Arrhythmia – A Case Study on early disease diagnostics

An Entrepreneur Idea on Application of Machine Learning over IOT Platforms for early diagnosis of Critical Medical Diseases

JANUARY 1, 2018

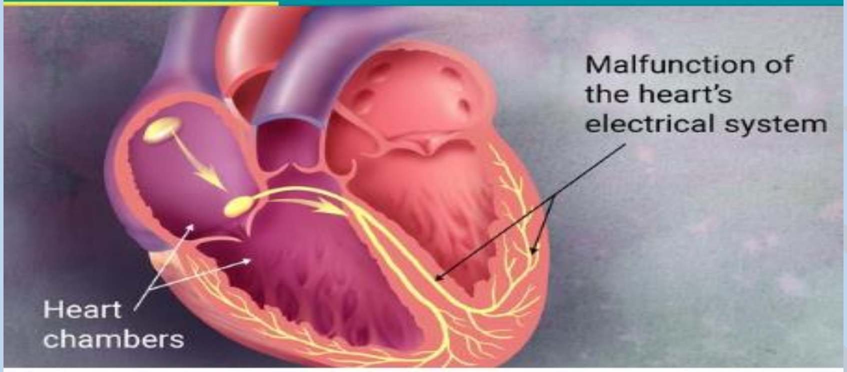
BY - SHREY MISHRA

New Delhi, India - 110017, E-mail: shrey.mishra@rocketmail.com,

What is Arrhythmia?

Arrhythmia
Also called: irregular heartbeat

ABOUT | SYMPTOMS | TREATMENTS



Improper beating of the heart, whether irregular, too fast or too slow.

Very common

More than **10 million** cases per year (India)

- Requires a medical diagnosis
- Lab tests or imaging often required
- Treatable by a medical professional

Cardiac arrhythmia occurs when electrical impulses in the heart don't work properly.

There may be no symptoms. Alternatively, symptoms may include a fluttering in the chest, chest pain, fainting or dizziness.

If required, treatment includes anti-arrhythmic drugs, medical procedure implantable devices and surgery.

DETECTION OF ARRHYTHMIA

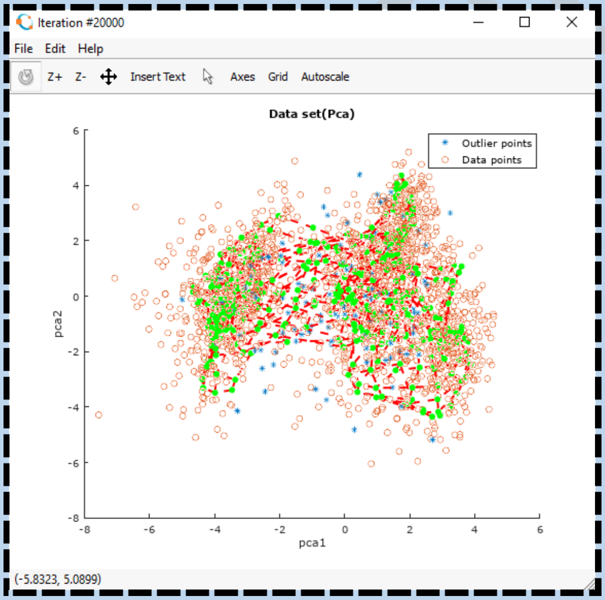
An arrhythmia is caused by a disruption of your heart's normal electrical system, which regulates your heart rate and heart rhythm. The severity of cardiac arrhythmias can vary tremendously. Most arrhythmias are completely benign and inconsequential, while others are extremely dangerous and life-threatening. And many of them, while not particularly dangerous, produce symptoms that can be quite disruptive to your life.

