying to classify the minority class. To face this problem, different proposals have emerged in the last years [6], which can be classified as:

- **Data sampling:** In this type of solution it is intended to modify the training set from which the classification algorithm starts. The objective is to obtain training data with a more balanced class distribution, so the classification can be performed in the standard way [7].
- **Algorithmic modification:** The objective pursued in this type of solution is to make an adaptation of the traditional classification algorithms in order to deal with the problem of imbalance in the data [8]. In these cases the data set is not modified, but it is the algorithm itself that adapts.
- **Cost-sensitive learning:** This type of solution can incorporate modifications both at the data level and at the algorithmic level, and is based on penalizing to a greater extent the mistakes made in classifying the minority classes than those committed when classifying the majority classes [9].

When dealing with imbalanced data, there are other factors of the same that must be taken into account, since they can greatly influence the results of the classification obtained. Some of these factors can be overlapping between the classes or noisy data.

3 Deep Learning

The need to extract higher-level information from the data analyzed through learning tools has led to the emergence of new areas of study, such as DL [10]. DL models are based on a deep architecture (multilayer) whose objective is to map the relationships between the characteristics of the data and the expected results [11]. There are some advantages to using DL-based techniques:

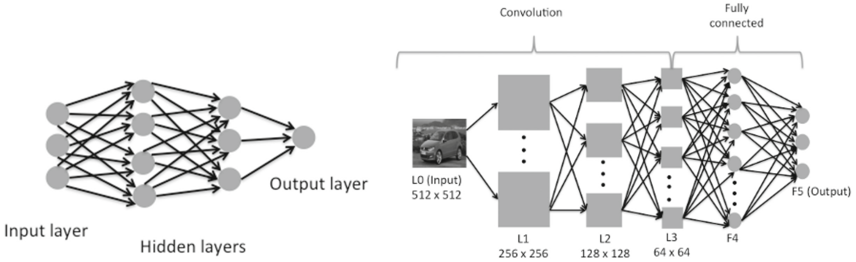
- DL models incorporate mechanisms for generating new characteristics by themselves, without having to develop them in an external phase.
- DL techniques improve yield in terms of time spent performing some of the more expensive tasks, such as feature engineering.
- DL-based models have proven successful in dealing with problems in certain fields such as image or sound recognition, improving over traditional techniques [15,16].
- DL-based solutions have a great ability to adapt to new problems.

Due to the good results obtained using DL-based proposals, recently different architectures have been developed, such as, CNN [17] or recurrent neural networks [18]. These architectures have been designed for multiple fields of application, producing very efficient results in the field of image recognition. In Sect. 3.1, we introduce CNN that will be used to perform the experimentation associated with this study.

3.1 Convolutional Neural Network

CNNs are a type of deep neural network based on the way some animals visualize. These networks have been shown to be very effective in certain areas such as image recognition and classification.

CNNs focus on the idea of spatial correlation by applying a series of local connectivity patterns between the neurons of the adjacent layers [19,20]. This implies that, unlike traditional networks where each neuron connects to all the neurons of the previous layer, in the CNNs each neuron only connects to a small region of the previous layer. Another fundamental difference is that the neurons of CNNs are arranged in three dimensions, whereas in the traditional networks they are realized in two dimensions. Figure 1 shows the difference between both types of networks.



(a) Traditional Neural Network (b) Convolutional Neural Network

Fig. 1. Difference between traditional neural network and CNN.

The architecture of a CNN is based on a sequence of layers, each of them transforming a volume of activations to another by a certain function. There are three types of layers:

- **Convolution layer:** On this layer lies the majority of the computational weight. It is formed by a set of filters that must be learned during the process. Each filter is spatially small, ensuring local connectivity, and scrolls along the entire input, generating a global map of activations. In this way, it is possible for the network to learn filters that activate when there is some determined pattern, for example, edges or curves.
- **Pooling Layer:** This type of layer is normally inserted between several successive convolution layers. The fundamental objective is to reduce the spatial size of the representation by applying the function determined to each slice of the depth of the input independently. In this way a reduction of the height and width is performed but without modifying the depth of the representation. The functions that can be applied are of different types, for example, maximum or average.
- **Fully connected layer:** In this type of layer each of the neurons has complete connections to all the neurons of the anterior layer, differentiated thus from the convolutional layer in which it connects to a local region. The calculations that are performed are the same in both types of layers.

