*MEDINFO 2015: eHealth-enabled Health*
*I.N. Sarkar et al. (Eds.)*


*doi:10.3233/978-1-61499-564-7-867*

# Identifying Clinical Study Types from PubMed Metadata: The Active (Machine) Learning Approach

**Adam G. Dunn [a], Diana Arachi [a], Florence T. Bourgeois [b,c]**

[a] *Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, NSW, Australia*
[b] *Children's Hospital Informatics Program, Boston Children's Hospital, Boston, MA, USA*
[c] *Department of Pediatrics, Harvard Medical School, Boston, MA, USA*

**Abstract**

*We examined a process for automating the classification of articles in MEDLINE aimed at minimising manual effort without sacrificing accuracy. From 22,808 articles pertaining to 19 antidepressants, 1000 were randomly selected and manually labelled according to article type (including, randomised controlled trials, editorials, etc.). We applied a machine learning approach termed 'active learning', where the learner (machine) selects the order in which the user (human) labels examples. Via simulation, we determined the number of articles a user needed to label to produce a classifier with at least 95% recall and 90% precision in three scenarios related to evidence synthesis. We found that the active learning process reduced the number of training instances required by 70%, 19%, and 14% in the three scenarios. The results show that the active learning method may be used in some scenarios to produce accurate classifiers that meet the needs of evidence synthesis tasks and reduce manual effort.*

***Keywords:***

Machine Learning; Databases, Bibliographic; Antidepressants.

## Introduction

When undertaking reviews or synthesis of clinical evidence, researchers are faced with the task of identifying all relevant published studies in bibliographic databases. This has become increasingly difficult and time-consuming with the dramatic increase in the number of studies being published each year.

To improve the efficiency of evidence synthesis tasks that require the comprehensive identification of specific types of articles, researchers have analytically devised search queries or filters, and applied them to large bibliographic databases such as MEDLINE [1-8]. Others have proposed methods for context-specific searches by partially or completely replacing the need for manual screening and classification using machine learning methods [9-11].

In an effort to maximise the proportion of relevant articles identified (recall) in these searches – which ideally should be 100% for systematic reviews – researchers are often forced to sacrifice the proportion of irrelevant articles also included as relevant (precision), leaving them with many articles that require further manual query and classification.

Here we consider the application of machine learning to the task of article classification, and propose the use of *active learning* [12], an approach aimed at reducing the number of examples that a user (human) needs to provide to the learner (machine). Active learning has been used to classify clinical studies for systematic reviews previously [11], and for well over a decade in biomedical research applications ranging from drug discovery to biomedical informatics [13, 14].

Our aim was to examine whether active learning could be used to reduce the manual workload for classifying articles in tasks without substantially reducing the quality of the process. To do this, we conducted experiments for three scenarios of evidence synthesis, using MEDLINE metadata on a large corpus of articles about antidepressants.

## Methods

### Data

We used PubMed to search MEDLINE on 9 December 2014 for articles that included in the title, abstract, or MeSH terms, the names or synonyms of one or more of the 19 antidepressants that were approved for use by the US Food and Drug Association (FDA) between 1995 and 2005. The search returned 40,098 unique articles. From this set, we selected the 22,808 articles for which the antidepressant of interest was already approved by the FDA when the article was published, and then randomly selected 1000 for inclusion in the study. We selected antidepressants for a study of author influence in postapproval opinion/commentary.

Two investigators (AD and DA) classified each article as one of eight article types: (1) systematic review; (2) non-systematic review and meta-analysis; (3) guideline; (4) randomised controlled trial; (5) other clinical study (e.g. non-randomised trial, cost-benefit analysis); (6) case study; (7) non-clinical study (animal models, chemistry for synthesis, testing, or where antidepressants are not the primary topic); and (8) comment or opinion piece. Agreement between the two investigators was 90% (Cohen's kappa = 0.87) and disagreements were resolved by discussion.

Specific characteristics (metadata) were abstracted from the articles to form the set of potential features, which are used as the basis for training the machine learning classifiers. These metadata included article titles and abstracts, MeSH terms, and publication types assigned in MEDLINE. For titles and abstracts, each word with more than three letters was extracted and binary features created to represent their presence or absence in each article. MeSH terms and publication types were used to create binary features without change. No other pre-processing methods (such as stemming or the use of stop words) were applied to the metadata prior to their use in the classifier training. The total number of features that could be used to train the classifiers was 5,606.

