Risk Factors and Prediction of ST-segment Elevation Myocardial Infarction

Joshua Emakhu, Leslie Monplaisir, Celestine Aguwa, and Sara Masoud
Department of Industrial and Systems Engineering
Wayne State University
Detroit, MI 48201, USA
fw7471@wayne.edu, leslie.monplaisir@wayne.edu, bd4891@wayne.edu, hb9225@wayne.edu

Suzan Arslanturk
Department of Computer Science
Wayne State University
Detroit, MI 48201, USA
gl9815@wayne.edu

Egbe-Etu Etu
Department of Marketing and Business Analytics
San Jose State University
San Jose, CA 95192, USA
egbe-etu.etu@sjsu.edu

Joseph Miller
Department of Emergency Medicine and Internal Medicine
Henry Ford Hospital
Detroit, MI 48202, USA
jmiller6@hfhs.org

Abstract

Studies have shown that the mortality rate due to ST-segment elevation myocardial infarction (STEMI) has drastically increased in developed and developing countries. The purpose of this study is to develop a diagnostic support tool to help classify STEMI patients and validate the predictive risk factors associated with STEMI using an ensemble learning approach. In this retrospective data-mining study, the data are retrieved from electronic health records of an urban emergency department between January 2017 and August 2020. A Random Forest model is trained to classify non-acute coronary syndrome (non-ACS) etiologies and STEMI patients using 38 features. Of the study cohort, 411 patients with chest pain fulfilled inclusion criteria, of whom 225 (55%) are STEMI, and 186 (45%) are non-ACS etiologies patients. The proposed framework successfully classifies the non-ACS etiologies and STEMI patients with recall and area under the receiver operating characteristic (AUROC) values of 73% and 90%, respectively.

Keywords

Acute coronary syndrome, ST-segment Elevation Myocardial Infarction, Prediction, BorutaShap, and Machine Learning.

1 Introduction

Cardiovascular disease (CVD) is a group of illnesses that comprise both heart and blood vessels (Mendis *et al.* 2011), including coronary heart disease (CHD) and coronary artery disease (CAD), and acute coronary syndrome (ACS). Although health professionals often use the terms CAD and ACS interchangeably with CHD, they are not equivalent. ACS is a subcategory of CAD, while CHD is an outcome of CAD. ACS typically comprise three manifestations of

ischemic heart disease (Danish 2016) – (i) ST-segment elevation myocardial infarction (STEMI), (ii) Non-ST-segment elevation myocardial infarction (NSTEMI), and (iii) Unstable angina (UA).

STEMI, usually resulting from thrombotic obstruction of a coronary artery, is a fatal cardiovascular emergency requiring early diagnosis and rapid reperfusion therapy. Clinical trial and registry data have shown that individuals with STEMI, and its counterpart NSTEMI, have differing short-term prognoses and responses to treatments (Cox *et al.* 2006; Abbott *et al.* 2007). Given the aging US population and the growing populace burden of obesity and diabetes, the attributes of hospitalization for acute myocardial infarction (AMI) patients have changed during recent years (Mokdad *et al.* 2001; Floyd *et al.* 2009). According to (Garoufalis *et al.* 1998), the frequency of AMI used to be as low as 2–6% but has increased in recent times due to the increased predominance of risk factors for CAD, such as earlier onset glucose resistance and obesity, in adolescence (Sinha *et al.* 2002). Early acknowledgment and risk factor modification are crucial to mitigating AMI's risk (Jamil *et al.* 2013).

The existing predictive models for mortality prediction following STEMI comprise the Thrombolysis in Myocardial Infarction (TIMI) score and the Global Registry of Acute Cardiac Events (GRACE) scores. These scores permit risk stratification, identification of high-risk patients, and appropriate treatment (Fox *et al.* 2006; Addala *et al.* 2004; Morrow *et al.* 2000). However, these predictive models are based on a conventional statistical approach which conveys inherent limitations, conceivably impeding performance (Breiman 2001; Shouval *et al.* 2014). In current practice, standard statistical methods for prediction depend on parametric regression methods. However, progressions in Computer Science, particularly in machine learning (ML), allow for the development of complex predictive models exposed to risk factors. ML techniques like the random forest, neural network, and support vector machines have been presented in epidemiologic studies for predictive analytics (Peng *et al.* 2010; Kruppa *et al.* 2012; Gurm *et al.* 2013). This study aims to develop a diagnostic support tool to help classify STEMI patients and validate the predictive risk factors associated with STEMI using an ensemble learning approach.

The rest of the paper proceeds as follows. After reviewing related literature in section 2, we propose a dynamic decision support tool in section 3. We then presented our findings from our results and gave insights in section 4. Finally, section 5 includes concluding remarks.

2 Literature Review

Machine learning (ML) is a technique of data analysis that automates logical model building. Utilizing algorithms that iteratively learn from data, ML permits computers to discover hidden insights without being expressly programmed about where to look. Thus, ML offers an alternative methodology for the development of prediction models. ML is scientific research standing at the convergence of statistics and computer science, relying on proficient computing algorithms. The significance of ML has been perceived through the difficulties of building statistical models from massive data sets, which required computational strategies (Deo 2015). ML may detect the intricate and non-linear interactions between variables by limiting the error between predicted and observed outcomes (Karthik *et al.* 2014).

