UNIVERSITI PUTRA MALAYSIA

BAYESIAN NETWORK MODELING OF GASTROINTESTINAL BLEEDING

NAZZIWA AISHA

FS 2013 73
BAYESIAN NETWORK MODELING OF GASTROINTESTINAL BLEEDING

By

NAZZIWA AISHA

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science.

November 2013
COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright ©Universiti Putra Malaysia
In the name of Allah, Most Gracious, Most Merciful

Dedicated to:

My beloved father & mother

&

My beloved spouse: Lwere Kamada
Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Master of Science

BAYESIAN NETWORK MODELING OF GASTROINTESTINAL BLEEDING

By

NAZZIWA AISHA

November 2013

Chair: Mohd Bakri Adam, Ph.D.
Faculty: Science

Acute gastrointestinal bleeding (GIB) is a common medical emergency with 50-150 per 100,000 people admitted per year. Although 80 percent of GIB cases stop spontaneously, it is important to determine the source of bleeding and establish a diagnosis such that possible recurrences are prevented and that the most suitable management may be given in future episodes. In the emergency room, when a patient shows signs of hematemesis (vomiting of red blood), it is obvious that the patient has upper gastrointestinal bleeding. In the absence of hematemesis however, the source of bleeding remains unclear. While the diagnosis of GIB is best done by a gastroenterologist, it is not always feasible, due to scarcity of resources and time. A reliable classification model would be very helpful in diagnosing patients more efficiently and effectively targeting the scarce resources.

Current review of the literature, did not reveal any model that predicts the source of GIB in the absence of hematemesis. This thesis uses a graphical modeling approach, specifically Bayesian networks, to model the different outcomes of GIB. One key advantage of Bayesian network models in this context is their ability to predict the outcome with partial observations on variables or attributes. The four outcome variables predicted are: source of bleeding, need for urgent blood resuscitation, need for urgent endoscopy, and disposition. Performance of the models is assessed by classification or prediction accuracy, area under curves, sensitivity and specificity values. The Bayesian network models provide good accuracy for the prediction of the source of bleeding and need for urgent blood resuscitation but did not do well on predicting need for urgent endoscopy, and disposition. The models require further validation if they are to be used in clinical settings.
Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Sarjana Sains

PEMODELAN RANGKAIAN BAYESIAN PENDARAHAN GASTROUSUS

Oleh

NAZZIWA AISHA

November 2013

Pengerusi: Mohd Bakri Adam, Ph.D.
Fakulti: Sains

Pendarahan gastrousus akut (GIB) adalah penyakit kecemasan biasa dengan 50-150 daripada 100,000 orang dimasukkan ke wad dalam tempoh setahun. Walaupun 80 peratus daripada kes-kes GIB berhenti secara spontan, adalah penting untuk menentukan punca pendarahan dan melakukan diagnosis supaya dapat menghalang perkara yang sama berulang dan langkah yang paling sesuai diambil pada masa akan datang. Di dalam bilik kecemasan, apabila pesakit menunjukkan tanda-tanda hematemesis (muntah darah merah), ia adalah jelas bahawa pesakit mempunyai pendarahan gastrousus atas. Walaubagaimanapun, ketiadaan hematemesis menyebabkan punca pendarahan tidak jelas. Walaupun diagnosis GIB adalah lebih baik dilakukan oleh gastroenterologi, ia tidak kerap dilaksanakan kerana kekurangan sumber dan masa. Satu model klasifikasi yang boleh dipercayai akan banyak membantu dalam mendiagnosis pesakit dengan lebih cekap dan berkesan berdasarkan kepada sumber-sumber yang terhad.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Dr. Mohd Bakri Adam, for his patience and guidance not only on the thesis but also with career decisions and for imparting his knowledge of teaching to me. I am also grateful for the opportunity he gave me to be a demonstrator here at University Putra Malaysia for one semester and for giving me the typesetting skills of LaTex. I am thankful that I have learned so much from Dr. Bakri and I feel that I have grown intellectually.

I would like to say a big thank you to Dr. Shamarina Shohaimi. It has been a pleasure to work with her. Without Dr. Shamarina, I would not have the statistical experience I do now. I have used and will continue to use the statistical software skills given to me by her.

