Construction of Multi-dimensional Arterial Health Status Map based on Molecular and Clinical Measurements, Fuzzy System and Data Cubes

Lawrence W.C. Chan, PhD1, Iris F.F. Benzie, PhD1, Thomas Y.H. Lau, PhD1, Yongping Zheng, PhD1, Alex K.S. Wong1, Y. Liu, PhD2, Phoebe S.T. Chan, PhD1
1Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong, China; 2Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Hong Kong, China

Abstract
Atherosclerosis results from inflammatory processes involving biomarkers, such as lipid profile, haemoglobin A1C, oxidative stress, coronary artery calcium score and flow-mediated endothelial response through nitric oxide. This paper proposes a health status coefficient, which comprehends molecular and clinical measurements concerning atherosclerosis to provide a measure of arterial health. An arterial health status map is produced to map the multi-dimensional measurements to the health status coefficient. The mapping is modeled by a fuzzy system embedded with the health domain expert knowledge. The measurements obtained from the pilot study are used to tune the fuzzy system. The inferred arterial health coefficients are stored into the data cubes of a multi-dimensional database. Due to this adaptability and transparency of fuzzy system, the health status map can be easily updated when the refinement of fuzzy rule base is needed or new measurements are obtained.

Introduction
The development of atherosclerosis involves multiple stages of endothelial dysfunction, inflammation, apoptosis and matrix alteration at the arterial walls, leading to reduced blood flow, vascular occlusion and thrombosis. As a common diabetic complication, atherosclerosis is further accelerated by diabetes mellitus through the glucose-mediated vascular damage predisposing to the development of vascular diseases such as stroke and coronary heart disease. According to the figures provided by Diabetes Hong Kong, a charitable voluntary organization founded in September 1996, there are about 0.7 million Hong Kong people having diabetes mellitus which comprise one-tenth of the total population in Hong Kong as of 2006. Biomarker profiling, including the measurement of plasma ascorbic acid (AA), Ferric Reducing Ability of Plasma (FRAP), Low Density Lipoprotein (LDL), Complete Blood Count (CBC) and C-Reactive Protein (CRP), can be used to assess the risk of diabetes-accelerated atherosclerosis. Further, analytical approaches were proposed to investigate atherosclerosis with respect to the thermodynamics, haemodynamics and mass transfer physical models. The role of haemodynamics in predicting the atherosclerosis using Doppler and B-mode ultrasound was also highlighted by a number of research studies. It is important to aggregate all these molecular and clinical measurements to assess the extent of atherosclerosis and thus the arterial health.

Statistical tests are commonly used to evaluate the association between the variables. However, the variation of one variable against the other cannot be illustrated through the test. Fuzzy systems were widely used for classification, modeling and reasoning of data. Having the linear parameterized structure, the fuzzy systems can be further extended to be a multiple regression of nonlinear functions. Excellent transparency and adaptability of fuzzy systems have been proved in biomedical and engineering applications. Furthermore, the emerging technology of multi-dimensional database facilitates the storage and retrieval of the multi-factorial arterial health status indicator through data cubes, forming an arterial health status map.

This paper proposes a novel three-stage approach which consists of the key feature identification using the Pearson’s correlation test, the fuzzy modeling and reasoning using the fuzzy system, and the production of arterial health map using data cubes.

Materials and Methods
A. Subjects
The pilot study is recruiting Type II diabetes patients, whose age between 46 and 60 years, non-smoking and without any records of stroke and chronic coronary heart disease from Diabetic Mutual Aid Society (DMAS) of Hong Kong. The data of 34 subjects have been collected. The data were split into the training dataset of 11 data points and the testing dataset of 23 data points. The mean age of these subjects is 54.2 years (SD 4.5, range 46.6-60.4).
These subjects consist of 12 males and 22 females. Six of them were identified by a radiologist as having atheromas at the common carotid artery, the internal carotid artery and the bifurcation. No carotid vascular problem was found in the rest of the subjects.

B. Data Collection

Data are collected from the medical laboratory tests of the fasting blood samples, extracranial carotid sonography and transcranial Doppler sonography of the subjects. Prior to the ultrasound examinations, the systolic and diastolic blood pressure, weight and height of the subjects were measured.

