CARDIOVASCULAR DISEASE PREDICTION USING MACHINE LEARNING: A BRIEF DISCUSSION ON ITS METHODOLOGIES

R.Saravanan[1], S.Pushparaj[2], V.K.Thulasiraja[3], N.Sathish Kumar[4]

[1]Assosiate Professor, Dept. of IT, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry.

[2,3,4]Student, Dept. of IT, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry

---***---

Abstract - Human heart is the major organ in human body which is used to siphon the blood and supply it to all organs. Cardiovascular diseases occur due to failure in action of heart functions, its vessels, muscles, valves etc. According to current scenario estimated by WHO, major deaths in India is due to cardiovascular disease, approximately one person dies per minute. In clinical field predicting of occurrences of heart diseases is tedious work which may help many of the people to save their life as quick as possible. Basically machine learning techniques are used to predict the heart diseases, by using various techniques including supervised or unsupervised techniques. So in this paper we discussed on the various methodologies used to predict heart diseases which gives us overall study of the existing work.

Key Words: *Machine learning, heart diseases, human heart*,

1. INTRODUCTION

Cardiovascular diseases(CVD) act as main source of death around the world. CVDs are the number one reason for death universally, taking an expected 17.9 million lives every year dependent on the estimation of WHO in 2019. CVDs are a gathering of scatters of the heart and veins and incorporate coronary illness, cerebrovascular malady, rheumatic coronary illness and different conditions. Four out of five CVD patients were passing because of coronary episodes and strokes, and 33% of these passing happen rashly in individuals under 70 years old. People in danger of CVD may show raised pulse, glucose, and lipids just as overweight and weight. [1]

The major challenge that the Healthcare industry faces now-a-days is superiority of facility in-order to predict the heart diseases. Diagnosing the disease correctly & providing effective treatment to patients will define the quality of service. Poor diagnosis causes disastrous consequences that are not accepted. [2] Regardless cardiovascular diseases founded as the significant wellspring of death in world in antiquated years, these have been declared as the most dangerous disease. The entire and precise administration of an infection lay on the all-around coordinated judgment of that sickness. A right and systematic approach should be made for perceiving high-chance and what's more, for the most part unique individual body can show various manifestations of coronary illness which may fluctuate as needs be. [3] However, they habitually incorporate back torment, jaw torment, neck torment, stomach issue and improper breath, chest torment, arms and shoulders torments. There are a wide range of heart diseases which incorporates cardiovascular breakdown and stroke and coronary conduit diseases.

