To test the proposed system, we implemented prototypes of some atomic microservices such as color deconvolution that can be used for staining separation and quantification in both PGR and ER image analysis. The result of the analysis classifies each case as "negative", "positive", or "suspect". The cases classified as "suspect" are then further analyzed by experts with a possible software support from other services. For example, the analysis could be refined using other atomic services to identify the tumor area and to evaluate positive cells only in the tissue classified as tumour.

### Discussion

The proposed architecture, although preliminary, seems to provide a relatively simple approach to generic digital slide analysis services in a distributed environment, taking into account the issues coming from the large size of the involved files. Specific security issues have not yet been examined, although SOA security has been abundantly investigated (e.g.,[8]) and apparently provides a sensible solution for security of health records.[9]

However, to describe and retrieve image analysis microservices according to the theoretical model of SOA, sufficiently abstract and shared terms should be adopted, that is, an ontology of operations, of slide contents at different levels (subcellular, cellular, tissue, organ…) and of diseases is needed, like proposed by the MICO project.[10]

The microservices are currently being integrated in the O3IMS PACS in two ways, to be considered representative of two image analysis modalities:

- Pull modality: the pathologist, while viewing the slide through a workstation, decides to carry out some automated analysis, and thus invokes a service from the graphical interface;

- Push modality: after digitization of a slide, the PACS may autonomously invoke some analysis on it, basing on available metadata. This way, when the pathologist will examine the slide, results are already available. Furthermore, some analysis could screen slides to avoid the pathologist the examination of obvious ones (i.e., frankly negative IHC results).

More microservices implemented and integrated within the PACS imply a higher possibility of analysis using either the microservices directly or in combination for a deeper analysis and automatization. The scalable architecture allows a high flexibility in digital slides image analysis, potentially covering a wide number of quantitative analysis over different types of stained images. The main advantage of such architecture is that many different analyses can be performed over potentially heterogeneous data with a limited computational load of the local machine as well as of the communication system.

The system could be updated with the implementation and integration of new atomic microservices as well as their combinations (even with existing ones) to fulfill the needs for image analysis of the clinicians.

### Acknowledgments

The work has been partially funded by the project HEaD – Higher Education and Development - FP1619942003 (Region Friuli – Venezia Giulia).

### References

1. Flynn AJ, Boisvert P, Gittlen N, Gross C, Iott B, Lagoze C, *et al.* Architecture and initial development of a knowledge-as-a-service activator for computable knowledge objects for health. Stud Health Technol Inform 2018;247:401-5.
2. Gao Y, Burns SS, Lauzon CB, Fong AE, James TA, Lubar JF, *et al.* Integration of XNAT/PACS, DICOM, and research software for automated multi-modal image analysis. Proc SPIE Int Soc Opt Eng 2013;8674.
3. Zerbe N, Hufnagl P, Schlüns K. Distributed computing in image analysis using open source frameworks and application to image sharpness assessment of histological whole slide images. Diagn Pathol 2011;6 Suppl 1:S16.
4. Rassias Andrikos C, Tsanakas P, Maglogiannis I. Versatile Cloud Collaboration Services for Device-Transparent Medical Imaging Teleconsultations. In Proceedings. Of CBMS. IEEE; 2017. p. 306-11.
5. Schuler R, Smith DE, Kumaraguruparan G, Chervenak A, Lewis AD, Hyde DM, *et al.* A flexible, open, decentralized system for digital pathology networks. Stud Health Technol Inform 2012;175:29-38.
6. Guo H, Birsa J, Farahani N, Hartman DJ, Piccoli A, O'Leary M, *et al.* Digital pathology and anatomic pathology laboratory information system integration to support digital pathology sign-out. J Pathol Inform 2016;7:23.
7. Inchingolo P, Beltrame M, Bosazzi P, Cicuta D, Faustini G, Mininel S, *et al.* O3-DPACS open-source image-data manager/Archiver and HDW2 image-data display: An IHE-compliant project pushing the e-health integration in the world. Comput Med Imaging Graph 2006;30:391-406.
8. Beer MJ, Hassan MF. Adaptive security architecture for protecting RESTful web services in enterprise computing environment. Serv Oriented Comput Appl 2017;12:111-21.
9. Rezaeibagha F, Win KT, Susilo W. A systematic literature review on security and privacy of electronic health record systems: Technical perspectives. Health Inf Manag 2015;44:23-38.
10. Racoceanu D, Capron F. Towards semantic-driven high-content image analysis: An operational instantiation for mitosis detection in digital histopathology. Comput Med Imaging Graph 2015;42:2-15.