I also wish to thank Dr. Michael D Witting and Professor Adrienne Chu for providing the data I used in this thesis.

I would especially like to thank my family and friends for all of their support, in particular, my parents (Mbabaali Sulaiman and Nanfuka Hadijah Mbabaali), husband (Dr Lwere Kamada), sister (Hadijah, Grace, Shifah, Hanifa), brothers (Ibrahim, Siraje, Abdunulu, Brian) and friends (Shakila Raha, Mariam, Shadijah, Aidah, Kiran, Rukia, Balqis, Phang, Waziri, Nur Aini Jammaludin, Atiya Wims, Yusuf, Ali, Arzuuka, Hadijah, Hajarah, Rehi and Abdulrahman). They have always been there for me, no matter what, and have always believed in me.
I certify that a Thesis Examination Committee has met on 28 November 2013 to conduct the final examination of Nazziwa Aisha on her thesis entitled “Bayesian Network Modeling of Gastrointestinal Bleeding” in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

**Zanariah binti Abdul Mjid, Ph.D.**
Professor
Fakulti of Science Universiti Putra Malaysia
(Chairperson)

**Noor Akma bt Ibrahim, Ph.D.**
Professor
Fakulti of Science Universiti Putra Malaysia
(Internal Examiner)

**Mohd Rizam bt Abu Bakar, Ph.D.**
Professor
Fakulti of Science Universiti Putra Malaysia
(Internal Examiner)

**Yong Zulina Zubairi, Ph.D.**
Professor
Center of fundamental science University of Malaya
Malaysia
(External Examiner)

---

**NORITAH OMAR, PhD**
Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:
This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of **Master of Science**. The members of the Supervisory Committee were as follows:

**Mohd Bakri Adam, PhD**  
Senior Lecturer  
Faculty of Science  
Universiti Putra Malaysia  
(Chairperson)

**Shamarina Shohaimi, Ph.D.**  
Senior Lecturer  
Faculty of Science  
Universiti Putra Malaysia  
(Member)

---

**BUJANG BIN KIM HuAT, Ph.D.**  
Professor and Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date:

vii
DECLARATION

Declaration by graduate student
I hereby confirm that:

• this thesis is my original work;

• quotations, illustrations and citations have been duly referenced;

• this thesis has not been submitted previously or concurrently for any other degree at any other institutions;

• intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;

• written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;

• there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: __________________________ Date: __________________________

Name and Matric No.: ____________________________________________
Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;

- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ___________________________  Signature: ___________________________
Name of Chairman of  Name of Member of
Supervisory Supervisory
Committee: ________________  Committee: ________________
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPYRIGHT</td>
<td>i</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>APPROVAL</td>
<td>vi</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xvi</td>
</tr>
</tbody>
</table>

CHAPTER

1 INTRODUCTION

1.1 Background 1
1.2 Objectives of the thesis 4
1.3 Methods 4
1.4 Expected outcome 5
1.5 Limitations of the study 5
1.6 Organization of the thesis 5

2 BAYESIAN NETWORKS

2.1 Introduction 6
2.2 Bayesian Network Formalism 7
2.3 Representation 9
  2.3.1 Qualitative Component of a Bayesian Network 9
  2.3.2 Quantitative Component of a Bayesian Network 11
2.4 Inference 12
  2.4.1 Prediction using a Bayesian Network 12
  2.4.2 Diagnosis Using a Bayesian Network 12
2.5 Modeling process of a Bayesian network 13
  2.5.1 Manual Construction 13
  2.5.2 Learning a Bayesian Network From Data. 14
2.6 Model selection by scoring 16
  2.6.1 Bayesian Quality Measures 16
  2.6.2 Minimum Description Length 16
2.7 Model selection by searching 17
2.8 Bayesian Network Classifiers. 19
  2.8.1 Naive Bayes Classifier 19
  2.8.2 Tree Augmented Naive Bayes Classifier 21
  2.8.3 Bayes Network Augmented Naive Bayes Classifier 22
3 BAYESIAN NETWORK MODELING OF GASTROINTESTINAL BLEEDING 23
  3.1 Introduction 23
    3.1.1 Source of bleeding 23
    3.1.2 Blood resuscitation 24
    3.1.3 Early endoscopy 24
    3.1.4 Disposition 25
  3.2 Data 26
    3.2.1 Method of Ascertaining Outcomes 26
    3.2.2 Explanatory and Outcome Variables 26
  3.3 Methods 28
    3.3.1 Model Evaluation and Validation Step 28
    3.3.2 Sensitivity Analysis 29
  3.4 Results 30
    3.4.1 Source of GIB 30
    3.4.2 Need for urgent blood resuscitation 36
    3.4.3 Need for urgent endoscopy 42
    3.4.4 Disposition 48
  3.5 Conclusion 53

