James Hizon

Data Science in Stratified Healthcare and Precision Medicine

Final Course Assessment

1) Discuss both the data science aspects and the medical healthcare aspects, including recent or future advances in technology.
   - Should explain how combining these two topics will particularly improve precision medicine or stratified healthcare.

2) Based on the same two topics, you should reflect on the challenges of these two topics combined. This should be from both a technical and medical or social point of view.
   - Should also discuss opportunities that will come about if such challenges are overcome.
   - Should include a fictitious example, demonstrating such challenges and how these can be overcome to provide new opportunities.

3) Finally, you should relate the same two topics to ethical issues and discuss these from both sides, providing an unbiased opinion.

Choose two topics from the list below and discuss how they can be combined to improve precision medicine or stratified healthcare:

- Sequence Processing
- Image Analysis
- Network Modelling
- Probabilistic Modelling
- Natural Language Processing
- Process Modelling

Topics for My Assessment:

- Image Analysis
- Process Modelling

Medical Image Analysis:

What is medical image analysis? Medical image analysis is the science of solving and analyzing medical imaging modalities and digital image analysis techniques. Different imaging modalities involve geometric, X-ray (2D/3D), MR-Images (2D, 3D, 4D, etc.), tomographic methods, microscopic images (standard staining method and HMC), SPECT (Radioactive isotopes), Ultrasound, and different artificially created images (bulls-eye for hearts).

Medical problems that exist include diagnosis (identifying the nature of an illness by examination of symptoms), follow up on treatments, comparing different treatments, patients and drugs, online imaging for active intervention and predicting development.

Image analysis problems include segmenting (dilineating different organs), classification (determining types of leukocytes/white blood cells), registration by comparing different modalities and patients, reconstruction (making 3D measurements), measuring flow (inside aorta, the main artery), reconstructing flow fields (inside the heart), building shape priors efficiently, and visualizing results.

Cancer Center AI

In terms of recent technologies, using Deep Learning and Machine Learning (ML) platforms, Cancer Center, an artificial intelligence platform, has developed specialized algorithmic solutions to analyze medical images for oncology and radiology for faster and better human accuracy. Available solutions both in API and Web Platforms, are designed to offer fast and better access to help develop a therapeutic approach quickly in order to help patients and physicians by reducing anxiety caused by not knowing what they are dealing with. Their solutions automate all the work, from segmentation of images, to finding regions of interests, generating statistical descriptions of images, counting cells, mitosis, and even recognizing the type of cell.

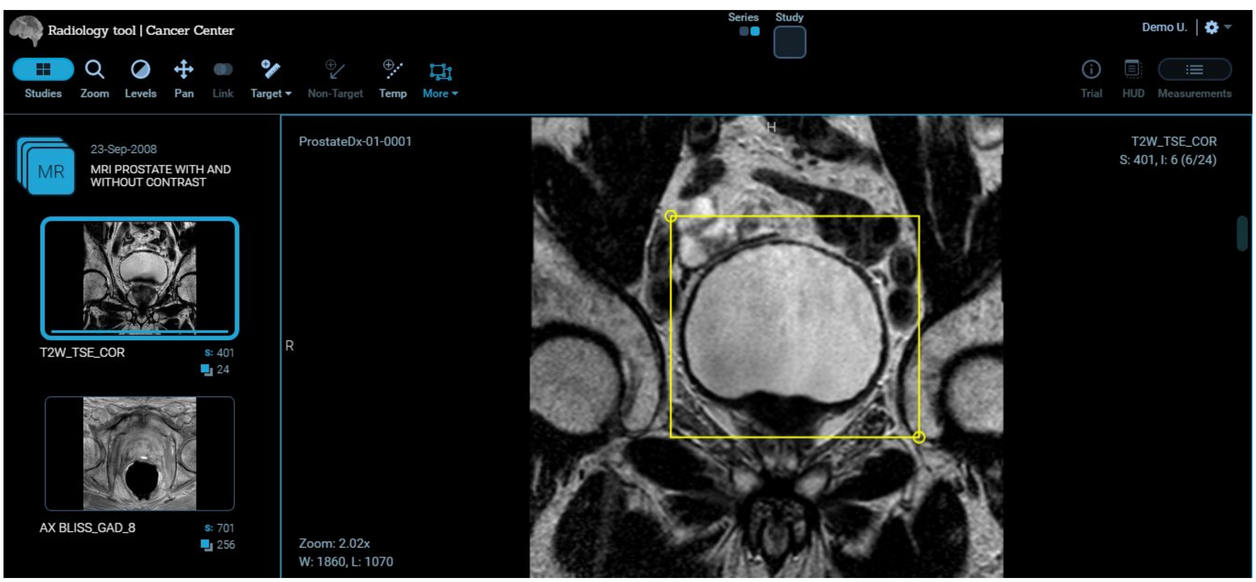
Technologies being used by Cancer Center AI include machine learning and deep learning, automated medical image analysis, MRI (Magnetic Resonance Imaging), PET/CT scans, Pathology Images (Whole slide imaging), and general planimetry.

