2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

Convolutional Neural Networks Learning from Respiratory data

Diego Perna
DIMES
University of Calabria
Rende(CS), Italy
d.perna@dimes.unical.it

***Abstract*—Convolutional neural networks (CNNs) have been successfully applied in a wide variety of fields, from image processing to genomic sequencing. In the context of biomedical data, we focus our attention on respiratory data, complex signals characterized by a high level of information richness and potential indicators of several common medical conditions. In this respect, we address the problem of identifying unhealthy indicators in respiratory sound data through the application of novel CNNs architecture and the extraction of Mel Frequency Cepstral Coefficients (MFCC), in order to unveil unhealthy indicators and provide doctors with a potentially life-saving tool.**

***Index Terms*—Deep learning, Neural networks, respiratory data, MFCC**

I. Introduction

Convolutional neural networks (CNNs) represent a class of deep, feed-forward artificial neural networks, usually applied, but not limited to, 2d images. CNNs are inspired by biological processes [1]; more precisely they mimic the visual cortex, a region of the brain in charge of elaborating electrochemical signals generated by the photo-receptors that constitute the cornea (i.e., our receptive field). Differently from other neural network architectures, CNNs are characterized to be shift-invariant or space invariant, based on their shared-weights architecture and translation invariance [2], [3].

In recent decades, thanks to advancements in high-throughput technologies, a deluge of biological and medical data has been and is being generated. In the field of bio-medicine, successful application of CNNs is mainly focused on problems such as medical image classification and segmentation [4]–[6] applied to heart left ventricle [7], [8], brain tumor [9], [10], pancreas [11] and prostate [12]. Moreover CNNs are also utilized to tackle problems such as genomic sequencing, gene expression analysis [13], [14] (RNA and DNA), and protein structure analysis [15]–[19].

In this work, we explore the use of convolutional neural networks for the detection of serious and less-severe diseases through the analysis of respiratory data. Starting with a brief introduction to CNNs and respiratory data, we then describe preprocessing steps, setting and regularization techniques that help us to apply CNNs to the previously mentioned problem.

II. Methodology

Convolutional neural networks are usually considered to be a deep architecture and they consist of an input and an output layer, as well as multiple hidden layers. While several different arrangements can be made, CNNs' building blocks can be identified in four type layers, namely: convolutional layer, pooling layer, fully connected layer, and normalization layer.

In the context of image classification, CNNs use relatively little preprocessing compared to other image classification algorithms. Convolutional-based architectures can also easily learn filters that in traditional algorithms are usually hand-engineered, relieving developers and engineer and at the same time limiting human errors. This highlights the benefit of using such architectures when dealing with complex data, limiting the need for prior knowledge and human effort in feature design. Conversely, when dealing with data of a different nature than images (i.e., data with a native structure different than 2D), a certain amount of preprocessing is needed in order to adapt the input data to the representation that better suits the needs of the network.

A. Preprocessing

1) Respiratory data: A single-channel respiratory sound, like the one shown in Figure 1, is a cycle of four main components, two pauses, and two distinctive patterns. Depending on when we start recording, the first part of the cycle can be the inspiratory phase or the expiratory phase. Discarding fine grain variations, mostly due to the conversion of air vibrations to electrical signal, we can summarize the cycle in the following way. Conventionally, we start the respiratory cycle from the inspiratory phase (i.e., air inflow), that is characterized by a lower amplitude and a regular pattern, whereas the expiratory component (i.e., air outflow) shows one or multiple peaks, a decreasing amplitude pattern, and it is usually characterized by a higher average energy.

In this work we base our experimental evaluation on the ICBHI 2017 Challenge dataset [20], this respiratory sound database was originally compiled to support the scientific challenge organized by the Int. Conf. on Biomedical Health Informatics - ICBHI 2017. The database consists of a total of 5.5 hours of recordings containing 6898 respiratory cycles, of which 1864 contain crackles, 886 contain wheezes, and

978-1-5386-5488-0/18/$31.00 ©2018 IEEE

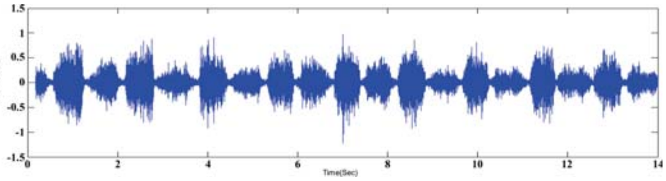

Fig. 1. Example respiratory cycle waveform of a healthy patient.

