Texture analysis of the developing human brain using customization of a knowledge-based system [version 1; referees: 2 not approved]

Hugues Gentillon¹,², Ludomir Stańczyk¹, Michał Strzelecki², Maria Respondek-Liberska³

¹Department of Radiology and Diagnostic Imaging, Barlicki University Hospital, Medical University of Lodz, Lodz, Poland
²Institute of Electronics, The Faculty of Electrical, Electronic, Computer and Control Engineering, Technical University of Lodz, Lodz, Poland
³Diagnosis and Prevention of Congenital Malformations, Instytut Centrum Zdrowia Matki Polki, Lodz, Poland

Abstract

Background: Pattern recognition software originally designed for geospatial and other technical applications could be trained by physicians and used as texture-analysis tools for evidence-based practice, in order to improve diagnostic imaging examination during pregnancy.

Methods: Various machine-learning techniques and customized datasets were assessed for training of an integrable knowledge-based system (KBS), to determine a hypothetical methodology for texture classification of closely-related anatomical structures in fetal brain magnetic resonance (MR) images. Samples were manually categorized according to the magnetic field of the MRI scanner (i.e. 1.5-tesla (1.5T), 3-tesla (3T)), rotational planes (i.e. coronal, sagittal and axial), and signal weighting (i.e. spin-lattice, spin-spin, relaxation, proton density). In the machine-learning sessions, the operator manually selected relevant regions of interest (ROI) in 1.5/3T MR images. Semi-automatic procedures in MaZda/B11 were performed to determine optimal parameter sets for ROI classification. Four classes were defined: ventricles, thalamus, grey matter, and white matter. Various textures analysis methods were tested. The KBS performed automatic data pre-processing and semi-automatic classification of ROIs.

Results: After testing 3456 ROIs, statistical binary classification revealed that combination of reduction techniques with linear discriminant algorithms (LDA) or nonlinear discriminant algorithms (NDA) yielded the best scoring in terms of sensitivity (both 100%, 95% CI: 99.79-100), specificity (both 100%, 95% CI: 99.79-100) and Fisher coefficient (≈E+4, ≈E+5, respectively).

Conclusions: LDA and NDA in MaZda can be useful data mining tools for screening a population of interest subjected to a clinical test.
Introduction

Medicine is not an exact science but an applied, interdisciplinary field. Therefore, the time to produce physicians specialized in radiology is very long. Moreover medicine is still and will still be evolving in years to come. As cures are being discovered or invented, new diseases become known and mutations surface along with new variants. Trainable knowledge-based systems (KBS) could be an answer to the global shortage of radiologists. Besides, there are other major obstacles, which also prevent KBS from being fully functional out of the factory. The body of medical knowledge, known to this date, is gathered and transferred through theoretical and clinical intuition as well as experience. Then physicians continue to expand their acquired knowledge with years of evidence-based practice. On top of the challenge is the fact that some conditions may not be symptomatic during medical examination. Hence, medicine is indeed a continuing learning process. That was why we proposed the approach to have the observer (in this case the physicians working in the field) as the direct trainer (the programmer) of a KBS designed as a customizable, conceptual framework. In computer science, a framework is a system which implements the process of abstraction — i.e. a technique where developers make computer modeling/programming simpler to understand, use and apply. Such a KBS should not be designed as “a hammer to drive a nail” but as an abstraction system with generic functionality — which can be changed with user-written codes and customized for unlimited applications. The purpose is to enhance human-computer interaction (HCI) in medicine.

The aforementioned points are introduced especially to show the need for customization in computer-aided-diagnosis (CAD). Pre-programmed CAD systems surely help radiologists and obstetricians, but they could be better and more useful with some room for customization — and could even improve the sharing and preservation of diagnostic innovations.

