*MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation*
*P. Otero et al. (Eds.)*

*© 2022 International Medical Informatics Association (IMIA) and IOS Press.*
*This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).*
*doi:10.3233/SHTI220112*

# Natural Language Processing to Identify Abnormal Breast, Lung, and Cervical Cancer Screening Test Results from Unstructured Reports to Support Timely Follow-up

**Courtney J. Diamond, BS[a], John Laurentiev, MS[b], Jie Yang, PhD[b], Amy Wint, MSc[a], Kimberly A. Harris, MM[a], Tin H. Dang, BA[a], Amrita Mecker, BS[a], Emily B. Carpenter, BS[a], Anna N. Tosteson, ScD[c], Adam Wright, PhD[d], Jennifer S. Haas, MD, MSc[a], Steven J. Atlas, MD, MPH[a], and Li Zhou, MD, PhD[b]**

[a] *Department of General Internal Medicine, Massachusetts General Hospital, Boston, MA, United States*
[b] *Department of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, United States*
[c] *The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, Lebanon, NH, United States*
[d] *Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States*

## Abstract

*Cancer screening and timely follow-up of abnormal results can reduce mortality. One barrier to follow-up is the failure to identify abnormal results. While EHRs have coded results for certain tests, cancer screening results are often stored in free-text reports, which limit capabilities for automated decision support. As part of the multilevel Follow-up of Cancer Screening (mFOCUS) trial, we developed and implemented a natural language processing (NLP) tool to assist with real-time detection of abnormal cancer screening test results (including mammograms, low-dose chest CT scans, and Pap smears) and identification of gynecological follow-up for higher risk abnormalities (i.e. colposcopy) from free-text reports. We demonstrate the integration and implementation of NLP, within the mFOCUS system, to improve the follow-up of abnormal cancer screening results in a large integrated healthcare system. The NLP pipelines have detected scenarios when guideline-recommended care was not delivered, in part because the provider mis-identified the text-based result reports.*

## Keywords:

Natural language processing; Early detection of cancer; Clinical decision support

## Introduction

Cancer is the second leading cause of death in the United States, with 241,400 deaths attributed to breast, cervical, colorectal (CRC), and lung cancers alone [1]. Screening for cancer can reduce cancer-specific mortality, but only if timely and appropriate follow-up is achieved for abnormal cancer screening test results ("abnormal results"). While many healthcare systems have reminder systems to screen eligible patients for preventive cancer tests, follow-up rates for abnormal results are modest and vary widely across cancer type, severity of initial abnormal findings, and clinic and patient demographics [2]. Barriers to follow-up are many: patient-level factors such as social determinants of health may hamper access to care, while system-level issues such as failures in care transitions between the specialists performing or interpreting the test and the patient's primary care provider may delay or prevent the patient from being notified of the result [3][4].

Underlying these potential obstacles to completing recommended follow-up once an abnormal result is known, however, is the ability to identify and reliably notify providers of the abnormal result [5]. Identification of abnormal values is becoming easier and more reliable in the age of electronic health records (EHRs) with structured data and alerts, though it is widely acknowledged that health information technology (HIT) interventions alone are not a panacea for improving follow-up [6]. Even in healthcare systems that do employ HIT in a clinical decision support system (CDSS), not all data elements are structured or easily integrated within the CDSS. Further, though standarized lexicons have been developed and widely adopted for breast, lung, and cervical cancer screening tests [7][8][9], reports themselves are not always structured. Free-text reports are still commonly used in a variety of settings, including certain cancer screening tests, and thus require the careful attention of the interpreting provider. The design and layout of such free-text reports may facilitate the accurate acquisition of information, or may be prone to erroneous interpretation.

Regardless of the ease of reading a free-text report, automated extraction of results can assist with the identification of abnormal results in the cancer screening process and facilitate the application of increasingly complex management guidelines, such as the ASCCP recommendations for cervical cancer screening [10]. Natural language processing (NLP) has been implemented previously to extract results from Papanicolau smear cytology reports [11][12][13], but largely in a retrospective fashion and for research purposes that were not integrated into clinical workflows. Here, we demonstrate the efficacy of a system that performs real-time automatic extraction of results from breast, lung, and cervical cancer screening tests, providing discrete results that are then integrated into the larger study architecture to assist with determining patient eligibility and need for intervention if overdue. Additionally, we demonstrate the utility of an NLP program that identifies specific gynecological procedures that are required for the follow-up of high risk cervical screening results that are otherwise not reliably identified within the EHR.

