EcgRuleML: A Rule-Based Markup Language for Describing Diagnostic ECG Criteria

RR Bond, DD Finlay, CD Nugent, G Moore

University of Ulster, Belfast, United Kingdom

Abstract

From the clinician's perspective, computerized interpretation of the electrocardiogram (ECG) is carried out in a 'black box'. That is, the rules used for interpretation are not easily accessible to the clinician. In this study we propose the ECG Rule Markup Language (ecgRuleML) as a way to externalize decision rules used to interpret the ECG. EcgRuleML utilizes the eXtensible Markup Language (XML) to provide a framework for articulating quantitative rules for measuring intervals, segments, widths, peaks, heart rate and the cardiac axis. Abstract features of the ECG such as slurred S waves cannot be easily represented numerically and are therefore articulated using codes. To test the ecgRuleML framework, rules have been defined to assess ST Elevation Myocardial Infarction (STEMI) in a Lux-192 Body Surface Potential Map (BSPM). An algorithm has been integrated into a BSPM viewer where the rules have been parsed from an ecgRuleML document and executed in 63ms (mean from 10 trials) on a PC (3GHz CPU, 3GB RAM).

1. Introduction

Clinical Decision Support Systems (CDSS) and computerized interpretation techniques are commonly used to assist humans in making rapid and accurate diagnostic decisions [1]. Complex signals such as the ECG require a lot of expert knowledge for accurate interpretation. As a result, automated ECG interpretation is important and it has been shown to improve diagnostic accuracy when used to assist a clinician [1].

A number of computerized techniques are available to interpret the ECG, each of which has their own advantages and disadvantages. These techniques include multivariate statistical analysis, decision trees, artificial neural networks (ANNs) and fuzzy logic to name but a few [2]. Nevertheless, traditional rule based ECG interpretation techniques are still widely used today [2, 3]. Although ANNs have been criticized for being a 'black box' solution [4], in reality even rule based techniques may be viewed as a 'black box' solution, at least from the clinician's perspective. This is due to the fact that rule based systems are assembled by a knowledge engineer who develops the rules from the clinician's description of ECG criteria and compiles such rules into unreadable code, only understood by the computer [5]. Therefore, the clinician receives the decision without having access to the rules that were used to make the decision. Another drawback to existing rule based techniques is that they often utilize static knowledge bases and as a result do not always incorporate the most up-to-date information [6]. This is unfortunate as medical knowledge and ECG criteria frequently change.

As a possible solution to the aforementioned problems, we propose ecgRuleML as a way to externalize decision rules. EcgRuleML was developed using XML which allows decision rules to be read by both machines and humans. Hence clinicians could use ecgRuleML or a graphical interface to ecgRuleML to articulate new rules and update existing rules without consulting a computer programmer to revise and recompile the system. This is a much needed feature since "the rapid evolution of medical knowledge makes knowledge base maintenance a particular problem" [6]. EcgRuleML also supports standardization and interoperability as rules can be shared and exchanged amongst heterogeneous systems and researchers. As a result, systems may incorporate rules from numerous experts. This would, in turn, improve the system itself and avoid duplication of work.

This idea of open and human readable decision rules is not necessarily a new ideology. The ECG Criteria Language (ECL) was developed in the 1980s by Hewlett-Packard [7]. ECL was designed to be a rule language that could be read and used by a cardiologist. Nevertheless, ECL was still quite similar to a programming language and did not see widespread use. Given that ecgRuleML exploits XML, we envisage that it has the potential to be more accessible than ECL. Furthermore, although ecgRuleML is the first rule based markup language for the ECG, similar XML languages have already been developed. The Home Rule Markup Language (HomeRuleML) and HomeCI was developed to allow healthcare professionals to create rules that can be used to infer decisions from the activity in a smart home environment [8]. These studies have influenced the development of ecgRuleML.

ISSN 0276−6574

2. Methodology

EcgRuleML was developed using XML which is a text based markup language used to structure information semantically [9]. XML has already been used to establish interoperable formats for storing ECG datasets such as ecgML [10] and XML-BSPM [11]. To date, XML has not yet been used to structure decision rules used to interpret the ECG. Nevertheless, XML is a suitable solution as reasoning can be optimized using XML's inherent tree structure. The ecgRuleML schema is presented in Figure 1.