Three scenarios were created as examples of evidence review, synthesis, and surveillance tasks that might benefit from automation. In Scenario 1, the aim was to find all clinical trials for use in a systematic review (4 and 5, above). In

Scenario 2, the aim was to identify all randomised controlled trials for inclusion in a meta-analysis (4, above). In Scenario 3, the aim was to find only opinion pieces, commentaries, and non-systematic reviews (2 and 8, above) for examining the distribution of contributions by individual authors.

**The active learning process**

We defined the ordered set of articles that the user is asked to label by the learner as the *labelling queue*. The unique feature of the active learning process is that the learner iteratively trains new classifiers each time it receives an additional label from the user, and uses the resulting information to determine the order of the labelling queue.

The aim of the learner is to minimise the length of the labelling queue by converging more quickly on a classifier that is able to exceed the performance requirements.

The proposed process for applying active learning to this task includes two phases (an alternative to a previous approach [11]): the first is for when there are too few examples in each class to examine the distribution of features across the two classes and identify meaningful features (Phase I); and the second when a new classifier is trained in each step and used to select the next article to place in the labelling queue (Phase II).

***Phase I***

In Phase I, the issue is that there is not enough information available for the learner to know enough about the distribution of features across relevant/irrelevant articles, and thus not enough identifiable features capable of discriminating between the two classes to effectively construct a classifier.

A simple solution in this stage is to progress at random until enough useful information is available for the active learning approach. To do this, we take advantage of the statistical distribution of features among known articles. The user is asked to label articles selected at random until there are a minimum of five features that exhibit a significant difference ($p<0.05$ in a Fisher's exact test) between the classes. These features then become available for the classifiers constructed in the second phase.

Alternative methods for finding useful articles at this stage involve clustering the articles based on the distribution of features, without knowing to which class any of the articles belong [15]. We did not consider these alternatives in this study but these could reduce the workload in Phase I.

***Phase II***

Phase II is the active learning phase. In each round, a new classifier is trained using the labels from the labelling queue, and the information produced by applying the new classifier to the pool of remaining unlabelled articles is used to determine the next article to be labelled by the user.

For selecting useful features, we applied a simple statistical method in which features from all labelled articles (so far) are compared across the two classes (either relevant or irrelevant articles). In the statistical testing, Fisher's exact tests were applied to each feature and those with the lowest p-values were then included as features to train the classifiers. In this study, we chose to select the features that remained after a Bonferroni correction was applied, or the five features with the lowest p-values, whichever was larger.

We used support vector machines as the machine learning algorithm in these experiments, as it is appropriate for tasks such as document classification [16]. The support vector machines were constructed using linear kernels and a least squares method. To apply the active learning approach to support vector machines, the learner selects the unlabelled article that is closest to the hyperplane found to best separate the relevant and irrelevant articles during training, and passes that article to the user for labelling [12, 17].

Note that we only considered one approach for feature selection, one type of machine learning algorithm, and applied the rules consistently across all simulations. Alternative choices for feature selection and for classifiers could have affected the performance of the active learning and passive learning approaches in different ways.

**Analysis**

We evaluated the active learning process by estimating the risk of not achieving a pre-specified performance for each number of labelled articles. The pre-specified performance criteria were 95% recall (the percentage of relevant articles that were identified), and 90% precision (the percentage of identified articles that were relevant). Using these measures, we compared the results of the active learning simulations to a passive learning baseline, in which the articles labelled by the user are selected at random from the pool of all articles. This is equivalent to a process that continues with Phase I and never progresses to Phase II.

To estimate the risk of imperfect recall at a given labelling queue length, we ran repeated simulations to determine how many articles would need to be labelled by the user before a classifier with 95% recall and 90% precision was trained. In each case, 500 articles were selected (by stratified random sampling) for training and the remaining 500 were used as a holdout test set (which differs from a previous approach [11]). The performance results were determined by applying the classifer to the holdout set. For each of the three scenarios, we ran 1000 simulations for the passive learning baseline, re-sampling 500 articles each time.

The features that were most often used to train the classifiers that first reached 95% recall and 90% precision were reported in the active learning process as 'positive' features if they were more often present in the relevant articles, and 'negative' features otherwise.

We also examined how the process might work in practice, by conducting simulations in which the user labels 100 articles by active learning or the passive learning baseline. In these experiments, we simulated the approach 1000 times using 500 articles, but we determined the performance by testing it using the 500 articles (labelled and unlabelled) to find out if labelling 20% of articles would ensure a reliable classifier.