MRI (DICOM Analysis):

Cancer AI's mission is to combine the state of the art computer science research with medical imaging to improve the quality of diagnosis, save doctor's time, and give patients a better chance.

Their case study involved segmentation and classification based on MRIs of brain. The samples were split into 2 categories: Oligodendroglioma and Astrocytoma (glioma tumor). Part of the study was to find the correlation between MRI and pathology based classification. Radiology images of the same patient varied in terms of shape, average intensity level and head position. The challenge was to select the proper cluster that corresponds to the tumor area. Symmetry analysis of the hemispheres proved to be a good tool for cluster classification (whether it is it a tumor or not).

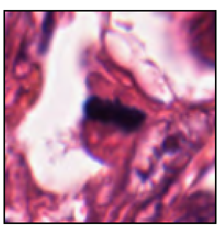
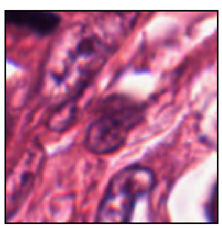
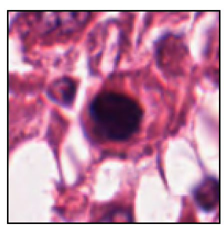
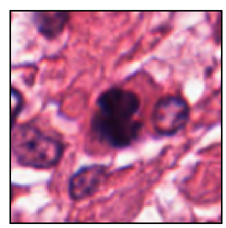
Intensity, shape, position and textural features were extracted from the T1, T1C, FLAIR, and T2 scans. Based on evolution and selection of features, tumor position and distribution of intensity in FLAIR image was most relevant, due to high correlation and highest importance. The lesions were particularly bright or dark in FLAIR vs. T2 scans. In terms of textural pattern based classification, more samples translated to higher effectiveness, and Random Forest classifier validated with k-fold cross validation with an average accuracy of 87.0%.



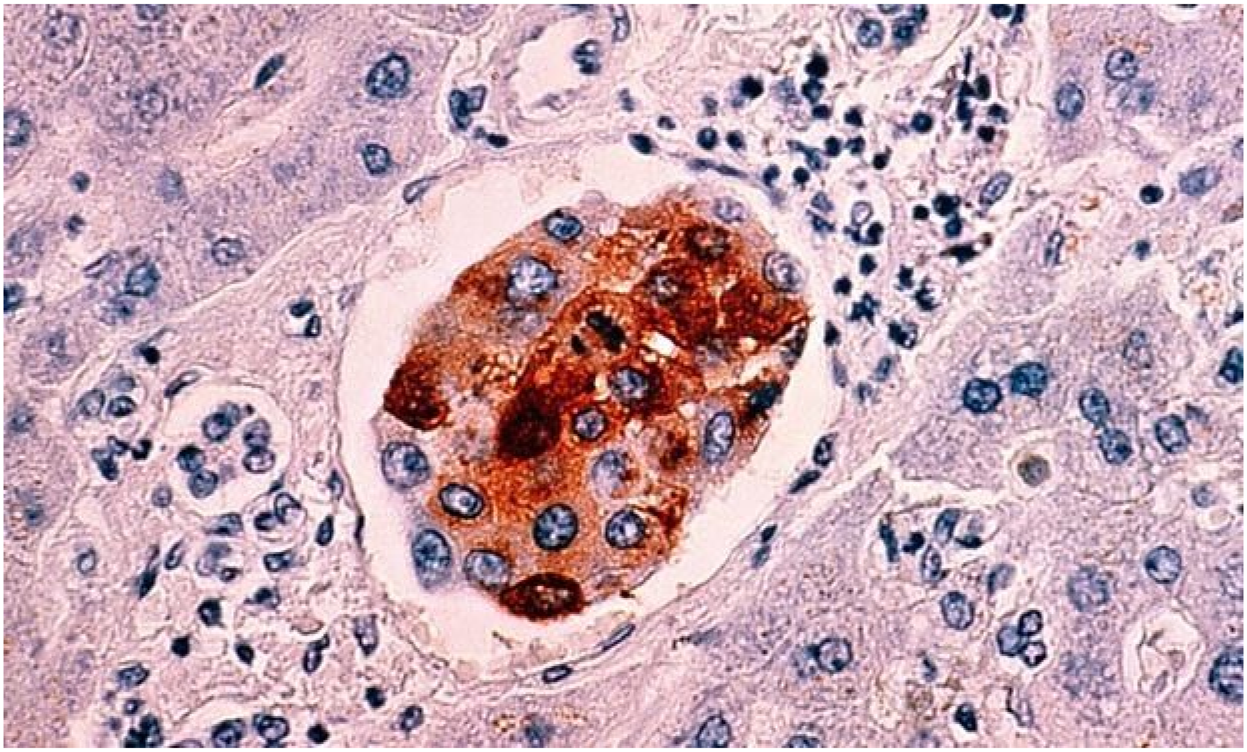
Digital Pathology

For precision, the best method of seeing a single tissue is to examine a tissue under the microscope. Every histopathologist would agree that sometimes finding the most descriptive part of the slide is a harder task than it is to make the diagnosis itself. In order to solve this problem, we can localize the most important parts of the slide. Moreover, we can extract the very factors that influence diagnostic models the most. Cancer Center AI offers Manual Whole Slide Imaging Software (Pathocam) to save pathology images by using your own microscope equipment and PathoViewer to zoom/view, annotate and share it by web/browser.