Motivated by modest and variable follow-up of abnormal cancer screening test results, we designed the multilevel Follow-Up of Cancer Screening (mFOCUS) study (NCI grant U01CA22545), a pragmatic intervention trial that is currently being conducted in 40 primary care practices that are part of three primary care networks. These NLP tools were developed for use in two of these networks and comprise a population management system that allows research staff to accurately identify patients who qualify for outreach to promote follow-up of abnormal results.

## Methods

### Clinical Setting and Patient Population

The mFOCUS trial is being conducted in three primary care networks: Massachusetts General Hospital (MGH), Brigham and Women's Hospital, (BWH) and Dartmouth-Hitchcock Health (D-HH). Of these three networks, only D-HH has codified test results for breast, cervical, colorectal, and lung cancer screening tests; MGH and BWH, which share an EHR infrastructure within the Mass General Brigham (MGB) system, only have a structured data source for colorectal cancer screening results. Thus, the NLP tool was needed for extraction of test results for breast, cervical, and lung cancer screening in the MGB system.

The fundamental criteria for inclusion in the mFOCUS study are: 1) an abnormal result on one of the aforementioned cancer screening tests and 2) being overdue for follow-up of the abnormal result. This includes patients who received follow-up that is clinically inappropriate given the specific abnormality found. Because there are no discrete data readily available in the data warehouse to determine whether a patient has an abnormal result, and because determination of follow-up depends upon knowing the value of the result of the screening test, the NLP tools themselves are crucial for the operation of our study to determine which patients are eligible for participation.

### mFOCUS System Architecture

The mFOCUS system infrastructure is summarized in Figure 1. Data from the Epic EHR is stored in the Epic Chonicles and Epic Clarity databases, and is addtionally loaded into the MGB Enterprise Data Warehouse (EDW). Daily, the EDW is queried for qualifying screening tests, and reports are extracted for storage in the study database. Patient demographic data, such as age, language, sex, and primary care practice affiliation are also queried and stored. An external C# program triggers the exterally stored Python NLP programs to load and read the text reports of newly queried screening tests from the study database. Results are written back to the study database, which then uses the codified results to determine which reports contain abnormal findings. For the subset of patients with abnormal results, the EDW is queried again for any appropriate follow-up procedure occuring after the initial screening test. Finally, patient eligibility is determined within the mFOCUS study database, taking into consideration demographic and past medical history data, the presence of an abnormal result, the absence of an appropriate follow-up procedure, and the time elapsed since the initial screening test. To enable patient outreach via workbenches in the EHR, health maintenance modifiers are added to the charts of eligible patients via Epic Webservices, flagging them for inclusion in the workbench. Additionally, data from the study database are pulled to an external web-application where researchers may access patient details to conduct outreach, record outcomes of outreach, or manually record changes to study status; these user-generated data are also written back to the study database and captured for outcome assessment.

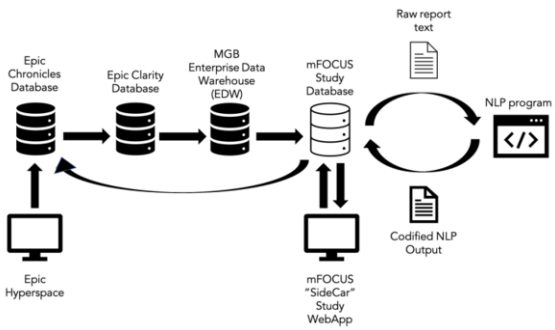


*Figure 1—mFOCUS system architecture*

### Cohort Definitions

#### *Breast, lung, and cancer screening test reports*

Patients in the mFOCUS study are identified from within the two MGB primary care networks. Using structured data from the EDW, patients were identified through affiliation with one of 32 primary care practices within the MGH and BWH networks. Addtionally, we searched for patients with one of three qualifying cancer screening tests: mammogram, low-dose chest CT scan, or Papanicolau (Pap) smear with or without an HPV test. Further, inclusion was limited to those patients with tests performed within the timeframe specifed in Table 1. The earliest permissible date allowed for each type of screening test was calculated based on potential severity of abnormal result, and the relevant recommended guidelines for timely follow-up. For example, the least severe abnormal mammogram result considered eligible is a BIRADS 3, for which a follow-up recommendation of six months is standard. Three additional months are afforded to allow standard care processes to occur; thus, a patient is not considered overdue for follow-up until nine months past the initial screening date. mFOCUS includes all patients who were eligible (and consequently overdue) beginning March 1, 2020; thus, the earliest permissible date for mammograms included in our study is June 1, 2019.

