Application of artificial intelligence in medicine

PB016 – Umělá intelligence I.

Roman Ďuriš - 485642

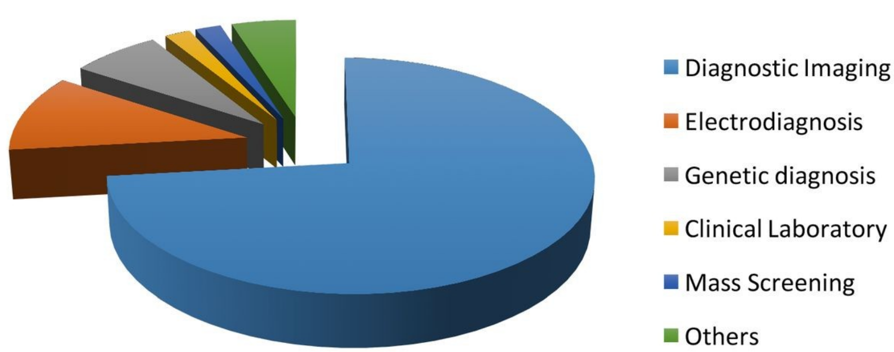
1. Introduction

Artificial intelligence is a part of computer science that tries to make computers more intelligent. One of the basic requirements for any intelligent behavior is learning. Therefore, machine learning is one of the major branches of artificial intelligence and, indeed, it is one of the most rapidly developing subfields of AI research. Machine learning algorithms were from the very beginning designed and used to analyze medical datasets. Modern hospitals are well equipped with monitoring and other data collection devices, and data is gathered and shared in large connected information systems. Machine learning technology is currently well suited for analyzing medical data and in particular, there is a lot of work done in medical diagnosis in small specialized diagnostics problems.

Data for learning algorithms are often available in the form of medical records in hospitals. All that has to be done is to input the records with known correct results to the learning algorithm, in principle, the medical diagnostic knowledge can be automatically derived from the description of cases solved by humans in the past. The derived classifier can be used in several ways:

- AI can learn features from a large volume of healthcare data, and then use the obtained insights to assist the clinical practice in treatment design or risk assessment;
- AI system can extract useful information from a large patient population to assist in making real-time inferences for health risk alert and health outcome prediction;
- AI can do repetitive jobs, such as analyzing tests, X-Rays, CT scans or data entry;
- AI systems can help to reduce diagnostic and therapeutic errors that are inevitable in human clinical practice;
- AI can assist physicians by providing up-to-date medical information from journals, textbooks and clinical practices to inform proper patient care;
- AI can manage medical records and analyze both the performance of an individual institution and the whole healthcare system;
- AI can help develop precision medicine and new drugs based on the faster processing of mutations and links to disease;
- AI can provide digital consultations and health monitoring services — to the extent of being “digital nurses” or “health bots”.

None of the sections is intended to provide a comprehensive overview, but rather to describe some subareas, algorithms, and applications which from my point of view seem to be important for medicine.