Data mining (DM) utilizes ML algorithms for knowledge discovery and predictive modeling in massive datasets (Shouval *et al.* 2014; Deo 2015). DM applications are continuously entering clinical use in everyday medicine and cardiology (Arsanjani *et al.* 2015; Gurm *et al.* 2014; Hsich *et al.* 2011; Shouval *et al.* 2015). As opposed to the standard statistical approach, ML does not begin with a pre-defined model; instead, it allows the data to develop the model by identifying underlying patterns.

DM is significant for various kinds of issues. Predicting a reliant variable from the implications of independent variables is one of the utilizations of this technique. Healthcare is one of the fields with big data accessible freely from various sources, and DM could be the best way to handle it. As per the Centers for Disease Control and Prevention (CDC), 647,457 people died of coronary illness in the US as of (2017) [Source: CDC, National Center for Health Statistics]. One reason for casualty due to heart disease is that risks are either not classified or distinguished distinctly later. In the present circumstance, ML could be the best way to deal with this issue and forecast at the beginning phase.

Several ML methods have effectively been applied for prediction. A portion of the methods utilized for such prediction problems is support vector machines (SVM), neural networks, decision trees, regression, and naive Bayes classifiers. SVM was perceived as the best predictor and precision, followed by neural networks with improved accuracy, and decision trees introduced a diminished accuracy. Chest pain, age, smoking, hypertension, and diabetes were estimated

to be the risk factors of heart disease (Latha and Jeeva 2019). Studies on DM techniques for CAD prediction uncover that neural networks, decision trees, and naive bayes are instrumental in predicting coronary illness.

Scott *et al.* (2002) examined the estimations and quality checking of care for patients with acute coronary syndromes. Massad *et al.* (2008) assessed the present state of the art of logical applications in clinical diagnosis. Quteishat and Lim (2008), in their paper, examined the intelligent DM techniques like min-max neural networks to medical diagnosis. They also chose real-life clinical records from which suspected ACS patients, mainly STEMI, are gathered and utilized for their analysis. Although, the utilization of DM when tackling the issue of forecasting in-hospital outcomes in ACS has not been investigated broadly in the existing literature (Velicki *et al.* 2014; Mihajlović *et al.* 2011).

3 Methods

The primary purpose of this research is to propose a framework capable of classifying STEMI and non-ACS etiologies in the emergency department (ED) and recognizing the leading risk factors contributing to this classification. The proposed framework is built on three phases. The first phase, data preprocessing, comprises feature engineering (i.e., cleaning the data, data imputation, and standardization). The second phase, feature selection, reduces the dimension of the dataset by identifying the most suitable features. The final phase, model training, and evaluation starts with 10-fold cross-validation to provide a robust evaluation of the performance for the trained model (i.e., Random Forest) based on metrics such as the AUROC, recall, precision, F1-score, and accuracy. Figure 1 below depicts the architecture of the proposed framework.

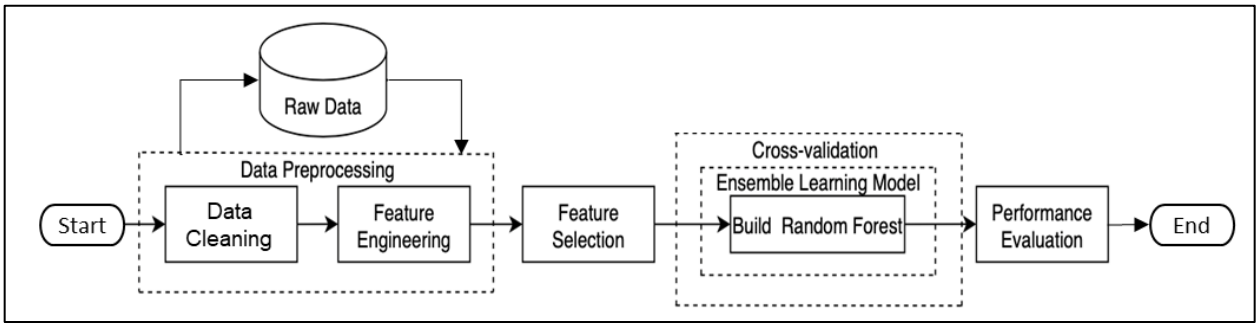

Figure 1: Architecture of proposed framework

3.1 Data Collection

The data used for the study analysis is extracted from the electronic health record (EHR) made available by Henry Ford Health Systems. A total of 411 patients with chest pain were selected for this study, of which 225 (55%) had STEMI and 186 (45%) had non-ACS etiologies of chest pain. 38 variables are selected after a broad literature search and following discussions with medical experts. Then, all available features, including demographic and clinical factors, are retrieved from the EHRs.

3.2 Data Preprocessing

This includes two main steps: i) data cleaning and ii) feature engineering

3.2.1 Data Cleaning

Data cleaning eliminates incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. The following stages are performed during our analysis. (i) *Eliminating redundant data:* missing data values and inconsistencies can negatively impact the integrity of the dataset; redundant data points are dropped. (ii) *Data imputation:* this is achieved through the use Multiple Imputations by Chained Equations (MICE) approach (Schafer and Graham 2002). MICE is a robust, informative method of handling missing data in datasets. The approach addresses the issue of missing data in a dataset through an iterative series of predictive models.

3.2.2 Feature Engineering

This process includes using space information to extricate variables from a dataset. The following stage is performed during our analysis. (i) *One-hot encoding:* to supplant encoded categorical features with recently created binary factors.