*Table 1—Test inclusion criteria*

| Screening Test | Earliest permissible date |
|---|---|
| Mammogram | 6/1/2019 |
| LDCT scan | 6/1/2019 |
| Pap smear | 9/1/2016 |

#### *Anatomic pathology reports*

Determining which patients are eligible for mFOCUS requires identifying an abnormal result, assigning appropriate follow-up procedures and timing, and finally determining whether an abnormal test has not had the appropriate follow-up in the specified time period. Determining appropriate follow-up was achieved by compiling and validating a list of relevant procedures ordered within the MGB network, and searching for these procedure types for a given patient in the EDW. However, at times, generic procedure codes were used to place orders for follow-up, i.e. "anatomic pathology" in place of "colposcopy", which presented challenges for our follow-up algorithm that searched for specific procedure codes. To mitigate this issue, we developed another NLP pipeline that processed anatomic

pathology reports for patients who were currently eligible for an overdue cervical cancer screening test. Only patients who were currently enrolled in the mFOCUS study for a qualifying cervical cancer screening test were queried for potential anatomic pathology reports.

**NLP Pipeline Development**

A NLP pipeline, as part of the BWH Medical Text Extraction, Reasoning, and Mapping System (MTERMS) NLP suite, was developed for this study, which comprises three parts: 1) Preprocessing: cleaning the pathology report text, including removing excess whitespace and non-alphanumeric characters, and normalizing the words/phrases in different formats (e.g. "follow up"-> "followup"); 2) Pattern matching: using the customized rules captured in a dictionary (including a set of keywords and phrases; details below) to match spans of text in the pathology reports; 3) for extracted patterns, assigning the pathology results according to the dictionary. The initial set of rules were based upon relevant cancer screening guideline recommendations and were reviewed by local specialists. The NLP pipeline was evaluated on a training set of real pathology reports, after which the output was manually reviewed, and used to inform updates to the rules.

A rule-based approach was chosen due to the nature of the task. The information being extracted was specific (test results). The reports relevant to the task are semi-structured, use a well-defined lexicon and contain similar semantic groups (e.g., body location, pathology features/findings, stage level). This approach also enabled negation detection through pattern matching, rather than requiring a large annotated training and data set, as is needed for a machine learning-based approach.

**Dictionary Development**

A dictionary of relevant result phrases was manually compiled for each of the report types analyzed. For each report type, a sample of 500 cases was reviewed, and relevant patterns were annotated and added to the dictionary. The dictionary was improved and reassessed for several iterations until suitable measures of accuracy and precision were achieved (i.e. >95%).

**Evaluation of NLP Performance**

For each report type, NLP performace was evaluated for precision, recall, and F-measure on a test set containing unique cases not used in the training set. Pap smear reports were evaluated on four different aspects: the primary cytological finding (e.g. NILM, ASCUS, LSIL), secondary cytological findings that can co-occur with a primary finding (e.g. presence of endometrial cells or atrophy), the value of the HPV test result if present, and finally the HPV genotype pool tested. To evaluate the NLP pipelines for mammograms, LDCT scans, and the primary cytological component of Pap smears, a training set of 1000 cases was used due to low prevalence of high-level abnormalities. For all other piplines, a test set of 500 cases was used. Additionally, we recorded the number of incidents in which the NLP output alerted our team to a high-level abnormal result that was not initially communicated to the patient by the interpreting or primary care provider.

**Integration and Implementation**

Upon validating the NLP tool, the program was subsequently integrated within the larger mFOCUS architecture. In August 2020, all mammogram and LDCT reports from the earliest permissible date as described in Table 1 up to two months prior were queried from the EDW and stored in the study database; the initial load size is listed in Table 2. Cerivcal reports were loaded into the system in October 2020. For each pipeline, a sample of the NLP outputs was selected for quality assurance and reviewed by study clinicians. We oversampled rare results, i.e. BIRADS 5, LRADS 4X, to ensure accuracy of detecting more severe findings. Once successful integration of the NLP tool was confirmed, we began daily automated queries of the EDW to identify new qualifying tests from the date two months prior to the date of the query. The two month delay allows standard outreach procedures and follow-up to occur, while still identifying potentially eligible patients prior to their becoming overdue. Tests that are determined to have abnormal results, as well as the corresponding NLP output, remain in the system until they are overdue, at which point they become eligible for study intervention. Once eligible, patient charts are manually reviewed by study staff, and any NLP errors are recorded and reported to determine what action is needed to rectify.

