Medical Knowledge Mining from Image Data – Synthesis of Medical Image Assessments for Early Stroke Detection

Waldemar W. Koczkodaj[1], Artur Przelaskowski[2], Kazimierz T. Szopinski[3]
[1]*Corresponding author: wkoczkodaj@cs.laurentian.ca*
[1]*Laurentian University, Department of Mathematics and Computer Science Sudbury ON P3E 2C6*
[2]*Institute of Radioelectronics, Warsaw University of Technology, Poland*
[3]*Department of Diagnostic Imaging, Second Faculty of Medicine Medical University of Warsaw, Poland*
[3]*Interdisciplinary Centre for Mathematical and Computational Modelling University of Warsaw, Poland*

Abstract. The key issue of this study is synthesis of medical images and expert knowledge for early detection of a medical condition, such as stroke or cancer. Such synthesis is a missing link for making decisions during the diagnostic process. Knowledge mining in image databases can be enhanced by computing the relative importance of image features using pairwise comparisons. Computed weights can be systematically used for synthesis of various image features present in the same or different images.

Key words: knowledge mining, image data, pairwise comparisons, inconsistency analysis, early stroke detection.

1. Introduction

The development of medical imaging systems in the past century was truly revolutionary. The main reasons for this amazing progress were considerable advances in digital image formation and the computer technology. However, image interpretation methods have only recently begun to benefit from the advances in computer technology. The correct and reasonably fast interpretation of collected image data is still the bottleneck. Referral of medical images to the clinician's attention takes place mostly after a manual analysis by a radiologist based on his/her experience and cognitive intuition. The interpretation of most medical images is still almost exclusively the task of human experts.

Our contribution is expected to serve as a helpful tool for radiologists. Sensitive recognition, specific understanding and accurate assessment of full imaging contents is a key issue for successful early detection of a medical condition and for a diagnostic process. For this purpose, it is necessary to extract crucial data from medical images, characterize and objectify human observers, compile all numerically accessible data, and consider physical and technological conditioning of currently available modalities.

We propose synthesizing all individual pieces of image data into one simple indicator which can be used as a red flag to pay more attention to the patient by possibly running more tests or by carrying out a physical examination of the patient. Computer-supported radiological interpretation will be more reliable and objective, repetitive, time- and cost-effective. The relation-based ontological models useful for this study are illustrated in [9]. The introduction of CBIR by Kato in 1992 [4] was recognized as an essential step forward in image management. Content-based access methods offer more than simple text-based queries. Image retrieval is based on similarity matching of diagnostic image content. It requires semantic descriptors synthesized on a higher abstraction level of human diagnostic procedures.

Additional medical knowledge completing or objectifying expert observations is useful for distinguishing or assessing human behaviors. Consensus among diversified opinions is possible through new knowledge extraction and formulation methods, etc. However, knowledge extracted by data processing and knowledge inference procedures is still synthesized at the human semantic level.

Synthesis of information from different imaging methods still poses a huge problem. Search for a consensus among experts' activities becomes extremely useful for diagnostic knowledge mining. The data mining framework combines elements from image feature analysis, selection, clustering and classification for accessible basic and modern modalities. On the other hand, expert opinions are collected and compared with one another. However, all these important pieces of knowledge have to be put together by manual interpretation.

In this study, we postulate to use pairwise comparisons for putting together scattered pieces of medical images and medical expertise. The purpose of our paper is to present a method supporting medical knowledge mining from image data. The proposed method adjusts the concept of consistency-driven pairwise comparisons to that challenging imaging problem. Our approach is general, and can be used for any problem where diagnostic criteria or conditions can be related to each other in pairs.

For the illustration of our approach we have chosen the stroke since it is not only a common condition and one of the largest contributors to hospital care costs, but is also hard to recognize even from medical images at its acute stage. Early stroke detection may save lives and considerable hospital costs at the national scale.

The basic contribution of our submission is the innovative concept of objectifying subjective evaluations, widely practised in medicine, especially in radiology, in order to solve the rather challenging task of supporting the diagnostic process by data and image mining.

2. Diagnostic criteria of acute ischemic stroke

Stroke is the clinical syndrome of rapid onset of a focal, or sometimes global, cerebral blood-flow deficit with a vascular cause lasting more than 24 hours or leading to death. It is the third leading cause of death and one of the leading causes of adult disability.

Recently, the introduction of thrombolysis has increased the role of early imaging. The rule "time is brain" has led to new emphasis on hyperacute changes in plain computed tomography (CT), providing anatomic information for indicating acute cerebral ischemia, or excluding lesions that produce symptoms or signs mimicking those of a stroke, such as hemorrhage and neoplasms. Additionally, the growing use of more complex imaging modalities providing physiologic information, has been noticed in the diagnosis of an in acute stroke. Significant emphasis has been placed on the delineation of the ischemic penumbra, also called tissue at risk (see the examples presented in Fig. 1).