## Results

Non-clinical studies made up the greatest proportion in the sample of 1000 articles, followed by other clinical studies, randomised controlled trials, case studies and non-systematic reviews (Table 1).

*Table 1– Distribution of study types*

| Article Type | Frequency | Proportion |
|---|---|---|
| Non-clinical study | 410 | 41% |
| Other clinical study | 201 | 20% |
| Randomised controlled trial | 130 | 13% |
| Case study | 105 | 10% |
| Non-systematic review | 101 | 10% |
| Systematic review | 30 | 3.0% |
| Opinion piece/commentary | 22 | 2.2% |
| Guideline | 1 | 0.1% |
| Total | 1000 | 100% |

**Scenario 1: clinical studies for use in systematic reviews**

In Scenario 1, the task was to identify the 331 randomised controlled trials and other clinical studies (4 and 5, above). Applying the active learning process, the median number of articles labelled by the user when the classifier first exceeded 95% recall and 90% precision was 108 (IQR 78-146). To reach the same level of performance, the passive learning process required a median of 366 articles (IQR 284-460). This corresponds to a workload reduction of 70% in the number of labelled articles required to train a classifier (Figure 1). The results also show that for a small proportion of active learning simulations, the classifier never reached 95% recall and 90% precision (even after the user had labelled 200 articles), while for the passive learning simulations, the risk of training a poor classifier for unseen articles remained even after the user had labelled the entire training set of 500 articles.

The positive features most commonly included in the final classifiers generated by the active learning process included the MeSH terms "female" (97% of final classifiers) and "adult" (96%), the publication types "Randomized Controlled Trial" (99%) and "Clinical Trial" (97%), and abstract words "study" (70%) and "patients" (47%). Negative features included MeSH terms "Animals" (100%) or "rats" (49%), and publication type "Review" (94%) or "Case Reports" (86%).

To examine how the active learning process might work in practice, we measured the performance of the classifiers after the user had labelled 100 articles. The performance is considered *in sample*, measuring the recall, precision, and $F_1$-score of the 500 sampled articles including those that were already labelled (as existing methods have done [11]). The active learning process produced a median recall of 93%, median precision 96%, and median $F_1$-score 0.94. In the passive learning process, the median recall was 90%, median precision 90%, and median $F_1$-score 0.90 (Table 2). The differences in recall, precision, and $F_1$-score were significant ($p<0.001$) under a two sample Kolmogorov-Smirnov test.

**Scenario 2: randomised controlled trials for meta-analyses**

In Scenario 2, the median number of articles labelled in the passive learning process was 28 (IQR 22-38) compared to 24 (IQR 20-30) in the active learning process, representing a 14% reduction in workload (Figure 2). The results likely reflect the standardised ways in which randomised controlled trials are published, rather than the unbalanced nature of the dataset (130 relevant articles). Among the positive features that were most often used by the classifiers in the active learning process was the publication types "Randomized Controlled Trial" (99%), abstract terms "blind" (64%), "placebo" (62%), "double" (60%), and "week" (21%), and the MeSH terms "DoubleBlind_Method" (61%) and "Adult" (38%).

The marginal difference between the two approaches is reflected in the results of the practical test after the user has labelled 100 articles (Table 2), where both active and passive methods produce similar results. The differences in recall, precision, and $F_1$-score were significant ($p<0.001$).

**Scenario 3: opinion pieces and non-systematic reviews for a task in pharmacosurveillance**

In Scenario 3, the median number of articles labelled in the passive learning process was 42 (IQR 30-58) and in the active learning process 52 (IQR 37-79). This represents a 19% workload reduction in the number of training instances required (Figure 3). The results in Scenario 3 also show that a small proportion of the passive learning classifiers never reached the pre-specified stopping criteria, while the risk was minimal in the active learning process after 150 articles had been labelled. The positive features that were most often used by the classifiers in this scenario were the abstract words "therapy" (25%), "trials" (21%), "agents" (23%), "effective" (17%), and the publication type "Review" (95%). The negative features included abstract words "were" (48%) and "study" (27%), and the MeSH term "Male" (34%).

Reflecting the small gap between the passive and active learning processes, the differences in performance after the user labelled 100 articles were also relatively small (Table 2). The differences in recall, precision, and $F_1$-score were significant ($p<0.001$).