These types of layers that have been listed above are used to form a complex CNN. To do this, layers of different types are stacked according to the model to be constructed.

4 Impact of Imbalanced Data on Convolutional Neural Networks

Once the theoretical principles necessary to establish the bases of this study have been introduced the initial hypothesis is presented, as well as the reasons that lead to this hypothesis. Also, some elements used in the experimentation are presented.

Classification is usually applied to real data. Therefore, the characteristics of this type of data must be taken into account. One of these traits is the imbalance of the data, which implies a great difference between the number of majority and minority classes of the examples used, as we have seen in Sect. 2.

There are different studies [6–9] that show that this characteristic of the data affects different classification models and propose solutions to reduce the effects of classifying data of this type. However, because CNN-based classification techniques are very recent, we have not found any studies on the influence of imbalance on them. We assume that this feature can affect the performance of CNNs. This leads us to propose the initial hypothesis of the work: the classification results obtained through CNNs will be affected by the imbalance of the data.

Section 5 is intended to verify the proposed hypothesis. To do this, a CNN will be used to classify real images of traffic signals, performing different executions in which various sets of examples from the same dataset are used with a decreasing imbalance ratio (IR).

5 Experimentation

The objective of this study is to analyze whether the data imbalance can negatively influence the classification of images made by CNNs. To do this, a set of tests will be performed using data with a different degree of imbalance in them. These data correspond to real images of traffic signals [12], the objective being to correctly classify the signal according to its type. In order to perform the classification a CNN will be used. The network used will be the same for all the experiments performed. Also, it should be noted that no mechanism is used to minimize the effect of the imbalance, since it is intended to analyze how it affects the CNN classification.

5.1 Experimental Framework

In performing this experiment, a traffic signal dataset [12] with a total of 11 910 images belonging to 43 different types of signals or classes has been used. First, it is necessary to perform a pre-processing of the images. This phase has two fundamental objectives: on the one hand, to trim the image in order to select only the traffic signal (Fig. 2); and on the other hand, to scale the images so that all of them have the same size. In this sense, it has been decided to scale them to a size of 32 × 32 since it is the most widely used in other studies [19].

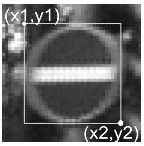


Fig. 2. Example of image crop (source: http://benchmark.ini.rub.de).

Once all the images included in the dataset have been pre-processed. The images corresponding to 10 classes have been selected in order to emphasize the imbalance between the data. Therefore, the classes selected have been the 5 with more examples and the 5 with fewer examples. In this way, we will obtain a subset of the original dataset.

An important aspect to keep in mind when working with imbalanced data is the IR. This measure is defined as the ratio between the number of examples of the majority class and the number of instances of the minority class [13,14].

Another aspect to be taken into account is that the number of images used in each experiment is the same even though the IR of the samples is different. This is important because if the number of examples varies significantly it can affect the results obtained, thus hiding the effects of the imbalance in them. In this way, it has been selected 2 700 images for each execution. The reason for selecting this value is given by the number of images of the minority classes in the original dataset, having 270 examples of the minority classes and about 2 700 of the majority. When balancing the dataset 270 images of each class are selected, so the total number of examples is 2 700, which is kept constant throughout all runs even if the IR changes.

Finally, it should be pointed out that in order to evaluate the efficiency in the classification of the CNN in each case, the Error Rate, Accuracy and Recall will be used.