The early stroke detection approach is based on synthesis of the effects of computer support, human observations based on imaging modalities, and clinical conditions. The assessment method can be applied either to subjective estimation of diagnostic methods efficiency based on personal experience, or to measurable data - e.g. sensitivity values obtained by clinical trials.

Stroke can be identified by a number of criteria or characteristics of pre-stroke conditions. Such features include image-based symptoms of stroke involving different neuroimaging modalities, clinical symptoms ordered according to the respective scales and results of routine screening, laboratory and other neurological tests. All of them are related to the time elapsed after onset of ischemia because time is brain [10].

Needless to say, these criteria do not equally contribute to a successful identification of an ischemic problem. Establishment of the relative importance of the individual features is not a trivial task, especially if there are many of such features. Numerical, quantitative assessment of feature importance is a challenging problem. By common sense, it is easier to perform when we compare these features and conditions in pairs than by looking at all of them together.

Our experiences in acute stroke knowledge mining are based on long term research directed at extraction of diagnostically important information from unenhanced CT image data. The designed tool provides more sensitive visualization of tissue density distribution [14]. However, the most important problem in practice is how to use this extracted knowledge for clinical diagnosis or how to merge the computed information with the imaged symptoms, other modality or laboratory test results. Regretfully, there is no reference method for "putting everything together". There is no uniform opinion of radiologists to which degree the results of different tests and images influence the diagnosis. To our knowledge, there is no widely accepted reference method to synthesize subjective medical expertise. Our study attempts to introduce a better method of diagnosis based on integrated indexes of many available images and tests. It is a data mining tool for

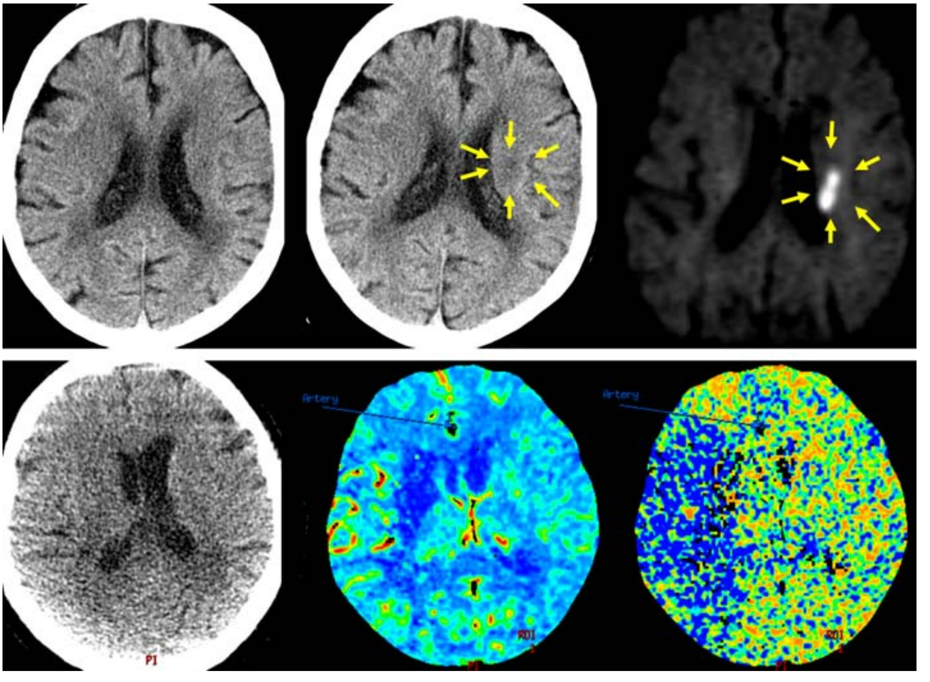

Fig. 1. Examples of ischemic stroke imaging: **top raw case** is early CT scan (7 hours after onset of ischemia) with hidden symptoms, next follow-up scan of the case (51 hours after early examination) with subtle hypodensity manifestation (indicated with arrows) and MR diffusion (DWI) scan (51.5 hours after early examination) with clearly noticeable infarct area; **bottom raw case** is early examination (3 hours after onset of ischemia) with hidden symptoms in noncontrast CT scan and clear manifestation of ischemia in perfusion-weighted CT showed a perfusion deficit in the mean transit time (MTT) parameter maps (middle) and on the regional cerebral blood volume (CBV) map with infarct (right).

an image database that helps combine partial data into one index which may not be decisive by itself but can serve as a medical indicator, or a red flag, for looking at a specific case more closely.

3. The method: pairwise comparisons of acute stroke criteria

The proposed solution is a method consisting in pairwise comparisons of all accessible diagnostic criteria. Such a method is particularly useful when we need to form our decision through subjective assessments based on expert experience. The sensitivity values obtained in prior clinical trials and other robust data have to be taken into consideration by experts formulating their opinions, if such are available.

