Context Identification in Electronic Medical Records

By

Reejis Stephen

M.B.B.S

Saint Johns National Academy of Health Sciences, 2001

Submitted to the Department of Health Sciences and Technology
in Partial Fulfillment of the Requirement for the Degree of
Masters of Science in Medical Informatics

At the

Massachusetts Institute of Technology

August 2004

Signature of Author: _____
Department of Health Sciences and Technology
August 31, 2004

Certified By: _____
Aziz Boxwala
Clinical Assistant Professor of Radiology, HMS
Affiliated Faculty, Division of Health Sciences and Technology
Thesis Supervisor

Accepted By: _____
Martha L. Gray
Director, HST
Edward Hood Taplin Professor of Medical and Electrical Engineering, MIT

Context Identification in Electronic Medical Records

By

Reejis Stephen

Submitted to the Department of Health Sciences and Technology
On August 31, 2004 in Partial Fulfillment of the
Requirements for the Degree of Masters of Science in
Medical Informatics

ABSTRACT

In order to automate data extraction from electronic medical documents, it is important to identify the correct context of the extracted information. Context in medical documents is provided by the layout of documents, which are partitioned into sections by virtue of a medical culture instilled through common practice and the training of physicians. Unfortunately, formatting and labeling is inconsistently adhered to in practice and human experts are usually required to identify sections in medical documents. A series of experiments tested the hypothesis that section identification independent of the label on sections could be achieved by using a neural network to elucidate relationships between features of sections (like size, position from start of the document) and the content characteristic of certain sections (subject-specific strings). Results showed that certain sections can be reliably identified using two different methods, and described the costs involved. The stratification of documents by document type (such as History and Physical Examination Documents or Discharge Summaries), patient diagnoses and department influenced the accuracy of identification. Future improvements suggested by the results in order to fully outline the approach were described.

Thesis Supervisor: Aziz Boxwala, M.B.B.S., Ph.D.
Title: Clinical Assistant Professor of Radiology, HMS

Acknowledgments

I would like to thank my advisor Dr. Aziz Boxwala for providing insightful feedback, criticism and direction to the work. This would not be possible without his support, guidance and patience.

None of this would have been possible without access to the database and text parsing tools provided by the Eclipsys Corporation. Thanks to Paul Gertman and Lynn Martin for their mentorship, encouragement and provision of these resources.

Thanks also to Dion Harmon who lent his ear to the ideas as they evolved and his skilled editing of my bulky code. Thanks to Vishwanath for his enthusiasm constant interest with such ideas, and for his help with using the documentation-free linguistic software and to Joyce who provided all her help and computation hardware when my system crashed. Thanks also to Anoop for his invaluable help with formatting. And thanks to Wai Ling and Ed for keeping my spirits high throughout.

Of course, none of this would have been possible without my family, so thank you Mom and Dad, Reena, Kumar, Keets and Rony.

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Background

The Nature and Use of Medical Information

Clinical encounters such as hospital stays or outpatient visits generate a large amount of information. This comprises demographic data, like age, gender, race, contact information, address, payer information and clinical data such as laboratory test results or reports, clinical history that is obtained from the patient or a proxy, findings from the physical examination, and treatment given to the patient. This information is highly useful to the patient care provider, and institution for several utilities:

- For reimbursement to the provider by the payers [1-3].
- For grouping into research cohorts.
- For measuring the quality of care meted out in a visit and whether it conformed to the standard practice guidelines.
- For measuring operational performance of an institution and planning of resources
- For determining whether established guidelines yield desired outcomes.
- For implementing clinical decision-support for the providers and patients.
- For longitudinal care of the patient

Many of the discrete data items such as visit outcomes, diagnoses, cost of items and services and diagnoses, needed for the above applications, are often abstracted from the documentation generated by health care delivery personnel. Hence, there is a large potential to extract this information from both perspectives of a cost of care and quality improvements in health care delivery.

Much of the information is available only as a narrative text and it needs to be converted into a codified standard form. Numerous schemes for codifying medical data exist- such as SNOMED (Systemized Nomenclature of Medical and Surgical concepts) and ICD-10 (International Statistical Classification of Diseases – tenth revision), all of which arose for the express need of standardizing the meaning of what is usually captured in natural language. The problem remains processing the data from narrative text into these codified forms. Except in a few situations, human abstractors currently do this, and the volume of work involved poses a significant cost to hospitals. Thus, a very small portion of the possibilities for utilizing these data has been realized.

Currently medical information is available in several hospitals through the Electronic Medical Record (EMR) systems. These are also known as Computer-based Patient Records (CPR). EMRs are a "repositories of electronically maintained information about an individual's lifetime health status and health care, stored such that it can serve the multiple legitimate users of the record" [4]. In its current form even though electronically available, much of the useful clinical data is still in pre-processed textual form in medical databases. A very small part of the information such as laboratory test results and

demographic details are captured directly as granular elements by the system. Much of the rest remains locked in the natural language of clinical documents in EMRs.

To obtain discrete data instead of natural language text, several attempts have been made earlier [2]. One approach is to captured codified data during clinical data entry using the above mentioned and similar coding schemes [5]. But such methods have been difficult to implement in clinical practice except in very limited domains as they prove to be time consuming and disruptive of the clinical workflow [6-8]. This produces a resistance towards the implementation of codified data entry in clinical environments. This is because it takes sufficient effort to make representation systems that capture all the clinical nuances, and these are very likely to be incomplete requiring additions that may be overlooked given the detail of what can be said using natural language. There is the added step of training users to pick the right codes for the intended meaning of terms, as the terms may not be obvious to them. This is an intensive task and if the users are not properly trained, there is potential for errors in data entry. Hence text-based electronic records are likely to persist in use long enough to warrant automated information-extraction techniques to be developed if the utilities mentioned earlier are to be realized.

Extraction of Information from Text Documents

There are several approaches to extracting granular information from documents. The most commonly used is by employing human abstractors to manually go through electronic medical files, reading the text and establishing context before picking the value

of granular data elements they are interested in. This process is time consuming and tedious. Abstraction by humans is also costly and error prone. Hence, human abstractors are typically employed to extract only the most essential information- typically diagnoses entries for the purposes of payer reimbursement.

This limited usage does not allow for the bulk of information in medical documents that may be used for various purposes such as the aforementioned benchmarking initiatives, cohort selection, physician profiling and the like to be captured in reasonable time as it takes the average coder time to go through a record for the most granular information such as diagnoses or specific quality control data like the JCAHO (Joint Commission for the Accreditation of Health Organizations) core measures [5, 9]. Hence, this does not scale to the extraction of all the potentially useful information from documents, which is subsequently lost.

Another method of extracting information from free text is by using Natural Language Processing (NLP). These techniques enable the rapid extraction of information from electronic text documents with an accuracy that is only limited by the ambiguity and form of the syntax employed in the documents. The technique promises to realize the range of possibilities that medical information may serve, as it allows automated extraction in bulk, across a variety of purposes at relatively low human effort. Much of the human effort is in the initial investment and subsequently allows rich yields.

Natural Language Processing

Natural language processing is done by either of two methods or a combination of both. These are linguistic-based parsing and statistical techniques. Linguistic parsers take the syntactic form of sentences into consideration. Every input sentence is broken into components and an attempt is made to match each part with a known database of words or groups of words that correspond to clear parts of speech- like the subject or object of a sentence. Meaning is inferred based on the known form of the syntax represented. The power of such tools increases with the number of syntactic sentential forms incorporated into the model. Statistical techniques on the other hand attempt to capture meaning by looking at the frequency of the association between words and a particular concept of interest. For example if the word pattern "heart attack" is preceded within some defined length of characters by the word pattern "not a", the inference might be made that a heart attack is not present. Other methods are used for statistical parsing such as Hidden Markov Models [9] and the application of more qualifier patterns to subjects, but the number of things that may be tried are open ended and beyond the scope of this discussion. The validation of the data extracted by this technique is intractable especially when more complex meanings than the simplest facts are desired. In several cases, the subject of pronouns is only attributable based on the context, and that level of establishing context from text is impossible to achieve in its entirety.

Both methods attempt to extract a value for concepts represented in documents. Concepts are concisely captured meanings which are the basis of reasoning [2]. They form the

elements whose values are granular data needed for the purposes mentioned above. For example, "aspirin" is a concept. It has synonyms like "acetyl salicylate" that map to the same concept. NLP parsers treat all synonym instances of a concept as the same element. Concepts are available readily from ontological hierarchies of concepts like from SNOMED-CT and the International Classification of Diseases (ICD) systems among others, and have found wide utilization in the standardization of medical data through codification [2, 10-12]. The system chosen depends on the specific use of the elements the parser is trying to capture. A typical example of a SNOMED concept hierarchy includes child concepts and how they relate to parent concepts. For example, "Aspirin" is a concept that is a child of the concept "medication" which in turn is a child of the concept "substance". Concepts may also belong to different trees- i.e. multiple inheritances are possible and a concept can be the child of more than one parent.

Before either method of NLP extraction may be applied, in medical documents, errors are reduced if the context in which the information was obtained from is first identified. For example, the term diabetes might have different implications in the context of a section on family history than in a section of history of present illness. Medical documents have a structure comprised of context-specific sections of text by virtue of the methodical approach health care professionals have towards patients that is usually consistent. As an illustration, the prominent clinical text book on the approach to a patient, the Hutchison's Clinical Methods, twentieth edition suggests the following sub headings for sections of a History and Physical Examination document [13]:

The presenting complaint.

The history of presenting illness.

The history of previous illness.

The menstrual history.

The obstetric history.

The treatment history.

The family history.

The social history.

The occupational history.

Review of systems.

These sections are followed by the Physical Examination. The physician will typically conclude with a conclusion or impression of the case and a plan for care. The structure of such a document is generally the same for all physicians who practice the allopathic system of medicine with some variations based on local formats or physicians' personal preferences. For instance, a few of the sections might be left out – such as obstetric history or treatment history. There is usually a "presenting complaint" section followed by a "history of presenting complaint" section and always a "physical examination" section.

Thus, there are two clear steps to the process of NLP-based extraction. Identifying the context, and then parsing the context for the concepts of interest. This work attempts to make progress on automating the problem of context identification. The next section discusses in detail the sections involved in different kinds of clinical text documents.

1.2 Problem Statement

Document Types and the Content of Sections

Two types of documents of interest for clinical information are the History and Physical Examination and the Discharge Summary documents. The first is generated during the initial assessment of the patient and the latter is a summary of that episode of care at the end of the visit. There are other kinds of text documents such are progress notes, consultation reports, clinical or laboratory test reports, and nursing notes, but the focus of this research is on History and Physical Examination and Discharge Summary documents. Both these are divided into sections. A History and Physical Examination document has the following sections:

1. The presenting complaint.

   This contains an explicit set of problems stated by the patient in his or her own words as to the reasons for the visit.

2. The history of presenting illness.

   This section goes into each of the complaints enumerated in the presenting complaints section and attempts to elucidate associations and distinctions that further help in identifying the causes for each.

3. The history of previous illness.

Any chronic illnesses or past illnesses that might change the approach to the patient are described in this section.

4. The menstrual history.

   This section is not always present, but identifies the details of the menstrual cycle, that may be pertinent and affect any kind of care, but tend to be overlooked.

5. The obstetric history.

   This section is usually very relevant in Obstetrics and Gynecology cases and affects decisions made about the patient in this context.

6. The treatment history.

   This section goes into details about the current illness and what treatment has already been received or self-administered.

7. The family history.

   Relevant illnesses in the family are obtained in this section as well as details necessary to decide whether intervention on the family level is necessary or if there are aspects of the illness complicated by association with the family.

8. The social history.

   In several instances, social circumstances complicate a case or offer insights into understanding the origination of the illness. These are recorded here.

9. The occupational history.

   Many diseases are consequential of certain occupations and these occupations can complicate some diseases.

10. Review of systems.

The purpose of this section is to identify history related to every other system than the ones that are complained about to ensure that nothing is missed.

11. Physical examination.

    In this section, the physician objectively examines the patient and records findings.

12. Conclusion or impression and plan.

    Here the physician's interpretation of the case, as it appears is summarized along with a plan for care.

A Discharge Summary document summarizes the initial visit and subsequent assessments. Hence, it has many sections in common with the History and Physical Examination document. This document type also has other sections. These are:

1. Hospital course.

   This comprises details of the stay and interventions carried out during this episode of care.

2. Discharge diagnoses

   This is a list of the diagnoses the patient has- both past chronic diagnoses and the ones identified during the visit.

3. Discharge plan.

   This section explains the next steps to be taken with regard to that episode of care, such as when a follow up visit is needed, medications that are prescribed, or what might be done in the future for the patient.

Several sections are not described in the traditional books, but health care personnel have found them convenient to record and these are consistently found within both kinds of documents mentioned above. Two good examples of these are:

1. Allergies

   It is sometimes grouped under personal history, but at other times independently noted and is where allergies particularly to medications are noted.

2. Medications.

   The medications the patient is on at the time of admission are noted here.

The best index for the structure of a document is the local practice at a place. The structure drifts over long periods but is quite consistent in the short term and adapted to the needs of the environment in which it develops.

Identifying Sections within Electronic Medical Text Documents

As illustrated above in section 1.1, ascertaining context in medical documents prior to parsing the text for data mining purposes greatly increases the accuracy of the data captured. Electronic medical texts are handwritten by physicians or transcribed from audio files. Rarely are they typed in directly as separate documents in the database. In most cases, they usually have section headings in them, and context identification is usually possible using simple rules. However, sometimes section headings are left out altogether; sometimes, wrong section headings are given and even non-standard section headings might be used. Sections may also seem to blend into each other without any clear-cut distinction. For these reasons, a simple parsing of section names is not sufficient

for labeling sections (assigning a fixed context marker to the section) prior to data mining efforts. This makes any attempt to parse out sections unreliable by itself at best and potentially misleading at worst.

Attempting to identify sections without relying on explicit headings given to them must then be based on other generalized features that can be consistently identified. These would have to range from the topological properties (see below) to the content within sections. This work explores the feasibility of an approach to identifying sections based on easily extractable content and topological features.

Prior Work in this Domain:

Prior work by Hahn et al addresses the problem of identifying section with a complex approach via linguistic methods which was implemented in a software tool [14]. There also have been attempts to solve the section identification problem in the commercial sector such as the Flemish company, Language and Computing, NV. These solutions are proprietary. The only other academic work that attempted to address this problem was communicated via a poster presentation at the AMIA 2001 [15] conference that promised a solution to the section-identification problem, but no approach has been published since. While other information extraction approaches have been successful, these usually circumnavigated the problem of section identification, by confining NLP-based data mining efforts to relatively single-context documents like test or laboratory reports [16].

Approaching a Solution using Artificial Neural Networks

Identifying sections based on the content and position of sections may be possible using a supervised learning approach when no rigid rule exists a priori, by which to determine a section. This allows for an overall pattern or regularity to be discerned and classification of sections done based on the characteristics of sections. One of the best pattern recognition techniques is the Artificial Neural Network (ANN) or Neural Network (NN) for short.

A neural network is a mathematical model for information processing based on a connectionist approach to computation. In a neural network model, simple nodes (or "neurons", or "units") are connected together to form a network of nodes - hence the term "neural network"[17].

A typical feedforward neural network is designated by a set of input nodes and output nodes connected by a set of hidden nodes in between. These nodes are processing units. Each layer of nodes may be connected to the subsequent layer in arbitrary ways under control by the user. Nodes pass on their outputs to the nodes they are connected to. Each node takes an input and adjusts it by a weight before producing an output with a function, typically a sigmoid function. The final output is compared against the true output for the case and the degree of difference or error is used to modify the weights by a feedback process such that the modification of the weights is proportional to the amount each contributes to the overall error [18].

With an adequate number of training cases, a suitable pattern of weights can be obtained provided the model is complex enough, to ensure a good characterization of output predictions that match or are close to the true values for the cases. This network can then be applied to unknown cases to classify them according to the experience gained during the learning process.

1.3 Study approach

It is possible to apply neural networks to solving the problem of section identification, provided suitable section characteristics can be identified in a quantifiable manner.

A manual review of the documents was done during this study to identify a set of such features. These were of two kinds:

1. Topological or surface features: Broad descriptives of a section, such as the distance from the start of the document or the size of the section.
2. Qualitative features: A metric of partially quantifying characteristic content within a section.

These two groups of features for each section together constituted the input set of variables to the neural network.

Topological Features

The surface features that enable identifying sections were obtained by inspection of electronic text documents used as the experimental data in this study. Sections were grouped physically into paragraphs, or groups of paragraphs. The size of the section was significant as some sections like Presenting Complaints are much smaller than Physical Examination section for instance. The distance from the beginning of the document was also found to be important (sections like Presenting Complaints are always first while Conclusion or Plan sections are usually the last in the document). When the content of adjacent sections are clubbed together, a single section in the document would present a larger size than its component sections individually. This can confound prediction. The start position and the size of a section, are not sufficient by themselves to differentiate such a variant from the standard cases. To allow a flexibility in recognizing such variants, a third surface feature variable representing the distance of the end of the section from the end of the document was added.

Hence the three surface features selected were:

1. The section size (Size).
2. The distance from the start of the document (Start).
3. The distance from the end of the document (End).

Qualitative Features (Content Variables):

To control for the difference in style of the language used in the document, an attempt was made to distinguish between identifiable medical concepts within text versus the size of the text - a ratio which would decrease with the increasing use of non-concept words. The SNOMED-CT vocabulary was used as a universal set of concepts. The number of concepts from the SNOMED-CT in a section were identified using a parsing tool, that identified synonyms of these from a vocabulary of concept synonyms. Concepts from each section that were characteristic of the subjects of particular sections were used to test if sections could be differentiated on the basis of this. The counts of concepts that were felt to capture the characteristics differentiating between sections constituted the remaining qualitative input variables listed below. Synonyms of SNOMED-CT concepts were used as it was felt that this hierarchy had sufficient detail to enable sets of concepts that distinguished between types of sections. As SNOMED-CT allows multiple inheritance, whenever concepts belonged to two groups, they were excluded from one of the groups as specified below, inorder to preserve the distinctness in the content represented by each of the qualitative variables.

The final set of qualitative feature variables were:

1. Medication concept count (Med)

   A count of all children of the SNOMED concept "drug, medicament or biological substance" – concept ID 311980000.

2. Procedure concept count (Proc)

   A count of all children of the SNOMED concept "procedures" concept ID – 71388002)

3. Investigation concept count (Inv)

A count of all children of the SNOMED concept "laboratory procedures -general" – concept ID 269814003.

4. Diagnoses-related concept count (Diag)

A count of all children of the SNOMED concept "disease" – concept ID 64572001.

5. Findings-related concept load (Finding)

A combined count of all the children of the SNOMED concepts "Clinical history and Observation findings" concept ID 250171008, "findings by method" concept ID 118240005, "finding by site" concept ID 118234003, "clinical history/examination observable"- concept ID 363788007, excluding all the children of "symptom" – concept ID 19019007.

6. Symptom-related concept count (Sympt)

A count of all the children of the SNOMED concept "symptom" – concept ID 19019007.

7. Family concept count (Fam)

A count of all the children of the concept "person in the family" – concept ID 303071001

Each section was evaluated with respect to these ten dimensions and a NN was used to predict test cases based on learning done on a training set.

The format of sections and the content changes with the cultural effects of local practice. Hence, the rules for characterizing sections are likely to vary with the department, kind of document and type of diagnoses under evaluation. Thus, it is not possible to identify every kind of section in every document for these reasons. The

goal of this project was to attempt identifying sections at the absolute granularity of labeled sections, a lesser granularity where certain sections would be considered together, and to estimate how accuracy varied across departments, diagnoses groups, and type of documents.

Chapter 2

Materials and Methods

2.1 Data Source and Initial Processing

This study was performed on the electronic text documents that are part of a large-scale data warehouse, developed by the Eclipsys Corporation, a company that makes EMR systems. A number of History and Physical Examination and Discharge Summary documents were randomly selected and analyzed. There were 109 History and Physical Examination documents and 79 Discharge Summary documents, from a 500 bed hospital used in the analysis. The data warehouse is implemented on a Microsoft SQL Server database management system. The documents in the database were accessed using SQL queries.

The documents were manually inspected to determine the beginning and ending locations of sections. Each section was identified manually by expert review (by the author, who is a physician) and section label assigned regardless of the actual labels in the document. During this process, the clinical department where the document originated and the principal diagnostic group of the case were also noted. A large text file of the documents with this additional information tagged on to each section in every document was produced.

A Perl script using Regular Expressions (Regex) was then used to parse out the sections from the documents and the information tagged with the identity of each section was used as the "gold standard". The set of Regex expressions that partition the documents into relevant sections was ascertained by inspection. The expression patterns varied for Discharge Summary and History and Physical Examination documents and the sets are displayed in Appendix 1 a.

2.2 Description of Documents and Sections

All the 109 History and Physical Examination documents were from the Emergency Medicine department. The documents were categorized by the diagnosis group of case being evaluated. The number of sections that this yielded for each group is shown in Table 1.