Research background & goals

The aim of this research was to find a customizable software framework (KBS) to mathematically determine an optimal combination of texture analysis methods to differentiate anatomical structures in the developing fetal brain (i.e. regions of interest (ROI)). Why did this experiment focus on normal development of central nervous system (CNS) rather than common conditions affecting fetal organogenesis? When considering fetal intervention to correct an anomaly in utero, the first ethical priority is actually mother safety. Prenatal well-being is clinically second, after evaluation of benefit-risk ratio. In the 1960s, fetal surgery was conceived and introduced in clinical practice. In spite of the improvement in surgical technology, the number of successfully-treated cases of congenital defects and life expectancy of survivors are still limited. The theoretical procedures are troublesome and have not been investigated enough. Hence, they are considered as experimental treatments. Whether invasive or minimally invasive, fetal intervention is not the ultimate solution of this societal conundrum. It is usually reserved for cases of severe fetal anomalies. In recent years, prenatal therapy is gaining popularity in religiously conservative territories particularly where abortion is prohibited. Fetal intervention is recommended for fetuses with mild and non-lethal defects, in order to discourage abortion and continue childbearing. In the country where this research was carried out (i.e. Poland), abortion is, once again, nationally banned — except in cases of rape, life-threatening pregnancy and childbirth, and grave malformations. Pressures are due to growing pro-life supporters demanding total anti-abortion, child-bearing at all costs, and capital punishment for illegal termination of pregnancy. Nevertheless, Poland neither sanctions nor executes the outlaws of illegal abortion. Despite the absence of penalty, abortion is respectfully performed as permitted by local authority. The aforementioned dilemma justified the usage of fetal MRI (fMRI) in this research. In theory, it was previously hypothesized and documented that the high electromagnetic fields (EMF) used for MR procedures can disrupt the early stage of organogenesis. To this date, the embryotoxic, fetotoxic, and teratogenic effects of MRI are not well known. Normally, fMRI is not recommended during the first and second trimester, unless it is absolutely necessary to confirm and/or supplement the diagnosis of fetal anomalies. Legal decision to prescribe abortion to a patient sometimes requires advanced clinical investigation, accompanied by psychological counseling before and after the operations. To this date, ultrasound (US) devices are preferred for obstetrical examinations. With 2D, 3D, and 4D US scans, physicians can effectively and efficiently diagnose the majority of life-threatening conditions affecting mother and fetus. Therefore, ultrasonography (USG) is sufficient for diagnosis of severe malformations affecting abdominal organs. Why was magnetic resonance imaging (MRI) prescribed? Fetal brain is where US devices struggle to produce desirable results. MRI was subsequently performed to rule out severe abnormalities in brain development, which are not visible on sonogram. Magnetic resonance (MR) samples came from fetuses with suspected heart and kidney defects. Disruption of organogenesis in the latter, depending on severity, might affect normal development of the brain. MR samples used in this experiment were visually investigated by specialists, structure-by-structure. No severe malformations were observed. In these cases, MRI studies did not add any further indication to legally fulfill the criteria to terminate pregnancy. The human visual apparatus has its limit. Unfortunately, missed diagnoses do occur. Malformations may not be apparent prior to birth. If fetal defects are suspected, bureaucracy may also restrict access to more advanced testing and healthcare. Consequently, pregnant women are deliberately forced to bear and deliver malformed babies. At the end of the day, physicians may still have to deal with the legal liability for failure to terminate pregnancy. Unwanted fetuses may become neglected, and foundling is also a growing problem in society. Congenital brain defects and its impacts on physical and cognitive development may not be detectable until after birth. Fetal outcome and mental retardation can be difficult for a physician to predict. The process requires access to better medical examination, development of more advanced tools and further investigation. That is why the ultimate goal of this feasibility study was to gather knowledge for the practicality of a proposed project seeking to improve diagnostic accuracy and precision, by extending HCI usage in medicine. There are many computer-aided diagnostic tools on the commercial shelves (e.g. Radiomics, Definiens Tissue Phenomics, CAD4TB Diagnostic Software). Sadly, they are primarily designed for pre-loaded applications but not much else.
Methods
This study was approved by the Research Ethics Committee of the Medical University of Lodz (MUL). Written informed consents were obtained from all subjects, as well as perpetual licenses pertaining to copyright and ownership bundle-of-rights of the medical records — for the purpose of education, research and publication. All experiments were performed in accordance with the relevant institutional and national guidelines, regulations, licenses and approvals — from, but not limited to MUL, Lodz University of Technology (TUL), Barlicki University Hospital (BUH), Central University Hospital (CUH), Kopernik Hospital, Biegarski Specialty Hospital, Polish Mother’s Memorial Hospital & Research Institute of Lodz (ICZMP) and National Health Fund (NFZ). The collective approval-certificate (number RNN/213/13/KE) and the research proposal were reviewed and endorsed by the senior officers of the Research Ethics Committee.

Ethics and informed consent
This feasibility study was approved by the Bioethics Research Board which regulates experiments carried out at the Medical University of Lodz and affiliated research hospitals in Poland (permit number: RNN/213/13/KE). Due to administrative and logistic delay as well as finding volunteers and expensive cost and availability of MRI—, it took us nearly five years to collect the magnetic resonance (MR) samples. The nature of the study was explained to patients in Polish by attending physicians, and individual written informed consent was obtained for this research and its publication. Agreement number: 3/2011 - concluded on December 6, 2011, between the experimenter (Hugues Gentillon), research supervisor (Ludomir Stefaničzyk) and rector of Medical University of Lodz (Radzisław Kordek), Faculty of Biomedical Science - Postgraduate research in Diagnostic Imaging and Radiotherapy. Consent for publication of these data was obtained. Furthermore all data from human participants were anonymously as per consent agreement. Informed Bioethics Committee Approval Number: RNN/213/13/KE, JULY 16, 2013. If you have any questions regarding the decision please include the above number and date in your letter. Send correspondence to: THE BIOETHICS COMMITTEE OF THE MEDICAL UNIVERSITY OF LODZ Al. Kościuszki 4, 90-419 Łódź, tel. 0 785 911 596, 42 272-59-05, fax 42 272-59-07.

Clinical trial registration
Though this research shares similarities with a clinical trial, it is “virtual” — i.e. it is non-interventional. Such medical study does not meet criteria for clinical trial registration[18–42]. Furthermore the investigational tools were merely used for technical exploration and to measure their feasibility in medical practice — by using simulation settings. Lastly, the results were not used to alter patients’ therapeutic care and outcome[18–42].

Advance notice to readers. Readers should not expect us to teach the entire science of artificial neural network (ANN) (feedforward neural networks, recurrent neural network, etc.) in just a manuscript. It is not possible. A full introduction is not even possible. Unfamiliar readers are expected to make an effort on their own to read and learn the basic principles of artificial neural network and know the basic terminology. Like a human brain, an ANN can store memories. ANN can also judge based on stored memories and logical rules. To run a naïve trial run, it was ideal to have the ANN in a condition like ‘permanent global amnesia’ – i.e. a phenomenon where a brain is in a state of total blackout and thus cannot judge based on prior memories. The KBS used in this experiment was lacking an automatic memory cleaner and optimizer. Hence, stored memories were manually deleted in the ANN for every trial run.