Fig.1 illustrates the types of diseases and causes occurred in heart which include stroke, mitral regurgitation, high BP, cardiac arrest, arrhythmia, congestive heart diseases, pulmonary stenosis and so on. [4] Heart expert's create a good and huge record of patient's database and store them. It also delivers a great prospect for mining a valued knowledge from such sort of datasets. There is huge research going on to determine heart disease risk factors in different patients, different researchers are using various statistical approaches and numerous programs of data mining approaches. Statistical analysis acknowledged the count of risk factors for heart diseases counting smoking, age, blood pressure, diabetes, total cholesterol, and hypertension, heart disease training in family, obesity and lack of exercise. For prevention and healthcare of patients who are about to have addicted of heart disease it is very important to have awareness of heart diseases.
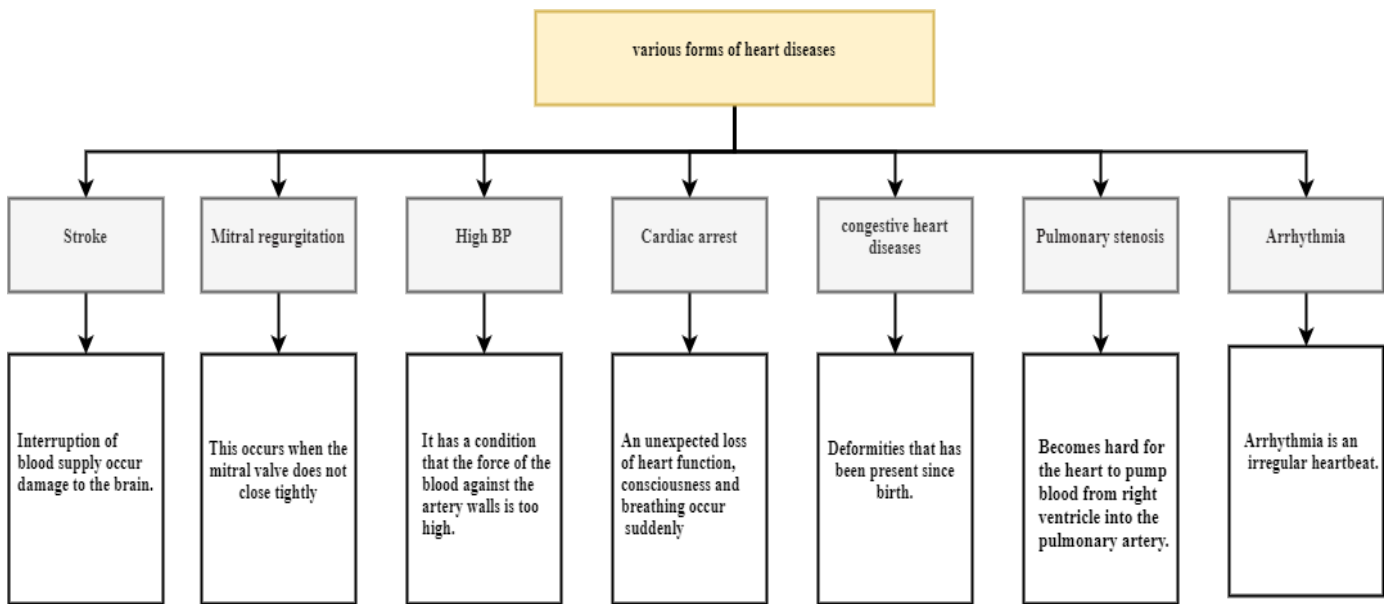
Fig.1 Various Forms of Heart Diseases and Its Causes

A stroke happens when the blood supply to some portion of your brain is hindered or diminished, forestalling cerebrum tissue from getting oxygen and supplements. Synapses start to fail the dust in minutes. A stroke is a health related crisis, and brief treatment is vital. Early activity can diminish cerebrum harm and different difficulties. Mitral regurgitation caused due to the reversal of blood flow from left ventricle (LV) to the left atrium. While cardiovascular diseases occur due to special coronary heart or blood vessel problems which may directly occurs your heart or blood vessels by way of atherosclerosis that is a buildup of fatty plaques in your arteries. Plaque buildup thickens and stiffens artery partitions, that may stop the flow of blood float thru your arteries for your organs and tissues. [5] Atherosclerosis is also a big reason of cardiovascular disease. It may be corrected by following, including an unhealthy diet, lack of exercising, being obese, less tension and smoking. Be that as it may, in a heart it's sick or disfigured, the coronary heart's electric driving forces may not pleasantly start or travel through the heart, making arrhythmias considerably more liable to expand Congenital heart regularly widen simultaneously as a little child is inside the belly inside its mom. Heart deformities can widen in light of the fact that the heart creates, about a month after idea, changing over the float of blood in the coronary heart. Some clinical circumstances, meds and qualities may play a situation in causing heart abandons. Heart deserts likewise can increment in grown-ups. As you age, your heart's shape can change, incurring a coronary heart issue.

Some risk factors for heart disease:

- Hypertension.
- Diabetes.
- High cholesterol.
- Smoking.
- Obesity.
- Physical inactivity.
- High BP.

This paper is organized as follows. Section II explains about the various machine learning algorithms used to predict the heart diseases. Section III comparison between each technique. Section IV summarises about the paper.

II. VARIOUS MACHINE LEARNING TECHNIQUES HELPS TO PREDICT HEART DISEASES

Basically machine learning techniques are classified into:

1. Supervised learning – The algorithm can learn from the dataset through a training process and then it can respond to any new input based on what it has learned.

