DS&AI 2023

DATA SCIENCE AND ARTIFICIAL INTELLIGENCE CONFERENCE 2023

1ST - 3RD FEBRUARY 2023

A Gradient Boosting Regression Model for the Prediction of an Individual's Short Term Blood Pressure

Ms Daisy Kiptoo

Dr. Moses Thiga

Dr. Peter Rugiri

Dr. Pamela Kimeto

Sponsored by Artificial Intelligence www.elsevier.com/locate/artint JOURNAL Google TensorFlow

Background

- Hypertension has no warning signs nor symptoms, and many people do not know they have it, measuring the BP level is the only way to know about a person's BP status.
- According to the World Health Organization (WHO), ischaemic heart disease (also referred to as coronary heart disease) and stroke are the world's biggest killers, accounting for a combined 15.2 million deaths in 2016 (Nichols, 2017).
- The two conditions can be derived from hypertension.

Problem

- BP varies throughout the day and can change with the physical and psychological state of an individual, their environment, the observer, and the measuring device.
- If an individual's BP level begins to rise consistently, and action is not taken due to factors such as lack of awareness, the person might develop high BP.
- Due to the pervasiveness of hypertension, individualized prediction of BP is very important as it may significantly reduce the rise of BP by giving precautions to an individual.
- The study helped in predicting an individual's short term BP based on their BP readings and activities. It also helped to pin-point why an individual's BP was raised.

Objectives of the Study

i. To develop a model for regular collection of blood pressure readings and activity data.

ii. To identify and train a machine learning algorithm to learn and predict future blood pressure readings of an individual using their past blood pressure readings and activity data.

iii. To validate the performance of the machine learning algorithm for learning and predicting future levels of an individual's blood pressure using their past readings and activity data.

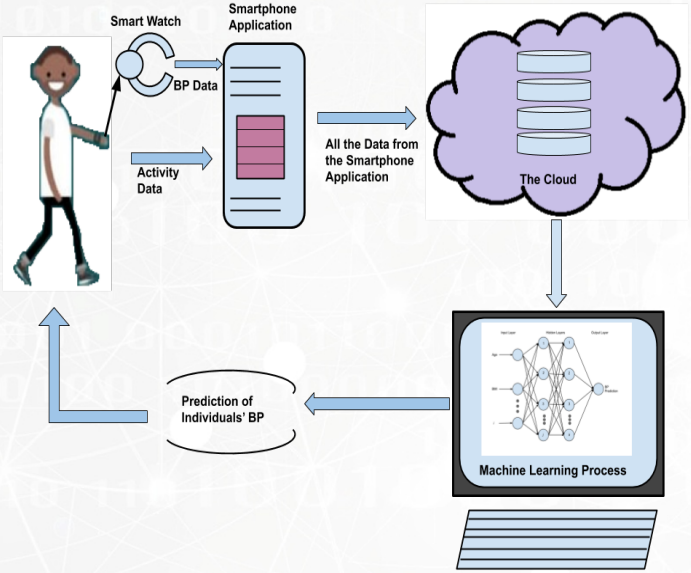
Significance

- Raised BP is a condition that a significant population who have the condition are not aware of; this study helped individuals be more aware of their BP status and therefore has the potential to greatly improve the monitoring and control of BP due to its prediction.
- The study also helped determine which activities contribute to raised BP for the individuals.

Literature Review

- The causes of hypertension are often not known though there are risk factors associated with the increase of BP. Some of these factors are; age, race, weight, alcohol use, tobacco use, gender and existing health conditions such as chronic kidney decease (MacGill, 2018).
- There is a growing number of hypertensive patients, therefore, “improving BP control is thus unquestionably a goal of fundamental importance for cardiovascular prevention worldwide” (Mancia *et al.*, 2017).
- This study made future short term predictions of an individual’s BP using data collected from the individuals inclusive of the mood and activities.

Conceptual Framework



- The smartwatch takes BP readings from the individual, which is then transmitted to the smartphone using Bluetooth technology.
- The individual also provides data such as the activity data, alcohol intake, tobacco intake and future calendar events to the smartphone application.
- The data is then sent to the cloud and then used by the machine learning model to be able to learn and predict the individual's BP.

Research Methodology

Study Design

- The methodology that was used for the study is design science and experimental methods.
- For the first objective, which are the data collection tools, Rapid Application Development (RAD) was used.
- For objective two, the implementation of the ML model was done using Iterative and Incremental Development.

Research Methodology

Sampling Procedure, Location and Population of the Study

- The sampling procedure for this study was quota sampling.
- This research took place in Uasin Gishu County, in all the 6 constituencies.
- The population for the study comprised of people who do not have known hypertension.