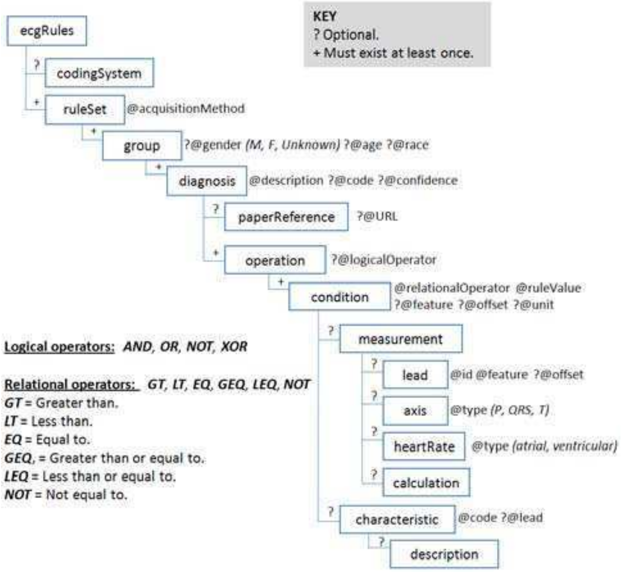

Figure 1: EcgRuleML XML schema.

The root element ***ecgRules*** contains one required child element called ***ruleSet*** which can exist multiple times. The root element also consists of one optional sub element called ***codingSystem***. This is used to specify the coding system used in identifying a diagnosis. Possible coding systems include SNOMED [12] and ICD-10 [13]. The ***ruleSet*** element has one required attribute called ***acquisitionMethod*** which retains the name of the ECG acquisition technique, i.e. 12-lead ECG, Frank leads. This attribute allows for the articulation of different rules for different ECG acquisition methods. This is novel as ECG interpretation systems normally cater for standard acquisition methods [14]. The ***ruleSet*** element has one sub element called ***group*** which can also exist multiple times. The ***group*** element has three optional attributes called ***gender***, ***age*** and ***race***. This allows for the definition of rules dependent on gender, age and ethnicity. The ***gender*** attribute can have one of three enumerated values (***M***, ***F***, ***Unknown***). The ***age*** attribute can specify age groups, i.e. '21-30'. The ***race*** attribute specifies an ethnic group using codes supplied by the Centers for Disease Control and Prevention (CDC) [15]. This aspect of ecgRuleML is important given that gender and age dependent criteria already exist [16]. If these attributes are not specified in a document, the inference engine would default to applying all the enclosed rules to all ages, genders and ethnic groups.

The ***group*** element has one sub element called ***diagnosis*** which can be used multiple times. The ***diagnosis*** element has one required attribute called ***description***, which can be used to describe a diagnosis e.g. 'Anterior Myocardial Infarction'. The ***diagnosis*** element also has two optional attributes called ***code*** and ***confidence***. The ***code*** attribute stores a code from the previously specified coding system. The ***confidence*** attribute stores a float value between 0 and 1. This denotes the level of certainty that the pathology is present when the encapsulated rules are true. The ***diagnosis*** element consists of one required sub element called ***operation*** which can exist multiple times. The ***operation*** element has one optional attribute called ***logicalOperator*** which can have one of four enumerated values (***AND***, ***OR***, ***NOT***, ***XOR***). The ***operation*** element consists of one required sub element called ***condition*** which can also exist multiple times. The purpose of the ***condition*** element is to encapsulate a rule that may be true or false. The ***condition*** element has one required element called ***relationalOperator*** which can have one of six enumerated values (***GT***, ***LT***, ***EQ***, ***GEQ***, ***LEQ***, ***NOT***). The ***condition*** element also has four optional attributes called ***ruleValue***, ***feature***, ***offset*** and ***unit***. In respect to the ***relationalOperator***, the ***ruleValue*** attribute can contain a numerical value or a Boolean value (***true***, ***false***). It may be necessary to measure one component of the ECG against another component, e.g. is the R wave in V1 greater than the S wave in V2. As a result, the ***feature*** attribute can specify which component of the ECG is to be measured and used as a rule value i.e. ***rPeak*** etc. These quantitative features have been listed in Table 1.

Table 1: Enumeration values of the ***feature*** attribute.

Intervals	Segments	Peaks	Widths	Other
prInterval stInterval qtInterval	prSegment stSegment	pPeak qPeak rPeak sPeak tPeak uPeak	pWidth qrsWidth tWidth uWidth	jPoint QTc

