Cardiac Arrest Prediction using Machine Learning Algorithms

R Karthikeyan1*, D. Vijendra Babu2, EKarthik3, R. Suresh4, MNalathambi5, S. Dinakaran6

1Asst. Professor(Gr.-II), Department of Electronics and Communication Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission’s Research Foundation-Deemed to be University, Paiyanoor, Kancheepuram, TamilNadu, India
2Professor & Vice Principal, Department of Electronics and Communication Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission’s Research Foundation-Deemed to be University, Paiyanoor, Kancheepuram, TamilNadu, India
3,4,5UG Student, 6Professor, Department of Electronics and Communication Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission’s Research Foundation-Deemed to be University, Paiyanoor, Kancheepuram, TamilNadu, India

Email:*karthikkanda95@gmail.com

Abstract. Cardiac arrest and other cardiovascular problems are the most prevalent issue among millions of men, and there are numerous causes that function as the basis of this crisis, such as people's wellbeing, mainly because of job stress, exhaustion, bad food quality, and an elevated cholesterol level as a consequence of the lack of technology cardiac disease. Many scientific and medical support programs change every day, yet every program has its own special features, advantages and disadvantages. The goal of this article is to research the probability of cardiac arrest based on various regulated or unregulated variables in specific data set machine learning algorithms.

Keywords: Cardiac arrest; diagnosis; medical supporting systems; machine learning algorithms

1. Introduction

Human body consists of species very susceptible to heart disease, brain disease, kidney disease, liver disease etc. One of the most significant disease resistant is cardiovascular disorder, given regular lifestyle, which results in pressure in the lungs, stroke and other signs attributable to damaged blood vessels. There are several causes that can or cannot regulate heart disease. For example, it is not feasible to regulate age, ethnicity, and family background though obesity, physical activity and diet may be regulated. Some physicians are also utilizing a machine-learning program for cardiac disease prediction, while most are not. According to the WHO, the world’s leading cause of death for cardiovascular disease is CVD and people with CVD or high-risk needs to be detected early to address this issue. Artificial intelligence is one of the approaches to this issue by utilizing machine learning to estimation the percentage of risks.

A number of classification algorithms, including Random Forest, Decision Tree, Linear Regression, SVM and Artificial Neural Networks (ANN) in machine learning are accessible in predictive mode. ANN is the strongest tool in artificial intelligence owing to its brain like functioning, which is the basic unit of the network and a set of inputs feeding on a neuron that provides output. ANN is the most effective tool in Artificial Intelligence. If we were able to predict cardiac arrest by using a certain algorithm, that would be a milestone both in engineering as well as in medical sciences. It does not only count as progress, but also could help to prevent heart arrest death to a significant degree and directly affect the life expectancy of a person.
2. Related Work

We used a prospective, population-based, nationwide registry of OHCA in Japan and a weather forecast dataset from Weather Company in United States. We developed OHCA prediction models from 2005 to 2013 by using the generalized linear models (GLM) which assume the Poisson distribution and evaluated whether the prediction model can work by validation datasets between 2014 and 2015 in the next few years. The results are available at the same time. Then, the Extreme Gradient Boosting (XGBoost) algorithm evaluated whether the prediction model [1] has been developed for machine education. We assessed the prediction models with a mean absolute percentage error (MAPE) that is a measure of previous results. OHCA cases handled by Dallas-FortWorth Centre of Resuscitation Science were examined. Patients of at least two minutes of ROSC were included. The presence of life-threatening ECG and/or chest compression [2] was determined by the re-arrest within 12 minutes of ROSC. A total of 18 HRV-features were calculated for 1 minutes and 2 minutes after ROSC. Tools were integrated into a list of 100 trees in Random Forest (RF) to forecast re-arrest events. The model testing and the results in terms of region under curved (AUC) field was accompanied by ten times cross-validation of thirty repetitions. The low survival rate of people with sudden heart arrest is one of the major issues in healthcare. Late cardiac arrest may have opportunities to interfere to avoid the occurrence to decrease mortality [3]. The heart arrest was expected by conventional mathematical techniques. Often they analyzed differences in group levels using a limited number of variables [4]. At the other side, machine learning methodology, which is part of a rising wave in predictive medical research, has generated customized predictive analytics on more complicated data and achieved promising results.