Image segmentation and classification of tumor cases from histopathological samples are important parts of today's medicine. This modality of medical image analysis is crucial in order to make each particular cancer diagnosis reliable. On a global scale, it is becoming more important to offer tools for digital pathology data sharing and organize community based activities. Cancer Center AI offers solutions that help not only in making the process of segmentation and classification easier and much faster, but also making it possible for specialist all over the world to share, discuss and analyze troubling cases from each other.

Challenges of pathology include segmentation of nuclei, classifying nuclei types, making a statistical summary of samples, finding regions of interest, classifying tumor type and analysis of mitosis. The solution to these problems involve case study with professionals, mathematical modelling, deep learning, machine learning, image processing, and data analysis.



PET Analysis:

Inside oncology, the study and treatment of tumors, Positron Emission Tomography (PET) is widely used in diagnostics of cancer metastases, in monitoring the progress within the process of cancer treatment and planning of radiotherapeutic interventions. The difficulty lies within accurate and reproducible study of the tumor in PET scans, yet is vital for delivery of appropriate radiation dose, minimizing adverse side-effects of the therapy, and reliable evaluation of the treatment. The goal is to use PET for providing clinicians with intelligent software supporting accurate, efficient and reproducible delineation of tumor.

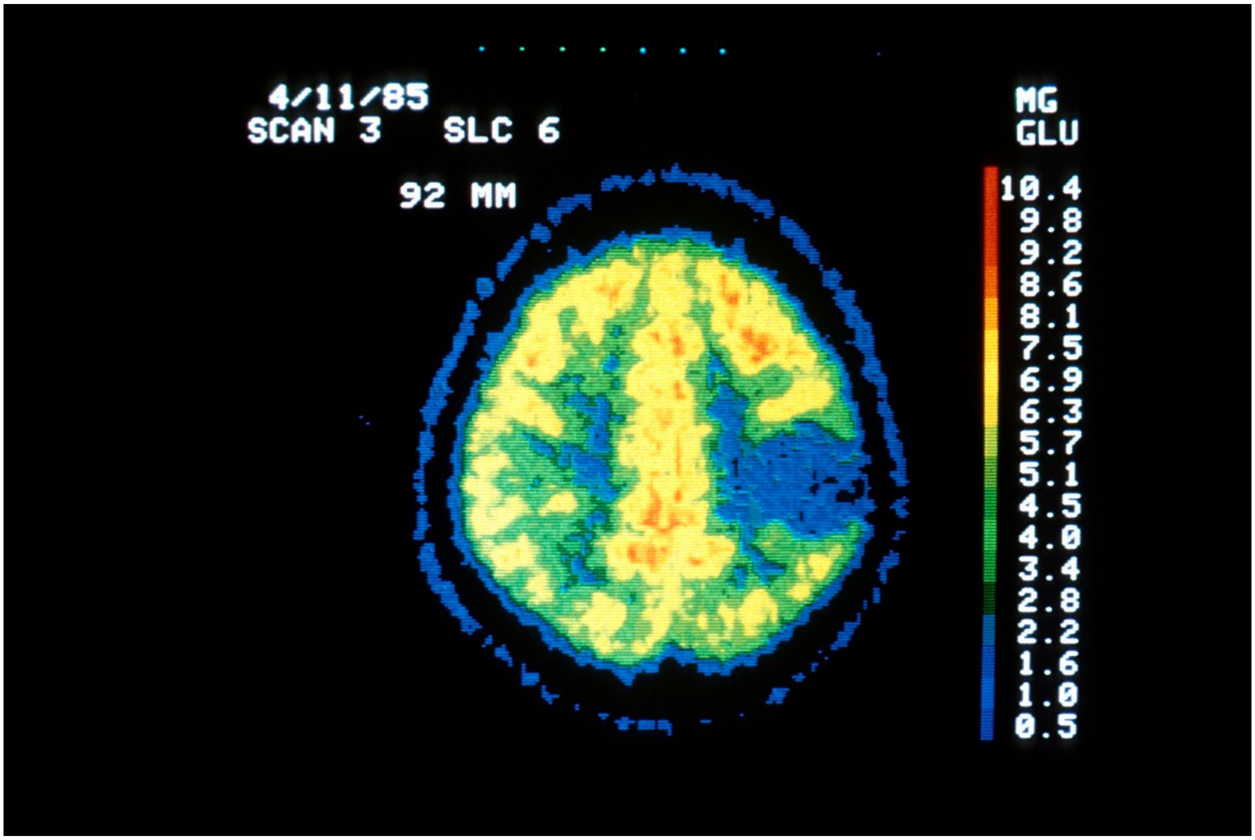
PET and CT (Computer Tomography) scans are standard imaging tools used by physicians to pinpoint disease states inside the body. A CT scan identifies a suspected tumor, whereas the PET scan confirms if the tumor is malignant and if it has spread. By combining PET and CT scans, physicians can more accurately diagnose and identify cancer, heart disease and brain disorders.

