

Unsupervised Machine Learning Elicits Patient Archetypes in a Primary Percutaneous Coronary Intervention Service

Aleeha Iftikhar
Computing, Engineering & Built Environment
 Ulster University Jordanstown
 Northern Ireland, UK
Iftikhar-a1@ulster.ac.uk

Raymond Bond
Computing, Engineering & Built Environment
 Ulster University Jordanstown
 Northern Ireland, UK
rb.bond@ulster.ac.uk

Victoria McGilligan
Centre for Personalised Medicine, Faculty of Life & Health Sciences
 Ulster University Jordanstown
 Northern Ireland, UK
v.mcgilligan@ulster.ac.uk

Khaled Rjoob
Computing, Engineering & Built Environment
 Ulster University Jordanstown
 Northern Ireland, UK
rjoob-k@ulster.ac.uk

Stephen J Leslie
Department of Diabetes & Cardiovascular Science
 University of the Highlands and Islands, Centre for Health Science,
 Inverness, UK
stephen.leslie@nhs.net

Charles Knoery
Department of Diabetes & Cardiovascular Science
 University of the Highlands and Islands, Centre for Health Science,
 Inverness, UK
c.knoery@nhs.net

Anne McShane
Letterkenny University hospital, Letterkenny, Donegal, Ireland
Anne.McShane@hse.ie

Aaron Peace
Western Health and Social Care Trust
 C-TRIC, Ulster University
 Northern Ireland, UK
a.peace@c-tric.com

Abstract—A primary percutaneous coronary intervention (PPCI) re-establishes blood flow in an obstructed coronary artery. PPCI referrals vary in admission criteria partly on the basis of ECG findings, hence, not all the referrals are accepted. The aim of the paper is to discover archetypes of accepted patients referred to the PPCI center. Cluster analysis was performed on a PPCI referral dataset to identify patient archetypes and identify any key patterns of patients who were accepted for PPCI. A k-means clustering algorithm was used with the elbow method for determining the optimum number of clusters (groups of patients). A silhouette plot was generated for within cluster validation. Among the accepted PPCI referrals, there were four different groups of patients. The patients within each group have similar characteristics. The largest cluster of patients include male patients being referred out of hours and with excessive door to balloon times (DTBTs) as compared to those referred in hours. Another cluster includes older female patients who are referred out of hour. Also, it was discovered that the false activation rate and DTBTs are higher in females as compared to male clusters. The smallest cluster include the most elderly patients in the whole referral dataset and mainly includes more males than female's who are referred out of hour and have the highest false activation rate, DTBT, and 30 days mortality rate. The cluster analysis of PPCI dataset revealed different patient archetypes. Each group of patients have a different mean age, out of hours referral rate, DTBT, false activation and 30 days mortality rate compared to other group. The identified clusters could be helpful for the clinicians to better understand their patients and utilize this information to aid the clinical decision making.

Keywords— *Clustering Methods, Machine Learning, K-means Algorithm, PPCI Referral, STEMI, Patient Archetypical*

The copyright notice: 978-1-7281-1867-3/19/\$31.00 ©2019 IEEE

I. INTRODUCTION

Acute chest pain patients are immediately assessed using a 12-lead electrocardiogram (ECG), which is performed and interpreted by paramedics, PPCI nurse activators or cardiologists. The ECG interpreters use the ST-segment elevation myocardial infarction (STEMI) criteria for the investigation of cardiac disorders by assessing the heart's electrical signals. Electrical changes in the heart correspond to depolarization and repolarization which leads to cardiac contraction and relaxation [1]. The ECG is a standardised diagnostic tool for cardiac care. When a patient is diagnosed with an acute myocardial infarction (AMI), they are referred to the catheterisation laboratory (CathLab) for a primary percutaneous coronary intervention (PPCI) for reperfusion. A preferred treatment for myocardial infarction with ST-segment elevation is PPCI which is effective for opening the infarct-related artery [2] [3][4]. On referral for PPCI, the activator (specialist nurse) decides whether to refer the case on to a cardiologist. The cardiologist then interprets the patient's case including the 12-lead ECG, demographics and other symptoms such as chest pain to make a decision on whether to accept or turn down the patient for referral to the CathLab. The admission criteria for PPCI is often variable, partly on the basis of poor ECG interpretation, hence not all the referrals are accepted. The strict criteria may result in some patients with acutely occluded coronary arteries not getting the treatment in time. It has been reported that a number of patients referred with negative STEMI criteria for PPCI still required angioplasty [5]. The time (in minutes) from the patient first being admitted to/entering a hospital to the time when a device is used to reopen the blocked artery is referred as door-to-balloon time. The door-to-balloon is a regulation which recommends as ≤ 90 minutes but now decreasing to ≤ 60 [6] [7]. Several measures have been used to try to monitor and