# Deep Digital Convergence of Radiology, Pathology, and Clinical Molecular Biology

**Harry B. Burke**[1]

[1]Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA. E-mail: harry.burke@usuhs.edu

### Abstract

Historically, most radiology, pathology, and clinical molecular biology information has been image-based and sequestered within each domain. This means that their information could not be combined to improve our knowledge of disease and treatment. This paper proposes the convergence of radiology, pathology, and clinical molecular biology through the integration of their digital data in powerful statistical models in order to create information synergies that can be used by trained models to improve risk estimation, diagnostic certainty,

treatment effectiveness, and clinical outcomes. This advance will improve the quality and safety of medical care. Historically, most radiology, pathology, and clinical molecular biology information has been image-based and sequestered within each domain. This means that their information could not be combined to synergistically improve our knowledge of the disease and its treatment. This paper proposes the convergence of radiology, pathology, and clinical molecular biology through the integration of their digital data in a statistical model and the use of the trained model to improve the quality and safety of patient care. During the 20th Century radiology, pathology, and clinical molecular biology were independent clinical domains. Each existed in its own ocular realm – radiologists looked at structural and functional anatomic images, pathologists looked at tissue-based images, and clinical molecular biology looked at biochemical false-color microarray images. Each looked for features that could be used to determine the risk of disease, diagnose disease, and assess the severity of disease. Radiology viewed anatomic images generated by: (i) Röntgen-ray images and (ii) algorithmically constructed images created either from either digital molecular data (for example, magnetic resonance imaging) or from analog cellular data (for example, ultrasound). Pathology viewed enhanced (stains, antibodies, etc.) images of molecular/cellular/multicellular material affixed to slides. Clinical molecular biology viewed analog false-color images of probe-detected gene expression – which could be converted into numeric data, but the resulting numbers were imprecise because of the imprecision inherent in the false-color data. There are several issues related to the visual detection of image-based clinical information. One issue is additivity; each clinical domain possesses some information that is not contained in the other domains (orthogonality), but there was no way to combine the images across the imaging domains in order to create an additive model of the patient and disease. Another issue is observability; there is more information in the analog data than can be seen by an observer, therefore, from a cybernetic perspective, the visual assessment of images results in a loss of information. A third issue is subjectivity; the visual assessment of analog data (signal detection) is subject to high intra and inter-observer variability (error). This lowers predictive accuracy for the three types of prediction, namely, risk/prevention, diagnosis, prognosis/treatment) and, as a consequence, decreases the clinical utility of the information.