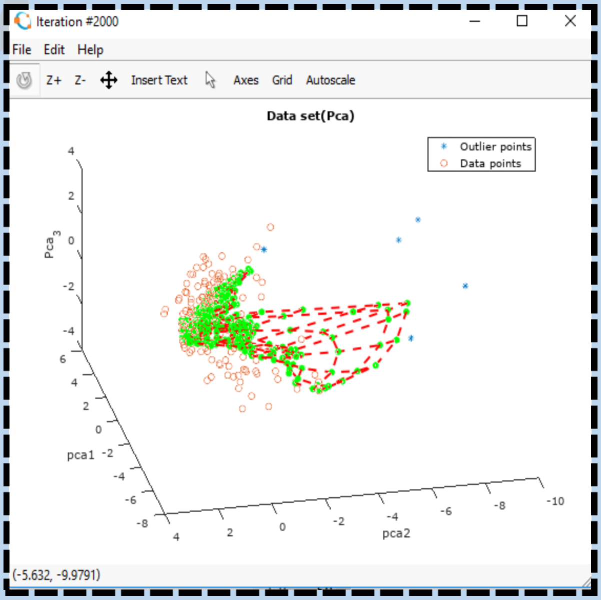
ENTREPRENEUR IDEA

Application Context - Early detection of potentially dangerous diseases such as Breast Cancer, Osteoarthritis, Diabetic Retinopathy, Arrhythmia (malfunction in the heart) and many other typical diseases using Machine Learning algorithms over IOT platform is the core concept of this idea. As a case study, the idea has been applied here on Arrhythmia, one of the most critical disease where generally medical science faces a lot of challenges due to lack of sufficient visible symptoms.

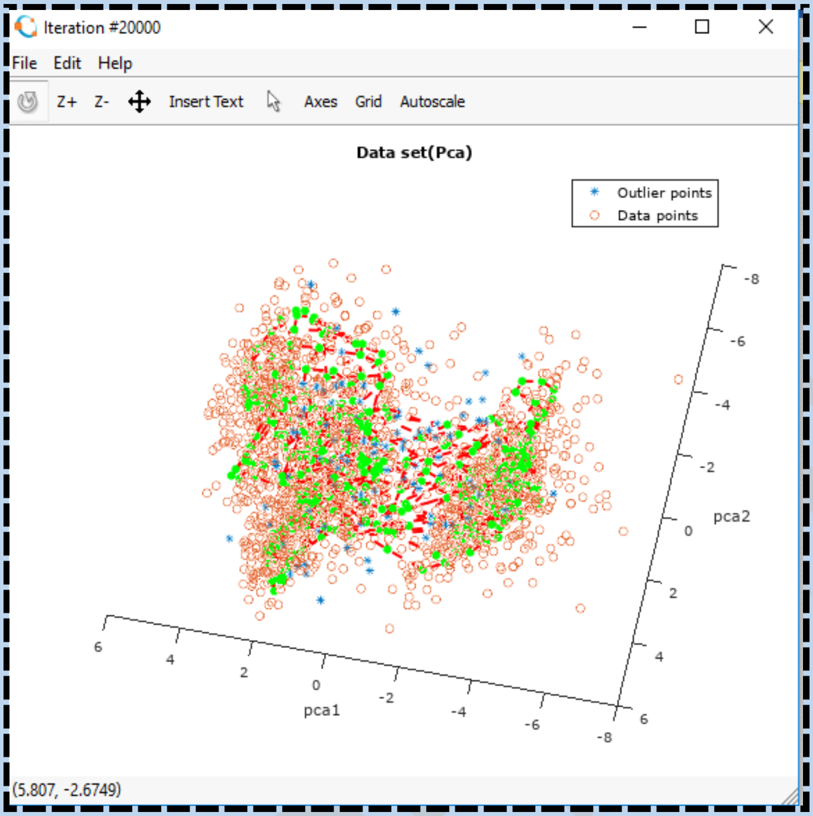
In order to prove the concept, at first visualization of the data shall be carried out and an observable pertaining pattern will be drawn using the Self Organizing Maps (A Machine Learning algorithm that can help plot the visualization pattern from large chunk of data in a form of a map). For better visualization, the data is compressed from multiple dimensions to reduced PCA (Principal component analysis) dimensions.

The results of SOM trained map, identifying various classes are drawn below –





The original datasets used for training can be downloaded from here dataset Or alternatively from the web page - http://odds.cs.stonybrook.edu/



SOM algorithm's working can be studied in more details from the documentation prepared and tested by me through following link –

[https://github.com/mv96/SOM-maps]

My working code link for the above documentation -

[https://github.com/mv96/SOM-maps]

As we can clearly see in the graphs of the datasets, trained on SOM algorithm, takes the shape of the data after a set number of iterations and hence the data possesses a certain pattern which is clearly observable.

In order to further identify this pattern with exact metric some standard Supervised Machine learning approaches have been used, such as -

1) Regression Based Models
2) Artificial Neural Networks
3) Support Vector Machines with various Kernels

The detail below is a result section of testing the code and efficient working of these algorithms on 18 different datasets –