minimise the door to balloon time. It was concluded that the use of a dashboard named H2H (home to hospital) helped improve the overall performance and process and approximately 90% of the patients achieved the recommended door-to-balloon time of 60 minutes [8]. Many studies report different aspects and analysis of the PPCI pathway, however, there is a lack of analysis to unveil unique patterns or different types/groups with respect to the acceptance and turnaround of patients referred to the CathLab. One research study was conducted to analyse mortality of the PPCI referrals in relation to sex and race resulting no significant difference [9]. Lawesson et al. [10] presented the gender dependency on the symptoms in STEMI patients. They found that chest pain was less prevalent in women as compared to men, furthermore, other symptoms like shoulder, throat and back pain are as twice as common in men. Another study investigated whether the effectiveness of in-hospital and long-term outcomes for STEMI patients treated by PPCI within normal working hours compared with those treated out-of-hours (OOHs) [11]. De Boer et al. conducted research to estimate the clinical effects of PPCI versus fibrinolysis (FL) in connection with age and concluded that the reduction in clinical end points was not influenced by age [12]. Other research investigated the medical treatment, reperfusion trends and in-hospital mortality during STEMI in older adults. The study determined that patients undergoing PPCI or thrombolysis (TL) had lower in-hospital mortality compared with patients with no reperfusion (NR) [13]. Jakobsen et al. studied the outcome of PPCI in relation to age and sex and also compared sex and age specific mortality of PPCI patients with that of the general population. After 90 days post-PPCI, the mortality of treated patients was comparable to the mortality of the general population, independent of sex and age [14]. Another study [15] stated that age is associated with impaired myocardial perfusion, but not epicardial perfusion, after PPCI treating AMI. To further improve the outcome of elderly AMI patients, efforts should be aimed at improving myocardial perfusion beyond epicardial recanalization. In gender-based mortality analysis Gevaert et al. concluded that PPCI-treated women had a higher in-hospital mortality rate even after correcting for TIMI risk scores. The TIMI risk score was effective in predicting in-hospital mortality but performed slightly better in men [16].

Referral analysis is not only limited to PPCI patients, the literature also presents the analysis of several other patient referral data studies. A meta-analysis of individual patient data from seven medical centres in the United States, Europe, and Japan was performed for the primary outcome of all-cause mortality. Standardized mortality ratios and predictors of mortality were estimated. They concluded that a high mortality risk for systemic sclerosis and internal organ involvement and anti-topoisomerase antibodies turned to be important determinants of mortality [17]. Other research analysed the literature on the rates of referral for intervention or assessment of students from three racial groups: Caucasian, African American, and Hispanic resulting in no significant differences between the referral rates of Hispanic patients and Caucasian patients whereas when comparing African American with Caucasian patients, the mean risk ratios comparing the referral rates were significantly different from zero [18]. This study [19] investigated relative occurrence, age at presentation, sex distribution, and skeletal distribution of malignant soft-tissue tumors. Literature shows almost every

field of medicine have analysed patient referrals for multiple investigations. Similarly, machine learning (ML) techniques have also been used to analyse patient referrals for patient groupings. Cluster analysis (unsupervised learning) is analysis of data with the absence of category information. The objective of cluster analysis is to discover homogeneous clusters in the data revealing new insights, patient archetypes and new knowledge including new classifications or labels [20]. The approach used within this current work has also been used in previous studies. This approach uses the k-means clustering algorithm for feature extraction to examine and determine STEMI in the ECG signal of a person [21]. The primary aim of this current work was to use the k-means clustering method (a form of unsupervised machine learning) on a PPCI dataset to investigate homogenous clusters of patients to reveal patient archetypes.