Computer vision. Under a reciprocal cooperation-agreement between MUL and TUL, we obtained permission to test a KBS consisting of a custom version of MaZda software package[5,6,4]. It contains algorithms for data classification and visualization. How did it come to us? In 1998, a group of medical scientists, engineers, physicists, mathematicians and others initiated the B11 project at the Cooperation in Science and Technology (COST). The purpose was to develop software frameworks for quantitative analysis of MR scans, to improve medical diagnosis. MaZda, a Delphi/C++ computer program, was originally built in 1996 for applications in mammography. It was later used by COST to complement its MRI software modules (e.g. B11, B21). MaZda 4.6 package was the last official release (B11 version 3.3 included). In this research, we collaborated with modular programmers to further upgrade, separate and amalgamate the functionality and compatibility of MaZda (version 5 RC HG) with other modules. In the MaZda 5 version used in this research, the new features names were introduced to maintain compatibility with WEKA software (www.cs.waikato.ac.nz/ml/weka/, a popular package for data mining, classification and analysis). This way data generated by Mazda may be used as the input to the WEKA. Also, 3D deformable model that are used for 3D volume-of-interest (VOI) generating was introduced — for details, see the publication by: Piotr M. Szczypiński, Ram D. Sriram, Parupudi V.J. Sriram, D. Nageshwar Reddy, A model of deformable rings for interpretation of wireless capsule endoscopic videos, Medical Image Analysis, Volume 13, Issue 2, April 2009, Pages 312–324). Moreover, to speed up and automate, some commands regarding 3D analysis were introduced. This way one can build a script (that contains e.g. commands for image loading, analysis option loading, performing analyses and storing results) for automatic analysis of a number of 3D data. Finally, this MaZda 5 version was compiled with Embarcadero software environment, which was not an ideal framework. Thus, the inventors decided to return to Builder C++. Currently, the latest version of Mazda being built is qMaZda (described and available at www.eletel.p.lodz.pl/pms/SoftwareQmaZda.html).

Figure 1 shows the main steps of texture analysis with MaZda and B11. A key change in this custom version is the integration of MaZda with a variety of computational geometric algorithms from Qhull (e.g. VSCH_1, VSCH_2, VSCH_3, VSCH_4, VSCH_5). For details about 1D, 2D, 3D, and 4D features, see Dataset 7. VSCH network algorithms are located in the ‘Convex Border’ menu. VSCH_1, for example, can be used to identify the strongest discriminant parameter in a large dataset.

Dataset 7. Parameter list
http://dx.doi.org/10.5256/f1000research.10401.d146788
Figure 1. Main steps of texture analysis with MaZda and B11.

Our trial-and-error experiments and texture-analysis software development spanned over five years of research. All the findings shared common denominators. Quantitative brain-tissue segmentation was affected by several factors, such as characteristics of fetuses (e.g. gestational age, shape, normal/abnormal development) and the quality of fMR images (e.g. 1.5/3T, resolution, slice thickness, rotational planes, artifacts, etc). Automatic segmentation of newborn brain MRI has been documented in the literature. The algorithmic contributions reported so far have achieved limited success, unfortunately. Automatic segmentation of prenatal brain is even more challenging and time-consuming.

Unsupervised segmentation. Once an image is acquired in a readable format (bitmap format (BMP)), the first step is texture segmentation – i.e. partitioning an image into ROIs. B11 can perform unsupervised segmentation and cluster analysis. In some instances, B11 achieved accuracy closed to that of clinicians. However, unsupervised segmentation was not reliable enough for therapeutic use. Fetal brain segmentation with B11 still required extensive expert interaction. In our observations, the key problems were maternal factors, environmental effects, growth variability, randomness of fetal movements and its detrimental effects on image quality. Therefore, automatic segmentation was used for new insight into the possibility of improving supervised segmentation. The steps of unsupervised segmentation are relatively simple: image acquisition and run analysis. ROIs and segment numbers can be manually adjusted. The drawback with B11 segmentation is limitation to 8-bit grayscale BMP. 16-bit DICOM was converted to visually lossless BMP, by dropping least significant bits. Note that B11 identified textures not anatomical structures. The information collected from the unsupervised trials was later used as guidance to manually estimate boundaries of anatomical structures (ROIs) for the supervised trials. The preliminary trials were single-blinded — i.e. the user knew the characteristics of the ROIs, and the KBS received no hints (no ROI selection).

Further information was gathered with a semi-automatic (unblended) segmentation by defining 4 classes (ROIs): thalamus, ventricles, grey matter and white matter. In unsupervised mode, the KBS performs quite well when brain images are from MR examination of the same subjects and same sequences — but performs poorly when they came from different subjects. The findings were likewise for same sequences of the same patient taken at a different time and MR scanner settings. The challenge was: how do we match macroscopic characteristics with electronic recognition, regardless of MR image shadings? MaZda and B11 are not yet designed to allow user to well define semantic rules and/or import plugins for fully electronic recognition of anatomical structures. Object-based image analysis tools such as eCognition work consistently well for geo-spatial applications (e.g. identification of a river in an image). In fetal radiology, it is still a challenge to achieve consistent results with automated-pattern recognition of prenatal anatomy. Programming a reliably effective system for such highly sensitive application is feasible but also time-consuming. Such a task would require taking into account all known variations due to pregnancy chronology and fetal developmental, to minimize segmentation errors. MaZda and B11 were originally built for HCI rather than fully-automated applications. Therefore, the best practical methodology, in this research, was for the operator to at least have prior knowledge of human embryogenesis, in order to manually and correctly identify and select fetal ROIs. Often, macroscopic appearance of many brain structures are not well differentiated in the first trimester. Hence, the selected samples were at least 20 weeks of maternal age.