Datasets tested	Rows(m)	Column (n)	SVM (Without Gaussian)	SVM (With Gaussian)	ANN	Logistic Regression	Anomaly %
Vowels	1456	12	100	100	99.77	98.351	3.4
Lympho	148	18	100	97.77	97.78	97.95	4.1
Letter	1600	32	93.94	93.94	94.79	91.938	6.25
Glass	214	9	98.46	98.46	98.46	94.419	9
Vertebral	240	6	84.72	86.11	88.11	90	12.5
Mammography	11183	6	98.36	98.59	98.68	98.42	2.32
Smtp	95156	3	99.98	99.98	99.96	99.98	0.03
Annthyroid	7200	6	94.4	94.44	92.12	94.7	7.42
Thyroid	3772	6	99.38	99.02	97.73	99.04	2.5
Pima	768	8	78.78	63.2	71.86	78.16	35
Wbc	278	30	99.12	99.12	95.61	95.47	5.6
Ionosphere	351	33	88.67	89.62	85.84	83.71	36
Optdigits	5216	64	99.36	98.4	99.87	N/A	3
Satellite	6435	36	N/A	68.41	88.67	87.956	32
Satimage	5803	36	99.54	98.85	96.56	99.58	1.2
Speech	3686	400	97.1	98.28	99.88	96.08	1.65
Breast	683	9	96.09	96.58	96.09	95.18	35
Arrhythmia	452	274	85.29	83.82	91.18	84.66	15

The results are based upon my research paper published in **International Journal of Scientific & Engineering Research Volume 9, Issue 11, November-2018 ISSN 2229-5518** [https://www.ijser.org/onlineResearchPaperViewer.aspx?Outlier-Detection-using-Supervised-Machine-Learning-Algorithms.pdf]

CONCLUSION –

ANN has surpassed the all the other approaches and currently holds the accuracy of 91.18% on Arrhythmia dataset.

(Note- that original data was trained on 70-30 model which only takes 70% of the total data to train, remaining 30% to test and unlike the standard 80-20 approach)

WAY FORWARD –

The state of the art **'Capsule Networks'** proposed by Geoffrey Hinton can predict better results than currently obtained from ANN approach [https://arxiv.org/abs/1710.09829]

ML MODEL DEPLOYMENT -

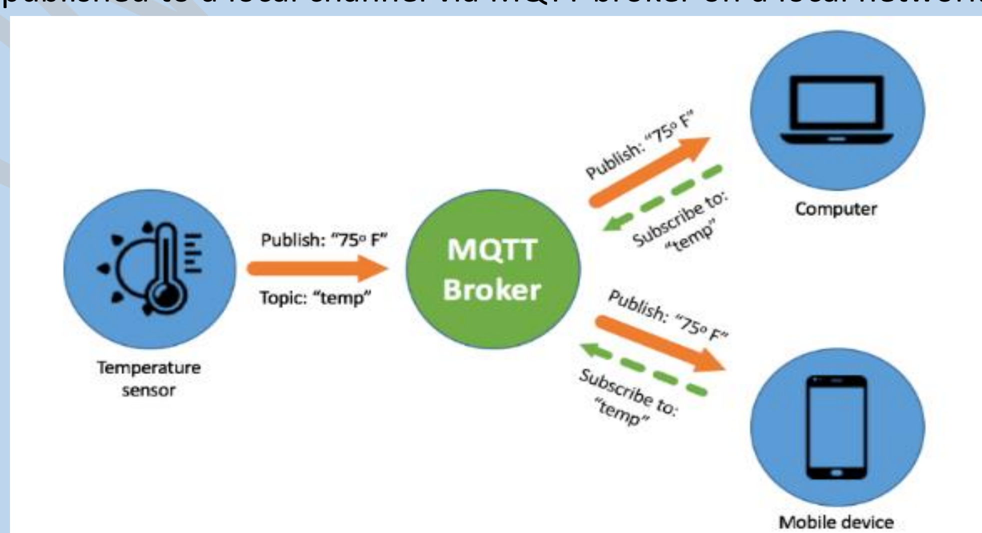
Upon getting satisfactory results from the prediction algorithms, the deployment of the concept in real time using Raspberry Pi with Arduino can be done over a smartphone application.

In order to prove the concept, I have performed a live test and video recording for same is linked below (In this recording the CNN identifies Cats and Dogs using a pre trained model on raspberry Pi) -
https://www.dropbox.com/s/iucnavrtnhsw4t7/2018-7-3-14-6-5.mp4?dl=0