3.3 Feature Selection

Feature Selection is the way toward diminishing features to those contributing the most to the output (Shaikh 2018). To distinguish the features with the most commitment in boosting the performance of a classification algorithm, BorutaShap (Keany 2020) is introduced. BorutaShap is a wrapper feature selection technique that merges the Boruta feature selection algorithm with SHAP (SHapely Additive exPlanations). SHAP links game theory with local explanations to create the only consistent and accurate explainer. From an exceptionally high level, SHAP analyses the average marginal contributions for each feature across all permutations at a local level. As this method is additive, the average value of these marginal contributions can then accomplish global feature importance. Furthermore, since most feature selection methods are biased towards continuous or categorical data, BorutaShap becomes suitable for this type of problem. It takes care of this limitation; thus, leading to our implementation in this study.

3.4 Ensemble Model

In statistics and ML, ensemble methods utilize several learning algorithms to acquire better predictive performance than any constituent learning algorithms alone (Maclin and Opitz 2011; Rokach 2010). In contrast to a statistical ensemble in statistical mechanics, which is usually endless, an ML ensemble comprises only a particular finite set of alternative models. Notwithstanding, it ordinarily permits a considerably more flexible structure to exist among those other options. The classifier utilized stated in the proposed framework is Random Forest. Random forests or random decision forests are ensemble learning techniques for classification and regression problems. A random forest is operated by building many decision trees at training time and outputting the class that is the mode of classification or regression of the individual trees (Ho 1995, 1998). In addition, random decision forests correct for decision trees' propensity for overfitting to their training set (Hastie *et al.* 2001).

4 Results and Discussion

4.1 Descriptive Statistics

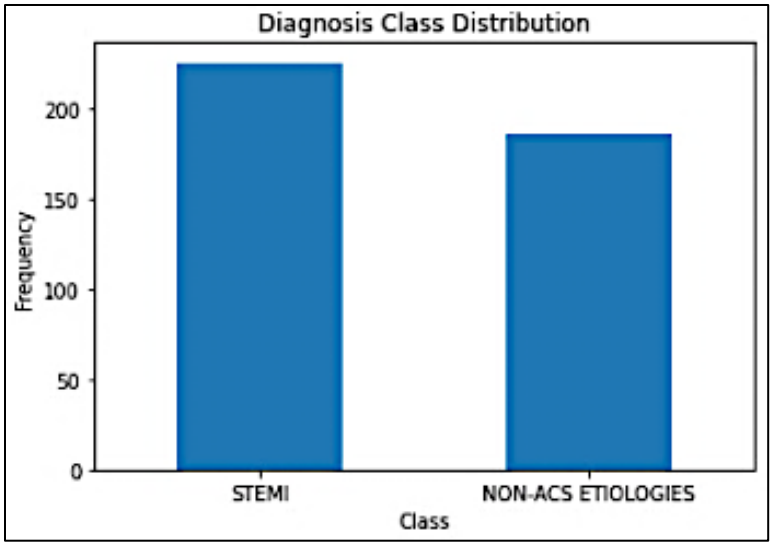

Figure 2: Diagnosis class distribution

Table 1 shows the characteristics baseline of patients with STEMI and non-ACS, where all measurable tests are two-tailed at a significant level of 5%. The mean age of the patient groups is 60.43±12.71 and 50.83±16.61years for STEMI and non-ACS etiologies, respectively. The ratio of female patients seems to be lower in STEMI (38%) compared to non-ACS (46%). Furthermore, dominance of coronary artery disease (69% vs 24%), chronic heart failure (28% vs 24%), chronic heart disease (29% vs 24%), cardiovascular disease (56% vs 39%), hypertension (78% vs

55%), heart attack (69% vs 12%), and diabetes (37% vs 25%) are observed in STEMI patients. Figure 2 shows the diagnosis class distribution.

Table 1: Baseline characteristics of selected variables

Variables		*STEMI*	*Non-ACS*	*p-value*
Demographic	*Age (years)*	60.43±12.71	50.83±16.61	<0.001
Sex	*Female*	0.38	0.46	0.126
	Male	0.62	0.54	
Ethnicity	*Decline*	0.02	0.03	0.127
	Do not know	0.06	0.02	
	Hispanic/Latino	0.04	0.04	
	Not Hispanic/Latino	0.88	0.91	
Comorbidities	*Abdominal aortic aneurysm*	0.00	0.02	N/A
	Atrial fibrillation	0.09	0.10	0.470
	Alcoholism	0.09	0.12	0.165
	Anemia	0.13	0.29	<0.001
	Asthma	0.06	0.21	<0.001
	Breast cancer	0.02	0.02	0.995
	CAD	0.69	0.24	<0.001
	CHF	0.28	0.24	0.521
	CHD	0.29	0.24	0.376
	Chronic liver disease	0.07	0.10	0.325
	Chronic lung disease	0.16	0.32	<0.001
	CKD	0.26	0.32	0.106
	COPD	0.11	0.21	0.005
	CVD	0.56	0.39	0.008
	Diabetes	0.37	0.25	0.024
	Type-2 diabetes	0.38	0.25	0.019
	Hypertension	0.78	0.55	<0.001
	History of hemoglobinopathies	0.00	0.02	0.158
	History of hypertension	0.78	0.55	<0.001
	Lipid disorder	0.00	0.00	N/A
	Nephrotic syndrome	0.00	0.00	N/A
	Noncardiac atherosclerosis	0.00	0.00	0.997
	Peripheral vascular disease	0.08	0.04	0.164
	Heart attack	0.69	0.12	<0.001
	Liver disease	0.00	0.00	N/A
	CAD risk factors	0.00	0.00	N/A
	Obesity	0.22	0.21	0.818
Vital Signs	*SBP*	140.31±32.82	137.38±25.56	0.446
	DBP	85.22±22.25	79.65±18.82	0.014
	HR	84.74±21.63	87.83±17.70	0.022
	Temperature	97.77±1.32	98.23±0.59	<0.001
	RR	19.63±4.94	18.10±2.42	0.004
	Oxygen saturation	97.82±2.84	97.68±7.41	0.936

4.2 Feature Engineering

After performing one-hot encoding on our dataset, we had a total of 75 features. This is due to the method producing vectors with lengths equal to the number of categories in the data set. We can see this increase from 38 to 75 features, as illustrated in Figure 3.

