

Understanding Automatic Diagnosis and Classification Processes with Data Visualization

Pierangela Bruno

*Department of Mathematics and Computer Science
University of Calabria
Rende, Italy
bruno@mat.unical.it*

Francesco Calimeri

*Department of Mathematics and Computer Science
University of Calabria
Rende, Italy
calimeri@mat.unical.it*

Alexandre Sébastien Kitanidis

*Department of Electronic, Information and Bioengineering
Politecnico di Milano
Milano, Italy
alexandresebastien.kitanidis@mail.polimi.it*

Elena De Momi

*Department of Electronic, Information and Bioengineering
Politecnico di Milano
Milano, Italy
elena.demomi@polimi.it*

Abstract—Providing accurate diagnosis of diseases generally requires complex analyses of many clinical, biological and pathological variables. In this context, solutions based on machine learning techniques achieved relevant results in specific disease detection and classification, and can hence provide significant clinical decision support. However, such approaches suffer from the lack of proper means for interpreting the choices made by the models, especially in case of deep-learning ones.

In order to improve interpretability and explainability in the process of making qualified decisions, we designed a system that allows for a partial opening of this black box by means of proper investigations on the rationale behind the decisions; this can provide improved understandings into which pre-processing steps are crucial for better performance.

We tested our approach over artificial neural networks trained for automatic medical diagnosis based on high-dimensional gene expression and clinical data. Our tool analyzed the internal processes performed by the networks during the classification tasks in order to identify the most important elements involved in the training process that influence the network's decisions.

We report the results of an experimental analysis aimed at assessing the viability of the proposed approach.

Index Terms—GradCAM, Heatmap, Hot-spot map, Convolutional Neural Networks

I. INTRODUCTION

Interpreting the decision-making processes of neural networks can be of great help at enhancing the diagnostic capabilities and providing direct patient- and process-specific support to diagnosis and disease classification. However, interpretability and explainability represent critical points for approaches based on deep learning models, that achieved great results in disease classification. In this context, the definition of proper models summarizing the mechanisms steering neural networks decisions can be crucial for an effective examination of the causes of the resulting choices.

In this context, we propose a novel method based on heatmaps and hot-spot maps for analyzing the internal pro-

cesses performed by a network during the training for a classification task; the aim is to identify the most important elements that will influence the network's decisions. In particular, we use gradient visualization techniques to produce a coarse localization map highlighting the image regions most likely to be referred to when taking the classification decision; the identified areas are then removed from the images in order to check whether the classification performance changes.

We evaluate the proposed approach over different “instances”, such as gene expression or other clinical data. Gene expression represents the amount of RNA produced in a cell under different biological states and clinical data provide both health-related information associated with patient care and features related to disease that are relevant and helpful to diagnosis classification, and even, early prediction process. Both instances consist of a set of “attributes” characterizing each patient. The attributes can be defined as: (1) proportion of genes in each gene expression and (2) clinical information related to treatment, pathology and patient characteristics.

We perform pre-processing operations (i.e., dimensionality reduction) and generate a set of 2-D images representing the instances of all patients that suffer from a specific pathology; in particular, we generate heatmaps and hot-spot maps to show features proportion level and provide spatial information among each variable, respectively.

The herein proposed method is based on a hybrid approach that relies on data visualization techniques and visual explanations for classification via a model based on Convolutional Neural Networks (CNNs); this lead to more transparent and explainable result process achievements.

II. RELATED WORK

Several methods were proposed to perform automatic diagnosis based on gene expression or clinical data. For instance, Support Vector Machines with Squared Loss (SVMSL) [1] were applied to perform cancer classification. The 1-norm