B1. Medical Laboratory Tests

Fasting blood samples were collected, and a panel of biomarkers as reflecting risk of vascular disease in these subjects were measured. The measured biomarkers include complete blood count (CBC) of white blood cells, red blood cells and platelets, plasma glucose, total cholesterol, HDL-cholesterol, LDL-cholesterol, triglycerides, high sensitivity C-reactive protein (hsCRP: a marker of inflammation), haemoglobin A1C (HbA1C), and a panel of markers of oxidant-to-antioxidant balance, including FRAP, plasma AA and uric acid.

B2. Extracranial Carotid Sonography

Pulsed wave and color flow Doppler and B-mode ultrasound is a safe and clinically useful signaling and imaging technology to quantify the presence of carotid arteries steno-occlusive disease, narrowing and intimal-media thickening. B-mode ultrasound scan images are able to predict and classify carotid atherosclerotic plaques and thus assess the risk for stroke. In the examinations, the transducer was moved cranially along the artery to the proximal portion of internal (ICA) and external carotid (ECA) arteries to check for any stenosis, narrowing, or intimal thickening. Then, the transducer was rotated 90 degrees to align with the CCA. The CCA, ICA and ECA were then examined starting from the origin of the CCA to the proximal portion of ICA and ECA. All the blood flow velocities at the proximal, middle and distal portions of the CCA, proximal portions of ICA and ECA were measured. Similarly, Doppler scanning will be performed on the contralateral CCA with the blood flow velocities be measured.

B3. Transcranial Doppler Sonography

Transcranial Doppler (TCD) ultrasound is able to assess atherosclerotic ischemic stroke and intracranial pressure. Blood flow velocities of both left and right Middle Cerebral Arteries (MCAs) were recorded and a 2 MHz transducer were used and placed on the temporal bone window, for the best insonation to obtain flow velocity waveform of MCA. The highest blood flow velocity was sought through the temporal window at a depth between 45-66 mm.

C. Feature Extraction

The molecular measurements and the medical laboratory test readings of blood samples are directly regarded as part of the features. The others involve indirect calculation using the clinical measurements. The Body Mass Index (BMI) is defined as,

\[
BMI = \frac{W}{H^2}
\]

where \( W \) is the subject’s weight and \( H \), height.

B-mode color flow Doppler images can be acquired during the extracranial carotid sonography examination. The B-mode ultrasound images show the morphology of carotid artery and facilitate the measurement of lumen diameter \( d \) and the thickness of atheroma \( h_a \) with the aid of color Doppler flow patterns. Waveform signals of blood flow velocities were obtained from the extracranial carotid sonography and the transcranial Doppler sonography examinations, Peak Systolic Velocity, End Diastolic Velocity and Time Average Velocity can be measured from the waveform signals.

Some important features can be derived from the above measurement, including Reynold number, resistance index, EDV ratio and percentage of luminal diameter reduction.

C1. Reynold number

Reynold number \( (Re) \) is dimensionless feature reflecting the arterial blood flow condition given by,

\[
Re = \frac{\rho du}{\eta}
\]

where \( \rho \) is the density of blood \((1 \text{ g/cm}^3)\), \( \eta \) is the viscosity of blood \((3.5 \times 10^{-3} \text{ Pa sec})\), \( d \) is the diameter of lumen of artery and \( u \) is the mean of PSV across the lumen. Turbulence is most likely established when the Reynold number is greater than 2000. The occurrences of turbulence can usually be found at the atheroma-prone sites, such as carotid bifurcation, vessel curvature and arterial branching. The mean of \( Re \) on the left and right was used for the data analysis.

C2. Resistance Index

Resistance Index \( (RI) \) is a major blood flow parameter derived from the Doppler frequency spectrum where the errors introduced by the Doppler angle and also the other possible instrument-dependent effects are normalized. These three parameters are defined by,
\[ RI = \frac{(PSV - EDV)}{PSV} \]  

where \( V \) is the time average of blood flow velocity over the cardiac cycle. The mean of each parameter on the left and right was used for the data analysis.