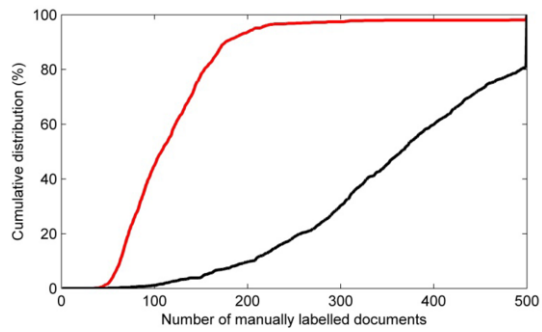


*Figure 1– Clincial studies (S1): the cumulative percentage of simulations in which the classifiers met the pre-specified performance within that number of labelled documents; passive learning (black); active learning (red).*

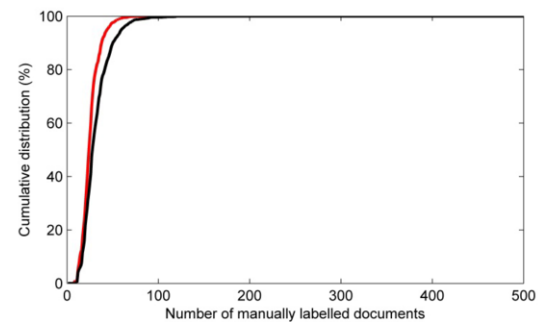


*Figure 2– Randomised Controlled Trials (S2): the cumulative percentage of simulations in which the classifiers met the pre-specified performance within that number of labelled documents; passive learning (black); active learning (red).*

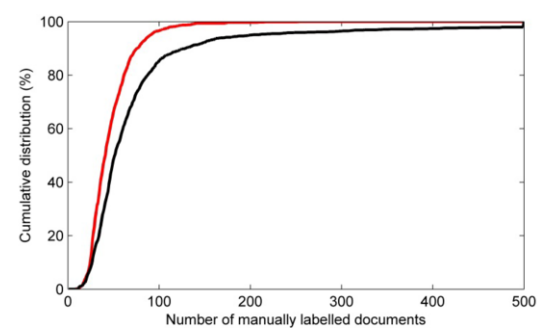


*Figure 3– Opinions & Commentaries (S3): the cumulative percentage of simulations in which the classifiers met the pre-specified performance within that number of labelled documents; passive learning (black); active learning (red).*

*Table 2– Comparing the performance (in the training set) after 100 user labellings to examine the process in practice.*

| Process and scenario | Median recall (interquartile range) | Median precision (interquartile range) | Median $F_1$-score (interquartile range) |
|---|---|---|---|
| **Passive learning process** | | | |
| S1: All clinical trials | 0.90 (0.88-0.93) | 0.90 (0.88-0.91) | 0.90 (0.89-0.91) |
| S2: Randomised controlled trials | 0.98 (0.97-0.98) | 0.99 (0.98-0.99) | 0.98 (0.97-0.99) |
| S3: Non-systematic reviews and opinion pieces | 0.95 (0.93-0.96) | 0.94 (0.92-0.96) | 0.94 (0.93-0.95) |
| **Active learning process** | | | |
| S1: All clinical trials | 0.93 (0.89-0.95) | 0.96 (0.94-0.97) | 0.94 (0.92-0.95) |
| S2: Randomised controlled trials | 0.98 (0.98-0.99) | 0.99 (0.99-0.99) | 0.99 (0.99-0.99) |
| S3: Non-systematic reviews and opinion pieces | 0.97 (0.96-0.97) | 0.97 (0.96-0.98) | 0.97 (0.96-0.97) |

## Discussion

We applied an active learning method to the article classification task, demonstrating that the approach can reduce the need for training instances without sacrificing recall in some scenarios. The results showed a reduction of 70%, 14%, and 19% in the number of training instances required to produce 95% recall and 90% accuracy in three evidence synthesis scenarios. These results suggest that active learning may be a feasible solution for article classification in some circumstances.

In previous methods for classifying articles in MEDLINE that did not employ machine learning methods, researchers analytically derived general seearch terms by examining very large numbers of initial articles [1-8]. These methods were generally unable to identify every relevant article, and recall could only be increased by sacrificing precision, often resulting in article lists in which fewer than half of the articles returned were relevant. As a result, extensive manual review would be required in a second stage of screening. It is likely that the active learning process presented here would improve on these approaches because it uses fewer labels to provide a low risk of missing a relevant study without sacrificing precision.