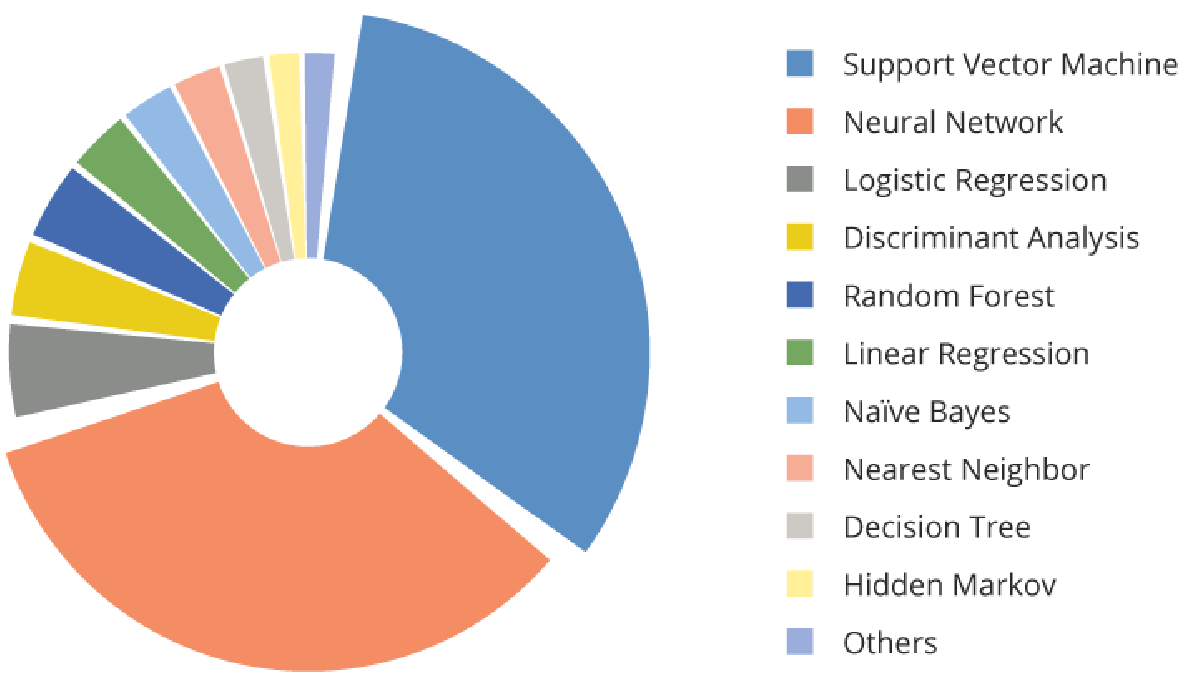


2. History

As soon as electronic computers came into use in the 1960s, the algorithms were developed that enabled modeling and analyzing large sets of data. From the very beginning, three major branches of machine learning emerged, symbolic learning, statistical methods, and neural networks. Through the years, all three branches developed advanced methods: statistical or pattern recognition, such as k-nearest neighbors, discriminant analysis, and Bayesian classifiers, inductive learning of symbolic rules, such as top-down induction of decision trees, decision rules and induction of logic programs, and artificial neural networks , such as the multilayered feedforward neural network with backpropagation learning, the Kohonen's self-organizing network , and the Hopfield's associative memory.

3. Present

Nowadays we have equipped the best suitable methods for machine learning and acknowledged requirements that any machine learning system has to satisfy to be used in the development of applications in medical diagnosis. We will briefly describe each requirement and several methods and algorithms which are used at present followed by some interesting applications of these methods in practice.



3.1 Requirements for machine learning in Healthcare

For a machine learning system, to be useful in solving medical tasks, the following feature are desired: good performance, the ability to appropriately deal with missing data and with noisy data (errors in data), the transparency of diagnostic knowledge, the ability to explain decisions and the ability of the algorithm to reduce the number of tests necessary to obtain reliable result.

3.1.1 Good Performance

The algorithm has to be able to extract significant information from the available data. The diagnostic accuracy of new cases has to be as high as possible. Algorithms should perform at least as well as physicians to perform better as they using the same description of patients. Therefore, if there is a possibility to measure the accuracy of physicians, their performance can be used as the lower bound of the required accuracy if the ML algorithm in the given problem.

3.1.2 Dealing with missing and noisy data

In medical diagnosis, very often records of patients are incomplete or suffer from uncertainty or errors. Therefore, machine learning algorithms have to be able to deal with incomplete descriptions of patients and they need to have effective means for handling noisy data.

3.1.3 Transparency of diagnostic knowledge

The generated knowledge should and the explanation of decisions be transparent to the physician. A physician should be able to analyze and understand the generated knowledge. Ideally, the automatically generated knowledge will provide to the physician a novel point of view on the given problem, and may reveal new interrelations and regularities that physicians did not see before in explicit form.

3.1.4 Explanation ability

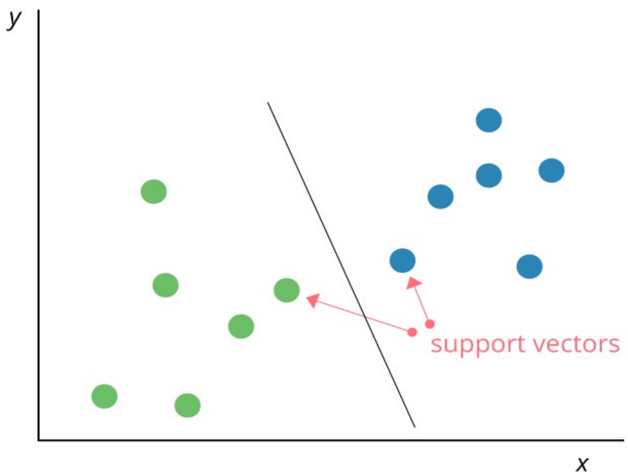
The system must be able to explain decisions when diagnosing new patients. When faced with and unexpected solution physician should be able to get a further explanation of how the decision was made. The only possibility to accept unexpected solution from "black box" classifier is in the situation where such classifier significantly outperforms all other classifiers, including the physicians themselves in terms of the classification accuracy.

3.1.5 Reduction of the number of tests

In medical practice, the collection of patient data is often expensive, time-consuming and harmful for the patients. Therefore, it is desirable to have a classifier that is able to reliably diagnose with the smallest possible set of data about the patient. However, the process of determining the right subset of data may be time-consuming, as it is essentially a combinatorial problem. Some ML systems are able to select appropriate subset themselves during the learning process and may be more appropriate than others that lack this facility.