Supervised segmentation. 3-tesla (3T) and 1.5-tesla (1.5T) magnetic resonance (MR) sequences of fetal brain were manually segmented into 3456 ROI. The categories were predefined as followed: ventricles (class 1), thalamus (class 2), white matter (class 3) and grey matter (class 4). The selected samples did not have any brain...
malformations. As previously mentioned, the anomalies were in the cardiovascular and/or renal systems. The focus of this research was on normal anatomy of fetal brain. Forward processing method — also known as “supervised segmentation”68 — was performed as delineated: (1) image acquisition from MR scanner, (2) selection of ROIs with MaZda, (3) image normalization with MaZda (4) feature extraction with MaZda (5) data preprocessing with B11 (6) texture classification with B1163,64,68. The first steps were done with MaZda and last two steps with B11 (Figure 1). After the preliminary trials, we became interested to learn what needs to be adjusted in order to reduce misclassification of MR images. The unsupervised segmentation revealed that the KBS was very sensitive to greenscale shading, artifacts, and image thickness as well as resolution quality. Thus, we trained the KBS accordingly. 1.5T images were originally encoded as 12-bit lossless JPEG (Joint Photographic Experts Group) format and wrapped in Digital Imaging and Communications in Medicine (DICOM). 1.5T images were transcoded from lossless JPEG to uncompressed DICOM. 3T images were natively uncompressed. 360 parameters were extracted with MaZda (histogram: 9; co-occurrence matrix: 220; run-length matrix: 20; gradient matrix: 5; auto-regression: 5; Haar wavelet: 28; geometry: 73). Parameters’ names are provided in the appendix at the end of this manuscript. We assessed three automated techniques of parameter selection — i.e. Fisher selection, POE + ACC (classification error probability and average correlation coefficients), MI+PA+F (combination of mutual information, pair analysis and fisher selection). Further details about the mechanics and functionality of these techniques are provided in the manufacturer’s user manual and tutorial guides69. For each technique, we also used the VSCH (vector supported convex hull) module in MaZda to enhance computation and visualization of geometric structures: 1) pre-reduction (before selection) to rule out insignificant parameters, 2) post-reduction (after selection) to identify strongest relevant parameters. All aforementioned parts of the process were computed automatically by the KBS without any manual interference. Data imported [SEL (Schweitzer Engineering Laboratories)] to CSV (comma-separated values) in B11 were preprocessed with PCA, LDA and NDA and classified by means of 1-nearest neighbor (1-NN) and ANN. In machine learning and cognitive science, 1-NN is also known as k-NN, where n=1. It is a “lazy-learning” algorithm, in which new ROIs are locally classified by getting interwove into the closest cluster in the training set. The rest of the computation is delayed until the end of the classification process. K-NN is implemented in B11 classifiers, as well as in the preprocessing procedures in MaZda69. On the other hand, ANN is a self-organizing algorithm with hidden layers and adjustable number of neurons69-72. It can be used for both supervised and unsupervised classification. Neural classification algorithms are implemented in MaZda/B11. ANN training, for example, is standardized for NDA analysis69,70 — i.e. a type of feedforward-artificial neural network, based on multilayer Perceptron (MLP)69. Nonlinear procedures and classifiers in B11 are MPL algorithms69,70. For optimal procedures, two sets of samples are needed: one for training and another one for validation71. The pitfall with this algorithm is its sensitivity to overtraining (too strong memorization)69,70. ANN training time is shorter with standardization. For continuation, training without standardization was carried out, in spite of long processing time. ANN (one-class/ n-class) and 1-NN training runs were conducted with different sequences of MR images: T2-weighted (T1), T1 weighted, and proton-density (PD) sequences. N-class training was discontinued due to repeated problems with overtraining and lack of reproducibility in F values and miss-classification errors.

Customization
Despite the usage of multi-level, automated selection/reduction techniques, some extracted values still did not match the controlled ROI values. Differentiating thalamus from other thalamic nuclei and grey matter was the key problem. That was when we manually intervened to customize and improve the extracted data. First ROI surface areas were manually increased, in order to limit the number of parameters reporting zero and infinity values. Parameters which couldn’t be correctly computed were manually omitted in the report file. Some pre-processing procedures in both MaZda and B11 couldn’t be performed when the report file contained erroneous values. We accessed MaZda generated report files by changing the extension format from SEL to CSV and then imported them into Excel 2013 for adjustment. Parameters measured with other CAD tools can also be entered in the report files by simply using Microsoft Excel. The edited file can then be imported in B11 to perform texture classification.