### C3. End Diastolic Velocity Ratio

Symmetry of haemodynamics could be diagnostic criteria for vascular diseases. Asymmetric blood flow implies high possibility of unilateral stenosis. End Diastolic Velocity (EDV) ratio in CCA proved to be an additional indicator for predicting unilateral haemodynamically significant stenosis in ICA. EDV ratio is defined as the side-by-side ratio of the end diastolic blood flow velocity at CCA given by,

\[ EDV \text{ ratio} = \frac{\max(EDV_{LCCA}, EDV_{RCCA})}{\min(EDV_{LCCA}, EDV_{RCCA})} \]  

where \( EDV_{LCCA} \) and \( EDV_{RCCA} \) are the EDVs in left and right CCA respectively.

### C4. Percentage of Luminal Diameter Reduction

Vascular laboratories make use of various criteria for the diagnosis of arterial disease, in which luminal diameter reduction is one of their standards. The percentage of luminal diameter reduction (%LDR) is defined by the following equation.

\[ \%LDR = \frac{h_a}{d} \]  

where \( h_a \) and \( d \) are the thickness of atheroma causing the reduction of luminal diameter and the radiologist’s estimate of the original luminal diameter at the site of atheroma respectively. If atheroma is not found, %LDR will be assigned a zero value.

### D. Statistical Analysis for Key Feature Identification

Since over 50 measured values and derived features were available for the evaluation, we performed correlations test among them in order to identify key features for further analysis and fuzzy reasoning.

Two-tailed Pearson’s test was applied to find out the correlation \( r \) and its significance \( p \) between the extracted features and the percentage of luminal diameter reduction (%LDR). The standard Pearson Product-Moment correlation \( r \) is defined by the following formula.

\[ r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{nS_XS_Y} \]  

where \( \overline{X} \) and \( \overline{Y} \) are the sample mean of \( X \) and \( Y \), the two variables to be correlated; \( S_X \) and \( S_Y \), the sample standard deviation of \( X \) and \( Y \); \( n \) is the number of \( (X,Y) \) pairs. The Bivariate Correlations procedure provided by the statistical software SPSS was used to compute the coefficient. Features with high correlation with %LDR yet reaching high significant levels are considered to be key features. Features having \( p \) value <0.05 were denoted by (*).

### E. Analytical Model using ANFIS

Fuzzy system is comprised of linguistic fuzzy rules representing the relationship between input and output variables. Adaptive Neuro-Fuzzy Inference System (ANFIS) is a kind of fuzzy system where the input is fuzzified but the output is crisply defined. Due to the parameterized structure and the linguistic representation, the modeling ability and transparency of ANFIS are superior to the other artificial intelligence tools. A typical fuzzy rule in ANFIS with zero-order Sugeno mode is given by,

Rule \( i \): If \( x \) is \( A_i \) and \( y \) is \( B_i \) then \( z = k_i \)

where \( x \) and \( y \) are input variables, \( A_i \) and \( B_i \) are their corresponding fuzzy sets, for example “Low”, “Medium” and “High”, \( z \) is output variable, and \( k_i \) is its corresponding crisply defined value of the \( i^{th} \) fuzzy rule. This linguistic rule can express any non-linear relationship between input and output variables.

By integrating all the fuzzy rules covering the whole input domain and output range, ANFIS can be formulated to a network structure comprised of input and output membership functions and the logical operators between them. Fuzzy Logic Toolbox of Matlab® is used to build and train the ANFIS in this study. The key features selected from the extracted features in the statistical test were the input of ANFIS. The output was the percentage of luminal diameter reduction. The ANFIS represents an analytic model formulating the relationship between the selected key features and the degree of atherosclerosis. After the fuzzy system was trained using the training dataset, the represented analytical model was validated using the testing dataset. The arterial health status coefficient is defined as the complement of the percentage of luminal diameter reduction, i.e. 1-%LDR.

### F. Arterial Health Status Map

Data cubes act as elementary logical memory units of the arterial health status map. The mapping from the key features to the corresponding arterial health status coefficient is logically stored by the data cubes.