[1] Fortunately, radiology, pathology, and clinical molecular biology data have become, or are rapidly becoming, digital. This means that the images will no longer be needed – they will be replaced by computational digital data. This transition from analog to digital data permits, for the first time, the convergence of the three domains. The union of these imaging domains will increase the amount, and quality, of clinically useful information through improvements in additivity and observability, and it will reduce error through the elimination of subjectivity. We can employ a conceptual framework, a set of variables and the relations among them that are thought to account for a phenomenon, to help us understand, and guide our modeling of, disease.[2] The body is a unitary, complex biological system[3] which has the following framework: (1) the body is an integrated, interdependent hierarchical organization that is composed of systems, each of which serves one or more biological functions, (2) the complete uncompensated failure of one of the body's necessary systems results in the body's failure – but the body has alternate systems for some functions and they may be able to take over for a failed system; (3) the body can be described in terms of four-dimensional, interconnected levels of analysis, including the molecular, cellular, and multicellular levels; (4) a level is defined in terms of its units and rules (the allowed interactions and activities) and the level's units and rules are the constituents of its functional systems at that level: (5) each level has different units and rules; (6) the levels are interrelated a hierarchical manner; (7) time is different at different levels, i.e., things occur at different rates at different levels, furthermore, body time is not equivalent to level time; (8) complex biological systems are dynamical in that their units are always interacting with each other in order to maintain homeostasis; (9) in terms of the functioning of complex biological systems, the inhibition of an activity is, many times, just as important as the existence of the activity; (10) the body's biological systems self-organize based on their constitutive units and rules; (11) complex biological systems have the ability to adapt to changes in their internal systems, their functions, and their external environment, which means that they evolve over time; (12) complex biological systems have the ability to maintain themselves, protect their existence, and to learn from experience; and (13) the functioning of a complex biological system depends on its present state, its environment, and its feedback and feedforward processes. Finally, higher biological/anatomical levels are constructed from four dimensional multicellular functional units (MFUs). MFUs have at least three characteristics: (1) they are composed of multiple cell types and their local environment, including the extracellular matrix, (2) their cells are co-dependent and spatio-temporally interact in an organized, cooperative manner and (3) they perform one or more biological functions that are required to maintain homeostasis. Biological systems are probabilistic (statistical) rather than deterministic (causal) because they are loosely coupled rather than tightly coupled systems. Tightly coupled means that what the system does in the future is not influenced by what it is currently doing. For example, a spring is governed by the rule of proportionality which means that its response to a stimulus is always proportional to the strength of the stimulus. Because tightly coupled systems are deterministic, they are inflexible, therefore, when conditions change they become maladapted and dysfunctional. Loosely coupled means that the forces affecting the present state of the system can determine, to a greater or lesser degree, the future state of the system. Biological systems are loosely coupled probabilistic systems that can adapt to changes in themselves, their functions, and their environment. For example, in physics, Newton and Einstein created deterministic systems whereas quantum

