DEEP LEARNING APPROACHES FOR SMALL DIMENSIONAL BIOMEDICAL DATA

Karin Proell [(a)]

(a) Department of Medical Informatics and Bioinformatics
Upper Austrian University of Applied Sciences Hagenberg, Softwarepark 11, Austria
(a) Karin.Proell@fh-hagenberg.at

ABSTRACT
In this paper we apply convolutional neuronal networks in different configurations to solve prediction tasks on medical data: Given 27 blood parameters obtained by labor blood examination the classes of tumor markers C153 and PSA should be predicted. Based on former work the results of trained Multi-Layer-Perceptrons (MLP) were moderate. Our major interest was now focused on the question if the prediction quality of CNN models outperforms MLPs. We had to transform the vector of input data into a two-dimensional pseudo image and augment it with different correlation values for increasing spatial structure. Various experiments with CNNs show that the prediction quality slightly increases compared to MLPs.

Keywords: deep learning, convolutional neural networks, Multi-Layer-Perceptron, transformation of vector into pseudo image

1. INTRODUCTION

Analyzing biomedical data is most times a classification task. Given a series of sample data a model should be provided which assigns each sample to one of several predefined output classes which can be used for prediction tasks.
Typically in labor blood examination numerous blood parameters such as HB, WBC, HKT, MCV, RBC, PLT, KREA, BUN, GT37, ALT, AST, TBIL, CRP, LD37, HS, CNEA, CMOA, CLYA, CEOA, CBAA, CHOL, HDL, CH37, FER, FE, BSG1, TF and tumor markers such as AFP, C125, C153, C199, C724, CEA, CYFRA, NSE, PSA, S100, SCC, TPS etc. are measured and used for diagnostic purposes (Djavan et al. 2002; Harrison et al. 2005; Jung at al. 2005). The value ranges of tumor markers are divided into four non-overlapping intervals, called classes. Class 1 includes all values less than *Normal Value* of marker, Class 2 includes all values between *Normal Value and Extreme Normal Value* of marker, Class 3 includes values between *Extreme Normal Value and Plausible Value* of marker and Class 4 includes all values *Greater than Plausible Value*.
The question is to find a model to predict the classes of each tumor marker separately using only the measurements of the blood examination as input.
One major problem of this task are missing values in the input data. It may be that a specific medical procedure was not considered necessary in a particular case or that the procedure was taken in a different laboratory with the values not available in the patient record, or that the measurement was taken but not recorded due to time constraints.
In previous work we focused on a variety of methods to handle missing data, including relatively simple approaches like discarding samples containing missing data values, replacing missing values with zero or applying mean imputation. We also applied different approaches for estimation of missing values in the input data: neural network based estimation of a specific marker value depending on existing values of a related marker and neural network based estimation of missing tumor markers depending on standard blood parameter measurements (Jacak et al. 2014; Markey et al. 2006, Liparini et al. 2005).
Additionally we trained Neural Networks (MLPs) with different configurations using the blood parameters as input and the tumor marker classes as output we could observe different prediction quality of the models for each marker type (Jacak et al. 2011; Jacak et al. 2010a; Jacak et al. 2010b).
In all experiments prediction quality of the models was moderate but the best results were obtained for models of tumor marker C153 and the worst for models of tumor marker PSA.
Our major interest in the current experiments was now focused on the question if the prediction quality of CNN models outperforms prediction quality of traditional MLPs. In our experiments we compared prediction quality of CNN Models (deep learning) to traditional MLPs (shallow approach) for C153 and PSA tumor markers. We did not apply sophisticated methods for missing value imputation but simply replaced missing value data with zero.
At our disposal examples of 27 blood parameters (see above) with known output classes of tumor marker C153 of approximately 6200 patients are provided.
So our data set comprises 27 blood parameters as input and C153 class values as output of approximately 6200 patients.
Secondly examples of 27 blood parameters with known output classes of tumor marker PSA of approximately 4300 patients are available.

