قائمة ملخص الابحاث العلمية المقدمة من الدكتور وليد أحمد يوسف /– كلية الحاسبات والمعلومات – جامعة حلوان
للحصول علي اللقب العلمى: أستاذ مساعد في تخصص – علوم الحاسب

No.	Title	Author(s)	Place	Date	Rank	ISSN / ISBN
1.	**Learning Algorithms May Perform Worse with Increasing Training Set Size: Algorithm-data Incompatibility**	**W.A. Yousef**; S. Kundu **مشاركون من تخصص آخر:** The George Washington University, Statistics Dep.	Computational Statistics & Data Analysis 74(c): p. 181-197.	2014	International Journal IF: 1.028 5-year IF: 1.373 5-Interdisciplinary Applications (A) 6-Statistics and Probability (A)	ISSN: 0167-9473

In machine learning problems a learning algorithm tries to learn the input–output dependency (relationship) of a system from a training dataset. This input–output relationship is usually deformed by a random noise. From experience, simulations, and special case theories, most practitioners believe that increasing the size of the training set improves the performance of the learning algorithm. It is shown that this phenomenon is not true in general for any pair of a learning algorithm and a data distribution. In particular, it is proven that for certain distributions and learning algorithms, increasing the training set size may result in a worse performance and increasing the training set size infinitely may result in the worst performance —even when there is no model misspecification for the input–output relationship. Simulation results and analysis of real datasets are provided to support the mathematical argument.

خوارزميات التعلم يمكن أن تؤدي سوءا مع زيادة حجم مجموعة التدريب: التعارض بين الخوارزميات و البيانات

في مشاكل التعلم الآلي يحاول خوارزمية تعلم كيفية تعلم تبعية المدخلات والمخرجات باستخدام بيانات مجموعة التدريب. تشوه هذه العلاقة بين المدخلات والمخرجات عادة عن طريق الضجيج العشوائي. من التجربة العملية، والمحاكاة، والنظريات الخاصة، معظم الممارسين يعتقدون أن زيادة حجم مجموعة التدريب يساعد على تحسين أداء الخوارزمية. وتبين أن هذه الظاهرة ليست صحيحة بشكل عام لأي زوج من خوارزمية التعلم وتوزيع البيانات. على وجه الخصوص، ثبت أن لبعض التوزيعات وخوارزميات التعلم، زيادة حجم مجموعة التدريب قد يؤدي إلى أداء أسوأ وزيادة حجم مجموعة التدريب بلا حدود قد يؤدي في أسوأ أداء وذلك حتى عندما لا يكون هناك خطأ فى نموذج المدخلاتو المخرجات. وتقدم نتائج المحاكاة وتحليل مجموعات البيانات الحقيقية لدعم حجة الرياضية.

2.	**Assessing classifiers in terms of the partial area under the ROC curve** http://www.sciencedirect.com/science/article/pii/S0167947313000881	**W.A. Yousef**	Computational Statistics & Data Analysis 64(0): p. 51-70.	2013	International Journal IF: 1.028 5-year IF: 1.373 5-Interdisciplinary Applications (A) 6-Statistics and Probability (A)	ISSN: 0167-9473

Assessing classifiers using the partial area under the ROC curve (PAUC) (or its equivalent, “separability”, that is a function of the chosen threshold of the decision variable) is considered. The population properties of the “separability” as a function only of the trained classifier and the selected threshold are derived. Next, the nonparametric estimation of the “separability” and its mean, for which we assume the availability of only one dataset, using the leave-pair-out bootstrap-based estimator, is considered. In addition, the influence function approach to estimate the uncertainty of that estimate is used. The major contributions are the inclusion of the effect of the training set on the properties of the “separability”, and also on its nonparametric estimator, in both the mean and the variance; this is a key difference from the PAUC literature and its use in medical community. The mathematical properties are confirmed by a set of experiments using simulated and real datasets. Finally, the true performance (not its estimate) of classifiers measured in “separability” may vary .significantly with varying the training set, while its estimate yet has a small estimated variance. This accounts for having “good” estimate for “bad” performance