II. METHODS

A. Dataset

This study involved analysis of an anonymised dataset from a cardiac care unit (CCU) at Altnagelvin Hospital (Northern Ireland, UK) for consecutive STEMI accepted patients from September 2014 to February 2019. The total study population consisted of 980 patients that were accepted for a PPCI intervention (n=714, 73% men). This referral data was routinely recorded by a staff nurse using a paper-based form and then digitised using a spreadsheet. Data contained some inconsistencies and missing values. The removal of some records and missing data imputation were applied on the dataset. For data imputation, any observation (patient) with 2 or more missing variables (NAs) was removed resulting in 828 observations. For data imputation the Multivariate Imputation by Chained Equations (MICE) package in R was used. The MICE package implements a method to deal with missing data. The package creates multiple imputations (replacement values) for multivariate missing data. The method is based on Fully Conditional Specification (FCS) implemented by the MICE algorithm as described in Van Buuren and Groothuis-Oudshoorn [22]. Each variable has its own imputation model. Built-in imputation models are provided for continuous data (predictive mean matching, normal), binary data (logistic regression), unordered categorical data (polytomous logistic regression) and ordered categorical data (proportional odds). MICE can also impute continuous two-level data (normal model, pan, second-level variables) [23].

For the '30 days mortality' variable a total of 219 values were imputed using Mode imputation. Whereas, for the DTBT variable a total of 320 values were imputed using MICE. The data cleansing was also performed where all the observations with any NA value were removed, as a result 508 observations were available. After the approval of an ethics application to undertake this analysis, the staff nurse removed all personal identifiable information such as names, date of birth and unique patient identifiers.

B. Data analysis

Clustering analysis entails statistically grouping similar cases in such a way that cases in a group (known as a cluster) are more similar to each other as compared to the other groups (clusters). This called within group similarity. In this study, clustering is used to help better understand the relationships between different groups of patients accepted for PPCI. This

information will be used to unveil the patterns and patient archetypes that exist within each of the clusters

C. Feature used

Only the patients accepted for PPCI were used (980 referrals collected in 2014–2019).

Patients were clustered using five attributes or features that we believe characterize a PPCI patient:

- Age
- Sex
- Out of hours
- 30 days mortality and
- Door to balloon time (DTBT)

We selected these features due to their explanatory power. Literature shows that gender and age distribution and their effects has been an interesting and important area of research [24][16][10][25]. Most of the PPCI being referred out of hours could be critical information for the CathLab service availability 24/7. The door to balloon time (DTBT) is an important regulation for any patient going to the CathLab, and the literature includes multiple studies that measure, monitor and control DTBT [11] [26] [27][28] [29][30] as it can be a KPI for STEMI patient care. Similarly, 30 days mortality is an obvious key variable in PPCI. From a machine learning perspective, these features provide the smallest and simplest possible feature set that captures both the patient and limited treatment information for each individual.

D. Clustering technique

Before running clustering algorithms on the data, each variable was standardized, i.e. each of the five features were centred at zero and scaled using variance. For clustering, the K-means clustering algorithm was used given that is the most widely used and established clustering algorithm in the unsupervised machine learning literature. Before applying the clustering algorithm, it is fundamental to choose the quantity of cluster centroids. We applied the heuristic elbow strategy, as exhibited in Figure 1, which takes the total within-cluster sum of squares as a function of the number of clusters. We then selected the number of clusters with the aim that adding or removing another cluster does not affect the total within-cluster sum of squares. Using elbow method, we distinguished that four a reasonable number of clusters that will demonstrate the groups in terms of explained variability. Throughout our clustering analysis, sums-of squares of the relevant deviation are used to measure statistical variability. Euclidean distance was used as a natural measure of deviations for our selected four features. The total sum of squares was computed using the deviations of individual patient from the centroid of the whole dataset. The within-groups sum of squares is computed using the distance of an individual patient from the centroid of the assigned cluster. The between groups sum of squares uses the distances of cluster centroids from the dataset centroid and is numerically equal to the difference between the total sum of squares and the within-groups sum of squares.