Regions of interest
Additional tests were carried out with same ROIs (i.e. thalamus, ventricle, grey matter and white matter) to dramatically improve accuracy and precision of the KBS: it was done with a customized dataset derived from MaZda algorithms, using semi-manual reduction and nearest-neighbor feature selection (see Data availability: Dataset 1–Dataset 2). The training data were used to orient the KBS to recognize what ROIs had the same tissue characteristics, in spite of being originated from different patients or different sequences of the same patients. The training was conducted with combination of two built-in classification tools (i.e. nearest neighbor (NN) and artificial neural network) and four data processing techniques (i.e. RAW: read as written; PCA: principal component analysis, LDA: linear discriminant analysis; NDA: nonlinear discriminant analysis). To measure the KBS sensitivity and specificity, we defined “normal” as “ROIs with identical tissue” characteristics and “abnormal” those with different tissue characteristics and “abnormal” those with different tissue characteristics.

Figure 2. White matter ROI selection. a) specificity scoring: ROI 1 = white matter control, ROI 2 = white matter | — b) sensitivity scoring: ROI 1: white matter control, ROI 2: thalamic nucleus other than thalamus.
characteristics (Figure 2–Figure 5). Apart from noise and artifacts, we found out that the preliminary results were also affected by the planes (axial, coronal, sagittal) — which refer to the rotational planes of the spinning MR scanner in relation to the mother, not the fetus. There flows the reason for the classification by rotational planes. In learning mode, we observed a consistent scoring for all the ROIs. Thus the logical and semantic information provided to the KBS was effective. Statistical binary tests (also known as classification function tests) were computed in STATISTICA version 10 to assess the performance of each procedure (combination of preprocessing techniques and classifiers). In medicine, binary scores (TP, FP, TN, FN etc.) are used to determine not just normal and abnormal characteristics but also classification property of an examination.

Results
With Fisher coefficient (F), we tested for difference between ROIs. It was nearly zero for ROIs which were alike. Therefore, the tissue anatomy was consistently the same among the normal ROI group. In testing mode, misclassification values, as low as 0%, were also recorded, in some trials (Table 1). RAW and PCA did not responded to the training, while LDA and NDA did. We obtained high F values, 100% sensitivity and 100% specificity for LDA and NDA (Table 1–Table 2) — which means that there was likely a real difference between the normal and the abnormal ROIs.

Dataset 1. 1.5T data
http://dx.doi.org/10.5256/f1000research.10401.d146782
10 parameters were retained out of 348. All parameters are listed in Dataset 7.

Dataset 2. 3T data
http://dx.doi.org/10.5256/f1000research.10401.d146783
10 parameters were retained out of 348. All parameters are listed in Dataset 7.

Dataset 3. 1.5T
http://dx.doi.org/10.5256/f1000research.10401.d146784
NORMAL was defined as ROIs with identical tissue - e.g. white matter in the occipital region vs white matter in the frontal region of the brain.

Dataset 4. 3T
http://dx.doi.org/10.5256/f1000research.10401.d146785
NORMAL was defined as ROIs with identical tissue - e.g. white matter in the occipital region vs white matter in the frontal region of the brain.
Table 1. Binary classification tests — 1.5T.

<table>
<thead>
<tr>
<th></th>
<th>RAW</th>
<th>PCA</th>
<th>LDA</th>
<th>NDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>938</td>
<td>956</td>
<td>1728</td>
<td>1728</td>
</tr>
<tr>
<td>TN</td>
<td>1595</td>
<td>1622</td>
<td>1728</td>
<td>1728</td>
</tr>
<tr>
<td>FP</td>
<td>133</td>
<td>106</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FN</td>
<td>790</td>
<td>772</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>3456</td>
<td>3456</td>
<td>3456</td>
<td>3456</td>
</tr>
<tr>
<td>Sn (%)</td>
<td>54.28</td>
<td>55.32</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Sn 95%CI</td>
<td>51.9 - 56.65</td>
<td>52.94-57.69</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>Sp (%)</td>
<td>92.3</td>
<td>93.87</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Sp 95%CI</td>
<td>90.94-93.52</td>
<td>92.63-94.95</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>87.58</td>
<td>90.02</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>PPV 95%CI</td>
<td>85.46-89.50</td>
<td>88.06-91.76</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>66.88</td>
<td>67.75</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>NPV 95%CI</td>
<td>64.95-68.76</td>
<td>65.84-69.62</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>PLR</td>
<td>7.05</td>
<td>9.02</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td>PLR 95%CI</td>
<td>5.96-8.35</td>
<td>7.46-10.90</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td>NLR</td>
<td>0.5</td>
<td>0.48</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NLR 95%CI</td>
<td>0.47-0.52</td>
<td>0.45-0.50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>P 95%CI</td>
<td>48.32-51.68</td>
<td>48.32-51.68</td>
<td>48.32-51.68</td>
<td>48.32-51.68</td>
</tr>
<tr>
<td>F</td>
<td>≈ 150</td>
<td>≈ 150</td>
<td>≈ 5000</td>
<td>≈ E+4</td>
</tr>
</tbody>
</table>


Discussion

To this date, no such research has been documented in the literature. The explanation could be derived from the difficulty of finding fetal MRI samples for medical research, as well as the common hindrance to their availability — i.e. continuing systematic concerns over the theoretical risks of MRI usage during pregnancy, in parallel to the lack of clinical studies and trials assessing such theoretical risks, plus the expensive cost of MRI examination and the scarcity of customizable CAD tools on the freeware shelves — just to list a few.