تقييم المصنف من حيث مساحة جزئية تحت منحنى ROC

يعتبر تقييم المصنفات باستخدام منطقة جزئية تحت منحنى ROC (PAUC) (أو ما يعادلها، "انفصالية"، وهى دالة فى عتبة الإختيار). الخصائص الإحصائية لل"انفصالية" بوصفها دالة فقط من المصنف المدرب وعتبة الإختيار. التخمين اللابارامترى لل"انفصالية" وكذلك متوسطها، (ونحن نفترض توافر مجموعة بيانات واحدة فقط) باستخدام المخمن leave-pair-out bootstrap. وبالإضافة إلى ذلك، يتم استخدام نهج وظيفة-النفوذ (influence function) لتقدير عدم التيقن من هذا التقدير. المساهمات الرئيسية هي إدراج تأثير التدريب على مجموعة خصائص ال"انفصالية"، وأيضا على المخمن اللابارامترى لها في كل من الوسط الحسابي والتباين، وهذا هو الفرق الرئيسي من الأدب PAUC واستخدامه في المجتمع الطبي. الخصائص الرياضية تأكدت بمجموعة من التجارب باستخدام مجموعات البيانات ومحاكاة حقيقية. وأخيرا، الأداء الحقيقي (وليس تقديراتها) من المصنفات التى تقاس بال"انفصالية" تختلف بشكل كبير مع اختلاف مجموعة التدريب، في حين أن تقديراتها لا تتباين بشكل كبير. وهذا يفسر بوجود تقدير "جيد" لأداء "سيئ".

3.	**Classifier variability: Accounting for training and testing** http://www.sciencedirect.com/science/article/pii/S0031320312000180	W. Chen; B. Gallas; **W.A. Yousef** **مشاركون من تخصص آخر:** The U.S. Food and Drug Administration / Digital Imaging and Applied Mathematics	Pattern Recognition 45(7): p. 2661-2671.	2012	International Journal IF: 2.292 5-year IF: 3.172 2-Artificial Intelligence (A)	ISSN: 0031-3203

We categorize the statistical assessment of classifiers into three levels: assessing the classification performance and its testing variability conditional on a fixed training set, assessing the performance and its variability that accounts for both training and testing, and assessing the performance averaging over training sets and its variability that accounts for both training and testing. We derived analytical expressions for the variance of the estimated AUC and provide freely available software implemented with an efficient computation algorithm. Our approach can be applied to assess any classifier that has ordinal (continuous or discrete) outputs. Applications to simulated and real data sets are presented to illustrate our methods.

تقلب المصنف: اعتبار التدريب والاختبار

نحن نقسم تقييم المصنفات الإحصائية إلى ثلاثة مستويات: تقييم أداء وتصنيف تقلباته مشروطا بمجموعة التدريب الثابتة، وتقييم الأداء وتقلباته التي تمثل كلا من التدريب والاختبار، وأخيرا تقييم الأداء في المتوسط على مجموعات التدريب والتي تمثل كلا من التدريب والاختبار. نحن قمنا بإشتقاق عبارات تحليلية لهذه الفروق من الـ AUC التقديرية وتوفير البرمجيات المتاحة مجانا مع تنفيذ خوارزمية حساب كفاءة. ويمكن تطبيق نهجنا لتقييم أي مصنف يحتوي على ناتج ترتيبي (متصلة أو منفصلة). نقوم كذلك بعرض التطبيقات لمجموعات البيانات ومحاكاة حقيقية لتوضيح طرقنا.

4.	**Uncertainty estimation with a finite dataset in the assessment of classification models** http://www.sciencedirect.com/science/article/pii/S016794731100209X	W. Chen; B. Gallas; **W.A. Yousef** **مشاركون من تخصص آخر:** The U.S. Food and Drug Administration / Digital Imaging and Applied Mathematics.	Computational Statistics & Data Analysis 56(5): p. 1016-1027.	2012	International Journal IF: 1.028 5-year IF: 1.373 5-Interdisciplinary Applications (A) 6-Statistics and Probability (A)	ISSN: 0167-9473

To successfully translate genomic classifiers to the clinical practice, it is essential to obtain reliable and reproducible measurement of the classifier performance. A point estimate of the classifier performance has to be accompanied with a measure of its uncertainty. In general, this uncertainty arises from both the finite size of the training set and the finite size of the testing set. The training variability is a measure of classifier stability and is particularly important when the training sample size is small. Methods have been developed for estimating such variability for the performance metric AUC (area under the ROC curve) under two paradigms: a smoothed cross-validation paradigm and an independent validation paradigm. The methodology is demonstrated on three clinical microarray datasets in the microarray quality control consortium phase two project (MAQC-II): breast cancer, multiple myeloma, and neuroblastoma. The results show that the classifier performance is associated with large variability and the estimated performance may change dramatically on different datasets. Moreover, the training variability is found to be of the same order as the testing variability for the datasets and models considered. In conclusion, the feasibility of quantifying both training and testing variability of classifier performance is demonstrated on finite real-world datasets. The large variability of the performance estimates shows that patient sample size is still the bottleneck of the microarray problem and the training variability is not negligible.