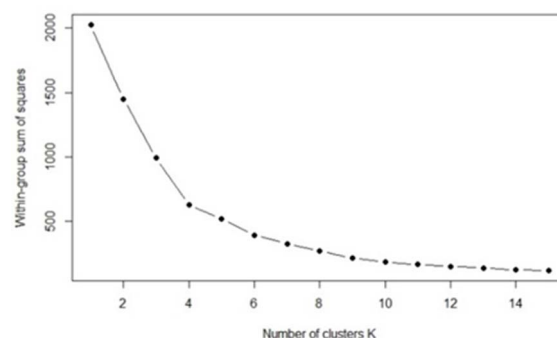


Figure 1. Elbow method illustration: the within-groups sum of squares drops from three clusters to four and comparatively flattens out for all the higher numbers of clusters. This suggests K = 4 is the best number of clusters for the problem at hand

Using elbow method (when using the plot shown in Figure 1), we consider the height of the curve at each K as the fraction of its height at K = 1 (leftmost dot on all elbow method plots). The elbow method is a visual inspection approach where we select k (the number of clusters) when we see the first clear elbow joint (when together with the previous point and subsequent point creates a line that looks like a human arm). The K-means clustering algorithm uses random sampling. The algorithm (implemented in R) outputs clusters as a numbered sequence, in order of extraction. The order in which the clusters are extracted varies depending on which slice of the dataset is inspected. Cluster package was used for the cluster analysis.

E. Ethical aspects

Permission for the study was obtained from the regional Ethical Review Board IRAS 251710, NHS (ORECNI) and complied with the Declaration of International Research Integrity Association [12].

III. RESULTS

A. Clusters

The elbow method for different portion of dataset as shown in figure 1, mostly resulted in K=4 as the optimal number of clusters. The four-cluster solution for the PPCI patients is explained in detail below. Figure 2 shows the clustering results with four-cluster split. Table 1 shows the cluster (type of PPCI referral patients) and their features.

1. *Cluster number*: number of clusters.
2. *Cluster size*: number of patients captured in the cluster.
3. *Mean age*: in-cluster average age of patients within that cluster.
4. *Sex*: sex distribution and dominance within that cluster.
5. *Out of hour*: number of referral within/without working hour within that cluster.
6. *False activation*: number of false activations within each cluster.
In accepted group there could be true activations as well as false activation (mainly patients who went to the CathLab, but no blockage was found).
7. *30days motility*: mortality rate within that cluster.
8. *Avg. DTB*: mean door to balloon time within that cluster.

Figure 3 shows the mean age and DTBT within each cluster with std. error bars whereas figure 4 depicts the

boxplot for the average door to balloon time and age within each cluster. Similarly, figure 5 shows the overall clusters distribution over each feature listed above

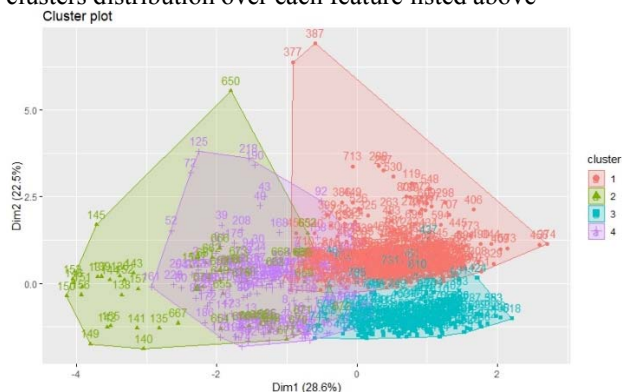


Figure 2 Clustering results. show four-cluster split.

Tables 1 show cluster averages for two features: age, door to balloon time, sex, 30days mortality, as well as cluster sizes

Cluster	Cluster name	Cluster size	Age	Sex	Out of hour	False Activation	30days Mortality	Avg DTB
1	Out of hour male	357	61	Male	100%	2.8%	0%	40m ins
2	Mortality group	53	73	30M : 23F	60%	22.6%	100%	42m ins
3	In hour males	210	60	Male	0%	1.4%	0%	30m ins
4	All females	202	70	Females	63%	5.9%	0%	39m ins