Table 1. Number of documents and number of sections within diagnoses groups for History and Physical Examination documents

History and Physical Examination		
Diagnosis group	Number of documents	Number of sections
General Medicine	33	522
Respiratory medicine	15	262
Neurology	12	217
Cardiology	16	282
Surgery	9	123
Gastroenterology	7	134
Oncology	4	49
Psychiatry	3	58
Endocrinology	3	30
Gynecology	2	36
Orthopedics	2	37
Obstetrics	1	15
Urology	1	16
ENT	1	18
Total	**109**	**1799**

The Discharge Summary documents were grouped by the department in which they were created. The numbers of documents for each department are shown in Table 2. The medical subspecialties were grouped together to give one group called "Other Medical Specialties".

Table 2 Number of documents and number of sections within departmental groups for Discharge Summary documents

Discharge Summary		
Department	Number of documents	Number of sections
General Medicine	22	189
Other medical Specialties	28	189
Obstetrics and Gynecology	16	113
Surgery	13	107
Total	**79**	**598**

For the History and Physical Examination documents, there was a slight difference between the expected format and the kind of sections available in this set, but the differences were found to be remarkably consistent across the whole set, possibly because they were all from the same department. The sections identified for this set of documents were:

1. Presenting Complaints (PC).
2. History of Presenting Complaints (HOPC).
3. Past Medical or Past Surgical History (PH).
4. Medications (M).
5. Allergies (A).
6. Family History (FH).
7. Personal History (PerH).
8. Social History (SH).
9. Occupational History (OH).
10. Review of Systems (RS).

11. Physical Examination (PE).
12. Laboratory Investigations (L).
13. Conclusion or Plan (P).

For the Discharge Summary documents, the sections obtained were generally a similar set across departments. These were:

1. History (H)
2. Past Medical or Surgical History (PH).
3. Social History (SH).
4. Personal History (PerH).
5. Family History (FH).
6. Medications (M)
7. Laboratory data (L).
8. Physical Examination (PE).
9. Allergies (A).
10. Hospital Course (HC).
11. Discharge Diagnoses (DD).
12. Discharge Plan (DP).

The accuracy of the section- identification method was evaluated on five groupings of the documents:

1. The group of Discharge Summary documents as a whole (at a granularity described in the next section).

2. The group of History and Physical Examination documents as a whole at low granularity (as described below in the next section).
3. The group of History and Physical Examination documents at highest granularity. In this grouping, every kind of section as seen in the original files was represented.
4. Grouping the Discharge Summary documents by the department the documents originated.
5. Grouping the History and Physical Examination documents by the principal diagnosis.

2.3 Grouping of sections

For group 3 above, the entire set of thirteen sections as described above was chosen for each History and Physical Examination document (high granularity grouping).
The remaining groups using the History and Physical Examination documents (groups 2 and 5) were evaluated on the low granularity grouping defined below as this was felt to capture context optimally:

1. Presenting Complaint.
2. History of Presenting Complaint.
3. Past History, Medications, Allergies (PhMdAg)

   In location, these three sections were in roughly the same region of the document but inconsistently present.
4. Personal History, Family History, Social History, Occupational History (PFSO).

These were also in a similar region of the documents and present to different degrees.

5. Review of Systems.
6. Physical Examination.
7. Laboratory and Plan (LP).

   These were usually together and of similar contextual significance.

The groups based on the Discharge Summaries (1 and 4) were also considered at a lower granularity than the granularity in the documents as defined below:

1. History (H).
2. Past Medical or Surgical History.
3. Social, Personal or Family History (SPFH).

   These three sections were variably present in the same location.
4. Medications (M).
5. Allergies (A).
6. Physical Examination (PE).
7. Laboratory Data (L).
8. Hospital course (HC).
9. Discharge diagnoses (DD).
10. Discharge plan (DP).

2.4 Preparation of Datasets

A prior set of concepts extracted by string-parsing using a commercial tool developed by the NLP company Language and Computing was already available for use with the Eclipsys database. These previously extracted concepts formed the input data for the qualitative feature variables. Seven feature variables, databases for each feature were implemented with Microsoft SQL Server database management system, based on groups of SNOMED-CT concept hierarchies described in section 1.3 above, to enable counts of the number of concepts within each variable-type for each section. This process was effected by running SQL scripts to count the number of parsed concepts in each section. The three surface feature variables were extracted using Perl scripts. The descriptive statistics of the input variables of both types of documents are in Appendix 1 b.

For each grouping studied, the sections were randomly partitioned into three sets. Two sets were used for training and the third was a holdout set used for evaluation.

2.5 Training of Neural Network

Training was done using the NevProp (Nevada University Back-Propagation) version 3 software. The model generated was then used to produce predictions that were tested using other methods. The Neural Network parameters for the training and prediction are in Appendix 2. Details of each are available from the NevProp3 user's manual [19] .

The training was performed well past the best epoch, using a heuristic that optimized for both discrimination and calibration (details in Appendix 2). The best model learnt was used to predict the unseen cases in a holdout set.

Perl scripts were used to parse the result files and evaluation of the output was done according to the various metrics described using both Perl and R scripts. Predictions for each of the available section-types in each of the five document groupings were made. The results were then evaluated according to metrics described in the next section.

2.6 Context Assignment Methods

Contexts are embodied in the partitioning of the document into well-defined regions or sections. Labels or section-types are the markers that identify the sections parsed from the documents (PE, A, PhMdAg etc). The neural network assigns predictive scores to each input section for each section label. Context identification can be done in two ways using the NN predictions:

Type 1 Method: In this method, scores are compared across all input sections in a document for every label. The label is assigned to the input section that receives the highest score for that label.

For example if a section had the prediction outputs 0.2, 0.3, 0.4, 0.1 for the labels PE, PC, HOPC and A respectively, then the label assigned by the method would be HOPC.

Rationale: With this kind of prediction, a document is partitioned into its constituent contexts, which can be stored in a database and based on which filtered extraction of concepts by NLP parsing can be done.

Type 2 Method: In this method, scores are compared within an input section for every label. The section is assigned a label that exceeds a predefined threshold. In this way, a section can be assigned more than one label.

For example, if all the predictions for the section PC are in the range 0.2 to 0.9 and if the chosen threshold is 0.8, then only sections with predictions between 0.8 and 0.9 are assigned the label PC. No sections with predictions below 0.8 would be considered PC sections. The same section might also be labeled HOPC if it crosses the threshold necessary to label it as an HOPC section.

Rationale: It is possible that a particular kind of section the user is interested in and which the NN predicts well, is mislabeled because another section-type has yielded a higher prediction value. When a user is interested in only one kind of section for extraction, such as only Discharge Diagnoses, for example, and chooses to ignore all other kinds of sections in the document, it is useful to see how the tool predicts the Discharge Diagnoses section to the exclusion of all other sections. This gives the user higher yields of correct context even when the section is mislabeled as per method 1 described above, but this approach cannot be used when the user is interested in more than one section-type for a given set of documents.

2.7 Evaluation Metrics

For the Type 1 Method:

Choosing an appropriate label for a section from a list of valid section labels is a task of discrimination. Hence the metrics used to evaluate the performance of the type 1 method were the measures of discrimination [20]:

Sensitivity (Sen):

It is the probability that the tool will recognize a section of a given type. With respect to a particular type of section, a True Positive (TP) is a correctly identified section of that type and a False Negative (FN) is a section of that type that the tool misclassified as a different kind of section.

$$Sen = TP \div (TP + FN)$$

Specificity (Spec):

It is the probability that the tool will correctly rule out sections when they are not of a given type under consideration. With respect to a particular section type, a True Negative (TN) is a section that is not of that type and that the tool did not misclassify as that type and a False Positive (FP) is a section misclassified as that type.

$$Spec = TN \div (TN + FP)$$

Positive Predictive Value (PPV):

It is the probability that a section labeled as a particular type is actually a section of that type.

$$PPV = TP \div (TP + FP)$$

Negative Predictive Value (NPV):

It is the probability that a section not labeled as a particular type, in actuality is not one of that type.

$$NPV = TN \div (TN + FN)$$

Accuracy (Acc):

It is the number of sections that the tool correctly classified. It is the number of true positives for the number of sections in the set.

$$Acc = TP \div SC$$

These metrics were evaluated for each type of section when the sections were labeled based on the highest prediction value. The average value for each grouping of documents was also computed.

For the Type 2 Method:

Within a given section type, the tool can predict a section variably, depending on what prediction threshold is chosen. Hence, the four discrimination indices can change with the

threshold chosen. However as each prediction tries to mirror the actual value, a calibration of performance is possible.

For each section-type in each of the five document groupings, the discrimination across all valid thresholds was measured using the Receiver Operator Curve (ROC), where the area under the curve, which is equivalent to the c-index, is an indication of how good the discrimination is.

The Calibration indices used were The Brier Score (BS) and the Hosmer Lemeshow (HL) statistic.

The Brier score is calculated as follows:

$$BS = \sum \frac{(P - A)^2}{n}$$

Where P is the prediction and A is the actual value (1 or 0) for a given section and n is the number of sections in the set.

The Hosmer Lemeshow goodness of fit statistic ($\hat{C}$) was the Pearson's Chi-Square statistic (with g-2 degrees of freedom) from the table of observed and predicted output frequencies:

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - nP_{av})^2}{nP_{av}(1 - nP_{av})}$$

where there are g = 10 bins with n predictions in each bin and P_{av} is the average of the predictions in the k^{th} bin and O is the observed sum for the k^{th} bin [21]. Both the Brier

score and the $\hat{C}$ test were implemented using scripts in R on the output generated by the Neural Network.

2.8 Method comparisons

The two methods were compared in performance against the actual label values for each section and against each other. For method 2, the choice of a best threshold had to be made for a single model to use for the comparison. This was done by computing every possible specificity and sensitivity for all thresholds and then picking the model with a threshold that maximized the vector of Sensitivity and Specificity. The discriminatory statistics for this model were then calculated.

Chapter 3

Results

3.1 Input Variable Relevance

For each grouping of variables, the NN utilized the input variables to different degrees. A statistic built into the NevProp software computed how much each variable contributed to the prediction. This is called the Automatic Relevance Determination (ARD) statistic and each variable's contribution to the model given in terms of the ARD Relevance statistic is in Table 3 below. The numbers are a percentage of the total contribution by the variables. Each row sums to 100 percent. Details of the statistic are in the NevProp user's manual [20].

Table 3. ARD relevance statistic for each of the topological and qualitative variables across all groupings of documents

			Variable									
		%	Start	End	Size	Med	Proc	Invest	Diag	Find	Sympt	Fam
History and Physical Examination		High Gran	8.84	4.8	6.44	6.48	12.9	10.97	7.89	20.1	10.5	10.97
		Low Gran	7.77	9.49	3.97	26.3	4.52	14.95	17.8	3.69	0.98	10.54
	Diagnoses Groupings	Card	10	10	10	10	10	10	10	10	10	10
		Gastro	9.26	11.4	9.91	9.06	3.7	25.77	12.2	14.2	3.11	1.44
		Med	10	10	10	10	10	10	10	10	10	10
		Mixed	11.4	10.7	13.9	3.53	11.9	4.53	5.59	23.7	3.74	11.13
		Neuro	9.63	10.4	16.5	9.6	2.79	21.46	6.6	16.8	0.99	5.2
		Resp	9.42	7.74	25.9	13	12.3	10.17	6.66	2.93	8.1	3.72
		Surg	11.5	10.4	12.4	17.8	1.63	14.22	8.74	5.69	3.24	4.28
		Entire set	10	10	10	10	10	10	10	10	10	10
Discharge Summary	Dept. Grouping	Med	13.1	13.8	29.7	9.95	3.44	14.24	6.21	3.43	4.46	1.6
		Med (other)	10	10	10	10	10	10	10	10	10	10
		Surg	10	10	10	10	10	10	10	10	10	10
		OBG	13.8	11.1	22.5	17.1	4.51	4.81	14.5	0.16	11.3	0.2
Average			10.3	9.98	13.7	11.6	7.7	12.22	9.73	10.1	6.89	7.077

3.2 Type 1 method

The Type 1 method labels every section in the file in favor of the section type with the highest value of the NN's predictions. It is useful when the documents need to be partitioned into sections when no specific data mining utility has yet been conceived and no section is preferred over the others. The results for the evaluation of this method showed it to be mediocre in its present form.

The discriminatory statistics across the different model groupings are presented in Appendix 3.

3.3 Type 2 method

This method is useful only when the data mining is going to be from a single kind of section and provided that other sections which might contain the same subject information do not get mistaken for the given section. The discrimination of this method was very good.

The discrimination statistics across the different groupings are in Appendix 4.

3.4 Comparison of method 1 and method 2 with actual labels

The comparison of the two methods against the actual labels was plotted as a series of cross tabulation matrices for easy elucidation of the common misclassifications made and to determine in what regions would one method be superior to the other. This was done for all document groupings. The results are presented in Appendix 5. For each grouping, the performance of method 2 is based on the best threshold selected. The discriminatory statistics for this model in each grouping are also presented.

Chapter 4

Discussion

4.1 Description of the input variable statistics

The ARD score on the input variables showed a roughly equal relevance on average across the document groups for the ten input variables. Although for a few groupings, there was equal relevance for all the variables, the pattern of variable-relevance was different across the other groups, and no general trend is defined. The section size appeared relatively more useful in several groupings and the symptom variable was least useful as indicated by the ARD score.

4.2 History and Physical Examination Documents

The History and Physical Examination documents were all generated by the Emergency Medicine department and as expected, a common structure prevailed in the document formatting on manual review of the documents and on inspecting the description statistics of the sections.

These statistics are given in Appendix 1 a. For History and Physical Examination documents the following findings are noteworthy:

1. The mean positioning and size of the sections is consistent with expectations. The section-types are arranged in the classical order expected in History and Physical

Examination documents. However, the middle sections of A, S, F and PerH sections are not ordered well among themselves and they are present inconsistently. The largest sections were PH and PE and this was consistently reflected in the statistics.

2. In the PC section, the most prominently represented content-related variable was the Finding input variable, which is expected as the SNOMED classes of symptoms and findings are similar, and this section is meant to capture the symptoms.
3. In the HOPC section none of the qualitative-feature variables have been parsed, and this is desirable as each of the content-related variables have been designed with a view to capture specific aspects of other sections.
4. The most represented qualitative variables for each section that were successfully parsed out are in the Table 4 below.

Table 4. Expected significant content variables for different sections in History and Physical Examination documents

Section-type	**Significant content variables**
Past History	Diagnoses, Procedures, Findings
Medications	Medications, Findings
Social History	Family, Findings
Family History	Diagnoses, Family, Findings
Review of Systems	Findings, Diagnoses, Symptoms
Physical Examination	Findings
Laboratory	Investigation, Procedure, Findings
Plan	Investigation, Procedure, Diagnoses, Finding, Medication

The findings variable is present in almost all the sections to variable degrees and has little apparent significance as a discriminator.

5. The variables that were prominently represented, but might not be of obvious significance are the following in Table 5 below.

6.

Table 5. Unexpected significant content variables for History and Physical Examination documents

Section-type	Significant content variables
Allergies	Diagnoses
Personal History	Diagnoses
Occupational History	Family

These can be explained since the diagnoses of hives, urticaria and allergic rhinitis are prominently featured in the A section, the family history is frequently included under the O section and negative histories of diseases featured in the PerH section instead of PH where they should have been if the traditional format was strictly adhered to.

4.3 Discharge Summary Documents

The DD sections were not as consistently formatted as in the History and Physical Examination documents. This is in part due to them being from different departments across which practiced methods can vary. The positioning of sections within documents was consistent, except for the DD section, which had a low start position, though not at the beginning of the documents, and a very high standard deviation. The high standard deviation was because some documents had this section at the beginning and some at the end of the document, but no documents had it in the middle regions.

The expected content variables of prominence for each of the sections are in Table 6 below.

Table 6. Expected significant content variables in Discharge Summary documents

Section-type	Significant content variables
History	Finding, Diagnoses, Procedure
Past History	Diagnoses
Social Personal and Family History	Finding, Family, Invest
Medications	Medications
Laboratory	Investigation, Procedure, Medication, Diagnoses, Finding
Hospital Course	Investigation, Diagnoses
Physical Examination	Findings
Discharge Diagnoses	Diagnoses

Some sections had unexpected yields of qualitative variables that were not obvious (Table 7 below). The DP section contained frequent references to counseling involving the family members, which explains this case.

Table 7. Unexpected significant content variables in Discharge Summary documents

Section-type	Significant content variables
Discharge Plan	Family
Physical Examination	Procedure
Allergies	Diagnoses

Several documents from specialty departments contained diagnostic investigations that were part of the initial examination and were labeled under PE. This is a deviation from recommended practice but is common in specialty routine. The A sections had high yields of diagnoses for similar reasons as in the History and Physical Examination documents.

4.4 Evaluation of the Type 1 Method

History and Physical Examination Documents

High Granularity Grouping

The average sensitivity (0.63) of the method for this group of documents was not sufficient to make the approach universally useful across all sections. The average PPV was even lower (0.58) as there were disproportionately greater number of false positives among the predictions.

The accuracy of predictions ranged from 0 to 95 percent. The most accurately predicted sections which are also the sections with the highest sensitivity were the PE section (0.95) followed by HOPC (0.91). These values are sufficiently high for practical utility. However, the PPV for these were significantly lower (0.72 and 0.52 respectively). It had zero success with PerH, L, F and S, probably due to the relatively small prevalence of these sections (2-5%) among the documents. The remaining sections were predicted with poor accuracy, though notably each of these also had a low prevalence in the set (<7%). The NPV was significantly higher than the sensitivity, indicating that the false positives were proportionately more than the false negatives.

A large number of sections were misclassified as PE sections, which also happened to be the most predominant section in the set. These misclassified sections were typically those expected to be located between the HOPC and PE sections.

Low Granularity Grouping

The overall accuracy as expected goes up when the granularity of section-labeling is reduced. This is because the binned groups of sections are inconsistently present, but when present show surface features with the same surface characteristics, (they are located in the same approximate region and are of comparable size to each other). The sensitivity goes up from 0.63 to 0.68, and the positive predictive value from 0.58 to 0.65, which is when compared to the high granularity grouping, proportional to the corresponding decrease in false positive predictions. The observations for the NPV verses the specificity remained similar to the values of these indices with the high granularity grouping.

An unfortunate side effect of decreasing the granularity is the complete misclassification of PC sections as PhMdAg sections or PE sections. The misclassification of many different sections as PE sections is still prominent here as it was in the high granularity grouping.

Within Diagnostic Groups

The accuracy within the diagnostic divisions keeping the department constant (Emergency Medicine) was much higher. The average sensitivity across the seven groups (Cardiology, Surgery, Respiratory, Neurology, Medicine, Gastroenterology and Mixed Medical Specialties) ranged from 0.63 to 0.86, and the specificity from 0.79 to 0.93. The

PPV ranged from 0.70 to 0.85, closely mirroring the sensitivity, which indicates that both the false positives and the false negatives were low. As observed for the undifferentiated grouping, there was a significantly higher NPV than sensitivity, consistently across the sections with the same implication that the false positive rate is higher than the false negative rate for this group.

The LP section showed high variability from group to group, being completely misclassified in the mixed medical specialties and in surgery where in both cases it was misclassified as PE. Only the PE section was consistently retrieved across the groups.

Cardiology

The high average indices are due in a large part to the more than 50 percent prevalence of PE sections and high accuracy in the HOPC section. The system performed very poorly in the PhMdAg, PFSO, LP, RS and PC sections, many of which are very useful, so the average performance statistics are too optimistic for this group.

Surgery

When restricted to this grouping, the model still performed best with only two sections- the PE and the PhMdAg sections. Since PE sections constituted 65 percent of the documents, this appeared to boost the overall performance on this set. The sensitivity was relatively low for this group (0.67), suggesting a high false negative rate. The performance on the remaining sections was poor.

Respiratory

The PE, PFSO, LP and RS sections had significantly high specificity and sensitivity. The remaining sections amounted to about 22 percent of the sections within this group, and did not yield good results.

Neurology

The successful sections were PhMdAg, PE and LP. For the PhMdAg section, the PPV was significantly lower than the sensitivity, suggesting that the false positive rate is high. The RS, HOPC, PFSO and PC sections showed poor results.

Medicine

In this group, the PhMdAg, PE and HOPC were the highest scoring sections. The RS, PFSO and PC scored low.

Gastroenterology

The most successful sections were PhMdAg, PE, LP, and HOPC. The remaining sections comprised 14 percent of the documentation and did not contribute much to the overall figures.

Other Medical Specialties

The PE section and HOPC were the only sections identified with reasonable accuracy. As expected, because this grouping does not differentiate within diagnoses groups, the performance here is lower than in the other sets.

Discharge Summary Documents

Entire Set

The sensitivity and PPV for the entire set were very low (0.49 and 0.49). The NPV and the specificity were low compared to the History and Physical Examination documents. The sections with a good sensitivity and specificity were DP, H and HC. L, DD, PE, PH, A, M and SPF sections had poor sensitivities ranging from zero to 0.52 but high specificities from 0.93 to 0.97.

The most consistently identified section was the H section. L, HC and DP sections were also obtained with high accuracy. The DD section was frequently misclassified as DP. The misclassification of DP as DD also occurred significantly, although not as frequently. Unlike in the History and Physical Examination documents, the PE section could only be successfully extracted half the time.