The pairwise comparisons method has been brought to the attention of the multimedia, machine graphics and vision community in [13] but in the context of visual object classification. Our novel proposal is to apply this method to improve medical diagnosis through objectified and more orderly assessment of the provided information and the extracted knowledge. For that purpose, a hierarchical structure of complete, and as formal as possible, domain knowledge regarding the medical problem should be established. Thus, a hierarchical model with criteria used to extract, merge and assess completely accessible knowledge for reliable stroke identification has been proposed.

3.1. Fundamentals of pairwise comparisons

The pairwise comparisons method was used, presumably for the first time, by Condorcet in 1785 [1] for his election method in which voters rank candidates in the order of preference. However, it was Fechner who specified pairwise comparisons as a scientific method in 1860, although only from the psychometric perspective (see [2]). Thurstone provided a mathematical analysis of this method, and called it the "law of comparative judgments" (LCJ) in 1927 (see [3]). LCJ can be used to scale a collection of stimuli based on simple comparisons between the stimuli taken two at a time. Although Thurstone referred to it as a law, it can be more appropriately identified as a measurement model which could be of great use for machine vision. The model allows us to synthesize diverse features involved in machine vision, such as human identification problem, seismic incident prediction etc..

The hierarchy reduces the number of comparisons from $\Theta(n^2)$ to $\Theta(n \ln n)$, making it finally applicable to a wide variety of problems. For example, a moderate case with 49 features would require 1,176 comparisons without a hierarchy and only 168 comparisons if these 49 features are arranged into a hierarchy by dividing them into group of seven features each.

Measurements of length or by mass/weight are commonly used and never questioned. We have become so accustomed to having standards that sometimes we find it difficult to imagine anything without a standard measure. In the case of stroke conditions, it may not be a single condition but several of them, often involving sophisticated medical images. We may not be able to come up with a yard stick for stroke but we may still express the relative importance of one image compared to another image.

Intuitively, it is obvious that the "two at a time" approach has a chance to be more accurate than the method of assessing "everything at once". However, to show that the pairwise comparisons method is superior to the common sense "by an expert's eye" approach is not entirely a trivial task since there are many hurdles to overcome. At the current stage of pairwise comparisons theory, it is impossible to prove, or disprove, by analytical means which method is superior. The necessity of using computer technology for Monte Carlo experiments in [6] and [8] may explain why the problem of accuracy was not properly addressed in the 1950s or 1960s, when most of the theoretical work on the pairwise comparisons method was done. However, the drop in the estimation error for lengths of randomly generated bars from approximately 15% to 5% [6, 8] is a clear indication of potential gains in precision through using pairwise comparisons.

In the context of stroke, the introduction of a hierarchical structure can express fundamental knowledge used for acute stroke diagnosis. Image-based diagnostic information completed with clinical conditions is analyzed, interpreted and evaluated within restrictive time limitations.

We will denote these conditions, information or criteria for recognizing stroke-like objects by $A1$, $A2$, ..., An. The pairwise comparison method does not impose any limit on the number of diagnostic criteria (as stimuli compared to each other). A widely accepted heuristic is setting the maximum on one level to seven, since seven items give 21 distinct pairs to compare.

3.2. Analysis of stroke criteria

Among the nine early signs of brain infarction in CT images, there is only one quantitative sign: decreased attenuation (measured in Hounsfield units) in basal ganglia [13]. There are two semi-quantitative signs: Alberta Stroke Program Early CT Score (ASPECT) value, and hypoattenuation in thirds of middle cerebral artery territory. The remaining six signs (obscuration of lentiform nucleus, cortical sulcal effacement, focal hypoattenuation, loss of insular ribbon, hyperattenuation of vessel, and loss of gray and white matter differentiation in the basal ganglia) are qualitative.

Identified signs of stroke and examination conditions are the criteria of subjective disease assessment. Other important criteria are laboratory and clinical quantitative markers, and artery occlusion, sinus or venous thrombosis and penumbra/infarct areas recognized in other imaging modalities.

Tab. 1 illustrates our proposal for a hierarchical model with the criteria used to extract fast and completely accessible knowledge, and merge it for reliable stroke identification and accurate treatment ordering.

It is not clear which parameter would be most suitable in grading the importance of diagnostic factors. The most often used diagnostic accuracy measure is the Receiver Operating Characteristics (ROC)-based test of pathology detection. The determined parameters of sensitivity, specificity, accuracy or predictive values reflects the usefulness of diagnostic methods in clinical practice. As far as early stroke detection is concerned, sensitivity seems to be the most suitable parameter. The sensitivity of diagnostic methods in ischemic stroke diagnosis assessed in clinical trials seems to be a more robust indicator than personal preference for specific imaging methods. However, individual preferences are mostly used, especially for difficult, abnormal cases. The influence of

Tab. 1. Hierarchical model of ischemia identification criteria; PWI - perfusion weighted MR imaging, CTA - CT angiography.