### Results and Discussion

#### A. Correlation Test Results
The following table shows the Pearson's correlation coefficients of features against %LDR and their significance levels (Table 1).

| Feature                  | r    | p   |
|--------------------------|------|-----|
| Mid. CCA Re              | 0.239| 0.480|
| Mid. CCA RI              | -0.433| 0.183|
| Mid. CCA EDV ratio       | 0.764 | 0.010|
| White blood cell         | 0.708 | 0.015|
| Neutrophils              | 0.621 | (*) 0.042|
| Lymphocyte               | 0.345 | 0.299|
| Eosinophils              | 0.403 | 0.219|
| Monocytes                | 0.283 | 0.399|
| Basophils                | 0.587 | 0.058|
| Platelets count          | 0.316 | 0.344|

**Table 1. Correlation of features against %LDR**

Among the features shown in Table 1, the EDV Ratio in middle CCA and Complete Blood Count (CBC) are significantly correlated with %LDR. Therefore the EDV ratio and WBC were chosen as key features and thus an input variable of the adaptive neuro-fuzzy inference system (ANFIS). As the features may be correlated with each other, it is suggested to adopt the Principal Component Analysis (PCA) to further enhance the accuracy of the analytical model.

B. Analytical model using ANFIS

Two key features, EDV ratio and WBC, selected by the above-mentioned statistical analysis are considered to be the input variables of adaptive neuro-fuzzy inference system (ANFIS). The output variable of the ANFIS is the %LDR which assembles the size of atheroma and also the extent of atherosclerosis. The following figure illustrates the mapping between the biomedical features and the %LDR through the ANFIS (Figure 1).

![ANFIS diagram](image)

**Figure 1. ANFIS relates features and %LDR.**

Three fuzzy membership functions were used to represent the “Low”, “Medium” and “High” values of each feature. Nine fuzzy rules were generated to take into accounts all combinations of membership functions of WBC and EDV ratio. The ANFIS is the integration of these nine fuzzy rules. The values of WBC, EDV ratio and %LDR of the subjects were used to adjust the parameters $k_i$ of the ANFIS in the training process. The following table illustrates the resulting rules obtained by the training process (Table 2). The testing error, which is defined as the root mean squared difference between the ANFIS output and the %LDR of the testing data, is 0.0963. If the two least significant features, MCCA Re and Monocytes, constituted the input variables of ANFIS, the testing error is 0.1578. It was found that a more generalized analytical model can be obtained if highly significant features are selected.

| Fuzzy Rule | WBC Ratio | EDV Ratio | %LDR |
|------------|-----------|-----------|------|
| 1          | Low       | Low       | 0    |
| 2          | Low       | Medium    | 0    |
| 3          | Low       | High      | 0.241|
| 4          | Medium    | Low       | 0    |
| 5          | Medium    | Medium    | 0    |
| 6          | Medium    | High      | 0    |
| 7          | High      | Low       | 0.739|
| 8          | High      | Medium    | 0    |
| 9          | High      | High      | 0.211|

**Table 2. Resulting ANFIS rules after training process**

C. Arterial Health Status Map

By dividing the ranges of WBC and EDV ratio into 99 intervals respectively, 10000 grid points in a two-dimensional space are obtained. These grid points are fed into the fuzzy system and the corresponding output values are generated. The arterial health status coefficient, which is the complement of the fuzzy system output, is stored in a data cube for each EDV ratio and WBC pair. The following figure shows a two-dimensional arterial health status map produced by the pilot study (Figure 2).

![Arterial Health Status Map](image)

**Figure 2. Arterial Health Status Map**
Conclusion

This paper proposes a three-stage approach for the key feature identification, the modeling of the relationship between the selected features and the extent of atherosclerosis using ANFIS, the construction of arterial health status map using data cubes. The white blood cell count and the end diastolic velocity ratio were found to be associated with the formation of carotid atheromas. The analytical model using ANFIS shows the relationship between these features and the extent of atherosclerosis. The two-dimensional arterial health status map with respect to these features is also constructed.