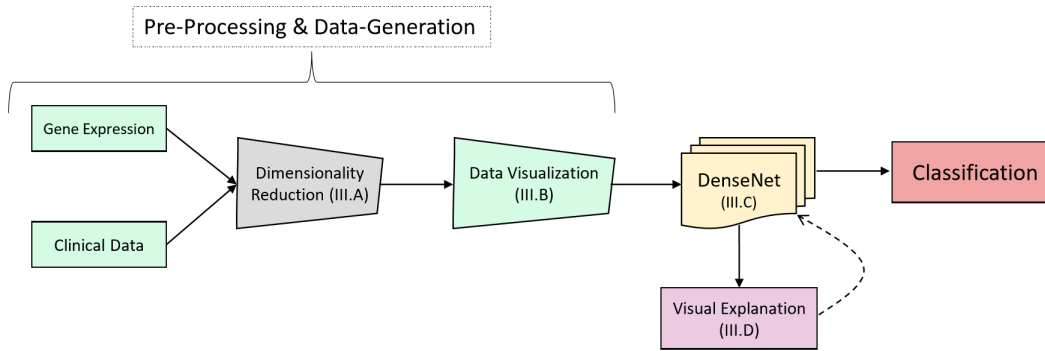


Fig. 1. Workflow of the proposed framework. Dimensionality reduction is performed on Gene Expression or Clinical Data. The reduced dataset is transformed into images and diagnosis classification is performed using DenseNet 169. GradCAM and Guided GradCAM are generated during CNN training. In brackets the manuscript section.

SVMSEL is first used to select “useful” genes and then to classify the resulting dataset. This method performs a very fast gene selection compared with other methods, but the classification performance is obtained using approximation methods that could affect the overall performance. Rotation Forest based-Genetic algorithms (GAs) were used to perform breast cancer classification based on clinical data [2] featuring several attributes such as “clump thickness”, “cell size”, “cell shape”, etc. GA is employed as a feature reduction mechanism to remove redundant data; the best feature subset is classified via Rotation Forest techniques. The approach achieves an accuracy mean value higher than 98%.

As for classification methods, deep-learning-based models recently achieved promising results. Deep learning models, such as Convolutional Neural Networks (CNN) [3], are proven to be appropriate and effective when compared to conventional methods. Interestingly, CNNs currently represent the most widely used method for image processing; nevertheless, the application of deep learning methods is not common in the context of gene expression, due to the well-known “large p , small n ” problem [4], where p refers to the number of features and n refers to the number of samples.

In the last year, understanding neural network decision-making was a topic of research. Several approaches were proposed to visualize the behavior of a CNN by sampling image patches that maximize the activation of hidden units [5], and by backpropagation to identify or generate salient image features [6]. Other researchers were trying to solve this problem by explaining neural network decisions by generating informative heatmaps such as GradCAM [7], or through layer-wise relevance propagation [8]. These methods present some limitations; indeed, the generated heatmaps were qualitative and not informative enough to specify which concepts have been detected. An improvement was provided using semantically explanation from visual representation [9] to decompose the evidence for a prediction for image classification into semantically interpretable components, each with an identified purpose, a heatmap, and a ranked contribution.

In this work, we propose a novel approach for analyzing the internal processes and the decision performed by a neural

network during the training phase. Our approach relies on data visualization techniques to represent gene expression or clinical data of all patients that suffer from a specific pathology. Data visualization represents an advantage w.r.t the state-of-the-art methods as it is able to improve diagnosis classification and, also, provides a better interpretability and explainability of the rationale behind the decisions performed.

The remainder of the paper is structured as follows. In Section III we provide a detailed description of our approach, that has been assessed via a careful experimental activity, reported in Section IV; we analyze and discuss results in Section V, eventually drawing our conclusions in Section VI.

III. PROPOSED APPROACH

The herein proposed approach is illustrated in Fig. 1:

- 1) Data Reduction using Principal Component Analysis (PCA), in order to reduce the dataset dimension;
- 2) Data Visualization for each patient, through:
 - a heatmap: visualize proportion level of each attribute;
 - a hot-spot map, based on Getis-Ord G_i^* , to visualize spatial information level of each attribute;
- 3) Disease Classification using CNNs;
- 4) Visual Explanations using Gradient-weighted Class Activation Mapping (GradCAM) to indicate the discriminative image regions used by the CNN;
- 5) Verification of patients clusterization by survival analysis.

A. Data Reduction

PCA is used to reduce data from the high p -dimensional variable space to a K -dimensional variable component space, with less than the number of samples n . We selected a cut-off value of 75% of explained variance by the remaining component as it appeared to be the best compromise between data readability and quality.