The ***offset*** attribute stores a numerical value representing milliseconds. This is used if a feature such as the ***jPoint*** is defined and an offset of 60ms is required to measure ST60 amplitudes. Finally, the ***unit*** attribute specifies the unit of measurement, i.e. ***ms***, ***mV, bpm***,

degrees etc. The ***condition*** element can have one of two optional sub elements called ***measurement*** and ***characteristic***. The ***measurement*** element specifies quantitative conditional values which can be measured against the numerical value specified in the aforementioned ***ruleValue*** attribute. The ***measurement*** element can have one of four optional sub elements called ***lead***, ***axis***, ***heartRate*** and ***calculation***. The ***lead*** element has two required attributes (***id***, ***feature***) and one optional element (***offset***). The ***id*** attribute defines the lead and the ***feature*** attribute defines a measurable feature of that lead. This ***feature*** attribute is a clone of the other ***feature*** attribute of the ***condition*** element. As a result they both share the same list of values in Table 1. The ***offset*** attribute stores a value in milliseconds for reasons previously explained. The ***axis*** element has one required attribute called ***type*** which is used to specify which axis measurement, i.e. ***P***, ***QRS***, ***T***. The ***heartRate*** element has one required attribute called ***type*** which also specifics whether to measure the atrial or ventricular heart rate. The ***calculation*** element can be used to define simple calculations. This is useful as simple calculations are often found within diagnostic ECG criteria, e.g. if the S wave in V1 + R wave in V5 is more than 35mm then the patient may have Left Ventricular Hypertrophy (LVH).

The ***measurement*** element is used to extract quantitative features of the ECG. A number of abnormal features cannot be easily represented numerically. As a result, the ***characteristic*** element can be alternatively used to represent abstract features such as slurred S waves and M shaped P waves. This element has one required attribute (***code***) and one optional attribute (***lead***). The ***code*** attribute specifies a code used to define an abstract feature of the ECG. The ***lead*** attribute is used to specify a lead that would be assessed to determine if the abstract feature exists. The ***characteristic*** element has one optional sub element called ***description***, which provides a textual description of the abstract feature. This ***description*** element sustains the human readability of the document as isolated codes are unintelligible.

3. Results

The ecgRuleML format has been shown to allow the storage of a range of ECG decision rules. For example, the ecgRuleML framework was successful in representing diagnostic criteria extracted from the ECG library website [17]. Furthermore, since ecgRuleML supports different ECG acquisition techniques, it can be used in more advanced scenarios. In this study, rules have been defined to assess STEMI in a Lux-192 BSPM. From the 192 leads available, a sample of 74 leads where identified for STEMI assessment. That is, 15 leads were selected to assess Anterior Myocardial Infarction (MI), four for Septal MI, six for Right Ventricular MI, six for Apical MI, ten for Lateral MI, six for High Posterior MI, 15 for True Posterior MI and six for Inferior Posterior MI. Defining rules for STEMI assessment using ecgRuleML was straight forward. STEMI is present if ST elevation exceeds a specified threshold in two or more contiguous leads [18]. For example, Anterior MI may be present if ST elevation exceeds 0.2mV in at least 2 of the specified 15 leads. Therefore rules must be defined to allow identification of each individual lead and its associated contiguous leads. This procedure involves a lot of rules as leads in some regions can have up to eight contiguous leads. An example rule from an ecgRuleML document for assessing STEMI (Anterior) in one BSPM lead and its contiguous leads can be viewed in Figure 2. The equivalent IF-THEN rule can also be viewed in Figure 3.

```
<diagnosis description="Acute Anterior Myocardial Infarction">

    <operation>

        <condition relationalOperator="GEQ" ruleValue="0.2" unit="mV">
            <measurement>
                <lead id="75" feature="jPoint" offset="60" />
            </measurement>
        </condition>

    </operation>

    <operation logicalOperator="OR">

        <condition relationalOperator="GEQ" ruleValue="0.2" unit="mV">
            <measurement>
                <lead id="63" feature="jPoint" offset="60" />
            </measurement>
        </condition>

        <condition relationalOperator="GEQ" ruleValue="0.2" unit="mV">
            <measurement>
                <lead id="64" feature="jPoint" offset="60" />
            </measurement>
        </condition>

        <condition relationalOperator="GEQ" ruleValue="0.2" unit="mV">
            <measurement>
                <lead id="76" feature="jPoint" offset="60" />
            </measurement>
        </condition>

        <condition relationalOperator="GEQ" ruleValue="0.2" unit="mV">
            <measurement>
                <lead id="87" feature="jPoint" offset="60" />
            </measurement>
        </condition>

        <condition relationalOperator="GEQ" ruleValue="0.2" unit="mV">
            <measurement>
                <lead id="88" feature="jPoint" offset="60" />
            </measurement>
        </condition>

    </operation>

</diagnosis>
```

Figure 2: An excerpt of an ecgRuleML file where one lead and five contiguous leads are assessed for STEMI.