5.2 CNN Architecture

Although the set of images used will change in distinct executions, the CNN used will be similar in all cases in order to see the effects of the different IRs. This CNN has an architecture whose sequence of layers is as follows:

- **Convolution layer 1:** 32 filters are applied on the original image. These filters are 5 × 5 size. This creates 32 feature maps.
- **Pooling layer 1:** 32 feature maps are subsampled using max pooling with a pooling window of size 2 and a stride of 2.
- **Convolution layer 2:** This layer applies 64 filters with a size of 5 × 5 to these subsampled images and generates new feature maps.

- **Pooling layer 2:** The previously generated features maps are subsampled. For this, max pooling with a window of size 2 and a stride of 2.
- **Fully connected layer:** In this layer all the generated features are combined and used in the classifier. This layer has as many individuals as classes have the problem.

In the training process, the cross-entropy error is used to evaluate the network and, later, the back-propagation is used to modify the weights of the network. The configuration chosen is the default setting used in the TensorFlow software[1], considering the initial size of the images and the different output values that the problem may have.

5.3 Results Analysis

In the experiments was carried out to classify the images through a CNN, starting from datasets different IR. In particular, four experiments with IR 1/10, 1/5, 1/3 and 1/1 were performed.

As mentioned in Sect. 5.1 the dataset used has a total of 11 910 images. However, it has also been indicated that each experiment must have a total of 2 700 images, with the aim that all experiments have the same number of examples. Therefore, the first step has been to reduce the number of images of each class proportionally and randomly, in order to reduce the desired number by maintaining the corresponding IR. So, the dataset presents the distribution of examples that can be seen in Table 1 for each case.

Table 1. Train and test examples per experimentation

Class	IR 1/10		IR 1/5		IR 1/3		IR 1/1	
	Train	Test	Train	Test	Train	Test	Train	Test
1	36	11	66	20	98	31	203	67
2	351	115	320	106	288	95	203	67
3	392	130	361	118	325	106	203	67
4	365	121	334	111	301	100	203	67
5	376	125	345	115	311	103	203	67
6	36	11	65	21	97	32	203	67
7	39	13	72	24	108	36	203	67
8	36	11	65	21	97	32	203	67
9	360	120	330	110	297	99	203	67
10	39	13	72	24	108	36	203	67
Total	2030	670	2030	670	2030	670	2030	670

[1] https://www.tensorflow.org/.

Table 1 shows the number of examples of each of the classes in the dataset for each experiment. It can be seen that there are a total of 2 700 images of which 2 030 will be used to train the network and 670 to evaluate the model in all cases. However, it can perceive how the ratio between examples of the majority and minority classes is different. The results obtained in each of the experiments are exposed below.

Results with IR 1/10

The first experiment was carried out to classify the images through a CNN, starting from a dataset with a high degree of imbalance between classes. Table 1 shows the number of examples of each of the classes in the dataset. Similarly, it can be seen how, in this first experiment, there is an IR of approximately 1/10 between the minority and majority classes. Once the dataset is set for experimentation, a CNN is used to perform the classification, obtaining the results presented in Tables 2 and 3.

Table 2 shows the number of test samples per class and the number of errors that the model has made in classifying these examples. Also, in Table 3 it can be seen the Error Rate, Recall and Precision by class. Both tables represent the results for the 4 experiments performed.

Table 2. Number of total and error examples in test per experimentation.

Class	IR 1/10		IR 1/5		IR 1/3		IR 1/1	
	Test	Error	Test	Error	Test	Error	Test	Error
1	11	4	20	4	31	2	67	1
2	115	1	106	3	95	2	67	0
3	130	1	118	2	106	2	67	3
4	121	2	111	0	100	2	67	0
5	125	0	115	2	103	0	67	0
6	11	4	21	2	32	2	67	0
7	13	2	24	0	36	0	67	0
8	11	4	21	1	32	0	67	0
9	120	1	110	1	99	1	67	1
10	13	3	24	0	36	0	67	3
Total	670	22	670	15	670	11	670	8

In this first experiment where the IR is 1/10, the results obtained show the Error Rate per class, with the percentage of global Error being 0.033, since 22 images of a total of 670 are classified badly. Also, the global Accuracy value is 0.963 and the global Recall value is 0.848. These results will serve as a basis for determining whether executions performed with lower imbalance improve them.