Several causes are believed to influence neural recovery and long-term operation following cardiac arrest (OHCA). Past findings found that the latter outcomes were closely linked to both pre-hospital and historical circumstances and the health status of hospitals [5]. The aim of this research is to define and analyze early prediction clinical variables of unconscious survivors after OHCA using the use of machine learning stats. Cardiac arrest is a critical problem for the health of patients with a ~80 percent in hospital mortality marked by lack of traceability, a loss of control as well as an apnea. Accurate assessments of high risk cases are important not only to increase the overall summary performance, but also to increase the [6] quality of life, since cardiac arrest cases have significant neurological consequences.

Research has focused on showing static risk scores without the physiological condition of the patient being taken into account. In this analysis, we are using the Multichannel Secret Markov Model to incorporate an optimized concept of sequential contrasts. Within the discrete-time survival research system [7], socioeconomic factors, test results, and vital signs were used to determine the cumulative outcome of cardiac attack, intensive unit transition and death. Two logistic regression models were contrasted with numerous methods of learning (one using linear and the second using limited cubic splines) [8]. The equations have been extracted by the date in the first 60% and then checked by the next 40%.

High mortality is related to resuscitate cardiac arrest; however, the likelihood of adverse effects dependent on current disease severity values is restricted. We have developed more accurate models of predictive risk using both logistic regression (LR) technology and machine learning (ML), using in-hospital data [9] available within the initial 24-hour reception, using a combination of demographic, physiological and biochemical data [10]. Heart attack is a life-threatening, sometimes lethal event. Whilst some of the cardiac arrests are recognized as somewhat inevitable by physicians, often after an examination, the remainder remains difficult to identify. [11] Differences in physiology in certain patients can therefore anticipate severe impairments contributing to heart or respiratory arrests when studied in depth. To order to determine the predictability of cardiac arrests to seriously ill paediatric patients in intensive care, the goal is to take account of causally-related shifts in signs such as heart rate, respiration rhythm, [12]systolic blood pressure and peripheral cutaneous oxygen
3. Proposed Method

3.1. Parameters Used

The parameters used to apply an algorithm were modifications, age, chestpain, remabp, chol, fbs, RestECG, MaxHR, ExAng and Oldpeak. The used data collection had missing values to be removed before the algorithms were applied. The absent values in the dataset have been removed by replacing the spaces with mean column values [13]. The outliers dataset was checked and during the mapping, none was discovered. Various algorithms are explained in the following section, including support for a vector machine, random forest, decision tree, logistic regression and the artificial neural network.

3.2. Support Vector Machine:

The vector machine follows an algorithm of classification. It is an algorithm for the controlled computer analysis. If the data is increasing, it is preferable. The kernels are one of the key factors for the usage of the algorithm. Specific kernel functions for the decision feature may be defined when implementing SVM [14]. Popular kernels are available, but custom kernels may also be listed that make them more flexible. The data is predicted in a wide field in SVM, which classifies these data as the data is linearly inseparable. There is a separator between the classes in front of the separator, which draws the hyperplane. The algorithm will then predict the class of inputs to be in as the next input is entered. Help vectors are observation points, while the SVM differs from each other. Figure 1 displays Graphical reorientation of Support Vector Machine Algorithm.

![Figure 1: Graphical reorientation of Support Vector Machine Algorithm](image)

3.2. Random Forest:

Random forests are flexible algorithms, which in their simplicity and benefit from their use in regression and classification are commonly used in machine learning. Forest is a multi-decision tree set that takes into account the introduction of additional alterability to the structure of trees and other random function subsets [15]. Random Forest finds the best function in the random subset, resulting in a large variety of outcomes, that is to say better versions. For evaluating the relative worth of properties, Sk learn provides the criteria. The issue of regression is overposition and can be minimized by means of this method because the normal decision-making processes the probability of over position. Another issue with regression is being a large variance not prevalent in a random forest when utilizing several trees, the possibility of stumbling upon a classifier that does not perform well because of the relation between train and test results.