## Results

The mFOCUS study and enrollment are ongoing; at the time of writing, a total of 384,495 screening test reports (mammogram, LDCT scan, Pap smear) have been analyzed and assessed for eligibility using the NLP pipelines. Additionally, 2,885 anatomic pathology reports have been queried and processed to determine whether they were appropriate follow-up procedures for mFOCUS-eligible cervical cancer screening patients. A summary of all reports processed is presented in Table 2.

*Table 2– Summary of Reports Processed*

| Report type | Total | Initial load size | Average processed per month |
|---|---|---|---|
| Mammogram | 219,761 | 161,886 | 5,278 |
| LDCT scan | 9,112 | 2,689 | 244 |
| Pap smear | 155,622 | 117,748 | 2,502 |
| Anatomic pathology | 2,885 | 2,352 | 388 |

For the NLP pipelines implemented to extract results from qualifying cancer screening tests, we report high measures of precision and recall for each report type (Table 3). Evaluation of the NLP pipeline for anatomic pathology cases indicates a similarly high measures of precision and recall.

*Table 3– Performance Metrics*

| Text source | Precision | Recall | F-Measure |
|---|---|---|---|
| Mammogram-BIRADS score (n = 1000) | 1.000 | 0.986 | 0.993 |
| LDCT scan-LungRADS score (n = 1000) | 1.000 | 0.999 | 0.999 |
| Pap smear- primary cytology (n = 1000) | 1.000 | 0.983 | 0.991 |
| Pap smear- other cytology (n = 500) | 1.000 | 0.970 | 0.985 |
| Pap smear- HPV test result (n = 500) | 1.000 | 1.000 | 1.000 |
| Pap smear- HPV genotype (n = 500) | 1.000 | 0.978 | 0.989 |
| Anatomic pathology-colposcopy (n = 500) | 0.975 | 0.845 | 0.905 |

In addition to standard measures of NLP performance, we are able to present a tangible measure of the impact of this tool: the number of abnormal results identified by the NLP program that were previously misinterpreted by the patient's provider. Of the 713 high-risk abnormal Pap smear patients manually reviewed after NLP classification, we identified 15 patients who were misinformed of their high-risk HPV positive test result, and thus required additional follow-up.

Error analysis indicated that the primary cause of incorrect outputs is the failure to extract a raw text report from the procedure in question. Many mammograms are performed outside of the MGH and BWH primary care networks, and while results are scanned into Epic Hyperspace, they are not entered into the same fields used to store the results of locally-performed tests. Thus, these tests yield a null NLP output, and require the study staff to manually review the patient's chart to determine the result. Similarly, some Pap smear samples are sent to external laboratories for processing, and reports are entered into a different set of fields in the EHR, creating a semi-structured report which requires additional processing. Another source of error identified is the appearance of a new pattern in a text report that was not previously seen in the original sample set or coded in the relevant dictionary. This occurred largely due to variation in clinician annotation, i.e. using "LungRADS 3", "Lung-RADS 3", and "LungRADS Category 3" interchangeably, and transcription-related errors, i.e. mistyping "ungRADS 3" in place of "LungRADS 3".

## Discussion

The NLP pipelines developed for the mFOCUS trial are crucial to the operations of the study for the two MGB primary care networks. For breast, lung, and cervical cancer screening patients, the NLP programs are the only automated means by which we are able to extract discrete results from free-text mammogram, LDCT, and Pap smear reports, respectively. Thus, it is essential for the automated detection of abnormal results and the identification of eligible patients. Further, the NLP pipelines demonstrated value in detecting abnormal results from free-text reports that were previously missed by the original interpreting provider. This underscores the value of a supplementary text extraction program used in conjunction with free-text reports of screening test results.

Additionally, our NLP pipeline for anatomic pathology reports has allowed us to increase the accuracy of identification of eligible cervical cancer screening patients by determining which patients had appropriate gynecological follow-up for an abnormal result on their initial screening test. Prior to implementation, the accuracy of identifying cervical cancer screening patients who did not have appropriate follow-up for an abnormal result was low, and required manual review of high-risk patients prior to study enrollment to avoid contacting individuals who already had follow-up. The new pipeline allows for automated detection of these patients and thus greatly improves the accuracy of our system without requiring human intervention.