506 contain both crackles and wheezes, in 920 annotated audio samples from 126 subjects. The cycles were annotated by respiratory experts and the annotations state the presence of crackles, wheezes, a combination of them, or no adventitious respiratory sounds. More in detail, the annotation style format includes the beginning of the respiratory cycle(s), as well as the end of the respiratory cycle(s), the presence or absence of crackles (presence=1, absence=0), and finally presence or absence of wheezes (presence=1, absence=0). The recordings were collected using heterogeneous equipment and their duration ranged from 10s to 90s. Sensory equipment from which the recordings were acquired is also provided, and one of the following locations characterize each sample: trachea (Tc), anterior left (Al), anterior right (Ar), posterior left (Pl), posterior right (Pr), lateral left (Ll), and lateral right (Lr). Furthermore, to increase the difficulty of the current identification task the noise levels in some respiration cycles emulate real-life conditions.

2) Feature extraction: The preprocessing step in charge of extracting the features of the audio signal relies on Mel Frequency Cepstral Coefficients (MFCCs) [21]. MFCCs have been the dominant features in the field of speech recognition and their success is due mostly to their ability to represent the speech amplitude spectrum in a compact form, and also thanks to the perceptual and computational considerations that are at the basis of the extraction of MFCCs coefficients.

The first step of this procedure is to divide the input signal into frames of equal length, usually applying a windowed function at fixed intervals. Then we generate a cepstral feature vector for each frame and apply the Direct Fourier Transform (DFT) on each frame. While the amplitude spectrum is retained, we discard the information about the phase of the signal. To mimic the way how the human brain perceives the loudness of a sound, we apply the logarithm function to the amplitude spectrum values. The next step is to smooth the spectrum and emphasize perceptually meaningful frequencies by aggregating the spectral component into a lower number of frequency bins. The last step of the procedure consists of applying the Discrete Cosine Transform (DCT) to reduce the number of parameters in the system, given the high correlation that usually characterizes the Mel-spectral vectors.

B. Neural network architecture

Convolutional neural networks represent a class of feed-forward artificial neural networks, usually applied 2d images. In the following paragraphs, we describe the building blocks of convolutional neural networks, including the different types of layer, activation functions, and regularization techniques.

a) Convolutional layer: Convolution is one of the most important operations in signal and image processing, it can be applied to one-dimensional (e.g. speech processing), two-dimensional (e.g. image processing) or three-dimensional (video processing) data. Although the natural form of sound data is one-dimensional, in this work we aim to analyze respiratory data in two-dimensional shape. In the context of deep learning, an image is usually considered as a matrix of pixels; in this work, we consider a respiratory sound as a matrix of Mel Frequency Cepstral Coefficients. Convolutional layers can be regarded as applying a moving filter onto a two-dimensional matrix, moving the filter over the matrix guarantee the benefit of re-using the learned weight, resulting in a more efficient architecture w.r.t. the number of weights being used in function of the size of input data.

b) Pooling layer: Pooling layers combine the outputs of a cluster of neurons at one layer into a single neuron in the next layer. Several functions can be used to select and propagate signals through the network, such as max or average pooling. Max pooling select only the maximum value from the cluster of neurons at the prior layer. Whereas average pooling propagates the average of the cluster of neurons values at the prior layer.

c) Fully-connected layer: A less efficient way of processing information coming from other layers, but often needed in some precise point of the architecture is represented by fully-connected layer. This layer connects every neuron in one layer to every neuron in another layer, and in principle is equal to the traditional multi-layer perceptron neural network (MLP).

d) Activation functions: Another key aspect of neural networks, which affect the performance of the network especially in terms of back-propagation, is represented by the choice of the activation function for each type of layer. Activation functions can be divided into two main categories: linear and non-linear activation functions.

Tanh (or hyperbolic tangent) activation function belongs to the family of sigmoid functions and it is defined in the range $\in [-1, 1]$, which makes it suitable especially for binary classification problems. Rectified Linear Unit (ReLU) applies the non-saturating activation function $max(0, x)$ and is often preferred to other functions because it trains the neural network faster and without penalizing generalization accuracy. The use of ReLU also increases the nonlinear properties of the decision function and of the overall network without affecting the receptive fields of the convolution layer.

