

# Comparison of MLP and Quadratic Neural Unit for Lung Tumor Motion Prediction

M.Sc. Ricardo Rodriguez

Supervisor: Prof. Ing. Jiří Bíla, DrSc.

Supervisor-specialist: Ing. Ivo Bukovský, Ph.D.

## **Abstract**

*This paper presents a study of lung tumor-motion time-series prediction, first, with the use of conventional static (feedforward) MLP neural network (with a single hidden perceptron layer) and, second, with the static quadratic neural unit (QNU), i.e., a class of polynomial neural network (or a higher order neural unit). We also demonstrate that QNU can be trained in a very efficient and fast way for real time retraining due to its linear nature of optimization problem. The objective is the prediction accuracy of 1 [mm] for 1-second prediction horizon. So it is well applicable for radiation tracking therapy.*

## **Keywords**

*Lung tumor-motion, time series prediction, radiation therapy, MLP, QNU, retraining.*

## **1. Introduction**

During lung-tumor radiation therapy, the radiation must be maintained in the region of tumor and it should be kept far away from the healthy tissue. However, this is a nontrivial task because of the body motion. The tumor motion can be associated with internal movements caused by cardiac cycles, and respiration, and also with patient's stochastic movements and systematic drifts [1] [2]. The respiration is dominant among them, thus respiratory motion has been widely studied. The respiratory motion becomes a complex non-stationary process (if the respiration changes amplitude and/or period over time over time).

In lung tumor motion it is well known to have amplitude between 0.5 and 2.5 cm, even some times 5 cm [4]. As a consequence, the dose distribution might be delivered significantly different from the prescribed one, and the radiation toxicity may increase dramatically [5] [6] [7] [8] [9] [10].

Several methods have been developed to model the respiratory motion gated radiation therapy or real time tumor tracking, but their use is still questioned [1][3]. Three general approaches have been achieved to predict respiration behaviour [3]. Biomechanical study of the breathing process is the basic approach. Other method consists of a respiratory mathematical model using harmonic functions. The most promising method is an approach based on learning algorithms which need to be trained with previously observed input-output patterns.

According to the above statements, feedforward neural networks have promising capabilities for implementation to lung motion time series prediction, and lung motion prediction with neural network (NN) is a subject of great interest in medicine due to the possibility of capturing dynamics and structural aspects [5] [3]. Some authors are convinced that deep analysis is still needed [5] [3] [11] [12].

Regarding the above issues of neural networks and our experience with higher-order nonlinear neural architectures [13] – [16] we extend our study on a second order neural unit so called quadratic neural unit (QNU). QNU can be considered a standalone second order neural unit of higher order neural networks (HONN) or a class of polynomial neural networks [17][18].