References

1. Dzau VJ, Braun-Dullaeus RC, Sedding DG. Vascular proliferation and atherosclerosis: New perspectives and therapeutic strategies, Nature Medicine 2002;8(4):1249-1256.
2. Brownlee M. Biochemistry and molecular cell biology of diabetic complications, Nature 2001;414:813-820.
3. Choi SW, Benzie IFF, Lam CSY, Chat SWS, Lam J, Yiu CH, Kwan JJ, Tang YH, Yeung GSP, Yeung VTF, Woo GC, Hannigan BM, Strain JJ. Inter-relationships between DNA Damage, Ascorbic Acid and Glycaemic Control in Type 2 Diabetes Mellitus, Diabetic Medicine 2005;22:1347-1353.
4. Benzie IFF, Strain JJ. The Ferric Reducing Ability of Plasma (FRAP) as a Measure of “Antioxidant Power”: The FRAP Assay, Analytical Biochemistry 1996;239:70-76.
5. Kullo IJ, Ballantyne CM. Conditional Risk Factors for Atherosclerosis, Mayo Clinic Proceedings 2005;80(2):219-230.
6. Tsimikas S. Emerging biomarkers of instability, In R. Virmani, J. Narula, M. B. Leon, J. T. Willerson (Eds.), The Vulnerable Atherosclerotic Plaque: Strategies for diagnosis and management, Malden, Mass.: Blackwell Futura, 2007; 187-210.
7. Wang HH, Wang XF. Analytical Methods of Atherosclerosis Research, In M. Schoenhagen (Ed.), Current Developments in Atherosclerosis Research, Hauppauge, N.Y.: Nova Science, 2006; 33-66.
8. Kamouchi M, Kishikawa K, Okada Y, Inoue T, Ibayashi S, Iida M. Reappraisal of Flow Velocity Ratio in Common Carotid Artery to Predict Hemodynamic Change in Carotid Stenosis, American Journal of Neuroradiology 2005;26(4):957-962.
9. Gaitini D, Soudack M. Diagnosing Carotid Stenosis by Doppler Sonography, Journal of Ultrasound in Medicine 2005;24(8):1127-1136.
10. Tan KS, Tan CT, Wong KS. Applications of transcranial Doppler ultrasound in atherosclerotic ischemic stroke: An Asian perspective, Neurology Asia 2005; 1-5.
11. Ubeyli ED, Guler I. Adaptive neuro-fuzzy inference systems for analysis of internal carotid arterial Doppler signals, Computers in Biology and Medicine 2008;35(8): 687-702.
12. Chan CW, Chan WC, Jayawardena AW, Harris CJ. Structure selection of neurofuzzy networks based on support vector regression, International Journal of Systems Science 2002;33(9): 715-722.
13. Brown M, Harris CJ. Neurofuzzy adaptive modelling and control 1994; New York: Prentice Hall.
14. Lal BK, Hobson RW, Pappas PJ, Kubicka R, Hameed M, Chakhtura EY, Jamil Z, Padberg FT, Haseer PB, Durán WN. Pixel distribution analysis of B-mode ultrasound scan images predicts histologic features of atherosclerotic carotid plaques, Journal of Vascular Surgery 2002;35(6): 1210-1217.
15. Denzel C, Fellner F, Wutke R, Bazler K, Müller KM, Lang W. Ultrasonographic analysis of arteriosclerotic plaques in the internal carotid artery, European Journal of Ultrasound 2003;16(3): 161-167.
16. Christodoulou CI, Pattichis CS, Pantziaris M, Nicolaides A. Texture-Based Classification of Atherosclerotic Carotid Plaques, IEEE Trans. Medical Imaging 2003;22(7): 902-912.
17. Bellner J, Rommer B, Reinstrup P, Kristiansson KA, Ryding E, Brandt L. Transcranial Doppler Sonography Pulsatility Index (PI) Reflects Intracranial Pressure (ICP), Surgical Neurology 2004;62(1): 45-51.
18. Neumyer MM. Vascular Technology, In R. A. Curry, B. B. Tempkin (Eds.), Sonography: Introduction to Normal Structure and Function 2007; St. Louis, Mo.: Saunders 480-496.
19. Chen PY, Popovich PM. Correlation – Parametric and Nonparametric Measures 2002; Thousand Oaks, CA: Sage.
20. Rafanelli M. Basic Notions, In M Rafanelli (Ed.), Multidimensional Databases: Problems and Solutions 2003; Idea Group Publishing 1-45.
21. Jang JSR, Sun CT, Mizutani E. Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence 1997; Upper Saddle River, NJ: Prentice Hall.
22. Jolliffe IT. Principal component analysis 2002; New York, NY: Springer.