Calibration of a portable PM2.5 measuring device using web-query and machine learning

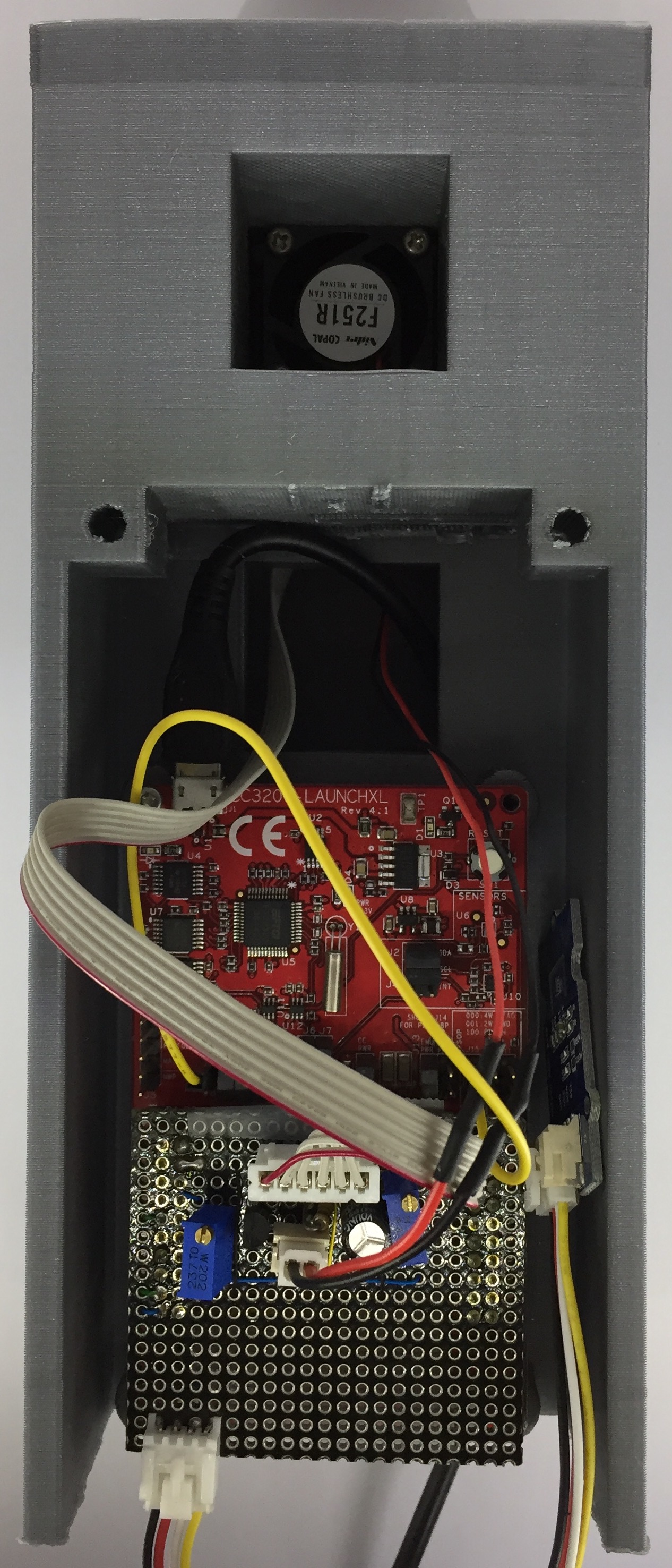
Byoung Gook Loh and Gi-Heung Choi

Abstract

Introduction

Device

|  |  |
| --- | --- |
| Fig.1 Photo of portable PM2.5 sensing device (PSD)  (dimension : 80x80x200 mm) | Fig.2 Functional block diagram of IoT PSD |



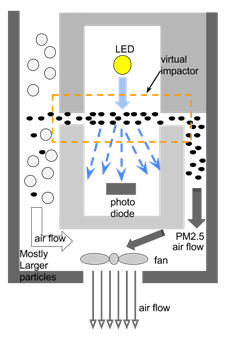


Fig. 4. Working principle of Sharp PM2.5 sensor module

Table 2. Technical specification of Sharp PM2.5 sensor(6) ( **to be removed** )

|  |  |
| --- | --- |
| size | 53 x 40 x 51mm |
| weight | 53 g ± 7g |
| sensing range | 25 ~ 500 μg/m3 |
| Sensitivity | 1 volt/100 μg |
| power consumption | 1.1 watt |
| noise | 38dB |

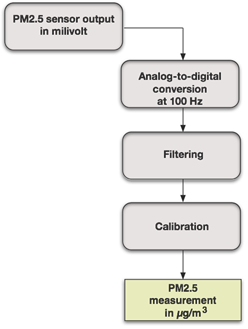
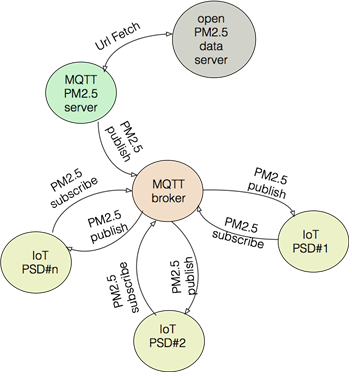