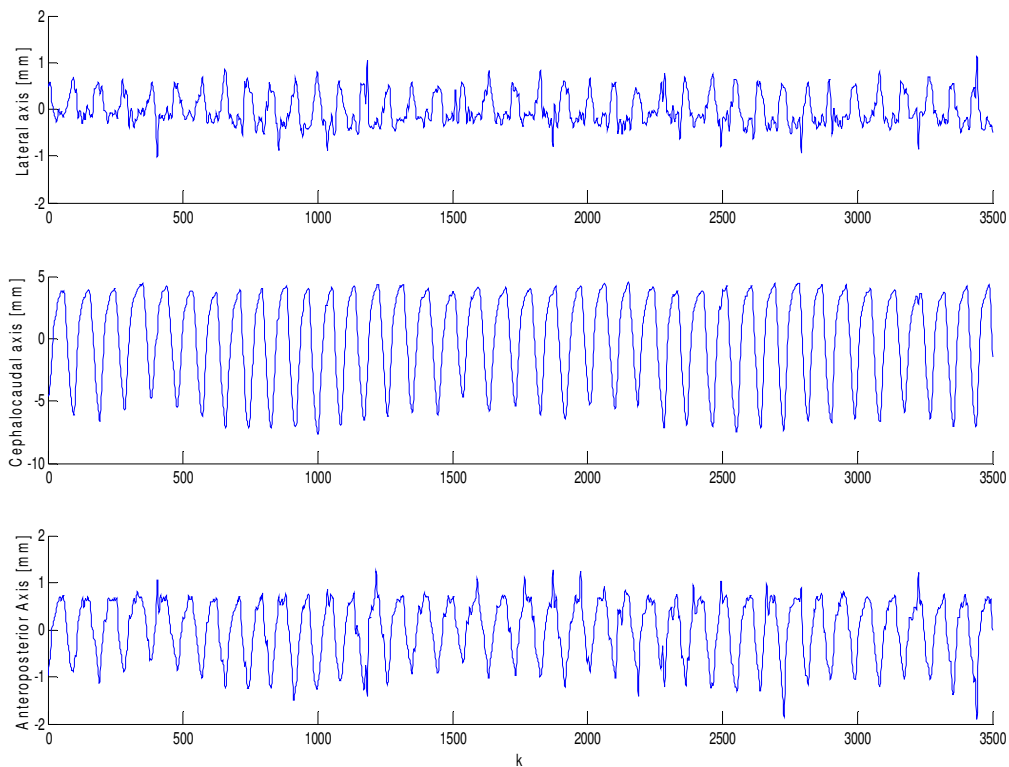
We study implementation of static neural networks (i.e. MLP and QNU), we use the most popular learning algorithm, i.e., the Levenberg-Marquardt (L-M) algorithm [19] [20] that is powerful optimization algorithm and it is easy to be implemented. L-M technique is used for non-linear least-squares problems. When the solution is far from the correct one, the algorithm behaves as a steepest descent method.

Also, because of the nonstationary nature of lung tumor motion in time, we implemented a sliding window retraining [21] to capture temporal variations in time series validity of the neural model at every sample of prediction.

In this paper, we propose a study of lung tumor motion time-series prediction, first, with the use of MLP neural network (with a single hidden perceptron layer) and, second, with the QNU. We also demonstrate that QNU can be trained in a very efficient and fast way for real time retraining. The objective of our study was to achieve the prediction accuracy bellow 1 [mm] for 1 second prediction horizon with our approach and to study capabilities of simplest yet powerful neural network models, i.e. static MLP networks and static QNU to achieve better prediction accuracy than in published and comparable works. The QNU was chosen for its high quality of nonlinear approximation and its excellence convergence [16] that is in this paper discussed in the light of its linear optimization nature(a unique minima for training).

## 2. Data description

The three-dimensional time series of lung tumor motion data were obtained from measurements by Hokkaido University Hospital [22][23].



**Figure 1.** Preprocessed time series of the observed tumor marker position of the lung.

The sampling period was 30Hz, and the spatial resolution was 0.01 [mm]. The time series were preprocessed in order to reduce the noise and avoid abnormal data included in rough data of the time series [22][24]. The preprocessed time series is shown in Figure 1.