4 CLASSIFICATION OF GASTROINTESTINAL BLEEDING IN PATIENTS WITHOUT HEMATEMESIS 54
  4.1 Introduction 54
  4.2 Data 54
  4.3 Model Information 55
  4.4 Methods 56
  4.5 Results 58
    4.5.1 Results of Naive Bayes model 58
    4.5.2 Comparison with other models 61
  4.6 Conclusion 64

5 CONCLUSION AND FUTURE WORK 65
  5.1 Conclusion 65
  5.2 Future work 67

REFERENCES/BIBLIOGRAPHY 68
APPENDICES 79
BIODATA OF STUDENT 84
LIST OF PUBLICATIONS 85
LIST OF TABLES

Table                                                                 Page
1.1 Key early decisions in the medical management of acute upper       3
  gastrointestinal bleeding.
2.1 Conditional probability table of the output node Turn up.         7
3.1 Outcome variables.                                                27
3.2 Evaluation Step-Model performance in prediction of the source of  32
  bleeding.
3.3 Validation Step - Model performance in the prediction of the source
  of bleeding.                                                        32
3.4 Sensitivity of source to a finding at another node in a Naive Bayes
  model.                                                               32
3.5 Sensitivity of source to a finding at another node in a tree aug-
  mented Naive Bayes model.                                            33
3.6 Sensitivity of source to a finding at another node in the Bayes
  network augmented Naive Bayes model.                                33
3.7 Evaluation Step - Model performance for resuscitation.             37
3.8 Validation Step - Model performance for resuscitation.             40
3.9 Sensitivity of resuscitation to a finding at another node in a Naive
  Bayes model.                                                        40
3.10 Sensitivity of resuscitation to a finding at another node in a tree
  augmented naive Bayes.                                               41
3.11 Sensitivity of resuscitation to a finding at another node in a Bayes
  net augmented Bayes model.                                           41
3.12 Evaluation Step - Model performance for endoscopy.                45
3.13 Validation Step - Model performance for endoscopy.                46
3.14 Sensitivity of endoscopy to a finding at another node in a naive
  Bayes model.                                                        46
3.15 Sensitivity of endoscopy to a finding at another node: Tree aug-
  mented naive Bayes.                                                 47
3.16 Sensitivity of endoscopy to a finding at another node: Bayes net augmented Bayes model.