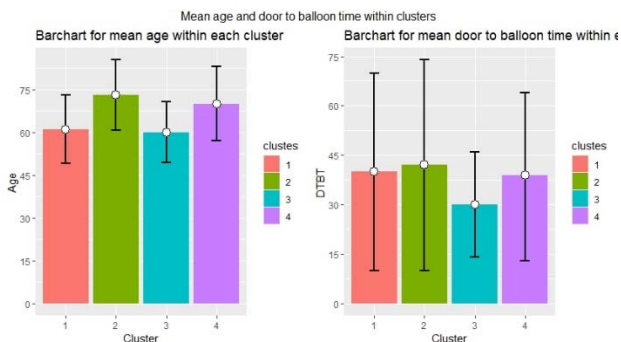


Figure 3: The mean age and DTBT within each cluster with std. error bars

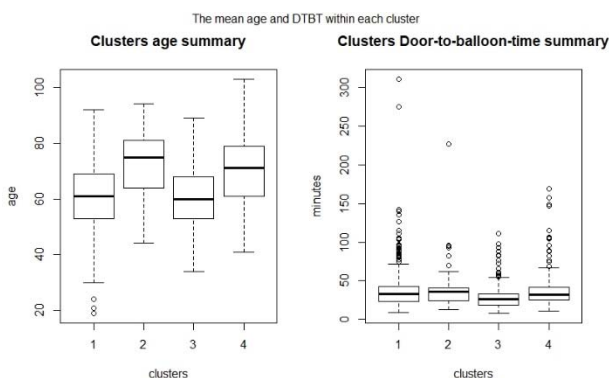


Figure 4: The Clusters distribution over age and DTBT feature

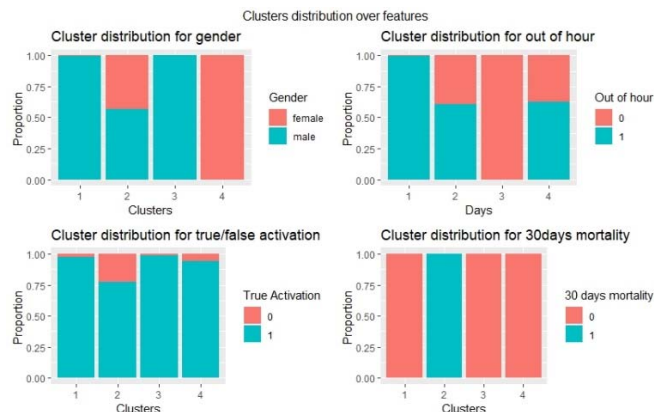


Figure 5: The Clusters distribution over each feature

B. Cluster analysis

Tables 2 show cluster name, description, as well as cluster sizes

Cluster no.	Cluster name	Cluster size	Cluster description
1	Out of hour older males	357 (43%)	This is the largest cluster representing the majority of patients. This cluster represents the “out of hour males”. The mean age for this group is 61 years old (SD=12) with a very low rate (2.8%) of false activations (patients who went to the CathLab, but no blockage was found) and a zero rate of 30 days mortality with the average of 40 mins (SD=30) door to balloon time. In summary, these represent out of hour males who are treated in good time and survive beyond 30 days.
2	Older mortality patients	53 (6%)	This is the smallest and most interesting cluster as it comprises of both males (57%) and females (43%) who all die within 30 days. This group are the oldest in age i.e. mean age of 73 years (SD =12) but also with the highest mean DTBT (42mins±32) with most of the patients referred out of hour. This group also have the highest false activation rate of 22.6%.
3	In hour rapidly reperfused younger males	210 (26%)	This is the second largest cluster representing males who are referred in-hour and who are treated sooner than any other group (mean DTBT 30 mins±16) with a 0% 30 days mortality rate. This group has a mean age of 60 years (SD=10) with only a 1% false activation rate.
4	Older females	202 (25%)	This cluster represents all females and ~63% of the patients are referred out of hour and the mean age for this group is higher than the male groups i.e. 70 years

			(SD=13). Although, 30 days mortality rate in this group is also zero but this group contains 5.9% of the false activations with (39mins±25) of mean DTBT.
--	--	--	---