Machine learning approaches used to classify articles from MEDLINE and other bibliographic databases have been proposed for different purposes [10, 18], and active learning has been proposed for identifying relevant articles to include in systematic reviews [11]. In one example, authors used 10,000 articles as a training set, achieving 73.7% precision and 61.5% recall when identifying articles that were scientifically rigorous [9]. In another, the aim was not to label studies, but rather to order the returned documents so that the most relevant articles were listed first [19]. In the study that applied active learning, the performance of the method was determined by describing a measure of workload reduction, and the results indicated the potential for a 40-50% reduction without sacrificing perfect recall [11].

Another group of studies considers a more general task: the automatic assignment of MeSH labels to articles [20-25]. This is similar to the tasks we considered here because an accurate mapping of MeSH labels to articles could be used to replace or complement a context-specific screening of article types for inclusion in an evidence synthesis task.

The main contribution present in this study is the application of the active learning approach to article classification for evidence synthesis tasks, and an evaluation of the approach using a holdout set, examining the risk of not meeting pre-defined performance criteria for a given number of labellings. Compared to other applications of machine learning, our approach considers a relatively simple problem with binary classification by document type. However, it would be simple to extend the approach to consider more complex screening requirements where only specific study designs were relevant (e.g. specific comparators or outcome measures). Our study is also relatively simple because it considers one form of feature construction (no stemming, stop words, or n-grams), only one type of machine learning algorithm (support vector machines, no stacking or boosting), and a simple method for selecting features (selecting from a list ranked by p-values). The approach may be improved by introducing more sophisticated machine learning methods.

The differences between the three scenarios suggest that class imbalance (for example, where very few articles are relevant), and the context of the task (which influences the features that may be suitable) may both influence the length of the labelling queue required to produce high enough levels of recall and precision to make the approach worthwhile. This means that tasks aimed at identifying rare article groups may benefit least from the active learning approach but this is speculation because we considered only one scenario in which the classes were heavily imbalanced. However, identifying rare types of articles using the active learning approach has been considered in depth elsewhere [17], warranting further investigation.

### Impossible guarantees and empirical estimates of risk

Tasks in evidence synthesis often require 100% recall in screening tasks [26], so it is worth considering whether an automated method can be developed to guarantee perfect recall. In inductive learning, the user and the learner are unable to calculate how well the trained classifiers will perform on unseen examples. This means that although we were able to stop the simulations when the classifier met the pre-specified performance criteria in the experiments here, this is not possible in practice.

This problem is exacerbated by the variability in the classifiers being produced using both passive and active learning within and across the scenarios. This means that the number of articles that need to be labelled by the user to meet a specific risk level may vary for different tasks. While the risk of missing relevant articles is reduced by the active learning process, the variability suggests that we do not yet have a way to (analytically or empirically) determine how many articles need to be labelled to guarantee an acceptable level of risk.

### Limitations

We used only one type of classifier and one method for selecting and constructing features. While support vector machines are appropriate for tasks with a large number of features and document classification tasks in particular, other classifiers and methods for selecting features could affect the results in unexpected ways. Equally, the stopping criteria for Phase I and the final performance criteria that we reported on could have been chosen differently, and these choices could also affect the results.

We examined only one class of clinical interventions (antidepressants), and three relatively simple scenarios in

evidence synthesis. While it would be reasonable to expect that other drug classes and conditions would have similar distributions for article types and produce the same levels of performance, this may not hold for atypical interventions or conditions. Specific search queries aimed at identifying studies that meet specific design specifications may not produce the same results. For example, it may be more difficult to construct a classifier that can separate from all articles those that report on clinical trials comparing two specific interventions and report a specific measurable outcome. Other than issues with class imbalance for cases where there are very few relevant articles, the approach is general so we have no reasons to expect that the method would perform any differently in these cases.

## Conclusion

In this study, we considered the problem of article classification as a shared task between a (human) user and a (machine) learner. Active learning may provide the potential to reduce the manual effort required to classify articles for context-specific tasks in evidence synthesis without reducing the performance. However, variability in the performance of classifiers within and across the scenarios suggests that we do not yet have a way to both ensure that no relevant articles are missed and substantially reduce the workload required to screen documents for inclusion in a study.

## References


[1] Haynes R, McKibbon K, Wilczynski N, Walter S, Werre S, Hedges Team, Optimal search strategies for retrieving scientifically strong studies of treatment from medline: Analytical survey. BMJ, 2005 **330**:1179.

[2] Haynes R, Wilczynski N, McKibbon K, Walker C, Sinclair J, Developing optimal search strategies for detecting clinically sound studies in MEDLINE. J Am Med Inform Assoc, 1994 **1**:447-58.