B. Data Visualizzazione

For each patient, we created heatmaps and hot-spot maps that are graphical representations of data where the individual values contained in a matrix are represented as colors. *Heatmaps* are directly generated from quantitative differences

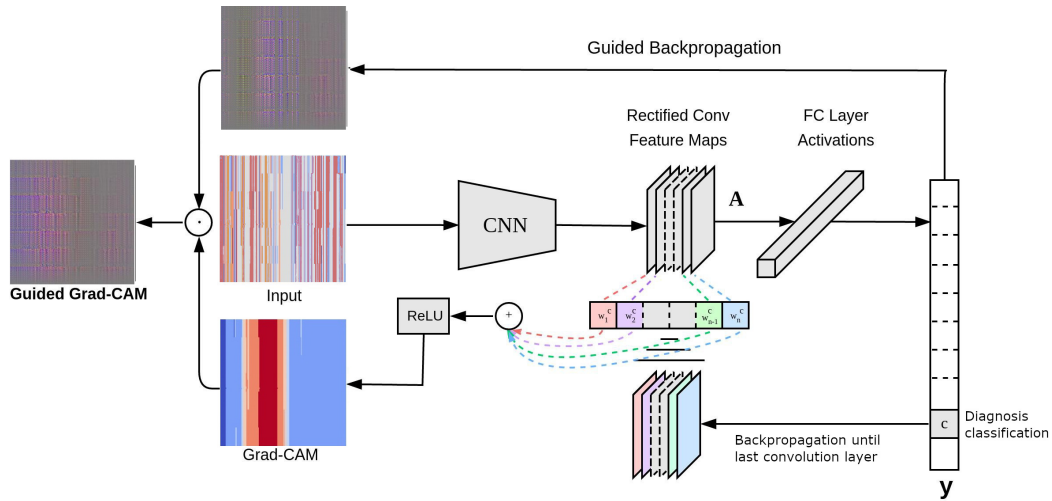


Fig. 2. An example of GradCAM structure. Given an image, and a category (“Diagnosis c ”) as input, we forward propagate the image through the model to obtain the raw class scores before softmax. The gradients are set to zero for all classes except the desired class (Diagnosis c), which is set to 1. This signal is then backpropagated to the rectified convolutional feature map (A) of interest, where we can compute the coarse GradCAM localization (blue heatmap). Finally, we pointwise multiply the heatmap with guided backpropagation to get Guided GradCAM visualizations which are both high-resolution and class-discriminative.

TABLE I

DATASET DESCRIPTION. FOR EACH DATASET, ON THE COLUMNS WE REPORT THE NUMBER OF PATIENTS, THE NUMBER OF ATTRIBUTES, THE NUMBER OF ATTRIBUTES AFTER PCA, THE TYPOLOGY OF DATASET AND CLASSIFICATION.

DATASET	NUMBER OF PER PATIENTS	NUMBER OF ATTRIBUTES PER PATIENTS	NUMBER OF ATTRIBUTES PER PATIENTS AFTER PCA	ISTANCE	CLASSIFICATION
BREAST CANCER	569	10	8	CLINICAL DATA	BENIGN OR MALIGNANT MASS
MAMMOGRAPHY MASS	961	45	28	CLINICAL DATA	BENIGN OR MALIGNANT MASS
PARKINSON DISEASE	197	754	52	CLINICAL DATA	HEALTHY OR NOT
BREAST OR KIDNEY CANCER	635	9834	112	GENE EXPRESSION	BREAST OR KIDNEY
LYMPHOMA CANCER	62	4026	102	GENE EXPRESSION	HEALTHY OR NOT
BREAST CANCER (BC-TCGA)	590	17814	133	GENE EXPRESSION	HEALTHY OR NOT
BREAST CANCER (GSE2034)	286	12634	127	GENE EXPRESSION	HEALTHY OR NOT

in expression levels, by coding numerical values of different types of data into colors. *Hot-spot maps* are generated using the Getis-Ord G_i^* statistics computed to preserve the spatial information contained in the matrices; for example, the hotspot show the highest concentration of genes by means of darker colors. Getis-Ord G_i^* statistic estimates the density distribution of features at the local level, and measures the degree of spatial association in a whole dataset. The method evaluates the degree to which each feature (i.e., gene) is surrounded by features with similarly high or low values within a specified geographical distance (neighborhood).