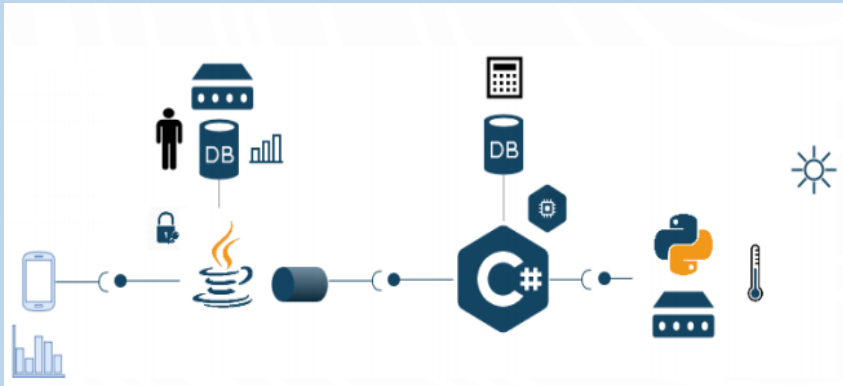
In order to explain architecture, all the data flow diagrams are included below:

Arduino directly interacts with most of the sensors and sends the information on a Serial Monitor which is further sent to Raspberry Pi over USB (UART protocol) following a Master-Slave model.

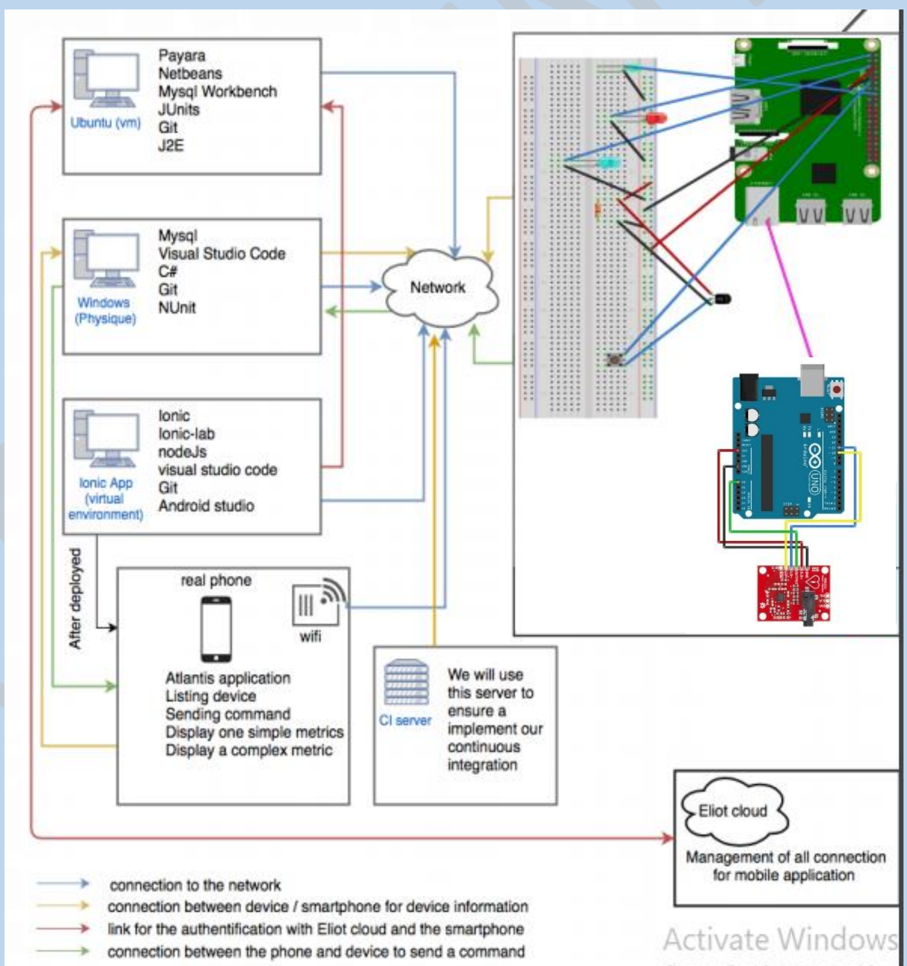
The data is then published to a local channel via MQTT broker on a local network. This Publisher-Subscriber model is supported by a large number of clients on most of the programing languages. The data sent using MQTT is TLS encrypted and also supports data transmission to large number of clients on low bandwidth.



The entire data flow channel from the sensors to smart phone is demonstrated below:

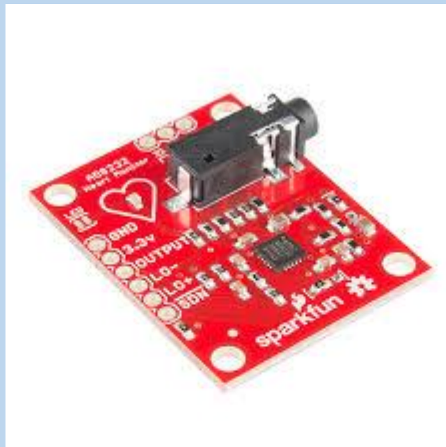


By using MQTT data can be fetched on a C# and Java client which will directly support the native platform of the android smart phone application used. Below diagram explains the working of the mobile development for smart city –