تقدير عدم اليقين مع مجموعة بيانات محدودة في تقييم نماذج التصنيف

لترجمة المصنفات الجينومية بنجاح إلى الممارسة السريرية فإنه من الضروري الحصول على قياسات موثوق بها وقابلة للتكرار لتصنيف الأداء. وجود نقطة تقدير لأداء المصنف يجب أن يترافق مع وجود قدر من عدم اليقين. بشكل عام، عدم اليقين هذا ينشأ ،على حد سواء، من محدودية حجم مجموعة التدريب و محدودية حجم مجموعة الاختبار. تباين التدريب هو دلالة على قدر من استقرار المصنف وله أهمية خاصة عندما يكون حجم عينة التدريب صغير. وقد تم تطوير أساليب لتقدير مثل هذه التقلبات للـ AUC كأداء متري للمساحة تحت منحنى الـ ROC وذلك بمنهجيتين: صحة النموذج ممهدة و نموذج صحة التحقق المستقلة. ونطبق المنهجية على ثلاثة مجموعات للبيانات السريرية مقدمة من اتحاد MAQC-II : سرطان الثدي والورم النخاعي المتعدد، والعصبية. بينت النتائج أنه يرتبط أداء المصنف مع تقلبية كبيرة وأداء المقدرة قد تتغير بشكل كبير على مجموعات البيانات المختلفة. وعلاوة على ذلك، فقد تبين أن تقلب التدريب له نفس الرتبة للتغير الناتج من مجموعات الإختبار. في الختام، يتجلى جدوى قياس كل من التدريب وتقلب اختبار أداء المصنف على مجموعات البيانات في العالم الحقيقي محدود. التباين الكبير في التقديرات تبين أن حجم عينة المرضى لا يمثل العائق الحقيقى للمشكلة وتباين التدريب أمر يستحق النظر.

5.	**The Microarray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models.** http://www.nature.com/nbt/journal/v28/n8/full/nbt.1665.html	Shi, L.; et al.; **W.A. Yousef** **مشاركون من تخصص آخر:** The U.S. Food and Drug Administration / Digital Imaging and Applied Mathematics.	Nature Biotechnology 28(8): 827-838.	2010	International Journal IF: 23.3 Unclassified	ISSN: 1087-0156

Gene expression data from microarrays are being applied to predict preclinical and clinical endpoints, but the reliability of these predictions has not been established. In the MAQC-II project, 36 independent teams analyzed six microarray data sets to generate predictive models for classifying a sample with respect to one of 13 endpoints indicative of lung or liver toxicity in rodents, or of breast cancer, multiple myeloma or neuroblastoma in humans. In total, >30,000 models were built using many combinations of analytical methods. The teams generated predictive models without knowing the biological meaning of some of the endpoints and, to mimic clinical reality, tested the models on data that had not been used for training. We found that model performance depended largely on the endpoint and team proficiency and that different approaches generated models of similar performance. The conclusions and recommendations from MAQC-II should be useful for regulatory agencies, study committees and independent investigators that evaluate methods for global gene expression analysis.

" مراقبة الجودة MAQC-II" ودراسة الممارسات الشائعة من أجل تطوير تحقيق النماذج التنبؤية للميكروأري.

يتم تطبيق البيانات من التعبير الجيني للتنبؤ فى التطبيقات قبل السريرية والسريرية، ولكن لم يثبت موثوقية هذه التوقعات بعد. فى مشروع MAQC-II تمم تحليل البيانات من قبل 36 فريقا مستقلا حيث قاموا بتحليل 6 مجموعات لتوليد نماذج تنبؤية لتصنيف عينة فيما لواحدة من 13 تطبيق يدل على الرئة أو تسمم الكبد في القوارض، أو من سرطان الثدي، المايلوما المتعددة أو العصبية في البشر. تم بناء أكثر من 30000 نموذج باستخدام تركيبات العديد من طرق التحليل. تم اختبار النماذج على البيانات التي لم تستخدم للتدريب و قامت فرق البحث بتوليد هذه النماذج التنبؤية دون معرفة المعنى البيولوجي لبعض النقاط النهائية وذلك من أجل تقليد الواقع السريري. وجدنا أن أداء النموذج يعتمد إلى حد كبير على كفاءة ونقطة النهاية ومهارة الفريق وبالرغم من ذلك إلا أنها ولدت نماذج لها أداء مماثل. الاستنتاجات والتوصيات من MAQC-II مفيدة للوكالات التنظيمية ولجان الدراسة والمحققين المستقلين من أجل تقييم طرق تحليل التعبير الجيني العالمي.

