

**Identifying UMLS concepts from ECG Impressions using KnowledgeMap**  
**Joshua C. Denny, MD<sup>1,2</sup>, Anderson Spickard, III, MD, MS<sup>1,2</sup>, Randolph A Miller, MD<sup>1</sup>,**  
**Jonathan Schildcrout, PhD<sup>3</sup>, Dawood Darbar, MD, PhD<sup>2</sup>, S. Trent Rosenbloom, MD,**  
**MPH<sup>1,2</sup>, Josh F. Peterson, MD, MPH<sup>1,2</sup>**

**<sup>1</sup>Department of Biomedical Informatics, <sup>2</sup>Department of Medicine, <sup>3</sup>Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN**

*Electrocardiogram (ECG) impressions represent a wealth of medical information for potential decision support and drug-effect discovery. Much of this information is inaccessible to automated methods in the free-text portion of the ECG report. We studied the application of the KnowledgeMap concept identifier (KMCI) to map Unified Medical Language System (UMLS) concepts from ECG impressions. ECGs were processed by KMCI and the results scored for accuracy by multiple raters. Reviewers also recorded unidentified concepts through the scoring interface. Overall, KMCI correctly identified 1059 out of 1171 concepts for a recall of 0.90. Precision, indicating the proportion of ECG concepts correctly identified, was 0.94. KMCI was particularly effective at identifying ECG rhythms (330/333), perfusion changes (65/66), and noncardiac medical concepts (11/11). In conclusion, KMCI is an effective method for mapping ECG impressions to UMLS concepts.*

## **INTRODUCTION**

Electrocardiogram (ECG) reports provide a wealth of medical information about myocardial or extracardiac disease and medication effects. While there are algorithms for automated image processing of the ECG tracing, accuracy rates for this approach are inconsistent at 42-96%.<sup>1,2</sup> Examples of errors from automated image processing include misrecognition of rhythms, incorrect ECG interval calculations during arrhythmias, incorrect assessment of ventricular hypertrophy, and missed myocardial infarctions.<sup>1</sup> For these reasons, cardiologists produce a formal “ECG impression” at most institutions. Typically these statements are created semiautomatically by a combination of editable preset phrases generated by ECG software and typed or dictated text from clinicians. The result is that a portion of the interpretation does not have a set vocabulary or structure.

Transforming ECG impressions into structured, coded form is essential to epidemiologic research or decision support around cardiovascular disease and drug safety or efficacy. The cardiac effects of medications, including sudden death from QT

prolongation and increased risk of myocardial infarction, are often discovered years after initial marketing.<sup>3,4,5</sup> Codifying this data would allow for computerized methods to study changes in a patient’s ECG, including QT prolongation or shortening, new or resolved arrhythmias, and identification of myocardial perfusion abnormalities. A reliable system for codifying ECGs could provide real-time decision support to physicians using the electronic medical record.

ECG impressions typically consist of two types of findings: description of the morphologic findings in the tracing (“ST elevation”) and the interpretation of those findings (“acute myocardial infarction”). Automated methods are typically more accurate for the former.<sup>1</sup> While much work has been done to standardize recording of ECGs for communicability<sup>6,7,8</sup>, we could find no encapsulated vocabularies of ECG findings and interpretations. Since significant portions of the statements are input directly by the cardiologist, relying on the preset list of strings in ECG software is not sufficient. Additionally, these statements can be customized, so a solution based on this methodology would not be scalable.

Numerous studies have studied application of standardized vocabularies or natural language processing techniques to clinical documents. The Unified Medical Language System (UMLS) has been applied to radiology reports<sup>9</sup>, discharge summaries<sup>10,11</sup>, problem lists<sup>12</sup>, and medical education documents<sup>14</sup> with recall rates from 0.76 to 0.83 across different studies. We have not found published applications of controlled vocabularies or Natural Language Processing (NLP) systems for ECG interpretations. Since ECGs represent a circumscribed area of knowledge with significant concept repetition, we anticipate that effective matching will leverage unambiguous ECG concepts to discern between ambiguous ones. Many concepts (such as acronyms) are likely to be repeated in other ECGs in an unambiguous way. In addition, concepts in ECGs are likely to cluster with other related concepts. For example, “PE” is likely to mean “pulmonary embolus” and not “pleural effusion”