$$\mathbf{y}(k) = [y_1(k) \quad y_2(k) \quad y_3(k)] \quad (1)$$

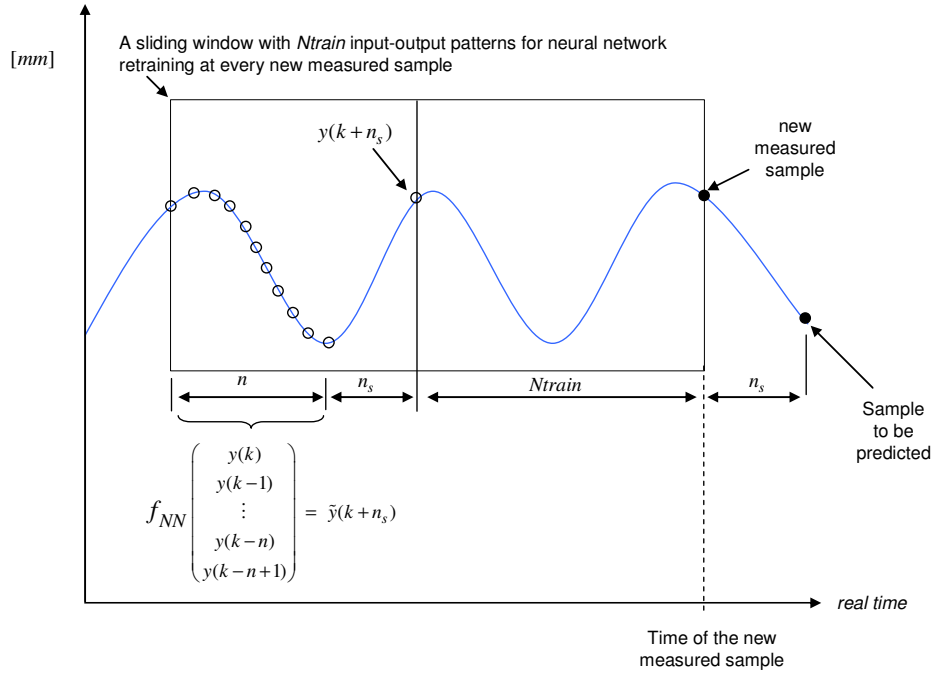
The vector  $\mathbf{y}(k)$  (eq. (1)) at discrete time  $k=1,2,\dots$ ,  $y_1(k)$ ,  $y_2(k)$  and  $y_3(k)$  are positions of marker on the lateral, cephalocaudal, and anteroposterior axes in [mm]. The dominant periods of the time series are approximately about 3 seconds.

### 3. Prediction Methods

#### 3.3 Sliding Window Retraining

Because the respiration time series are highly nonstationary (with time varying frequency, mean and amplitudes), it is impossible to obtain a generally valid model from a single training data set. Therefore, we investigate the effect of retraining of the MLP and the QNU predictive models to their prediction accuracy. We retrained the models with most recent history of measured values at every new measured sample, i.e. before each new sample prediction. This approach is sometimes referred in literature as a sliding window approach [21].

For every retraining we choose  $epochs = 30$  (for MLP) and  $epochs = 300$  (for QNU) as we could notice that the mean absolute error was not improved with more number of training epochs into the window especially for long term prediction. After the current window training is performed, the neural network predicts the unknown  $n_s$  samples ahead from the new measured value. The training window for the predictive models is shown in Figure 2.



**Figure 2.** A sliding (retraining) window for model retraining at every new measured sample, the window slides ahead with each new measured sample.

We used the well known L-M algorithm for the weight increments of  $i$ -th hidden neuron at every epoch of training.

In the L-M algorithm for MLP neural network,  $\mu$  can be automatically decreased or increased at every training epoch depending on the convergence of the training performance of the network, which is for  $N$  training samples given as sum of square errors

$$Q(\text{epoch}) = \sum_{k=1}^N e(k)^2, \quad \text{where } e(k) = y(k) - \tilde{y}(k), \quad (2)$$

and from trial searches we concluded the following scheme

$$\begin{aligned} \text{if } & Q(\text{epoch}) > Q(\text{epoch}-1) \text{ then } \mu = \mu/10 \\ \text{else } & \mu = \mu \cdot 10 \end{aligned} \quad (3)$$

The L-M algorithm for the perceptron-type network, requires computation of the Jacobian matrix  $\mathbf{J}_i$  at each epoch, so the matrix inverse has to be always calculated according to the basic L-M formula in epoch-times [19][20]. The inverse matrix calculation for the network results in slowing down the real-time computation.