3.3. Decision Tree:

The Decision Tree Algorithm has been developed to resolve regression and classification issues for supervised learning algorithms. The key benefit of decision-making bodies is that both computational and classification data can be analyzed. The training model is used to decide the value or class of the
target/attribute as well as other regular algorithms decision tree algorithms but in this case, the learning decision rules of the prior training date are employed. The algorithm uses a tree structure that is sometimes called a decision node. The algorithm and each inner node is equivalent to two or more leaf nodes. The largest node, or root node, is the greatest indicator of the data collection. This algorithm divides the whole data set into fragments or subparagraphs and generates a tree with leaf nodes, internal nodes and root nodes. It creates a decision tree. The model gets fit and fitter as its tree becomes larger and more complex. Figure 2 display Dataflow of Random Forest Algorithm and Figure 3 dataflow of Decision Tree Algorithm.

3.4. Logistic Regression:
Regression in logistics is a regression type, in that "y" is the objective variable (binary) in linear weighing or the input variable values 'X.' In this case the objective variable is 'Decision,' which determines whether a person is likely to be arrested cardiacly and can call a doctor. The best curve for
a vector 'x' oriented variable, which contains different parameters depending on that curve, is obtained. The likelihood function is the function which is subsequently converted into binary values (0, 1) for actual probability estimation after the training of a data collection. One important reason is that more than an explaining variable can be included, which can be dichotome, ordinal or continuous. The explanatory variable can therefore be considered. A quantified value for the strength of the association adjustment of other variables also provides logistic regression. Figure 4 Data position of Logistic Regression Algorithm in details.

4. Result And Discussion
The used data collection had missing values to be removed before the algorithms were applied. The absent values in the dataset have been removed by replacing the spaces with mean column values. The outlier’s dataset was checked and during the mapping, none was discovered. Various algorithms are explained in the following section, including support for a vector machine, random forest, decision tree, logistic regression and the artificial neural network. There are different types of machine learning algorithm (SVM, Decision Tree, Random Forest, Logistic Regression and ANN) is applied and accuracy are displayed in Figure 5, Figure 6, Figure 7, Figure 8 and Figure 9 with details.
Figure 5: Precision obtained (using SVM: 53.8462 percent)

Figure 6: Precision obtained (Random Forest: 59.4405 percent)
Figure 7: Precision obtained (DecisionTree: 49.0566 percent)

Figure 8: Precision obtained (Logistic Regression: 56.3106 percent)
Figure 9: Precision obtained (ANN: ~85.00%)

Table 1: Accuracy Table

| S.No | ALGORITHM         | ACCURACY |
|------|-------------------|----------|
| 1    | Artificial Neural network | 85.4%    |
| 2    | Logistic Regression  | 57.310%  |
| 3    | Decision Tree       | 50.5%    |
| 4    | Random Forest       | 60.5%    |
| 5    | Support Vector Machine | 53.8%    |

Table 1 shows the different types of machine learning algorithm (SVM, Decision Tree, Random Forest, Logistic Regression and ANN) with accuracy are displayed. The performance of the neural network was considered to be the best. When the neural network learns during each run, the mean value of the precision outcomes was taken and used as the absolute precision of ANN. When applying the algorithm to avoid heart arrest, the only drawback encountered is that the data collection is not accessible greater than the one used. Had a bigger dataset been available, the neural network may have been more adequately trained and the tests more reliably than they are now. A much wider dataset may be used to boost the ANN’s and the times and unknown layers that would enhance the accuracy and accurate performance of the neural network.

5. Conclusion

The article represents article is to research the probability of cardiac arrest based on various regulated or unregulated variables in specific data set machine learning algorithms. Following the implementation of Help Vector Machine, Random Forest, Logistic Regression and Artificial Neural Network algorithm on a dataset, it was observed that the Artificial Neural Network precision is higher (~85%) to predict the frequency of cardiac arrest. Also, the algorithm's accuracy is poor because the dataset is small. If ANN has been more reliable than the actual average value of the test should have been. Has the data collection been greater? The Artificial Neural Network will therefore be used as a framework for forecasting cardiac arrest in patients based on data collection algorithms. When the neural network learns during each run, the mean value of the precision outcomes was taken and used as the absolute precision of ANN. When applying the algorithm to avoid heart arrest, the only drawback encountered is that the data collection is not accessible greater than the one used. Had a bigger dataset been available, the neural network may have been more adequately trained and the tests are more reliable.
References

[1] Nakashima, Takahiro, Soshiro Ogata, Teruo Noguchi, Satoshi Yasuda, Yoshio Tahara, Ken Nagao, Kunihiro Nishimura, Hiroshi Nonogi, Koji Ishara, and Sunao Kojima. "The Cardiac Arrest Prediction Model Based on Weather Forecast by Machine Learning." *Circulation* 140, no. Suppl_1 (2019): A13297-A13297.