In order for PET to work, a patient is first given a radioactive glucose through an IV. All cells need glucose for energy, however, cancer cells use glucose faster than normal cells. A PET scan is used to further measure metabolic activity by detecting radiation emitted when cancerous cells absorb this glucose. 3D-images can then be generated by the computer showing the detected activity throughout the body.

Benefits of PET/CT scan include higher accuracy with better image quality, higher conveniency to have both scans performed simultaneously, and less error due to patient positioning changes.

The scan results are evaluated by a radiologist or physician trained in nuclear medicine. For ensured accuracy, the Cancer Center uses an additional step of having another read the scan.

Cancer Center AI has published their own machine learning methods for accurate delineation of tumors in PET images. In addition, they have developed their own solution for PET images segmentation: SterSEG.



Breast Cancer Database

Furthermore, Cancer Center AI has official image databases for breast, lung, prostate and skin cancer. One example from *The Cancer Imaging Archive (TCIA) Public Access*, is the CBIS-DDSM (Curated Breast Imaging Subset of DDSM). This particular dataset is updated and a standard version of Digital Database for Screening Mammography (DDSM), a database of 2,620 scanned mammography studies. It contains normal, benign and malignant cases with verified pathology information. The images have been decompressed and converted to DICOM format. Updated ROI segmentation and bounding boxes, and pathologic diagnosis for training data are also included. Given public access of imaging data, this allows for the support of clinical research. In my opinion, I would agree that data sharing is a crucial component to strengthen the future of cancer research.

Business Process Modelling:

Business process modelling and notation (BPMN) is a graphical representation of a company's business processes or workflows, as a means of identifying potential improvements. The focus is on discovering, designing, deploying, interacting with, operating, optimizing and analyzing end to end processes.

BPMN diagrams are composed of five main categories of elements: flow objects, data objects, connecting objects, swim lanes and artifacts. Flow objects include events, activities, and gateways. Data objects provide information about what activities require to be performed and/or what they produce. Data objects are also represented by data inputs and outputs, and use the data store element for retrieval. Connecting objects include sequence flow, message flow and association to show the order in which the activities are performed in addition to showing communication flow. Swim lanes include pools and lanes to partition and organize a set of activities. Lastly, artifacts include groups of graphical elements within the same category, and text annotations to provide additional text information for the reader.

Three basic models of processes inside BPMN Diagrams include Private Processes, Public Processes, and Choreographies. Different types of diagrams are used in the process. Private BPs, which are internal to the organization and can be executable (if modeled for such purpose) or non executable. Public BPs, on the other hand, represents the interactions between a private BP and another Process or Participant. The following is an example of a Public Process.

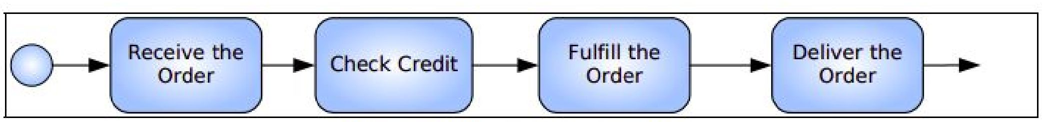


Figure 4. Private Process example.

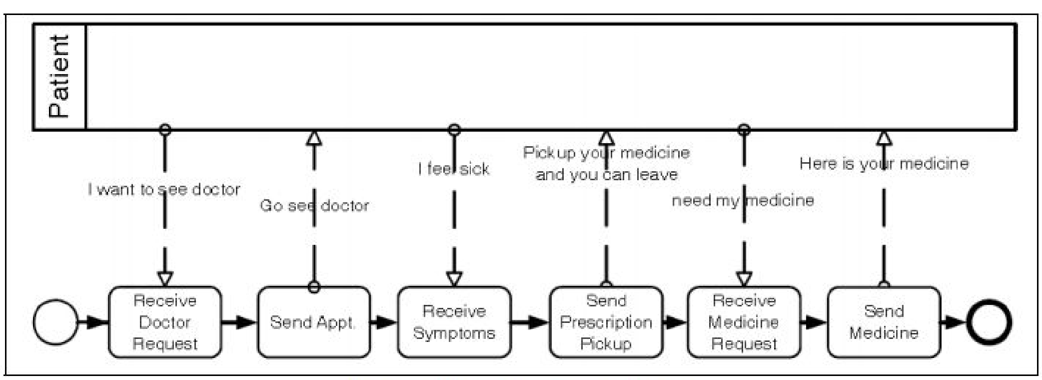


Figure 5. Public Process example [3].