Selecting KBS tools

The majority of the KBS we came across were designed for technical use and not easily customizable. Such programs required paying for marketing company maintenance and for in-house-developed customization service, on top of the annual license fee. Thus this option was not feasible for application in real-world settings, where resources are often sparse (e.g. eCognition, Media Cybernetics, Radiomics, Definiens Tissue Phenomics, CAD4TB Diagnostic Software, etc.).

Logic and reasoning behind the research design

Previous medical studies done with MaZda include inflammation, brain cancer detection, multiple sclerosis, electrophoresis, etc. Herein, we defined some test samples as “abnormal” ROIs. However, they were, in reality, normal tissue. Not testing directly for a common anomaly doesn’t necessarily mean that there is no real...
Table 2. Binary classification tests — 3T.

<table>
<thead>
<tr>
<th></th>
<th>RAW</th>
<th>PCA</th>
<th>LDA</th>
<th>NDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>1424</td>
<td>1440</td>
<td>1728</td>
<td>1728</td>
</tr>
<tr>
<td>TN</td>
<td>1651</td>
<td>1636</td>
<td>1728</td>
<td>1728</td>
</tr>
<tr>
<td>FP</td>
<td>77</td>
<td>288</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FN</td>
<td>304</td>
<td>92</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>3456</td>
<td>3456</td>
<td>3456</td>
<td>3456</td>
</tr>
<tr>
<td>Sn (%)</td>
<td>82.41</td>
<td>83.33</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Sn 95%CI</td>
<td>80.53-84.18</td>
<td>81.49-85.06</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>Sp (%)</td>
<td>95.54</td>
<td>94.68</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Sp 95%CI</td>
<td>94.46-96.47</td>
<td>93.51-95.69</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>94.87</td>
<td>93.99</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>PPV 95%CI</td>
<td>93.63-95.93</td>
<td>92.69-95.13</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>84.45</td>
<td>85.03</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>NPV 95%CI</td>
<td>82.77-86.03</td>
<td>83.36-86.6</td>
<td>99.79-100</td>
<td>99.79-100</td>
</tr>
<tr>
<td>PLR</td>
<td>18.49</td>
<td>15.65</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td>PLR 95%CI</td>
<td>14.85-23.03</td>
<td>12.82-19.12</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td>NLR</td>
<td>0.18</td>
<td>0.18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NLR 95%CI</td>
<td>0.17-0.20</td>
<td>0.16-0.20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>P 95%CI</td>
<td>48.32-51.68</td>
<td>48.32-51.68</td>
<td>48.32-51.68</td>
<td>48.32-51.68</td>
</tr>
<tr>
<td>F</td>
<td>≈500</td>
<td>≈500</td>
<td>≈1500</td>
<td>≈1E+5</td>
</tr>
</tbody>
</table>


medical application. Though the tests were simulated, the research design was conceived for real-world medical applications. For example, this research design could be used to detect ectopic tissue migration, neurogenesis and neuronal migration (brain function migration as a result of natural process or after injury), metaplasia and interference with brain development. Last but not least, this simulation research followed standards used in clinical trials.

Statistical test and interpretation
The choice of binary classification (sensitivity/specificity) was favored over frequentist inference (p-value) because it provides more information in terms of statistical relevance to medical diagnosis, prognosis and disease prevalence. One key difference between Fisher, POE+ACC, and MI+PA+F is the number of parameters. To perform MI+PA+F, the dataset must contain at least 30 parameters strongly matching its selection-reduction criteria. Otherwise, the KBS reported an error. In our study, it was a common occurrence when the surface area of a ROI was insufficient to extract 30 parameters meeting the MI+PA+F semantics. RAW and PCA were not so affected by the training process and thus remained very sensitive to minute difference in greyscale shading.

Recommendations
A solution to high misclassification (M) was to exclude some parameters which were very sensitive to post-editing sharpness. In this research, the images were, however, processed without post-editing sharpness because high M was not regarded as a problem. Instead, we used RAW and PCA as reference tests (results before the training of the KBS). On the other hand, LDA and NDA responded well to the training, and M was consistently zero.

KBS memory clearing
During the study, we had to obviously clear the KBS memory several times for every trial run. We hope that the software developers will soon implement a more effective and efficient way (e.g. one-click) to clear specific random-access memory (RAM) without closing module(s) or without manually dumping the entire RAM or restarting the computer.

Constraints, limitations, and assumptions
Patients gave consent to perform MRI examination and use of images in research and for the manuscript publication. Nevertheless, this authorization was not enough, as ownership and copyright of
medical records are not always exclusively attributed to patients — and such rights may not be assignable. For the sake of prudence, we had to also seek institutional/research- hospitals’ clearance and approval — which in turn were then subject to different administrative and logistic factors and regulations. Consequently, it took nearly five years to collect sufficient MRI samples to make this research possible.

Conclusion
In brief, the findings show that better results were obtained with LDA and NDA. The observed difference between the two imaging modalities was previously and repeatedly proven to be due to 3T MRI having higher resolution and able to capture more details. Lastly, LDA and NDA could be useful tests for pre-screening — provided ruling-in/ruling-out semantics are well defined and the KBS is well trained.