Table 3. Results for test dataset

Class	IR 1/10			IR 1/5			IR 1/3			IR 1/1		
	Error	Precision	Recall	Error	Precision	Recall	Error	Precision	Recall	Error	Precision	Recall
1	0.364	**1.000**	0.636	0.200	**1.000**	0.800	0.065	**1.000**	0.935	**0.015**	0.985	**0.985**
2	0.009	0.966	0.991	0.028	0.954	0.972	0.021	0.989	0.979	**0.000**	**1.000**	**1.000**
3	**0.008**	0.963	**0.992**	0.017	0.951	0.983	0.019	0.990	0.981	0.045	**1.000**	0.955
4	0.017	0.983	0.983	**0.000**	**1.000**	**1.000**	0.020	0.990	0.980	**0.000**	0.985	**1.000**
5	**0.000**	0.977	**1.000**	0.017	0.983	0.983	**0.000**	0.956	**1.000**	**0.000**	**0.985**	**1.000**
6	0.364	0.875	0.636	0.095	**1.000**	0.905	0.062	0.968	0.937	**0.000**	0.985	**1.000**
7	0.154	0.917	0.846	**0.000**	0.960	**1.000**	**0.000**	0.947	**1.000**	**0.000**	**1.000**	**1.000**
8	0.364	**1.000**	0.636	0.048	**1.000**	0.952	**0.000**	**1.000**	**1.000**	**0.000**	0.985	**1.000**
9	**0.008**	0.952	**0.992**	0.009	0.991	0.991	0.010	**1.000**	0.990	0.015	0.956	0.985
10	0.231	**1.000**	0.769	**0.000**	**1.000**	**1.000**	**0.000**	**1.000**	**1.000**	0.045	**1.000**	0.955
Total	0.033	0.963	0.848	0.022	0.984	0.958	0.016	0.984	0.980	**0.012**	**0.988**	**0.988**

Results with IR 1/5

The next step is to perform a new execution with a lower IR in concrete, reducing the IR to 1/5. To do this, the first step is to start from the initial dataset that has 11 910 images, and then perform a random deletion of 50% of examples of all major classes. This way we get a dataset with 6 510 images. Once this is done only 2 700 images should be selected, so that all experiments have a similar number.

In Table 1 can be seen as the number of examples per class for this experiment and can be checked that the IR is 1/5. The results obtained from the CNN with this new distribution of examples in the dataset can be seen in Tables 2 and 3.

These results obtained with an IR 1/5 show a decrease of the Error Rate with respect to the experimentation realized with a IR 1/10. The Error value obtained is 0.022, since 15 images of a total of 670 are classified badly. Also, the results show an increase in both Accuracy and Recall. The global Accuracy value obtained is 0.984 and the global Recall value is 0.958. These results reinforce the initial idea, so the next step is to continue to reduce the IR to verify if the improvement is broadened.

Results with IR 1/3

In order to continue to verify the initial hypothesis, the following experimentation focuses on reducing the IR to 1/3. Therefore, starting from the initial dataset with 11 910 images, a random selection of 30% of the examples of each of the majority classes is performed, obtaining a dataset with 4 350 images. Once this is done, and in order to perform all executions with the same number of examples, 2 700 images of that set are selected.

In Table 1, it can be verified that the distribution of examples per class for this experiment has an IR of 1/3, since a greater reduction of the examples of the majority classes has been performed. From this new distribution of the dataset examples, the results shown in the Tables 2 and 3 are obtained using a CNN.