Choreographies express the expected behavior in an interaction between participants. The following is an example:

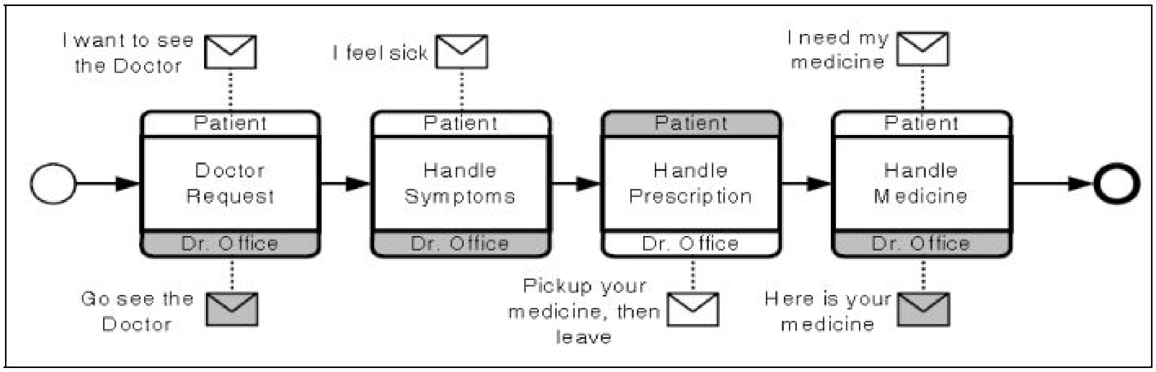


Figure 6. Example of Choreography [3].

Collaborations, including processes and/or choreographies represent the interaction between two or more business units. Usually, collaborations contain two or more participants (Pools) who interchange messages.

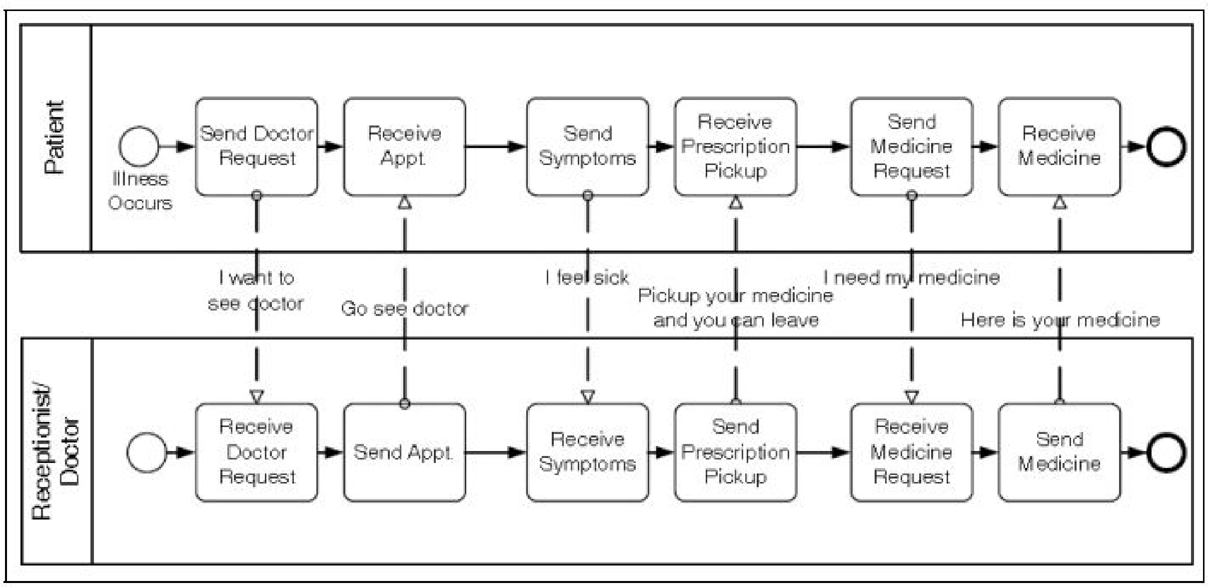


Figure 7. Example of Collaboration [3].

Application of BPM are vital inside a competitive healthcare market, where hospitals have to re-organize their structure and operations to become more responsive organizations with a patient service as efficient as possible. Thus, not only will labor and capital, but also information will be a critical resource: availability, correctness, and facilities to process information, are crucial for an efficient patient. Furthermore, implementation of properly designed processes are a key aspect for a good use of information systems. The lack of control in processes used to deliver medical care is clearly a major problem in the context of preventable medical errors with lethal damages and high economic costs in hospitals. However, care safety assessment, activity-based analysis, workflow technology and knowledge of care pathways provide the methodological basis for continuous quality improvements in the health sector.