Image-based merits related to onset time for stroke incident identification		
general diagnostic factors ↓	basic imaging features ↓	advanced imaging features ↓
✓onset time	✓CT - hypodense area	✓CT perfusion - MTT/CBV (penumbra)
✓full imaged information (penumbra plus infarct)	✓CT - hyperdense artery and hypodense area	✓PWI/DWI mismatch (penumbra)
✓neuro-scales (NIH)	✓CT - effacement of the gray/ white junction along the cortex	✓DWI (infarct)
✓laboratory markers	✓CT - obscuration of the lentiform nucleus	✓CT perfusion - CBV (infarct)
	✓CT - insular ribbon sign	✓SPECT (penumbra/infarct)
	✓CTA - artery occlusion	✓PET (penumbra/infarct)
	✓CTA - sinus or venous thrombosis	

selected criteria on the sensitivity of early stroke diagnosis varies widely in the literature [13] – exemplary results were presented in Table 2. Diffusion-weighted imaging is more sensitive to acute brain ischemia than plain CT, with reported sensitivity up to 95% [11].

Tab. 2. Sensitivity of early stroke diagnosis based on selected signs in CT.

Sign	Sensitivity
Hyperdense Middle Cerebral Artery (MCA)	54.5-78.5%
Hypoattenuation	36-83%
Hypoattenuation and Hyperdense MCA	60-87%

Therefore, normalized consensus in grading importance of diagnostic factors is an inaccessible challenge. The proposed pairwise comparison methodology is a way to make such consensus more realistic.

The first step of pairwise comparisons is to establish the relative preference of two criteria for situations in which it is impractical (or meaningless) to provide absolute estimations of the criteria. The relative comparison coefficients a_{ij} for criteria $A1, A2, \ldots, An$ are expected to satisfy the conditions $a_{ii} = 1$ and $a_{ij} = 1/a_{ji}$. The first constraint is related to comparing a given attribute with itself. The second constraint is a consequence of the obvious fact that $x/y = 1/(y/x)$ for $x, y \neq 0$.

A scale from 1 to 5, presented by Tab. 3, is used for expressing the importance of one criterion with respect to the other criterion in the pair. Other scales also exist but all of them are isomorphic.

Tab. 3. Comparison scale

Code	Definition of intensity or importance
1	Equal or unknown importance
2	Weak prevalence of one over the other
3	Moderate to essential importance
4	Demonstrated importance
5	Absolute importance
2.5 etc	Intermediate importance

Absolute estimations of the weights defining the importance of the analyzed stroke criteria are practically unobtainable with statistical or more formal procedures. It would be nice to have them, and probably more clinical trials would contribute to the accuracy of the current estimates, which are based on the results of experts' consensus or approximate opinions. However, we cannot wait until such clinical trials (taking sometimes 10 years or more) take place. This approach allows us to improve the processing of (often subjective) expert assessments of medical images.

We have proposed the use of the following comparison scale in table 3 for the subjective expression of relative preferences of stroke criteria, compared in pairs by an radiologist. Let us consider the selected part of the model presented in Tab. 1. The values of relative importance of basic imaging criteria, listed in the Table 4, have been entered by a panel of two radiologists and a biomedical engineer experienced in computer-aided support of stroke diagnosis for the illustration of the method. Intermediate scores were used to express the relative importance more precisely. For clinical trials, the values should be established by an extended panel of medical and biomedical engineering experts.

Tab. 4. Relative importance of the considered basic imaging criteria

	Criterion	A1	A2	A3	A4	A5	A6	A7
A1	CT - hypodense area	1	1	4	3	4	4	4
A2	CT - hyperdense artery and hypodense area		1	4	3	4	3.5	3.5
A3	CT - effacement of the gray/white junction along the cortex			1	0.5	1	2	2
A4	CT - obscuration of the lentiform nucleus				1	2	3	3
A5	CT - insular ribbon sign					1	1.5	1.5
A6	CTA - artery occlusion						1	1
A7	CTA - sinus or venous thrombosis							1

3.3. Knowledge mining in image data for stroke indicators using consistency-driven pairwise comparisons method

In the pairwise comparisons method, diagnostic criteria are considered in pairs by one or more experts. It is necessary to evaluate the individual alternatives, derive weights for the criteria, construct the overall rating of the alternatives, and finally identify the best alternative. Let us denote the criteria by $A1, A2, \ldots, An$ (n is the number of compared criteria), their actual weights by $\gamma_1, \gamma_2, \ldots, \gamma_n$, and the matrix of the ratios of all weights by $\Gamma = [\gamma_i/\gamma_j]$. The matrix of pairwise comparisons $\mathbf{A} = [a_{ij}]$ represents the relative intensities of assessments between the individual pairs of alternatives (Ai versus Aj, for all $i, j = 1, 2, ...n$), chosen usually from a given scale. The elements a_{ij} are considered to be estimates of the ratios γ_i/γ_j, where $\boldsymbol{\gamma}$ is the vector of the actual weights of the criteria (which is what we want to find). All the ratios are positive and satisfy the reciprocity property $a_{ij} = 1/a_{ji}$, $i, j = 1, 2, \ldots, n$. The practical challenge confronting the pairwise comparisons method comes from the lack of consistency in the pairwise comparisons (PC) matrices [12]. Basically, the distance-based inconsistency indicator is defined as the maximum over all triads $\{a_{ik}, a_{kj}, a_{ij}\}$ of elements of $\mathbf{A}$ (with all indices i, j, k distinct) of their inconsistency indicators, which in turn are defined as follows:

$$ii = \min\left(\left|1 - \frac{a_{ij}}{a_{ik}a_{kj}}\right|, \left|1 - \frac{a_{ik}a_{kj}}{a_{ij}}\right|\right). \tag{1}$$

The minimal number of alternatives which may cause inconsistency is three. A comparison of two criteria often results in inaccuracy, that is, inexact knowledge; however, it does not involve inconsistency.

The distance-based inconsistency is the minimum distance from three "ideal" triads with no inconsistency when the "third" value is substituted using the consistency condition $a_{ij} \times a_{jk} = a_{ik}$. Since we are not in a position to say which ratio is incorrect a priori, all three assessments must be reconsidered before we attempt to find a consistent approximation for a given pairwise comparisons matrix. The stress on locating the most inconsistent assessments is expressed by adding the term *consistency-driven* to the method's name since it is easier to remedy the consequences of an error (in the judgment) when we are able to locate it. When we find an error, we have a chance to fix it.

Let us consider, for example, the consistency analysis of the matrix $\mathbf{A}$ given in Table 4. The results are presented in the screen image in Fig. 2.

In essence, we need to find a consistent $n \times n$ matrix $\mathbf{B}$ which differs from matrix $\mathbf{A}$ "as little as possible". The normalized vector of the geometric means of rows (or columns) can be used to produce a fully consistent PC matrix which is the nearest approximation for a given (usually inconsistent) PC matrix $\mathbf{A}$. There is a strong relationship between accuracy and consistency. This is why the inconsistency analysis is the main focus of the consistency-driven approach.

When making comparative assessments of intangible criteria (such as, for example, "looking exhausted or disoriented"), we face not only imprecise or inexact knowledge, but also the inconsistency of our own subjective assessments. More importantly, the improvement of knowledge elicitation by controlling the inconsistency is not only desirable, but even needed for the refinement of our own expertise. It simply forces us to reconsider our domain knowledge by studying the case in more depth, or by using better data mining techniques.

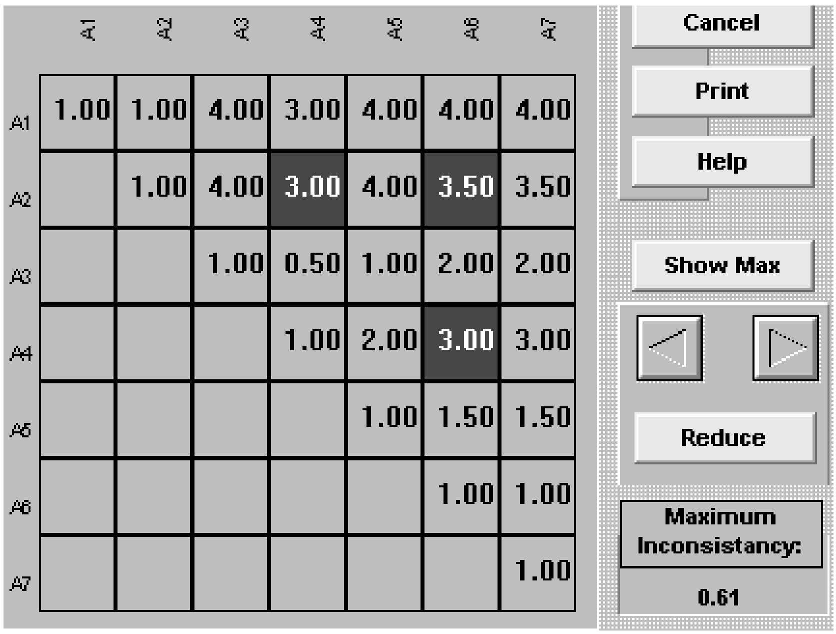

Fig. 2. Screen image with the pairwise comparisons matrix - initial form with significant maximum inconsistency; the analysis tool used was the Concluder v.2.72 [13].