The sigmoid (also known as logistic) function is a non-linear, monotonic, and differentiable functions. It is defined in the range $\in [0, 1]$ and its use it is mainly due to this characteristic. However, it is one of the functions that is affected by the vanishing gradient problem: the value of the gradient of this function tends to zero at both fringes, limiting the learning capability of the neural network.

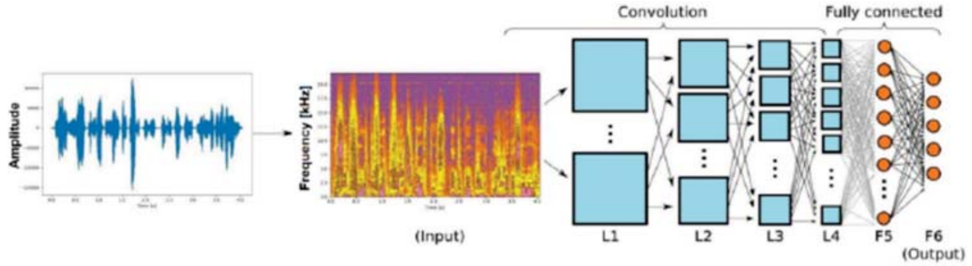

Fig. 2. Overview of the whole architecture: firstly the respiratory sound data are transformed into the frequency domain, then the CNN is applied to the input data to learn how to distinguish healthy and unhealthy patient.

Conversely, the softmax function transforms a K-dimensional vector z of real values to a K-dimensional vector $\sigma(z)$ of real values in the range $(0, 1)$, and in addition, all the entries sum up to 1. In practice, the output of the softmax function is usually used to represent a probability distribution over K different possible outcomes, and it represents a more general logistic function.

$$\sigma(\mathbf{z})_j = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}} \qquad \forall j = 1, ..., K. \tag{1}$$

C. Implementation details

a) Regularization techniques: In order to reduce overfitting, we use a fairly recent technique that aims to prevent complex co-adaptations to training instances. Dropout [22], as the name suggests, consists of dropping out units, both visible and hidden, in a neural network, in order to limit the connection between the learned weights and the training data.

We also applied elastic net regularization [23], a regression method that linearly combines the L1 and L2 penalties of the lasso and ridge methods into the loss function in order to make the learned model more robust.

b) Optimization algorithm: For what concerns the optimization technique we rely on ADAM [24], a simple and computationally efficient algorithm for gradient-based optimization of stochastic objective functions. This method is particularly useful when dealing with large datasets or high-dimensional parameter space. The method combines the advantages of two recently popular optimization methods: the ability of AdaGrad to deal with sparse gradients, and the ability of RMSProp to deal with non-stationary objectives.

c) Class imbalance: Often, real-world datasets present certain characteristics that can affect prediction accuracy and validity. One common feature is *class imbalance*, i.e., the difference in the number of positive and negative examples in the case of a binary problem, or more in general, the problem exists when one or more classes are dominant with respect to each other. Here we try to overcome the issue of class imbalance through two different techniques, namely SMOTE and RUS. The former is an over-sampling technique that aims to increase the number of under-represented classes through the generation of new artificial samples through interpolation techniques. Conversely, the latter approach belongs to the under-sampling group of methods and pursue the goal of decrease the number of examples associated with the most represented classes, through random deletion.

To be able to deal with more than two classes, the model has been implemented with the multi-class version of the well-known cross-entropy loss function. The cross-entropy between two probability distributions p and q over the same underlying set of events measures the average number of bits needed to identify an event drawn from the set, if a coding scheme is used that is optimized for an ”artificial” probability distribution q, rather than the “true” distribution p.

In the end, the proposed neural network is composed of two convolutional layers using both max-pooling and dropout. The two convolutional layers are followed by two dense layers, and then a final layer based on the softmax activation function.

III. Evaluation

We divided the previously described dataset into an 80% train and 20% testing set, and we devised two testbeds: the first one is a simplified classification task with only two classes, namely healthy and unhealthy, where all the diseases have been grouped into a single class. In the second, more challenging, testbed we try to train a neural network to distinguish between chronic and non-chronic respiratory diseases; to this purpose we devised three classes: healthy, chronic and non-chronic diseases.

In this preliminary experimental evaluation, to evaluate the performance of the two convolutional-based classifiers we resort to from well-known evaluation measures, namely: precision, recall, and f1-score.