1.1. State of the art:

Deep learning has in recent years set an exciting new trend in machine learning. The theoretical foundations

of deep learning are well rooted in the classical neural network (NN) literature. But different to more traditional use of NNs, deep learning accounts for the use of many hidden neurons and layers—typically more than two—as an architectural advantage combined with new training paradigms. While resorting to many neurons allows an extensive coverage of the raw data at hand, the layer-by-layer pipeline of nonlinear combination of their outputs generates a lower dimensional projection of the input space. Every lower-dimensional projection corresponds to a higher perceptual level. Provided that the network is optimally weighted, it results in an effective high-level abstraction of the raw data or images. This high level of abstraction render an automatic feature set, which otherwise would have required hand-crafted or bespoke features (Miotto et al. 2017; Schmidhuber 2015).

In domains such as health informatics, the generation of this automatic feature set without human intervention has many advantages:

Clinical Imaging: Following the success in computer vision, the first applications of deep learning to clinical data were on image processing, especially on the analysis of brain Magnetic Resonance Imaging (MRI) scans to predict Alzheimer disease and its variations. In other medical domains, CNNs were used to infer a hierarchical representation of low-field knee MRI scans to automatically segment cartilage and predict the risk of osteoarthritis. Deep learning was also applied to segment multiple sclerosis lesions in multi-channel 3D MRI and for the differential diagnosis of benign and malignant breast nodules from ultrasound images. More recently, CNNs were used to identify diabetic retinopathy in retinal fundus photographs, obtaining high sensitivity and specificity over about 10 000 test images with respect to certified ophthalmologist annotations. CNNs also obtained performances on classifying biopsy-proven clinical images of different types of skin cancer over a large data set of 130 000 images.

Electronic health record (EHR): More recently deep learning has been applied to process aggregated EHRs, including both structured (e.g. diagnosis, medications, laboratory tests) and unstructured (e.g. free-text clinical notes) data. The greatest part using deep architectures is applied for a specific, usually supervised, predictive clinical task. In particular, a common approach is to show that deep learning obtains better results than conventional machine learning models with respect to certain metrics, such as Area Under the Receiver Operating Characteristic Curve, accuracy and F-score. Most applications present end-to-end supervised networks, some works also propose unsupervised models to derive latent patient representations, which are then evaluated using shallow classifiers (e.g. random forests, logistic regression).
Several works applied deep learning to predict diseases from the patient clinical status. A four-layer CNN was used to predict congestive heart failure and chronic obstructive pulmonary disease and showed significant advantages over the baselines.
A comprehensive literature review about deep learning for healthcare, including a summary of articles can be found in (Miotto et al. 2017; Rav et al. 2017).

2. EXPERIMENTAL DATA

2.1. Initial Situation:

Compared to typical input data sets used in deep learning tasks we have to deal with the following problems in connection with our available data set:

- Input vectors comprise only 27 parameters (small dimension) per patient and the number of those sample vectors is limited to about 6200 for C153 and 4300 for PSA records.
- Many of the input vectors have a huge number of missing parameter values as blood examinations are expensive and therefore only a small subset of parameters is of interest for a specific diagnosis. The C153 data sets contains about 40 % and the PSA data set about 30% missing values in the blood examination data.
- The input vectors are not uniformly distributed among the four output classes. The major part of the input vectors is assigned to class value 1 which indicates no clinical evidence. We discarded samples of class 1 to obtain a proportion of not more than 50 % data sets of class 1 and 50 % in classes 2, 3 and 4.

2.2. Transformation of dimension:

The major problem for the CNN approach was the transformation of the one dimensional vector of 27 parameters of blood examination into a two dimensional augmented matrix without having additional information to be included. So the task was to transform the one dimensional vector of blood examination parameters of every patient into a two dimensional pseudo image. These images are considered as a matrix of pixel values with grey scaled values representing the values of blood examination. To increase the number of dimensions and to obtain a local spatial structure among every parameter value the following transformation methods for increasing the number of dimensions were applied:
Correlation was calculated

- between input parameters and
- between input parameter and tumor markers

and added as further dimensions to the pixel matrix.
This results in a non-quadratic matrix which was finally resized to a quadratic one with 28x28 dimensions. These pseudo images were used as input data to train MLPs and CNNs with different configurations.
In Table 1 correlation of blood parameters to tumor markers C153 and PSA are shown.