In practice, inconsistent assessments are unavoidable when at least three criteria are independently compared against each other. For example, let us look closely at the three highlighted criteria in matrix **A**. Eq. 1 gives inconsistency 0.61, which is considered (as a heuristic proposed in [5]) as too high, and needs to be reduced. By changing the relative importance of $A4$ against $A6$ from 3 to 2.5, and $A4$ against $A7$ from 3 to 2.5, we reduce the local inconsistency (of the highlighted triad) to 0.56 (Fig. 3). When this is done, the software (by using the same equation) localizes another triad, $A2/A3, A2/A6, A3/A6$ with values $[4, 3.5, 2]$. Changing 4 to 3 makes this triad more consistent, but the next triad, $A2/A4, A2/A6, A4/A6$ with values $[3, 3.5, 2.5]$ has inconsistency indicator $ii = 0.53$. Such a procedure is continued up an acceptable consistency of the modified matrix. This inconsistency reduction process should be realized by medical experts who must base their new assessments on better understanding and medical knowledge.

The inconsistency concept is easier to explain by considering three criteria (or more generally - stimuli or information objects) C, D, and E and their respective areas. Let us assume the following initial assessments: C/D is 2, D/E is 3, and C/E is 5. Evidently, the above assessments violate $C/E = C/D \times D/E$, so we may try to correct the last assessment to 6 since 2×3 gives 6. Unfortunately, we do not know which assessment is inaccurate. In particular, as is frequently the case in practice, each original assessment might have been (and usually is) just a little inaccurate. Hence from the practical viewpoint, it is safe to assume that every assessment is somewhat inaccurate.

In concrete applications, a high value of the inconsistency indicator is a symptom of potential problems. A distance-based inconsistency reduction algorithm focuses, at each step, on an inconsistent triad and "corrects" it by replacing it with a consistent (or, more generally, less inconsistent) triad. This resembles "whack-a-mole," a popular arcade game. One difference is that instead of one mole, we have three array elements as explained above. After "hitting the mole" (which generally results in some other "moles" coming out), the next triad is located and the process is repeated. In case of a 7 by 7 array, this may be a tedious task but it needs to be done only once.

The final results of the presented example are shown in Fig. 3.

To sum up, the initial pairwise comparison assessments are entered into the pairwise comparisons matrix **A**. The next step is to refine these assessments by modifying the matrix elements according to inconsistency analysis. The refined relative assessments of all pairs are used for computing weights. The final weights of diagnostic criteria could be applied for assessment of new stroke cases based on the information collected in medical records of patients including the knowledge extracted from image examinations.

The solution to the above pairwise comparisons matrix is a normalized vector of geometric means:

$$V = [v_1, v_2, ..., v_n] \text{ where } v_i = \sqrt[n]{\prod_j a_{ij}}. \tag{2}$$

Geometric means have obvious interpretation as the arithmetic means in the corresponding space obtained by a logarithmic mapping. It does make sense to get the "average" of partial comparisons in order to get the global weights. In case of the matrix shown in Fig.3 (bottom), the final weights were calculated from Eq. 2, and the normalized vector is presented in Fig. 4.

These weights, corresponding to criteria $A1, A2, \ldots, A7$, should be applied (as multipliers) to the values of attributes for each patient suspected of having stroke conditions,