3.2 Description of algorithms

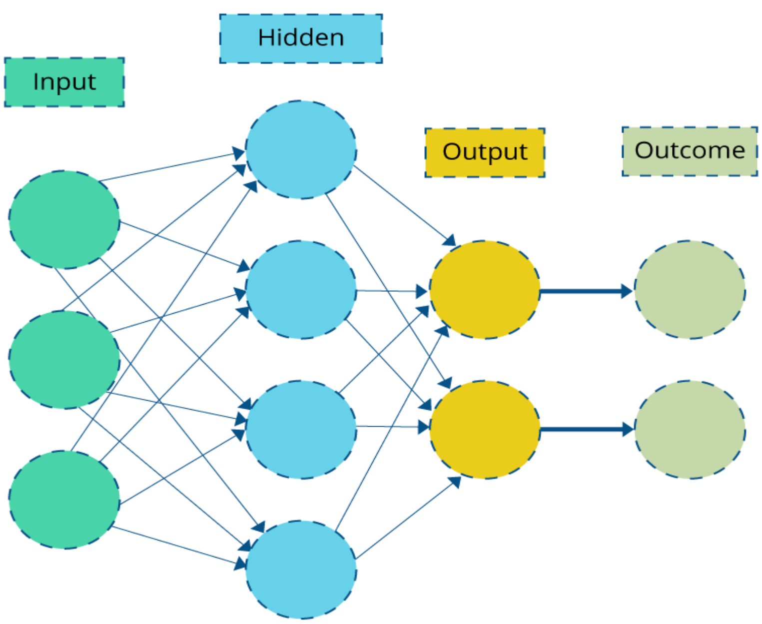
3.2.1 Support Vector Machine



Support Vector Machines (SVM) can be employed for classification and regression, but this algorithm is chiefly used in classification problems that require division of a dataset into two classes by a hyperplane. The goal is to choose a hyperplane with the greatest possible margin, or distance between the hyperplane and any point within the training set, so that new data can be classified correctly. Support vectors are data points that are closest to the hyperplane and that, if removed, would alter its position. In SVM, the determination of the model parameters is a convex optimization problem so the solution is always global optimum.

SVMs are used extensively in clinical research, for example, to identify imaging biomarkers, to diagnose cancer or neurological diseases and in general for classification of data from imbalanced datasets or datasets with missing values.

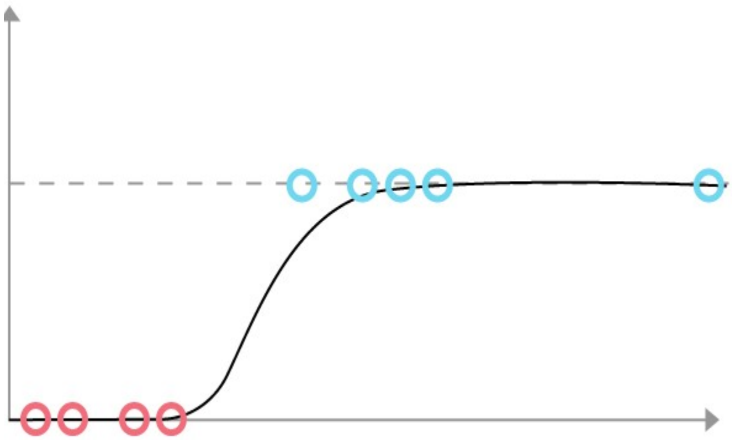
3.2.2 Neural networks



In neural networks, the associations between the outcome and the input variables are depicted through hidden layer combinations of prespecified functionals. The goal is to estimate the weights through input and outcome data in such a way that the average error between the outcome and their predictions are minimized.

Neural networks are successfully applied to various areas of medicine, such as diagnostic systems, biochemical analysis, image analysis, and drug development, with the textbook example of breast cancer prediction from mammographic images.

3.2.3 Logistic Regression



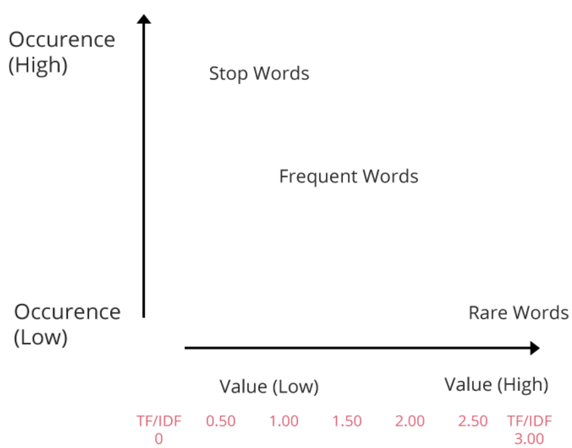
Logistic Regression is one of the basic and still popular multivariable algorithms for modeling dichotomous outcomes. Logistic regression is used to obtain the odds ratio when more than one explanatory variable is present. The procedure is similar to multiple linear regression, with the exception that the response variable is binomial. It shows the impact of each variable on the odds ratio of the observed event of interest. In contrast to linear regression, it avoids confounding effects by analyzing the association of all variables together.