C. Classification

We classified the patients using DenseNet 169 [10]. This network is made of dense blocks, where for each layer the inputs are the feature maps of all the previous layers. Networks can be more efficient to be trained if they have shorter connections between layers close to the input and than close to the output. The DenseNet architecture features several advantages, since each layer has direct access to the gradients from the loss function and the original input signal,

the flow of information and gradients, ensuring alleviation of the vanishing gradient problem.

D. Visual Explanations

We used GradCAM to identify visual features in the input able to explain result process achieves during the classification. In particular, it uses the gradient information flowing into the last convolutional layer of the CNN to assign importance values to each neuron. As shown in Fig. 2, in order to obtain the class-discriminative localization map ($L_{GradCAM}^c$), GradCAM computes the gradient y_c (i.e. output score y for class of interest c - instance to be predicted) w.r.t. feature maps activation matrix A^k of a convolutional layer. It performs a weighted combination of forward activation maps A^k , and is followed by ReLU to obtain:

$$L_{GradCAM}^c = ReLU\left(\sum_k a_k^c A^k\right)$$

where a_k^c is a value computed to express the “importance” of feature map A^k for a target class c (i.e. tumor).

We combined GradCAM with existing pixel-space visualizations to create a high-resolution class-discriminative visualization and generated a Guided GradCAM. GradCAM localizes relevant regions on original image but it was not able to highlight fine-grained details like pixel-space gradient visualization. We computed Guided Backpropagation [11] that visualizes gradients w.r.t. the image where negative gradients are suppressed when backpropagating through ReLU layers.

E. Survival Analysis

According to the classification results, we grouped patients of all datasets into two diagnostic groups. For each group we computed the Kaplan-Meier [12] analysis in order to obtain the overall survival using:

- 1) Images generated on original dataset;
- 2) Images generated after removing the highlighted regions identified by GradCAM and Guided-GradCAM.

We compared the survival curves computed on 1 and 2 images to show if there is a statistical difference after removing the highlighted regions and then if the selected elements can be used as the diagnosis markers. The probability of survival is computed as:

$$P(t) = P(s_1 s_2 | s_1)$$

where s_1 is the probability of surviving after the first day and s_2 is the probability of surviving after the second day.

IV. EXPERIMENTAL PROTOCOL

A. Dataset description and training phase

For the experimental analysis we used these datasets:

- the publicly available dataset from the Gene Expression Omnibus (GEO)¹, a database consisting of microarray, next generation sequencing (NGS) and other high-throughput data;
- the dataset from the UC Irvine Machine Learning Repository², a database consisting of heterogeneous set of features about different diseases and different categories of DNA microarray;
- the dataset from The Cancer Genome Atlas (TCGA)³, a publicly available database hosting many types of data including genomic, epigenomic, transcriptomic, and proteomic data.

All dataset used in our experiments are described in Table I. Each dataset is converted into 2-D images according to Section III-B; we split them into training (80%) and testing (20%) sets; the 20% of the training set is used as validation set, in order to monitor the training process and prevent overfitting.

All experiments are performed on a machine equipped with a 12 x86 64 Intel(R) Core(TM) CPUs @3.50GHz, running GNU/Linux Debian 7 and using CUDA compilation tools, release 7.5, V 7.5.17 NVIDIA Corporation GM 204 on GeForce GTX 970.

¹<https://www.ncbi.nlm.nih.gov/geo/>

²<https://archive.ics.uci.edu/ml/index.php>

³<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>

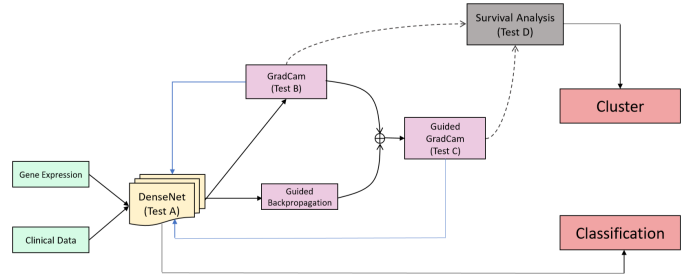


Fig. 3. Workflow of the test description. Each test is performed on heatmap and hot-spot map. Blu arrows indicate network re-training using "new" dataset obtained by removing highlighted elements founded by visual explanations technique.