4.3 Feature Selection

Figure 3 displays the output values from the feature selection algorithm, BorutaShap, where the green, yellow, and red bars symbolize accepted, uncertain, and rejected features, respectively. A total of 11 variables, including age, anemia, cardiovascular disease, chest pain, chronic lung disease, chronic obstructive pulmonary disease, coronary artery disease, diastolic blood pressure, heart attack, systolic blood pressure, and temperature level, are reported as significantly contributing.

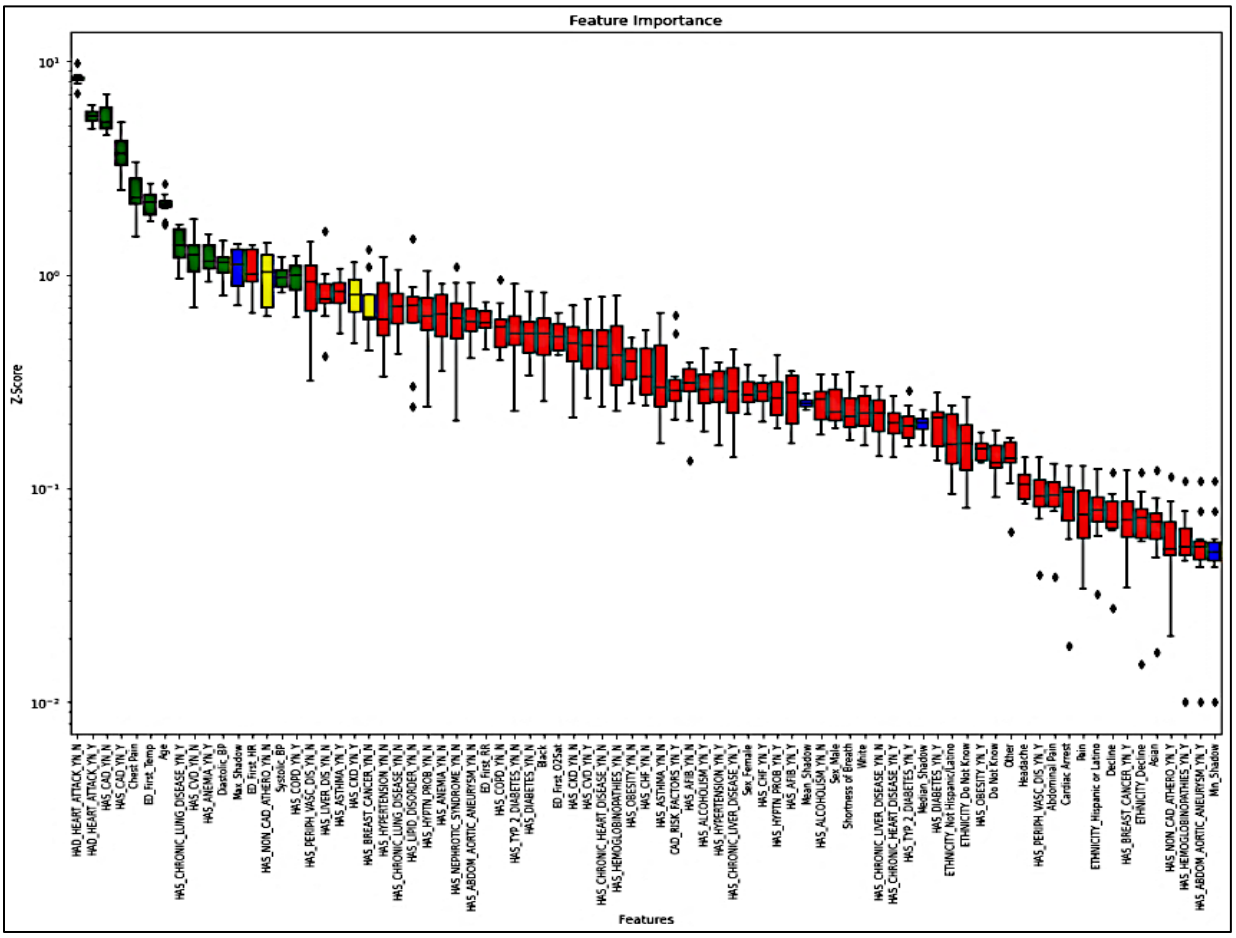

Figure 3: Selection of potential variables by BorutaShap

4.4 Model Performance

Given the impacts of the features on the performance of the random forest model, the features recommended by the BorutaShap algorithm are compared to eleven random selections of features. The performances of the classification model based on different sets of features are presented in Figure 4. As displayed in Figure 4, the features suggested by BorutaShap are indeed the most significant. Furthermore, Table 2 fully illustrates the evaluation criteria for each feature subset.