EKG # 430			
<b>Input sentence #312:</b> Atrial fib. w/rapid ventr. response.			
<u>Exact</u>	<u>Incorrect</u>	<u>Category</u>	<u>Candidate Match</u>
<input checked="" type="radio"/>	<input type="radio"/>	EKG Rhythm	Atrial fibrillation rapid ventricular response (Atrial fibrillation with rapid ventricular response) [Disease or Syndrome]
Unidentified concepts (if applicable): <input type="text"/>			
<b>Input sentence #313:</b> Rate=167			
<u>Exact</u>	<u>Incorrect</u>	<u>Category</u>	<u>Candidate Match</u>
<input checked="" type="radio"/>	<input type="radio"/>	EKG Rhythm	Heart rate (Chronotropism, Cardiac) [Organism Attribute]
Unidentified concepts (if applicable): <input type="text"/>			
<b>Input sentence #314:</b> Diffuse ST-T abnormalities,			
<u>Exact</u>	<u>Incorrect</u>	<u>Category</u>	<u>Candidate Match</u>
<input checked="" type="radio"/>	<input type="radio"/>	EKG findings	Electrocardiographic abnormality (Electrocardiogram (ECG) abnormal) [Finding]
<input checked="" type="radio"/>	<input type="radio"/>	EKG findings	T wave (T wave feature (observable entity)) [Finding]
<input checked="" type="radio"/>	<input type="radio"/>	EKG findings	St interval (ST interval, function (observable entity)) [Laboratory or Test Result]
Unidentified concepts (if applicable): <input type="text" value="diffuse"/>			

**Figure 1:** Sample markup of an ECG in the test set. In this example, “diffuse” was entered as an unidentified concept.

based on repetition in other ECGs and related concepts such as “right ventricular strain.”

Vanderbilt University Medical Center (VUMC) has greater than 10 years of ECG impressions stored in free-text electronic form. In this study, we seek to assess the accuracy of a system for identifying UMLS concepts from ECG impressions and the coverage of ECG concepts in the UMLS.

## METHODS

### *Selection of EKGs*

We selected a test set of ECGs from an anonymized database of all 140,000 ECGs performed at Vanderbilt University Medical Center (VUMC) between 1999-2003. Every ECG includes a free-text cardiologist-generated impression that was created at the time of the reading using Philips TraceMaster ECG Management System<sup>14</sup> and then stored in our electronic medical record in an XML-like form. Each ECG also included calculated intervals and a coded severity (normal, otherwise normal (i.e. a single or mild abnormality such as “frequent PVCs”), abnormal, or defective) that is selected by the interpreting cardiologist. ECG impressions are created from stock phrases stored in the TraceMaster system and free-text entered by the end-user. These stock phrases can be customized by end-users. As an example, “Normal sinus rhythm” is a stock phrase, while “Normal sinus rhythm with frequent PVCs” was created by the end-user by editing the stock phrase at the time of ECG interpretation.

We randomly selected 260 ECGs of severity “normal”, “abnormal”, or “otherwise normal” to be reviewed by three authors (AS, DD, JP) not familiar with the concept identification software. Since impressions from abnormal ECGs include more unique concepts than normal ECGs, we weighted the set to include 75% “abnormal” ECGs and 25% “normal” or “otherwise normal” ECGs.

### *Concept identification of ECGs*

The ECG dataset was processed with the KnowledgeMap concept identifier (KMCI). KMCI is a rigorous scored-based algorithm for mapping free text to Unified Medical Language System (UMLS) concepts.<sup>13</sup> It was designed initially for mapping UMLS concepts in medical education documents, and this is the first application of the software to clinical narrative. When processing a free text document, KMCI first identifies sentences and then noun phrases using a rule-based part-of-speech tagger. By using approximate natural language processing techniques, KMCI rearranges lexically-connected noun phrases for improved matching (“left ventricle is enlarged” is transformed into “enlarged left ventricle”). KMCI generates numerous word variants. Finally, KMCI uses document-based statistical techniques using unambiguous concepts to help select from among possible candidates for ambiguous phrases.<sup>13</sup> At this time, KMCI does not interpret negation.