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

Table 1: Correlation of blood parameters to tumor markers C153 and PSA

Blood Parameter	Correlation to C153	Correlation to PSA
ALT	0,2	-0,1
AST	0,4	0,0
BSG1	0,2	0,2
BUN	0,1	0,1
CBAA	-0,1	-0,1
CEOA	-0,1	-0,1
CH37	-0,1	-0,1
CHOL	0,0	-0,1
CLYA	-0,2	-0,2
CMOA	0,0	0,0
CNEA	0,2	0,2
CRP	0,3	0,2
FE	-0,3	-0,1
FER	0,3	0,1
GT37	0,4	0,0
HB	-0,4	-0,2
HDL	-0,1	0,1
HKT	-0,4	-0,2
HS	0,1	-0,1
KREA	0,1	0,1
LD37	0,5	0,2
MCV	0,0	0,1
PLT	0,2	0,1
RBC	-0,3	-0,2
TBIL	0,0	0,0
TF	-0,3	-0,1
WBC	0,0	0,1

These pseudo images were used as input data to train MLPs and CNNs with different configurations. In Figure 1 und Figure 2 examples of such pseudo images are presented.

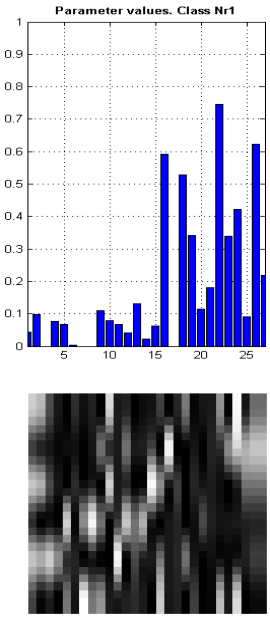

Figure 1: Example of a pseudo image for a Class 1 patient. Black fields indicate missing values.

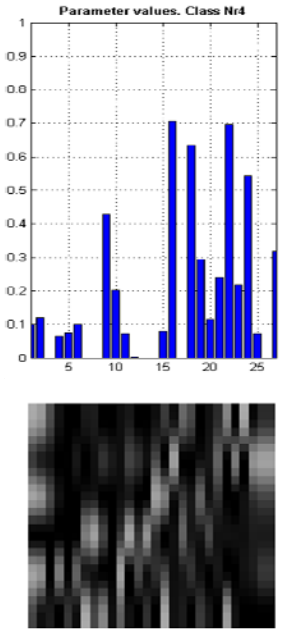

Figure 2: Example of a pseudo image for a Class 4 patient. Black fields indicate missing values.

2.3. Experiments

Our major interest in the experiments was focused on the prediction quality of CNN models in comparison to traditional MLPs. To perform the experiments we separated the data set with the 6200 and 3400 pseudo images into a collection of training and test set. The experiments with MLPs were conducted using the Neural Network Toolbox™ of MATLAB 2016a. The CNN experiments are based on the DeepLearnToolbox (Berg 2012) with slightly modified code.

After processing a training set its performance was measured on the corresponding test set. The following experiments were performed:

1. Experiments with MLPs: Traditional MLPs with various numbers of neurons were applied to train the pseudo images. The best result could be obtained using a two hidden layer MLP with 54 neurons.
2. Experiments with CNNs: For CNNs we used different configurations concerning:
   - Number of filters
   - Dimension of filters
   - Type of filters
   - Number of convolutional layers
   - Different scales of maxpooling

2.3.1. Experiments with tumor marker C153

For experiments with marker C153 we divided all samples (=6200) into 70% training data and 30% test data. Table 2 presents the different configurations for CNNs used for 10 experiments for marker type C153.