Within Departmental Sub Groups

The overall performance of these groups did not differ much from the undifferentiated group of Discharge Summary documents above. The unexpected best performer was the mixed group of Other Medical Specialties which of the lot is expected to have the worst performance if the hypothesis is true that departmental practices cause variations on the form of documents.

In the General Medicine group, the HC section was always confused with the DP section. A large proportion of the DP sections were misclassified as PE. Most DD sections were classified as H sections. For the grouped medical specialty departments, there was a consistent misclassification of HC as DP. For the most part, the other sections were properly classified except when the testing sample size of sections in the group was small, when a tendency to misclassify as the more prevalent DP section was noted. The performance for the Obstetrics and Gynecology group was generally good. Misclassifications there tended to be more frequently as DP or DD, which is acceptable. In the surgical group, confusion between DP and DD was also prominent. Many of the other sections were confused as DD or DP. PE, L and PH were often mislabeled as H sections.

General Medicine

The overall performance for this group was lower than that for the entire group. The sensitivity was only 0.29 and the PPV was 0.41, with a high false negative rate. The two

sections with high sensitivity and specificity pairs were the PE and SPF sections. However, these had very low PPV values of about 10 percent. Hence, the method has not worked for any sections in this group.

Surgery

The H and DP section had high specificity and sensitivity. Both had low PPV values of around 0.5 but were still good enough for use. These two groups accounted for around half of the sections available.

Obstetrics and Gynecology

The average sensitivity and specificity were similar to the whole group, but three sections provided useful results-DD, HC and PE. DD had a relatively low PPV, because of a high false positive rate.

Other Medical Specialty Departments

The sections that were predicted well for this group included L, DP and DD. The DP section had a low PPV of around 0.5.

4.5 Evaluation of the Type 2 Method

History and Physical Examination Documents

High Granularity Grouping

The sections with high c-indices were PC, HOPC, PH and PE. The sections M, A, PerH, P and RS had moderately high c-indices. It did not perform well on L, S, O and F sections. The calibration scores were excellent, indicating that the tool was definitely sensitive to the nature of the sections.

With this group, there was a tendency to classify the same section as different types of sections when each of those are looked for. No sections were identified as a particular label exclusively. There was a tendency for a large proportion of PC sections to be classified as PH, A, PerH and O sections. HOPC sections were equally likely to be picked up as PerH or O and to a lesser extent as M sections. PH sections were mistaken as A, PerH, PC and O sections. PE sections were often identified as OH. LP sections were consistently misclassified as PerH. The worst selectivity was for F, S and PerH. Almost all sections ran a high risk of being identified as PerH sections.

Low Granularity Grouping

Compared to the high granularity grouping, the PE and HOPC sections remained at the same accuracy. The c-index on the PC and RS sections dropped significantly. The PhMdAg, PFSO and LP sections improved in c-indices significantly. The calibration indices were very good for this grouping.

Only HOPC and PE sections had a high PPV.

HOPC sections were consistently picked up as PC sections. However, the rest were often picked up as two or more other kinds of sections.

Within Diagnostic Sub Groups

The PPV was consistently high for PE and HOPC sections. The remaining sections showed variation in the PPV. Performance for the Obstetrics and Gynecology group was good in this regard.

Cardiology

In this grouping, the tool performed well on all sections except on the PFSO section and with excellent calibration indices throughout.

Gastroenterology

In this group, the performance across all sections was excellent. However, the calibration indices for all except the PE section were unsatisfactory, possibly due to the small number of document samples available.

General Medicine

In this grouping, the performances were excellent in both discrimination and calibration except for the RS section. Here, discrimination ability was non-existent (c-index was only 0.52), but the calibration was outstanding. This can only mean that as far as this section is concerned, there is too much variability even though the input variables are sensitive to the section type.

Surgery

The discrimination indices were good, but the poor calibration was probably due to the low number of cases. The only section in which the discrimination was poor was the PFSO history section.

Respiratory

The discrimination and calibration for this group were excellent across all sections.

Neurology

In this group, the discrimination performance across all section types was impressive except for the PFSO section, where both discrimination and calibration were not satisfactory.

Mixed Medical Specialties

For this grouping, the RS section showed poor discrimination even though the calibration was satisfactory. The remaining sections were satisfactorily identified.

Discharge Summary Documents

Entire set

The discrimination across the entire set was very satisfactory. The c-indices ranged from 0.75 to 0.96. Except for the H section, where the calibration was poor, the remaining ones showed excellent calibration.

However, there was a strong tendency to mislabel sections across all section types. Consequently, the positive predictive value of HC was the only sufficiently high one.

Department groupings

When grouped by department types, the Discharge Summary documents had better results. Unfortunately, the false positive rate remained high.

Surgery

For the sections H, PH, HC, DP, PE, the discrimination indices were excellent. However, the calibration indices for DP were not satisfactory. The performance on DD and L sections were poor.

The sections with high PPV were H, PH, HC and PE sections.

General Medicine

The discrimination statistics for the H, PH, SPF, L, HC, PE and A sections were satisfactory. However, the calibration statistics for HC were poor. The method failed to be useful on the sections DD and DP.

Only L and HC had high PPV values.

Obstetrics and Gynecology

The sections L, HC, PE and A had excellent discrimination and calibration indices. The sections H, DD and DP had good discrimination indices but poor calibration indices. The performance was poor on PH sections.

HC was the only section with a high PPV value.

Grouped Medicine Specialties

The discrimination statistics across all sections were excellent. The calibration for H, DP and DD were low.

The PPV values for all sections were good.

Chapter 5

Conclusion

5.1 Implications of Results and Current Utility

The most important result from this work is the demonstration that an underlying pattern exists in electronic documents by virtue of the similarities in physicians training. This pattern can now be utilized to automate the identification of context using machine-learning tools. There are also considerations discussed below, to be taken into account before implementing such tools.

The most desirable way of classifying context in a document is by method 1, where each section will be labeled only once. This method worked best when the documents are grouped according to the principle diagnoses, showing that the content of the sections is what affects performance more than the department that generated the document or the granularity of the section grouping. However, any formalized method must incorporate grouping the documents by type and department of origin. Fixing the granularity at an appropriate level also contributed to the performance appreciably.

Method 2 is of limited use. When a user is interested in only a single context from the document, and when that section is not misclassified as one where the meaning of the concept extracted can be confounded, this method is useful. For example, if it is of interest to find cases of a family history of diabetes, then the section of interest is the F

section. However, PH is another section where diabetes could be mentioned. If PH is often confused with F, then the method cannot be used here.

Certain kinds of apparent failure of classification may however be acceptable. Misclassifying PC sections as HOPC are in many cases due to the physicians combining both into one section. There are a few cases where this occurs almost uniformly in the set: between DP and DD sections and between S, O, F and PerH sections. It is essential to determine the amount of differentiability of section types prior to judging the discerning power offered by the tool. In this case, significant inherent confusion existed between the following groups of sections in Table 8.

Table 8. Each row in the table shows sections that are frequently combined or mixed up in documents by physicians

Frequently Mixed Up or Combined Groups (in the documents)
Document type: History and Physical Examination
Presenting Complaint, History of Presenting Complaint
Social, Personal, Family, Occupational
History of Presenting Complaint, Review of Systems
Laboratory, Conclusion and Plan
Document type: Discharge Summary
Discharge Diagnoses, Discharge Plan
History, Past History, Social Personal Family, Physical Examination
Laboratory, Hospital Course

With the understanding of the limitations this intrinsic confusion poses, useful extraction might be carried out within those limits. However, the methods used did confuse sections outside of these groupings in several instances. The methods are not applicable in these cases as mentioned in the discussion, where they do not correspond to the mappings in the table above.

The results also show that the variation in the accuracy of extraction of a particular section varies from group to group quite arbitrarily, and hence extrapolation of its performance across groupings cannot be made.

The sample size used for this study was very small. Several section-types existed only in small numbers. This becomes particularly significant when groupings by department type or diagnoses were made. The results in these sections appeared to be better than when the cases were not grouped, but the actual gain in accuracy cannot be established with certainty unless a larger sample of documents is obtained. It is also possible that accuracy will improve when more cases are used for learning.

The most useful sections from a practical standpoint from History and Physical Examination documents are PC, A, PH, PE and P. In the case of Discharge Summary documents, the interesting sections are DD, DP, L, M and A. Acceptable performances concerning these are in Table 9 below:

Table 9. Sections of practical value that were successfully identified within groupings for both types of documents

Method 1	
Document Type: Discharge Summary	
Grouping	**Section Name**
Entire Set	None
By Department:	
General Medicine	Laboratory and Plan
Obstetrics and Gynecology	Hospital Course, Physical Examination, Laboratory and Plan
Surgery	
Other Medical Specialties	Laboratory, Discharge Diagnoses
Document Type: History and Physical	
Entire Set	Physical Examination, Laboratory and Plan History of Presenting Complaints
By Diagnoses grouping	
Cardiology	Physical Examination, Laboratory and Plan, History of Presenting Complaints
Surgery	Physical Examination
Respiratory	Physical Examination, Laboratory and Plan
Neurology	Physical Examination, Laboratory and Plan
Medicine	Physical Examination, Laboratory and Plan, History of Presenting Complaints
Gastroenterology	Physical Examination, Laboratory and Plan, History of Presenting Complaints
Mixed Medical Specialties	Physical Examination, History of Physical Examination
Method 2	
Document Type: Discharge Summary	
Entire Set	None
By Department:	
General Medicine	Laboratory
Obstetrics and Gynecology	None
Surgery	Physical Examination
Other Medical Specialties	All sections
Document Type: History and Physical	
Entire Set	Physical Examination, Laboratory and Plan, History of Presenting Complaints
By Diagnoses grouping	
Cardiology	Physical Examination, History of Presenting Complaints
Surgery	Physical Examination, Laboratory and Plan, History of Presenting Complaints

Respiratory	Physical Examination, Laboratory and Plan, History of Presenting Complaints
Neurology	Physical Examination, Past History, Medicines, Allergies, History of Presenting Complaints
Medicine	Physical Examination, Laboratory and Plan, History of Presenting Complaints
Gastroenterology	Physical Examination, Laboratory and Plan, History of Presenting Complaints
Mixed Medical Specialties	Physical Examination, History of Presenting Complaints

HOPC is included in this table because presenting complaints are often binned with their history as mentioned above, for this sample. Consistent and desirable performance seems localized to PE, L and P and HOPC sections except in a few exceptions. Thus in the present form, the methods have limited utility, except in the Discharge Summary departmental group- Other Medical Specialties. It is possible that if sufficient cases are found within a narrow grouping that considers document type, department and diagnoses, the classification of sections will show a much better performance.

Using the tool to automatically exclude groups of sections that are unlikely to produce high yields for specific data-mining purposes, prior to some other method of section-identification (like manual review) is currently possible. This is dependent on a high NPV, which is very consistent across all the sets. This would reduce the numbers of documents to be perused by a factor of ten on the average.

In cases where the sensitivity is low but the PPV is sufficiently high (PPV > 0.7 and PPV> Sensitivity), the methods could be used for cohort selection in research studies, as the actual number of missed cases (false negatives) need not necessarily be a

consideration. This is applicable to the following sections in Table 10. No cases for Method 2 showed up with this condition.

Table 10. Sections with PPV value higher than Sensitivity, useful for cohort selection

Cohort Selection	
Method 1	
Document type: Discharge Summary	
Grouping	**Section**
General Medicine	Laboratory
Other Medical Subspecialties	History, Past History, Discharge Diagnoses
OBG	History, Laboratory
Surgery	Hospital Course
Document type: History and Physical Examination	
Entire Set	History of Presenting Complaints, Laboratory and Plan
Diagnoses groups:	
Cardiology	Laboratory and Plan
Respiratory	Laboratory and Plan, Personal Family Social Occupational
Neurology	Personal Family Social Occupational
Medicine	History of Presenting Complaints, Laboratory and Plan
Gastroenterology	Laboratory and Plan
Mixed Medical Specialties	Personal Family Social Occupational

5.2 Future Directions

As already mentioned, the sample size was too small to draw reliable conclusions for the study. Furthermore, to be trusted, the results for the smaller sub-groups need to be reproduced on larger sets. It is also possible that when the study is extended beyond the range of a single hospital's records, the results will vary in unpredictable ways. The unpredictability of results when the grouping changes is supportive of the tenet that if the process of section-identification is to be automated, all relevant groupings need to be

identified and considered and sufficient sample size ensured before any learnt model can be trusted.

Another direction for future work involves narrowing down the set of input variables necessary for section-identification. This would involve identifying variables that are correlated with each other, and then choosing all but one of these to keep in the model. The advantage of reducing the input variables is twofold. First, the processes needed to produce the input set can be reduced and second, the size of the input to the NN and correspondingly the learning time can be reduced. This may or may not be possible for different document groupings.

Combining the method with the identification of section-labels through Regex parsing of files for these, is a yet unexplored aspect that may enhance the success of section-labeling. Now, though section labels were available in most of files, the attempt was to be able to identify context without relying on label patterns, as several documents do not have them and naming was inconsistent.

The method performed better with respect to certain sections. These were usually the HOPC and PE. In History and Physical Examination documents, these separate the files into three distinct regions: A presenting complaint region, a group of middle sections and a concluding group of sections like L and P sections. This suggests the possibility that if the NN had used the knowledge of those sections positions and labels when attempting to predict unknown sections from the three groups, prediction success might have improved.

This could be implemented in two ways. In the first, a NN attempts to identify the two sections in a document. Subsequently a second NN feeds a variable derived from this information into the prediction of unknown sections (such as the positions of the PE or HOPC sections). Alternately, a second technique can utilize information about adjacent sections. This would involve a more complex set of input variables, where in addition to the ones already present, each section's input contains the variable values of sections adjacent to it (or even further removed, if the results are promising). In effect, the identities of adjacent sections would influence the prediction of sections.

A future implementation venture would attempt to automate the extraction of input variable details from document sections and run the section-identification tool within database-derived groupings of document type, department and diagnoses grouping in a single step. Much improvement in the accuracy and a reliability of the methods needs to be established before such a product can be realized.

References

1. Guyer, S., *Clinical data repositories: an overview.* Nurs Case Manag, 2000. **5**(1): p. 2-9; quiz 10-2.
2. Alexander, S., T. Conner, and T. Slaughter, *Overview of inpatient coding.* Am J Health Syst Pharm, 2003. **60**(21 Suppl 6): p. S11-4.
3. Malatestinic, W., et al., *Components of Medicare reimbursement.* Am J Health Syst Pharm, 2003. **60**(21 Suppl 6): p. S3-7.
4. Paul C. Tang, C.J.M., *Computer-Based Patient-Record Systems*, in *Medical Informatics*, L.E.P. Edward H. Shortliffe, Gio Wiederhold Lawrence M. Fagan, Editor. 2001, Springer, New York. p. 327-357.
5. Bensadoun, H., *[The Medical Information Systems Project clinical coding and surgeons: why should surgeons code and how?].* J Chir (Paris), 2001. **138**(1): p. 6-14.
6. Laerum, H., G. Ellingsen, and A. Faxvaag, *Doctors' use of electronic medical records systems in hospitals: cross sectional survey.* Bmj, 2001. **323**(7325): p. 1344-8.
7. Sullivan, F. and E. Mitchell, *Has general practitioner computing made a difference to patient care? A systematic review of published reports.* Bmj, 1995. **311**(7009): p. 848-52.
8. Wyatt, J.C., *Clinical data systems, part 3: development and evaluation.* Lancet, 1994. **344**(8938): p. 1682-88.
9. Majoros, W.H., G.M. Subramanian, and M.D. Yandell, *Identification of key concepts in biomedical literature using a modified Markov heuristic.* Bioinformatics, 2003. **19**(3): p. 402-7.
10. Maas, M.L., C. Delaney, and D. Huber, *Contextual variables and assessment of the outcome effects of nursing interventions.* Outcomes Manag Nurs Pract, 1999. **3**(1): p. 4-6.
11. Utz, S.W., *Computerized documentation of case management. From diagnosis to outcomes.* Nurs Case Manag, 1998. **3**(6): p. 247-54.
12. Voytovich, A.E., *Reduction of medical verbiage: fewer words, more meaning.* Ann Intern Med, 1999. **131**(2): p. 146-7.
13. Swash, M., *Doctor and Patient*, in *Hutchison's Clinical Methods*, M. Swash, Editor. 1996, W. B. Saunders Company Ltd. p. 4-26.
14. Hahn U, R.M., Schulz S., *Discourse structures in medical reports--watch out! The generation of referentially coherent and valid text knowledge bases in the MEDSYNDIKATE system.* International Journal of Medical Informatics, 1999. **53**(1): p. 1-28.
15. Heinze, D.T., M.L. Morsch, and J. Holbrook, *Mining free-text medical records.* Proc AMIA Symp, 2001: p. 254-8.
16. Hripcsak, G., et al., *Unlocking clinical data from narrative reports: a study of natural language processing.* Ann Intern Med, 1995. **122**(9): p. 681-8.
17. Winston, P.H., *Simulated Neural Nets*, in *Artificial Intelligence*. 1993, Addison-Wesley Publishing Company. p. 443-448.
18. Winston, P.H., *Back-Propagation Characteristics*, in *Artificial Intelligence*. 1993, Addison-Wesley Publishing Company. p. 457-468.

19. Philip H. Goodman, D.B.R., *Nevada backPropagation version 3, User Manual; Artificial Neural Network Simulator for Statistical Inference*. 1996, University of Nevada. p. 129-162.
20. D.K. Owens, H.C.S., *Measurement of the Operating Characteristics of Diagnostic Tests*, in *Medical Informatics*, L.E.P. Edward H. Shortliffe, Gio Wiederhold Lawrence M. Fagan, Editor. 2001, Springer, New York. p. 88-90.
21. David W. Hosmer, S.L., *The Hosmer-Lemeshow Tests*, in *Applied Logistic Regression*, N.I.F. Noel A. C. Cressie, Ian M. Johnstone, J.B. Kadane, David W. Scott, Bernard W. Silverman, Adrian F. M. Smith, Jozef L. Teugels, Editor. 2000, John Wiley & Sons, Inc.: New York. p. 147-149.

Appendices

Appendix 1 a

The Regular expression set used for parsing sections from History and Physical Examination documents:

```
([A-Z ]+:)|([A-Z ]+)|([A-Z]+:)|(FOLLOW)|(LABORATORY)|(LABORATORY,)|
(HOSPITAL)|(CONDITION)|(DISPOSITION)|(PHYSICAL)|(CLINICAL)|(PRINCI
PAL)|(ALLERG)|(DISCHARGE)|(ADMITTING)|(SURGERIES)|(ASSESSMENT)|(
MEDICATION)|(REVIEW OF)|(CHIEF)|(HISTORY OF)|(PAST
MEDICAL)|(FAMILY)|(DIAGNOSIS)|(RECOMMENDATION)|(PAST
SURGICAL)|(CURRENT)|(SOCIAL)| (PLAN)|(INDICATION)|(VITAL)|(REVIEW
OFSYSTEMS)|
(PERSONAL)|(IMPRESSION)|(REASON)|(REFERRING)|(PREOPERATIVE)|(OB
STETRIC)|(CHILDHOOD)|(ADULT)|
(OB)|(HABITS)|(HOSPITALIZATION)|(INITIAL)|(FINAL DIAGNOSIS)|([A-Z
]+:)
```

The Regular expression set used for parsing sections from Discharge Summary documents:

```
[A-Z ]+:)|\=(?=[A-Z ]+:)|[\n\r](?=[A-Z ]+:)|
\n(?=FOLLOW)|\n(?=LABORATORY)|\n(?=LABORATORY,)|\n(?=HOSPITAL)|
\n(?=CONDITION)|\n(?=DISPOSITION)|\n(?=PHYSICAL)|\n(?=CLINICAL)|
\n(?=PRINCIPAL)|\n(?=ALLERG)|\n(?=DISCHARGE)|\n(?=ADMITTING)|
\n(?=SURGERIES)|\n(?=ASSESSMENT)|\n(?=MEDICATION)|\n(?=REVIEW OF)|
\n(?=CHIEF)|\n(?=HISTORY OF)|\n(?=PAST MEDICAL)|\n(?=FAMILY)|
\n(?=DIAGNOSIS)|\n(?=RECOMMENDATION)|\n(?=PAST SURGICAL)|
\n(?=CURRENT)|\n(?=SOCIAL)|\n(?=PLAN)|\n(?=INDICATION) |
\n(?=VITAL)|\n(?=REVIEW OF
SYSTEMS)|\n(?=PERSONAL)|\n(?=IMPRESSION)|
\n(?=REASON)|\n(?=REFERRING)|\n(?=PREOPERATIVE)|\n(?=OBSTETRIC)|
\n(?=CHILDHOOD)|\n(?=ADULT)|\n(?=OB)|\n(?=HABITS)|\n(?=HOSPITALIZAT
ION)|
\n(?=INITIAL)|(?=FINAL DIAGNOSIS)|(\=+\=)
```

Appendix 1 b

This is a presentation of the descriptive variables of each section within the two kinds of documents.
Note: Valid N is the number of sections of the type considered.