The cluster analysis grouped PPCI referral patients into 4 clusters. The largest cluster named “out of hour males” shows the overall dominance of males who are referred out of hour with a mean age of 61 (SD=12) with a very low rate (2.8%) of false activation (patients who went to the CathLab, but no blockage was found) and 0% rate of 30 days mortality with the average DTBT of 40 mins. The second largest cluster was again a male dominant cluster where all the referrals were made in hour. This group had mean age 60 years old with almost no false activations and a 0% rate of 30 days mortality with the least average DTBT of only 30 mins. The third largest cluster named “all females” - as the name suggests this cluster is female dominated. 63% of the patients were referred out of hour with the mean age for this group being higher than the male groups i.e. 70 years. Although, 30 days mortality rate in this group is also 0% but this group contains 5.9% of false activations with a mean 39mins±25 DTBT. The fourth and the smallest cluster of PPCI patients is the only group that comprises of both males (57%) and females (43%) all of which die within 30 days. This group has the highest average age i.e. 73 years old±12 with the highest mean DTBT i.e. 42mins±32, this group also has the highest false activation rate of 22.6% with 100% 30 days mortality. Looking at these clusters we can identify different archetypes of patients referred for a PPCI / CathLab intervention. Among all the patients, the group of patients with highest mean age were mostly referred out of hour and experienced the maximum DTBT and had the highest 30 days mortality rate. False activation was found more common in female dominant clusters.

C. Clustering validation and evaluation

Cluster validation is basically a procedure to evaluate the validity of the results of the clustering solution. The silhouette plot is a measure used to evaluate and validate the clusters distributions. Silhouette (SiSi) measures the similarity of an object with the other objects of its own cluster against the objects in neighbour clusters. SiSi values range from 1 to -1. SiSi values close to 1 refers to well clustered solution whereas close to -1 indicates that the object is poorly clustered, and that assignment to some other cluster would probably improve the overall results. Silhouette plot was generated using K-means clustering. Figure 6 shows the cluster silhouette plot with an average value 0.4. The plot shows that all the observations have positive values which supports that the solution is well fitted within the clusters. There are no negative values in any of the clusters which again validates the clustering results.

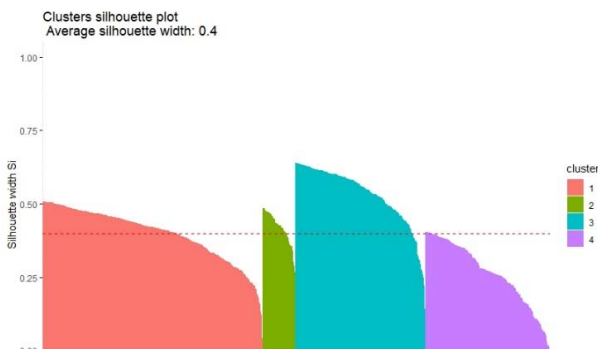


Figure 6: The clustering silhouette plot

IV. DISCUSSION

Knowing the types of different patients and their characteristics could be an interesting information for the clinicians in cardiac care. The patient’s categorization might aid clinicians in decision making for accepting or turning down the patients for the CathLab. In this research the clustering analysis was performed with a cleaned up original PPCI patient dataset. Whilst a small portion of the data included missing data imputations, we carried out the analysis using completed case analysis (where there was no data imputations) and the analysis resulted in the same cluster features for both the datasets (imputed and original/complete case analysis). Hence, the results do not change when removing all cases with any NA.

It was observed that PPCI dataset is dominated by the male gender. All the male patients who were referred in hours had the least (0%) false activation rate and a 0% 30 days mortality rate. They also attained the least DTBT which is 30mins. Subsequently, the males who referred out of hour have low false activation but 0% 30 days mortality however they took little more DTBT which is 40 mins $P < 0.001$. We uncovered a group of female patients who are older than both the male groups. Among this group about 63% were referred out of hour which also led to false activations (5.9%). The 30 days mortality rate however in this group is 0%. The DTBT is also slightly higher than the male group. From the results, we can conclude that females are older than males and tend to be referred more out of hour as compared to male groups, also there are more false activations in the female clusters taking more DTBT ($P < 0.001$) as compared to male clusters. The smallest and most interesting cluster represents the oldest patients with 100% 30 days mortality and highest DTBT. This is the only cluster that includes both males and females, slightly more than half are male patients and around 60% of the patients were out of hour. The mortality cluster also had highest percentage of false activations. For the validation of these clustering results the silhouette plot were used to measure the similarity of an object with the other objects. SiSi values range from 1 to -1. SiSi values close to 1 refers to well clustered and our clusters silhouette plot with average value 0.4 which can be considered as good result. The objective of cluster analysis is to discover homogeneous clusters in the data revealing new insights, patient archetypes. The approach used within this current work has also been used in previous studies for different medical field. However, the patient’s archetypes for the PPCI referral couldn’t be found in the literature. The primary aim of this current work was to use the k-means clustering method (a form of unsupervised machine learning) on a PPCI dataset to investigate homogenous clusters of patients to reveal patient archetypes. These findings are based on one dataset of a single hospital of Northern Ireland which can limit the results and may not be a depiction of all PPCI services. However, in future analysis, enhancement has been planned in terms of including the dataset from other UK hospitals.