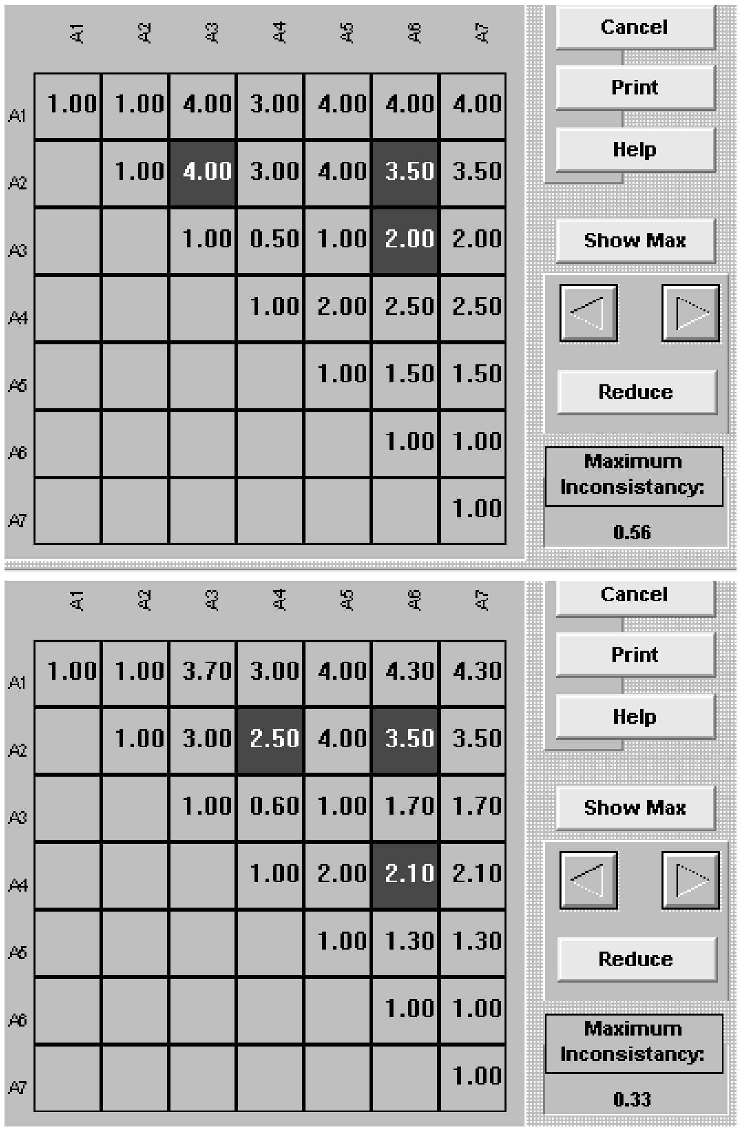

Fig. 3. Screen image with the pairwise comparisons matrix - one step modification with reduced inconsistency (up) and final modification with acceptable consistency.

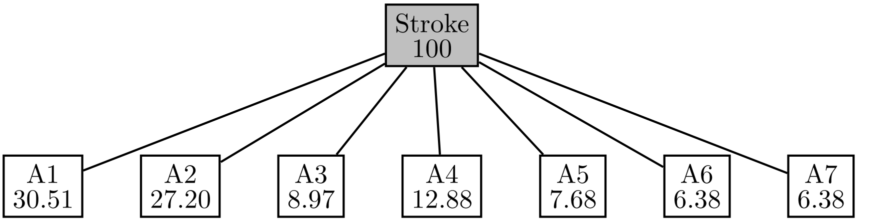

Fig. 4. Screen image with the final weights pairwise comparisons methods

to establish a merit index that can be used for making a decision on further examination. The threshold will be established by clinical trials which are expected to follow this study.

Let us see how our computed weights work by a demonstration on a hypothetical case. For this purpose, we will use records of three patients. Their individual scores (ranging from 0 to 5) on criteria $A1$, $A2$, ..., and $A7$ are given in Table 5. The scores could be calculated by a computer program (e.g. based on the percentage of damaged tissue) or assessed by a physician during examination (with 0 given for complete absence of any stroke feature, and 5 – for full stroke symptoms conviction). The total score for the patients are shown in the last column.

Tab. 5. Relative importance of considered basic imaging criteria for three patients, with exemplary mean scores of stroke conviction.

Patient	A1	A2	A3	A4	A6	A5	A7	Total
1	1	2	4	3	5	5	5	25
2	4	5	4	3	1	1	2	20
3	5	5	1	1	0	3	1	16

Patient number 1 has the highest score, and seems to be the best candidate for extra care. However, when we use our weights:

.305	.272	.129	.0897	.0768	.0638	.0638

as multipliers for the patients scores, we obtain different results:

Patient	1	2	3
Score	2.65	3.63	3.36

Now it is evident that patient 2 and 3 should be given more attention than patient 1.

Discussion and conclusions

This study attempts to help medical and health professionals to synthesize valuable clinical information obtained by image mining. The proposed methods, adjusted and exhaustively verified by clinical trials, can be potentially used in hospital emergency departments or in neurological clinics for screening and prevention of stroke.

Finding a magic formula for image mining is nothing but a scientific dream. It is more and more evident that a long process of gradual improvements is the way to go. As a consequence, ”a little bit of this and a little bit of that” is utilized for solving practical problems such as accurate early stroke diagnosis.

For the assessment of patient outcome, several different assessment scales exist, such as the Rankin score (RS), Barthel Index (BI), or mortality alone. On top of it, different thresholds for these assessment scales are arbitrarily used (e.g., RS of <= 1 versus <= 2). As far as we know, the presented method is the only one which allows for combining the medical knowledge from different medical images, imaging modalities, laboratory and clinical records into one index.

The transition from the mechanistic approach ”by a formula” or measurement to solutions based on ”judgment call” or ”expert opinion” is well supported by the pairwise comparisons approach, in which the complexity of deciding on ”everything at once” is reduced to comparing two features against each other to establish their relative importance. This approach reduces the complexity to the bare minimum of two features which is an irreducible number in practice (comparing the same feature with itself does not make much sense as a trivial case of identity). The pairwise comparisons method supports synthesis of partial solutions into one global solution. This paper shows that it can be done, and that it should be done.

Ideally, the elements of a pairwise comparisons matrix should be based on measurements and/or statistical studies of past observations. However, relying on opinions of a panel of experts is an acceptable compromise solution, until such reliable statistics are available. In case of medical diagnosis, expert opinions are the only way to reach final decisions.

The strongest conclusions are probably induced by the ”what if analysis.” In the absence of weights computed by the proposed method of pairwise comparisons, one would either add all scores received for the individual characteristics or try to guess some weights. The authors are quite confident that in both cases the proposed method would classify visual information more accurately, as it is intuitively evident. If not, the calibration of weights by what is known as *clinical trials* and/or manual tune up in medicine would improve the precision (bringing machine learning and artificial neural networks into the enhanced model) towards achieving our goal to improve image mining by a better way of classifying visual objects.