6.	**On Detecting Abnormalities in Digital Mammography.** http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5759684&contentType=Conference+Publications&queryText%3DOn+Detecting+Abnormalities+in+Digital+Mammography	**W. A. Yousef**; W.A. Mustafa; A.A. Ali; N.A. Abdelrazek; A.M. Farrag. **مشاركون من تخصص آخر:** الأشعة التشخيصية -كلية طب القصر العينى- جامعة القاهرة	Applied Imagery Pattern Recognition Workshop, 2010. Proceedings. 39rd; IEEE Computer Society; Washington DC, USA.	2010	International Conference Unclassified	ISSN: 1550-5219

Breast cancer is the most common cancer in many countries all over the world. Early detection of cancer, in either diagnosis or screening programs, decreases the mortality rates. Computer Aided Detection (CAD) is software that aids radiologists in detecting abnormalities in medical images. In this article we present our approach in detecting abnormalities in mammograms using digital mammography. Each mammogram in our dataset is manually processed—using software specially developed for that purpose—by a radiologist to mark and label different types of abnormalities. Once marked, processing henceforth is applied using computer algorithms. The majority of existing detection techniques relies on image processing (IP) to extract Regions of Interests (ROI) then extract features from those ROIs to be the input of a statistical learning machine (classifier). Detection, in this approach, is basically done at the IP phase; while the ultimate role of classifiers is to reduce the number of False Positives (FP) detected in the IP phase. In contrast, processing algorithms and classifiers, in pixel-based approach, work directly at the pixel level. We demonstrate the performance of some methods belonging to this approach and suggest an assessment metric in terms of the Mann Whitney statistic.

حول الكشف عن الأمور الغير طبيعية في تصوير الثدي بالأشعة الرقمية

سرطان الثدي هو السرطان الأكثر شيوعا في بلدان كثيرة في جميع أنحاء العالم. الكشف المبكر عن السرطان، في برامج التشخيص أو الفحص سواء، يقلل من معدلات الوفيات. الكشف بمساعدة الحاسوب (CAD) هو برنامج يساعد في الكشف عن الأماكن الغير طبيعية في صور الأشعة الطبية. في هذه المقالة نقدم أسلوبنا في الكشف عن تشوهات في تصوير الثدي بالأشعة السينية باستخدام التصوير الشعاعي الرقمي للثدي. كل صور الأشعة في مجموعة البيانات لدينا معالجة يدويا باستخدام البرمجيات المتقدمة، المصممة خصيصا لهذا الغرض، من قبل طبيب الأشعة لتسمية أنواع مختلفة من التشوهات. خوارزميات الكمبيوتر الحالية تقوم بوضع علامة مرة واحدةعلى أماكن التشوهات. غالبية تقنيات الكشف الحالية تعتمد على معالجة الصور لاستخراج مناطق الاهتمام ثم استخراج ملامح من تلك المناطق لتكون مدخلات آلة التعلم الإحصائية (المصنف). الإكتشاف، في هذا النهج، يتم بالأساس في مرحلة المعالجة في حين أن الدور النهائي من المصنفات هو الحد من عدد من ايجابيات كاذبة. فى المقابل، خوارزميات المعالجة والمصنفات، في منهجنا، العمل مباشرة يكون على مستوى البكسل. فى هذا البحث نبرهن على أداء بعض الطرق المنتميه لهذا النهج، ونقترح إجراء تقييم متري من حيث إحصائية مان ويتنى.