[2] Elola, Andoni, Elisabete Aramendi, Unai Irusta, Naroa Amezaga, Jon Urteaga, Pamela Owens, and Ahamed H. Idris. "Machinery for forecasting cardiac arrest in outpatient environment Machine learning technologies." *Circulation* 140, no. Suppl_2 (2019): A127-A127.

[3] Javan, Samaneh Layeghian, Mohammad Mehdi Sepehri, and Hassan Aghajani. "To interpret and synthesise early-hearted research using machine learning based on a multi-layered integrative architecture." *Journal of biomedical informatics* 88 (2018): 70-89.

[4] Kim, Hanbi, Hieu Nguyen, Qingchu Jin, Sharmila Tamby, Tatiana Gelaf Romer, Eric Sung, Ran Liu et al. "19: A CARDIAC ARREST LEARNING-BASED ARREST USING A LASSE MULTICENTER PREDICTION." *Critical Care Medicine* 48, no. 1 (2020): 10.

[5] Johnsson, J., S. Hoerberg, A. Holm, S. Gustafzelius, J. Dankiewicz, and N. Nielsen. "5228 Identification of reliable clinical variables for early prediction of outcome after out-of-hospital cardiac arrest using algorithm-based machine learning statistical methods." *European Heart Journal* 40, no. Supplement_1 (2019): ehz746-0076.

[6] Akrivos, E., V. Papaioannou, N. Maglaveras, and I. Chouvarda. "Prediction of heart arrest by machine learning of intensive care patients." In *International Conference on Biomedical and Health Informatics*, pp. 25-29. Springer, Singapore, 2017.

[7] Churpek, Matthew M., Trevor C. Yuen, Christopher Winslow, David O. Meltzer, Michael W. Kattan, and Dana P. Edelson. "Comparison multicenter of computer training procedures and classical regression for clinical ward prediction." *Critical care medicine* 44, no. 2 (2016): 368.

[8] Nanayakkara, Shane, Sam Fogarty, Michael Tremeer, Kelvin Ross, Brent Richards, Christoph Bergmeir, Sheng Xu et al. "A historical international registry study: characteristic risk of inhospital mortality after cardiac arrest using machine learning." *PLoS medicine* 15, no. 11 (2018).

[9] Stevens, Robert D. "Machine teaching post-cardiac arrest neuroprognosis electroencephalographic decoding." *Critical care medicine* 47, no. 10 (2019): 1474-1476.

[10] Matam, B. R., Heather Duncan, and David Lowe. "Machine learning system for the detection of heart arrest in an intensive paediatric health centre." *Journal of clinical monitoring and computing* 33, no. 4 (2019): 713-724.

[11] Silvana, M., Akbar, R. and Gravina, H., 2020, October. Optimization of Lung Recognition Characteristics by Artificial Neural Network. In 2020 *International Conference on Information Technology Systems and Innovation (ICTSI)* (pp. 236-241). IEEE.

[12] Miller, A., 2018, June. ICHOM Lung Cancer Data Compilation in the OIS Mandatory data storage. In 2018 IEEE 31st *International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 436-437).

[13] Sakamoto, M., Sora, S. and Kuroki, F., 2020, September. Flat forms of electrodes using resonant circuit technologies to predict lung cancer location. In 2020 IEEE International Symposium on Radio-Frequency Integration Technology (RFTI) (pp. 61-63). IEEE.

[14] Hoque, A., Farabi, A.A., Ahmed, F. and Islam, M.Z., 2020, June. CT scan images for automatic diagnosis of lung cancer. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1030-1033). IEEE.

[15] Xingliang, M. and Fangfeng, L., 2019, July. Association Factors analysis that causes Apriori algorithm dependent lung cancer. In 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 292-295). IEEE.