Data and software availability
MaZda Package v5 RC HG available from: http://dx.doi.org/10.17632/dkxyrzwpzs.1

F1000Research: Dataset 1. 1.5T data, 10.5256/f1000research.10401.d146782.
F1000Research: Dataset 2. 3T data, 10.5256/f1000research.10401.d146783.
F1000Research: Dataset 3. 1.5T, 10.5256/f1000research.10401.d146784.
F1000Research: Dataset 4. 3T, 10.5256/f1000research.10401.d146785.
F1000Research: Dataset 5. 1.5T, 10.5256/f1000research.10401.d146786.
F1000Research: Dataset 6. 3T, 10.5256/f1000research.10401.d146787.


Author contributions
HG: experimenter, research designer, and writer; LS: supervisor, sample provider, clinical feedback; MRL: coordinator, sample provider, clinical feedback; MS: software maker, updates, technical feedback.

Competing interests
No competing interests were disclosed.

Grant information
Medical University of Lodz & Polish Research Committee and affiliated institutions and hospitals; Self-funded; MRI cost was covered by Polish National Health Fund, grants and financial aid from Swedish Ministry of Education and Research/Centrala Studiostödsnämnden and from U.S. Department of Education.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
The authors gratefully acknowledge prof. Rafał Pawliczak and MUL staff for research coordination, logistics, and administration; prof. Paweł Liberski and MUL neuropathology department for funds to cover MRI expenses; prof. Ludomir Stańczyk and Barlicki hospital staff for sample supply and clinical feedback; prof. Maria Respondek-Liberska and Matki Polki Hospital for sample supply and clinical feedback — as well as prof. Tadeusz Biegański, ICZMP director; prof. Michał Strzelecki and TUL staff for providing MaZda software and technical feedback.

References
Open Peer Review

Current Referee Status: ✗ ✗

Version 1

Referee Report 02 August 2017
doi:10.5256/f1000research.11207.r24724

Sanjay Kalra¹, Abdullah Ishaque²
¹ Department of Medicine, University of Alberta, Edmonton, AB, Canada
² Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, Canada

The study aims to investigate whether texture analysis can be used as an ROI classification tool in fetal MRIs.

The authors used data collected from 1.5T and 3T MRI scanners. Commercially provided software Mazda was used to perform texture analysis and B11 was used for classification.

There are several concerns regarding this manuscript:

1. The exact objectives and aims are not clear, either in the Abstract or Introduction.

2. In the Introduction there is too much information on topics which are not relevant to the paper. The dialogue verges into the arenas of ethics, religion, politics, and medical jurisprudence. What is required here, yet missing, is a more focused rationale for the study, and background material on the methods used, namely texture analysis and artificial networks. Thus, I disagree with the stance of the authors’ (as stated in the Methods, “Advance notice to readers”) to not even offer a brief description given that these methods are otherwise quite abstract to the uninformed reader.

3. The authors do not provide basic information about their data including the number of subjects, age, gender, etc. The clinical and demographic information should be provided to allow the readers a sense of the size and characteristics of their datasets. Additionally, no information regarding the imaging dataset is provided including image resolution and acquisition parameters.

4. The manuscript is hard to follow. The method section is fairly complex and convoluted. The authors should consider removing information from the Methods that does not add value to the manuscript in allowing a third party to replicate the study. Furthermore, results should be removed from the methods section as they should be provided in the Results section.

5. The authors used normal tissue to test for classification of potential “abnormal” tissue. This is an interesting proof of concept approach; however, to be clinically relevant, a dataset with abnormal anatomy should be tested as well to investigate the power of the texture analysis in differentiating between tissue classes and potential diagnoses.
6. The authors should compare texture classification against clinical evaluation and classification of the MRI images to investigate if texture analysis adds any value to the current practices. It is of note that the authors achieved very high classification rates with texture analysis and should be commended on the study as it does seem to be a large project.

7. Constraints, limitations, and assumptions section: A discussion on the administrative and logistical challenges of this study is not needed and should be removed.

Overall, it is an interesting study. Sections need to be considerably streamlined and focused to the science and methods. More information needs to be added that are crucial in a neuroimaging study. The authors should consider applying texture analysis techniques to anatomically abnormal data and compare their results against human classification.

Is the work clearly and accurately presented and does it cite the current literature?  
Partly

Is the study design appropriate and is the work technically sound?  
Partly

Are sufficient details of methods and analysis provided to allow replication by others?  
No

If applicable, is the statistical analysis and its interpretation appropriate?  
Yes

Are all the source data underlying the results available to ensure full reproducibility?  
No

Are the conclusions drawn adequately supported by the results?  
Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Neuroimaging

We have read this submission. We believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 02 Aug 2017

Hugues Gentillon, Medical University of Lodz, Poland

Thank you for your comments and suggestions.

We do not agree with some of your comments and suggestions from a practical (clinical) point of view, for the following reasons below.

This article briefly considers existing legislations relevant to protecting patient privacy and clinical data.
You stated that “exact objectives and aims are not clear” to you. However, you did not explain why, and we feel that some of your comments mentioned are a matter of personal preference. For example, you ask us to publish details about ‘patient demographics’, but we feel that this is unnecessary, in this case.

We strongly recommend you to read:


Competing Interests: None.
Provided information on methods is insufficient

I would like to refer to this report http://www.humanbrainmapping.org/COBIDASreport for guidelines on what to report for MRI studies in general.

In particular, there is no information provided on how the MR images were obtained, this includes missing information on the type of MR sequence, its parameters, vendor of the equipment, etc.

There is no information on the nature of the MR image preprocessing. One of the issues with fetal MRI is the impact of unavoidable motion of the fetus during the scan. This aspect is not touched upon in the manuscript.