In healthcare, logistic regression is widely used to solve classification problems and to predict the probability of a certain event, which makes it a valuable tool for a disease risk assessment and improving medical decisions.

3.2.4 Natural Language Processing

In healthcare, a large proportion of clinical information is in the form of narrative text, such as physical examination, clinical laboratory reports, operative notes and discharge summaries, which are unstructured and incomprehensible for the computer program without special methods of text processing. Natural Language Processing addresses these issues as it identifies a series of disease-relevant keywords in the clinical notes based on the historical databases that after validation enter and enrich the structured data to support clinical decision making.

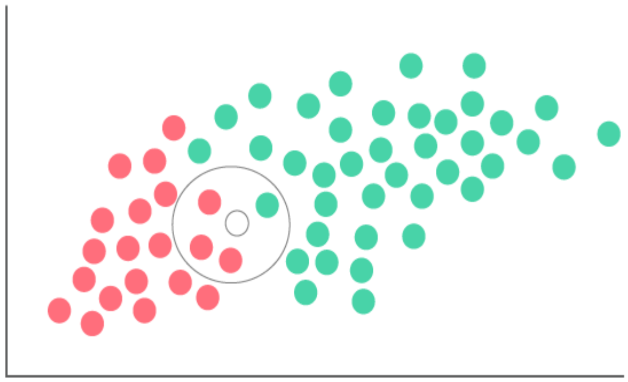
3.2.5 TF-IDF



The basic algorithm for extracting keywords, TF-IDF stands for term frequency-inverse document frequency. The TF-IDF weight is a statistical measure of word importance to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus.

In healthcare, TF-IDF is used in finding patients' similarity in observational studies, as well as in discovering disease correlations from medical reports and finding sequential patterns in databases.

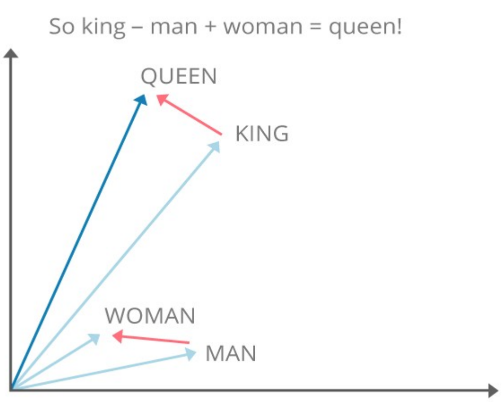
3.2.6 Naïve Bayes



Naïve Bayes classifier is a baseline method for text categorization, the problem of judging documents as belonging to one category or the other. Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features are interdependent, all of these properties independently contribute to the probability of belonging to a certain category.

It remains one of the most effective and efficient classification algorithms and has been successfully applied to many medical problems, such as the classification of medical reports and journal articles.

3.2.7 Word Vectors



Considered to be a breakthrough in NLP, word vectors, or word2vec, is a group of related models that are used to produce word embeddings. In their essence, word2vec models are shallow, two-layer neural networks that reconstruct linguistic contexts of words. Word2vec produces a multidimensional vector space out of a text, with each unique word having a corresponding vector. Word vectors are positioned in the vector space in a way that words that share contexts are located in close proximity to one another.

Word vectors are used for biomedical language processing, including similarity finding, medical terms standardization and discovering new aspects of diseases.

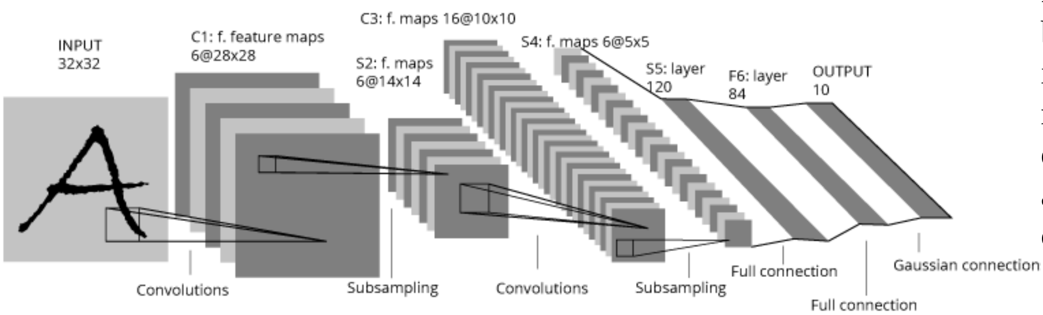
3.2.8 Deep Learning

Deep Learning is an extension of the classical neural network technique, being, to put it simply, as a neural network with many layers. Having more capacities compared to classical ML algorithms, Deep Learning can explore more complex non-linear patterns in the data. Being a pipeline of modules each of them is trainable, Deep Learning represents a scalable approach that, among others, can perform automatic feature extraction from raw data.

In the medical applications, Deep Learning algorithms successfully address both Machine Learning and Natural Language Processing tasks. The commonly used Deep Learning algorithms include convolution neural network (CNN), recurrent neural network, deep belief network and multilayer perceptron, with CNNs leading the race from 2016 on.