3.17 Accuracy and AUC for GIB scores


3.20 Sensitivity of disposition to a finding at another node in a naive Bayes model.

3.21 Sensitivity of disposition to a finding at another node in a tree augmented naive Bayes model.

3.22 Sensitivity of disposition to a finding at another node in a Bayes net augmented Bayes model.

4.1 Conditional probability table for the node HGIB in the Naive Bayes model in percentage.

4.2 Conditional probability table for the node BUN/CR in the Naive Bayes model in percentage.

4.3 Sensitivity of source to findings at other nodes.

4.4 Individual factor percent probabilities of source of bleeding.

4.5 Posterior probabilities of GIB given more than one observation from patients.

4.6 Model Performance on the real gastrointestinal bleeding data.

4.7 Simulated study results: Independent gastrointestinal bleeding data.

4.8 Simulated study results: Correlated gastrointestinal bleeding data.

A.1 Explanatory variables in the data set.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Gastrointestinal Tract.</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Endoscopy.</td>
<td>2</td>
</tr>
<tr>
<td>2.1 A Bayesian network for predicting turn up among substance abusers in a longitudinal study.</td>
<td>6</td>
</tr>
<tr>
<td>2.2 An example of a Bayesian network with 5 nodes showing the conditional probability distributions for each node.</td>
<td>9</td>
</tr>
<tr>
<td>2.3 The three types of connections among variables.</td>
<td>9</td>
</tr>
<tr>
<td>2.4 The Bayesian network for the burglary example.</td>
<td>10</td>
</tr>
<tr>
<td>2.5 (a); A naive Bayesian network and (b); A tree augmented Bayesian network.</td>
<td>20</td>
</tr>
<tr>
<td>3.1 Naive Bayes model for predicting the source of bleeding.</td>
<td>30</td>
</tr>
<tr>
<td>3.2 Tree augmented Naive Bayes for predicting the source of bleeding.</td>
<td>31</td>
</tr>
<tr>
<td>3.3 Bayes net augmented Naive Bayes for predicting the source of bleeding.</td>
<td>31</td>
</tr>
<tr>
<td>3.4 Posterior probability of the source of bleeding in the Naive Bayes with evidence of hematemesis.</td>
<td>34</td>
</tr>
<tr>
<td>3.5 Posterior probability of the source of bleeding in the Tree augmented Naive Bayes with evidence of hematemesis.</td>
<td>35</td>
</tr>
<tr>
<td>3.6 Posterior probability of the source of bleeding in the Bayes net augmented Naive Bayes with evidence of hematemesis.</td>
<td>35</td>
</tr>
<tr>
<td>3.7 Naive Bayes model for predicting the need for urgent blood resuscitation.</td>
<td>36</td>
</tr>
<tr>
<td>3.8 Posterior distribution of the Naive Bayes model when there is evidence that patient has syncope.</td>
<td>37</td>
</tr>
<tr>
<td>3.9 Naive Bayes model with evidence that patient does not have syncope.</td>
<td>38</td>
</tr>
<tr>
<td>3.10 Naive Bayes model with evidence that patient has hematocrit drop of between 3.5 to 7.5, heart rate $\geq 88.5$ and no melena.</td>
<td>38</td>
</tr>
</tbody>
</table>
3.11 Tree augmented naive Bayes model for predicting the need for urgent blood resuscitation. 39