Interesting examples of BPMN implementation involve modelling anatomic pathology processes in a public hospital in Spain (Garcia et al). Similar experiences have been published in processes like telemedicine (remote diagnosis and treatment of patients using telecommunications technology) and post-transplant hepatitis (Parra et al). A benefit is seen in modelling care processes and that it can benefit from adaptations to healthcare environments to be optimally exploited. Another more recent work of BP models in hospitals involves a new process management system for defining and reconstructing clinical care processes by using the BPMN language for modelling and executing processes (Strasser et al). The main goal of the system is to assist hospital operators and clinical process managers to detect discrepancies between defined and actual clinical processes, in addition to identifying the main causes for high medical costs.

Two more hospitals include The Chester County Hospital experience in BPM and a Dutch academic hospital for implementation of BPM. The first hospital experience reports a workflow system integrating clinical, operational, and financial processes to provide a significant improvement in patient's safety, efficiency, and particularly in bed management and infection control. A recent application of BP models at a Dutch hospital evaluates interaction-centric process modelling method, which is a case study of healthcare human collaboration processes aimed to study the care pathway performed by the head and neck oncology team. More hospital BPs modelling has also been employed in radiology interpretation. Overall, BP modelling and simulation has reached a high level of maturity.

In combination with medical image analysis, BPMN models support the communication between domain experts and a computer scientist. Healthcare applications include anatomic pathology, telemedicine, chronic patient care, infection control, and patient's safety.

Difficulties that arise from BPM in healthcare include the complexity and special characteristics of processes involved in this sector. The traditional way to structure an organization is the formation of departments and vertical functional units, consisting of individuals with a similar area of expertise. A transition, however, is needed from viewing the healthcare organization as a number of departments, a patient-centered care model focusing on performed business processes. Another issue is working alongside multiple professionals from different departments working on shared tasks with a common goal. BPMN does not support explicit modelling of shared activities, and so activity repetition in different lanes is needed. The complexity of patients' treatments at a hospital may involve many exceptions that occur in healthcare processes. In the healthcare sector, we also see that most of these experiences of providing information relevant to care of patients in the right place and right time are solely in the initial phases, and the development of efficient information systems in this sector is growing slowly. Limitations can be found in discontinuing healthcare processes between hospitals and primary care settings when applying BP modelling and simulation. Thus, a challenge is presented to bridge this gap for process efficiency in patient care organization.

Information systems interoperability in healthcare organizations, which is developing slowly, can benefit from the use of BPMN to achieve its main goal of providing continuity care and ensuring security and privacy.

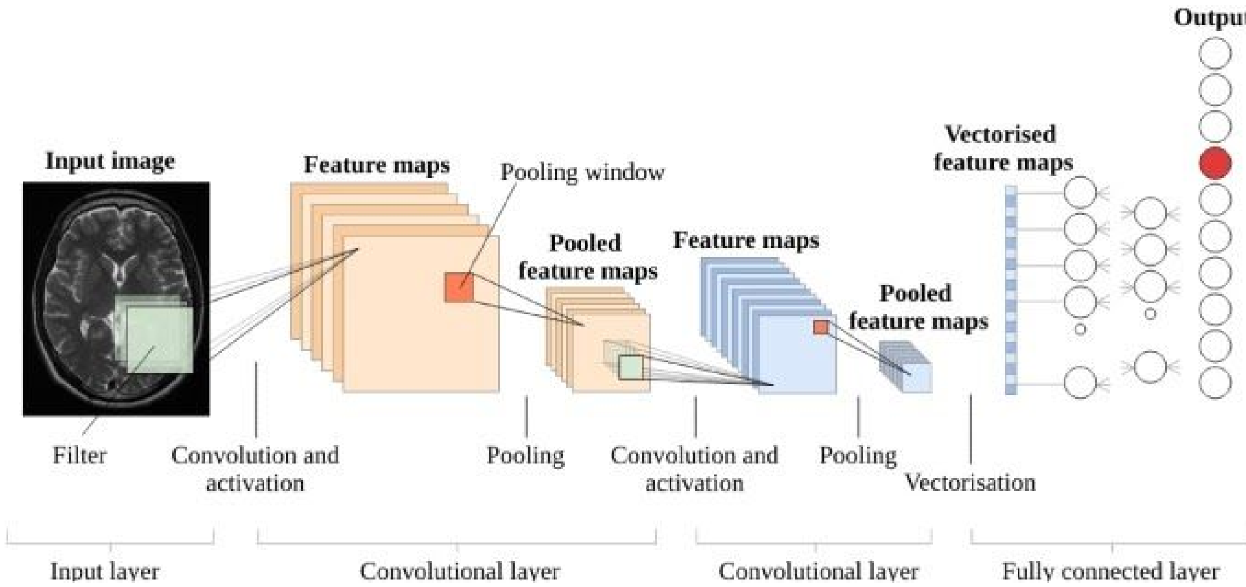
Combining Medical Image Analysis and Process Modelling

In combination of medical image analysis and process modelling, we can observe how medical image computing is growing in significance in terms of medical diagnostics and image-guided therapy. Recently, image analysis systems are seen at use to integrate advanced image computing methods and are used in practice for the extraction of quantitative image parameters or supporting the surgeon during a navigated intervention.