Table 2: Evaluation criteria for the feature subsets

Feature Subsets	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	***F1-score (%)***	***AUC (%)***
BorutaShap Subset	**82**	**83**	73	**82**	**90**
Subset 1	62	58	63	63	70
Subset 2	61	59	60	62	66
Subset 3	67	70	54	66	72
Subset 4	56	52	53	57	59

Subset 5	56	73	7	42	53
Subset 6	64	60	61	64	68
Subset 7	77	74	**78**	77	83
Subset 8	*75*	72	71	*74*	81
Subset 9	65	62	61	65	72
Subset 10	68	65	67	69	74
Subset 11	65	63	62	65	71

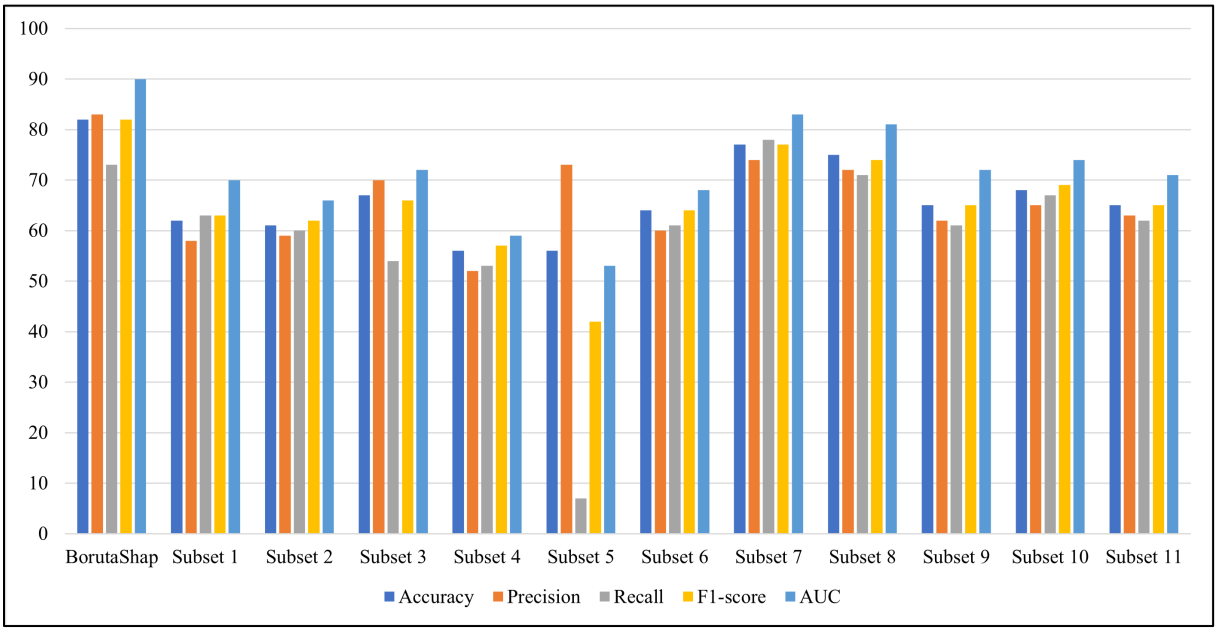

Figure 4: Bar chart comparing the ensemble learning methods

The BorutaShap selected features outperformed other features based on all the evaluation criteria from our proposed approach. The BorutaShap selected features have accuracy, precision, recall, F1-score, and AUROC values of 82%, 83%, 73%, 82, and 90%, respectively. Therefore, the AUROC curve for BorutaShap selected features is illustrated below in Figure 5.

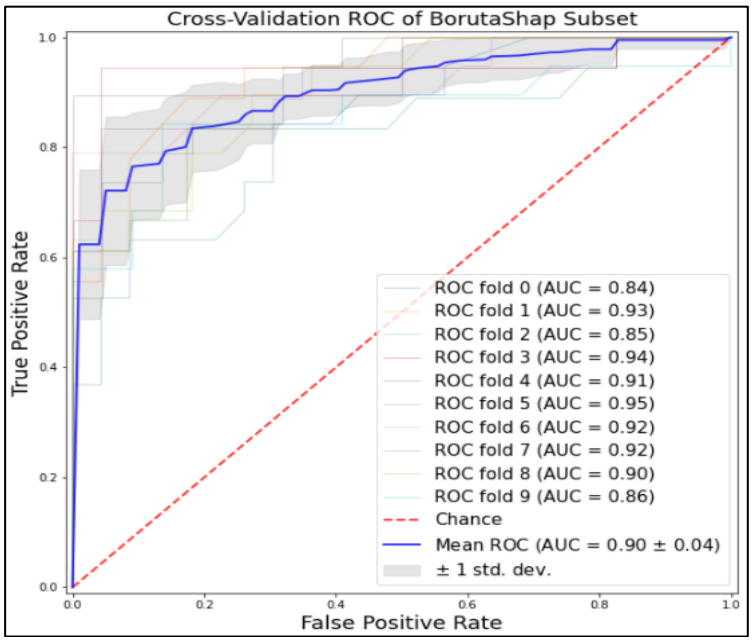

Figure 5: ROC for Random Forest on the BorutaShap selected features

The feature selection process in our current study merged data-driven systems and domain knowledge ensuing comprehensive clinical variables. From prior related works, several risk factors, such as age, coronary artery disease, cardiovascular disease, are exceptionally connected with STEMI. This result correlates well with 8 out of 11 features selected being clinically understood as typical risk factors of STEMI. This study also showed that having a history of heart attack is a strong indicator of STEMI cases. Prior related works have demonstrated that advanced age is associated with more significant morbidity than younger patients (Zorbozan *et al.* 2018). Additionally, anemia, chest pain, chronic lung disease, chronic obstructive pulmonary disease, and systolic blood pressure were regarded as strong indicators of STEMI case prediction (Bailey *et al.* 2003; Goedemans *et al.* 2020; Granot *et al.* 2019). Furthermore, diastolic blood pressure and temperature were also distinguished as indicators, showing their influence on the disease spectrum.