Table 2: Configuration of CNNs in experiments for tumor marker C153

	Number of filters	Dimension of filters	Number of convolutional layers	Scale of maxpooling	Types of filter
exp 1	1	5x5	1	4	average
exp 2	1	5x5	1	4	gaussian
exp 3	1	5x5	1	4	log
exp 4	1	5x5	1	4	prewitt
exp 5	1	5x5	1	4	random
exp 6	3	5x5	1	4	average, log, gaussian
exp 7	3	5x5	1	4	average, random, gaussian
exp 8	3	5x5	1	4	prewitt, log, random
exp 9	3	5x5	1	4	random
exp 10	3	5x5	1	4	prewitt, prewitt, log

The best result could be obtained using the configuration of CNN in experiment 9:

- Number of filters: 3
- Dimensions of filters: 5x5
- Types of filters: random
- Number of convolutional layers: 1
- Scale of maxpooling: 4

Performance in all experiments was measured by the following criteria:

- Percent of correct classification of samples (true positives)
- Distribution quality: Distribution of samples to correct classes
- Cohen's kappa coefficient

The results of the C153 experiments are shown in Figure 3.

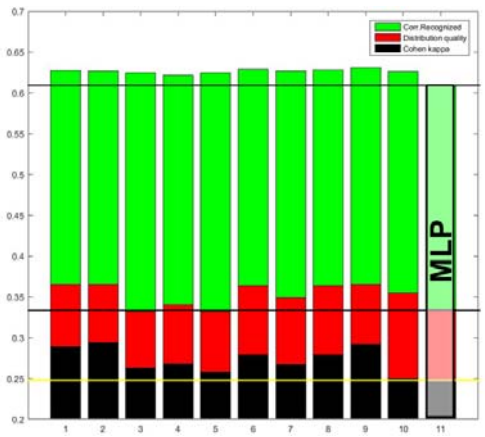

Figure 3: Comparison of results for marker C153 between ten different configuration setups of CNNs and one MLP based on the quality criteria: Percent of correct classification of samples, distribution quality among classes and Cohen's kappa coefficient

2.3.2. Experiments with tumor marker PSA

From previous work we know that the prediction quality of models for the PSA data is low. Therefore our experiments were focused on an increasing number of kernels working with random filters. The samples (=3412) were divided into 70% training data and 30% test data. Table 3 presents the different configurations for CNNs used for 4 different experiments for marker type PSA. The best result could be obtained using the configuration of CNN in experiment 3 with 7 random filters. Filter numbers greater 7 tend to over fit, the recognition rate of the test samples decreases.

Table 3: Configuration of CNNs in experiments for tumor marker PSA

	Number of filters	Dimension of filters	Number of convolutional layers	Scale of maxpooling	Types of filter
exp 1	3	5x5	1	4	random
exp 2	5	5x5	1	4	random
exp 3	7	5x5	1	4	random
exp 4	9	5x5	1	4	random

The best result could be obtained using the configuration of CNN in experiment 4:

- Number of filters: 7
- Dimensions of filters: 5x5
- Types of filters: Random
- Number of convolutional layers: 1
- Scale of maxpooling: 4

The results of the PSA experiments are presented in Figure 4.

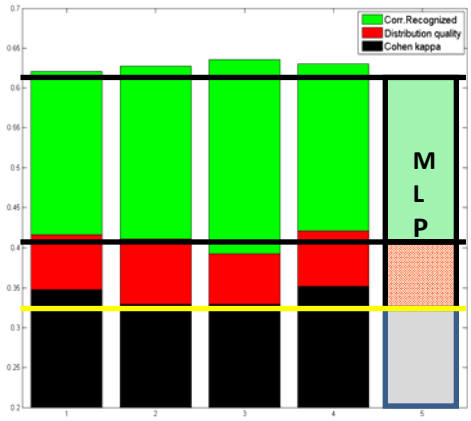

Figure 4: Comparison of results for marker PSA between four different configuration setups of CNNs and one MLP based on the quality criteria: Percent of correct classification of samples, distribution quality among classes and Cohen's kappa coefficient

ISBN 978-88-97999-85-0; Affenzeller, Bruzzone, Jiménez, Longo and Piera Eds.