V. CONCLUSION

The PPCI referral data analysis provided four patient archetypes. The clustering solution depict the prevalence of male patients who are being referred out of hour taking more

DTBT as compared to those referred in hour ($P < 0.001$). Female patients are older and referred more out of hour. Also, the false activation rate and DTBT are higher in females as compared to male clusters ($P < 0.001$). The cluster with the oldest patient's majority referred out of hour, have highest false activation rate and DTBT with 100% 30days mortality rate.

ACKNOWLEDGMENT

This research is supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB).

DISCLAIMER

The views and opinions expressed in this paper do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB).

REFERENCES

- [1] S. Meek and F. Morris, "ABC of clinical electrocardiography. Introduction. I-Leads, rate, rhythm, and cardiac axis.," *BMJ*, vol. 324, no. 7334, pp. 415–8, Feb. 2002.
- [2] P. Widimský, "Long distance transport for primary angioplasty vs immediate thrombolysis in acute myocardial infarction Final results of the randomized national multicentre trial—PRAGUE-2," *Eur. Heart J.*, vol. 24, no. 1, pp. 94–104, Jan. 2003.
- [3] E. C. Keeley, J. A. Boura, and C. L. Grines, "Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials," *Lancet*, vol. 361, no. 9351, pp. 13–20, Jan. 2003.
- [4] S. Silber *et al.*, "Guidelines for Percutaneous Coronary Interventions," *Eur. Heart J.*, vol. 26, no. 8, pp. 804–847, Apr. 2005.
- [5] A. Apps *et al.*, "High incidence of acute coronary occlusion in patients without protocol positive ST segment elevation referred to an open access primary angioplasty programme.," *Postgrad. Med. J.*, vol. 89, no. 1053, pp. 376–81, Jul. 2013.
- [6] "National Clinical Guideline Centre Myocardial infarction with ST-segment elevation The acute management of myocardial infarction with ST-segment elevation."
- [7] "7 Day Services Clinical Guidance-STEMI"
- [8] O. Drexhage *et al.*, "Call-to-balloon time dashboard in patients with ST-segment elevation myocardial infarction results in significant improvement in the logistic chain," *EuroIntervention*, vol. 13, no. 5, pp. e564–e571, 2017.
- [9] A. Krishnamurthy *et al.*, "Clinical outcomes following primary percutaneous coronary intervention for ST-elevation myocardial infarction according to sex and race," *Eur. Hear. J. Acute Cardiovasc. Care*, p. 204887261773580, 2017.
- [10] S. Sederholm Lawesson, R.-M. Isaksson, I. Thylén, M. Ericsson, K. Ångerud, and E. Swahn, "Gender differences in symptom presentation of ST-elevation myocardial infarction – An observational multicenter survey study," *Int. J. Cardiol.*, vol. 264, pp. 7–11, Aug. 2018.
- [11] K. S. Rathod *et al.*, "Out-of-hours primary percutaneous coronary intervention for ST-elevation myocardial infarction is not associated with excess mortality: a study of 3347 patients treated in an integrated cardiac network.," *BMJ Open*, vol. 3, no. 6, p. e003063, Jun. 2013.
- [12] S. P. M. de Boer *et al.*, "Mortality and Morbidity Reduction by Primary Percutaneous Coronary Intervention Is Independent of the Patient's Age," *JACC Cardiovasc. Interv.*, vol. 3, no. 3, pp. 324–331, Mar. 2010.