In the QNU predictive model, the inverse during the calculation of the Jacobian Matrix is calculated only once before the training in epochs starts, and then we also calculate the weight updates only with varying  $\mathbf{e}$  that is the only vector that is recalculated at every epoch in the modified L-M formula [19][20]. Moreover, we concluded that it was not necessary to implement increasing learning rate for QNU (as it was desirable for MLP and that can be reasoned by the superior convergence of QNU during training due to its linear nature of optimization problem).

Also, we notice that with the use of least square method we could avoid the inverse during the calculation of the Jacobian matrix, so we highlight merely division for a single weight. The computation time is much faster because the increment of the weights can be calculated in a for loop using merely division rather than calculate all weight updates once by the original L-M formula with the inverse of a large matrix. Thus, the only vector that is recalculated every epoch in the modified L-M formula is the error  $\mathbf{e}$ .

## 4. Experimental Analysis

### 4.1 MLP

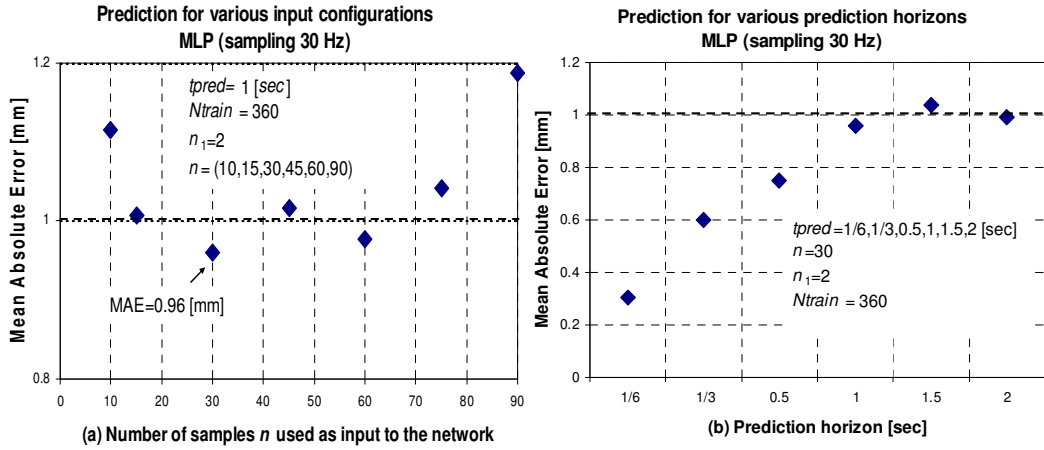
Experiments were carried out for 30 Hz and 15 Hz sampling. We investigated the effect of various setups of  $n$ ,  $n_s$ , and  $N_{train}$  to predict accuracy that was calculated as

$$MAE = \frac{1}{N} \sum_{k=1}^N \sqrt{e_1^2(k) + e_2^2(k) + e_3^2(k)} \quad (4)$$

where  $N$  is the number of testing samples, and  $e_1$ ,  $e_2$ , and  $e_3$  are the predicting errors of every axis 1, 2 and 3. For the predictive model applied on 30 Hz lung tumor motion recorded data, the prediction horizons and the different input configuration analysis are presented on Figure 3 (a)-(b). Figure 3(a) summarizes prediction errors at 1 [sec] (=30 steps) ahead by sliding window technique during testing for the time series. The results show prediction errors (MAE) for prediction horizon 1 [sec] when lengths of input vector were  $n = (10, 15, 30, 45, 60, 90)$  of  $\mathbf{Y}(k)$ . The number of input-output training patterns of the window is settled at  $N_{train}=360$ .

According to our testing, the prediction performs more precise results with  $n=30$  samples back from the time series taken as input to the neural network that are present in the current wave of the respiratory dynamics. For instance,  $n=30$  samples back of the time series includes useful information for the prediction, less or more inputs to the predictive model can affect badly to the prediction error, as can be seen in Figure 3(a).