2. Unsupervised learning – The dataset does not contain the responses in this technique. So, the algorithm tries to recognize the similarities between input values and categorizes them based on their similarities

3. Reinforcement learning – The technique is in the middle of supervised and unsupervised learning, where the model improves its performance as it

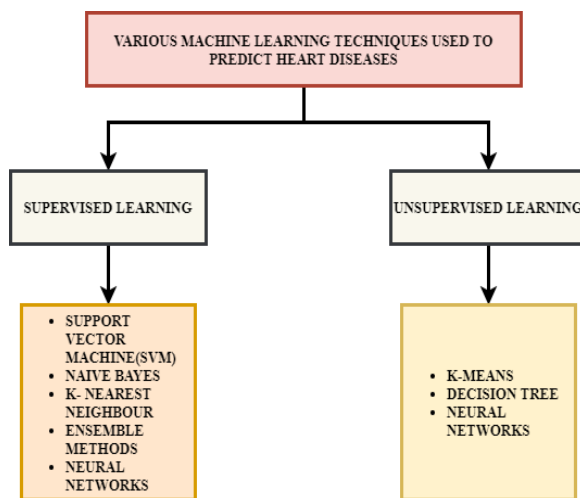interacts with the environment. Hence, learn how to correct its mistakes

Fig.2 illustrates the various machine learning techniques to predict the heart diseases.

2.1 Support Vector Machine (SVM):

A Support Vector Machine (SVM) is a discriminative classifier officially characterized by an isolating hyperplane. At the end of the day, given named preparing information (supervised learning), the calculation yields an ideal hyperplane which orders new models. In two dimensional space this hyperplane is a line separating a plane in two sections where in each class lay in either side. [6]. Based on the study, in order to predict the heart diseases various SVM types such as Binary Tree Support Vector Machine(BTSVM), One-Against-One (OAO), One-Against-All (OAA), Decision Direct Acyclic Graph (DDAG) and Exhaustive Output Error Correction Code (ECOC) are used. The information was first pre-handled by utilizing a min-max scaler. The following stage was preparing the calculation on the dataset which was finished utilizing the SVM calculations referenced previously. On account of execution assessment, SVM far performed better than some other calculations with 61.86% generally speaking precision rate.

Based on the proposed work, the use of non-linear classification set of rules has been made for the heart ailment prediction. [7] It is proposed to use big data tools including Hadoop Distributed File System (HDFS), Map reduce together with SVM for prediction of heart disease with optimized characteristic data set. This work of art made an examination on the utilization of different data digging methods for anticipating heart diseases. It proposes to utilize HDFS for putting away enormous realities in uncommon hubs and executing the forecast calculation utilizing SVM in extra than one hub all the while the use of

SVM. The SVM algorithm is used in equal design which yielded higher calculation time than consecutive SVM. "Support Vector Machine" (SVM) is a supervised device mastering set of rules which may be used for both type or regression challenges. However, it's miles often used in classification troubles. In the SVM set of rules, we plot every information item as a point in n-dimensional area (where n is wide variety of functions you've got) with the fee of each function being the cost of a specific coordinate. Then, we perform category by way of finding the hyper-aircraft that differentiates the 2 classes thoroughly. The intention of the SVM algorithm is to create the nice line or decision boundary that may segregate n-dimensional area into lessons in order that we are able to effortlessly placed the new facts factor in an appropriate category in the future. This best selection boundary is known as a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These severe instances are known as as support vectors, and consequently set of rules is called as Support Vector Machine. Consider the underneath diagram in which there are two one-of-a-kind classes which can be categorised the usage of a choice boundary or hyperplane:

SVM may be of two kinds:

Linear SVM: Linear SVM is used for linearly separable facts, because of this if a dataset may be classified into classes by the usage of a unmarried instantly line, then such records is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, this means that if a dataset cannot be categorized with the aid of the use of a straight line, then such statistics is termed as non-linear information and classifier used is called as Non-linear SVM classifier.