B. Fine-tuning

For the training phase we performed hyperparameters optimization. DenseNet 169 was trained with both optimizers Adam and SGD and for each optimizer 7 learning rate were tried. The best performance is obtained with the following configuration, trained for 150 epochs: Adam optimizer, learning rate 10^{-5} and batch size 32. The configuration of each network is obtained by hyperparameter tuning. We performed 10-fold cross-validation in order to choose the parameter value that gives the lowest cross-validation average error; experiments were performed on the very same machine with the same configuration of the other approaches.

C. Tests

Figure 3 lists the different approaches that we performed during our experiments. For all the approaches we used both heatmap and hot-spot map as input of CNN after the application of data reduction and data visualization.

- *Test A*: Apply GradCAM to diagnosis on both heatmap and hot-spot map using DenseNet 169.
- *Test B*: Perform diagnosis on a reduced dataset generated by removing the elements corresponding to the highlighted areas identified by Test A.
- *Test C*: Apply Guided-GradCAM to the classification performed by DenseNet, identify a set of the elements corresponding to the highlighted areas and remove from the original image input. Re-train on reduced dataset.
- *Test D*: Perform survival analysis and comparison among survival curves obtained from *Test A-Test B* and *Test C*

D. Performance Metrics

We assessed the effectiveness of our approach using Recall, as in this context, the most important thing is to minimize False Negatives (i.e., disease is present but is not identified).

Recall is computed as $(Rec = \frac{TruePositive}{TruePositive+FalseNegative})$. It considers prediction accuracy among only actual positives and explain how correct our prediction is among ill people. We perform paired t-test [13] among heatmap and hot-spot map in order to check whether the population distributions are similar.

TABLE II

VALIDATION RECALL (AND STANDARD DEVIATION) OF DENSENET 169 AFTER 10-FOLD CROSS-VALIDATION WITH THE DIFFERENT INPUTS, GRADCAM AND GUIDED-GRADCAM APPLIED DURING THE TRAINING, AND DENSENET 169 PERFORMED ON A NEW "REDUCED" DATASET OBTAINED BY REMOVING THE HIGHLIGHTED ELEMENTS SELECTED BY GUIDED-GRADCAM.

	DATASET	DENSENET	DENSENET 169	DENSENET 169
		169 (TEST A)	ON REDUCED DATASET (TEST B)	ON REDUCED DATASET (TEST C)
HEATMAPS	BREAST CANCER	0.9889 (0.0109)	0.9454 (0.0126)	0.9389 (0.0114)
	MAMMOGRAPHY MASS	0.8601 (0.0511)	0.8564 (0.0168)	0.8399 (0.0209)
	PARKINSON DISEASE	0.9886 (0.0113)	0.9394 (0.0169)	0.9372 (0.0099)
	BREAST OR KIDNEY	0.9068 (0.0360)	0.8914 (0.0136)	0.9012 (0.0301)
	LYMPHOMA CANCER	0.9681 (0.0183)	0.9469 (0.0121)	0.9411 (0.0118)
	BREAST CANCER (BC-TCGA)	0.9899 (0.0099)	0.9565 (0.0103)	0.9482 (0.0115)
	BREAST CANCER (GSE2034)	0.9372 (0.0142)	0.8730 (0.0221)	0.8817 (0.0341)
	HOT-SPOT MAPS	BREAST CANCER	0.9899 (0.0106)	0.9416 (0.0102)
MAMMOGRAPHY MASS		0.8991 (0.0211)	0.8514 (0.0141)	0.8402 (0.0123)
PARKINSON DISEASE		0.9949 (0.0051)	0.9414 (0.0149)	0.9399 (0.0181)
BREAST OR KIDNEY		0.9168 (0.0360)	0.8928 (0.0249)	0.9097 (0.0191)
LYMPHOMA CANCER		0.9881 (0.0203)	0.9589 (0.0151)	0.9549 (0.0147)
BREAST CANCER (BC-TCGA)		0.9983 (0.0034)	0.9299 (0.0241)	0.9378 (0.0206)
BREAST CANCER (GSE2034)		0.9345 (0.0204)	0.7000 (0.0371)	0.9091 (0.0141)

We also used *log-rank test* for comparing the survival distributions of two groups. The statistic test z is:

$$z = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

where E_1 and E_2 are the expected number of events in each group while O_1 and O_2 are the total number of observed events in each group. Log-rank tests were computed within a level of significance of 5% [14].