Figure 3(b) shows the mean absolute error for prediction horizons  $t_{pred}=[1/6, 1/3, 0.5, 1, 1.5, 2]$  [sec] for a single length of input vector  $\mathbf{x}$  with  $n=30$ , see (3). The number of training samples for the windows were settled to  $N_{train}=360$ . As it can be seen in Figure 3(b), the prediction MAE becomes larger as the prediction horizons increase.

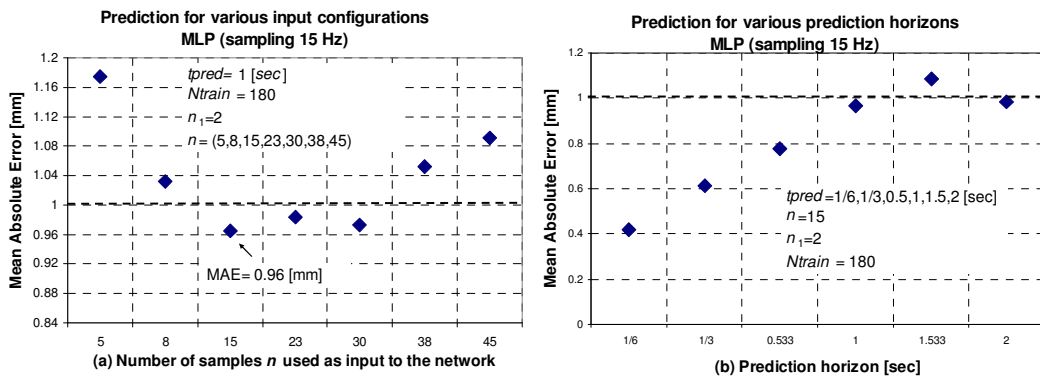


**Figure 3.** (a) The smallest MAE of 1 second prediction was obtained with  $n=30$  input samples to a 2-hidden-neuron neural network, sampling of 30 Hz). (b) MAE is increasing with prediction horizon  $t_{pred}=[1/6, 1/3, 0.5, 1, 1.5, 2]$  [sec].

The resulting time series prediction (30 Hz sampling) was carried out for  $k=1000$  to 2300 by MLP predictive model. The configuration of the model was  $n=30$  samples back used as inputs to the network,  $n_1=2$  neurons were used for the prediction, and  $N_{train}=360$  were considered as input-output training patterns for the sliding window. Thus, the MAE of the 0.5 [sec] prediction horizon was 0.75 [mm] with computing time of 304 [sec] and the MAE of the 1 [sec] prediction horizon was 0.96[mm] with 303 [sec] of computing time.

In the case of MLP predictive model applied on 15Hz sampling data, the predicting results are presented on Figure 4 (a)-(b). Figure 4(a) summarizes prediction errors at 1 [sec] (=15 steps) ahead by sliding window technique during prediction of time series. The results show prediction errors (MAE) for prediction horizon 1[sec] when lengths of input vector configuration were  $n=(5,8,15,23,30,38,45)$  of the time series signal  $\mathbf{Y}(k)$ . The number of input-output training patterns of the window is settled at  $N_{train}=180$ . In Figure 4(a) is shown that according to our testing for 1 [sec] prediction horizon, MAE = 0.96 [mm] was the smallest prediction error, which was obtained with  $n=15$  samples used as inputs to the network.

Figure 4(b) shows the MAE for prediction horizons  $t_{pred}=[1/6, 1/3, 0.533, 1, 1.533, 2]$  [sec] for a single length of input vector  $\mathbf{x}$  with  $n=15$ . The prediction MAE becomes larger as the prediction horizons increase as can be verified in Figure 4 (b).



**Figure 4.** (a)The smallest MAE of 1 second prediction was obtained with  $n=15$  input samples to a 2-hidden-neuron neural network, sampling of 15 Hz). (b)MAE is increasing with prediction horizon  $t_{pred}=[1/6, 1/3, 0.533, 1, 1.533, 2]$  [sec].