[3] Haynes R, Wilczynski N, Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. BMJ, 2004 **328**:1040.

[4] Montori V, Wilczynski N, Morgan D, Haynes R, Optimal search strategies for retrieving systematic reviews from medline: Analytical survey. BMJ, 2005 **330**:68.

[5] Robinson K, Dickersin K, Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. Int J Epidemiol, 2002 **31**:150-3.

[6] Shojania K, Bero L, Taking advantage of the explosion of systematic reviews: An efficient medline search strategy. Eff Clin Pract, 2000 **4**:157-62.

[7] Wilczynski N, Haynes R, Hedges Team, Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. BMC Med, 2004. **2**:23.

[8] Wilczynski N, Haynes R, Lavis J, Ramkissoonsingh R, Arnold-Oatley A, Hedges Team, Optimal search strategies for detecting health services research studies in MEDLINE. CMAJ, 2004. **171**:1179-85.

[9] Kilicoglu H, Demner-Fushman D, Rindflesch T, Wilczynski N, Haynes B, Towards automatic recognition of scientifically rigorous clinical research evidence. J Am Med Inform Assoc, 2009 **16**:25-31.

[10] Aphinyanaphongs Y, Tsamardinos I, Statnikov A, Hardin D, Aliferis C, Text categorization models for high-quality article retrieval in internal medicine. J Am Med Inform Assoc, 2005 **12**:207-16.

[11] Wallace B, Trikalinos T, Lau J, Brodley C, Schmid C, Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics, 2010 **11**:55.

[12] Tong S, Koller D, Support vector machine active learning with applications to text classification. J Mach Learn Res, 2002 **2**:45-66.

[13] Warmuth M, Liao J, Rätsch G, Mathieson M, Putta S, Lemmen C, Active learning with support vector machines in the drug discovery process. J Chem Inf Comput Sci, 2003 **43**:667-73.

[14] Liu Y, Active learning with support vector machine applied to gene expression data for cancer classification. J Chem Inf Comput Sci, 2004 **44**:1936-41.

[15] Nguyen H, Smeulders A, Active learning using pre-clustering. In: Proceedings of the 21st International Conference on Machine Learning, ACM. Banff, Canada 2004 623-30.

[16] Joachims T, Text categorization with support vector machines: Learning with many relevant features. In: Proceedings of the 10th European Conference on Machine Learning. Chemnitz, Germany 1998 137-42.

[17] Ertekin S, Huang J, Bottou L, Giles L, Learning on the border: Active learning in imbalanced data classification. In: Proceedings of the 16th Conference on Conference on Information and Knowledge Management, ACM. Lisbon, Portugal 2007 127-36.

[18] Sebastiani F, Machine learning in automated text categorization. ACM Comput Surv, 2002 **34**:1-47.

[19] Cohen A, Optimizing feature representation for automated systematic review work prioritization. AMIA Annu Symp Proc, 2008 **2008**:121-5.

[20] Trieschnigg D, Pezik P, Lee V, de Jong F, Kraaij W, Rebholz-Schuhmann D, MeSH up: effective MeSH text classification for improved document retrieval. Bioinformatics, 2009 **25**:1412-18.

[21] Névéol A, Shooshan S, Claveau V, Automatic inference of indexing rules for MEDLINE. BMC Bioinformatics, 2008 **9**:S11.

[22] Gay C, Kayaalp M, Aronson A, Semi-automatic indexing of full text biomedical articles. AMIA Annu Symp Proc, 2005 **2005**:271-5.

[23] Yetisgen-Yildiz M, Pratt W, The effect of feature representation on medline document classification. AMIA Annu Symp Proc, 2005 **2005**:849-53.

[24] Rak R, Kurgan L, Reformat M, Multilabel associative classification categorization of MEDLINE articles into MeSH keywords. IEEE Eng Med Biol Mag, 2007 **26**:47-55.

[25] Jimeno-Yepes A, Mork J, Demner-Fushman D, Aronson A, Comparison and combination of several mesh indexing approaches. AMIA Annu Symp Proc, 2013 **2013**:709-18.

[26] Tsafnat G, Glasziou P, Choong M, Dunn A, Galgani F, Coiera E, Systematic review automation technologies. Syst Rev, 2014 **3**:74.