3.12 Bayes net augmented Bayesian model for predicting the need for urgent blood resuscitation. 39

3.13 Naive Bayes model for endoscopy. 42

3.14 Tree augmented naive Bayes for endoscopy. 43

3.15 Bayes net augmented Bayes model for endoscopy. 43

3.16 Naive Bayes for endoscopy with evidence that an endoscopy was done. 44

3.17 Naive Bayes for endoscopy with evidence that a patient has orthostasis and HR $\geq 93$ percent. 44

3.18 Naive Bayes for endoscopy when a patient has no orthostasis and HR $\leq 93$ percent. 45

3.19 Naive Bayes for disposition. 49

3.20 Tree augmented naive Bayes for disposition. 49

3.21 Bayes net augmented naive Bayes model for disposition. 50

3.22 Posterior probability of the nodes in the naive Bayes model when a patient is in ICU. 51

3.23 Posterior probability of the nodes in the naive Bayes model when a patient is home. 51

4.1 Knowledge flow of the model building process. 57

4.2 Naive Bayes for predicting source of gastrointestinal bleeding in absence of hematemesis. 58

4.3 Posterior probability of the source of bleeding when BUN/CR $\geq 30$. 60
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Prediction accuracy</td>
</tr>
<tr>
<td>ANB</td>
<td>Augmented naive Bayes</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under receiver operating characteristic curves</td>
</tr>
<tr>
<td>BAN</td>
<td>Bayes net augmented naive Bayes</td>
</tr>
<tr>
<td>BN</td>
<td>Bayesian network</td>
</tr>
<tr>
<td>BNCs</td>
<td>Bayesian network classifiers</td>
</tr>
<tr>
<td>BUN/CR</td>
<td>Ratio of blood urea nitrogen to creatinine.</td>
</tr>
<tr>
<td>CPT</td>
<td>Conditional probability table</td>
</tr>
<tr>
<td>Cr</td>
<td>Creatinine</td>
</tr>
<tr>
<td>CV</td>
<td>Cross validation</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed acyclic graph</td>
</tr>
<tr>
<td>DBP</td>
<td>Diastolic blood pressure</td>
</tr>
<tr>
<td>ED</td>
<td>Emergency department</td>
</tr>
<tr>
<td>GI</td>
<td>Gastro intestinal</td>
</tr>
<tr>
<td>GIB</td>
<td>Acute gastrointestinal bleeding</td>
</tr>
<tr>
<td>Hct</td>
<td>Hematocrit</td>
</tr>
<tr>
<td>HGIB</td>
<td>History of gastrointestinal bleeding</td>
</tr>
<tr>
<td>HR</td>
<td>Heart rate</td>
</tr>
<tr>
<td>ICD-9</td>
<td>International Statistical Classification of Diseases</td>
</tr>
<tr>
<td>ICU</td>
<td>Intensive care unit</td>
</tr>
<tr>
<td>J48</td>
<td>Decision tree</td>
</tr>
<tr>
<td>KNN</td>
<td>K nearest neighbour</td>
</tr>
<tr>
<td>LGB</td>
<td>Lower gastrointestinal bleeding</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum a posteriori</td>
</tr>
<tr>
<td>MGIB</td>
<td>Middle gastrointestinal bleeding</td>
</tr>
<tr>
<td>NB</td>
<td>Naive Bayes</td>
</tr>
<tr>
<td>NBC</td>
<td>Naive Bayes Classifier</td>
</tr>
<tr>
<td>NGlavage</td>
<td>Naso gastric lavagePlatelet count</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative predictive value</td>
</tr>
<tr>
<td>NSAID</td>
<td>Non steroidal anti-inflammatory drug</td>
</tr>
<tr>
<td>NSAIDs</td>
<td>Nonsteroidal anti-inflammatory drugs</td>
</tr>
<tr>
<td>Plt</td>
<td>Platelet count</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive predictive value</td>
</tr>
<tr>
<td>RF</td>
<td>Random forrest</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristics</td>
</tr>
<tr>
<td>SBP</td>
<td>Systolic blood pressure</td>
</tr>
<tr>
<td>SN</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>SP</td>
<td>Specificity</td>
</tr>
<tr>
<td>TAN</td>
<td>Tree augmented naive Bayes</td>
</tr>
<tr>
<td>UGIB</td>
<td>Upper gastrointestinal bleeding</td>
</tr>
<tr>
<td>Weka</td>
<td>Waikato Environment for Knowledge Analysis</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 Background

Acute gastrointestinal bleeding (GIB) is a potentially life-threatening abdominal emergency that remains a common cause of hospitalization (Fallah et al., 2000; Cerulli, 2013). According to Kanwal et al. (2010) nonvariceal upper gastrointestinal bleeding results in 400,000 hospital admissions per year, costing more than $2 billion annually in the US. The rising incidence of GIB is associated with increasing non steroidal anti-inflammatory drug (NSAID) use such as aspirins and the high prevalence (64 percent) of Helicobacter pylori infection in patients with peptic ulcer bleeding (Sánchez-Delgado et al., 2011; Simon et al., 2013). Previous research has shown that the use of NSAIDs, for doses $\geq 325$ mg/d increases the risk of bleeding from peptic ulcer disease up to fivefold (Lanas et al., 2006). The advent of new drugs and endoscopic interventions have however reduced the number of surgeries for bleedings from peptic ulcer, but not mortality (Di Fiore et al., 2005). GIB is twice as common in men as in women and increases in prevalence with age (Longstreth, 1995).

The signs of GIB depend on the rate of blood loss. Patients with GIB will show signs of iron-deficiency anemia or hemoccult-positive stools when there is microscopic blood loss (Manning-Dimmitt et al., 2005). Some GIB patients will have coffee-ground emesis, which is the vomiting of altered black blood, hematemesis which is defined as the vomiting of fresh blood, melena which is the passing of black tarry stools or hematochezia which is is the passing of red blood via the rectum (usually from the lower gastrointestinal tract, but sometimes from a briskly bleeding upper gastrointestinal source) (Watson and Church, 2013; Palmer, 2002). These presentation signs are important in the diagnosis of GIB. When patients arrive in the emergency room with GIB, they are often met by a front line physician (not a gastroenterologist). The physician must make a prompt and accurate clinical assessment of the patient. This assessment involves checking the severity of the bleeding, its acuity, activity and location or source of bleeding. The source of bleeding could be from the upper part of the gastrointestinal tract or digestive tract (esophagus, stomach, and first part of the small intestine) as shown in Figure 1.1, and is known as upper gastrointestinal bleeding (UGIB) or from the lower part (includes much of the small intestine, large intestine or bowels, rectum, and anus) known as lower gastrointestinal bleeding (LGIB). Physician’s make several decisions based on their prediction of the source of bleed (UGIB or LGIB). The source of bleed will determine the management of bleed i.e., consultant to be assigned and the timing for consultation.