For technical reasons (the text would balloon to an acceptable size), this is mainly a presentation of the concept of knowledge mining from medical image. Subjective assessments are numerically verified, compared and modified according to inconsistency control procedure. Thus, comprehensive consensus for the problem solution is obtained, and defined more objectively. The weights obtained are useful for automatic inference for computer-aided diagnosis purposes. The diagnostic image information extracted in effective image processing procedures can be combined with other semantic descriptors of the analyzed objects obtained in neurological, clinical and laboratory procedures to synthesize diagnostic opinions and probabilistic suggestions on a higher abstract level of human understanding.

The innovative approach to looking at numerous medical images from the perspective of a "big picture" may be perceived as speculative, but the alternative is even worse and expressed by an old adage related to an alternative: "do something or do nothing". Since doing nothing is safe, it is often the preferred choice. In this study, the authors have decided to do something in this very difficult research area.

The presented numbers have been entered after only short-time consultations with a limited range of medical professionals for the purpose of illustrating the method. The obtained results need to be followed by clinical trials, sometimes taking years, whose results will lead to clinically useful tools and procedures. A jury model is strongly recommended as further improvement. A panel of experts should be called to set the priorities of criteria in pairs. It is in itself a complex task, as the rule of majority voting must not be used. A better approach is based on the Delphi method, or on a weak order as in [7]. The composition of such expert panel should be diversified. The majority the expert panel for stroke application should be radiologists and neurologists. However, other medical researchers specializing in strokes, including engineers, and officials from the Ministry of Health can also be included in the panel.

There is room for improvement in stroke analysis studies. There is also a need for more uniform definitions of multiple variables, such as collateral flow, degree of recanalization, assessment of perfusion, and infarct size.

Acknowledgments

We acknowledge invaluable input provided by Dr Katarzyna Sklinda for determining the relative importance of medical criteria.

References

1785

[1] de Caritat, M. J. A. N., Marquis de Condorcet: Essai sur l'Application de L'Analyse à la Probabilité des Décisions Rendues à la Pluraliste des Voix. Facsimile reprint of original published in Paris, 1972, by the Imprimerie Royale.

1860
[2] Fechner, G.T.: Elements of Psychophysics, Vol. 1. Translation by H.E. Adler of Elemente der Psychophysik (Leipzig Breitkopf und Hartel, 1860). Holt, Rinehart and Winston, New York, 1965.
1927
[3] Thurstone, L.L., Law of Comparative Judgements, Psychological Review, 34, 273–286.
1992
[4] Kato, T. (1992). Database architechture for content-based image retrieval. In Proceedings of the SPIE – The International Society for Optical Engineering, volume 1662 (pp. 112-113). San Jose, CA, USA.
1993
[5] Koczkodaj, W.W. (1993), A New Definition of Consistency of Pairwise Comparisons. Mathematical and Computer Modelling, Vol. 18, 7, 79–84.
[6] Koczkodaj, W.W., (1996), Statistically Accurate Evidence of Improved Error Rate by Pairwise Comparisons, Perceptual and Motor Skills, 82, 43–48.
1998
[7] Janicki, R., Koczkodaj, W.W., A Weak Order Solution to a Group Ranking and Consistency-driven Pairwise Comparisons, Applied Mathematics with Computation, 94(2/3), 227–236, 1998.
1998
[8] Koczkodaj, W.W., Testing the Accuracy Enhancement of Pairwise Comparisons by a Monte Carlo Experiment, Journal of Statistical Planning and Inference, 69(1), 21–32.
2005
[9] Kulikowski J.L., The role of ontological models in pattern recognition, Computer Recognition Systems. Proc. of the 4th International Conference on Computer Recognition Systems CORES'05 (M. Kurzynski et al eds), Springer, 43-52
[13] Wardlaw J.M., Mielke O., Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment - systematic review, Radiology 253(2), 444–453
2006
[10] Srinivasan A., Goyal M., Al Azri F., Lun C., State-of-the-art imaging of acute stroke, RadioGraphics 26, S75–S95
[11] Muir K.W., Buchan A., von Kummer R., Rother J., Baron J.C., Imaging of acute stroke, Lancet Neurol 5, 755–768
2008
[12] Bozsóki, S., Rapcsák, T., On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices, Journal of Global Optimization, 42(2), 157–175
2009
[13] Koczkodaj, W.W., Robidoux, N., Tadeusiewicz, R., Classifing Visual Objects by the Consistency-driven Pairwise Comprisons Method, to appear in MG&V.
[14] Przelaskowski A., Ostrek G., Sklinda K., Walecki J., Jozwiak R., Stroke slicer for CT-based automatic detection of acute ischemia, Advances in Intelligent and Soft Computing, Computer Recognition Systems, Springer 2009, in press