The resulting XML document is 80 kilobytes and is

```
IF (Lead75_ST60 >= 0.2) &&
((Lead63_ST60 >= 0.2) ||
(Lead64_ST60 >= 0.2) ||
(Lead76_ST60 >= 0.2) ||
(Lead87_ST60 >= 0.2) ||
(Lead88_ST60 >= 0.2))

THEN {
Diagnosis = "Acute Anterior
MI"
```

Figure 3: Example IF-THEN rule where one lead and five contiguous leads are assessed for STEMI.

equivalent to approximately 74 IF-THEN rules each utilising multiple relational operators (i.e. '>=') and multiple logical operators (i.e. '&&'). An inference algorithm has been integrated into a BSPM viewer [19] where STEMI rules have parsed from the ecgRuleML document and executed within 63ms (mean from 10 trials) on a PC (3GHz CPU, 3GB RAM). This result is reasonable as it was anticipated, prior to the study, that externalizing decision rules would dramatically increase the processing time of an inference engine because of the added requirement of parsing rules.

4. Conclusion and further work

An ecgRuleML document can be described as a human and machine readable rule based representation of expert knowledge. The goal of this study was to allow for the efficient management of an expert ECG knowledge base. However, for this to be fully realized, additional work must be performed. This will involve a drag and drop graphical interface that would allow clinicians to interactively and intuitively define rules. These graphically defined rules can then be converted into the ecgRuleML format to be interpreted by an inference engine. In summary, such developments have the potential to grant clinicians more control, and trust, over computerized ECG interpretation.

References

[1] Salerno SM, Alguire PC, Waxman HS. Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. Ann.Intern.Med. 2003;138(9):751-760.

[2] Papaloukas C, Fotiadis DI, Likas A, Stroumbis CS, Michalis LK. Use of a novel rule-based expert system in the detection of changes in the ST segment and the T wave in long duration ECGs. J.Electrocardiol. 2002;35(1):27-34.

[3] Mahesh V, Kandaswamy A, Venkatesan R. A Rule-based Expert System for ECG Analysis. International Journal of Engineering and Technology 2009;1(3):194-200.

[4] Itchhaporia D, Snow PB, Almassy RJ, Oetgen WJ. Artificial neural networks: current status in cardiovascular medicine. J.Am.Coll.Cardiol. 1996;28(2):515-521.

[5] Dunsmuir D, Daniels J, Brouse C, Ford S, Ansermino JM. A knowledge authoring tool for clinical decision support. J.Clin.Monit.Comput. 2008;22(3):189-198.

[6] Shortliffe EH, Cimino JJ. Biomedical informatics: computer applications in health care and biomedicine. Springer; 2006.

[7] Doue JC, Vallance AG. Computer-Aided ECG Analysis. Hewlett-Packard Journal 1985;36(9):29-31.

[8] Nugent CD, Davies RJ, Hallberg J, Donnelly MP, Synnes K, Poland M, Wallace J, Finlay D, Mulvenna M, Craig D. HomeCI-A visual editor for healthcare professionals in the design of home based care. Proceedings of the IEEE Engineering in Medicine and Biology Society; 2007.

[9] Extensible Markup language (XML). Available from: http://www.w3.org/XML/

[10] Wang H, Azuaje F, Jung B, Black N. A markup language for electrocardiogram data acquisition and analysis (ecgML). BMC Medical Informatics and Decision Making 2003;3(1):4.

[11] Bond R, Finlay D, Nugent C, Moore G. XML-BSPM: an XML format for storing Body Surface Potential Map recordings. BMC Medical Informatics and Decision Making 2010;10(1):28.

[12] SNOMED CT. Available from: http://www.ihtsdo.org/snomed-ct/

[13] WHO: International Classification of Diseases (ICD). Available from: http://www.who.int/classifications/icd/en/.

[14] Macfarlane PW, Devine B, Clark E. The university of Glasgow (Uni-G) ECG analysis program. Proceedings of Computers in Cardiology 2005; 32:451-455.

[15] CDC Race Code List. Available from: http://www.cdc.gov/nchs/data/dvs/RaceCodeList.pdf

[16] Macfarlane PW, Hampton DR, Clark E, Devine B, Jayne CP. Evaluation of age and sex dependent criteria for ST elevation myocardial infarction. Proceedings of Computers in Cardiology 2007; 34.

[17] Jennings D, Gerred S. Electrocardiogram (ECG, EKG) library. Available from: http://www.ecglibrary.com

[18] Wagner GS, Macfarlane P, Wellens H, Josephson M, Gorgels A, Mirvis DM, et al. AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part VI: Acute Ischemia/Infarction. J.Am.Coll.Cardiol. 2009;53(11):1003.

[19] Bond R. Body Surface Potential Map Viewer. Available from: http://bspm.raymondbond.com

Address for correspondence
Raymond Bond (Room 16J26),
University of Ulster (UUJ)
Shore Road,
Newtownabbey,
Co. Antrim,
BT370QB,
bond-r@email.ulster.ac.uk