mechanics (a type of statistics) is probabilistic.[4] Complex biological systems are loosely coupled to the extent that they do not jeopardize their existence. A deterministic system does not allow for choice, there is only one outcome, so there is no uncertainty. A probabilistic system always has choice, there is always has more than one possible outcome, so it is inherently uncertain. Since information is anything that reduces uncertainty[5] (uncertainty is mathematically equivalent to Boltzmann entropy) there is no information in a deterministic system. The amount of information in a probabilistic system is a function of its ability to reduce uncertainty. We are interested in the uncertainty related to the accuracy of our risk, diagnosis, and treatment predictions. For example, a patient's biomarker is informative if knowing its numerical value reduces our uncertainty regarding the patient's outcome (e.g., prognosis). Once we have a framework for complex biological systems, we can begin to understand non-traumatic disease. The homeostatic principle is necessary, but not sufficient, for the normal functioning of a biological system. An important homeostatic mechanism is deviation reduction (negative feedback) – which maintains the system's normal functioning. Disease is a biological system that violates the body's homeostatic principle through the use of deviation amplification (positive feedback) – it is this deviation amplification away from homeostasis that allows the disease to cause the body's failure. Disease has many of the characteristics of a chaotic system. Chaotic systems are a special class of loosely coupled systems that employ deviation amplification processes. In some situations, the body is able to contain the disease but, in other situations, the disease takes control of the body and destroys it. Our job is to help the body regain homeostasis through traumatic interventions, for example, surgery, and/or through molecular interventions, for example, medications. To accomplish this task, we must attend to Sir William Osler who said, "Medicine is the science of uncertainty and the art of probability" – in other words, our goal is to reduce uncertainty by creating statistical (probabilistic) models which integrate multilevel biological information in order to create accurate disease representations which help us understand and defeat the disease. It might be thought that the convergence of radiology, pathology, and clinical molecular biology is about the collapsing of three levels of analysis related to anatomic scale – from larger to smaller scale. But our levels of analysis are not related to scale, rather, they are the hierarchical levels related to the molecular, cellular, and multicellular systems and subsystems. These systems and subsystems are fantastically complex. Even the smallest subsystems are extremely complicated, for example, researchers are just beginning to try to model a single signal transduction pathway using supercomputers.[6] Modeling every system in the body is currently not possible, so what is the point of combining orthogonal biological information? The answer is that we are not going to model all the workings of normal biological systems, rather, we are going to model deviant biological systems (disease) – we are going enter the disease-related data into a statistical algorithm, let the statistical method learn the relevant-for-the-disease relationships (rather than all permissible biological system relationships), create a trained model that accurately represents the disease, and use that model of the disease to provide accurate probability estimates related to disease for risk/prevention, diagnosis, and prognosis/treatment. In other words, we do not need to know all the aspects of a system in order to model those aspects related to the disease and to use that model to improve the quality and safety of patient care. Radiology, pathology, and clinical molecular biology provide the digital molecular, cellular, multicellular data we need to create our disease models. Radiomics is radiographic digital data at the molecular, cellular, multicellular levels.[7] Non-Röntgen-ray radiology is noninvasive and some modalities, such as magnetic resonance imaging, are inherently digital and they can operate at the molecular, cellular, multicellular levels. Pathomics, which is invasive, is the use of statistical algorithms to digitize and learn key features of pathology images and it can operate at the molecular, cellular, multicellular levels.[8] Proteogenomics, which is also invasive, is the acquisition of disease-related digital gene and protein information and, of course, it operates at the molecular, cellular, multicellular levels.[9] Radiomic, pathomics, and proteogenomic digital data can be combined with other digital data, including physiologic and laboratory data, to create an optimal statistical model of the disease. The resulting trained model acts as a surrogate for the disease: to the extent that the model is an accurate representation of the disease, it can inform us regarding the disease's natural course and it can tell us which treatment to select to slow or truncate the disease's progression. In order to select the best statistical method to model the disease we need to formally understand complex biological systems. One way to do so is through network theory, i.e., nodes (functional units) and their connections (the relationships and interactions between nodes). Biological systems can be represented as a multiplex network, i.e., the units at different levels are not separate entities, which means that the data cannot be compressed into a single level model or represented as an aggregated model. Furthermore, we are modeling a dynamical rather than structural network, therefore, a level's units interact with each other and they are constitutive within and across levels (the effects of one level pass through from one level to another), and they can be codependent.[10] Furthermore, we have to be aware of multilevel information redundancy, i.e., that the same information can be shared across levels.[11] Finally, we would like to use as few levels as possible to explain the phenomenon because the more levels there are the sparser the representation. These network characteristics of biological systems have important implications for the convergence of multilevel digital data. One implication is that the statistical method must be able to deal with multilevel, interactional data. Another is that the method must be able to learn from the data. Finally, the method should make few, if any, parametric assumptions. What is required is a universal classifier, one that can learn multilevel dynamical data, including capturing the nonlinearities and interactions, and that will converge on the correct solution

