Head and Neck Cancer Metastasis Prediction via Artificial Neural Networks

Eveleen Darby, Tezeswari Nettimi, Shilpa Kodali and Liwen Shih
Computer Engineering, U of Houston – Clear Lake (UHCL)
shih@uhcl.edu
Lincoln Gray
University of Texas – Medical School, Houston

Abstract

Cancers in head and neck tends to spread to nearby lymph nodes. Lymph nodes trap the spreading tumor cells but then the tumor starts to grow in these nodes and then spread further. Our project is aimed to (1) predict the secondary regions of lymph nodes where a given primary tumor can metastasize, and (2) to generalize a pattern for lymph nodes metastasis of head and neck cancer by using artificial neural networks (ANN). The raw data for the analysis is provided by Dr. Lincoln Gray, Acta Otolaryngologica 2000, of 130 cases of pathologically-positive oral squamous cell carcinoma (SCC) from UK. Seven primary sites for tumors are identified: 1) buccal mucosa, 2) tongue, 3) retromolar trigone, 4) floor of mouth, 5) ventral tongue, 6) oropharynx, 7) lower alveolus. Ten secondary regions of lymph nodes metastasis are observed: five regions each on the same/opposite side (ipsilateral/contralateral) as the primary tumor site. In our oral squamous cell carcinoma study using ANN, we explore data analysis approach with two ANN methods: (1) a supervised multilayer feed forward back propagation (back-prop) method, and (2) an unsupervised self-organizing map (SOM) method. This experience provides insight into implementation of ANN and directions to future investigation. The results from back-prop are comparable to that using multidimensional scaling (MDS) with respect to prediction of lymph nodes that have highest percentage of being metastasized, while SOM requires further work to identify clustering for individual primary cancer as well as next level of lymph node metastases.

Keyword: *Cancer Metastasis Prediction, Artificial Neural Networks, Back-Prop, SOM, Data Visualization*

1. Cancer Metastasis Prediction

Cancer metastasis is the primary cause leading to mortality. In head and neck cancer, metastases first appear in regional lymph nodes. The lymph nodes trap the spreading tumor cells but then the tumor starts to grow in these nodes and then spread further. [1] Therefore "the most important factor in the prognosis of squamous carcinomas of the upper aero digestive tract is the status of the cervical lymph nodes." [1] Many studies have been conducted in recognizing the pattern of lymph node metastases [2]. However, their findings are reported in a number of different ways, such as in text or in tabular fashion that make it difficult to compare and summarize the results. Thus it would be beneficial to generate visualization of the resulting data from the studies. The benefits are especially great for large or high dimensional data.

Several methods are used in visualizing high-dimensional data. One of these more traditional methods is Multidimensional Scaling (MDS) with which the previous research on the same data was conducted. The key idea of MDS is to produce low dimensional visualization of the data so that the distance of data in the new space approximates the distance of data in the original space. It has been shown that MDS can summarize disease progression and produce a map of the metastases patterns in head and neck cancer [3]. While MDS demonstrates that it can summarize the cancer data, it faces difficulties in detecting the intrinsic dimension of the data, and to discover complicated nonlinear structure. Since artificial neural networks (ANN) has been successfully applied in bio-medical research and medicine decision support, a great deal of interest exists in trying to overcoming MDS' drawbacks by analyzing and visualization data using ANN. This project attempts cancer metastasis prediction with a few ANN methods.