History Document Descriptive Statistics

1. Presenting Complaint

	N	Minimum	Maximum	Mean	Std. Deviation
Start	72	0	661	35.68	84.684
Size	72	18	852	72.57	115.994
Med	72	0	2	.03	.236
Proced	72	0	2	.21	.442
Invest	72	0	2	.04	.262
Diag	72	0	6	.69	1.182
Finding	72	0	7	1.15	1.096
Sympt	72	0	3	.31	.597
Family	72	0	3	.07	.387
Valid N	72				

2. History of Presenting Complaint

	N	Minimum	Maximum	Mean	Std. Deviation
Start	114	0	3329	135.32	381.440
Size	114	31	2522	819.23	479.398
Med	114	0	0	.00	.000
Proced	114	0	0	.00	.000
Invest	114	0	0	.00	.000
Diag	114	0	0	.00	.000
Finding	114	0	0	.00	.000
Sympt	114	0	0	.00	.000
Family	114	0	0	.00	.000
Valid N	114				

3. Past History

	N	Minimum	Maximum	Mean	Std. Deviation
Start	110	304	3069	1058.71	535.154
Size	110	21	1696	279.92	252.076
Med	110	0	4	.25	.612
Proced	110	0	9	1.19	1.594
Invest	110	0	4	.30	.614
Diag	110	0	13	2.75	2.717
Finding	110	0	8	.82	1.342
Sympt	110	0	2	.11	.367
Family	110	0	3	.43	.748
Valid N	110				

4. Medications

	N	Minimum	Maximum	Mean	Std. Deviation
Start	84	365	3498	1346.26	618.388
Size	84	18	1942	241.19	312.121
Med	84	0	9	2.06	2.262
Proced	84	0	3	.30	.655
Invest	84	0	5	.15	.668
Diag	84	0	6	.69	1.280
Finding	84	0	19	1.25	3.150
Sympt	84	0	3	.20	.576
Family	84	0	6	.21	.945
Valid N	84				

5. Allergies

	N	Minimum	Maximum	Mean	Std. Deviation
Start	73	388	4404	1414.16	694.557
Size	73	17	391	52.41	62.039
Med	73	0	2	.16	.441
Proced	73	0	4	.10	.505
Invest	73	0	1	.03	.164
Diag	73	0	3	1.03	.799
Finding	73	0	1	.23	.426
Sympt	73	0	1	.01	.117
Family	73	0	3	.12	.470
Valid N (listwise)	73				

6. Family History

	N	Minimum	Maximum	Mean	Std. Deviation
Start	77	602	3686	1524.12	648.826
Size	77	30	476	98.42	90.640
Med	77	0	1	.08	.270
Proced	77	0	1	.10	.307
Invest	77	0	1	.25	.434
Diag	77	0	9	1.06	2.022
Finding	77	0	3	.39	.652
Sympt	77	0	0	.00	.000
Family	77	0	7	.75	1.349
Valid N (listwise)	77				

7. Personal History

	N	Minimum	Maximum	Mean	Std. Deviation
Start	16	891	2324	1523.69	570.011
Size	16	19	736	184.00	213.023
Med	16	0	5	.31	1.250
Proced	16	0	0	.00	.000
Invest	16	0	1	.25	.447
Diag	16	0	5	1.06	1.237
Finding	16	0	9	1.31	2.272
Sympt	16	0	3	.25	.775
Family	16	0	2	.44	.727
Valid N	16				

8. Social History

	N	Minimum	Maximum	Mean	Std. Deviation
Start	80	627	4162	1479.04	662.926
Size	80	31	642	130.10	112.865
Med	80	0	1	.09	.284
Proced	80	0	3	.24	.621
Invest	80	0	2	.24	.509
Diag	80	0	2	.36	.680
Finding	80	0	5	1.10	1.249
Sympt	80	0	1	.01	.112
Family	80	0	4	.68	.925
Valid N	80				

9. Occupational History

	N	Minimum	Maximum	Mean	Std. Deviation
Start	3	1018	2521	1884.33	777.375
Size	3	36	541	246.33	262.850
Med	3	0	0	.00	.000
Proced	3	0	0	.00	.000
Invest	3	0	0	.00	.000
Diag	3	0	0	.00	.000
Finding	3	0	2	.67	1.155
Sympt	3	0	0	.00	.000
Family	3	1	1	1.00	.000
Valid N	3				

10. Review of Systems

	N	Minimum	Maximum	Mean	Std. Deviation
Start	83	494	4746	1701.52	777.095
Size	83	28	1691	263.02	282.893
Med	83	0	3	.18	.521
Proced	83	0	4	.30	.694
Invest	83	0	2	.29	.482
Diag	83	0	11	1.33	1.945
Finding	83	0	27	3.64	4.560
Sympt	83	0	6	.86	1.515
Family	83	0	11	.55	1.382
Valid N	83				

11. Physical Examination

	N	Minimum	Maximum	Mean	Std. Deviation
Start	913	543	7974	2412.10	1056.649
Size	913	13	1117	101.76	110.312
Med	913	0	2	.02	.132
Proced	913	0	4	.33	.600
Invest	913	0	11	.11	.487
Diag	913	0	8	.28	.621
Finding	913	0	10	.85	1.249
Sympt	913	0	4	.10	.355
Family	913	0	3	.11	.338
Valid N	913				

12. Laboratory

	N	Minimum	Maximum	Mean	Std. Deviation
Start	63	1081	5538	2792.32	1028.779
Size	63	15	850	333.65	190.619
Med	63	0	6	1.02	1.540
Proced	63	0	5	1.11	1.321
Invest	63	0	13	3.70	3.295
Diag	63	0	4	.76	.875
Finding	63	0	4	1.05	1.156
Sympt	63	0	1	.06	.246
Family	63	0	3	.17	.493
Valid N	63				

13. Plan

	N	Minimum	Maximum	Mean	Std. Deviation
Start	111	8	6201	2846.86	1164.984
Size	111	25	1746	424.85	375.943
Med	111	0	8	.78	1.384
Proced	111	0	7	1.42	1.832
Invest	111	0	9	1.50	1.808
Diag	111	0	12	2.75	2.542
Finding	111	0	17	1.49	2.408
Sympt	111	0	3	.29	.578
Family	111	0	4	.79	1.153
Valid N	111				

Discharge Summary Descriptive Statistics

1. History

	N	Minimum	Maximum	Mean	Std. Deviation
Start	93	0	3526	298.95	534.722
Size	93	0	3640	424.42	511.167
Dx	93	0	11	1.14	1.965
Family	93	0	6	.87	1.287
Finding	93	0	8	1.71	2.109
Invest	93	0	27	.58	2.879
Medic	93	0	4	.12	.486
Proced	93	0	20	1.02	2.275
Sympt	93	0	4	.37	.791
Valid N	93				

2. Past Medical Or Surgical History

	N	Minimum	Maximum	Mean	Std. Deviation
Start	27	142	2512	764.30	550.920
Size	27	0	500	156.33	125.491
Dx	27	0	9	1.93	2.827
Family	27	0	2	.33	.555
Finding	27	0	3	.56	.934
Invest	27	0	1	.11	.320
Medic	27	0	1	.04	.192
Proced	27	0	5	.52	1.252
Sympt	27	0	2	.19	.557
Valid N	27				

3. Social Personal Family

	N	Minimum	Maximum	Mean	Std. Deviation
Start	15	522	2365	1095.20	543.087
Size	15	15	126	68.27	31.883
Dx	15	0	1	.20	.414
Family	15	0	1	.33	.488
Finding	15	0	2	.40	.632
Invest	15	0	1	.33	.488
Medic	15	0	1	.07	.258
Proced	15	0	1	.07	.258
Sympt	15	0	0	.00	.000
Valid N	15				

4. Medications

	N	Minimum	Maximum	Mean	Std. Deviation
Start	8	195	2108	1073.63	561.628
Size	8	56	265	155.75	73.669
Dx	8	0	2	.38	.744
Family	8	0	0	.00	.000
Finding	8	0	4	.50	1.414
Invest	8	0	1	.50	.535
Medic	8	0	5	2.13	2.232
Proced	8	0	1	.13	.354
Sympt	8	0	0	.00	.000
Valid N	8				

5. Laboratory

	N	Minimum	Maximum	Mean	Std. Deviation
Start	44	16	5278	1402.70	1059.066
Size	44	0	2969	659.39	677.380
Dx	44	0	12	2.05	2.477
Family	44	0	4	.48	1.089
Finding	44	0	9	1.45	1.731
Invest	44	0	32	8.20	8.894
Medic	44	0	13	2.43	3.015
Proced	44	0	17	3.05	3.543
Sympt	44	0	1	.02	.151
Valid N	44				

6. Hospital Course

	N	Minimum	Maximum	Mean	Std. Deviation
Start	73	61	6851	1359.62	1153.742
Size	73	137	4077	950.04	643.063
Dx	73	0	4	.23	.717
Family	73	0	3	.04	.351
Finding	73	0	5	.16	.646
Invest	73	0	10	.26	1.225
Medic	73	0	2	.04	.260
Proced	73	0	6	.15	.739
Sympt	73	0	4	.05	.468
Valid N	73				

7. Discharge Diagnosis

	N	Minimum	Maximum	Mean	Std. Deviation
Start	107	0	5117	689.21	1075.878
Size	107	18	945	149.07	141.003
Dx	107	0	4	.13	.631
Family	107	0	1	.01	.097
Finding	107	0	2	.06	.302
Invest	107	0	0	.00	.000
Medic	107	0	0	.00	.000
Proced	107	0	4	.10	.475
Sympt	107	0	0	.00	.000
Valid N	107				

8. Discharge Plan

	N	Minimum	Maximum	Mean	Std. Deviation
Start	154	43	6660	1784.90	1273.736
Size	154	15	1197	193.95	227.364
Dx	154	0	1	.02	.139
Family	154	0	7	.12	.656
Finding	154	0	5	.10	.538
Invest	154	0	9	.09	.795
Medic	154	0	3	.06	.328
Proced	154	0	6	.10	.654
Sympt	154	0	3	.02	.242
Valid N (listwise)	154				

9. Physical Examination

	N	Minimum	Maximum	Mean	Std. Deviation
Start	63	141	5789	1768.75	1091.965
Size	63	0	1073	202.57	232.368
Dx	63	0	3	.43	.689
Family	63	0	3	.24	.560
Finding	63	0	7	1.13	1.601
Invest	63	0	15	.79	2.824
Medic	63	0	8	.41	1.552
Proced	63	0	5	1.14	1.293
Sympt	63	0	1	.06	.246
Valid N (listwise)	63				

10. Allergies

	N	Minimum	Maximum	Mean	Std. Deviation
Start	16	299	2224	932.62	548.576
Size	16	17	344	68.13	87.908
Dx	16	0	3	1.13	1.025
Family	16	0	1	.13	.342
Finding	16	0	2	.31	.602
Invest	16	0	1	.06	.250
Medic	16	0	1	.06	.250
Proced	16	0	3	.38	.885
Sympt	16	0	0	.00	.000
Valid N (listwise)	16				

Appendix 2:

StandardizeInputs 1

Sets all the prediction values to the same scale.

ShuffleData YES

Selects rows randomly from the sets for training or testing.

CalccIndex YES

Calculates the C-index for each set.

ScoreThreshold 0.5

Sets a prediction threshold minimum of 0.5 (from the range 0 to 1)

1ofN YES

Normalizes the predictions so that all predictions sum to 1 for a given section.
OutputUnitType 3

This senses the units of the output variable automatically (in this case as a dichotomous variable).

WeightRange 0.001

Initial weights are set from a range between -0.001 and +0.001

TrainCriterion 3

This is the error, residual, loss or objective function setting. Since all the dependent variables are dichotomous, the cross entropy criterion is used.

WeightDecay -0.001

It is the fixed fraction of the weights magnitude subtracted at each weight update, to prevent the weight magnitude from getting excessive unless the data reinforces the growth.

OptimizeMethod 1

Uses the gradient descent function.

LearnRate 0.01

During optimization, the weights are changed in proportion to the gradient of error criterion. The factor of proportionality is the LearnRate.

Momentum 0.0

It is the fraction of the previous change in weight to be added to the next update.

AutoTrain YES

In this setting, a set of learning runs are made and a mean error per case on the training run for which the holdout subset showed the least error is determined. The training is then restarted from scratch on the test set until that error minimum is attained.

MinEpochs 50

This is the minimum number of epochs that training has to perform for before a model is assumed. It is a guard against local minima.

BeyondBestEpoch 5.5

This forces the NN to train beyond a factor of 5.5 times the best epoch in the training set, as a precaution against attaining a local minimum.

Appendix 3

Each table represents a document grouping. The sub tables display the discriminatory indices and the accuracy index for each section. At the end of the table, the average value for the indices across the sections is displayed. The figure in brackets next to the PPV value in the average PPV field is the number of true positives for the group.

History and Physical Examination Documents

Entire set- High Granularity Grouping

Section Name:	PH
Counted:	29
Accuracy:	51.7241379
PPV:	0.41666667
NPV:	0.96766744
Sensitivity:	0.51724138
Specificity:	0.95227273
Total Sections:	448
Percentage:	6.47321429

Section Name:	S
Counted:	22
Accuracy:	0
PPV:	0
NPV:	0.95089286
Sensitivity:	0
Specificity:	1
Total Sections:	448
Percentage:	4.91071429

Section Name:	PE
Counted:	226
Accuracy:	95.1327434
PPV:	0.72635135
NPV:	0.9527897
Sensitivity:	0.95132743
Specificity:	0.73267327
Total Sections:	448
Percentage:	50.4464286

Section Name:	P
Counted:	25
Accuracy:	20
PPV:	0.27777778
NPV:	0.95485327
Sensitivity:	0.2
Specificity:	0.97018349
Total Sections:	448
Percentage:	5.58035714

Section Name:	M
Counted:	22
Accuracy:	18.1818182
PPV:	0.57142857
NPV:	0.95945946
Sensitivity:	0.18181818
Specificity:	0.99300699
Total Sections:	448
Percentage:	4.91071429

Section Name:	F
Counted:	19
Accuracy:	0
PPV:	0
NPV:	0.95758929
Sensitivity:	0
Specificity:	1
Total Sections:	448
Percentage:	4.24107143

Section Name:	RS
Counted:	19
Accuracy:	26.3157895
PPV:	0.27777778
NPV:	0.96839729
Sensitivity:	0.26315789

Section Name:	HOPC
Counted:	23
Accuracy:	91.3043478
PPV:	0.525
NPV:	0.99531616
Sensitivity:	0.91304348

Specificity:	0.97058824
Total Sections:	448
Percentage:	4.24107143

Specificity:	0.95720721
Total Sections:	448
Percentage:	5.13392857

Section Name:	PC
Counted:	21
Accuracy:	76.1904762
PPV:	0.76190476
NPV:	0.98842593
Sensitivity:	0.76190476
Specificity:	0.98842593
Total Sections:	448
Percentage:	4.6875

Section Name:	L
Counted:	20
Accuracy:	0
PPV:	0
NPV:	0.95535714
Sensitivity:	0
Specificity:	0.997669
Total Sections:	448
Percentage:	4.46428571

Section Name:	PerH
Counted:	9
Accuracy:	0
PPV:	0
NPV:	0.97991071
Sensitivity:	0
Specificity:	1
Total Sections:	448
Percentage:	2.00892857

Section Name:	A
Counted:	13
Accuracy:	23.0769231
PPV:	0.25
NPV:	0.97752809
Sensitivity:	0.23076923
Specificity:	0.97972973
Total Sections:	448
Percentage:	2.90178571

Section Name:	O
Counted:	1
Accuracy:	0
PPV:	0
NPV:	0.99776786
Sensitivity:	0
Specificity:	1
Total Sections:	448
Percentage:	0.22321429

Total Sections:	448
Average Sensitivity:	0.63392857
Average Specificity:	0.85759947
Average PPV(398):	0.58381254
Average NPV:	0.96242599
Average Accuracy:	63.3928571

Entire Set- Low Granularity			
Section Name:	PhMdAg	Section Name:	PFSO
Counted:	132	Counted:	91
Accuracy:	53.78787879	Accuracy:	31.86813187
PPV:	0.550387597	PPV:	0.475409836
NPV:	0.92623942	NPV:	0.928653625
Sensitivity:	0.537878788	Sensitivity:	0.318681319
Specificity:	0.92961165	Specificity:	0.961859356
Total_Sections:	898	Total_Sections:	898
Percentage:	14.69933185	Percentage:	10.13363029
Section Name:	PFSO	Section Name:	RS
Counted:	91	Counted:	38
Accuracy:	31.86813187	Accuracy:	0
PPV:	0.475409836	PPV:	0
NPV:	0.928653625	NPV:	0.957683742
Sensitivity:	0.318681319	Sensitivity:	0
Specificity:	0.961859356	Specificity:	1
Total_Sections:	898	Total_Sections:	898
Percentage:	10.13363029	Percentage:	4.231625835
Section Name:	PE	Section Name:	LP
Counted:	456	Counted:	93
Accuracy:	92.3245614	Accuracy:	44.08602151
PPV:	0.708754209	PPV:	0.683333333
NPV:	0.926624738	NPV:	0.939323221
Sensitivity:	0.923245614	Sensitivity:	0.440860215
Specificity:	0.718699187	Specificity:	0.976941748
Total_Sections:	898	Total_Sections:	898
Percentage:	50.77951002	Percentage:	10.35634744
Section Name:	HOPC	Section Name:	PC
Counted:	55	Counted:	34
Accuracy:	94.54545455	Accuracy:	0
PPV:	1	PPV:	0
NPV:	0.996453901	NPV:	0.962138085
Sensitivity:	0.945454545	Sensitivity:	0
Specificity:	1	Specificity:	0.996539792
Total_Sections:	898	Total_Sections:	898
Percentage:	6.124721604	Percentage:	3.786191537
Total_Sections:	898		
Average Sensitivity:		0.683741648	
Average Specificity:		0.841539759	
Average PPV:(861)		0.647683365	
Average NPV:		0.93605642	
Average Accuracy:		68.37416481	

Cardiology				
Section Name:	PE		Section Name:	PFSO
Counted:	48		Counted:	8
Accuracy:	93.75		Accuracy:	12.5
PPV:	0.833333333		PPV:	0.333333333
NPV:	0.9375		NPV:	0.923913043
Sensitivity:	0.9375		Sensitivity:	0.125
Specificity:	0.833333333		Specificity:	0.977011494
Total_Sections:	93		Total_Sections:	93
Percentage:	51.61290323		Percentage:	8.602150538
Section Name:	LP		Section Name:	PhMdAg
Counted:	12		Counted:	14
Accuracy:	50		Accuracy:	71.42857143
PPV:	0.75		PPV:	0.5
NPV:	0.931034483		NPV:	0.951807229
Sensitivity:	0.5		Sensitivity:	0.714285714
Specificity:	0.975903614		Specificity:	0.887640449
Total_Sections:	93		Total_Sections:	93
Percentage:	12.90322581		Percentage:	15.05376344
Section Name:	HOPC		Section Name:	RS
Counted:	5		Counted:	3
Accuracy:	100		Accuracy:	33.33333333
PPV:	0.714285714		PPV:	0.5
NPV:	1		NPV:	0.97826087
Sensitivity:	1		Sensitivity:	0.333333333
Specificity:	0.977777778		Specificity:	0.989010989
Total_Sections:	93		Total_Sections:	93
Percentage:	5.376344086		Percentage:	3.225806452
Section Name:	PC			
Counted:	4			
Accuracy:	0			
PPV:	0			
NPV:	0.956989247			
Sensitivity:	0			
Specificity:	1			
Total_Sections:	93			
Percentage:	4.301075269			
Total_Sections:	93			
Average Sensitivity:		0.731182796		
Average Specificity:		0.901180898		
Average PPV:(90)		0.708201058		
Average NPV:		0.95324472		
Average Accuracy:		73.11827957		

Surgery	
Section Name:	PhMdAg
Counted:	3
Accuracy:	100
PPV:	0.375
NPV:	1
Sensitivity:	1
Specificity:	0.880952381
Total_Sections:	40
Percentage:	7.5

Section Name:	HOPC
Counted:	3
Accuracy:	0
PPV:	0
NPV:	0.925
Sensitivity:	0
Specificity:	1
Total_Sections:	40
Percentage:	7.5

Section Name:	PE
Counted:	26
Accuracy:	100
PPV:	0.787878788
NPV:	1
Sensitivity:	1
Specificity:	0.666666667
Total_Sections:	40
Percentage:	65

Section Name:	LP
Counted:	5
Accuracy:	0
PPV:	0
NPV:	0.875
Sensitivity:	0
Specificity:	1
Total_Sections:	40
Percentage:	12.5

Section Name:	RS
Counted:	1
Accuracy:	0
PPV:	0
NPV:	0.975
Sensitivity:	0
Specificity:	1
Total_Sections:	40
Percentage:	2.5

Section Name:	PC
Counted:	2
Accuracy:	0
PPV:	0
NPV:	0.95
Sensitivity:	0
Specificity:	1
Total_Sections:	40
Percentage:	5