3. RESULTS

We have achieved moderate increase of performance (about 3%) applying convolutional neural networks compared to MLPs trained on human blood parameters as input and C153 and PSA tumor marker as output.

The quality of the learned systems is primarily dependent of quality and size of the training sets. In the available blood examination data sets in our experiments we had to deal with about 40% missing values and input vectors are not uniformly distributed among the four output classes. We replaced missing values by imputation with zero value and discarded samples of class 1 to get a proportion of 50% class 1 samples and 50% samples of class 2, 3 and 4.

Standard neural networks are state-of-the-art classifiers that operate on vectors, without knowledge of the input topology. However, convolutional neural network exploit the knowledge that the inputs are not independent elements, but arise from a spatial structure.

Therefore we did several experiments with different configurations for CNNs augmenting the blood examination vector with correlation values between input parameter and tumor marker values and transformed it into a pseudo image.

In all experiments we found a CNN configuration which outperforms MLPs. Therefore we will apply further work in data preparation especially in context of missing value imputation and perform continue experiments with different CNN configurations on tumor markers other than C153 and PSA.

REFERENCES

Berg Palm R., 2012. Prediction as a candidate for learning deep hierarchical models of data, Available from: https://de.mathworks.com/matlabcentral/fileexchange/38310-deep-learning-toolbox

Djavan et al., 2002. Novel Artificial Neural Network for Early Detection of Prostate Cancer. Journal of Clinical Oncology, Vol 20, No 4, 921-929

Harrison, R.F., Kennedy, R.L., 2005, Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation. Ann Emerg Med.; 46(5):431-9.

Jacak W., Proell K., Winkler, S., 2014. Neural Networks based Feature Selection in Biological Data Analysis. in: Klempous, R., J. Nikodem, W. Jacak, Z. Chacko, eds. Advanced Methods and Applications in Computational Intelligence, Springer, pp.79-93

Jacak W., Proell K., 2010a, Data Driven Tumor Marker Prediction System, Proceedings of EMSS 2010, Fes, Marokko

Jacak W., Proell K., 2010b, Neural Network Based Tumor Marker Prediction, Proceedings of BroadCom 2010, Malaga, Spain

Jacak W., Proell K, 2011. Neural Networks Based System for Cancer Diagnosis Support - Proceedings of International Conference on Computer Aided Systems Theory EUROCAST 2011, Las Palmas, Spanien, pp. 283-285

Jung I., Thapa D., Wang GN., 2005. Neural Network Based Algorithms for Diagnosis and Classification of Breast Cancer Tumor. In: Hao Y. et al. (eds) Computational Intelligence and Security. CIS 2005. Lecture Notes in Computer Science, vol 3801. Springer, Berlin, Heidelberg

Liparini A., Carvalho S., Belchior J.C., 2005. Analysis of the applicability of artificial neural networks for studying blood plasma: determination of magnesium on concentration as a case study. *Clin Chem Lab Med*.; 43(9):939-46

Markey M.K., Tourassi G.D., Margolis M., DeLong D., 2006. Impact of missing data in evaluating artificial neural networks trained on complete data. Computers in Biology and Medicine, 5, 516

Miotto R., Wang F., Wang S., Jiang X., Dudley J., 2017. Deep learning for healthcare: review. Briefings in Bioinformatics, 2017, 1–11

Rav D., Wong C., Deligianni F., Berthelo M., Andreu-Perez J., Lo B., Yang G., Fellow, 2017. Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, Vol 21, No. 1, January

Schmidhuber J., 2015. Deep learning in neural networks: An overview. Neural Networks, Volume 61, January 2015, Pages 85-117