2.2 Naive Bayes:

The Naïve Bayes technique is based on Bayes theorem which is a basic strategy for developing classifiers: models that dole out class marks to issue occasions, spoke to as vectors of highlight esteems, where the class names are drawn from some limited set. There is no algorithm for solitary calculation for preparing such classifiers, yet a group of calculations dependent on a typical standard: all Naive Bayes classifiers accept that the estimation of a specific value is autonomous of the estimation of some other value, given the class variable. [8] Based on the study Naive Bayes has achieved an accuracy of 84.15% there by predicting the heart disease as quick as possible. It is a class technique based on Bayes' Theorem with an assumption of independence among predictors. In easy terms, a Naive

Bayes classifier assumes that the presence of a particular characteristic in a class is unrelated to the presence of any other feature.

For instance, a fruit can be considered to be an apple if it's miles red, round, and approximately 3 inches in diameter. Even if those features depend on each different or upon the lifestyles of the opposite capabilities, all of those houses independently contribute to the possibility that this fruit is an apple and this is why it's miles known as 'Naive'.

Naive Bayes model is straightforward to construct and specifically useful for extremely big statistics units. Along with simplicity, Naive Bayes is understood to outperform even enormously state-of-the-art type techniques.

2.3 K – Nearest Neighbour:

K- Nearest Neighbor (KNN)) depends on named input information to gain proficiency with a capacity that creates a fitting yield when given new unlabeled information. [9] It makes no presumptions about the information and is by and large be utilized for order assignments when there is exceptionally less or no earlier information about the information appropriation. This calculation includes finding the k-nearest information focuses in the preparation set to the information point for which an objective worth is inaccessible and doling out the normal estimation of the discovered information focuses to it. In view of the investigation [10] it provides us the accuracy rate of about 87.5%.The version illustration for KNN is the complete training dataset.

KNN has no version other than storing the whole dataset, so there is no studying required.

Efficient implementations can save the data the usage of complex facts systems like okay-d timber to make look-up and matching of new styles throughout prediction green.

Because the whole schooling dataset is stored, you may need to assume cautiously about the consistency of your schooling records. It might be a good concept to curate it, replace it often as new records turns into available and dispose of faulty and outlier records.

2.4 Ensemble Classifier:

Ensemble methods is a type of machine learning technique that that consolidates a few base models so as to deliver one ideal prescient model. This method is distinguished into types such as BAGG*ing* and **Random Forest** Models. In BAGGing [11] some sample measure of information, multiple bootstrapped subsamples are pulled. A Decision Tree is framed on each of the bootstrapped subsamples. After each subsample Decision Tree has been shaped, a calculation is utilized to total over the Choice Trees to frame the most productive indicator. In Random Forest Models [12] choose where to part dependent on an irregular determination of highlights. Instead of parting at comparable highlights at every hub all through, Random Forest models actualize a degree of separation in light of the fact that each tree will part dependent on various highlights. Ensemble modeling is a powerful way to enhance the overall performance of your version. It commonly can pay off to use ensemble getting to know over and above various fashions you might be constructing. Time and again, people have used ensemble models in competitions like Kaggle and benefited from it.

Ensemble learning is a wide topic and is only limited with the aid of your own imagination. For the cause of this newsletter, I will cowl the basic ideas and thoughts of ensemble modeling. This have to be enough as a way to begin constructing ensembles at your personal cease. As normal, we've tried to preserve things as easy as viable.

2.5 Neural Networks:

Machine learning techniques that make utilization of neural network systems for the most part don't should be customized with explicit principles that characterize what's in store from the information. The neural net taking in calculation rather gains from preparing many marked models that are provided during preparing and utilizing this answer key to realize what qualities of the information are expected to build the right yield. [13] Once an adequate number of models have been prepared, the neural system can start to process new, concealed sources of info and effectively return exact outcomes. In light of the examination the neural network technique system gives a 100% exactness by utilizing the human instinct like sex, weight, cholesterol, etc.