3.2.9 Convolutional Neural Network

The CNN was developed to handle high-dimensional data, or data with a large number of traits, such as images. Initially, the inputs for CNN were normalized pixel values on the images. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex, with individual cortical neurons responding to stimuli only in a restricted region of the receptive field. However, the receptive fields of different neurons partially overlap such that they cover the entire visual field. The CNN then transfers the pixel values in the image by weighting in the convolution layers and sampling in the subsampling layers alternatively. The final output is a recursive function of the weighted input values.

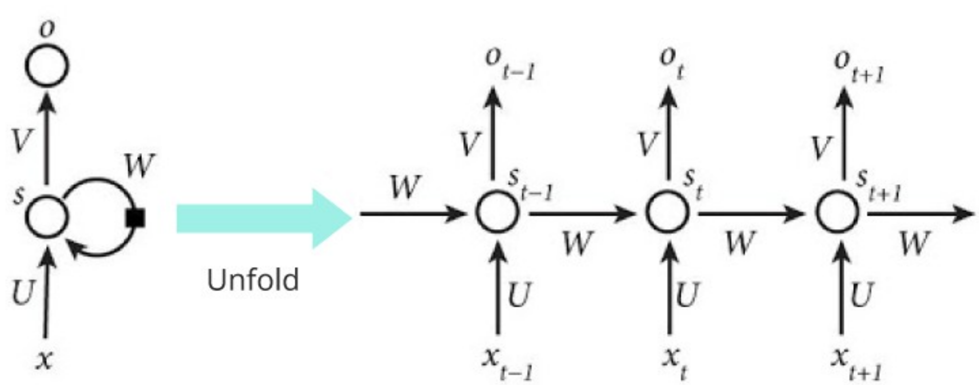


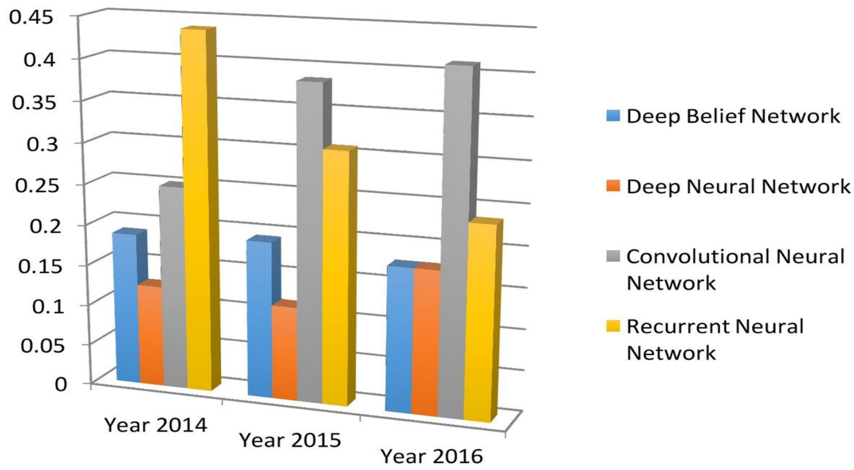
Recently, CNN has been successfully implemented in the medical area to assist disease diagnosis, such as skin cancer or cataracts.

3.2.10 Recurrent Neural Network

The second in popularity in healthcare, RNNs represent neural networks that make use of sequential information. RNNs are called *recurrent* because they perform the same task for every element of a sequence, and the output depends on the previous computations. RNNs have a “memory” which captures information about what has been calculated several steps back (more on this later).

Extremely popular in NLP, RNNs are also a powerful method of predicting clinical events.





Different from the classical neural network, deep learning uses more hidden layers so that the algorithms can handle complex data with various structures. In the medical applications, the commonly used deep learning algorithms include convolution neural network (CNN), recurrent neural network, deep belief network and deep neural network. This graph depicts their trends and relative popularities from 2013 to 2016. One can see that CNN is the most popular one in 2016.

3.3 Interesting AI implementations in practice

3.3.1 Classification of tumors and its stages in brain MRI using support vector machine and artificial intelligence

Introduction

New incorporation of SVM and ANN for tumor classification is introduced in this research. Brain MRI images with normal and abnormal behavior are firstly enhanced through some filter and preprocessing steps. Therefore, for the detection and classification of the tumor in the brain there be incorporated segmentation processes namely temper based K-means and modified Fuzzy C-means (TKFCM) clustering algorithm is used. In these techniques, the K-values vary from 1 to 8 which was limited to only 1 to 3 in conventional K-means and the automatically updated membership function eradicates limitation of FCM. Then, two kinds of features are extracted from these segmented images between these one kind is used to detect the tumor with SVM as it is easy to classify two kinds and other is used to classify tumor with ANN into five categories along with four malignant stages

Methodology

The brain MRI image data of normal brain and tumor brain with Lower Grade Glioma or Glioblastoma Multiforme is collected. All images are converted to .jpg form by using the MATLAB conversion tool. After the conversion, all images have the same size and direction using the Photoshop CS5 application with 256 x 256 pixels. Then the image is enhanced using adjusted, adaptive and histogram imaging.