V. RESULTS AND DISCUSSION

Table II reports classification results after 10-fold cross validation for all dataset, using either heatmaps or Getis-Ord G_i^* hot-spot maps. Also it lists the validation results obtained using GradCAM and Guided-GradCAM on original images and images generated after removing the highlighted regions founded by visual explanation technique.

The herein proposed approach achieves promisingly results using both gene expression and clinical data. For instance, taking into account Test A and gene expression data, the results of *DenseNet 169* trained on heatmap show the highest Recall mean value on Breast Cancer (BC-TCGA) (i.e., Rec 0.9899); among clinical datasets, instead, the highest Recall mean value is obtained on Parkinson Disease (i.e., Rec 0.9886) and Breast Cancer (i.e., Rec 0.9889). Furthermore, the evaluation of the prognosis based on *DenseNet 169* trained on hot-spot map shows a general improvement of Recall. Heatmap and hot-spot map have different population distribution. This means that the spatial information generated by Getis-Ord G_i^* statistic is

critical for providing knowledge and improving both precise population information and classification accuracy.

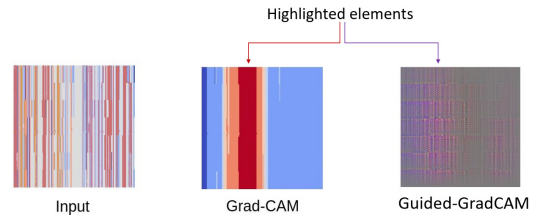


Fig. 4. Example of the markers identification on a heatmap using GradCAM and Guided-GradCAM. Elements highlighted in red for GradCAM and in purple for Guided-GradCAM are considered key features in classification process.

In Test B and Test C we selected and removed the 40% of highlighted elements; this threshold appeared to be the best compromise between data readability and quality. Figure 4 shows an example of the results. Both tests show a worsening of Recall value; DenseNet 169 trained on new heatmaps achieves on the 0.9454 and the 0.9354 of Recall after Test B and Test C, respectively, w.r.t. 0.9889 obtained on the original images. The similar worsening is shown using hot-spot maps, indeed the accuracy does not exceed the 0.96 in all the experiments. In general, a substantial decrease of Recall is shown using both Heatmap and Hot-spot map; this suggests that Test B and Test C are able to identify the important elements involved in the training process and, consequently, responsibility for this diminishment is due to images cutting by which we removed the peculiar characteristic of the disease.

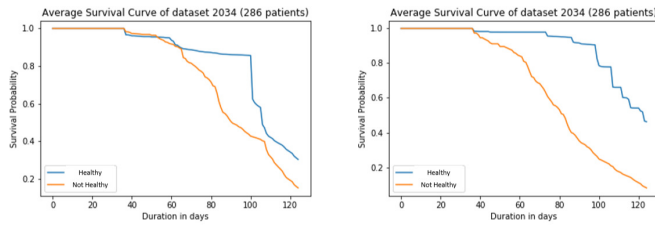


Fig. 5. Plots of Kaplan-Meier product limit estimates of survival of a group of patients (on the left), and after removing genes according to the Test B (on the right) computed on Breast Cancer (GSE2034) dataset.

TABLE III

LOG-RANK TEST COMPUTED ACCORDING BREAST CANCER (GSE2034) DATASET BEFORE (I) AND AFTER Approach C (II) and after Approach D (III) on survival probability (i.e Y axis) (right) and on survival time (i.e. X axis) (left)

LOG-RANK	OBSERVED	CRITICAL VALUE	P-VALUE
I	2.73	3.84	0.17
II	3.98	3.84	0.04
III	4.18	3.84	0.02

Figure 5 shows an example of survival graph computed according to the Test A and Test B, on the left and on the right, respectively.