Observing the results for this experiment with an IR value of 1/3, it can be seen that the results continue to improve previous experiments. In this case, a global Error of 0.016 is obtained, since 11 images of a total of 670 are classified badly. Also, the global Accuracy is 0.984 and the global Recall is 0.980. It can be seen that the trend continues to confirm that as the IR of the dataset decreases the classification results obtained through CNN improve. This fact confirms the initial hypothesis, and therefore, it gives rise to a last experiment in which the dataset is completely balanced.

Results with IR 1/1

The objective of the last test is, as it has been mentioned before, to verify the behavior of the CNN starting from a balanced dataset with IR 1/1. Therefore, the first step is to balance the initial dataset of 11 910 images. In order to do so the class with the fewest number of examples is selected and randomly delete examples of the rest of the classes until they match them. In this way, in order to perform the experimentation a dataset with 2 700 images is obtained with the distribution by class that is shown in Table 1.

Tables 2 and 3 show the results obtained using a CNN with the balanced dataset. In this case, the percentage of global Error is 0.012, since 8 images of a total of 670 are classified badly, the global Accuracy is 0.988 and the global Recall is 0.988. These results improve again to those obtained in previous experiments, which confirms the initial hypothesis, since as the imbalance decreases, better results are obtained.

Results Discussion

In the previous subsections, the initial hypothesis established in this study has been confirmed. Next, a visual representation of the results is shown through the Figs. 3 and 4 and a discussion of these results is made.

On the one hand, Fig. 3 shows the Error obtained by class for each of the experiments performed, it can be seen as in all cases, except class 3 and 9, the results obtained with the balanced dataset are the best. On the other hand, Fig. 4 shows the mean Error for each experiment, it can be seen that as IR decreases, better results are obtained.

Analyzing the presented results, different conclusions can be drawn. From the point of view of the overall results, it can be seen that the reduction of the degree of imbalance causes better results to be obtained when classifying with CNN, which confirms the initial hypothesis proposed in this study. If a class-to-class analysis is conducted, it can be seen that the best results are obtained with the balanced dataset for most classes. There are 2 exceptions, meaning classes 3 and 9, where the imbalanced dataset obtains better results. The explanation for this is that both classes are majority classes, so the imbalanced dataset has a greater number of examples of them than the balanced dataset, a fact that affects the classification.

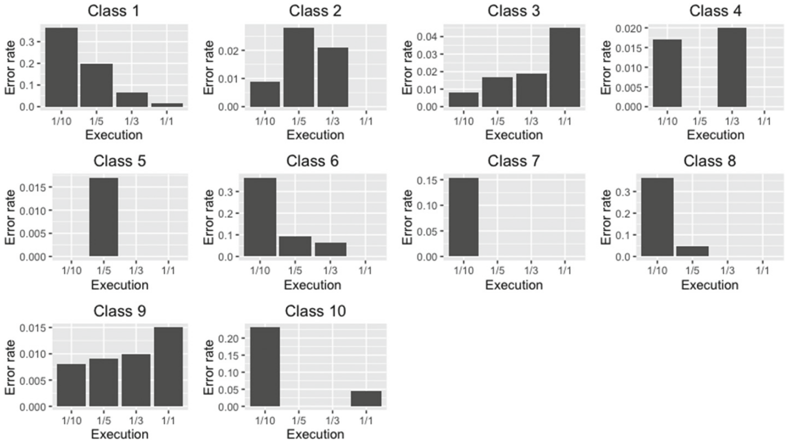


Fig. 3. Error rate for class and execution.

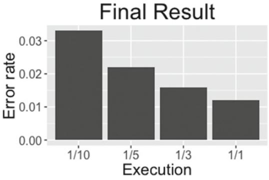


Fig. 4. Average error rate for execution.

Once the results are seen, it can be said that one of the aspects that could most affect the classification is the calculation of weights of the last layer of the network, the fully connected layer that determines the class in a last supervised phase, since the weights obtained could prioritize the majority classes with respect to the minority classes.