Beyond enabling us to accurately identify eligible patients for the mFOCUS project in real-time, the implementation of the NLP pipelines has provided an opportunity to compare the accuracy of these breast, cervical, and lung cancer screening test result outputs with the semi-structured colorectal cancer screening system that did not require use of NLP. Within the MGB network, several sources are used to document the findings of the colonoscopy: 1) the narrative of the procedure, as dictated by the performing gastroenterologist, 2) an anatomic pathology report, in the event that a polyp is removed, 3) a result letter that is sent to the patient to notify them of the findings and 4) a health maintenance (HM) topic that can be added to the patient's chart and adjusted by the addition of HM modifiers that shorten the recommended follow-up period based upon the findings of the procedure. The standard colonoscopy HM topic has a follow-up interval of 10 years, which is automatically calculated based off the date of the latest colonoscopy found in the patient's chart; when overdue, the topic appears red and appears as a "Care Gap". Modifiers can be added for any time interval at the discretion of the provider, but the mFOCUS study only considers patients with 1, 2, 3, and 5 year modifiers. In theory, a system such as the HM module that relies on discrete data and is well-integrated within the existing EHR infrastructure should have high accuracy of detecting patients who are overdue for their follow-up. However, we found from a sample of 735 patients identified as being eligible for mFOCUS due to an overdue colorectal HM topic that 61 had the incorrect health maintenance modifier applied to their chart. Of those, 5 were several years overdue for their recommended follow-up because the modifier set a longer follow-up interval than what was recommended by the endoscopist and pathologist. This is often caused by a failure to manually update the HM modifier to reflect the patient's most recent result. Given that the sample was drawn only from those patients identified as overdue by the HM topic system, it does not include those patients who are overdue according to their pathology results but have the incorrect follow-up interval modifier applied to their chart. Thus, we have only discovered a subsample of cases in which the HM topic system is inaccurately determining overdue status, and are ultimately understating the issue. For the mFOCUS study, an NLP pipeline that parsed the results of the colonoscopy narratives and anatomic pathology reports may have had greater accuracy in detecting patients overdue for follow-up to an abnormal colonoscopy finding.

### NLP lessons learned

The use of a rule-based NLP approach, while more straightforward to implement than a machine learning-based approach, requires the maintenance of an up-to-date keyword/phrase dictionary. Minor variations in wording between the pathology report text and dictionary patterns caused information to be missed by the NLP algorithm and were a common source of errors during validation. Although new patterns to capture these variations were added to the dictionaries, this approach requires an iterative process of identification and validation in order to ensure a comprehensive dictionary.

Additionally, due to the design of the NLP program, assigning the appropriate pathology result for certain reports necessitated the inclusion of more patterns for negation purposes (i.e. "unmatching" a previous text match in the same report). This introduced complexity that could likely be reduced by reworking the algorithm, given the necessary time and resources.

### Limitations

One limitation of our system is the rule-based nature of the NLP programs used. We needed to randomly select 500 reports to create the initial dictionary and logic rules. Initial iterations of the NLP programs produced many errors owing to the fact that many patterns seen in the test set were not reflected in the training set. As the mFOCUS study will be conducted for a duration of over two years, we anticipate changes in the language used in the reports for each of the cancer screening test types, which will require routine maintenance of dictionaries used.

Further, the results extracted by the NLP pipelines in the mFOCUS system are not immediately transferred to the existing EHR infrastructure, and are only readily available to study

staff with access to the mFOCUS database. To transfer the relevant results and related guideline-based follow-up recommendations back into the Epic infrastructure, Webservices must be called daily for each individual patient with a newly extracted result. An internal module within the EHR that is capable of performing the same text extraction and pattern matching would be of greater value to clinicians who do not have access to external servers to host the program and raw text files, or the requisite Webservices to transfer these results back to the EHR.

Lastly, the NLP pipelines developed for the mFOCUS system are specific to the data available in the Mass General Brigham network, and may not yield accurate results in other settings without tailoring of dictionaries. Many of the report narratives that were ultimately recorded in the dictionaries are specific to providers and their shorthand; even in the case of breast, lung, and cervical cancer screening tests where a standardized lexicon is used to describe findings, variations in punctuation, spelling, and even use of phrases like “and”, “or”, and “and/or” required the addition of multiple phrases which, while providing coverage of reports deriving from the MGB system, may not reflect the patterns found in other networks.