**Address for correspondence**

Adam G. Dunn, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 2109, NSW, Australia.
Email: adam.dunn@mq.edu.au

MEDINFO 2015

# Studies in Health Technology and Informatics

This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch. A driving aspect of international health informatics is that telecommunication technology, rehabilitative technology, intelligent home technology and many other components are moving together and form one integrated world of information and communication media. The series has been accepted by MEDLINE/PubMed, SciVerse Scopus, EMCare, Book Citation Index – Science and Thomson Reuters' Conference Proceedings Citation Index.



## Volume 216

*Recently published in this series*



ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

# MEDINFO 2015: eHealth-enabled Health

Proceedings of the 15th World Congress on Health and Biomedical Informatics

Edited by

Indra Neil Sarkar

*Brown University*

Andrew Georgiou

*Macquarie University*

and

Paulo Mazzoncini de Azevedo Marques

*University of São Paulo*

IOS Press

Amsterdam • Berlin • Tokyo • Washington, DC

ISBN 978-1-61499-563-0 (print)
ISBN 978-1-61499-564-7 (online)

*Publisher*
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

*Distributor in the USA and Canada*
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1 703 323 3668
e-mail: iosbooks@iospress.com

# Introductory Remarks from the Scientific Program Chairs

**Fernando Martin-Sanchez[a], Kaija Saranto[b]**

[a] *Health and Biomedical Informatics Centre, Melbourne Medical School, The University of Melbourne, Australia*
[b] *Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland*

MEDINFO is the premier international Health and Biomedical Informatics event. MEDINFO 2015 is hosted by SBIS (Brazilian Health Informatics Society) on behalf of the International Medical Informatics Association (IMIA) and will take place in the city of São Paulo from the 19th to 23rd August 2015. MEDINFO 2015 continues a 41-year tradition of bringing together world leaders, policy makers, researchers, practitioners, educators, and students to exchange ideas and contribute to the latest developments, innovations, and global trends in this rapidly advancing, multidisciplinary field of Health and Biomedical Informatics. This is the first MEDINFO that has been organized to reflect the new two-yearly cycle approved by IMIA. We were thus very happy when we reached the submission and registration deadlines with numbers very similar to previous MEDINFOs that had been organized in three-yearly cycles.

Under the theme: "*eHealth-enabled Health*", the world leaders in this field will gather in Brazil to share knowledge and analyze how Health and Biomedical Informatics is contributing to address some of the most challenging problems in health care, public health, consumer health and biomedical research. Researchers, clinicians, technologists and managers will attend and share experiences on the use of information methods, systems and technologies to promote patient-centered care, improve patient safety, enhance care outcomes, facilitate translational research, enable precision medicine and improve education and skills in health informatics.

This is an historical event as MEDINFO is hosted in Latin America for the first time. Inclusiveness has been a main goal in MEDINFO 2015 with affordable registration fees for the regional audience and use of Spanish and Portuguese language in tutorials and simultaneous translation in sessions held in the main auditorium. MEDINFO 2015 features a pre-congress offering of an extensive tutorial program by leading experts and a student paper competition that draws the best young talent from all over the world. The main program includes keynote talks, papers, posters, panels, workshops, and scientific demonstrations that span a broad range of topics from emerging methodologies that contribute to the conceptual and scientific foundations of Health and Biomedical Informatics, to successful implementations of innovative application, integration, and evaluation of eHealth systems and solutions.

The conference program features five keynote presentations, 178 paper presentations, 248 poster abstract presentations, 27 panels, 30 workshops and 17 scientific demonstrations.

The contributions and presentations included in the program were carefully selected through a rigorous review process involving almost 400 reviewers for a large number of submissions (793) sent by 2500 authors from 59 countries all over the world. The Scientific Program Committee Co-Chairs are grateful to the four Track Chairs, the members of the Scientific Program Committee and all the reviewers who have contributed to the process, and thank the Editorial Committee, the Local Organizing Committee and the IMIA officers (in particular CEOs and VP Medinfo) for assisting us in putting this program together.

The conference participants come to São Paulo from all continents and 60 different countries. We hope that you will enjoy the published proceedings and the overall program!

Sincerely,

Fernando Martin-Sanchez, PhD, FACHI, FACMI &
Kaija Saranto, PhD, FACMI, FAAN
Co-Chairs, MEDINFO 2015 Scientific Program Committee

This page intentionally left blank

# Introductory Remarks from the Editorial Committee Chair

**Indra Neil Sarkar**[a]

[a] *Center for Biomedical Informatics, Warren Alpert Medical School of Brown University, Providence, RI, USA*

Let me join the rest of the organizing committees in welcoming you to MEDINFO 2015 in São Paulo. As the Editorial Committee Chair, I had the distinct honor to review every accepted submission to this year's congress. I personally wish to extend a thanks to the authors for their fine contributions. Together with the meeting participants, MEDINFO 2015 is positioned to be an unprecedented exposition of the finest biomedical informatics innovations with global impact.