Section Name:	PFSO
Counted:	1
Accuracy:	0
PPV:	0
NPV:	0.975
Sensitivity:	0
Specificity:	1
Total_Sections:	40
Percentage:	2.5

Total_Sections:	40	
Average Sensitivity:		0.725
Average Specificity:		0.799404762
Average PPV:(29)		0.745167189
Average NPV:		1
Average Accuracy:		72.5

Respiratory			
Section Name:	LP	Section Name:	PFSO
Counted:	6	Counted:	13
Accuracy:	66.66666667	Accuracy:	61.53846154
PPV:	0.8	PPV:	0.888888889
NPV:	0.975903614	NPV:	0.936708861
Sensitivity:	0.666666667	Sensitivity:	0.615384615
Specificity:	0.987804878	Specificity:	0.986666667
Total_Sections:	87	Total_Sections:	87
Percentage:	6.896551724	Percentage:	14.94252874
Section Name:	PhMdAg	Section Name:	PE
Counted:	12	Counted:	46
Accuracy:	58.33333333	Accuracy:	97.82608696
PPV:	0.411764706	PPV:	0.9
NPV:	0.9375	NPV:	0.976190476
Sensitivity:	0.583333333	Sensitivity:	0.97826087
Specificity:	0.882352941	Specificity:	0.891304348
Total_Sections:	87	Total_Sections:	87
Percentage:	13.79310345	Percentage:	52.87356322
Section Name:	RS	Section Name:	HOPC
Counted:	4	Counted:	4
Accuracy:	100	Accuracy:	0
PPV:	0.571428571	PPV:	0
NPV:	1	NPV:	0.954022989
Sensitivity:	1	Sensitivity:	0
Specificity:	0.965116279	Specificity:	1
Total_Sections:	87	Total_Sections:	87
Percentage:	4.597701149	Percentage:	4.597701149
Section Name:	PC		
Counted:	3		
Accuracy:	0		
PPV:	0		
NPV:	0.965517241		
Sensitivity:	0		
Specificity:	1		
Total_Sections:	87		
Percentage:	3.448275862		
Total_Sections:	87		
Average Sensitivity:		0.781609195	
Average Specificity:		0.933358579	
Average PPV:(81)		0.802252424	
Average NPV:		0.975862557	
Average Accuracy:		78.16091954	

Neurology

Section Name:	PhMdAg
Counted:	12
Accuracy:	75
PPV:	0.375
NPV:	0.952380952
Sensitivity:	0.75
Specificity:	0.8
Total_Sections:	72
Percentage:	16.66666667

Section Name:	PE
Counted:	35
Accuracy:	80
PPV:	0.777777778
NPV:	0.840909091
Sensitivity:	0.8
Specificity:	0.822222222
Total_Sections:	72
Percentage:	48.61111111

Section Name:	LP
Counted:	9
Accuracy:	88.88888889
PPV:	0.666666667
NPV:	0.984375
Sensitivity:	0.888888889
Specificity:	0.940298507
Total_Sections:	72
Percentage:	12.5

Section Name:	RS
Counted:	3
Accuracy:	0
PPV:	0
NPV:	0.958333333
Sensitivity:	0
Specificity:	1
Total_Sections:	72
Percentage:	4.166666667

Section Name:	HOPC
Counted:	5
Accuracy:	0
PPV:	0
NPV:	0.930555556
Sensitivity:	0
Specificity:	1
Total_Sections:	72
Percentage:	6.944444444

Section Name:	PFSO
Counted:	7
Accuracy:	14.28571429
PPV:	1
NPV:	0.915492958
Sensitivity:	0.142857143
Specificity:	1
Total_Sections:	72
Percentage:	9.722222222

Section Name:	PC
Counted:	2
Accuracy:	0
PPV:	0
NPV:	0.972222222
Sensitivity:	0
Specificity:	1
Total_Sections:	72
Percentage:	2.777777778

Total_Sections:	72
Average Sensitivity:	0.638888889
Average Specificity:	0.886673116
Average PPV:(63)	0.709876543
Average NPV:	0.911117188
Average Accuracy:	63.88888889

General Medicine			
Section Name:	PhMdAg	Section Name:	LP
Counted:	26	Counted:	22
Accuracy:	69.23076923	Accuracy:	54.54545455
PPV:	0.529411765	PPV:	0.75
NPV:	0.948387097	NPV:	0.937888199
Sensitivity:	0.692307692	Sensitivity:	0.545454545
Specificity:	0.901840491	Specificity:	0.974193548
Total_Sections:	173	Total_Sections:	173
Percentage:	15.02890173	Percentage:	12.71676301
Section Name:	PE	Section Name:	HOPC
Counted:	86	Counted:	10
Accuracy:	96.51162791	Accuracy:	70
PPV:	0.775700935	PPV:	1
NPV:	0.966666667	NPV:	0.981927711
Sensitivity:	0.965116279	Sensitivity:	0.7
Specificity:	0.783783784	Specificity:	1
Total_Sections:	173	Total_Sections:	173
Percentage:	49.71098266	Percentage:	5.780346821
Section Name:	RS	Section Name:	PFSO
Counted:	7	Counted:	17
Accuracy:	0	Accuracy:	29.41176471
PPV:	0	PPV:	0.5
NPV:	0.959537572	NPV:	0.928571429
Sensitivity:	0	Sensitivity:	0.294117647
Specificity:	1	Specificity:	0.968944099
Total_Sections:	173	Total_Sections:	173
Percentage:	4.046242775	Percentage:	9.826589595
Section Name:	PC		
Counted:	6		
Accuracy:	0		
PPV:	0		
NPV:	0.965317919		
Sensitivity:	0		
Specificity:	1		
Total_Sections:	173		
Percentage:	3.468208092		
Total_Sections:	173		
Average Sensitivity:	0.722543353		
Average Specificity:	0.877211364		
Average PPV:(161)	0.717235939		
Average NPV:	0.962650868		
Average Accuracy:	72.25433526		

Gastroenterology

Section Name:	PhMdAg
Counted:	5
Accuracy:	100
PPV:	0.714285714
NPV:	1
Sensitivity:	1
Specificity:	0.952380952
Total_Sections:	45
Percentage:	11.11111111

Section Name:	PE
Counted:	28
Accuracy:	100
PPV:	0.875
NPV:	1
Sensitivity:	1
Specificity:	0.80952381
Total_Sections:	45
Percentage:	62.22222222

Section Name:	LP
Counted:	4
Accuracy:	75
PPV:	1
NPV:	0.976190476
Sensitivity:	0.75
Specificity:	1
Total_Sections:	45
Percentage:	8.888888889

Section Name:	HOPC
Counted:	2
Accuracy:	100
PPV:	1
NPV:	1
Sensitivity:	1
Specificity:	1
Total_Sections:	45
Percentage:	4.444444444

Section Name:	RS
Counted:	3
Accuracy:	0
PPV:	0
NPV:	0.933333333
Sensitivity:	0
Specificity:	1
Total_Sections:	45
Percentage:	6.666666667

Section Name:	PFSO
Counted:	2
Accuracy:	50
PPV:	0.5
NPV:	0.977272727
Sensitivity:	0.5
Specificity:	0.977272727
Total_Sections:	45
Percentage:	4.444444444

Section Name:	PC
Counted:	2
Accuracy:	0
PPV:	0
NPV:	0.955555556
Sensitivity:	0
Specificity:	1
Total_Sections:	45
Percentage:	4.444444444

Total_Sections:	45	
Average Sensitivity:		0.866666667
Average Specificity:		0.897402597
Average PPV:(41)		0.855400697
Average NPV:		1.012675966
Average Accuracy:		86.66666667

Mixed Medical Subspecialties

Section Name:	PE
Counted:	46
Accuracy:	97.82608696
PPV:	0.725806452
NPV:	0.975609756
Sensitivity:	0.97826087
Specificity:	0.701754386
Total_Sections:	86
Percentage:	53.48837209

Section Name:	PhMdAg
Counted:	10
Accuracy:	50
PPV:	0.5
NPV:	0.938271605
Sensitivity:	0.5
Specificity:	0.938271605
Total_Sections:	86
Percentage:	11.62790698

Section Name:	HOPC
Counted:	7
Accuracy:	100
PPV:	1
NPV:	1
Sensitivity:	1
Specificity:	1
Total_Sections:	86
Percentage:	8.139534884

Section Name:	PFSO
Counted:	11
Accuracy:	63.63636364
PPV:	0.875
NPV:	0.949367089
Sensitivity:	0.636363636
Specificity:	0.986842105
Total_Sections:	86
Percentage:	12.79069767

Section Name:	LP
Counted:	8
Accuracy:	0
PPV:	0
NPV:	0.906976744
Sensitivity:	0
Specificity:	1
Total_Sections:	86
Percentage:	9.302325581

Section Name:	RS
Counted:	2
Accuracy:	0
PPV:	0
NPV:	0.976744186
Sensitivity:	0
Specificity:	1
Total_Sections:	86
Percentage:	2.325581395

Section Name:	PC
Counted:	3
Accuracy:	0
PPV:	0
NPV:	0.965116279
Sensitivity:	0
Specificity:	1
Total_Sections:	86
Percentage:	3.488372093

Total_Sections:	86	
Average Sensitivity:		0.744186047
Average Specificity:		0.843240476
Average PPV:(74)		0.743406713
Average NPV:		0.974516907
Average Accuracy:		74.41860465

Discharge Summary Documents

Entire Set			
Section_Name:	Laboratory	Section_Name:	Discharge_Plan
Counted:	20	Counted:	75
Accuracy:	55	Accuracy:	76
PPV:	0.523809524	PPV:	0.467213115
NPV:	0.969594595	NPV:	0.928
Sensitivity:	0.55	Sensitivity:	0.76
Specificity:	0.966329966	Specificity:	0.781144781
Total_Sections:	307	Total_Sections:	307
Percentage:	6.51465798	Percentage:	24.42996743
Section_Name:	Discharge_Diagnoses	Section_Name:	History
Counted:	56	Counted:	47
Accuracy:	16.07142857	Accuracy:	70.21276596
PPV:	0.333333333	PPV:	0.611111111
NPV:	0.842281879	NPV:	0.948905109
Sensitivity:	0.160714286	Sensitivity:	0.70212766
Specificity:	0.933085502	Specificity:	0.925266904
Total_Sections:	307	Total_Sections:	307
Percentage:	18.24104235	Percentage:	15.30944625
Section_Name:	Physical_Examination	Section_Name:	Hospital_Course
Counted:	29	Counted:	38
Accuracy:	41.37931034	Accuracy:	71.05263158
PPV:	0.48	PPV:	0.771428571
NPV:	0.942372881	NPV:	0.960714286
Sensitivity:	0.413793103	Sensitivity:	0.710526316
Specificity:	0.95532646	Specificity:	0.971119134
Total_Sections:	307	Total_Sections:	307
Percentage:	9.446254072	Percentage:	12.37785016
Section_Name:	Past_Medical_History	Section_Name:	Allergies
Counted:	16	Counted:	6
Accuracy:	18.75	Accuracy:	0
PPV:	0.230769231	PPV:	0
NPV:	0.957236842	NPV:	0.980456026
Sensitivity:	0.1875	Sensitivity:	0
Specificity:	0.966777409	Specificity:	1
Total_Sections:	307	Total_Sections:	307
Percentage:	5.211726384	Percentage:	1.954397394

Section_Name:	Medications
Counted:	2
Accuracy:	0
PPV:	0
NPV:	0.993485342
Sensitivity:	0
Specificity:	1
Total_Sections:	307
Percentage:	0.651465798

Section_Name:	Social_Personal _Family
Counted:	8
Accuracy:	0
PPV:	0
NPV:	0.973941368
Sensitivity:	0
Specificity:	1
Total_Sections:	307
Percentage:	2.605863192

Total_Sections:	307	
Average Sensitivity:		0.495114
Average Specificity:		0.878594
Average PPV: (281)		0.497625
Average NPV:		0.897626
Average Accuracy:		49.5114

Departmental Groupings:

General Medicine			
Section_Name:	Hospital_Course	Section_Name:	Discharge_Plan
Counted:	10	Counted:	20
Accuracy:	0	Accuracy:	50
PPV:	0	PPV:	0.384615385
NPV:	0.838709677	NPV:	0.807692308
Sensitivity:	0	Sensitivity:	0.5
Specificity:	1	Specificity:	0.724137931
Total_Sections:	62	Total_Sections:	62
Percentage:	16.12903226	Percentage:	32.25806452
Section_Name:	History	Section_Name:	Discharge_Diagnoses
Counted:	9	Counted:	11
Accuracy:	44.44444444	Accuracy:	0
PPV:	0.4	PPV:	0
NPV:	0.913793103	NPV:	0.822580645
Sensitivity:	0.444444444	Sensitivity:	0
Specificity:	0.898305085	Specificity:	1
Total_Sections:	62	Total_Sections:	62
Percentage:	14.51612903	Percentage:	17.74193548
Section_Name:	Past_Medical_History	Section_Name:	Physical_Examination
Counted:	2	Counted:	1
Accuracy:	0	Accuracy:	100
PPV:	0	PPV:	0.083333333
NPV:	0.967741935	NPV:	1
Sensitivity:	0	Sensitivity:	1
Specificity:	0.983606557	Specificity:	0.847222222
Total_Sections:	62	Total_Sections:	62
Percentage:	3.225806452	Percentage:	1.612903226
Section_Name:	Laboratory	Section_Name:	Social_Personal_Family
Counted:	4	Counted:	1
Accuracy:	50	Accuracy:	100
PPV:	1	PPV:	0.125
NPV:	0.966666667	NPV:	1
Sensitivity:	0.5	Sensitivity:	1
Specificity:	1	Specificity:	0.897058824
Total_Sections:	62	Total_Sections:	62
Percentage:	6.451612903	Percentage:	1.612903226
Section_Name:	Allergies		

Counted:	1
Accuracy:	0
PPV:	0
NPV:	0.983870968
Sensitivity:	0
Specificity:	1
Total_Sections:	62
Percentage:	1.612903226

Total_Sections:	62	
Average Sensitivity:		0.290323
Average Specificity:		0.84321
Average PPV: (37)		0.418936
Average NPV:		0.816121
Average Accuracy:		29.03226

Other Medical Specialties					
Section_Name:	Laboratory		Section_Name:	Allergies	
Counted:	6		Counted:	1	
Accuracy:	100		Accuracy:	0	
PPV:	0.857142857		PPV:	0	
NPV:	1		NPV:	0.983870968	
Sensitivity:	1		Sensitivity:	0	
Specificity:	0.98245614		Specificity:	1	
Total_Sections:	62		Total_Sections:	62	
Percentage:	9.677419355		Percentage:	1.612903226	
Section_Name:	Hospital_Course		Section_Name:	Past_Medical_History	
Counted:	8		Counted:	3	
Accuracy:	0		Accuracy:	33.33333333	
PPV:	0		PPV:	1	
NPV:	0.870967742		NPV:	0.967213115	
Sensitivity:	0		Sensitivity:	0.333333333	
Specificity:	1		Specificity:	1	
Total_Sections:	62		Total_Sections:	62	
Percentage:	12.90322581		Percentage:	4.838709677	
Section_Name:	Discharge_Plan		Section_Name:	Social_Personal_Family	
Counted:	16		Counted:	2	
Accuracy:	100		Accuracy:	0	
PPV:	0.470588235		PPV:	0	
NPV:	1		NPV:	0.967741935	
Sensitivity:	1		Sensitivity:	0	
Specificity:	0.71875		Specificity:	1	
Total_Sections:	62		Total_Sections:	62	
Percentage:	25.80645161		Percentage:	3.225806452	
Section_Name:	History		Section_Name:	Discharge_Diagnoses	
Counted:	11		Counted:	10	
Accuracy:	63.63636364		Accuracy:	80	
PPV:	0.777777778		PPV:	1	
NPV:	0.927272727		NPV:	0.962962963	
Sensitivity:	0.636363636		Sensitivity:	0.8	
Specificity:	0.962264151		Specificity:	1	
Total_Sections:	62		Total_Sections:	62	
Percentage:	17.74193548		Percentage:	16.12903226	
Section_Name:	Physical_Examination		Section_Name:	Medications	
Counted:	3		Counted:	1	
Accuracy:	33.33333333		Accuracy:	0	

PPV:	0.5
NPV:	0.967213115
Sensitivity:	0.333333333
Specificity:	0.983333333
Total_Sections:	62
Percentage:	4.838709677

PPV:	0
NPV:	0.983870968
Sensitivity:	0
Specificity:	1
Total_Sections:	62
Percentage:	1.612903226

Total_Sections:	62	
Average	Sensitivity:	0.629032
Average	Specificity:	0.902091
Average	PPV: (49)	0.729139
Average	NPV:	0.943611
Average	Accuracy:	62.90323

Obstetrics and Gynecology

Section_Name:	Discharge_Diagnoses
Counted:	7
Accuracy:	71.42857143
PPV:	0.416666667
NPV:	0.935483871
Sensitivity:	0.714285714
Specificity:	0.805555556
Total_Sections:	36
Percentage:	19.44444444

Section_Name:	Hospital_Course
Counted:	5
Accuracy:	100
PPV:	0.833333333
NPV:	1
Sensitivity:	1
Specificity:	0.96875
Total_Sections:	36
Percentage:	13.88888889

Section_Name:	Discharge_Plan
Counted:	6
Accuracy:	66.66666667
PPV:	0.5
NPV:	0.9375
Sensitivity:	0.666666667
Specificity:	0.882352941
Total_Sections:	36
Percentage:	16.66666667

Section_Name:	Physical_Examination
Counted:	2
Accuracy:	50
PPV:	1
NPV:	0.971428571
Sensitivity:	0.5
Specificity:	1
Total_Sections:	36
Percentage:	5.555555556

Section_Name:	History
Counted:	6
Accuracy:	33.33333333
PPV:	1
NPV:	0.882352941
Sensitivity:	0.333333333
Specificity:	1
Total_Sections:	36
Percentage:	16.66666667

Section_Name:	Past_Medical_History
Counted:	4
Accuracy:	25
PPV:	0.333333333
NPV:	0.914285714
Sensitivity:	0.25
Specificity:	0.941176471
Total_Sections:	36
Percentage:	11.11111111

Section_Name:	Laboratory
Counted:	4
Accuracy:	50

Section_Name:	Allergies
Counted:	1
Accuracy:	0

PPV:	1
NPV:	0.941176471
Sensitivity:	0.5
Specificity:	1
Total_Sections:	36
Percentage:	11.11111111

PPV:	0
NPV:	0.972222222
Sensitivity:	0
Specificity:	1
Total_Sections:	36
Percentage:	2.777777778

Total_Sections:	36	
Average Sensitivity:		0.555556
Average Specificity:		0.90393
Average PPV: (34)		0.688725
Average NPV:		0.911234
Average Accuracy:		55.55556

Surgery

Section_Name:	Discharge_Diagnoses
Counted:	12
Accuracy:	25
PPV:	0.333333333
NPV:	0.8
Sensitivity:	0.25
Specificity:	0.857142857
Total_Sections:	48
Percentage:	25

Section_Name:	History
Counted:	7
Accuracy:	100
PPV:	0.538461538
NPV:	1
Sensitivity:	1
Specificity:	0.872340426
Total_Sections:	48
Percentage:	14.58333333

Section_Name:	Discharge_Plan
Counted:	17
Accuracy:	88.23529412
PPV:	0.6
NPV:	0.939393939
Sensitivity:	0.882352941
Specificity:	0.756097561
Total_Sections:	48
Percentage:	35.41666667

Section_Name:	Laboratory
Counted:	2
Accuracy:	0
PPV:	0
NPV:	0.958333333
Sensitivity:	0
Specificity:	1
Total_Sections:	48
Percentage:	4.166666667

Section_Name:	Physical_Examination
Counted:	3
Accuracy:	0
PPV:	0
NPV:	0.9375
Sensitivity:	0
Specificity:	1
Total_Sections:	48
Percentage:	6.25

Section_Name:	Past_Medical_History
Counted:	3
Accuracy:	0
PPV:	0
NPV:	0.9375
Sensitivity:	0
Specificity:	1
Total_Sections:	48
Percentage:	6.25

Section_Name:	Hospital_Course

Counted:	5
Accuracy:	40
PPV:	1
NPV:	0.934782609
Sensitivity:	0.4
Specificity:	1
Total_Sections:	48
Percentage:	10.41666667

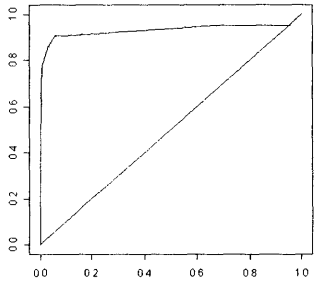
Total_Sections:	48	
Average Sensitivity:		0.5625
Average Specificity:		0.88012
Average PPV: (41)		0.560225
Average NPV:		0.933027
Average Accuracy:		56.25