Our findings show that ML can effectively help diagnose patients with STEMI from non-ACS etiologies given clinical information. Not only can the proposed framework classify these diseases with high accuracy, but it can also detect the leading features that essentially add to the classification process. Though our proposed framework shows high accuracy, and AUROC scores, the combination of the physicians' remarks may further improve the performance of our proposed framework. What's more, the proposed framework has not been implemented in actual clinical practice yet, but the potential variables identified would probably update clinical dynamics and improve the well-being of the designated patients.

5 Conclusion

Our findings support our proposed model to distinguish the significant variables and diagnose STEMI and non-ACS with high accuracy. With these features identified, it will enhance treatment decisions for STEMI patients requiring medical intervention. Physicians will not only have to rely on electrocardiogram readings, but with the inclusion of the selected variables, they may enhance decision-making in evaluating ACS. Future analysis may also be enhanced by considering more extensive free-text data in the EHR, as it could aid in uncovering hidden patterns and improve our predictive performance.

Reference:

Abbott, J Dawn, Hanna N Ahmed, Helen A Vlachos, Faith Selzer, and David O Williams. 2007. 'Comparison of outcome in patients with ST-elevation versus non–ST-elevation acute myocardial infarction treated with percutaneous coronary intervention (from the National Heart, Lung, and Blood Institute Dynamic Registry)', *The American journal of cardiology*, 100: 190-95.

Addala, Srinivas, Cindy L Grines, Simon R Dixon, Gregg W Stone, Judith A Boura, Anthony B Ochoa, Gregory Pellizzon, William W O'Neill, and Joel K Kahn. 2004. 'Predicting mortality in patients with ST-elevation myocardial infarction treated with primary percutaneous coronary intervention (PAMI risk score)', *The American journal of cardiology*, 93: 629-32.

Arsanjani, Reza, Damini Dey, Tigran Khachatryan, Aryeh Shalev, Sean W Hayes, Mathews Fish, Rine Nakanishi, Guido Germano, Daniel S Berman, and Piotr Slomka. 2015. 'Prediction of revascularization after myocardial perfusion SPECT by machine learning in a large population', *Journal of Nuclear Cardiology*, 22: 877-84.

Bailey, David, Y Wady Aude, Paul Gordon, and Douglas Burtt. 2003. 'ST-segment elevation myocardial infarction, severe anemia and nonobstructive coronary disease: case report and brief comment', *Connecticut medicine*, 67: 3-5.

Breiman, Leo. 2001. 'Statistical modeling: The two cultures (with comments and a rejoinder by the author)', *Statistical science*, 16: 199-231.

Cox, David A, Gregg W Stone, Cindy L Grines, Thomas Stuckey, Peter J Zimetbaum, James E Tcheng, Mark Turco, Eulogio Garcia, Giulio Guagliumi, and Robert S Iwaoka. 2006. 'Comparative early and late outcomes after primary percutaneous coronary intervention in ST-segment elevation and non–ST-segment elevation acute myocardial infarction (from the CADILLAC trial)', *The American journal of cardiology*, 98: 331-37.

Danish, MI 2016. *Short Textbook on Medical Diagnosis and Management* (Paramount Publishing Enterprise: Pakistan).

Deo, Rahul C. 2015. 'Machine learning in medicine', *Circulation*, 132: 1920-30.

Floyd, Kevin C, Jorge Yarzebski, Frederick A Spencer, Darleen Lessard, James E Dalen, Joseph S Alpert, Joel M Gore, and Robert J Goldberg. 2009. 'A 30-year perspective (1975–2005) into the changing landscape of patients hospitalized with initial acute myocardial infarction: Worcester Heart Attack Study', *Circulation: Cardiovascular Quality and Outcomes*, 2: 88-95.

Fox, Keith AA, Omar H Dabbous, Robert J Goldberg, Karen S Pieper, Kim A Eagle, Frans Van de Werf, Álvaro Avezum, Shaun G Goodman, Marcus D Flather, and Frederick A Anderson. 2006. 'Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE)', *bmj*, 333: 1091.

Garoufalis, Stefanos, George Kouvaras, George Vitsias, Kyriakos Perdikouris, Paraskevi Markatou, Jhon Hatzisavas, Neophytos Kassinos, Kostas Karidis, and Stefanos Foussas. 1998. 'Comparison of angiographic findings, risk factors, and long term follow-up between young and old patients with a history of myocardial infarction', *International Journal of Cardiology*, 67: 75-80.

Goedemans, Laurien, Jeroen J Bax, and Victoria Delgado. 2020. 'COPD and acute myocardial infarction', *European Respiratory Review*, 29.

Granot, Michal, Pnina Dagul, and Doron Aronson. 2019. 'Resting blood pressure modulates chest pain intensity in patients with acute myocardial infarction', *Pain reports*, 4.