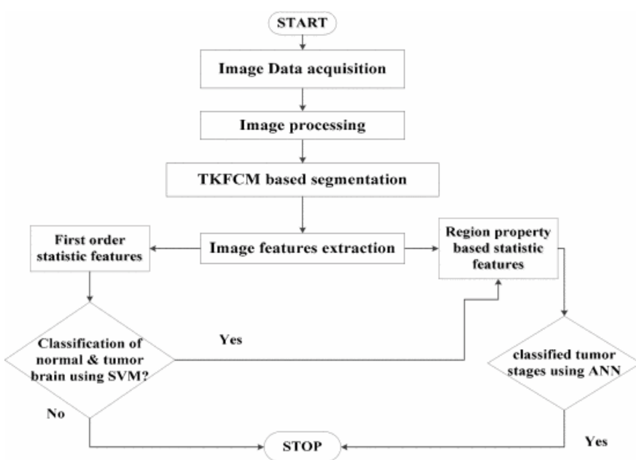
In the segmentation process, the K-means algorithm is used to segment MRI images based on the gray level. This gray level is selected depending on the temper of the image. The membership of modified Fuzzy c-means is updated with the cluster distances from centroid defined by the features of the tumor MRI image.

The TKFCM algorithm is the integration of the K-means and Fuzzy c-means with some modification. The temper is added along with the conventional K-means, which is identified by the temper or gray level intensity in the brain image. Besides, the Fuzzy c-means membership and Euclidean distance are modified by the image features.

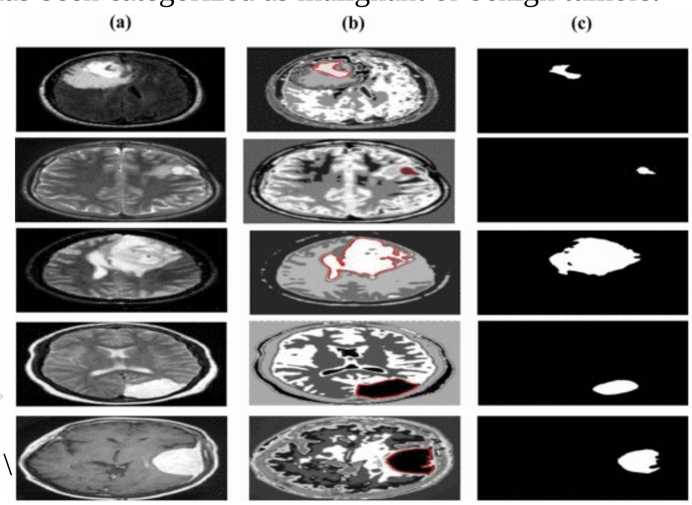
The system will extract the first and second-order statistic. The first order statistic feature is used to detect the exact tumor and its position in the brain MRI image and second-order region-based statistic feature is required for distinguishing the malignant tumor and benign tumor.

The key concept of SVM is the use of hyperplanes to define decision boundaries separating between data points of different classes. The hyperplanes for SVMs are used to separate these classified data as normal data and tumor data

Artificial Neural Network Backpropagation is one of the artificial neural networks with supervised learning. After attaining the information from the MRI tumor brain, the next process is the classification process which uses ANN-BP. The input towards the ANN is the information of feature extraction; the first and second-order statistic. Then, the ANN will generate the output in terms of classification results on MRI brain image which has been categorized as malignant or benign tumors.



\



conclusion

The proposed method showed better results as conventional methods like thresholding, region growing, SVM, ANN, FCM, TKFCM, and K-means. The accuracy of this method is 97,37 % and computational time is approximately 2 mins. per image. Therefore, it can be useful for both detecting tumor and classifying tumor stages of brain MRI images for experts.

Algorithms	Sensitivity (%)	Specificity (%)	Accuracy (%)	BER	Computational time
Thresholding	85	80	81.3	0.175	~3 min
Region Growing	88.46	75	86.47	0.182	~6 min
ANN	95.42	100	95.07	0.022	~8 min
FCM	86.95	85.7	86.4	0.136	~5 min
SVM	96.2	66.67	90.44	0.0234	~4 min
K-means	75	92.85	83.7	0.160	~160-170 sec
TKFCM [12]	88.9	100	91.89	0.055	~100 sec
Proposed SVM+ANN Method	98	100	97.37	0.0294	~ 2 min

a. Image size=256x256, Software=MATLAB2014a, Processor= Core2duo, RAM=2GB, windows=7

3.3.2 Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs

Introduction

One of the major objectives of chest radiography in the detection of pulmonary nodules that are often the initial radiologic manifestation of lung cancers. However, to date, pulmonary nodule detection on chest radiographs has not been completely satisfactory, with a reported sensitivity ranging between 36% - 84%, varying widely according to tumor size and study population. For this reason there has been increasing dependence on chest CT images, however, even low dose CT scans require approximately 50 -100 times higher radiation dose as single view radiographic examinations. The purpose of this study was to develop a deep learning-based automatic detection(DLAD) of malignant nodules on chest radiographs and to help specialists with their evaluation of scans.