2. ANN Data Analysis and Visualization

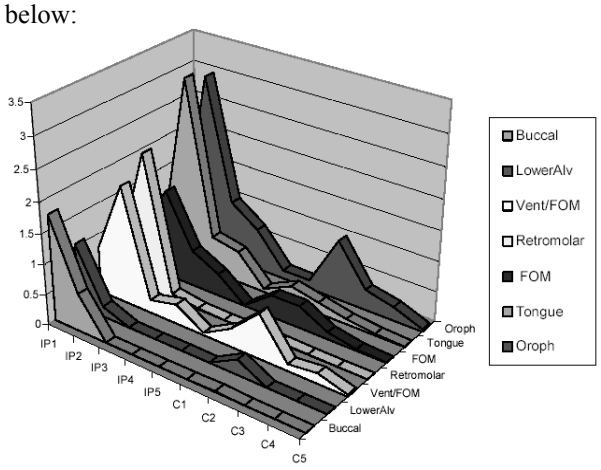
Material from 279 sequential cases of oral squamous cell carcinoma (SCC) treated at the Mersey Regional Center for Maxillofacial Surgery in Liverpool, UK between October 1989 and February 1998 provided the bases for this study. Out of the 279 cases, 130 cases are identified as having pathologically-positive neck nodes. Lymph nodes that can be metastasized are located in 10 regions; 5 regions each are in the same/opposite side (ipsi/contra lateral) as the occurrence of the primary cancer. IPSI Region 1 is the sub mental and sub maxillary region; IPSI region II is upper jugular chain; IPSI region III is middle nodes of the jugular chain; IPSI region IV is lower nodes of the jugular chain, and IPSI region V is the posterior triangle. Similarly, contra regions are located in the opposite side to the primary cancer.

Method 1 – supervised back-prop ANN: We conduct two experiments with Back-prop using EasyNN-plus software. **Experiment #1A** aims to classify which secondary site each of the primary tumors will metastasize. **Experiment #1B** takes the secondary sites as input to the network, and aims to classify which is the primary tumor it originates from. Parameters to be adjusted to arrive at optimal network architecture are *the learning rate, momentum, target error, number of iterations, and validation rules*. The best architecture (7-4-5-10) for Experiment #1A contains 7 input nodes (7 primary tumor sites), two hidden layers with 4 and 5 nodes in each layer, and 10 output nodes (10 secondary targeted nodes regions). The best architecture for Experiment #1B is 10-6-5-7.

Method 2 – unsupervised SOM: Our second method employs an unsupervised algorithm called the self-organizing maps (SOM), with experiments conducted using ANN software tool box in the MATLAB® 7. The training process of SOM describes a topology-preserving mapping from a high-dimensional input space onto a two-dimensional output space where patterns that are similar in terms of the input space are mapped to geographically close locations in the output space. **Experiment #2A** uses clustering of output data for a given primary tumor. Experiment #2A considers data only in an individual primary tumor, while **Experiment #2B** accepts all 130 cases. The input is the data shown in the secondary sites. The first SOM architecture consists of 1 input node, and 10 output nodes, while the second SOM architecture consists of 10 input and 70 output nodes.

3. Metastasis Correlation Prediction

Correlations obtained by back-prop between 7 primary sites and 10 spreading sites are depicted below:



Though cancer metastasis prediction from back-prop is comparable with previous MDS in terms of which secondary sites are more probable to be metastasized by a given primary tumor, more data details are still required to be analyzed. The results from the self-organizing map are less conclusive as the labels of each cluster are not easily identified.

4. Future Automatic Metastasis Prediction

The experiments with a feed forward multilayer back propagation method are comparable to that using multidimensional scaling with respect to prediction of lymph nodes that have highest percentage of being metastasized. The unsupervised self-organizing map method, helped us plot some cluster plots for each primary and some 3-D plots, but more efforts are needed generalize the patterns as the input data information we have is insufficient to do so. These methods require further work to identify clustering for individual primary cancer as well as next level of lymph node metastases. SOM is also compared with MDS and a few features have been found to be superior as compared to MDS.

5. References

[1] Shah JP., "Patterns of cervical lymph node metastasis from squamous carcinomas of the upper aerodigestive tract," *Am J Surg* 1990: 160: 405-9.

[2] Dickson R, Lipman M., "Molecular biology of breast cancer," in: DeVita V, Hellman S, Rosenberg S, eds. *Cancer: principles and practice of oncology,* Philadelphia, PA: Lippincott-Raven, 1997: 1541-56.

[3] Gray, L, Woolgar, J., and Brown, J.: A functional map of cervical metastases from oral squamous cell carcinoma. *Acta Otolaryngologica,* 120:885-890, 2000.