- [13] H. M. Medina *et al.*, "Reperfusion Strategies and Quality of Care in 5339 Patients Age 80 Years or Older Presenting With ST-Elevation Myocardial Infarction: Analysis from Get With The Guidelines-Coronary Artery Disease," *Clin. Cardiol.*, vol. 35, no. 10, pp. 632–640, Oct. 2012.
- [14] L. Jakobsen *et al.*, "Sex- and age-related differences in clinical outcome after primary percutaneous coronary intervention.," *EuroIntervention*, vol. 8, no. 8, pp. 904–911, Dec. 2012.
- [15] J. Pu *et al.*, "Comparison of epicardial and myocardial perfusions after primary coronary angioplasty for ST-elevation myocardial infarction in patients under and over 75 years of age," *Aging Clin. Exp. Res.*, vol. 22, no. 4, pp. 295–302, Aug. 2010.
- [16] S. A. Gevaert *et al.*, "Gender, TIMI risk score and in-hospital mortality in STEMI patients undergoing primary PCI: results from the Belgian STEMI registry," *EuroIntervention*, vol. 9, pp. 1095–1101, 2014.
- [17] J. P. A. Ioannidis *et al.*, "Mortality in systemic sclerosis: An international meta-analysis of individual patient data," *Am. J. Med.*, vol. 118, no. 1, pp. 2–10, Jan. 2005.
- [18] J. L. Hosp and D. J. Reschly, "Referral Rates for Intervention or Assessment," *J. Spec. Educ.*, vol. 37, no. 2, pp. 67–80, Aug. 2003.
- [19] M. J. Kransdorf, "Malignant soft-tissue tumors in a large referral population: distribution of diagnoses by age, sex, and location.," *AJR. Am. J. Roentgenol.*, vol. 164, no. 1, pp. 129–34, Jan. 1995.
- [20] A. K. Jain, "Data Clustering: 50 Years Beyond K-means," in *Machine Learning and Knowledge Discovery in Databases*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 3–4.
- [21] Bensusin, C. Kezi Selva vijila, and C. Hubert, "Detection of st segment elevation myocardial infarction (stemi) using bacterial foraging optimization technique," *Int. J. Eng. Technol.*, vol. 6, no. 2, pp. 1212–1223, 2014.
- [22] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *J. Stat. Softw.*, vol. 45, no. 3, 2011.
- [23] "Package 'mice' Title Multivariate Imputation by Chained Equations," 2019.
- [24] M. GAUB and C. L. CARLSON, "Gender Differences in ADHD: A Meta-Analysis and Critical Review," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 36, no. 8, pp. 1036–1045, Aug. 1997.
- [25] F. M. Chen, G. E. Fryer, and T. E. Norris, "Effects of comorbidity and clustering upon referrals in primary care.," *J. Am. Board Fam. Pract.*, vol. 18, no. 6, pp. 449–52, Nov. 2005.
- [26] E. H. Bradley *et al.*, "Achieving Door-to-Balloon Times That Meet Quality Guidelines," *J. Am. Coll. Cardiol.*, vol. 46, no. 7, pp. 1236–1241, Oct. 2005.
- [27] "Door to balloon times for STEMI - BCIS - BCIS." [Online]. Available: <https://www.bcis.org.uk/patient-area/door-balloon-times/>. [Accessed: 05-Jun-2019].
- [28] J. T. Levis, M. P. Mercer, M. Thanassi, and J. Lin, "Factors Contributing to Door-to-Balloon Times of ≤90 Minutes in 97% of Patients with ST-Elevation Myocardial Infarction: Our One-Year Experience with a Heart Alert Protocol.," *Perm. J.*, vol. 14, no. 3, pp. 4–11, 2010.
- [29] A. Taddei *et al.*, "Information system for assessing health care in acute myocardial infarction," *2012 Comput. Cardiol.*, pp. 205–208, 2012.
- [30] R. Estévez-Loureiro *et al.*, "Direct transfer of ST-elevation myocardial infarction patients for primary percutaneous coronary intervention from short and long transfer distances decreases temporal delays and improves short-term prognosis: the PRÓGALIAM Registry," 2010.