When UGIB is suspected, the diagnostic tool of choice is esophagogastroduodenoscopy or upper endoscopy and is done by a gastroenterologist while if LGIB is suspected, a colonoscopy or lower endoscopy is done, Figure 1.2, or arteriography if the bleeding is too brisk (Zuccaro, 1998; Eisen et al., 2001). The evaluation and management of LGIB may also necessitate consultation with a nuclear medicine specialist, a general surgeon or an intervention radiologist (Brackman et al., 2003).
Taking into consideration the numerous diagnostic and management options, an incorrect prediction of the source of bleeding can result in unwarranted consultations and procedures and in delays in delivery of proper care. Jensen et al. (1988) found out as many as 11 percent of patients suspected initially to have LGIB are ultimately found to have UGIB. Physicians also make several other decisions based on the initial assessment. They determine the endoscopy timing, consultation requirements, resuscitation requirements, triage, and prognostication (Kollef et al., 1997). Other decisions are listed in Table 1.1, (Cappell and Friedel, 2008).

Although a gastroenterologist would be the most preferred to diagnose every single case of GIB, this is not always feasible due to time and cost constraint. Physicians are thus left to the task of identifying patients who are at risk of adverse outcomes (re bleeding, death) and to make key decisions on preventing these adverse outcomes. There are several clinical and laboratory variables that are available in the first few hours of evaluation in the emergency department.
Table 1.1: Key early decisions in the medical management of acute upper gastrointestinal bleeding.

<table>
<thead>
<tr>
<th>Triage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admit to hospital versus discharge from emergency room</td>
</tr>
<tr>
<td>Admit to ICU versus monitored bed versus unmonitored hospital bed</td>
</tr>
<tr>
<td>Emergency versus routine gastroenterology consult</td>
</tr>
<tr>
<td>Surgical consult or not</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intensive monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasogastric tube insertion or not</td>
</tr>
<tr>
<td>Central venous line or Swann-Ganz catheter or not</td>
</tr>
<tr>
<td>Foley insertion or not</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General supportive therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endotracheal intubation or not</td>
</tr>
<tr>
<td>Transfuse packed erythrocytes or not</td>
</tr>
<tr>
<td>Transfuse other blood products or not</td>
</tr>
<tr>
<td>PPI therapy or not</td>
</tr>
<tr>
<td>Octreotide therapy or not</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endoscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency versus elective endoscopy</td>
</tr>
<tr>
<td>EGD versus colonoscopy</td>
</tr>
<tr>
<td>Endoscopic therapy or not</td>
</tr>
<tr>
<td>Specific modality of endoscopic therapy</td>
</tr>
</tbody>
</table>

Abbreviations: EGD, esophagogastroduodenoscopy; PPI, proton pump inhibitor. (Cappell and Friedel, 2008)

that can be used to identify patients at risk of adverse outcomes. (Adamopoulos et al., 2003) found that variables like haemoglobin level, haemodynamic status, that are available at presentation can be used to distinguish between patients in need of urgent endoscopy and those who don’t. Using symptoms alone however, physicians can predict the location of GIB with only 40 percent accuracy compared to endoscopy on other studies. Physicians use scoring systems to help them identify patients who are at risk of adverse outcomes and their fore in need of urgent treatment. i.e., urgent blood resuscitation or endoscopy. Several scoring systems and models have been developed to identify patients at risk of adverse outcomes. The Rockall score (Rockall et al., 1996b) uses clinical data from endoscopic findings to identify UGIB patients who are at high risk. The Blatchford score (Blatchford et al., 2000), is applicable for UGIB patients as well.