Recent developments and advances in the field of modelling and model-based image analysis have been discovered. Using process modelling, we can see enabled improvements of image analysis algorithms in terms of automation, accuracy, reproducibility and robustness. In addition, we can expect model-based image computing techniques to open up new perspectives for prediction of organ changes and risk analysis of patients. Complex models are used in different medical applications and medical images like radiographic images, dual-energy CT images, MR images, diffusion tensor images, as well as microscopic images. Thus, given these implementations, we can observe the high potential and range of modelling applied to medical images.

The usage of model-based image analysis methods helps us improve our segmentation quality, in addition to accuracy and reproducibility of quantitative image analysis. Moreover, these models enable new, powerful insights towards a deeper understanding of complex dynamic mechanisms within the human body. Therefore, model-based image computing methods are crucial for improving medical diagnostics and patient treatment in the years to come.

In terms of data modelling, convolutional neural networks are a powerful implementation for representation of images and other structured data. A typical CNN has multiple layers of convolutions and activations, often combined with pooling layers, and is trained using backpropagation and gradient descent for standard artificial neural networks. In addition, CNNs can implement dropout regularization and batch normalization.



When architecting a CNN for a particular task, there are multiple factors to consider, including understanding the task to be solved and the requirements to be met. Several famous CNN architecture worthy of checking out include AlexNet, VGG, and GoogleNet. Notice the use of RELUs, dropout regularization, splitting the computations on multiple GPUs, and using data augmentation during training within AlexNet model. This is an example of optimization techniques to consider for tweaking our convolutional neural network for higher accuracy.

The future, indeed, is in the hands of AI. By AI, we mean MRI attached to deep learning and medical imaging. Convolutional neural networks are used for efficiency improvement in radiology practices through protocol determination based on short-text classification. In addition, applications may include reduction of gadolinium dose in contrast-enhanced brain MRI by an order of magnitude without significant reduction in image quality. Applications of deep learning in radiotherapy, PET-MRI attenuation correction, radiomics, and theranostics in neurosurgical imaging to combine confocal laser endomicroscopy with deep learning models for automatic detection of intraoperative CLE images on-the-fly, are to be observed.

Another key important application of interest, includes deformable image registration that enables quantitative analysis across physical imaging modalities across time. An example includes a "cue-aware deep regression network" that learns from a given set of training images and real-time prostate segmentation during prostate biopsy, which utilizes temporal information in the series of ultrasound images.

Overall, process modelling applied to medical images will be expected to be in the rise in the years to come. We should expect a stronger desire for clinical practices to include process automation to produce more accurate insights within medical images.

Technical and Medical Challenges

Some of the issues that may arise in image analysis is the validation of segmentation methods For some scientists, the image data and reference is usually not available and it is impossible to reproduce the conclusions.

The solution to his problems is found inside databases, which provide access to medical datasets and segmentation results are emerging. Another solution may include hardware and software phantoms, which are artificial data that exhibit characteristics similar to real data.
A more flexible options involves the creation of software phantoms through geometric modelling software. Physical phantoms play an essential role in medical imaging to verify an imaging device and calibrate its parameters.

More challenging problems from Kilo- to Terabytes lie within factors due to large data volumes. These may include medical image management and image data mining, bioimaging, virtual reality medical visualizations and neuroimaging. Given the increasing amount of data, image processing and visualization algorithms need adjustment. The solution to this lies in scalable algorithms and advanced parallelization techniques using GPUs.

From "Deep Learning and Challenges in Medical Image Processing" by Jorrit Glastra, we can observe three main challenges in medical image processing. The first factor lies inside the need for access to large high-quality labelled datasets for training. A possible fictitious solution may involve augmentation. Augmentation is used by scientists to double, triple or even quadruple datasets, by modification of the available images so that new images are created which still shows the same characteristics, but slightly different and seen as training material by the neural network.

The next challenge is that the most successful deep learning models are currently trained on simple 2D pictures. CT-and MRI-images are usually 3D, adding literally an extra dimension to the problems. Thus, conventional x-ray images may be 2D, where current deep learning models are not adjusted to these either. Thus, in order to solve this problem, we must gain more experience with the application of neural networks to 3D medical images.

Lastly, due to non-standardized acquisition of medical images, we face one of the biggest challenges. The more variety inside the data, the larger the dataset needs to be to ensure the deep learning network results in a robust algorithm. A solution for solving this problem lies in application of transfer learning, a pre-processing technique applied to overcome scanner and acquisition specifics.

Overall, deep learning will play a major factor in the reality of medical image analysis, as companies begin to apply it for production. We must see the sizes of medical image datasets increase by aggregation (clustering) or augmentation, gain experience with 3D images, and have the medical community standardize medical acquisition for real progress.

Ethical Issues of AI, Process Modelling and Medical Image Analysis

Adoption of AI practices to process medical images for classification will face ethical concerns. Challenges may include addressing how databases and algorithms may introduce bias into the diagnostic process and that AI may not perform as intended. This can lead to a potential for patient harm. In this case, major implications are seen in how clinical laboratories and anatomic pathology groups may use artificial intelligence. Thus, medical laboratory executives and pathologists should be aware of possible drawbacks to the use of AI and machine-learning algorithms within the diagnostic process.