7.	**Nonparametric estimation of the threshold at an operating point on the ROC curve.** http://www.sciencedirect.com/science/article/pii/S0167947309002369	**W.A. Yousef**; S. Kundu; R.F. Wagner **مشاركون من تخصص آخر:** The George Washington University, Statistics Dep. The U.S. Food and Drug Administration / Digital Imaging and Applied Mathematics.	Computational Statistics & Data Analysis. 53(12): p. 4370-4383.	2009	International Journal IF: 1.028 5-year IF: 1.373 5-Interdisciplinary Applications (A) 6-Statistics and Probability (A)	ISSN: 0167-9473

In the problem of binary classification (or medical diagnosis), the classification rule (or diagnostic test) produces a continuous decision variable which is compared to a critical value (or threshold). Test values above (or below) that threshold are called positive (or negative) for disease. The two types of errors associated with every threshold value are Type I (false positive) and Type II (false negative) errors. The Receiver Operating Curve (ROC) describes the relationship between probabilities of these two types of errors. The inverse problem is considered; i.e., given the ROC curve (or its estimate) of a particular classification rule, one is interested in finding the value of the threshold ξ that leads to a specific operating point on that curve. A nonparametric method for estimating the threshold is proposed. Asymptotic distribution is derived for the proposed estimator. Results from simulated data and real-world data are presented for finite sample size. Finding a particular threshold value is crucial in medical diagnoses, among other fields, where a medical test is used to classify a patient as ‘‘diseased’’ or ‘‘nondiseased’’ based on comparing the test result to a particular threshold value. When the ROC is estimated, an operating point is obtained by fixing probability of one type of error, and obtaining the other one from the estimated curve. Threshold estimation can then be viewed as quantile estimation for one distribution but with the utilization of the second one.

تقدير غير باراميترى لتقدير عتبة عند نقطة على منحنى التشغيل

في مشكلة التصنيف الثنائي (أو التشخيص الطبي)، قاعدة التصنيف (أو الاختبار التشخيصي) تنتج متغير القرار المستمر الذي يقارن بالقيمة الحرجة (أو عتبة) بعد ذلك. اختبار القيم الأعلى (أو الأقل) من هذه العتبة تسمى القيم الإيجابية (أو السلبية) للمرض. النوعين من الأخطاء المرتبطة مع كل قيمة للعتبة هي: نوع الخطأ الأول (إيجابية كاذبة) ونوع الخطأ الثاني (سلبية كاذبة). منحنى التشغيل (ROC) يصف العلاقة بين احتمالات هذين النوعين من الأخطاء. المشكلة العكسية هى التى اعتبرناها فى هذا البحث ألا وهى بالنظر إلى منحنى التشغيل (أو تقدير منحهى التشغيل) لقاعدة تصنيف ما، كيف يتم العثور على قيمة عتبة ما ξ والذي يؤدي إلى نقطة محددة على منحنى التشغيل. ويقترح طريقة اللامعلمية لتقدير العتبة. نقترح فى هذا البحث طريقة لاباراميترية للتقدير. يتم عرض النتائج من البيانات المحاكاة وبيانات العالم الحقيقي لحجم عينة محدودة. إن العثور على قيمة العتبة خاصة أمر بالغ الأهمية في التشخيص الطبي، بين غيرها من المجالات، حيث يتم استخدام اختبار طبي لتصنيف المريض على أنه مصاب أو غير مصاب على أساس المقارنة بين نتيجة الاختبار إلى قيمة عتبة معينة. عندما يقدر منحنى التشغيل يتم الحصول على نقطة تشغيل عن طريق تحديد احتمال نوع واحد من الخطأ، والحصول على واحدة أخرى من المنحنى المقدر. ويمكن بعد تقدير عتبة أن ينظر إليه على أنه تقديرللكوانتيل لتوزيع ما ولكن مع الاستفادة من التوزيع الآخر.

8.	**Estimating Cross-Validation Variability** http://www.amstat.org/meetings/jsm/2009/onlineprogram/index.cfm?fuseaction=abstract_details&abstractid=304403	**W.A. Yousef**; W. Chen. **مشاركون من تخصص آخر:** The U.S. Food and Drug Administration / Digital Imaging and Applied Mathematics.	Proceedings of the 2009 Joint Statistical Meeting (JSM), Section on Statistics in Epidemiology; pp: 3318-3326; Washington DC, USA.	2009	International Conference Unclassified	ISBN: 978-0-9791747-7-3

We propose a novel influence-function-based approach to estimating the variance of classification performance that is estimated using the Monte Carlo K-fold Cross Validation (MKCV) procedure. In MKCV, the data is partitioned into K exclusive folds and the classifier is trained with K −1 folds of the data and then tested on the remaining one fold to estimate the performance; this process is repeated M times by shuffling the data each time and the M performance values are averaged to obtain the final performance estimate. A naive variance estimator is to use the sample variance of the M performance values, which does not account for the correlations and hence is biased. Our preliminary simulation results show that our new approach is promising.