Table III shows the comparison between log-rank test results obtained from original dataset (I), the resulting dataset after Test B (II) and after Test C (III). In particular, Test B and Test C indicate a significant difference between the population survival curves (p-value 0.0391 and p-value 0.0194); Test A, instead, does not show a significant difference between the two curves (p-value 0.1670).

VI. CONCLUSION

We presented a novel approach for analyzing the internal processes performed by a neural network during the training phase, with the aim of improving explainability in the process of making qualified decisions; more in detail, we try to identify the most important regions that influence the network’s decisions. We tested the proposed approach over DenseNet 169 trained for the task of automatic medical diagnosis based on images representing high-dimensional gene expression and clinical data. The use of images for representing data presents two relevant advantages: first of all, it significantly eases and improves data visualization; furthermore, it allows taking advantage of effective techniques explicitly geared towards image processing in order to perform classification tasks.

Starting from numerical “raw” data, we make use of PCA to reduce the dimensions, getting rid of redundant or irrelevant information and paving the way to a proper 2-D image-based representation.

Experimental results show that not only our proposal is comparable to current state-of-the-art methods, proving to be effective and robust, but it is also able to identify specific regions that are crucial in the neural network decision-making process, thus improving explainability. Indeed, classification

accuracy is lower when highlighted regions are removed from the input images; this suggests the importance of these areas in disease classification and the possibility to consider the set of elements identified as potential disease markers.

In general, as one might expect, dataset quality, along with careful dimensionality reductions and feature selection schemes, are of great importance for effective CNNs, and subsequently for accurate disease predictions. Hence, the choice of the most relevant and informative feature subset for training a model is the basis of robust and competitive models.

In contexts where early and accurate medical diagnosis of specific pathologies are essential, our method proves that data visualization combined with machine learning techniques can be used to analyze high-dimensional multivariate data and automatically discover new bio-markers by interpreting network decisions.

As future work is concerned, we plan to investigate misclassification errors and build more robust classifiers with better generalization.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of NVIDIA Corporation for the donation of GPUs used in this research.

REFERENCES

- [1] L. Zhang, et al., “Applying 1-norm SVM with squared loss to gene selection for cancer classification,” *Applied Intelligence*, vol. 48(7), pp. 1878–1890, 2018.
- [2] E. Aličković, and S. Abdulhamit, “Breast cancer diagnosis using GA feature selection and Rotation Forest,” *Neural Computing and Applications*, vol. 28(4), pp. 753–763, 2017.
- [3] Moccia, S., et al., “Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images. *Magnetic Resonance Materials in Physics*,” *Biology and Medicine*, 32(2), 187-195, 2019.
- [4] M. West, et al., “Predicting the clinical status of human breast cancer by using gene expression profiles,” *Proceedings of the National Academy of Sciences*, vol. 98(20), pp. 11462–11467, 2001.
- [5] M.D. Zeiler, and R. Fergus, “Visualizing and understanding convolutional networks,” *European conference on computer vision*, Springer, Cham, 2014.
- [6] A. Mahendran, A. Vedaldi, “Understanding deep image representations by inverting them,” In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5188–5196, 2015.
- [7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, “GradCAM: Visual explanations from deep networks via gradient-based localization,” In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.
- [8] S. Bach, A. Binder, G. Montavon, F. Klauschen, K. R. Müller, W. Samek, “On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation,” *PloS one*, vol. 10(7), pp. e0130140, 2015.
- [9] B. Zhou, Y. Sun, D. Bau, A. “Torralba, Interpretable basis decomposition for visual explanation,” In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 119–134, 2018.
- [10] G. Huang, Z. Liu, L. v. d. Maaten, K. Q. Weinberger, “Densely connected convolutional networks,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [11] J. T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, “Striving for simplicity: The all convolutional net,” *arXiv preprint arXiv*, vol.1412(6806), 2014.
- [12] D.G. Altman, “Analysis of Survival times,” *Practical statistics for medical research*, CRC press, pp. 365-93, 1990.
- [13] H. Hsu, A. L. Peter, “Paired t test,” *Wiley encyclopedia of clinical trials*, pp.1–3, 2007.
- [14] J.M. Bland, D.G. Altman, “The logrank test,” *Bmj*, vol. 328(7447), p. 1073, 2004.