AI’s ability to improve diagnosis, precisely target therapies and leverage healthcare data has lead to important predictions for precision medicine and personalized healthcare. Accenture offers a prediction that hospitals will spend $6.6 billion on AI by 2021, resulting in an annual growth rate of 40%, leading to a potential creation of $150 billion in annual savings for the United States healthcare economy by 2026. For this reason, healthcare providers may be too quick to adopt AI, yet find that AI may not perform as intended. The issue at hand, is that computer algorithms based off limited data sources and questionable methods can put patients at risk. An example may include housing data predominantly about people of Northern European descent, which can be a problem of misrepresentation and ultimately, bias. Limited information and diversity will limit the accuracy of medical insights. Making decisions about women’s health where women have been “under-treated vs. men” is another example.

Combating bias will be a major factor. Bias exists inside healthcare by humans, design and implementation. A problem that may arise may occur in implementing algorithms that have different interests such as the goal of making money or the fact that different treatment decisions about patients are made depending on insurance status or inability to pay.

In light of process modelling, the issue lies within the uncertain nature of processing models. In addition, there lies an absence of accountability and ownership for the management of ethical risks.

In terms of medical image analysis, it may be helpful to include datasets of patients of the age 18 or older. Another ethical issue lies in the need for consent for the use case of patient images. In addition, anonymization of image datasets was involved to remove the following items from DICOM headers: date of birth, institution name, patient name, patient ID, examination number and study date. To ensure correct and only scientific usage of data, benchmark participants have to sign an end-user agreement. An example of safe storage in the cloud is found in implementation of VISCERAL Benchmarks running on Microsoft (Azure), where only authorized participants who signed the end-user agreement has access to stored data. Lastly, long-term usage of data may involve an additional amendment to allow BRATS dataset for computer-based segmentation of brain lesions be maintained by VISCERAL.

Furthermore, acquiring medical imaging research data is not an easy process. Data privacy needs to be respected and agreed upon by medical ethics commissions of participating institutions. Safe storage and access of data in the cloud may lead to a continual, protected medical data analysis given the risks of data misuse can be reduced in a straightforward way.

Resources:

http://prac.im.pwr.edu.pl/~hugo/HSC/imprezy/EMS_school2010/heyden_lecture.pdf
https://cancercenter.ai/
https://www.academia.edu/22012641/Business_process_modeling_in_healthcare
https://www.ncbi.nlm.nih.gov/pubmed/23052259
https://www.ncbi.nlm.nih.gov/pubmed/23052259
https://www.quantib.com/blog/deep-learning-its-potential-and-challenges-in-radiology-ai
https://www.darkdaily.com/could-biases-in-artificial-intelligence-databases-present-health-risks-to-patients-and-financial-risks-to-healthcare-providers-including-medical-laboratories/
https://link.springer.com/content/pdf/10.1007%2F978-3-319-49644-3_3.pdf

More Applications of Deep Learning Models towards Medical Images:

Another important application area is advanced deformable image registration, enabling quantitative analysis across different physical imaging modalities and across time.[22] For example elastic registration between 3D MRI and transrectal ultrasound for guiding targeted prostate biopsy [95]; deformable registration for brain MRI where a "cue-aware deep regression network" learns from a given set of training images the displacement vector associated with a pair of reference-subject patches [96]; fast deformable image registration of brain MR image pairs by patch-wise prediction of the Large Deformation Diffeomorphic Metric Mapping model [97][23]; unsupervised convolutional neural network-based algorithm for deformable image registration of cone-beam CT to CT using a deep convolutional inverse graphics network [98]; deep learning-based 2D/3D registration frame-work for registration of preoperative 3D data and intraoperative 2D X-ray images in image-guided therapy [99]; real-time prostate segmentation during targeted prostate biopsy, utilizing temporal information in the series of ultrasound images [100].

This is just a tiny sliver of the many applications of deep learning to central problems in medical imaging. There are several thorough reviews and overviews of the field to consult for more information, across modalities and organs, and with different points of view and level of technical details. For example the comprehensive review [101],[24] covering both medicine and biology and spanning from imaging applications in healthcare to protein-protein interaction and uncertainty quantification; key concepts of deep learning for clinical radiologists [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], including radiomics and imaging genomics (radiogenomics) [112], and toolkits and libraries for deep learning [113]; deep learning in neuroimaging and neuroradiology [114]; brain segmentation [115]; stroke imaging [116], [117]; neuropsychiatric disorders [118]; breast cancer [119], [120]; chest imaging [121]; imaging in oncology [122], [123], [124]; medical ultrasound [125], [126]; and more technical surveys of deep learning in medical image analysis [41], [127], [128], [129]. Finally, for those who like to be hands-on, there are many instructive introductory deep learning tutorials available online.