Although a number of scoring systems have been developed, these scoring systems have focused on identifying patients with UGIB that are at risk of adverse outcomes. Little attention has been given to the identification of LGIB patients that are at risk of adverse outcomes. Classification or prediction models need to
be developed to help identify those patients in need of urgent resuscitation and endoscopic intervention (Baradarian et al., 2004; Jensen et al., 2000). There are several classification models that have been developed that can be used to predict several outcomes of GIB in patients (Chu et al., 2008). The Bayesian network (BN) models have however not been considered as predictors of outcomes of GIB. In this thesis, Bayesian network (BN) models are developed, to help in the diagnosis or prediction of outcomes of GIB. BNs have the following advantages over other methods.

1. BN’s represent powerful tools for graphically representing the relationships among a set of variables and for dealing with uncertainties.

2. The graphical structure of the BN provides a simple way to visualize the relationships between the variables.

3. BN’s are interactive and offer a graphical modeling mechanism that researchers can use to understand the behavior of a system or situation. When variables are observed or more evidence is obtained, the information can be propagated throughout the model and the effects on particular variables of interest can be inspected.

1.2 Objectives of the thesis

The objectives of the thesis are to:

1. Develop BN models for predicting the source of bleeding, need for urgent endoscopy, need for urgent blood resuscitation and disposition in patients with GIB.

2. Develop a naive Bayes model (NB) to predict the source of bleeding in GIB patients presenting without hematemesis.

3. Compare the NB model for predicting the source of bleeding in patients presenting without hematemesis with other classification models.

1.3 Methods

In order to achieve the objectives, a BN approach to modeling GIB is selected after a broad review of numerous approaches which include; rule induction, traditional statistics, random forest, and other modeling algorithms. Three BN models are considered: the naive Bayes (NB) model, tree augmented naive Bayes model (TAN), and Bayes network augmented naive Bayes (BAN). These models are developed with the help of two modeling shells. The first is Weka modeling shell which is selected because of its ability to incorporate case files, provide the classification accuracy and because it has a variety of other classifiers with which we can use to compare our model. Weka cannot perform sensitivity analysis and provides no option for updating the model. These two functions are found in the Netica (Norsys) modeling shell. Netica solves the network by performing standard belief updating which looks for the marginal posterior probability for each node. In BN modeling, a prior probability is the likelihood of some input parameters being in a particular state e.g. a patient having a history of GIB. A
conditional probability is the likelihood of a state of a parameter given the states of other parameters that affect it, i.e., how the presence of a history of UGIB affects the current presence of GIB. A posterior probability is the likelihood of a certain parameter being in a particular state e.g. the patient having UGIB given the input parameters, conditional probabilities and the rules governing the combination of the probabilities (Marcot et al., 2001). A network is solved when the nodes are updated.

1.4 Expected outcome

The expectation of this thesis is that the BNs developed will assist the physician in the emergency room to diagnose the patients more efficiently and effectively. BN models have the potential to identify patients with both UGIB and LGIB and ascertain their need for urgent treatment or disposition. The model can be used to predict the source of bleeding, disposition and need for intervention (endoscopy and resuscitation) in patients with UGIB or LGIB.

1.5 Limitations of the study

In order to use the BN model, it would need to be validated on larger real datasets and more simulation studies done. It takes time to get approval for use of a large established datasets from organizations. This has limited the testing of the BN models on datasets from other clinics. This study has thus been based on two datasets. A BN model is interactive. It would need to be developed as a web interface so that its performance be compared to the prediction of a physician at the hospital. Due to time and cost constraint, the web interface is not developed.

1.6 Organization of the thesis

This thesis is divided into five chapters. Chapter 1, explains the motivation of the thesis and how the contents throughout the document are organized. Chapter 2, introduces the methods used in the thesis beginning with the most general concepts and gradually focusing on those most related to the remaining chapters. First the general concept of BN is explained, and then details of the modeling process and their construction is given. Finally a specific type of BNs, called Bayesian network classifiers (BNCs) which is the main theme of the thesis, are emphasized. In Chapter 3, BNCs for predicting the four outcomes of GIB are developed. In Chapter 4, BNs for the special case of predicting the source of GIB in the absence of hematemesis are developed and the BNs are compared with other classification models. Finally, Chapter 5 summarizes the main contributions of the thesis and gives the future work.
BIBLIOGRAPHY