Gurm, Hitinder S, Judith Kooiman, Thomas LaLonde, Cindy Grines, David Share, and Milan Seth. 2014. 'A random forest based risk model for reliable and accurate prediction of receipt of transfusion in patients undergoing percutaneous coronary intervention', *PloS one*, 9: e96385.

Gurm, Hitinder S, Milan Seth, Judith Kooiman, and David Share. 2013. 'A novel tool for reliable and accurate prediction of renal complications in patients undergoing percutaneous coronary intervention', *Journal of the American College of Cardiology*, 61: 2242-48.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. 'The elements of statistical learning. Springer series in statistics.' in, *:* (Springer).

Ho, Tin Kam. 1995. "Random decision forests." In *Proceedings of 3rd international conference on document analysis and recognition*, 278-82. IEEE.

———. 1998. 'The random subspace method for constructing decision forests', *IEEE transactions on pattern analysis and machine intelligence*, 20: 832-44.

Hsich, Eileen, Eiran Z Gorodeski, Eugene H Blackstone, Hemant Ishwaran, and Michael S Lauer. 2011. 'Identifying important risk factors for survival in patient with systolic heart failure using random survival forests', *Circulation: Cardiovascular Quality and Outcomes*, 4: 39-45.

Jamil, Gohar, Mujgan Jamil, Hind AlKhazraji, Amber Haque, Fares Chedid, Manjula Balasubramanian, Bahaa Khairallah, and Anwer Qureshi. 2013. 'Risk factor assessment of young patients with acute myocardial infarction', *American Journal of Cardiovascular Disease*, 3: 170.

Karthik, L, Gaurav Kumar, Tarun Keswani, Arindam Bhattacharyya, S Sarath Chandar, and KV Bhaskara Rao. 2014. 'Protease inhibitors from marine actinobacteria as a potential source for antimalarial compound', *PloS one*, 9: e90972.

Keany, E. 2020. "Is this the Best Feature Selection Algorithm “BorutaShap”?" In. Analytics Vidhya.
Kruppa, Jochen, Andreas Ziegler, and Inke R König. 2012. 'Risk estimation and risk prediction using machine-learning methods', *Human genetics*, 131: 1639-54.
Latha, C Beulah Christalin, and S Carolin Jeeva. 2019. 'Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques', *Informatics in Medicine Unlocked*, 16: 100203.
Maclin, Richard, and David Opitz. 2011. 'Popular ensemble methods: An empirical study', *ArXiv e-prints*: arXiv: 1106.0257.
Massad, Eduardo, Neli Regina Siqueira Ortega, Laecio Carvalho de Barros, and Claudio José Struchiner. 2008. '... and Beyond: Fuzzy Logic in Medical Diagnosis.' in, *Fuzzy Logic in Action: Applications in Epidemiology and Beyond* (Springer).
Mendis, Shanthi, Pekka Puska, Bo Norrving, and World Health Organization. 2011. *Global atlas on cardiovascular disease prevention and control* (World Health Organization).
Mihajlović, Bogoljub, Svetozar Nićin, Pavle Kovačević, Stamenko Šušak, Lazar Velicki, Dragan Kovačević, and Mikloš Fabri. 2011. 'Evaluation of results in coronary surgery using EuroSCORE', *Srpski arhiv za celokupno lekarstvo*, 139: 25-29.
Mokdad, Ali H, Barbara A Bowman, Earl S Ford, Frank Vinicor, James S Marks, and Jeffrey P Koplan. 2001. 'The continuing epidemics of obesity and diabetes in the United States', *JAMA*, 286: 1195-200.
Morrow, David A, Elliott M Antman, Andrew Charlesworth, Richard Cairns, Sabina A Murphy, James A de Lemos, Robert P Giugliano, Carolyn H McCabe, and Eugene Braunwald. 2000. 'TIMI risk score for ST-elevation myocardial infarction: a convenient, bedside, clinical score for risk assessment at presentation: an intravenous nPA for treatment of infarcting myocardium early II trial substudy', *Circulation*, 102: 2031-37.
Peng, S-Y, Y-C Chuang, T-W Kang, and K-H Tseng. 2010. 'Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination', *European journal of neurology*, 17: 945-50.
Quteishat, Anas, and Chee Peng Lim. 2008. "Application of the fuzzy min-max neural networks to medical diagnosis." In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 548-55. Springer.
Rokach, Lior. 2010. 'Ensemble-based classifiers', *Artificial intelligence review*, 33: 1-39.
Schafer, Joseph L, and John W Graham. 2002. 'Missing data: our view of the state of the art', *Psychological methods*, 7: 147.
Scott, IA, CP Denaro, JL Flores, CJ Bennett, AC Hickey, and AM Mudge. 2002. 'Quality of care of patients hospitalized with acute coronary syndromes', *Internal medicine journal*, 32: 502-11.
Shaikh, R. 2018. "Feature Selection Techniques in Machine Learning with Python." In. Towards data science.
Shouval, R, O Bondi, H Mishan, A Shimoni, R Unger, and Arnon Nagler. 2014. 'Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT', *Bone marrow transplantation*, 49: 332-37.
Shouval, Roni, Myriam Labopin, Ori Bondi, Hila Mishan-Shamay, Fabio CICERI, Jordi ESTEVE, Sebastian GIEBEL, Norbert C GORIN, Christoph SCHMID, and Emmanuelle POLGE. 2015. 'Prediction of allogeneic hematopoietic stem-cell transplantation mortality 100 days after transplantation using a machine learning algorithm: a European Group for Blood and Marrow Transplantation Acute Leukemia Working Party retrospective data mining study', *Journal of Clinical Oncology*, 33: 3144-52.
Sinha, Ranjana, Gene Fisch, Barbara Teague, William V Tamborlane, Bruna Banyas, Karin Allen, Mary Savoye, Vera Rieger, Sara Taksali, and Gina Barbetta. 2002. 'Prevalence of impaired glucose tolerance among children and adolescents with marked obesity', *New England Journal of Medicine*, 346: 802-10.
Velicki, Lazar, Nada Cemerlic-Adjic, Katica Pavlovic, Bojan B Mihajlovic, Dragic Bankovic, Bogoljub Mihajlovic, and Miklos Fabri. 2014. 'Clinical performance of the EuroSCORE II compared with the previous EuroSCORE iterations', *The Thoracic and cardiovascular surgeon*, 62: 288-97.
Zorbozan, Onur, Arif A Cevik, Nurdan Acar, Engin Ozakin, Hamit Ozcelik, Alparslan Birdane, and Fikri M Abu-Zidan. 2018. 'Predictors of mortality in ST-elevation MI patients: A prospective study', *Medicine*, 97.

