000

001

002

003

004

Prognostic Significance of Dobutamine Stress Echocardiography in Cardiac Events Prediction

055

056

057

058

059

005

006

007

008

009

010

011

060

061

062

063

064

065

066

012

013

014

015

016

017

018

019

020

021

022

023

024

025

026

027

028

029

030

031

032

033

034

035

036

037

038

039

040

041

042

043

044

045

046

047

048

049

050

051

052

053

# Abstract

**This study attempts to determine the prognostic value of a technique called dobutamine stress echocardiography in prediction of cardiac events and more specifically identify which variables that were measured during the test best predict the occurrence of these events in patients over the next year. Logistic regression analysis and ran- dom forests techniques will be used for the prediction.**

# Description of Applied Problem

This study tries to determine the prognostic value of a widely used technique called dobutamine stress echocardiography (DSE) in detection of is- chemia.([J. Krivokapich](#_bookmark7), [1999](#_bookmark7)) Cardiac ischemia occurs because of the restriction of blood flow to the heart muscle which is usually due to obstruction of coro- nary arteries. If the blood flow gets completely blocked it can result in myocardial infarction (MI) commonly known as heart attack. ([Med](#_bookmark3)) Generally, a technique called stress echocardiography is used to predict oc- currence of ”cardiac events” in Patients. It involves increasing the patient’s heart rate by exercise- by hav- ing them run on the treadmill- and taking different measurements of the heart. The test uses ultrasound imaging of the heart for assessing pumping action of the heart as well as detecting abnormalities in any parts of the heart muscle due to receiving insufficient blood flow when the patient’s heart is under stress. However, if the patient is unable to exercise, a medi- cation called dobutamine is used that makes the heart beat faster and have the similar effects as exercise on the heart([J.A. Castillo Morenoa](#_bookmark0), [2005](#_bookmark0)).

Project report for CS4437/CS9637: Applied Machine Learning.

University of Western Ontario, Winter 2016.

Our goal is to determine whether stress echocardiog- raphy is still valuable to predict cardiac events when instead of exercise, dobutamine is used to put patient’s heart under stress and more precisely, to find out which information gained and measurements that were taken during the test will help us predict if the patient ex- perienced cardiac events over the following 12 months. These cardiac events are classified into four categories: cardiac death, myocardial Infarction (MI), angioplasty (PTCA) and coronary artery bypass grafting surgery (CABG).

# Description of Available Data

The data consists of 30 variables on 558 subjects from 1183 patients who were referred to UCLA Adult Cardiac Imaging and Hemodynamics Laboratories for DSE between March 1991 and March 1996 ([dat](#_bookmark4)). For each patient, only the first DSE test was considered for analysis in case of having more than one test in that time interval. 625 patients were excluded from the study for different reasons such as noncardiac death. The final study population included 220 men and 338 women. The data contains both numerical and cate- gorical variables. The data collected for the patients is explained in more details in Appendix[A](#_bookmark8).

# Analysis and Visualization Techniques

## Pre-proccesing

Data pre-processing is required before mining and dis- covery can begin. First, we need to convert nominal variables to numeric because some techniques such as regression require only numeric input. As an exam- ple, ”ECG” variable has three different levels: Normal, Equivocal and MI. In order to convert it to numeric values, two binary variables are created and replaced by this variable. So, we will have ”PosECG” variable which is 0 if there are signs of heart attack on ECG and ”equivECG” variable which would be 0 when the re-

067

068

069

070

071

072

073

074

075

076

077

078

079

080

081

082

083

084

085

086

087

088

089

090

091

092

093

094

095

096

097

098

099

100

101

102

103

104

105

106

107

108

110 165

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 111 |  | *Table 1.* Description of Variables |  |  | 166 |
| 112 | Variable | Description | Unit | Level | 167 |