which, in this case, is the creation of an accurate model of the disease. Fortunately, there is such a universal classifier, namely, the neural network.[12] A three-layer backpropagation neural network with an arbitrarily large number of sigmoidal hidden layer units can fit any real continuous function and, given that the solution is in the data and that there is sufficient data, it will find the correct solution.[13] The discriminative accuracy of the trained neural network is a function of how well it models the disease and it is measured by the receiver operating characteristic (ROC).[14] To be clinically useful a model's ROC should be at least 0.70.9. Our remaining task is to create clinical decision support systems (CDSS) that: (1) function in the prediction domains of risk/prevention, diagnosis, and prognosis/treatment; (2) contain powerful neural networks that use the patient's digital data to make accurate individualized patient predictions and that learn in order to improve their performance over time; and (3) effectively communicate to the clinician and patient in real time the individual patient clinical information and predictions required for better shared decision-making and optimal treatment. In summary, the transformation of the clinical domains of radiology, pathology, and clinical molecular biology from analog to digital data, and the integration of their digital data in powerful statistical models, will create information synergies that will improve risk estimation, diagnostic certainty, treatment effectiveness, and clinical outcomes. This advance will improve the quality and safety of patient care.

### Acknowledgement

Presented at the 14th European Congress on Digital Pathology and the 5th Nordic Symposium on Digital Pathology, Helsinki, Finland on 31 May 2018.

### Disclaimers

The views expressed in this manuscript are those of the author and do not represent the views of the U.S. Government or any of its agencies.

### Funding

Support for this manuscript was provided by the Patient Safety and Quality Academic Collaborative, a joint Defense Health Agency – Uniformed Services University program.

### Conflicts of interest

The author does not declare any conflicts of interest.

### References

1. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN, *et al.* Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA 2015;313:1122-32.
2. Carpiano RM, Daley DM. A guide and glossary on post-positivist theory building for population health. J Epidemiol Community Health 2006;60:564-70.
3. von Bertalanffy L. General System Theory: Foundations, Development, Applications. Revised edition. New York: George Braziller; 1976.
4. Schrodinger E. What Is Life? New York: The Macmillan Company; 1945.
5. Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27:379–423, 623-56.
6. Ganesan N, Li J, Sharma V, Jiang H, Compagnoni A. Process simulation of complex biological pathways in physical reactive space and reformulated for massively parallel computing platforms. IEEE/ACM Trans Comput Biol Bioinform 2016;13:365-79.
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278:563-77.
8. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, *et al.* Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep 2018;8:3395.
9. Burke HB. Predicting clinical outcomes using molecular biomarkers. Biomark Cancer 2016;8:89-99.
10. DeFord DR, Pauls SD. A new framework for dynamical models on multiplex networks. J Complex Netw 2018;6:353-81.
11. Stolarczyk S, Bhardwaj M, Bassler KE, Ma WJ, Josić K. Loss of information in feedforward social networks. J Complex Netw 2018;6:448-69.
12. Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE Jr., *et al.* Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 1997;79:857-62.
13. Hornik K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Netw 1990;3:551-60.
14. Swets JA. Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers. Mahwah, NJ: Lawrence Erlbaum Associates; 1996.

# Predicting Prostate Cancer Progression Using a Network of Bivariate Prognostic Models

**Guenter Schmidt[1], Hadassah Sade[1], Nathalie Harder[1], Harald Hessel[2], Maria Athelogou[1], Alexander Buchner[2], Christian Stief[2], Thomas Kirchner[2], Ralf Huss[1]**

[1]Definiens AG, Munich, Germany, [2]Ludwig Maximilian University, Munich, Germany. E-mail: gschmidt@definiens.com

### Abstract

**Background:** The accurate identification of prostate cancer patients with prostate specific antigen (PSA) biochemical recurrence (BCR) after radical prostatectomy is unsolved in oncology. We present a set of prognostic bivariate models discovered by a novel graph-based method integrating tissue phenomics data from immunohistochemistry (IHC) and mRNA-based gene expression profiling. **Methods:** Automated image analysis and co-registration determined spatial properties of cell populations detected in consecutive tissue sections stained for CD3/CD8, CD68/CD163, CK18/p63 and CD34 (Definiens Tissue Phenomics, Munich). For each of the 23 patients (Gleason-score 6-9, pT2, age<=80 years), the resulting tissue phenomics feature vector was expanded with gene expression measurements (nCounter PanCancer Immune Profiling Panel, NanoString Technologies, Seattle) from the same tissue sample. A minimal spanning tree was constructed based on graph nodes representing univariate prognostic features by adding edges representing bivariate prognostic features. Results: The edges of the prognostic network linking IHC with gene expression features comprise: (1) a large distance from CD163+ to CD3+CD8- cells and low MAGEC2 expression indicates low BCR risk, and (2) a small distance from CD34+ to