IV. Experimental results

We tested the proposed neural network in two different scenarios, binary and ternary classification, and with different settings, i.e., with and without the use of regularization techniques, i.e., elastic net and dropout.

Table I reports overall performance for both scenarios and both settings. The neural network has been set up to use the ReLU activation function in the intermediate layers and the softmax activation function in the final layer, and to tackle the problem of class-imbalance, the SMOTE over-sampling technique.

In both the binary and ternary classification tasks, performance results justify the use of dropout and elastic net, since best values are achieved with the use of both. Comparing

#classes	Method	Loss	Accuracy	Precision	Recall	F1-score
2	Without regularization	0.62	0.78	0.91	0.80	0.85
2	With regularization	0.66	0.83	0.96	0.83	0.88
3	Without regularization	0.55	0.76	0.82	0.79	0.80
3	With regularization	0.56	0.82	0.87	0.82	0.84

TABLE I
NEURAL NETWORK PERFORMANCE IN BINARY AND TERNARY CLASSIFICATION TASK, WITH AND WITHOUT THE USE OF REGULARIZATION TECHNIQUES.

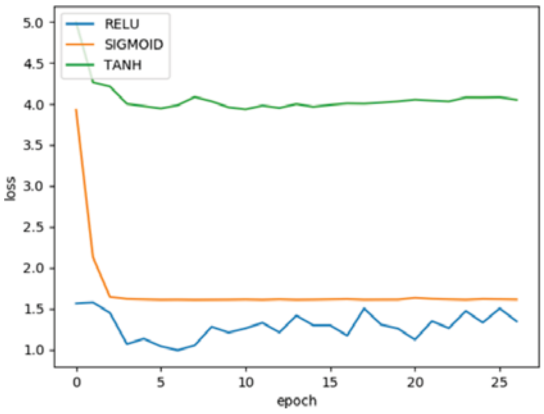

Fig. 3. Loss performance comparison of three activation function.

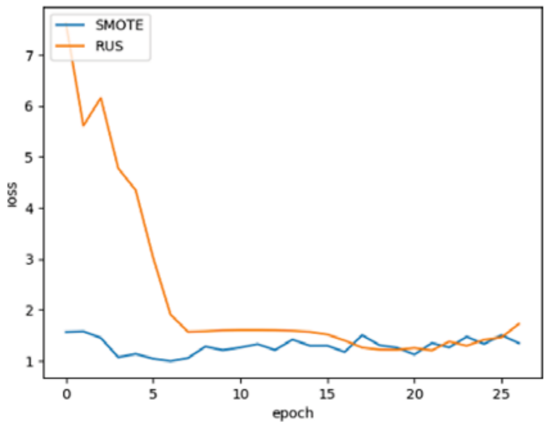

Fig. 4. Loss performance comparison using over- (SMOTE) and under-sampling (RUS) technique.

the two types of task, we observe a marginal deterioration of performance in correspondence of the more challenging ternary classification problem, probably due to the lack of enough instances of chronic or non-chronic diseases, or due to the inherent diversity of the diseases in terms of distinctive patterns and features. Overall, despite the complexity of the task at hand, the relatively small size of the neural network and of the dataset, satisfying results have been achieved in both binary and ternary tasks.

In addition to the above performance evaluation, we show results of the convolutional neural network with different settings, to highlight the benefits of the approach in use. Figure 3 depicts loss function values with increasing number of epochs with the use of three activation functions in the intermediate layers. The hyperbolic tangent function offers poor results in comparison to the other two functions and shows the inability to learn, fast enough, the appropriate weights to minimize the loss function. Sigmoid and ReLU show similar results, even if the latter is characterized by the ability to promptly reduce the loss function from the early epochs and also to maintain an advantage in term of performance.

We further compared the performance of the proposed neural network with the use of both under- and over-sampling techniques. While the final results are similar in value, both the quality of the weights learned and the number of epochs required to achieve these results put in clear advantage the over-sampling approach SMOTE. This is justified by the fact that the dataset is affected by class-imbalance in a severe manner. Thus, the use of the under-sampling technique means to limit the number of instances of the under-represented classes to only a few samples.

V. Conclusion

In this preliminary work, we explored the application of convolutional neural network to respiratory data for the detection of chronic and non-chronic diseases. We described the network architecture as well as the crucial preprocessing phase. The achieved performance results suggest that convolutional neural networks are a viable tool for the detection of specific features in respiratory data, and more in general sounds, that characterize chronic and non-chronic diseases.