Fig. 5. Flow-chart of PM2.5 sensing algorithm



Experimental Results

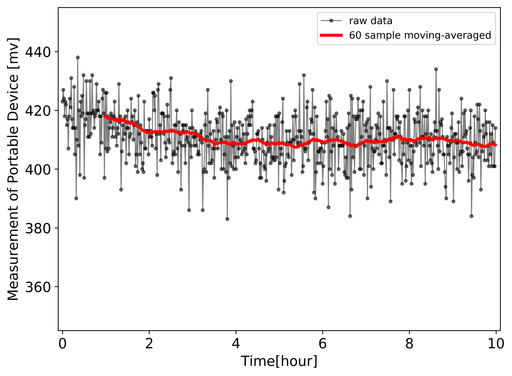


Figure Raw measurement of portable PM2.5 measuring device

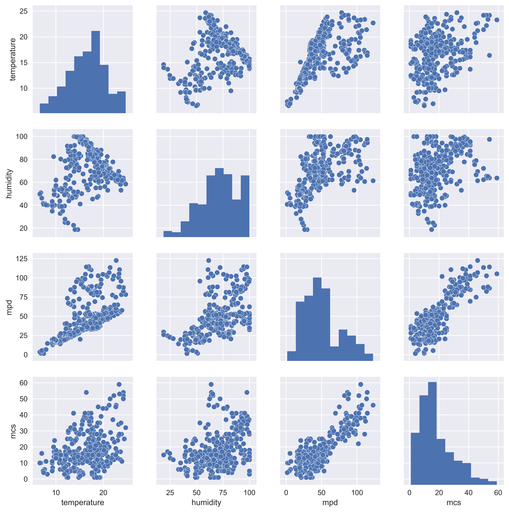
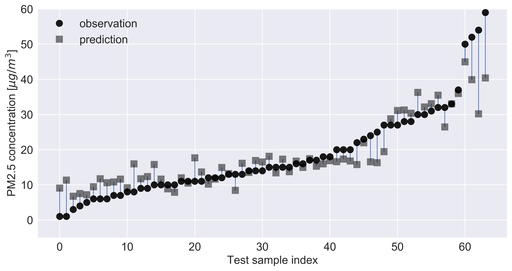


Figure Pairplot of variables studied in regression mgd : mcs:

Linear regression method is used a base case.

|  |  |
| --- | --- |
| Fig. Linear regression using regression samples | Fig. Linear regression using test samples |

Linear regression was performed on test data sets and evaluated against unseen verification samples which was 20 % of the total data set. (**What Is case I and case II in the data sets ??**)





Four widely used machine learning algorithms were investigated. \*\* number of data samples were used for regression. To gauge performance, mean coefficient of determination, R2 and mean of root mean squared error (RMSE) were compared.

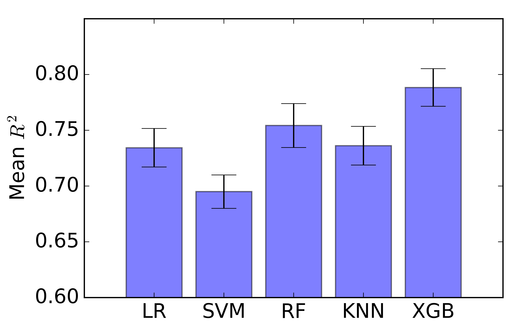


Fig Mean R2 of different machine learning algorithms

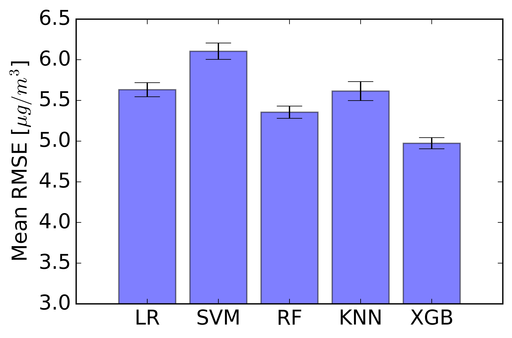


Fig Mean RMSE of different machine learning algorithms

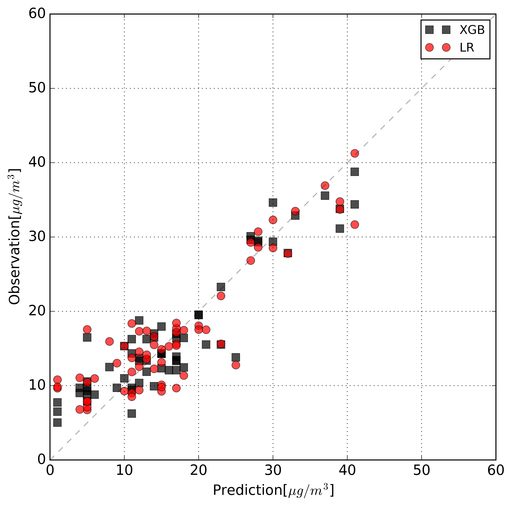


Figure regression result of XGB ( \*\* **remove Linear regression result )**

The effect of the number of experimental data used for regression on R2 is shown in Fig.\*. The value of R2 oscillates between 0.1 and 0.6 as the number of the data included in regression approaches 60 beyond which the R2 progressively converges to a value of 0.8 and shows a miniscule variation as the number of the data increases.