Prior to testing, we briefly reviewed a set of 100 ECGs to optimize the matching algorithm by

	ECG		Perfusion Changes	Noncardiac Medical	Medical Qualifiers		Nonmedical	Total
	Rhythm	findings			Structure	Medical		
Total Concepts	333	477	66	66	11	155	63	1171
Correct	330	424	58	65	11	123	48	1059
Incorrect	2	31	7	0	0	14	11	65
Unidentified	1	22	1	1	0	18	4	47
Recall	0.99	0.89	0.88	0.98	1.00	0.79	0.76	0.90
Precision	0.99	0.93	0.89	1.00	1.00	0.90	0.81	0.94

**Table 1:** Recall and precision by category. Five percent of ECG findings and 11% of medical qualifiers were unidentified concepts, most of which were not present in the UMLS.

expanding common abbreviations (with for “w/”, without for “w/o”, rule out for “r/o”, due to for “d/t”, and consistent with for “c/w”). In addition, we generated a list of “ignorable words” for UMLS candidate concepts. These included ECG, electrocardiographic, electrocardiogram, myocardial, cardiac, heart, segment, wave, interval, and lead. The KMCI algorithm does not significantly penalize candidate UMLS concepts when these “ignorable” words are absent in a document phrase (e.g., the candidate concept “ECG low voltage” is not penalized significantly when matching the document phrase “low voltage” since “ECG” is an ignorable word).

After tuning the system, we processed the test set of ECGs as a document. We used the 2005AA version of the UMLS for this study.<sup>15</sup>

#### ECG scoring by reviewers

Three reviewers (two internists and a cardiologist) scored the ECGs through a web interface that processed the XML output from the KMCI in sets of 10 ECGs each. **Figure 1** shows the ranking interface. Reviewers were presented with the original ECG sentence followed by the KMCI matches with matching string, preferred concept name, and the semantic type of the concept. The three reviewers marked each identified concept as “exact” or “incorrect.” If a concept could be expressed accurately as components, each component was considered individually. For example, “anterior MI” could be matched as “anterior” and “myocardial infarction” but also as “anterior myocardial infarction.” Reviewers manually entered unidentified concepts during scoring. They also categorized the concept represented by the original text string into one of seven categories: ECG rhythm (e.g., sinus rhythm, atrial fibrillation), ECG finding (e.g., QT prolongation, ST elevation), myocardial perfusion abnormality (e.g., myocardial infarction or ischemia),

myocardial structures (e.g., ventricular hypertrophy), noncardiac medical concepts (e.g., pulmonary embolus), medical qualifier (e.g., borderline, anterior), or nonmedical concept. For purposes of scoring, we ignored negation, such that a concept would be marked “exact” if it matched “change” instead of “no change.” We cataloged all statements manually for presence of negation or other qualifiers (e.g., “possible”, “cannot exclude”).

Because we expected a high concordance between reviewers, we chose percent agreement to assess interrater reliability; it was calculated as the number of items agreed upon divided by total items scored. We randomly selected twenty ECGs (10 normal, 10 abnormal) to be read by 2 of 3 reviewers in order to assess agreement. Recall was calculated as number of correct concepts (true positives) divided by total number of concepts in the document (true positives + false negatives [incorrect and unidentified concepts]). Precision was calculated as number of correctly identified concepts (true positives) divided by all concepts identified by the indexer (true positives + false positives).

After evaluation, we analyzed the ECGs to determine the number of unique statements. We replaced all references to ECG leads and sequence of leads (e.g., “V1-3”), dates, location specifiers (e.g., “anterior”), and other numbers with placeholders. We then counted the number of unique statements.

## RESULTS

Of 140 ECGs scored by reviewers, 1172 concepts were either identified by KMCI (1124/1171, 96%) or identified by reviewers as unidentified concepts (47/1171, 4%). One hundred five of the ECGs were classified as abnormal; 35 were classified as normal. The 20 ECGs double-scored represented 138 concepts. There was a 99% agreement between reviewers as to “exact” and “incorrect” rankings. There was a 74% agreement between the type

Category	N	Examples
Rhythm	2	Volume rate (“v-rate”)
ECG Findings	31	Level R (“R/S”), S [immunologic factor], short PR interval (“shorter PR”), WPW pattern (“repolarization pattern”), T wave (“wave”)
Myocardial structure	7	Vanadium (abbreviation “V”)
Medical qualifiers	14	Double vision (“double”). Periodicals, lead high (“high QRS voltage”)
Nonmedical	11	diabetic retinopathy (“DRS” - readers initials), Apr gene (date)
Unidentified concepts	47	Tracing (11), leftward (6), diffuse (5), various ECG leads (4), age (2), nondiagnostic (2), supraventricular (1), anterolateral infarct (misspelled document word: “arterolateral infact”), unusual (1), possible (1)