Appreciating the international scope of the MEDINFO congresses, it is essential to embrace principles to support scientific inclusivity. Therefore, in contrast to many scientific meetings, the general criteria used for selection into the MEDINFO proceedings is based mostly on scientific merit; language issues are not reason alone for a submission to be not selected. The cost of this inclusivity is that each accepted submission must be carefully reviewed and edited to adjust for language that does not impact the scientific contribution. It is important to note that even submissions from native English speakers may require editing due to variance from the required template, typographical errors, or grammatical issues.

Building on the framework developed by Christoph Lehmann for MEDINFO 2013, Assistant Editors (AEs) were recruited from biomedical informatics training programs (Table 1). The minimum criterion for selection as an AE was at least one first author peer-reviewed English publication (ideally in an informatics conference or journal). Poster submissions were reviewed by one AE; paper submissions by two AEs. The edits were then finalized and assembled into the final proceedings that are in front of you now.

It is important for authors to understand the costs associated with the editing and overall production efforts to ensure the MEDINFO proceedings are of the highest quality possible. Following my esteemed colleagues who served as Editorial Committee Chairs for previous MEDINFOs, I make a plea to each of you to consider the work that is involved when aiming to circumvent the standards established by the organizing committees.

Even moreso than in previous MEDINFOs, strict adherance to the template guidelines was deemed an essential criterion for inclusion in the proceedings. Nonetheless, a number of submissions did clear the peer-review process that still required formatting edits to ensure consistency in font size, spacing, and overall style. In some instances, text had to be significantly edited or figures drastically shrunken or eliminated all together to ensure page limits were respected. Even with such edits, a good faith effort was still made for preserving the scientific message of the contributions. I am thankful for the dedication and hard work of 26 AEs that worked, word-by-word, through each submission and made edits that were ultimately vetted and approved by me.

*Table 1– Assistant Editors (AEs) for MEDINFO 2015*

| Assistant Editor | Institution |
|---|---|
| Samira Y. Ali | Carlow University |
| Andrew B.L. Berry | University of Washington |
| Haresh L. Bhatia | Vanderbilt University |
| Richard Brandt | Texas Tech University |
| Matthew K. Breitenstein | Mayo Clinic |
| Jacqueline E. Brixey | University of Texas at El Paso |
| David Chartash | Indiana University |
| Perry M. Gee | Dignity Health |
| Mattias Georgsson | Blekinge Institute of Technology |
| Anupama Edward Gururaj | University of Texas Health Science Center at Houston |
| Zhe He | Columbia University |
| Kate Fultz Hollis | Oregon Health & Sciences University |
| Silis Y. Jiang | Columbia University |
| Saeed Mehrabi | Mayo Clinic |
| Amir Mohammad | Yale University |
| Tiffany Nicole Moncur | University of South Florida |
| Shauna Marie Overgaard | University of Minnesota |
| Jennifer Elizabeth Prey | Columbia University |
| Lisiane Pruinelli | University of Minnesota |
| Balaji Polepalli Ramesh | University of Massachusetts |
| Joseph D. Romano | Columbia University |
| Charlene Ronquillo | University of British Columbia |
| Ning Shang | Columbia University |
| Harry Tunnell | Indiana University-Purdue University Indianapolis |
| Mary Regina Wysocki | University of Texas Health Science Center at Houston |
| Rafeek Adeyemi Yusuf | University of Texas Health Science Center at Houston |

Finally, I wish to acknowledge the other members of the Editorial Committee (Paulo Mazzoncini de Azevedo Marques and Andrew Georgiou), along with Alvaro Margolis (IMIA Vice President for MEDINFO), Peter Murray (Immediate Past IMIA CEO), Elaine Huesing (Interim IMIA CEO), the leadership of the Local Organizing Committee (Beatriz de Faria Leão and Claudio Giulliano Alves da Costa) and the Scientific Program Committee Co-Chairs (Fernando Martin Sanchez and Kaija Saranto). These proceedings and this meeting are the product of a true team effort– I hope you enjoy MEDINFO 2015 in São Paolo!

Sincerely,

Indra Neil Sarkar, PhD, MLIS, FACMI
Chair, Editorial Committee