Biographies

Joshua Emakhu is a Ph.D. student and Graduate Teaching Assistant in the Industrial and Systems Engineering department at Wayne State University, Michigan, USA. He earned a B.Sc. in Civil Engineering from Covenant University, Ogun, Nigeria, and a Master’s in Industrial Engineering from Wayne State University, Michigan, USA. His research interests include data analytics, medical devices, recalls, healthcare, resource allocation, and the voice of

the customer. He is a member of IEOM, INFORMS, SAVE Int'l, and IISE. He is also a peer reviewer for IISE, RIED, and Frontiers for Psychology.

Leslie Monplaisir, Ph.D., is a Professor in the Industrial and Systems Engineering Department at WSU. He is the Lead Researcher and Director of the Product Development and Systems Engineering Consortium (PDSEC) at WSU. His research interests include Lean Product Development, Design for lean Systems and Services and Design reuse, New Product Technology Decision modeling, Product Architecture Optimization, and Healthcare Technology System Design. He has authored over 100 publications in these areas with funded research from NSF, Veterans Administration, and Ford Motor Company. Dr. Monplaisir joined the College of Engineering at WSU in the Department of Industrial and Manufacturing Engineering in 1996 from Florida A & M University, where he was a visiting assistant professor. He earned his Ph.D. in Engineering Management from the Missouri University of Science and Technology (MUST) a Master's in Computer Integrated Manufacturing from the University of Birmingham in Great Britain.

Celestine Aguwa, Ph.D., is currently at WSU as an Associate Professor Research involved with research and teaching graduate courses in Industrial and Systems Engineering. His background includes engineering education, lean and value methodology in product development and advanced manufacturing, and customer voice analysis. He is currently working on several funded research projects in data analytics. He has cross-functional industrial experience at Ford Motor Company and extensive professional experience as an Architect. Dr. Aguwa has a Ph.D. and MSIE in Industrial and Manufacturing Engineering from the University of Pittsburgh and Massachusetts, Amherst, respectively. He also has a B.Arch. Degree in Architecture from the University of Nigeria. Dr. Aguwa has several awards, including a patent, and has written several published papers. He is a member of the Institute of Industrial & Systems Engineering (IISE), SAVE International, and the Institute of Operation Research and Management Sciences (INFORMS).

Sara Masoud, Ph.D., is an assistant professor of industrial and systems engineering at Wayne State University. She received her doctorate in systems and industrial engineering and master's in statistics from the University of Arizona in 2019. She received her bachelor's degree in industrial and systems engineering from Sharif University of Technology, Iran, in 2014. Her research focuses on mixed reality, virtual reality, augmented reality, and dynamic, data-driven application systems by utilizing machine learning, simulation, and optimization models in agro-industry, transportation, health care, and manufacturing. She is a member of the Institute of Industrial and Systems Engineers (IISE) and the Institute of Operation Research and Management Sciences. In 2019, she received the IISE Annual Meeting Best Paper Award in the Data Analytics and Information Systems track.

Suzan Arslanturk, Ph.D., is an Assistant Professor in the Department of Computer Science, Wayne State University. She has received her Ph.D. from Oakland University, Michigan. Her research interests and expertise fall in the broad area of health informatics and healthcare systems engineering. In particular, she is motivated to tackle challenging theoretical and applied research problems that enable developing innovative machine learning and clinical solutions to enhance individual and population health outcomes, improve patient care, and optimize the operational performance of healthcare delivery systems. Before joining Wayne State, she worked at a healthcare informatics and analytics services company in Michigan. In addition, she has authored several referred research papers and conference proceedings.

Egbe-Etu Etu is an Assistant Professor in the Department of Marketing & Business Analytics at San Jose State University, California. His current research focuses on healthcare operations management, emphasizing improving the emergency department's performance during a surge event. He is an active member of Industrial Engineering and Operations Management (IEOM), IISE, and SAVE International. He has received the IEOM Conference Best Paper Award in Healthcare Track.

Joseph Miller, MD., is an Emergency Medicine Physician at Henry Ford Hospital. He specializes in Emergency Medicine, Internal Medicine, and Clinical Research. Dr. Miller's research interest is on emergency neurological conditions, such as acute stroke, epilepsy, and traumatic brain injury. He also researches hypertensive emergencies and teaches research methodology within the health system.