**Table 2:** Examples of incorrect and unidentified concepts by category. Original ECG text is in parentheses where applicable. For the unidentified concepts, the numbers in parentheses indicated the number of times the concept was reported by the reviewers.

categorizations; however, if “medical qualifier” and “nonmedical concept” were grouped, there was a 91% agreement between categorizations.

Overall recall for KMCI for all concepts was 90%; precision was 94%. As seen in **Table 1**, KMCI most frequently missed nonmedical concepts and medical qualifiers; it was the most accurate for medical concepts. **Table 2** lists all unidentified concepts. Additionally, 290 (27%) of the correctly-matched candidate UMLS phrases contained extra words that made matching more difficult (i.e., the document phrase was “low voltage” but the best UMLS match was “ECG low voltage”). We termed these candidate phrases “overmatches.” Of the medical concepts, the most difficulty was with ECG findings; many of these errors were due to incorrect overmatches (see table 2). There was one missed rhythm (“IVCD”) and one missed perfusion abnormality due to misspelling. In three cases, an incorrect overmatch with “WPW pattern” was made due to a high co-occurrence score with other concepts. The most common correct overmatches were “heart rate”, “ECG abnormality”, “ST segment”, and “T wave.” Overmatches including the word “lead” were generally errors.

Of 734 phrases in the study set, 247 were unique statements when stripped of date, time, location specifiers, and numbers. There were 16 negated phrases (3 were medical concepts, 9 “no change”, 4 “no prior tracing”). Thirty-six phrases were qualified as possible, identified by the words “possible” (11), “cannot rule out” or “cannot exclude” (6), and “consider” (19).

## DISCUSSION

KnowledgeMap, an automated concept identifier based on the UMLS, successfully identified the majority of clinically significant concepts within the free text interpretation of an ECG. The engine was most accurate with concepts related to cardiac rhythm and ischemic changes with few incorrect mapping and missed concepts. KMCI’s ability to accurately

match candidate concepts with extra words (i.e., “overmatches”) proved essential to its performance since they comprised 27% of the correct matches. KMCI was able to match these concepts due to their relation to other matched concepts in the document. However, overmatching induced medically significant error on three occasions (0.25%), where KMCI returned “WPW pattern” as an overmatch for the document word “pattern” due to the document-based scoring techniques. As expected, UMLS’ coverage of diagnoses (as manifested by rhythm, perfusion changes, and other medical concepts) was superior to its performance with ECG findings. This is ultimately a result of less synonymy for ECG findings. Augmenting the UMLS with a more formal ECG glossary would likely improve the performance of KMCI in this domain.

The vast synonymy and granularity of the UMLS generally leads to improved concept matching but can present certain challenges for information retrieval. Other algorithms, such as MedLEE, which codify text in a post-coordinated fashion, may prove superior for retrieval since it preserves hierarchical relationships.<sup>11</sup> For example, KMCI would match “anterior myocardial infarction” as a single concept whereas MedLEE would match it as “myocardial infarction” with location “anterior.” The parent-child type relationships defined in the UMLS provide a possible method for overcoming this limitation.<sup>15</sup>

This validation study of KMCI performance has several notable limitations. We did not formally test UMLS coverage of ECG concepts, which would require validating the engine’s matches with a much larger set of unique phrases. However, of the unidentified and incorrect concepts, only a few modifiers such as “leftward”, “diffuse”, and “nondiagnostic” were missing from the UMLS. In this study, we did not discriminate between partial matches and complete matches of clinical concepts (e.g., “ST-T wave abnormalities” and matched component concepts “T wave”, “ST segment”, and “ECG abnormality”). KMCI links component

concepts together; however, additional NLP-type processing will be needed to coordinate these concepts for future interpretation. Because the tested ECG interpretations were partially derived from standard phrases in the Philips TraceMaster software, the performance of the software might differ if used to identify concepts in ECG interpretations without standard phraseology. In general, the performance is expected to worsen if more of the interpretation included text entered directly by the cardiologist since one would expect more misspellings and nonstandard abbreviations. However several KMCI enhancements including a spell checker and a larger variety of abbreviations may compensate for the greater variability in free text. Restricting the UMLS concepts to only heart-related concepts would likely also lessen matching errors. Finally, the concept identification is only the first step in a 2-step process that will utilize NLP techniques to classify the concepts as absent or present, possible or probable. In a real implementation, a system would need to distinguish between “no change” and “change.”