[14] The proposed Coronary diseases prediction make the utilization of ANN calculation in records mining. Because of developing expenses of coronary heart affliction visualization ailment, there has been a need to grow new framework which can foresee coronary heart ailment. Forecast adaptation is utilized to are expecting the circumstance of the patient after evaluation based on different parameters like heart beat expense, blood pressure, cholesterol, etc. The precision of the device is demonstrated in java. Learning of the neural community

takes region on the idea of a sample of the population underneath have a look at. During the course of mastering, evaluate the fee added by using the output unit with actual cost. After that alter the weights of all devices so to improve the prediction.

There are many Neural Network Algorithms are to be had for training Artificial Neural Network. Let us now see a few essential Algorithms for schooling Neural Networks:

Gradient Descent — Used to find the neighborhood minimum of a feature.

Evolutionary Algorithms — Based on the concept of natural choice or survival of the fittest in Biology.

Genetic Algorithm — Enable the maximum appropriate rules for the answer of a trouble and pick it. So, they send their 'genetic material' to 'infant' guidelines.

2.6 Decision Trees:

Decision tree strategies performs easily with ceaseless and straight out traits. This calculation separates the populace into at least two comparable sets dependent on the most critical indicators. The decision Tree calculation, first ascertains the entropy of every single characteristic. At that point the dataset is part with the assistance of the factors or indicators with most extreme data addition or least entropy. In view of the exactness it gives 75% approx. [15] which is expressed to be terrible showing estimated. Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving **regression and classification problems** too.The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by **learning simple decision rules** inferred from prior data(training data). In Decision Trees, for predicting a class label for a record we start from the **root** of the tree. We compare the values of the root attribute with the record's attribute. Types of selection timber are primarily based at the form of goal variable we've got. It may be of two sorts:

Categorical Variable Decision Tree: Decision Tree which has a express goal variable then it called a Categorical variable decision tree.

Continuous Variable Decision Tree: Decision Tree has a continuous target variable then it's far referred to as Continuous Variable Decision Tree

III. COMPARISON OF MACHINE LEARNING TECHNIQUES BASED ON HEART DISEASES PREDICTION:

This section provides a tabular comparison between the methodologies listed above based on the research works made by the different authors. The comparison is made on the basis of accuracy and can be seen in table 2.

The table has 3 elements which are as follow:

1) Author: This shows the author/s of the paper and the reference number.
2) Technique used: This represents the classification algorithm used in the research; whether it was a single algorithm, a comparison or a hybrid model
3) Accuracy level: This represents the accuracy of the results of the proposed model

Table 1: Comparison Based On the Prediction of Cardiovascular Diseases Using Various Machine Learning Techniques

S.no	Author	Purpose	Techniques used	Accuracy	Advantage	Disadvantage
1	R.Sharmila[16]	To enhance the prediction techniques for cardio diseases	SVM in parallel fashion	85%	SVM is effective in cases where number of dimensions is greater than the number of samples.	SVM does not perform very well, when the data set has more noise. In cases where number of features for each data point exceeds the number of training data sample , the SVM will under perform.
2	Noura Ajam [17]	Heart Disease Diagnoses using Artificial Neural Network	ANN	88%	It works better even if the data is incomplete.	Little bit complex in determining proper network. Duration of prediction is unknown it depends on data input.
3	Sairabi H. Mujawar [18]	Prediction of Heart Disease using Modified K-means and by using Naive Bayes	Modified k-means algorithm integrating with naive bayes algorithm	93%	With less amount of data it can easily estimate the result and thus it performs in less amount of time.	It uses the additional technique called Laplace estimation in order to smoothen the zero frequency
4	Kanika Pahwa and Ravinder Kumar[19]	Prediction of cardio Disease using hybrid techniques Hybrid Technique Disease Using For Selecting Features	Naive Bayes	84.15%	It works well in small scale data's, can handle multi-category tasks and explains the result easily.	It is very sensitive and throws more error rate in classification of data.
5	Ahmad Shahin, Walid Moudani, Fadi Chakik, Mohamad Khalil [12] al.	Data mining in health care industry	Random Forest	97.7%	It is well suitable to solve classification and regression technique. It is less impact to noise.	It is complex and takes longer amount of time processing.