Another characteristic of the CNN that could influence the classification is the calculation of the weights corresponding to the different filters in the convolutional layers. These filters move along the different inputs, modifying the weights throughout the process, so that these weights could be overly adapted to the majority classes in cases where there is a greater imbalance. Both this conclusion and the previous one open new avenues of study, with the aim of verifying whether they are fulfilled by a more detailed study.

6 Conclusions

One of the problems that usually arises when trying to tackle the task of classification based on real data is that the data are not balanced, which negatively affects the results obtained with most of the classification models. In this study, a series of tests have been carried out with the aim of demonstrating that this problem also affects the classification by means of CNN.

Tests have shown how, as the imbalance between the data used to classify is minimized, the results obtained through CNN are better. Thus, the hypothesis initially established is fulfilled. These results imply that the distribution of the data must be taken into account when using this type of technique to classify images, since an excessive imbalance can negatively influence the results obtained.

The results derived from this study open up new possibilities for future work. A first approximation to the solution of the problem derived from the use of data intrinsically imbalanced with CNN is the application of classical methods: resampling techniques, cost-sensitive learning or ensembles methods. These techniques aim to reduce the effects of imbalance on the classification algorithms applied later. There is also the possibility of creating new models that combine traditional techniques that face the imbalance with CNN to perform the classification.

However, this is a first approximation to the problem performed with a particular dataset and a given technique, and this fact must be analyzed in more detail in future work.

Acknowledgments. The work of F. Pulgar was supported by the University of Jaén under the Action 15: Predoctoral aids for the encouragement of the doctorate. This work was partially supported by the Spanish Ministry of Science and Technology under project TIN2015-68454-R.

References

1. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. Wiley, USA (2000)
2. Kotsiantis, S.: Supervised machine learning: a review of classification techniques. In: Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, pp. 3–24 (2007)
3. Chawla, N.V., Japkowicz, N., Kotcz, A.: Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor. **6**(1), 1–6 (2004)
4. He, H., García, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
5. Sun, Y., Wong, A.K.C., Kamel, M.S.: Classification of imbalanced data: a review. Int. J. Pattern Recogn. Artif. Intell. **23**(4), 687–719 (2009)
6. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: A review on ensembles for class imbalance problem: bagging, boosting and hybrid based approaches. IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. **42**(4), 463–484 (2012)
7. Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.: A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explor. **6**(1), 20–29 (2004)
8. Zadrozny, B., Elkan, C.: Learning and making decisions when costs and probabilities are both unknown. In: Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining (KDD01), pp. 204–213 (2001)
9. Zadrozny, B., Langford, J., Abe, N.: Costsensitive learning by costproportionate example weighting. In: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM03), pp. 435–442 (2003)

10. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEE Trans. Pattern Anal. Mach. Intell. **3**(8), 1798–1828 (2013)
11. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning (2016)
12. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: Man vs. Computer. Neural Netw. **32**, 323–332 (2012)
13. García, V., Sánchez, J.S., Mollineda, R.A.: On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowl. Based Syst. **25**(1), 13–21 (2012)
14. Orriols-Puig, A., Bernad-Mansilla, E.: Evolutionary rule-based systems for imbalanced datasets. Soft Comput. **13**(3), 213–225 (2009)
15. Ciresan, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification, Technical Report No. IDSIA-04-12 (2012)
16. McMillan, R.L.: How Skype used AI to build its amazing new language translator, wire (2014)
17. LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time-series. In: Arbib, M.A. (ed.) The Handbook of Brain Theory and Neural Networks (1995)
18. Sak, H., Senior, A., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: Proceedings of Interspeech, pp. 338–342 (2013)
19. Sermanet, P., LeCun, Y.: Traffic sign recognition with multi-scale convolutional networks. In: Proceedings of International Joint Conference on Neural Networks (2011)
20. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on in Circuits and Systems (ISCAS), pp. 253–256 (2010)