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

bhr Basal heart rate bpm

basebp Basal blood pressure mmHg basedp Basal double product (*bhr × basebp*) *bpm × mmHg* pkhr Peak heart rate bpm

sbp Systolic blood pressure mmHg

dp Double product(*pkhr × sbp* ) *bpm × mmHG*

dose Dose of dobutamine given mg

maxhr Maximum heart rate bpm

pctMphr Percent maximum predicted heart rate achieved %

mbp Maximum blood pressure mmHg dpmaxdo Double product on maximum dobutamine dose *bpm × mmHg* age age years

gender male/female

baseEF Baseline cardiac ejection fraction %

dobEF Ejection fraction on dobutamine %

chestpain 1 means experienced chest pain

restwma Cardiologist sees wall motion anomaly on echocardiogram (1 = yes) posSE Stress echocardiogram was positive (1 = yes)

newmi New myocardial infarction, or heart attack (1= yes) newptca Recent angioplasty(1=yes)

newcabg Recent bypass surgery (1=yes) death died(1=yes)

hxoft history of hypertension (1 = yes) hxofdm history of diabetes (1= yes)

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

134

hxofcig history of smoking heavy/moderate/non-smo1k8e9r

hxofmi history of heart attack (1 = yes)

135

136

137

hxofptca history of angioplasty (1 = yes)

hxofcabg history of bypass surgery (1= yes) any.event Death, newMI, newPTCA, or newCABG

190

191

192

138

139

140

ecg Baseline electrocardiogram diagnosis normal/Equivocal/MI193

194

195

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

sult of ECG is equivocal. Moreover, the outliers must

be detected and dealt with.

Most classification techniques do not perform well, If the response variables have very unequal frequencies. For example, only 16% of the patients experienced any of the cardiac events while the other 84% didn’t. We can use under-sampling or over-smapling techniques to fix this problem([Chawla et al.](#_bookmark5), [2002](#_bookmark5)).

The next step would be to generate some plots such as box plots and histograms to get an overview of the dis- tribution of values for explanatory and response vari- ables. These plots can also, help us to identify outliers. Furthermore, summary statistics can provide us with simple and quick description of the data.

To help us determine which explanatory variables best predict the patient outcome and to decrease the risk of overfitting, a feature selection method called ”Quadratic Programming Feature Selection” will be used that reduces the number of explanatory vari- ables([Rodriguez-Lujan et al.](#_bookmark2), [2010](#_bookmark2)).

## Analysis

The first classifier that will be applied to the data is bivariate logistic regression to test the association of each possible explanatory variables with the re- sponse variable by looking at their corresponding co- efficients([James et al.](#_bookmark1), [2014](#_bookmark1)). In order to asses the accuracy of our model, the training error rate can be used.

Based on the result of bivariate logistic regression model, we can choose the explanatory variables that were significantly associated with the response vari- able and include them in a multivariate logistic regres- sion. Then, the polynomial functions of the variables (quadratic, cubic, quartic,...) will be used to get a non- linear decision boundary that better fits the data. In order to help us decide which is the best model, we can use k-fold cross validation with k=5 and k=10 by com- puting k-fold CV error test that come from applying a few logistic regression models to the data using differ- ent degrees of polynomial functions of the explanatory variables.

Additional analysis can be done using a non-

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

parametric technique called random forests which cre- ates multiple decision trees using bootstrapping([Hastie](#_bookmark6) [et al.](#_bookmark6), [2001](#_bookmark6)). Random forests method will result in sig- nificant decrease of variance in comparison to a single decision tree. We can use Gini index to get an overall understanding on the significance of explanatory vari- ables. Also, out-of-bag error estimation can be used to assess the accuracy of this technique.

## Visualization

It is often difficult to explore and make sense of large amounts of data; yet with the use of different visual- ization techniques, we will be able to examine the data to discover relationships between different explanatory variables and how each of them could affect the re- sponse variable.

We plan to use scatter plots to display the relationship between two numerical explanatory variables and their association with response variable. Also, parallel coor- dinates can be used to help with multivariate analysis where each of the best predictors will be mapped to each axis and the response variable can be displayed using two distinct colours. Moreover, stacked bar chart is another useful technique to display the relationship between a categorical explanatory variable and the re- sponse variable.

# References

URL [https://www.nlm.nih.gov/medlineplus/](https://www.nlm.nih.gov/medlineplus/medlineplus.html) [medlineplus.html](https://www.nlm.nih.gov/medlineplus/medlineplus.html).

URL [http://biostat.mc.vanderbilt.edu/](http://biostat.mc.vanderbilt.edu/DataSets) [DataSets](http://biostat.mc.vanderbilt.edu/DataSets).