IV. CONCLUSION

This paper overviews the literature of machine learning classification strategies for diagnosis heart condition. several mimetic papers on victimisation machine learning classification techniques were surveyed and categorised. The accuracy of the planned models vary looking on the tool used, the dataset used, the quantity of attributes and records within the dataset, the pre-processing techniques, furthermore because the classifier implemented within the model. We conclude that to build associate correct heart disease prediction model, a dataset with sufficient samples

and proper data should be used. The dataset should be pre-processed consequently as a result of it's the foremost important half to organize the dataset to be employed by the ma-chine learning algorithmic program and get sensible results. Also, a appropriate algorithmic program should be used once developing a prediction model

V. REFERENCES:

[1] Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, pp.2137-2159, July 2017.

[2] V.Krishnaiah, G.Narsimha, N.Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.

[3] Guizhou Hu, Martin M. Root, "Building Prediction Models for Coronary Heart Disease by Synthesizing Multiple Longitudinal Research Findings", European Science of Cardiology, 10 May 2005.

[4] Deeanna Kelley, "Heart Disease: Causes, Prevention, and Current Research", *JCCC,* Vol. 5 : Iss. 2 , Article 1, 2014.

[5] I. Kononenko, "Machine learning for medical diagnosis: History", Artificial intelligence in medicine vol. 23, no. 1, pp. 89–109, 2001.

[6] Madhura Patil, Rima Jadhav, Vishakha Patil, Aditi Bhawar, Geeta Chillarge, "Prediction and Analysis of Heart Disease Using SVM Algorithm", ISSN: 2321-9653, Jan 2019.

[7] W. Wiharto, H. Kusnanto, and H. Herianto, "Performance analysis of multiclass support vector machine classification for diagnosis of science and applications", vol. 5, no. 5, pp. 27–37, Jan 2015.

[8] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.

[9] M. Shouman, T. Turner, and R. Stocker, "Applying KNN in diagnosing heart patients"   International Jour-
neighbour in diagnosing heart disease patients."
journal of Information and Education Technology, vol. 2, no. 3, pp. 220, 2012.

[10] Jabbar MA," Prediction of heart disease using k-nearest neighbor and particle swarm optimization" Research Article - Biomedical Research Volume 28, Issue 9,Feb 2017

[11] J. Patel et al,"Heart disease prediction using machine learning and data mining technique," Heart Disease, vol. 7, no. 1, pp. 129–137, 2015.

[12]. Ahmad Shahin, Walid Moudani, Fadi Chakik, Mohamad Khalil et al," Data Mining in health care information system", ISBN: 978-1-4799-3166-8 ©2014 IEEE.

[13]. R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart through neural networks ressembles", Expert system with applications , vol. 36, no. 4, pp. 7675–7680, 2009.

https://doi.org/10.1016/j.eswa.2008.09.013 systems with. disease through neural networks ensembles," Expert

[14]. C. Dangare and S. Apte, "A data mining approach for prediction of heart disease using neural networks", International journal of Computer Engineering

& Technology, vol. 3, no. 3, pp. 30–40, 2012.

[15]. Amir Hussain, Peipei Yang, Mufti Mahmud and Jan Karasek et al., " A novel Cardiovascular Decision support framework for effective clinical risk assessment", 978-1-4799-4527- ©2014 IEEE.

[16]. R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data's ",International journal of Computer science and Engineering, May 2018.

[17]. Noura Ajam, "Heart Disease Diagnoses using Artificial Neural Networks", International institute of
Neural Network", The International Insitute of Science,
Technology and Education, vol.5, No.4, pp.7-11, 2015.

[18]. Sairabi H. Mujawar1 , P. R. Devale, "Prediction of Heart Disease using Modified K-means and by using Naive Bayes", Vol. 3, Issue 10, October 2015.

[19]. Kanika Pahwa and Ravinder Kumar et al. "Prediction of Heart Using Hybrid Technique For Selecting Features", IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON). ,2014.

[20]. M.A.JABBAR , B.L Deekshatulu and Priti Chndra et al. , "Alternating decision trees for early diagnosis of heart disease", Proceedings of International Conference on Circuits, Communication, **Control and** Computing (I4C 2014)