Appendix 4

History and Physical Examination

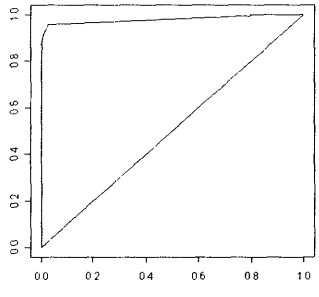
Entire- set. High Granularity Grouping

1. Presenting Complaint
Test data c-index: 0.91
Test data brier score: 0.039
Test data HL: 28.22
p value: 0

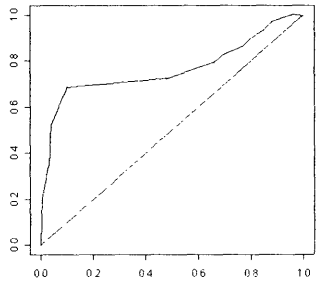


2. History of Presenting Complaint

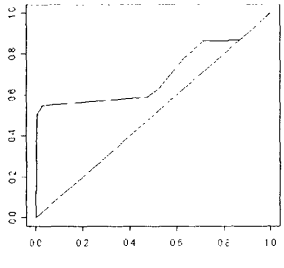
Test data c-index: 0.996
Test data brier score: 0.040
Test data HL: 37.00
p value: <0.001



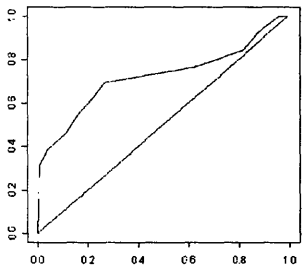
3. Past Medical or Surgical History
Test data c-index: 0.74
Test data brier score: 0.056
Test data HL: 15.98
p value: 0.043



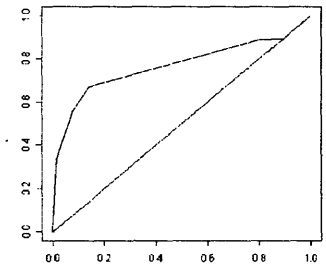
4. Medications
Test data c-index: 0.67
Test data brier score: 0.043
Test data HL: 16.70
p value: 0.033



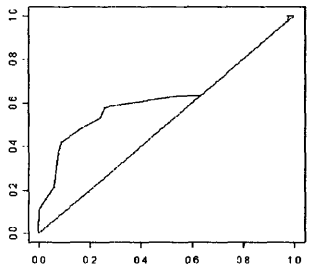
5. Allergies
Test data c-index: 0.68
Test data brier score: 0.030
Test data HL: 21.58
p value: 0.006



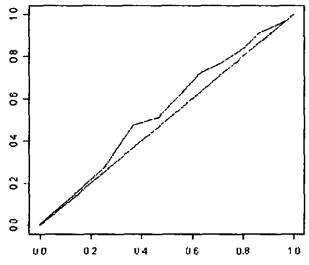
7. Personal History
Test data c-index: 0.68
Test data brier score: 0.02
Test data HL: 22.94
p value: 0.003



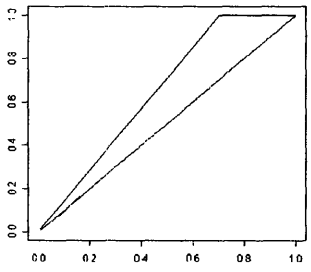
6. Family History
Test data c-index: 0.59
Test data brier score: 0.04
Test data HL: 14.08
p value: 0.080



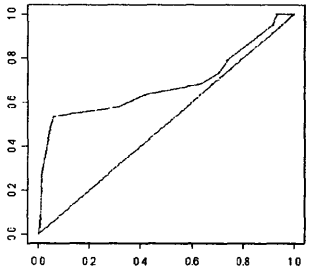
8. Social History
Test data c-index: 0.55
Test data brier score: 0.047
Test data HL: 9.62
p value: 0.293



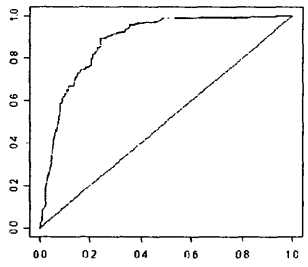
9. Occupational History
Test data c-index: 0.59
Test data brier score: 0.006
Test data HL: 28.757
p value: 0.000



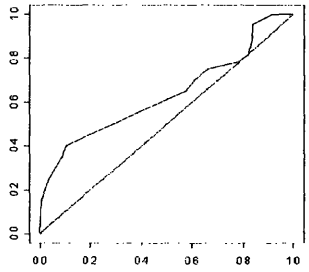
10. Review of Systems
Test data c-index: 0.65
Test data brier score: 0.039
Test data HL: 14.92
p value: 0.061



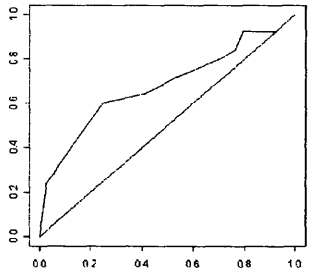
11. Physical Examination
Test data c-index: 0.88
Test data brier score: 0.35
Test data HL: 621.70
p value: <0.001



12. Laboratory
Test data c-index: 0.60
Test data brier score: 0.042
Test data HL: 10.75
p value: 0.216

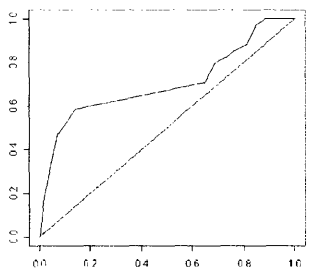


13. Plan
Test data c-index: 0.66
Test data brier score: 0.051
Test data HL: 10.13
p value: 0.256

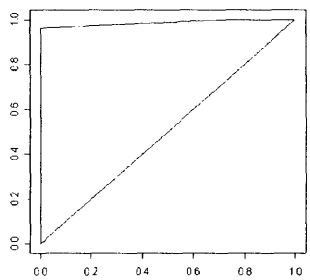


Entire set. Low Granularity Grouping

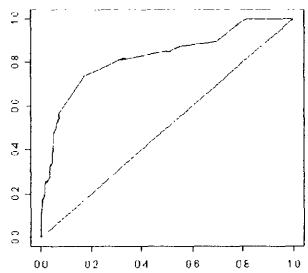
1. Presenting Complaint
Test data c-index: 0.66
Test data brier score: 0.042
Test data HL: 71.23
p value: <0.001



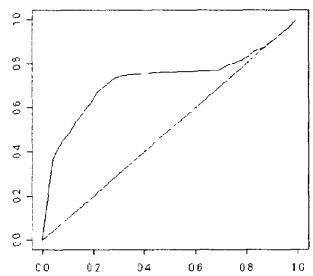
2. History of Presenting Complaint
Test data c-index: 0.99
Test data brier score: 0.042
Test data HL: 143.158
p value: <0.001



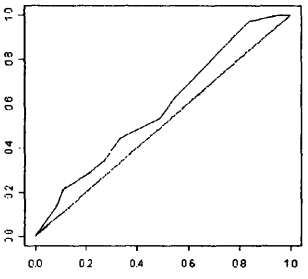
3. Past Histories Medicines and Allergies
Test data c-index: 0.82
Test data brier score: 0.107
Test data HL: 75.86
p value: <0.001



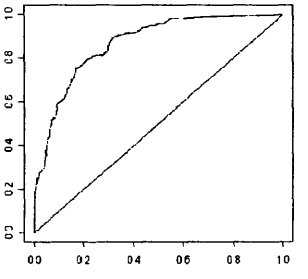
4. Personal Family Social Occupational
Test data c-index: 0.72
Test data brier score: 0.084
Test data HL: 47.56
p value: <0.001



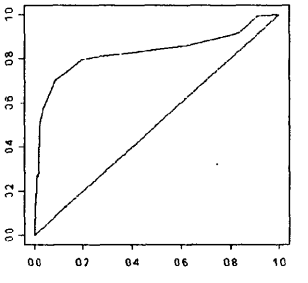
5. Review of Systems
Test data c-index: 0.56
Test data brier score: 0.045
Test data HL: 58.40
p value: <0.001



6. Physical Examination
Test data c-index: 0.87
Test data brier score: 0.27
Test data HL: 520.18
p value: <0.001

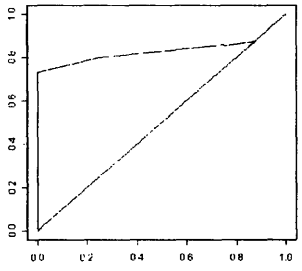


7. Laboratory and Plan
Test data c-index: 0.82
Test data brier score: 0.079
Test data HL: 84.91
p value: <0.001

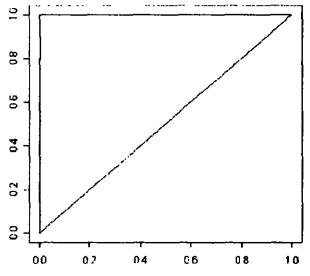


Cardiology diagnoses grouping

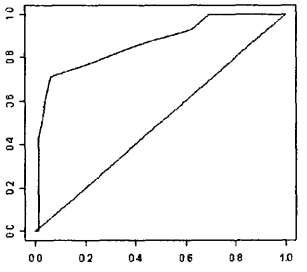
1. Presenting Complaint
Test data c-index: 0.80
Test data brier score: 0.046
Test data HL: 9.11
p value: 0.333



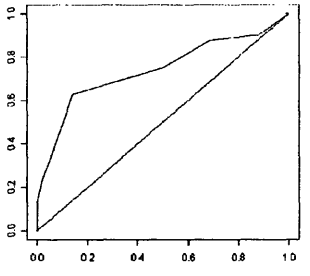
2. History of Presenting Complaint
Test data c-index: 1
Test data brier score: 0.039
Test data HL: 10.67
p value: 0.221



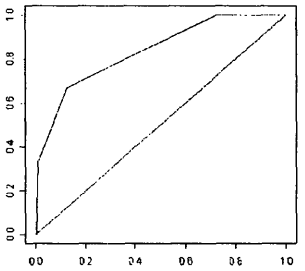
3. Past Histories Medicines and Allergies
Test data c-index: 0.84
Test data brier score: 0.104
Test data HL: 7.17
p value: 0.518



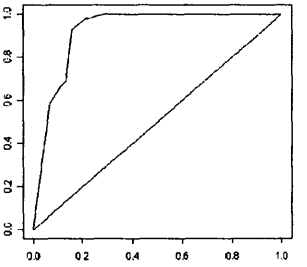
4.Personal Family Social Occupational
Test data c-index: 0.69
Test data brier score: 0.075
Test data HL: 5.99
p value: 0.648



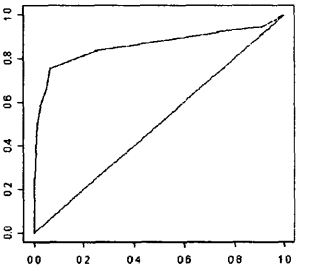
5. Review of Systems
Test data c-index: 0.71
Test data brier score: 0.036
Test data HL: 9.06
p value: 0.337



6. Physical Examination
Test data c-index: 0.90
Test data brier score: 0.27
Test data HL: 79.93
p value: <0.001

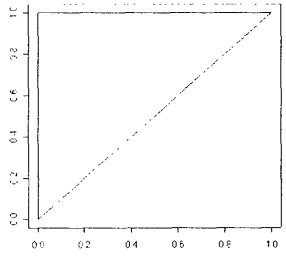


7. Laboratory and Plan
Test data c-index: 0.82
Test data brier score: 0.092
Test data HL: 11.04
p value: 0.200

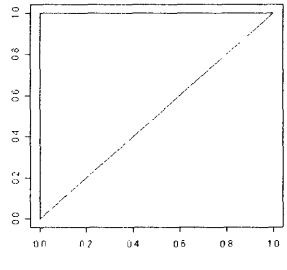


Gastroenterology diagnoses grouping

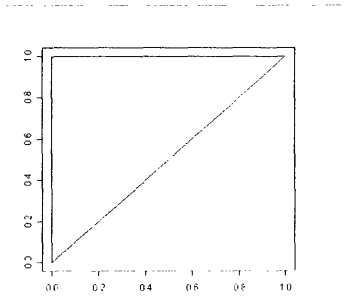
1. Presenting Complaint
Test data c-index: 1
Test data brier score: 0.047
Test data HL: 11.73
p value: 0.164



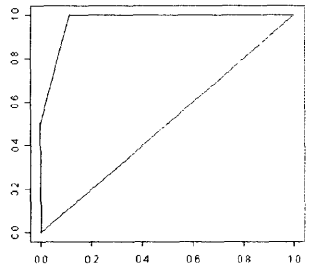
2. History of Presenting Complaint
Test data c-index: 1
Test data brier score: 0.034
Test data HL: 5.09
p value: 0.748



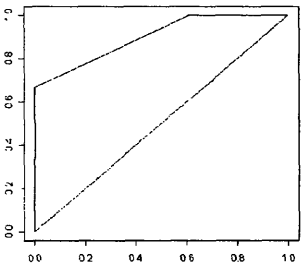
3. Past Histories Medicines and Allergies
Test data c-index: 1
Test data brier score: 0.069
Test data HL: 5.55
p value: 0.698



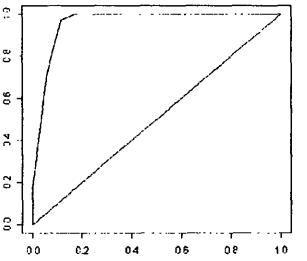
4.Personal Family Social Occupational
Test data c-index: 0.94
Test data brier score: 0.044
Test data HL: 5.066
p value: 0.751



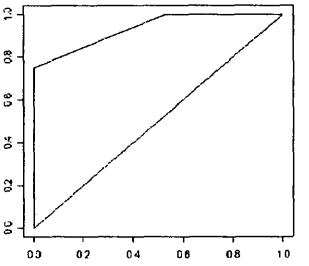
5. Review of Systems
Test data c-index: 0.80
Test data brier score: 0.063
Test data HL: 5.27
p value: 0.728



6. Physical Examination
Test data c-index: 0.94
Test data brier score: 0.302
Test data HL: 44.526
p value: <0.001

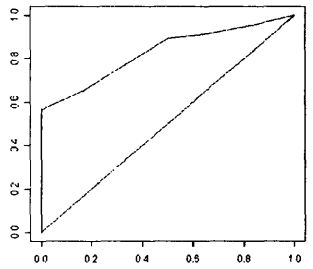


7. Laboratory and Plan
Test data c-index: 0.87
Test data brier score: 0.060
Test data HL: 5.29
p value: 0.726

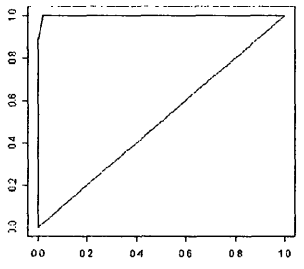


General Medicine Diagnoses Grouping

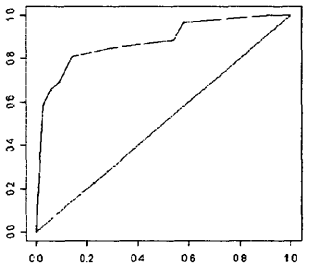
1. Presenting Complaint
Test data c-index: 0.79
Test data brier score: 0.040
Test data HL: 12.86
p value: 0.117



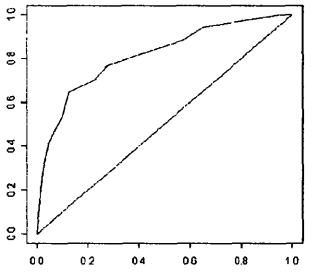
2. History of Presenting Complaint
Test data c-index: 0.99
Test data brier score: 0.045
Test data HL: 24.037
p value: 0.002



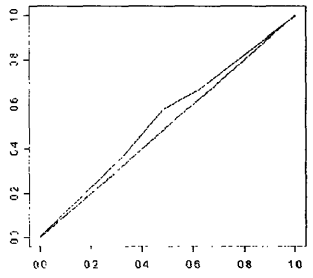
3.Past Histories Medicines and Allergies
Test data c-index: 0.86
Test data brier score: 0.104
Test data HL: 19.38
p value: 0.013



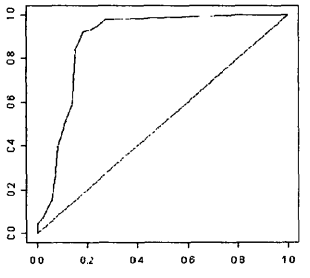
4.Personal Family Social Occupational
Test data c-index: 0.78
Test data brier score: 0.08
Test data HL: 10.47
p value: 0.233



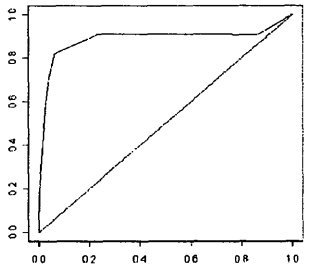
5. Review of Systems
Test data c-index: 0.52
Test data brier score: 0.043
Test data HL: 9.53
p value: 0.300



6. Physical Examination
Test data c-index: 0.87
Test data brier score: 0.260
Test data HL: 115.34
p value: <0.001

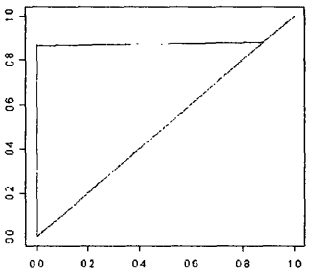


7. Laboratory and Plan
Test data c-index: 0.87
Test data brier score: 0.088
Test data HL: 24.82
p value: 0.002

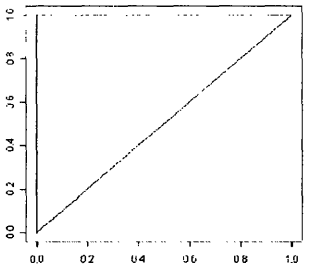


Mixed Medical Specialty Diagnoses Groupings

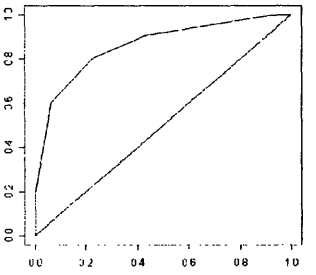
1. Presenting Complaint
Test data c-index: 0.88
Test data brier score: 0.040
Test data HL: 14.87
p value: 0.062



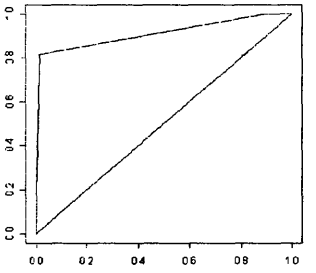
2. History of Presenting Complaint
Test data c-index: 1
Test data brier score: 0.050
Test data HL: 10.48
p value: 0.233



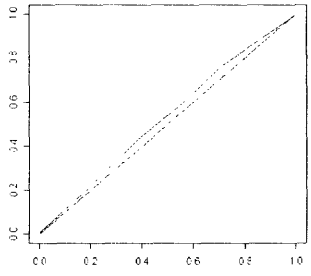
3.Past Histories Medicines and Allergies
Test data c-index: 0.80
Test data brier score: 0.089
Test data HL: 6.77
p value: 0.562



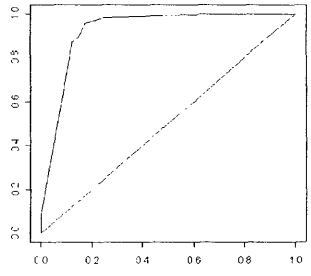
4.Personal Family Social Occupational
Test data c-index: 0.83
Test data brier score: 0.089
Test data HL: 12.75
p value: 0.121



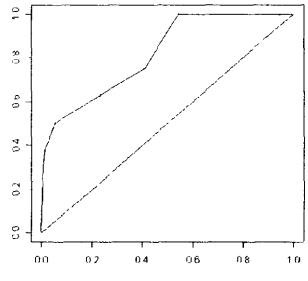
5. Review of Systems
Test data c-index: 0.51
Test data brier score: 0.030
Test data HL: 9.41
p value: 0.309



6. Physical Examination
Test data c-index: 0.89
Test data brier score: 0.275
Test data HL: 71.22
p value: <0.001

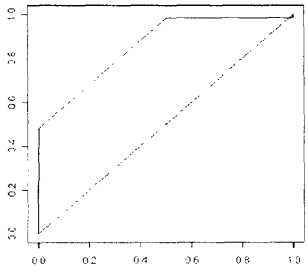


7. Laboratory and Plan
Test data c-index: 0.79
Test data brier score: 0.083
Test data HL: 7.53
p value: 0.481

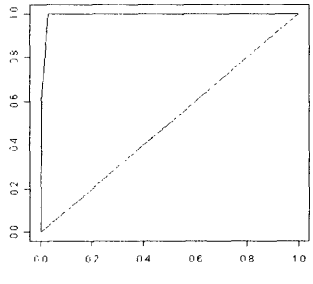


Neurology Diagnoses Grouping

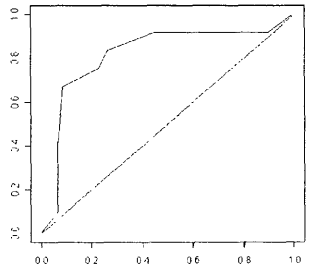
1. Presenting Complaint
Test data c-index: 0.73
Test data brier score: 0.034
Test data HL: 7.30
p value: 0.505