Analysis description assumes familiarity with the MaZda package. Here is an example:

"360 parameters were extracted with MaZda (histogram: 9; co-occurrence matrix: 220; run-length matrix: 20; gradient matrix: 5; auto-regression: 5; Haar wavelet: 28; geometry: 73). Parameters' names are provided in the appendix at the end of this manuscript."

Dataset 7 contains a plain list of names such as "GeoUg" that are uninterpretable without familiarity with the MaZda package (which in turn only runs on outdated windows machine (98,2000,XP, according to the website), and source code is not available).

Structure of manuscript

Especially the methods section does not contain typical sections, such as "MRI acquisition", "Participants", etc. Instead, it has "Advance notice to readers" that states that it is impossible to provide an introduction to machine learning. While that may or may not be true, I consider it problematic that the KBS is only described at a conceptual level, while there is extensive space devoted to the development history of MaZda, which seems irrelevant in the context of this study. (Note that the statement: "ANN is a self-organizing algorithm" is not true in its generality)

The heading levels seem to be off at times. "Computer vision" is a subsection of "Clinical trial registration".

In general the section heading should be more indicative of the content. The is "Customization" which reports on adjustments in the original procedure, but also on how files were renamed. The discussion has a section "Constraints, limitations, and assumptions" which essentially restates that data acquisitions took several years.

The authors should reconsider the components of the manuscript. What is presented as "datasets" are actually Excel tables with text content in their cells that are pretty much tables that should go into the manuscript. Proprietary formats are inappropriate for sharing data. Additionally data types of shared data should focus on re-use, i.e. numerical values of result statistics should be shared as such, and not embedded in textual descriptions.

Citations
The author often cite several publication in a single context. It was frequently difficult for me to see why particular choices were relevant. For example, in the context of "Due to administrative and logistic delay as well as finding volunteers and expensive cost and availability of MRI, it took us nearly five years to collect the magnetic resonance (MR) samples." Stipp H: Using neuroscience to improve ad impact: How new research tools can advance cultural marketing. Journal of Cultural Marketing Strategy. 2016; 1(2):193–202. and a news item on "Portable MRI developed at Los Alamos".

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 13 Feb 2017

Hugues Gentillon, Medical University of Lodz, Poland

Thank you for the report. We totally agree with your comments, but the purpose of our paper was to determine a hypothetical methodology for texture classification of closely-related anatomical structures in fetal brain magnetic resonance images.

**Competing Interests:** No competing interests

Discuss this Article

Version 1

Author Response 06 Sep 2017

Hugues Gentillon, Medical University of Lodz, Poland

In version 2, we did not change the scientific content of this paper for the following reasons:

1. In this feasibility study, we proposed that "the observer (in this case the physicians working in the field) as the direct trainer (the programmer) of a KBS designed as a customizable, conceptual framework." In the introduction, we further reiterate (re-emphasize) that "the ultimate goal of this feasibility study was to gather knowledge for the practicality of a proposed project seeking to improve diagnostic accuracy and precision, by extending HCI usage in medicine." Last but not least, we clearly stated that this research is not a clinical trial and registration was not required.

2. We feel that the reviewers were not so familiar with the process of clinical trials (esp. the various stages). In this manuscript, we cannot teach everything about clinical trials, just as we cannot also teach everything about the science of artificial neural network.

3. Herein we are taking the opportunity to further explain the purpose of a “feasibility study”. A feasibility study is the earliest stage of clinical trials. Its focus is on pre-clinical development, pre-clinical studies and/or non-clinical studies. A feasibility study itself also has different parts. This paper focuses on the initial phase. Herein we are not testing a drug but a "robotware" -- i.e. a software, a piece of a potentially larger
A.I. system.

4. We appreciate all the referees' critics and suggestions. Unfortunately, we had to strongly rebut the negative comments because they were influenced by misinterpretations and earlier negative comments. This manuscript is not, per se, a report of a clinical research but a feasibility study. In terms of clinical trials, a feasibility study may contain pre-clinical and/or non-clinical components. In the future, another experiment may be done to extensively test for pre-clinical components, such as biocompatibility and safety. Before we even get there, we had to carry out this preliminary research to investigate the non-clinical aspects of the project. That was the intended interpretation being sought from the readers.

5. We must not confuse scientific medicine with clinical medicine. Scientific medicine is hypothetical and experimental. Clinical medicine directly deals with applications, diagnosis, treatment, prognosis, etc. At this stage, the research focus is primarily on the non-clinical aspects of the project. As a result, we decided to omit unnecessary clinical information from this publication.

6. Why did we deviate from the philosophy of testing pathological versus non-pathological? The reason is computational semiotics. Unlike adolescents and adults, children naturally learn the architecture of normally spoken language without conscious learning. They perfectly imitate pronunciation and figure out the rules by themselves. In this experiment, our training approach was somewhat based on the aforementioned phenomenon. At this stage, we do not need various types of cancer tissues to train the software to recognize what is a normal texture for thalamus, for example.

7. It is important to also understand that "abnormal" does not necessarily mean pathological. In terms of neuroplasticity, our testing methods do have a clinical relevance.

We invite referees to read this paper with an open mind. Here are some inspirational words of wisdom from Charles F. Kettering: "People are very open-minded about new things… as long as they're exactly like the old ones." If we have to please publishing expectations, then it is not a feasibility research.

**Competing Interests:** None.