تقدير التباين فى التحقق الداخلى

نقترح طريقة جديدة مستمدة من دوال التأثيروذلك لتقدير التباين فى أداء المصنفات بإستحدام التحقق الداخلى من الصحة بموذتي كارلو (MKCV). في MKCV، يتم تقسيم البيانات إلى طيات حصرية عددها K ويتم تدريب المصنف على K-1 من الطيات ثم يتم الإختبار على الطية المتبقية. لتقدير الأداء. وتتكرر هذه العملية مرات M مع خلط البيانات كل مرة. ثم يؤخذ المتوسط للحصول على تقدير الأداء النهائي. مقدر التباين الساذج هو استخدام تباين العينة على عدد القيم M. ولكن هذه القيم، والتي لا تمثل الارتباطات الشرطية منحازة. نتائجنا الأولية من المحاكاة تبين أن وجهة نظرنا الجديدة واعدة.

9.	**Comparison of classifier performance estimators: a simulation study** http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=816020	W. Chen; R.F. Wagner; **W.A. Yousef**; B. Gallas. **مشاركون من تخصص آخر:** The U.S. Food and Drug Administration / Digital Imaging and Applied Mathematics.	Proceedings SPIE 7263 of IEEE Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment; Lake Buena Vista, FL, USA.	2009	International Conference Unclassified	ISSN: 0277-786X

We aim to compare resampling-based estimators of the area under the ROC curve (AUC) of a classifier with a Monte Carlo simulation study. The comparison is in terms of bias, variance, and mean square error. We also examine the corresponding variance estimators of these AUC estimators. We compared three AUC estimators: the hold-out (HO) estimator, the leave-one-out cross validation (LOOCV) estimator, and the leave-pair-out bootstrap (LPOB) estimator. Each performance estimator has its own variability estimator. In our simulations, in terms of the mean square error, HO is always the worst and the rating of the other two estimators depends on the interplay of sample size, dimensionality, and the population separability. In terms of estimator variability, the LPOB is the least variable estimator and the HO is the most variable estimator. The results also show that the estimation of the variance of LPOB using the influence function approach with a finite data set is unbiased or conservatively biased whereas the estimation of the variance of the LOOCV or the HO is downwardly, i.e., anticonservatively biased.

مقارنة أداء المصنفات: دراسة محاكاة

نحن نهدف إلى المقارنة المقدرات المبينة على إعادة التشكيل والتى تقرب المساحة تحت منحنى التشغيل لمصنف ما وذلك بدراسة محاكاة مونتي كارلو. والمقارنة من حيث التباين، والتحيز، ومتوسط مربع الخطأ. ندرس أيضا مقدرات التباين لهذه المقدرات. قارنا 3 مقدرات للمساحة تحت المنحنى: مقدر القبضة (HO) ، ومقدر ترك-الواحد (LOOCV) ، و مقدر البوتستراب (LPOB). كل مقدر له مقدر لقياس تباينه. في المحاكاة التى أجريناها، ومن حيث الخطأ في مربع الوسط، المقدر HO هو دائما أسوأ وتصنيف المقدرين الأخريين يعتمد على حجم العينة، الأبعاد، وانفصال التوزيعين. من حيث تقلب المقدر، فإن المقدر LPOB هو الأقل تغيرا والمقدر HO هوالأكبر. وتظهر النتائج أيضا أن تقدير التباين للمقدر LPOB باستخدام دوال التأثير مع مجموعة بيانات محدودة يعطى نتائج غير منحازة (أو متحفظه فى إنحيازها) في حين أن تقدير التباين لكل من LOOCV أو HO هو منحاز إلى أسفل.

10.	**Finite Training of Radiologists and Statistical Learning Machines: Parallel Lessons** http://www.amazon.com/Advances-Medical-Physics-Anthony-Wolbarst/dp/1930524382	R.F. Wagner; **W.A. Yousef**; W. Chen **مشاركون من تخصص آخر:** The U.S. Food and Drug Administration / Digital Imaging and Applied Mathematics.	Appears in: Advances in Medical Physics: 2008, A.B. Wolbarst, K.L. Mossman, and W.R. Hendee, Editors, Medical Physics Pub.: Madison, WI, USA.	2008	International Book Unclassified	ISBN: 9781930524385