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Syn- thetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357, June 2002. ISSN 1076-

9757. URL [http://dl.acm.org/citation.cfm?](http://dl.acm.org/citation.cfm?id=1622407.1622416) [id=1622407.1622416](http://dl.acm.org/citation.cfm?id=1622407.1622416).

Trevor Hastie, Robert Tibshirani, and Jerome Fried- man. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

D.O. Walter A. Garfinkel J. Krivokapich, J.S. Child. Prognostic value of dobutamine stress echocardiog- raphy in predicting cardiac events in patients with known or suspected coronary artery disease. *Jour- nal of the American College of Cardiology*, 33(3): 708–716, March 1999.

E. Molina Labordaa R. Florenciano S´ancheza J. Ortega Bernala J.A. Castillo Morenoa, J.L. Ramos Mart´ına. Dobutamine stress echocardiogra- phy in patients with stable chronic angina and a low- or medium-risk on exercise testing: Usefulness for assessing long-term prognosis. *Rev Esp Cardiol.*, 08(58):916–23, 08 2005.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statisti- cal Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014. ISBN 1461471370, 9781461471370.

Irene Rodriguez-Lujan, Ramon Huerta, Charles Elkan, and Carlos Santa Cruz. Quadratic programming fea- ture selection. *J. Mach. Learn. Res.*, 11:1491–1516, August 2010. ISSN 1532-4435. URL [http://dl.](http://dl.acm.org/citation.cfm?id=1756006.1859900) [acm.org/citation.cfm?id=1756006.1859900](http://dl.acm.org/citation.cfm?id=1756006.1859900).

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

336

|  |  |  |
| --- | --- | --- |
| 330 | **A.** | 385 |
| 331 | **Data Discription** | 386 |
| 332 |  | 387 |
| 333 |  | 388 |
| 334 | *Table 2.* Description of Variables | 389 |
| 335 | Variable Description | Unit Level 390 |

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

bhr Basal heart rate bpm

basebp Basal blood pressure mmHg basedp Basal double product (*bhr × basebp*) *bpm × mmHg* pkhr Peak heart rate bpm

sbp Systolic blood pressure mmHg

dp Double product(*pkhr × sbp* ) *bpm × mmHG*

dose Dose of dobutamine given mg

maxhr Maximum heart rate bpm

pctMphr Percent maximum predicted heart rate achieved %

mbp Maximum blood pressure mmHg dpmaxdo Double product on maximum dobutamine dose *bpm × mmHg* age age years

gender male/female

baseEF Baseline cardiac ejection fraction %

dobEF Ejection fraction on dobutamine %

chestpain 1 means experienced chest pain

restwma Cardiologist sees wall motion anomaly on echocardiogram (1 = yes) posSE Stress echocardiogram was positive (1 = yes)

newmi New myocardial infarction, or heart attack (1= yes) newptca Recent angioplasty(1=yes)

newcabg Recent bypass surgery (1=yes) death died(1=yes)

hxoft history of hypertension (1 = yes) hxofdm history of diabetes (1= yes)

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

357

hxofcig history of smoking heavy/moderate/non-smo4k1e2r

358

359

hxofmi history of heart attack (1 = yes)

hxofptca history of angioplasty (1 = yes)

hxofcabg history of bypass surgery (1= yes)

|  |  |  |  |
| --- | --- | --- | --- |
| 360  361 | any.event  ecg | Death, newMI, newPTCA, or newCABG  Baseline electrocardiogram diagnosis | 415  normal/Equivocal/MI416 |
| 362 |  |  | 417 |
| 363 |  |  | 418 |
| 364 |  |  | 419 |
| 365 |  |  | 420 |
| 366 |  |  | 421 |
| 367 |  |  | 422 |
| 368 |  |  | 423 |
| 369 |  |  | 424 |
| 370 |  |  | 425 |
| 371 |  |  | 426 |
| 372 |  |  | 427 |
| 373 |  |  | 428 |
| 374 |  |  | 429 |
| 375 |  |  | 430 |
| 376 |  |  | 431 |
| 377 |  |  | 432 |
| 378 |  |  | 433 |
| 379 |  |  | 434 |
| 380 |  |  | 435 |
| 381 |  |  | 436 |
| 382 |  |  | 437 |
| 383 |  |  | 438 |

413

414