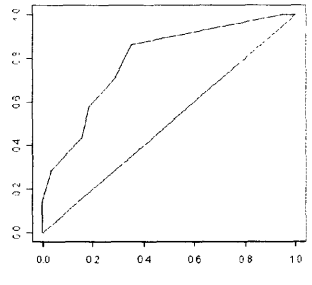
2. History of Presenting Complaint
Test data c-index: 0.99
Test data brier score: 0.063
Test data HL: 9.63
p value: 0.292



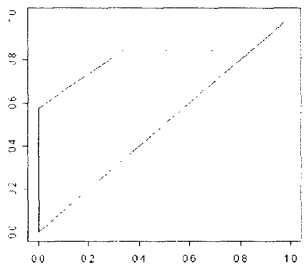
3.Past Histories Medicines and Allergies
Test data c-index: 0.79
Test data brier score: 0.118
Test data HL: 10.62
p value: 0.224



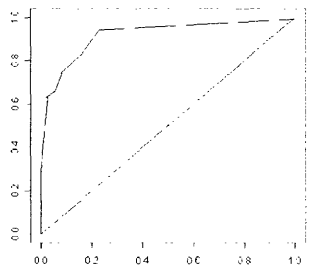
4.Personal Family Social Occupational
Test data c-index: 0.72
Test data brier score: 0.084
Test data HL: 5.78
p value: 0.672



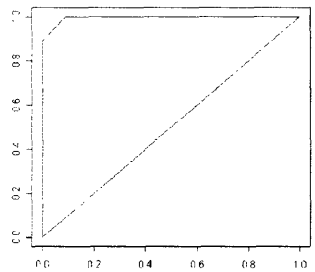
5. Review of Systems
Test data c-index: 0.75
Test data brier score: 0.045
Test data HL: 9.92
p value: 0.271



6. Physical Examination
Test data c-index: 0.90
Test data brier score: 0.263
Test data HL: 44.90
p value: <0.001

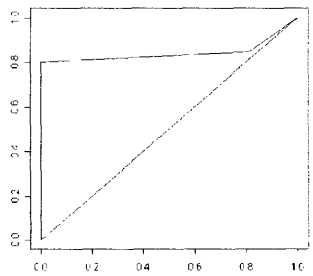


7. Laboratory and Plan
Test data c-index: 0.99
Test data brier score: 0.076
Test data HL: 44.90
p value: <0.001

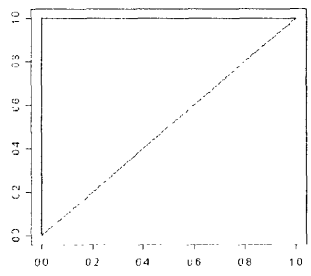


Respiratory Diagnoses Grouping

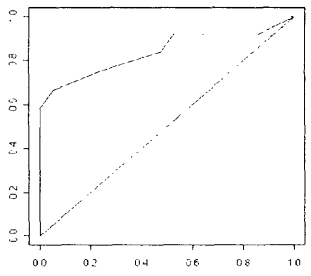
1. Presenting Complaint
Test data c-index: 0.87
Test data brier score: 0.039
Test data HL: 16.02
p value: 0.042



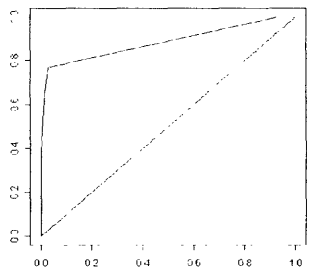
2. History of Presenting Complaint
Test data c-index: 1
Test data brier score: 0.038
Test data HL: 10.11
p value: 0.258



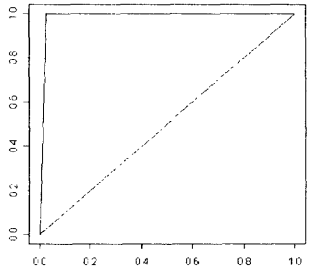
3.Past Histories Medicines and Allergies
Test data c-index: 0.81
Test data brier score: 0.099
Test data HL: 6.96
p value: 0.541



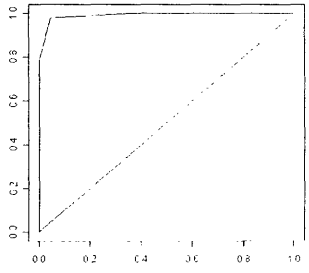
4.Personal Family Social Occupational
Test data c-index: 0.78
Test data brier score: 0.099
Test data HL: 11.31
p value: 0.185



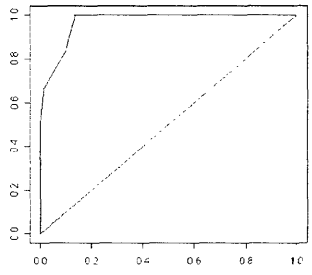
5. Review of Systems
Test data c-index: 0.98
Test data brier score: 0.039
Test data HL: 10.53
p value: 0.230



6. Physical Examination
Test data c-index: 0.98
Test data brier score: 0.266
Test data HL: 79.41
p value: <0.001

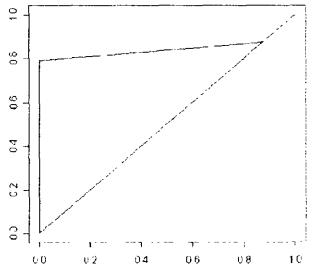


7. Laboratory and Plan
Test data c-index: 0.96
Test data brier score: 0.055
Test data HL: 7.96
p value: 0.437

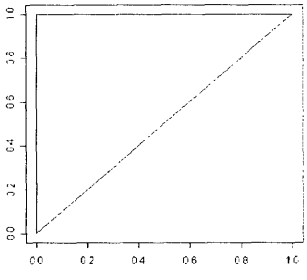


Surgery Diagnoses Groupings

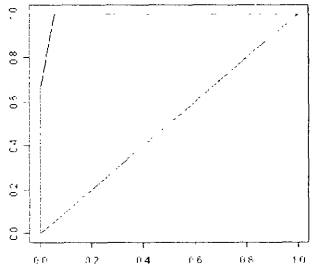
1. Presenting Complaint
Test data c-index: 0.85
Test data brier score: 0.051
Test data HL: 11.62
p value: 0.169



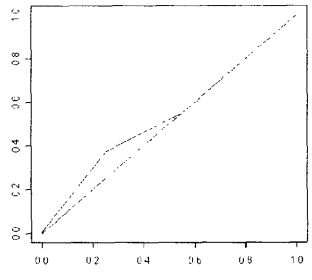
2. History of Presenting Complaint
Test data c-index: 1
Test data brier score: 0.0574
Test data HL: 6.40
p value: 0.602



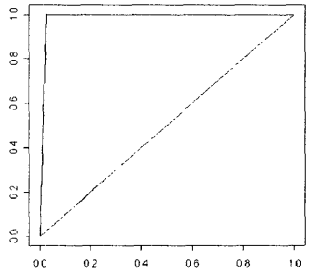
3.Past Histories Medicines and Allergies
Test data c-index: 0.98
Test data brier score: 0.057
Test data HL: 6.39
p value: 0.604



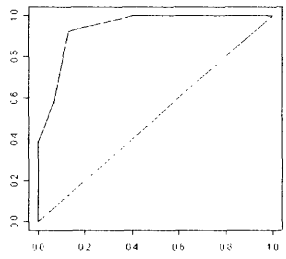
4.Personal Family Social Occupational
Test data c-index: 0.53
Test data brier score: 0.031
Test data HL: 5.21
p value: 0.735



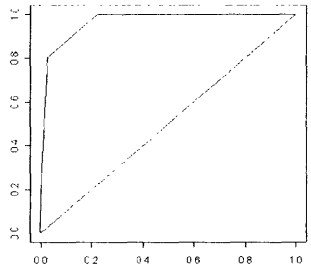
5. Review of Systems
Test data c-index: 0.98
Test data brier score: 0.031
Test data HL: 5.151
p value: 0.741



6. Physical Examination
Test data c-index: 0.91
Test data brier score: 0.323
Test data HL: 45.59
p value: <0.001



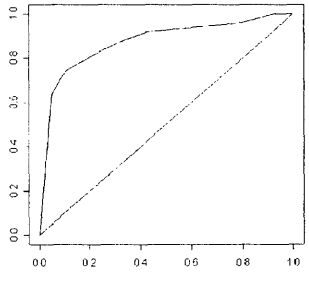
7. Laboratory and Plan
Test data c-index: 0.93
Test data brier score: 0.095
Test data HL: 12.75
p value: 0.121



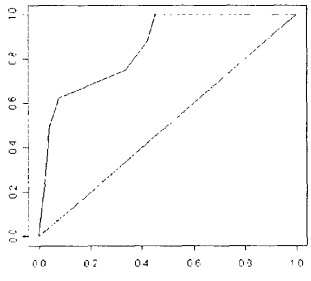
Discharge Summary Documents

Entire Set

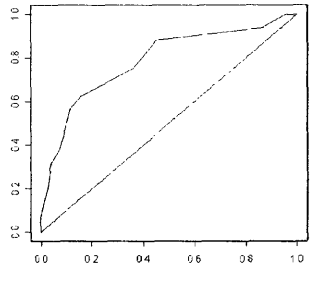
1. History
Test data c-index: 0.86
Test data brier score: 0.11
Test data HL: 39.10
p value: <0.001



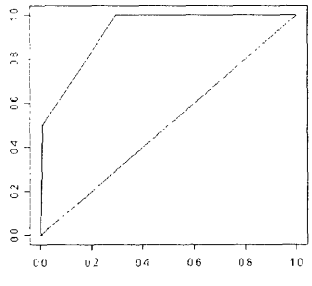
3. Social Personal Family
Test data c-index: 0.82
Test data brier score: 0.028
Test data HL: 19.26
p value: 0.014



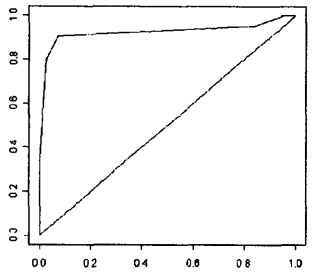
2. Past Medical Or Surgical History
Test data c-index: 0.75
Test data brier score: 0.048
Test data HL: 11.67
p value: 0.167



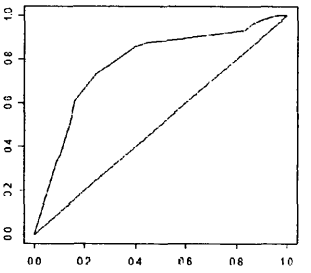
4. Medications
Test data c-index: 0.85
Test data brier score: 0.012
Test data HL: 23.85
p value: 0.002



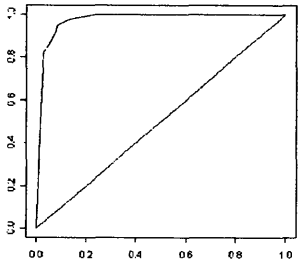
5. Laboratory
Test data c-index: 0.89
Test data brier score: 0.05
Test data HL: 28.87
p value: 0.000



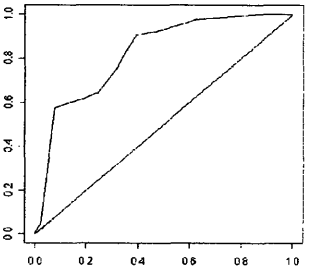
7. Discharge Diagnosis
Test data c-index: 0.75
Test data brier score: 0.139
Test data HL: 40.79
p value: <0.001



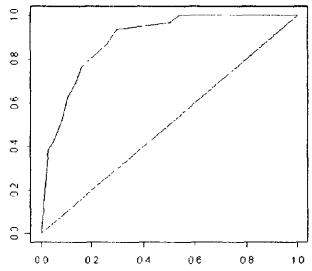
6. Hospital Course
Test data c-index: 0.97
Test data brier score: 0.087
Test data HL: 71.14
p value: <0.001



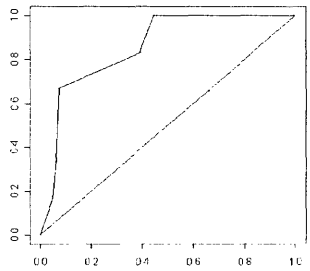
8. Discharge Plan
Test data c-index: 0.80
Test data brier score: 0.18
Test data HL: 104.96
p value: <0.001



9. Physical Examination
Test data c-index: 0.87
Test data brier score: 0.076
Test data HL: 17.87
p value: 0.022

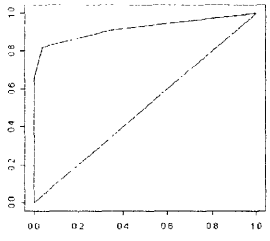


10. Allergies
Test data c-index: 0.82
Test data brier score: 0.022
Test data HL: 20.06
p value: 0.010

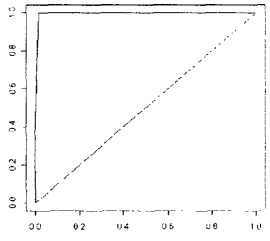


Medical Specialty Departments

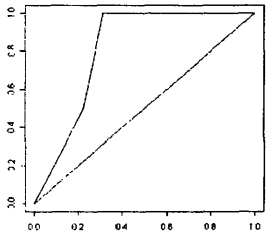
1. History
Test data c-index: 0.87
Test data brier score: 0.126
Test data HL: 13.98
p value: 0.082



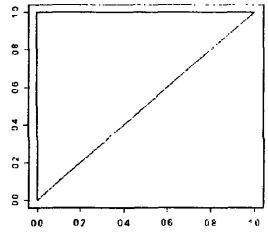
2. Past Medical Or Surgical History
Test data c-index: 0.99
Test data brier score: 0.038
Test data HL: 4.82
p value: 0.776



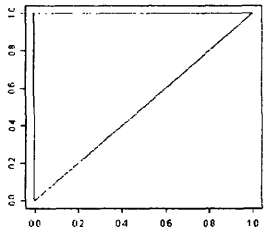
3. Social Personal Family
Test data c-index: 0.76
Test data brier score: 0.033
Test data HL: 5.050
p value: 0.752



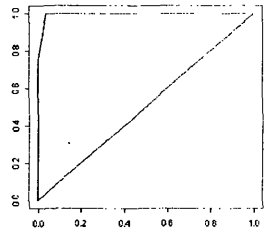
5. Laboratory
Test data c-index: 1
Test data brier score: 0.063
Test data HL: 9.77
p value: 0.281



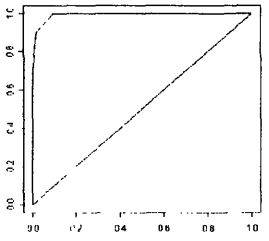
4. Medications
Test data c-index: 1
Test data brier score: 0.020
Test data HL: 5.15
p value: 0.742



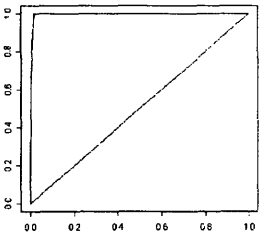
6. Hospital Course
Test data c-index: 0.99
Test data brier score: 0.09
Test data HL: 13.36
p value: 0.100



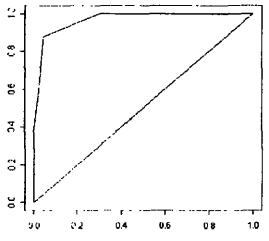
7. Discharge Diagnosis
Test data c-index: 0.99
Test data brier score: 0.11
Test data HL: 20.19
p value: 0.010



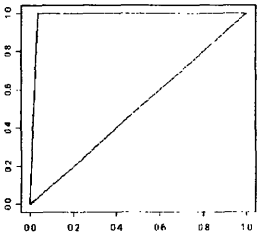
9. Physical Examination
Test data c-index: 0.99
Test data brier score: 0.038
Test data HL: 4.77
p value: 0.782



8. Discharge Plan
Test data c-index: 0.95
Test data brier score: 0.169
Test data HL: 30.64
p value: 0.000

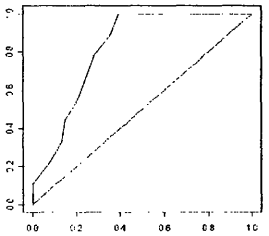


10. Allergies
Test data c-index: 0.98
Test data brier score: 0.019
Test data HL: 5.15
p value: 0.742

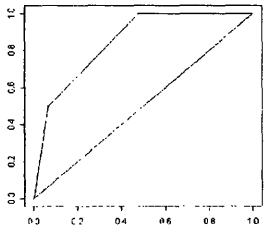


General Medicine Department

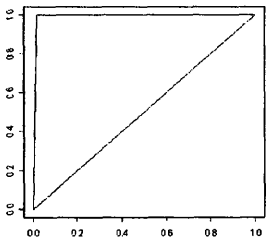
1. History
   Test data c-index: 0.79
   Test data brier score: 0.116
   Test data HL: 11.08
   p value: 0.197



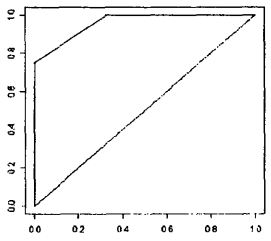
2. Past Medical Or Surgical History
   Test data c-index: 0.73
   Test data brier score: 0.033
   Test data HL: 5.19
   p value: 0.737



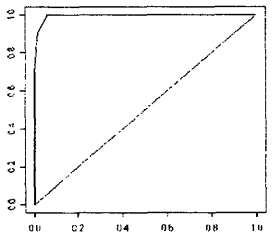
3. Social Personal Family
   Test data c-index: 0.98
   Test data brier score: 0.023
   Test data HL: 7.05
   p value: 0.531



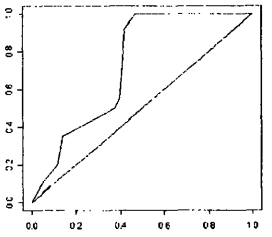
4. Laboratory
   Test data c-index: 0.92
   Test data brier score: 0.05
   Test data HL: 5.28
   p value: 0.727



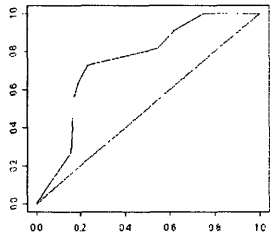
5. Hospital Course
Test data c-index: 0.99
Test data brier score: 0.113
Test data HL: 23.75
p value: 0.003



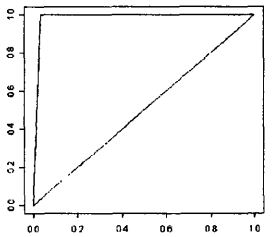
7. Discharge Plan
Test data c-index: 0.70
Test data brier score: 0.235
Test data HL: 52.84
p value: <0.001



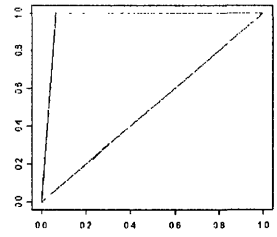
6. Discharge Diagnosis
Test data c-index: 0.70
Test data brier score: 0.149
Test data HL: 39.70
p value: <0.001



8. Physical Examination
Test data c-index: 0.97
Test data brier score: 0.023
Test data HL: 7.08
p value: 0.528

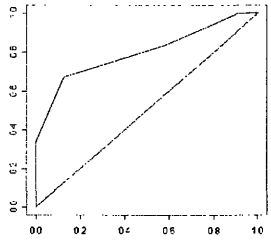


9. Allergies
Test data c-index: 0.94
Test data brier score: 0.020
Test data HL: 5.11
p value: 0.746

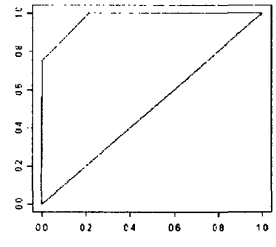


Obstetrics and Gynecology Department

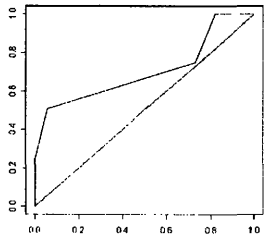
1. History
Test data c-index: 0.71
Test data brier score: 0.13
Test data HL: 9.62
p value: 0.293



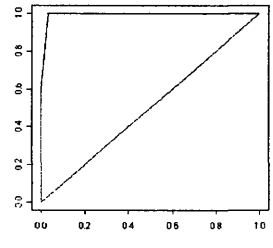
3. Laboratory
Test data c-index: 0.95
Test data brier score: 0.08
Test data HL: 4.69
p value: 0.790



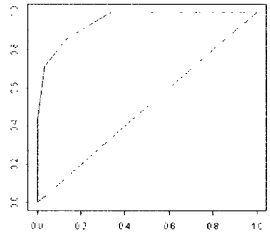
2. Past Medical Or Surgical History
Test data c-index: 0.60
Test data brier score: 0.092
Test data HL: 7.32
p value: 0.503



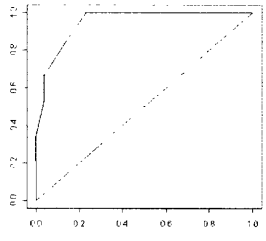
6. Hospital Course
Test data c-index: 0.99
Test data brier score: 0.088
Test data HL: 2.61
p value: 0.956