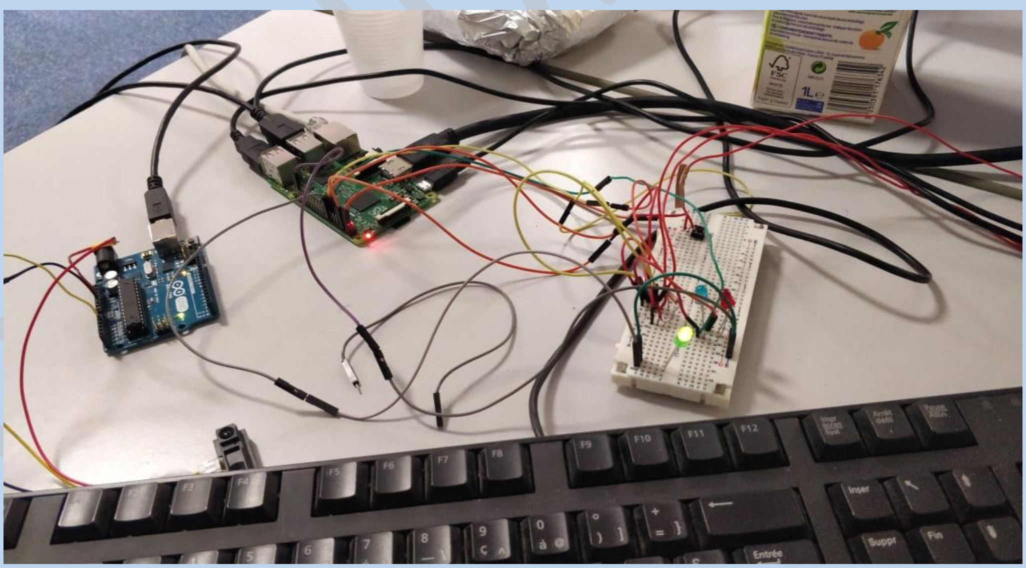
To explain the concept better here is a link to Fritzing shareable project file that I have made-[https://github.com/mv96/project-Atlantis]

For ECG (Electro Cardiogram) pulse detection, ECG Module (AD8232) component is used –



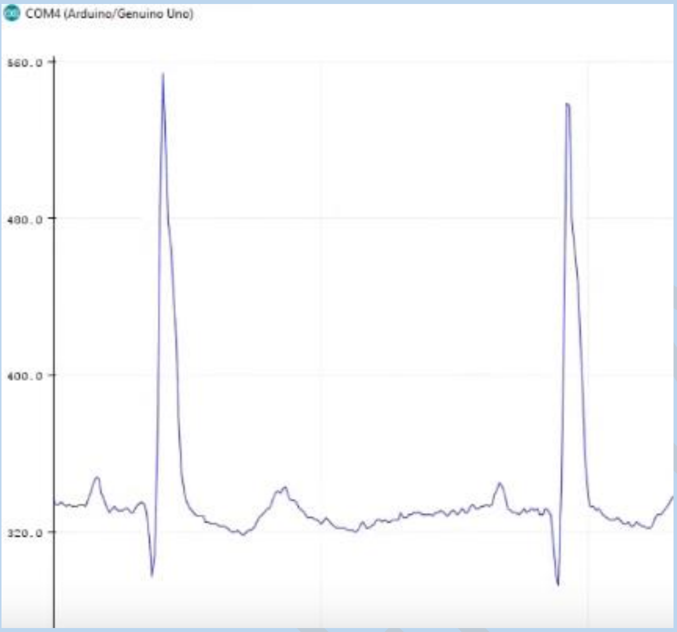
The fully functional code link that is used to extract data from various sensors from Arduino to Raspberry Pi is –

[https://github.com/sparkfun/AD8232_Heart_Rate_Monitor]



The picture shows data feed from different sensor types.

As we can see that the Arduino clearly identifies the ECG pattern which is ready to be tested with the original ML model as discussed in first section.



Arduino Serial Plotter Reading

SYNOPSIS -

According to reports every one in four adult (nearly 2.2 million in just America) can develop Arrhythmia over the age of 40. The current medical science practices of more dependency on human interface limits to the level of individual competency, experience and limited examinable parameters.

The above solution can drastically bridge up the gap between technology and medical science. The solution proposed in the above entrepreneur idea, costing less than 100$ may give a more précised micro identification of many such diseases followed by a timely treatment shall not only help in economic sense but also help in drastically reducing heavy life losses.

------End of report-------