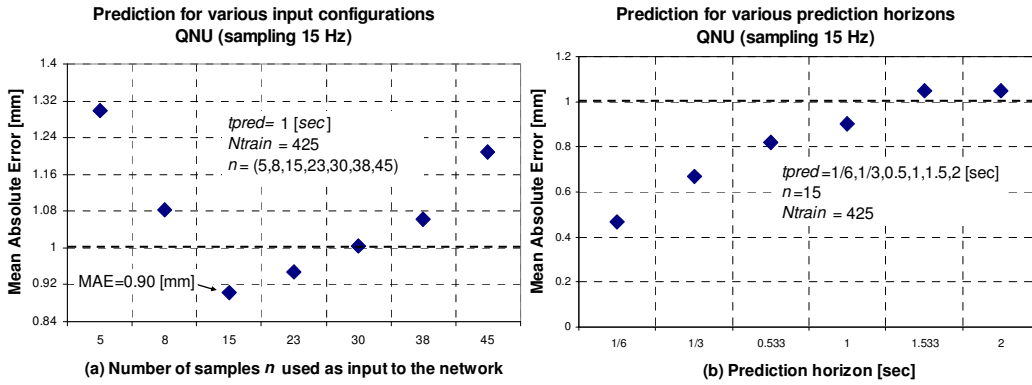
The time series prediction (15 Hz sampling) was carried out for  $k=500$  to 1150 samples by MLP predictive model. In our results, we highlight the prediction for  $tpred=0.5$  [sec] (i.e.  $n_s=8$ ) and  $tpred=1$  [sec] (i.e.  $n_s=15$ ) prediction horizons. The model was configured with  $n=15$  samples back used as inputs to the network,  $Ntrain=180$  for the sliding window, and  $n_1=2$ . Accordingly, the MAE of the 0.5 [sec] prediction horizon was 0.77 [mm] with 79 [sec] of computing time and the MAE of 1 [sec] prediction horizon was 0.96 [mm] with 97 [sec] of computing time.

## 4.2 QNU

In the QNU predictive model applied on 15 Hz sampling data, the predicting results are presented on Figure 5 (a)-(b). Figure 5 (a) shows prediction errors at 1 [sec] prediction horizon (= 15 steps ahead) by QNU adaptive model using sliding window technique during prediction of time series. The results present prediction errors (MAE) for prediction horizon 1 [sec] when lengths of input vector configuration were  $n= (5,8,15,23,30,38,45)$  of the time series signal  $Y(k)$ . The number of input-output training patterns of the window is settled at  $Ntrain=425$ . According to the prediction result for 1 [sec] prediction horizon, the smallest prediction error was MAE = 0.90 [mm] when we used  $n=15$  as inputs to the network.

Figure 5(b) presents MAE of prediction horizons  $tpred=[1/6, 1/3, 0.533, 1, 1.533, 2]$  [sec] for a single length of input vector  $x$  with  $n=15$ . The prediction MAE becomes larger as the prediction horizons increase (see Figure 5 (b)).

The prediction using the QNU predictive model (15 Hz sampling) was carried out for  $k=500$  to 1150 samples (43 [sec]). The model was configured with  $Ntrain=425$  input-output patterns for the sliding window,  $n=15$  samples back used as inputs to the network. Accordingly, the prediction results of 0.5 [sec] prediction horizon was MAE= 0.81 [mm] with duration of 18 [sec] of computing time for 43 [sec] of treatment time, and MAE = 0.90 for 1 [sec] prediction horizon with duration of 26 [sec] of computing time at 43 [sec] of treatment time as well.



**Figure 5.** (a) The smallest MAE of 1 second prediction was obtained with  $n=15$  input samples to a 2-hidden-neuron neural network, sampling of 15 Hz). (b) MAE is increasing with prediction horizon  $tpred=[1/6, 1/3, 0.533, 1, 1.533, 2]$  [sec].