Acknowledgements

This research work has been partly supported by the Start-(H)Open POR Grant No. J28C17000380006 funded by the Calabria Region Administration, and by the NextShop PON Grant No. F/050374/01-03/X32 funded by the Italian Ministry for Economic Development

References

[1] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, “Subject independent facial expression recognition with robust face detection using a convolutional neural network,” *Neural Networks*, vol. 16, no. 5-6, pp. 555–559, 2003.

[2] W. Zhang, “Shift-invariant pattern recognition neural network and its optical architecture,” in *Proceedings of annual conference of the Japan Society of Applied Physics*, 1988.

[3] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, “Parallel distributed processing model with local space-invariant interconnections and its optical architecture,” *Appl. Opt.*, vol. 29, no. 32, pp. 4790–4797, Nov 1990.

[4] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, “Medical image classification with convolutional neural network,” in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014, pp. 844–848.

[5] E. Vocaturo and P. Veltri, “On the use of networks in biomedicine,” in *14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) / 12th International Conference on Future Networks and Communications (FNC 2017) / Affiliated Workshops, July 24-26, 2017, Leuven, Belgium*, 2017, pp. 498–503.

[6] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[7] G. Luo, S. Dong, K. Wang, and H. Zhang, “Cardiac left ventricular volumes prediction method based on atlas location and deep learning,” in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec 2016, pp. 1604–1610.

[8] T. A. Ngo, Z. Lu, and G. Carneiro, “Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance,” *Medical Image Analysis*, vol. 35, pp. 159 – 171, 2017.

[9] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in mri images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[10] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. Jodoin, and H. Larochelle, “Brain tumor segmentation with deep neural networks,” *Medical Image Analysis*, vol. 35, pp. 18 – 31, 2017.

[11] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers, “Deep convolutional networks for pancreas segmentation in CT imaging,” in *Medical Imaging 2015: Image Processing, Orlando, Florida, USA, February 24-26, 2015*, 2015, p. 94131G.

[12] S. Liao, Y. Gao, A. Oto, and D. Shen, “Representation learning: A unified deep learning framework for automatic prostate mr segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Springer Berlin Heidelberg, 2013, pp. 254–261.

[13] C. Angermueller, H. J. Lee, W. Reik, and O. Stegle, “Deepcpg: accurate prediction of single-cell dna methylation states using deep learning,” *Genome biology*, vol. 18, no. 1, p. 67, 2017.

[14] G. Aoki and Y. Sakakibara, “Convolutional neural networks for classification of alignments of non-coding rna sequences,” *Bioinformatics*, vol. 34, no. 13, pp. i237–i244, 2018.

[15] S. Wang, S. Weng, J. Ma, and Q. Tang, “Deepcnf-d: predicting protein order/disorder regions by weighted deep convolutional neural fields,” *International Journal of Molecular Sciences*, vol. 16, no. 8, pp. 17 315–17 330, 2015.

[16] L. Yang and S. Tetsuo, “Malphite: a convolutional neural network and ensemble learning based protein secondary structure predictor,” in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1260–1266.

[17] Z. Lin, J. Lanchantin, and Y. Qi, “Must-cnn: A multilayer shift-and-stitch deep convolutional architecture for sequence-based protein structure prediction,” in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[18] J. Zhou and O. Troyanskaya, “Deep supervised and convolutional generative stochastic network for protein secondary structure prediction,” in *International Conference on Machine Learning*, 2014, pp. 745–753.

[19] W. Sheng, L. Wei, L. Shiwang, and X. Jinbo, “Raptorx-property: a web server for protein structure property prediction,” *Nucleic acids research*, vol. 44, no. W1, pp. W430–W435, 2016.

[20] B. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques *et al.*, “A respiratory sound database for the development of automated classification,” in *Precision Medicine Powered by pHealth and Connected Health*. Springer, 2018, pp. 33–37.

[21] M. H. Shirali-Shahreza and S. Shirali-Shahreza, “Effect of mfcc normalization on vector quantization based speaker identification,” in *Signal processing and information technology (isspit), 2010 ieee international symposium on*. IEEE, 2010, pp. 250–253.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[23] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[24] D. Kinga and J. B. Adam, “A method for stochastic optimization,” in *International Conference on Learning Representations (ICLR)*, vol. 5, 2015.