Assessment and comparison of medical imaging systems in the last decade have been characterized by the emergence of methods that account for the randomness or variability of the radiologists as well as that of the patients whose images they read. These methods have been naturally extended to the assessment of radiologists without and with computer-assist adjuncts, so-called computer-aided diagnosis. We are rapidly moving into an era in which image-based and nonimage-based biomarkers in high-dimensional space will be fused into diagnostic decision-making tools that may be used independently of human readers. The most obvious examples are multiple-gene-expression microarrays (“DNA chips”) whose large number of outputs will be applied to the task of detecting or diagnosing various disease conditions, as well as predicting who will be the responders and who the nonresponders to a specified drug or therapy (“personalized medicine”). Statistical learning machines must be trained from example cases how to combine these many outputs and then be tested for their performance of these tasks on previously unseen cases. The randomness or variability from training and testing of learning machines are in a one-to-one analogy with the randomness or variability from radiologists and patients in the conventional imaging task. This chapter reviews our understanding of this analogy and recent developments in assessment of both the conventional and the algorithmic modes of diagnosis.

التدريب المحدود لأطباء الأشعة وآلات التعلم الإحصائية: دروس متوازية

اتسمت تقييم ومقارنة نظم التصوير الطبي في العقد الماضي بظهور الأساليب التي تأخذ فى إعتبارها العشوائية أو التقلب من أطباء الأشعة الذين يقرءون الصور وكذلك لمرضى. وقد تم تمديد هذه الأساليب بشكل طبيعي في تقييم الأشعة دون ومع مساعدة الحاسوب، ما يسمى التشخيص بمساعدة الحاسوب. نحن نتحرك بسرعة إلى هذا العصر الذي ستتصهر فيه المؤشرات الحيوية التي تعتمد على الصور والتى لا تعتمد على الصور وذلك لتكوين أدوات إتخاذ قرار فى أبعاد عالية في والتى ممكن أن تستخدم فى التشخيص بشكل مستقل عن قراءة الإنسان. الأمثلة الأكثر وضوحا هي ميكروأريس متعددة جينات التعبير ("رقائق الحمض النووي") والتى عدد كبير من مخرجاتها سيتم تطبيقها على مهمة الكشف عن أو تشخيص الحالات المرضية. وكذلك توقع أى من المرضى سيستجيب وأى سوف لا يستجيب لعلاج ما ("الطب الشخصي"). يجب تدريب آلات التعلم الإحصائي من الحالات مع كيفية الجمع بين العديد من هذه المخرجات ومن ثم يتم اختبار أدائها لهذه المهام في قضايا لم يسبق لها مثيل. عشوائية أو تقلب الأداء الناتج من التدريب واختبار آلات التعلم هي مشابهة تماما مع تقلب العشوائية بسبب المرضى أو أطباء الأشعة. يستعرض هذا الفصل فهمنا لهذا القياس والتطورات الأخيرة في تقييم التشخيص التقليدي والخوارزمى.

11.	**Statistical Learning Machines from ATR to DNA Microarrays: Design, Assessment, and Advice for Practitioners.**	**W.A. Yousef**	Proceedings of 6th International Conference on Electrical Engineering (ICEENG), Military Technical College (MTC), Cairo, Egypt.	2008	International Conference Unclassified	

Statistical Learning is the process of estimating an unknown probabilistic input-output relationship of a system using a limited number of observations; and a statistical learning machine (SLM) is the machine that learned such a process. While their roots grow deeply in Probability Theory, SLMs are ubiquitous in the modern world. Automatic Target Recognition (ATR) in military applications, Computer Aided Diagnosis (CAD) in medical imaging, DNA microarrays in Genomics, Optical Character Recognition (OCR), Speech Recognition (SR), spam email filtering, stock market prediction, etc., are few examples and applications for SLM; diverse fields but one theory.

The field of Statistical Learning can be decomposed to two basic subfields, Design and Assessment. We mean by Design, choosing the appropriate method that learns from the data to construct an SLM that achieves a good performance. We mean by Assessment, attributing some performance measures to the designed SLM to assess this SLM objectively. To achieve these two objectives the field encompasses different other fields: Probability, Statistics and Matrix Theory; Optimization, Algorithms, and programming, among others.

Three main groups of specializations—namely statisticians, engineers, and computer scientists (ordered ascendingly by programming capabilities and descendingly by mathematical rigor)—exist on the venue of this field and each takes its elephant bite. Exaggerated rigorous analysis of statisticians sometimes deprives them from considering new ML techniques and methods that, yet, have no "complete" mathematical theory. On the other hand, immoderate add-hoc simulations of computer scientists sometimes derive them towards unjustified and immature results. A prudent approach is needed that has the enough flexibility to utilize simulations and trials and errors without sacrificing any rigor. If this prudent attitude is necessary for this field it is necessary, as well, in other fields of Engineering.