## 6. Conclusion

In this paper, we have developed a time series predictive models for lung tumor radiation therapy. An MLP with one hidden layer and a QNU were developed as predictive models. The adaptation rule for the models was the batch optimization (L-M) implemented on retraining (sliding window) technique (i.e. a moving window for every new sample as a retraining method). Then we demonstrated the predictive capability of the models by the moving window on periodic and highly nonlinear respiratory time series for prediction of real data of lung movement. The prediction results obtained by the predictive models satisfies the goals of our work for the prediction accuracy of 1 [mm] with at

most 1 [sec] prediction horizon. Also, study and comparison of the MLP and QNU predictive model performances were presented. For the long term prediction horizon 1 [sec], MLP performed faster at 15 Hz sampling rate (97 [sec]) than 30 Hz sampling (303 [sec]) at 43 [sec] of treatment time, and with the same MAE=0.96 [mm]. Accordingly, QNU predictive model (15 Hz) achieved faster computing time (26 [sec]) at 43[sec] of treatment time comparing with MLP (15 Hz). Also, the prediction accuracy was improved for the QNU predictive model (MAE = 0.90 [mm]) comparing with MLP predictive model (0.96 [mm]). L-M technique was used for non-linear least-squares problems for the QNU adaptive model in order to avoid the inverse calculation of the common L-M algorithm. Accordingly, the QNU predictive model became computationally faster than the MLP predictive model and even faster than the treatment time (43 [sec]). The lung tumor motion predictions were achieved by the proposed predictive models based on the adaptation to the time variant period involved in the cyclic dynamics of respiration.

### Acknowledgment

This work was supported grant No. SGS10/252/OHK2/3T/12 and SGS12/177/OHK2/3T/12, and also by PROMEP Mexican grant number UTCJ001.

### Symbols

$t_{pred}$	prediction horizon in seconds	[sec]
$t$	denotes continuous time	[sec]
$k$	denotes discrete time	[sample]
$N_{train}$	number of input-output training patterns	[sample]
MAE	3-D Euclidean distance	[mm]

### References

- [1] Sharp G C, Jiang S B, Shimizu S and Shirato H 2004 Prediction of respiratory tumour motion for real-time image-guided radiotherapy *Physics Medicine and Biology* 49 425-40
- [2] Hoisak J D P, Sixel K E, Tirona R, Cheung P C F and Pignol J F 2006 Prediction of lung tumour position based on spirometry and on abdominal displacement: Accuracy and reproducibility *Radiotherapy and Oncology* 78 339-46
- [3] Murphy M J and Dieterich S 2006 Comparative performance of linear and nonlinear neural networks to predict irregular breathing *Physics Medicine and Biology* 51 5903-14
- [4] Ma L, Herrmann C and Schilling K 2007 Modeling and Prediction of Lung Tumor Motion for Robotic Assisted Radiotherapy *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems (San Diego, CA)* pp 189-194
- [5] Li X A, Keall P J, and Orton C G 2007 Respiratory gating for radiation therapy is not ready for prime time *Med. Phys.* 34 867-87
- [6] Yan H, Yin F -F, Zhu G -P, Ajlouni M and Kim J H 2006 Adaptive prediction of internal target motion using external marker motion: a technical study *Physics in Medicine and Biology* 51 31-44
- [7] Isaksson M, Jalden J and Murphy M J 2005 On using and adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications *Med. Phys.* 32 3801-9
- [8] Ruan D 2010a Kernel density estimation-based real time prediction for respiratory motion *Physics in Medicine and Biology* 55 1311-26
- [9] Ruan D 2010b Prospective detection of large prediction errors: a hypothesis testing approach *Physics in medicine and biology* 55 3885
- [10] Demachi K, Zhu H, Ishikawa M and Shirato H 2009 Predictive Simulation of tumor movement for Chasing Radiotherapy *Journal of the Japan Society of Applied Electromagnetics and Mechanics* 17 222-6.
- [11] Kakar M, Nystrom H, Aarup L R, Nottrup T J and Olsen D R 2005 Respiratory motion prediction by using the adaptive neuro fuzzy inference system(ANFIS) *Physics in Medicine and Biology* 50 4721-28
- [12] Riaz N, Shanker P, Wiersma R, Gudmundsson O, Mao W, Widrow B and Xing L 2009

- Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression *Physics in Medicine and Biology* 54 5735-48
- [13] Gupta M M, Liang J, and Homma N 2003 *Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory*, New Jersey: John Wiley & Sons, Inc.
  - [14] Gupta M M, Homma N, Hou Z-G, Solo M G and Bukovsky I, 2010 *Higher Order Neural Networks: Fundamental Theory and Applications*, in *Artificial Higher Order Neural Networks for Computer Science and Engineering: Trends for Emerging Applications*, ed. M. Zhang, IGI Global, ISBN 13: 978-1-61520-711-4, pp 397-422
  - [15] Bukovsky I, Bila J, Gupta M M, Hou Z-G and Homma N 2009 *Foundation and Classification of Nonconventional Neural Units and Paradigm of Nonsynaptic Neural Interaction in Discoveries and Breakthroughs in Cognitive Informatics and Natural Intelligence in the ACINI book series* ed. by Yingxu Wang, University of Calgary, Canada, IGI Publishing, Hershey PA, USA, ISBN: 978-1-60566-902-1
  - [16] Bukovsky I, Bila J 2010a *Adaptive Evaluation of Complex Dynamic Systems using Low-Dimensional Neural Architectures* Springer's book on *Advances in Cognitive Informatics and Cognitive Computing*, Series: *Studies in Computational Intelligence* Vol. 323, eds. D. Zhang, Y. Wang, W. Kinsner, ISBN: 978-3-642-16082-0, pp.33-37
  - [17] Ivakhnenko A G 1971, *Polynomial Theory of Complex Systems* *IEEE Tran. on Systems, Man, and Cybernetics* Vol. SMC-1 (4). pp. 364-378
  - [18] Nikolaev N Y and Iba H 2003 *Learning Polynomial Feedforward Neural Network by Genetic Programming and Backpropagation* *IEEE Trans. on Neural Networks* vol. 14 no. 2, pp.337-350, March
  - [19] Marquardt D W 1963 *An Algorithm for the Least-Squares Estimation of Nonlinear Parameters* 1963 *SIAM Journal of Applied Mathematics* vol 11 issue 2 pp. 431-41
  - [20] Moré J J 1978 *The Levenberg–Marquardt algorithm: implementation and theory* In *Numerical Analysis*, *Lecture Notes in Mathematics*, Watson GA (ed.), vol. 630. Springer: Berlin,105–16
  - [21] Morgado D F, Antunes A, Vieira J and Mota A (2004) *Implementing the Levenberg-Marquardt algorithm on-line: A sliding window approach with early stopping* 2nd IFAC Workshop on *Advanced Fuzzy/Neural Control*
  - [22] Homma N, Sakai M, Endo H, Mitsuya M, Takai Y and Yoshizawa M 2009a *A New Motion Management Method for Lung Tumor Tracking Radiation Therapy* *WSEAS Trans. Systems* 8 471-80
  - [23] Homma N, Sakai M and Takai Y 2009b *Time Series Prediction of Respiratory Motion for Lung Tumor Radiation Therapy* *Proceedings of the 10th WSEAS International Conference on Neural Networks* (Prague, Czech Republic) pp 126-31
  - [24] Ichiji K, Sakai M, Homma N, Takai Y and Yoshizawa M 2010a *Lung Tumor Motion Prediction Based On Multiple Time-Variant Seasonal Autoregressive Model for Tumor Following Radiotherapy* 2010 *IEEE/SICE International symposium on System Integration (SII)* (Sendai, Japan)