Methodology

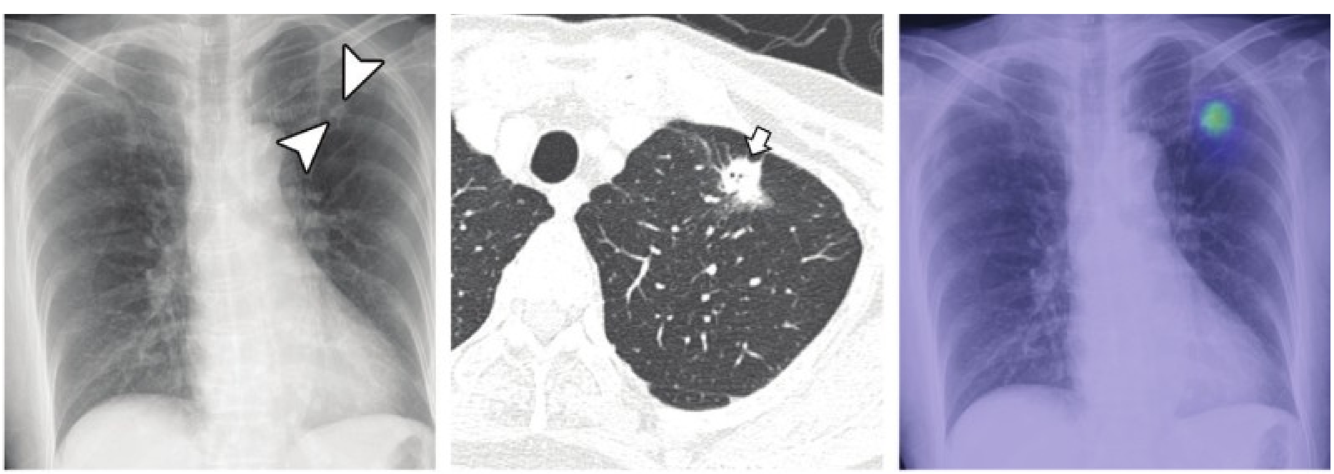
The algorithm was fed with 43 292 radiographic images, where those with malignant nodules were proven by pathologic analysis. Thereafter, the chest radiography data were randomly assigned into one of the following three data sets: Training data (42 092), tuning data (600) to optimize hyperparameters and internal validation data (600).

DLAD uses the pixel intensity of scans as an input and outputs location and the presence of a malignant nodule. For training of DLAD both, image-based information and position of nodules were used. DLAD was trained in a semisupervised learning manner because it is not only cost-effective but also enables the trained model to learn features in nodules that may be missed by radiologists.

A deep convolutional network with 25 layers and eight residual connections was designed. Brightness, contrast, and image size on input chest scans were randomly adjusted to make DLAD irrelevant to the variations. To speed up training, batch normalization was used. The outputs of three networks trained on the same data but with different hyperparameters were averaged for the final prediction.

Conclusion

DLAD accuracy of detection malignant nodules is between 92 – 99 % and it outperforms 17 out of 18 specialists. These results suggest that DLAD can help reduce human error and improve the accuracy of chest radiograph interpretation. Because DLAD was trained under the supervision of radiologists, the limitations of human perception are reflected in DLAD performance, for example, DLAD is unable to detect nodules smaller than 1 cm. For the future, establishing an algorithm supervised by a radiologically undetectable nodule by using CT scans as the reference standard is warranted.



3.3.3.Artificial Intelligence-Based Breast Cancer Nodal Metastasis Detection

Introduction

Reviewing sentinel lymph node biopsies for evidence of metastasis is an important feature of breast cancer staging, currently impacting clinical staging and treatment decisions. Reviewing lymph nodes for the presence of tumor cells is a tedious, time-consuming and potentially error-prone process. Present techniques to obtain additional sections from tissue block for further testing are associated with the increased workload, costs, and reporting delays. The modern approach to solve these problems is using deep learning for the detection of metastatic breast cancer in lymph nodes. In this study LYmph Node Assistant(LYNA) was applied to detect cancer metastases in lymph nodes.