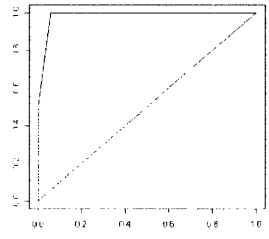
7. Discharge Diagnosis
Test data c-index: 0.92
Test data brier score: 0.13
Test data HL: 4.87
p value: 0.771



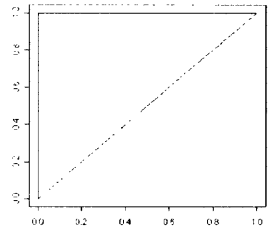
8. Discharge Plan
Test data c-index: 0.93
Test data brier score: 0.11
Test data HL: 3.26
p value: 0.917



9. Physical Examination
Test data c-index: 0.97
Test data brier score: 0.04
Test data HL: 2.67
p value: 0.953

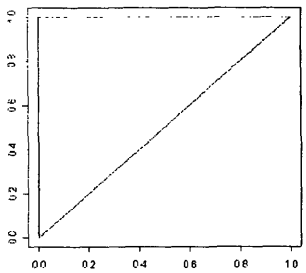


10. Allergies
Test data c-index: 1
Test data brier score: 0.024
Test data HL: 2.63
p value: 0.956

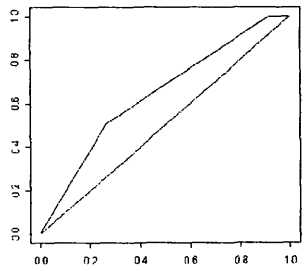


Surgery Department Grouping

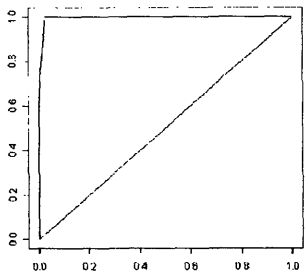
1. History
Test data c-index: 1
Test data brier score: 0.087
Test data HL: 5.44
p value: 0.709



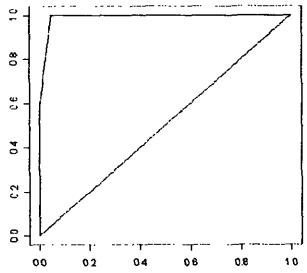
3. Laboratory
Test data c-index: 0.57
Test data brier score: 0.043
Test data HL: 5.42
p value: 0.712



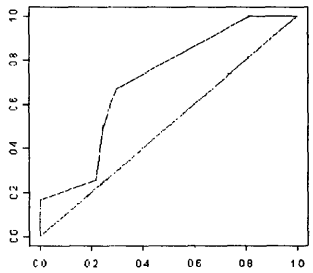
2. Past Medical Or Surgical History
Test data c-index: 0.99
Test data brier score: 0.057
Test data HL: 4.60
p value: 0.800



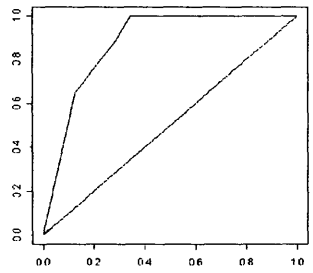
4. Hospital Course
Test data c-index: 0.98
Test data brier score: 0.073
Test data HL: 4.78
p value: 0.781



5. Discharge Diagnosis
Test data c-index: 0.60
Test data brier score: 0.18
Test data HL: 25.71
p value: 0.001



6. Discharge Plan
Test data c-index: 0.81
Test data brier score: 0.20
Test data HL: 19.24
p value: 0.014



Appendix 5

The column "Actual Counts" is a list of the counts of the sections within the grouping that were present. When a section was not present in the documents of that group, it is not presented here.

In the cross-tabulations, the horizontal represents the actual labels and the vertical represents the method.

Discharge Summary Documents

Discharge Summary Entire Set

Actual Counts

H	PH	SPF	M	L	HC	DD	DP	PE	A
47	16	8	2	20	38	56	75	29	6

Method 1 vs Actuals

	H	PH	SPF	M	L	HC	DD	DP	PE	A
H	44	6	1	0	1	1	3	4	5	0
PH	1	3	3	0	1	0	0	1	0	4
SPF	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0
L	1	2	0	1	11	1	2	1	2	0
HC	5	0	0	0	0	27	2	1	0	0
DD	2	2	0	0	1	2	9	11	0	0
DP	3	2	2	0	1	6	40	57	10	1
PE	2	1	2	1	5	1	0	0	12	1
A	0	0	0	0	0	0	0	0	0	0

Method 2 vs Actuals

	H	PH	SPF	M	L	HC	DD	DP	PE	A
H	16	6	2	0	5	12	15	18	6	2
PH	17	6	3	0	5	13	12	15	12	4
SPF	3	0	1	0	3	5	3	7	6	0
M	22	5	2	0	7	12	12	19	13	1
L	10	0	0	0	6	8	6	5	1	2
HC	12	0	0	1	4	10	14	17	1	0
DD	17	9	4	0	5	10	21	33	3	1
DP	24	10	4	2	8	14	27	43	13	2
PE	6	0	3	0	7	8	14	9	16	2
A	3	0	1	0	3	5	2	7	5	0

Method 2 Discriminatory Statistics

	TP	TN	FP	FN
H	37	216	45	10
PH	11	216	76	5
SPF	5	277	23	3
M	2	215	91	0
L	18	268	20	2
HC	36	247	23	2
DD	41	190	62	15
DP	63	149	84	12
PE	22	236	43	7
A	4	280	22	2

General Medicine Department

Actual Counts

H	PH	SPF	L	HC	DD	DP	PE	A
9	2	1	4	10	11	20	1	1

Method 1 vs Actuals

	H	PH	SPF	L	HC	DD	DP	PE	A
H	8	0	0	0	0	6	0	0	0
PH	0	0	0	0	0	1	0	0	0
SPF	1	1	1	0	0	2	2	0	1
L	0	0	0	2	0	0	0	0	0
HC	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0
DP	2	1	0	1	10	2	10	0	0
PE	2	0	0	1	0	0	8	1	0
A	0	0	0	0	0	0	0	0	0

Method 2 vs Actuals

	H	PH	SPF	L	HC	DD	DP	PE	A
H	5	1	0	1	4	5	6	0	0
PH	9	0	0	0	5	3	12	1	1
SPF	2	0	0	0	0	0	0	0	0
L	1	0	1	0	0	0	0	0	1
HC	3	1	0	0	4	2	3	0	0
DD	2	1	0	3	2	3	8	1	0
DP	7	1	0	3	5	6	13	1	0
PE	0	0	0	0	1	1	1	0	0
A	3	0	0	0	0	1	1	0	0

Method 2 vs Method 1

	H	PH	SPF	L	HC	DD	DP	PE	A
H	5	1	2	1	0	0	9	4	0
PH	5	0	6	0	0	0	12	8	0
SPF	2	0	0	0	0	0	0	0	0
L	1	0	2	0	0	0	0	0	0
HC	3	0	0	0	0	0	8	2	0
DD	3	1	1	2	0	0	7	6	0
DP	8	1	1	1	0	0	17	8	0
PE	0	0	1	0	0	0	2	0	0
A	2	0	2	0	0	0	0	1	0

Method 2 Discriminatory Statistics

	TP	TN	FP	FN
H	7	39	15	2
PH	2	32	29	0
SPF	1	61	1	0
L	3	59	0	1
HC	10	50	3	0
DD	8	40	12	3
DP	18	25	18	2
PE	1	60	2	0
A	1	58	4	0

Medicine Specialty Departments

Actual Counts

H	PH	SPF	M	L	HC	DD	DP	PE	A
11	3	2	1	6	8	10	16	3	1

Method 1 vs. Actuals

	H	PH	SPF	M	L	HC	DD	DP	PE	A
H	9	1	0	0	0	0	1	0	0	0
PH	0	1	0	0	0	0	0	0	0	0
SPF	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0
L	0	0	0	1	6	0	0	0	0	0
HC	0	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	8	0	0	0
DP	3	1	2	0	0	8	1	16	2	1
PE	1	0	0	0	0	0	0	0	1	0
A	0	0	0	0	0	0	0	0	0	0

Method 2 vs. Actuals

	H	PH	SPF	M	L	HC	DD	DP	PE	A
H	8	1	0	0	0	0	1	1	0	0
PH	1	3	0	0	0	0	0	0	0	0
SPF	3	0	2	1	5	1	1	7	0	1
M	0	0	0	1	0	0	0	0	0	0
L	1	0	0	0	5	0	0	0	0	0
HC	2	0	0	0	1	6	1	0	0	0
DD	3	0	0	0	0	1	7	3	1	0
DP	0	0	2	0	0	1	2	11	0	0
PE	1	0	0	0	0	0	0	1	2	0
A	0	0	0	1	0	0	0	0	0	1

Method 2 Discriminatory Statistics

	TP	TN	FP	FN
H	9	50	2	2
PH	3	59	1	0
SPF	2	42	19	0
M	1	62	0	0
L	6	57	0	0
HC	8	53	2	0
DD	10	48	5	0
DP	14	45	2	2
PE	3	59	1	0
A	1	61	1	0

Obstetrics and Gynecology Department

Actual Counts

H	PH	L	HC	DD	DP	PE	A
6	4	4	5	7	6	2	1

Method 1 vs. Actuals

	H	PH	L	HC	DD	DP	PE	A
H	4	0	0	0	0	0	0	0
PH	1	1	0	0	0	0	0	1
L	0	0	2	0	0	0	0	0
HC	1	0	0	5	0	0	0	0
DD	2	2	0	0	5	2	1	0
DP	0	1	1	0	2	4	0	0
PE	0	0	1	0	0	0	1	0
A	0	0	0	0	0	0	0	0

Method 2 vs. Actuals

	H	PH	L	HC	DD	DP	PE	A
H	4	0	0	0	2	2	0	0
PH	1	2	0	0	0	1	0	0
L	3	1	3	2	1	1	0	0
HC	3	0	0	3	0	0	0	0
DD	1	0	1	1	5	2	0	0
DP	1	0	2	2	0	4	0	0
PE	0	0	0	0	1	0	2	1
A	0	1	0	0	0	0	0	0

Method 2 Discriminatory Statistics

	TP	TN	FP	FN
H	4	27	4	2
PH	2	31	2	2
L	4	26	7	0
HC	5	31	1	0
DD	6	26	4	1
DP	5	27	4	1
PE	2	33	2	0
A	1	36	0	0

Surgery Department

Actual Counts

H	PH	L	HC	DD	DP	PE
7	3	2	5	12	17	3

Method 1 vs Actuals

	H	PH	L	HC	DD	DP	PE
H	7	2	1	0	0	0	3
PH	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0
HC	0	0	0	2	0	0	0
DD	0	1	1	2	3	2	0
DP	0	0	0	1	9	15	0
PE	0	0	0	0	0	0	0

Method 2 vs Actuals

	H	PH	L	HC	DD	DP	PE
H	7	0	0	0	0	0	0
PH	0	3	0	1	0	0	0
L	0	0	1	0	0	0	2
HC	0	0	0	5	2	0	0
DD	0	1	1	2	8	7	0
DP	0	0	0	0	9	15	0
PE	0	0	1	0	0	0	3

Method 2 Discriminatory Statistics

	TP	TN	FP	FN
H	7	42	0	0
PH	3	45	1	0
L	1	45	2	1
HC	5	42	2	0
DD	8	26	11	4
DP	15	23	9	2
PE	3	45	1	0

History and Physical Examination Documents

Entire Set High Granularity Grouping

Actual Counts

PC	HOPC	PH	Md	Ag	FH	PrH	SH	OH	RS	PE	L	P
21	23	29	22	13	19	9	22	1	19	226	20	25

Method 1 vs Actuals

	PC	HOPC	PH	Md	Ag	FH	PrH	SH	OH	RS	PE	L	P
PC	13	0	0	0	1	0	3	1	0	0	1	0	0
HOPC	1	23	1	1	0	1	1	2	0	3	0	1	0
PH	3	0	15	2	0	1	0	0	0	0	4	1	2
Md	0	0	1	7	3	3	0	1	0	1	2	4	2
Ag	3	0	0	1	2	5	0	2	0	0	0	0	0
FH	0	0	3	0	0	1	0	0	0	0	1	0	1
PrH	0	0	0	0	0	0	0	0	0	0	0	0	0
SH	0	0	0	0	0	0	0	0	0	0	0	0	0
OH	0	0	0	0	0	0	0	0	0	0	0	0	0
RS	0	0	0	0	0	0	0	0	0	0	0	0	0
PE	1	0	7	7	7	8	4	15	1	11	217	10	15
L	0	0	0	0	0	0	0	1	0	0	0	0	0
P	0	0	2	4	0	0	1	0	0	4	1	4	5

Method 2 vs Actuals

	PC	HOPC	PH	Md	Ag	FH	PerH	SH	OH	RS	PE	L	P
PC	21	0	15	0	1	8	3	4	0	0	4	1	3
HOPC	0	23	1	1	0	0	0	1	0	0	0	0	0
PH	13	6	22	13	3	9	1	6	0	3	13	5	7
Md	4	11	8	13	6	11	2	11	1	5	33	6	5
Ag	13	0	20	6	8	12	3	10	0	2	15	2	5
FH	2	0	5	0	0	5	0	4	0	1	2	1	2
PrH	16	22	17	12	4	11	7	17	1	16	207	13	20
SH	4	0	6	2	0	5	0	10	0	8	19	3	7
OH	10	21	10	10	4	8	4	15	1	14	191	10	17
RS	4	0	6	2	0	5	0	10	0	10	27	4	8
PE	0	0	2	3	4	3	3	8	0	4	176	6	7
L	4	14	7	11	4	9	4	13	1	10	167	13	16
P	4	4	8	10	2	5	2	9	0	12	51	14	19

(Optimal) Threshold-associated values

	TP	TN	FP	FN
PC	21	389	39	0
HOPC	23	423	3	0
PH	22	341	79	7
Md	13	324	103	9
Ag	8	348	88	5
FH	5	413	17	14
PrH	7	84	356	2
SH	10	373	54	12
OH	1	134	314	0
RS	10	364	66	9
PE	176	183	40	50
L	13	169	260	7
P	19	303	121	6

Entire Set (Low Granularity Grouping)

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
34	55	132	91	38	456	93

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	53	0	0	0	0	0
PhMdAg	14	0	67	11	2	10	13
PFSO	1	0	11	23	5	6	7
RS	0	0	0	0	0	0	0
PE	19	2	48	57	29	437	36
LP	0	0	6	0	2	3	37

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	29	41	52	26	11	14	11
HOPC	0	54	0	0	0	1	1
PhMdAg	16	0	92	23	4	36	37
PFSO	10	2	57	59	20	179	25
RS	11	2	19	23	14	134	12
PE	16	1	32	38	18	368	18
LP	3	4	49	24	6	77	72

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	29	710	155	5
HOPC	54	842	2	1
PhMdAg	92	651	116	40
PFSO	59	515	293	32
RS	14	660	201	24
PE	368	320	123	88
LP	72	643	163	21

General Medicine

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
6	10	26	17	7	86	22

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	7	0	0	0	0	0
PhMdAg	3	0	18	6	2	1	4
PFSO	1	0	0	5	1	2	1
RS	0	0	0	0	0	0	0
PE	1	2	7	6	3	83	5
LP	1	1	1	0	1	0	12

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	3	9	19	9	5	81	7
HOPC	0	10	2	0	0	1	0
PhMdAg	4	0	21	7	2	3	5
PFSO	5	8	9	13	4	11	5
RS	2	8	11	6	5	76	6
PE	1	0	5	5	3	79	2
LP	1	1	2	1	1	3	18

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	3	38	130	3
HOPC	10	161	3	0
PhMdAg	21	127	21	5
PFSO	13	115	42	4
RS	5	58	109	2
PE	79	72	16	7
LP	18	143	9	4

Cardiology

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
4	5	14	8	3	48	12

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	5	0	1	0	0	1
PhMdAg	4	0	10	2	0	0	4
PFSO	0	0	0	1	0	2	0
RS	0	0	0	0	1	0	1
PE	0	0	4	4	1	45	0
LP	0	0	0	0	1	1	6

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	4	5	14	4	3	42	9
HOPC	0	5	0	0	0	0	0
PhMdAg	2	0	10	1	0	0	2
PFSO	2	1	0	5	1	6	2
RS	1	4	1	0	2	3	2
PE	0	0	4	2	1	44	0
LP	0	0	0	1	2	2	9

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	4	13	77	0
HOPC	5	89	0	0
PhMdAg	10	75	5	4
PFSO	5	74	12	3
RS	2	80	11	1
PE	44	39	7	4
LP	9	77	5	3

Gastroenterology

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
2	2	5	2	3	28	4

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	2	0	0	0	0	0
PhMdAg	2	0	5	0	0	0	0
PFSO	0	0	0	1	1	0	0
RS	0	0	0	0	0	0	0
PE	0	0	0	1	2	28	1
LP	0	0	0	0	0	0	3

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	1	0	0
HOPC	0	2	0	0	0	0	0
PhMdAg	0	0	5	0	0	0	0
PFSO	2	0	2	2	1	0	0
RS	0	0	0	0	2	0	0
PE	0	0	0	0	1	27	1
LP	0	0	0	0	0	0	3

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	0	43	1	2
HOPC	2	44	0	0
PhMdAg	5	41	0	0
PFSO	2	39	5	0
RS	2	43	0	1
PE	27	16	2	1
LP	3	42	0	1

Mixed Medical Specialties

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
3	7	10	11	2	46	8

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	7	0	0	0	0	0
PhMdAg	2	0	5	2	1	0	0
PFSO	0	0	0	7	0	1	0
RS	0	0	0	0	0	0	0
PE	1	0	5	2	1	45	8
LP	0	0	0	0	0	0	0

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	3	7	7	7	0	46	8
HOPC	0	7	0	0	0	0	0
PhMdAg	3	2	8	6	2	1	3
PFSO	0	0	0	9	0	1	0
RS	0	0	0	0	1	0	0
PE	0	0	3	0	0	44	4
LP	0	1	3	2	0	11	5

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	3	9	75	0
HOPC	7	80	0	0
PhMdAg	8	60	17	2
PFSO	9	75	1	2
RS	1	85	0	1
PE	44	34	7	2
LP	5	62	17	3

Neurology

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
2	5	12	7	3	35	9

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	0	0	0	0	0	0
PhMdAg	2	2	9	2	3	6	0
PFSO	0	0	0	1	0	0	0
RS	0	0	0	0	0	0	0
PE	0	1	3	3	0	28	1
LP	0	2	0	1	0	1	8

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	1	5	1	0	0	23	5
HOPC	0	5	0	0	0	1	1
PhMdAg	2	2	10	2	3	7	0
PFSO	1	3	9	6	2	6	2
RS	1	5	10	4	3	30	9
PE	0	0	2	3	0	29	1
LP	0	3	0	1	0	2	9

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	1	37	34	1
HOPC	5	66	2	0
PhMdAg	10	45	16	2
PFSO	6	43	23	1
RS	3	11	59	0
PE	29	32	6	6
LP	9	58	6	0

Respiratory Medicine

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
3	4	12	13	4	46	6

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	0	0	0	0	0	0
PhMdAg	3	4	7	3	0	0	0
PFSO	0	0	0	8	0	0	1
RS	0	0	1	0	4	1	1
PE	0	0	3	2	0	45	0
LP	0	0	1	0	0	0	4

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	3	0	10	11	4	45	5
HOPC	0	4	0	0	0	0	0
PhMdAg	3	0	8	1	0	0	0
PFSO	0	0	0	10	0	0	2
RS	0	0	1	0	4	0	1
PE	0	0	2	0	0	45	0
LP	0	0	1	4	2	4	6

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	3	10	75	0
HOPC	4	84	0	0
PhMdAg	8	72	4	4
PFSO	10	73	2	3
RS	4	82	2	0
PE	45	40	2	1
LP	6	71	11	0

Surgery

Actual Counts

PC	HOPC	PhMdAg	PFSO	RS	PE	LP
2	3	3	1	1	26	5

Method 1 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	0	0	0	0	0	0	0
HOPC	0	0	0	0	0	0	0
PhMdAg	2	3	3	0	0	0	0
PFSO	0	0	0	0	0	0	0
RS	0	0	0	0	0	0	0
PE	0	0	0	1	1	26	5
LP	0	0	0	0	0	0	0

Method 2 vs Actuals

	PC	HOPC	PhMdAg	PFSO	RS	PE	LP
PC	2	0	2	1	1	25	4
HOPC	0	3	0	0	0	0	0
PhMdAg	2	0	3	0	0	0	0
PFSO	0	0	0	1	1	19	1
RS	0	0	0	0	1	1	0
PE	0	0	0	1	0	24	1
LP	0	0	0	0	0	1	4

(Optimal) Threshold associated values

	TP	TN	FP	FN
PC	2	6	33	0
HOPC	3	38	0	0
PhMdAg	3	36	2	0
PFSO	1	19	21	0
RS	1	39	1	0
PE	24	13	2	2
LP	4	35	1	1