## Conclusions

We demonstrate the high precision and recall of three rule-based NLP pipelines implemented within the greater multilevel Follow-Up of Cancer Screening study to identify patients overdue for follow-up for abnormal breast, lung, and cervical cancer screening test results in the MGH and BWH primary care networks. Additionally, we report the utility of a rule-based NLP pipeline to identify relevant gynecological follow-up procedures from free-text anatomic pathology reports found in the charts of potentially eligible patients with qualifying abnormal cervical cancer screening test results. Further, we show the pragmatic value of these pipelines in identifying patients whose abnormal results were previously undetected by their provider and required additional follow-up to that which was already prescribed. We have discussed the advantage of utilizing a supplementary system such as NLP in conjunction with free-text reports over using semi-structured data sources for identification of patients overdue for follow-up of abnormal results.

## Acknowledgements

This study is sponsored by the National Cancer Institute (U01CA22545).

## References

[1] R.L. Siegel, K.D. Miller, H.E. Fuchs, and A. Jemal, Cancer Statistics, 2021., *CA Cancer J Clin*, **71** (2021) 7–33.

[2] A.N. Tosteson, E.F. Beaber, J. Tiro, et al., Variation in Screening Abnormality Rates and Follow-Up of Breast, Cervical and Colorectal Cancer Screening within the PROSPR Consortium., *J Gen Intern Med*, **31** (2016) 372–379.

[3] C. Moore, O. Saigh, A. Trikha, and J. Lin, Timely Follow-Up of Abnormal Outpatient Test Results: Perceived Barriers and Impact on Patient Safety. *Journal of Patient Safety,* **4** (2008) 241–244.

[4] J. Zapka, S.H. Taplin, R.A. Price, C. Cranos, and R. Yabroff. Factors in quality care–the case of follow-up to abnormal cancer screening tests–problems in the steps and interfaces of care. *Journal of the National Cancer Institute. Monographs,* **2010** (2010) 58–71.

[5] E.G. Poon, T.K. Gandhi, T.D. Sequist, H.J. Murff, A.S. Karson, and D.W. Bates, "I wish I had seen this test result earlier!": Dissatisfaction with test result management systems in primary care, *Arch Intern Med,* **164** (2004) 2223–2228.

[5] A. Georgiou, J. Li, J. Thomas, M.R. Dahm, and J.I, Westbrook., The impact of health information technology on the management and follow-up of test results - a systematic review., *J Am Med Inform Assoc,* **26** (2019) 678–688.

[7] C.J. D’Orsi, E.A. Sickles, E.B. Mendelson, E.A. Morris, et al., *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*, Reston, VA, American College of Radiology, 2013.

[8] American College of Radiology Committee on Lung-RADS®. *Lung-RADS Assessment Categories version1.1.* Available https://www.acr.org//media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf. Accessed on May 14, 2021.

[9] D. Solomon, D. Davey, R. Kurman, A. Moriarty, D. O'Connor, M. Prey, S. Raab, M. Sherman, D. Wilbur, T. Wright, Jr, and N. Young, The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA,* **287** (2002) 2114–2119.

[10] R.B. Perkins, R.S. Guido, P.E. Castle, D. Chelmow, M.H. Einstein, F. Garcia, W.K. Huh, J.J. Kim, A.B. Moscicki, R. Nayar, M. Saraiya, G.F. Sawaya, N. Wentzensen, M. Schiffman, and 2019 ASCCP Risk-Based Management Consensus Guidelines Committee, 2019 ASCCP Risk-Based Management Consensus Guidelines for Abnormal Cervical Cancer Screening Tests and Cancer Precursors., *J Low Genit Tract Dis,* **24** (2020) 102–131.

[11] K.B. Wagholikar, K.L. MacLaughlin, M.R. Henry, R.A. Greenes, R.A. Hankey, H. Liu, and R. Chaudhry, Clinical decision support with automated text processing for cervical cancer screening., *J Am Med Inform Assoc,* **19** (2012) 833–839.

[12] K.E. Ravikumar, K.L. MacLaughlin, M.R. Scheitel, M. Kessler, K.B. Wagholikar, H. Liu, and R. Chaudhry, Improving the Accuracy of a Clinical Decision Support System for Cervical Cancer Screening and Surveillance, *Appl Clin Inform*, **9** (2018) 62–71.

[13] C. Moore, A. Farrag, and E. Ashkin, Using Natural Language Processing to Extract Abnormal Results from Cancer Screening Reports., *J Patient Saf,* **13** (2017) 138–143.

**Address for correspondence**

Li Zhou, MD, PhD

lzhou@bwh.harvard.edu