Methodology

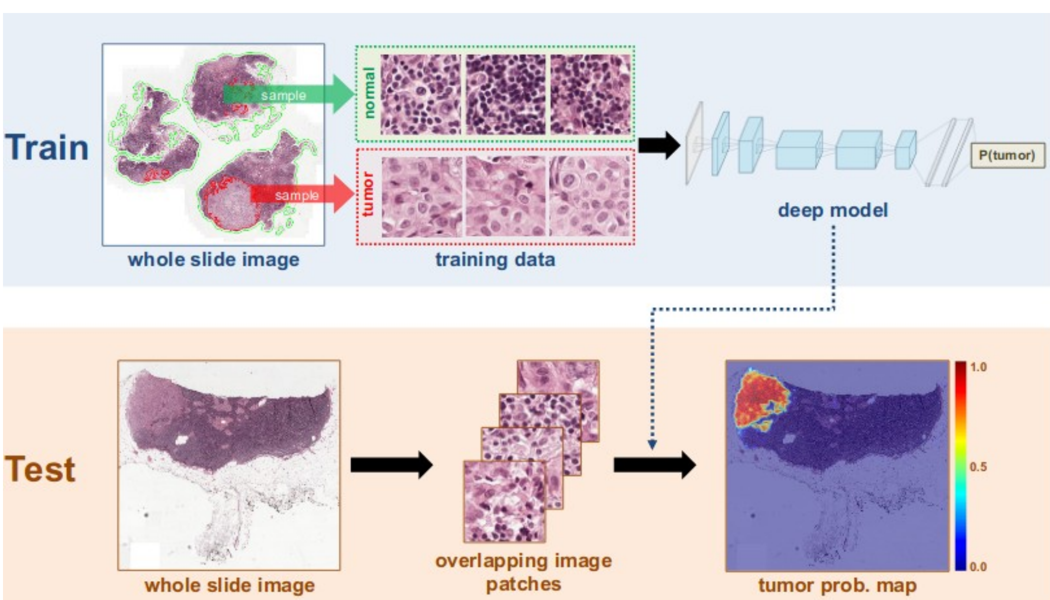
As input data, 399 whole-slide images were selected. During model training, the patch-based classification stage takes as input whole slide images and the ground truth image annotation, indicating the locations of regions of each slide containing metastatic cancer. Then millions of small positive and negative patches are selected from the training set.

These patches are then fed to GoogLeNet deep neural network for training to discriminate between the positive and negative patches. After the training stage, all training results are embedded in a heatmap image. On these heatmaps, each pixel contains a value between 0 and 1, indicating the probability that the pixel contains the tumor.

This map is then transformed into a colored heatmap of the corresponding slide with dark red color on patches with the highest probability of cancer cells.

Conclusion

The prediction heatmaps produced by the LYNA reached a localization score of 89%, which significantly exceeded the score of 73% for a pathologist with no time constraint. LYNA also ignored many types of artifacts and benign mimics of cancer. Because of very good results, LYNA can be used to significantly improve accuracy, speed, and error-rate of metastasis diagnosis.



4. Sources

1. LIU, Yun, Timo KOHLBERGER, Mohammad NOROUZI, et al. Artificial Intelligence-Based Breast Cancer Nodal Metastasis Detection: Insights Into the Black Box for Pathologists. *Archives of Pathology & Laboratory Medicine* [online]. 2019, **143**(7), 859-868 [cit. 2019-11-26]. DOI: 10.5858/arpa.2018-0147-OA. ISSN 0003-9985. http://www.archivesofpathology.org/doi/10.5858/arpa.2018-0147-OA

2. https://ai.googleblog.com/2017/03/assisting-pathologists-in-detecting.html

3. Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H. Beck, Deep learning for identifying metastatic breast cancer, 2016. https://arxiv.org/abs/1606.05718

4. NAM, Ju Gang, Sunggyun PARK, Eui Jin HWANG, et al. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology* [online]. 2019, **290**(1), 218-228 [cit. 2019-11-26]. DOI: 10.1148/radiol.2018180237. ISSN 0033-8419. http://pubs.rsna.org/doi/10.1148/radiol.2018180237

5. KONONENKO, Igor. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine* [online]. 2001, **23**(1), 89-109 [cit. 2019-11-26]. DOI: 10.1016/S0933-3657(01)00077-X. ISSN 09333657. https://linkinghub.elsevier.com/retrieve/pii/S093336570100077X

6. JIANG, Fei, Yong JIANG, Hui ZHI, et al. Artificial intelligence in healthcare: past, present, and future. *Stroke and Vascular Neurology* [online]. 2017, **2**(4), 230-243 [cit. 2019-11-26]. DOI: 10.1136/svn-2017-000101. ISSN 2059-8688. http://svn.bmj.com/cgi/doi/10.1136/svn-2017-000101

7. AHMMED, Rasel, Anirban Sen SWAKSHAR, Md. Foisal HOSSAIN a Md. Abdur RAFIQ. Classification of tumors and it stages in brain MRI using support vector machine and artificial neural network. In: *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)* [online]. IEEE, 2017, 2017, s. 229-234 [cit. 2019-11-26]. DOI: 10.1109/ECACE.2017.7912909. ISBN 978-1-5090-5627-9. http://ieeexplore.ieee.org/document/7912909/