KMCI is a promising tool for identifying UMLS concepts from semi-structured or free-text clinical narrative. Several domains of research including quality improvement, safety and efficacy of drug treatments, rapid identification of patients qualifying for clinical trials are dependent on reliably transforming unstructured text into coded format. Additionally, our ability to organize and process information for real-time decision support is highly dependent on the degree that the source data is accurately coded. We anticipate that new research and projects in these domains will be stimulated by the availability of a large database of coded ECG information.

#### ACKNOWLEDGEMENTS

We would like to thank Mark Arrieta and Dario Giuse, Dr. Ing., for their assistance in assembling the set of ECGs. This study was supported in part by grants LM08576 and LM007995 from the National Library of Medicine.

#### REFERENCES

1. Willems JL, Abreu-Lima C, Arnaud P, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med.* 1991;325:1767-73.
2. Theodore L. Tsai, Douglas B. Fridsma, and Guido Gatti. Computer Decision Support as a Source of Interpretation Error: The Case of Electrocardiograms. *J. Am. Med. Inform. Assoc.* 2003;10(5):478-483.
3. Ray WA, Murray KT, Meredith S, Narasimhulu SS, Hall K, Stein CM. Oral erythromycin and the risk of sudden death from cardiac causes. *N Engl J Med.* 2004 Sep 9;351(11):1089-96.
4. Bresalier RS, Sandler RS, Quan H, et al. Cardiovascular Events Associated with Rofecoxib in a Colorectal Adenoma Chemoprevention Trial. *N Engl J Med* 2005; 352:1092-1102.
5. Solomon SD, McMurray JJV, Pfeffer MA, et al. Cardiovascular Risk Associated with Celecoxib in a Clinical Trial for Colorectal Adenoma Prevention. *N Engl J Med* 2005; 352:1071-1080.
6. Wang H, Azuaje F, Jung B, Black N. A markup language for electrocardiogram data acquisition and analysis (ecgML). *BMC Med Inform Decis Mak.* 2003 May 7;3(1):4.
7. OpenECG. Available at: <http://www.openecg.net/>. Accessed 3/16/2005.
8. FDA XML Data Format Design Specification Available at: [http://www.edisc.org/discussions/EGC/FDA\\_XML\\_Data\\_Format\\_Design\\_Specification\\_DRAFT\\_B.pdf](http://www.edisc.org/discussions/EGC/FDA_XML_Data_Format_Design_Specification_DRAFT_B.pdf). Accessed 3/5/2005.
9. Lowe HJ, Antipov I, Hersh W, Smith CA, Mailhot M. Automated semantic indexing of imaging reports to support retrieval of medical images in the multimedia electronic medical record. *Methods Inf Med.* 1999;38(4-5):303-7.
10. Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. *J Am Med Inform Assoc.* 2001;8(1):80-91.
11. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc.* 2004 Sep-Oct;11(5):392-402.
12. Tuttle MS, Olson NE, Keck KD, et al. Metaphrase: an aid to the clinical conceptualization and formalization of patient problems in healthcare enterprises. *Methods Inf Med.* 1998 Nov;37(4-5):373-83.
13. Denny JC, Smithers JD, Miller RA, Spickard A 3rd. "Understanding" medical school curriculum content using KnowledgeMap. *J Am Med Inform Assoc.* 2003 Jul-Aug;10(4):351-62.
14. Philips TraceMaster ECG Management Software. Information available at: [http://www.medical.philips.com/main/products/ardiography/products/ecg\\_management/tracemaster/index.html](http://www.medical.philips.com/main/products/ardiography/products/ecg_management/tracemaster/index.html). Accessed: 3/13/05.
15. National Library of Medicine. UMLS Knowledge Sources, 2005AA edition. Available at: <http://www.nlm.nih.gov/research/umls/>. Accessed 3/06/05.