In the spirit of this prelude, this article is intended to be a pilot-view of the field that sheds the light on SLM applications, the Design and Assessment stages, necessary mathematical and analytical tools, and some state-of-the-art references and research.

آلات التعلم الإحصائية : من ATR إلى DNA microarray: تصميم وتقييم، وتقديم المشورة للممارسين

التعلم الإحصائي هو عملية تقدير لاحتمالية علاقة المدخلات والمخرجات الغير معروف باستخدام عدد محدود من الملاحظات. و آلة التعلم الإحصائي هي الدالة التي تعلمت هذه العملية. بينما جذورها تنمو عميقا في نظرية الاحتمالات ، SLMs موجودة في كل مكان في العالم الحديث . التعرف التلقائي على الهدف (ATR) في التطبيقات العسكرية ، التشخيص بمساعدة الكمبيوتر (CAD) في التصوير الطبي ، ميكروأى الحمض النووي في مشروع الجينوم، التعرف الضوئي على الحروف (OCR) ، التعرف على الكلام (SR) ، و فلترة البريد الإلكتروني، و التنبؤ بسوق الأوراق المالية ، وما إلى ذلك، هي بعض الأمثلة والتطبيقات؛ مجالات متنوعة ولكن نظرية واحدة .

مجال التعلم الإحصائية يمكن أن تنقسم لاثنين من المجالات الفرعية: تصميم و تقييم . ما نعنيه بالتصميم ، هواختيار الأسلوب المناسب لبناء SLM يحقق أداء جيدا . وما نعنيه بالتقييم ، هو بعض مقاييس الأداء. لتحقيق هذين الهدفين يشمل المجال غيرها من المجالات المختلفة : احتمال والإحصاء و نظرية المصفوفات ، الخوارزميات، والبرمجة وغيرها.

ثلاث مجموعات رئيسية من التخصصات ، وهي الإحصائيين والمهندسين و علماء الكمبيوتر مهتمة بهذا المجال. التحليل الدقيق المبالغ فيه من الإحصائيين يحرمهم أحيانا من النظر في تقنيات جديدة. من ناحية أخرى ، الإستخدام المفرط للمحاكاة من قبل علماء الكمبيوتر في بعض الأحيان يفرز نتائج غير مبررة وغير ناضجة. وهناك حاجة إلى النهج الوسط.

12.	**Microcalcification detection with and without prototype CAD system (LIBCAD): a comparative study**	N.A. Abdelrazek; W.A. Mustafa; **W. A. Yousef** **مشاركون من تخصص آخر:** الأشعة التشخيصية -كلية طب القصر العينى- جامعة القاهرة	The Egyptian Journal of Radiology and Nuclear Medicine 44(2): p. 397-404.	2013	International Journal Unclassified	ISSN: 0378-603X

Breast cancer is the most common cancer all over the world. Early detection of cancer, in either diagnosis or screening programs, decreases the mortality rates. The survival rate is greatly influenced by how early the cancer is treated, thus it is important to discover the disease at an early stage. Clusters of microcalcifications are early sign of possibly cancer and are in general not palpable. Computer Aided Detection (CAD) is software that aids radiologists in detecting breast abnormalities, including microcalcifications. *LIBCAD* is a recent CAD that is designed to be importable to any image viewer or any other software. We measure the performance of the CAD microcalcifications detection, and compare it to the readings of an experienced radiologist.

كشف التكلسات الصغيرة مع وبدون نظام الكشف الحاسوبى LIBCAD : دراسة مقارنة

سرطان الثدي هو أكثر أنواع السرطان شيوعا في جميع أنحاء العالم. الكشف المبكر عن السرطان، فى التشخيص وإما في برامج الفحص، يقلل من معدلات الوفيات. ويتأثر معدل البقاء على قيد الحياة إلى حد كبير من قبل كيف يتم التعامل في وقت مبكر لمرض السرطان، وبالتالي فإنه من المهم اكتشاف المرض في مرحلة مبكرة. مجموعات من microcalcifications هي علامة مبكرة من السرطان وربما هي بشكل عام الأقل وضوحا. الكشف بمساعدة الكمبيوتر (CAD) هو البرنامج الذي يساعد أطباء الأشعة في الكشف عن سرطان الثدي. نقيس أداء الكشف لبرنامج ال LIBCAD وذلك لمقارنة القراءات من أطباء الأشعة ذوي الخبرة.