The total number of participants for the study therefore was 45

Research Methodology

Instrumentation

- The data for this study was collected through the smartwatches and the smartphone application.
- i. The smartwatch specifications;
- It was able to read a person's BP.
- It was Bluetooth enabled so as to be able to send the BP readings to the smartphone application.
- It had a long lasting battery, for the sake of consistency.
- ii. The specifications for the smartphone;
- It was able to take in a new smartphone application

Research Methodology

Validity and Reliability of the Instrument

- Test-retest reliability was done by measuring the BP of individuals several times within a short period of time to see whether the smartwatch was reliable in its measurements.
- Internal Consistency Reliability was also used as a measure of reliability of the instrument. The test was done by measuring the accuracy of the smartwatches and comparing them to a doctor's office BP monitoring device several times.
- The smartwatch readings were close to the readings of the sphygmomanometer, which shows that it is quite accurate.

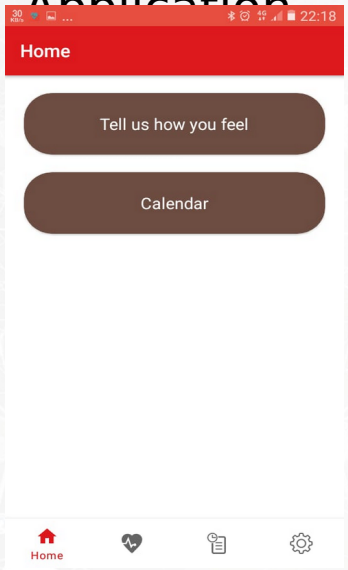
Research Methodology

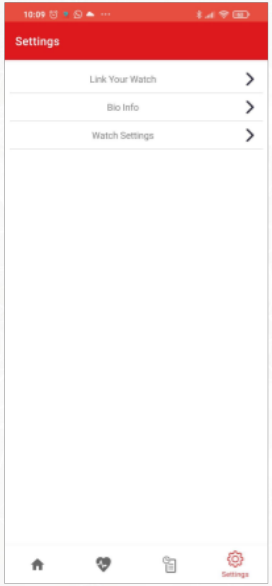
Data Analysis

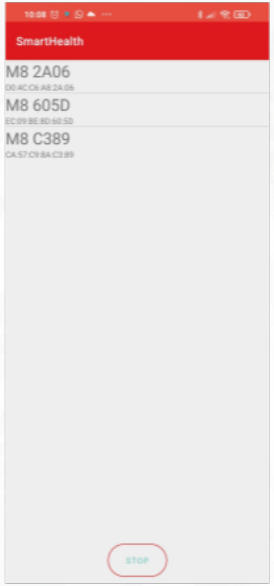
- This process involves identification of null values, inaccurate data, record matching and de-duplication of data.
- Therefore, Pandas, NumPy and Scikit-learn libraries were used for data transformation, data manipulation and data pre-processing; this included checking for missing values, categorical data, standardization of data, data normalization and splitting of data.
- Once the data was analysed, it was then translated into a format that the ML model could understand.

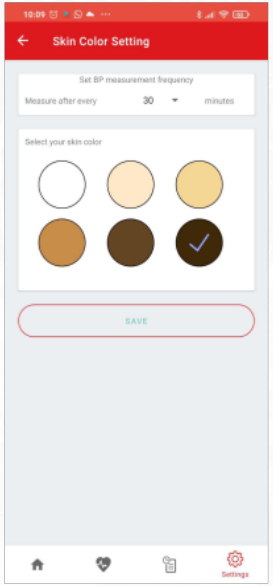
Results

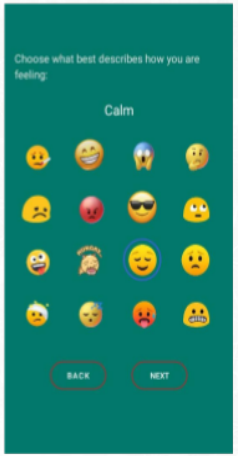
Objective One: The Smarthealth Application

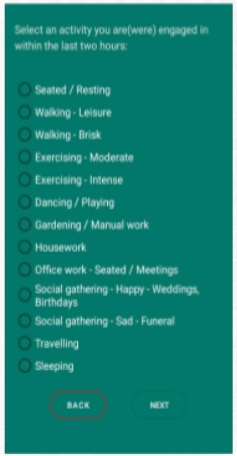












Results

Objective One

Advantages noted by the research participants:

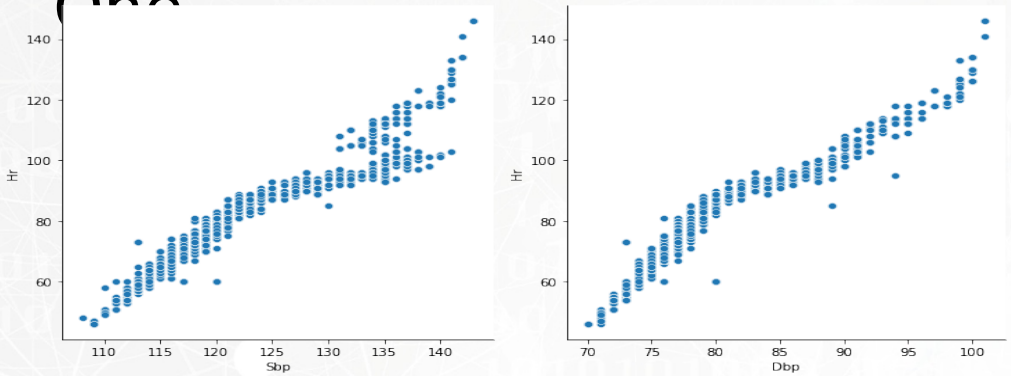
- Participants liked the fact that the watch reduced the need for them to go to the hospital to take their BP readings. They liked the fact that they could take BP anytime, anywhere.
- Some participants said that they didn't know their BP because they rarely got checked, so it proved to be very helpful to them because they now had knowledge of their BP status.
- Some participants said that the application helped them stay alert, live more consciously, they were more cautious of what they let themselves feel, for example.

Concerns raised by the research participants:

- Some participants found it stressful to sleep with the smartwatch.
- The application drained the smartphone battery fast.
- The smartwatch vibrates at times, especially after detecting the individual's temperature is quite high, that tends to surprise the participants.
- Some participants were concerned about the green light, PhotoPlethysmoGraphy (PPG), emitted while taking some readings like BP and HR.

Results

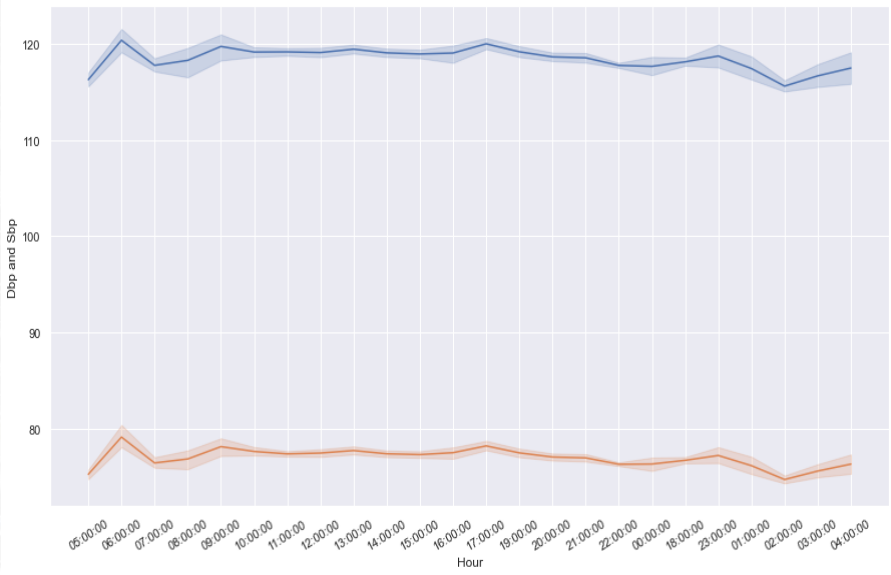
Objective One



- This relational plot shows the relationship between the variables, HR, Systolic blood pressure (SBP) and Diastolic blood pressure (DBP).
- As HR increases, the SBP and the DBP also increases. This shows that there is a high correlation between the HR, the SBP and the DBP.

Results

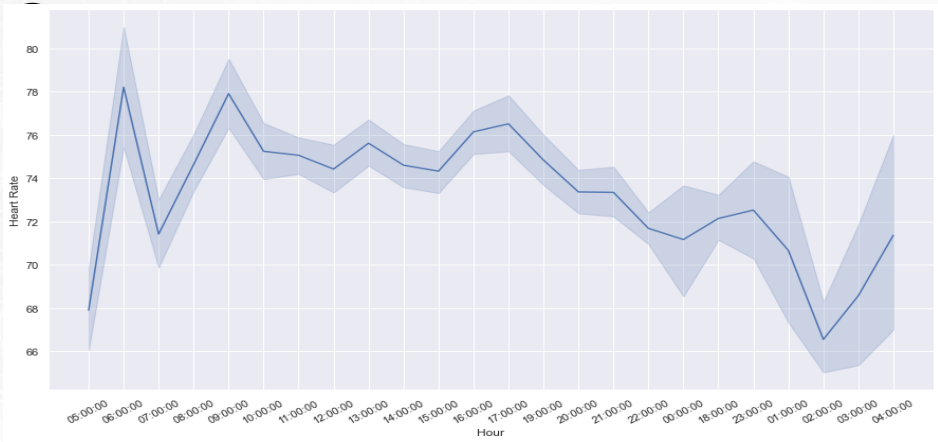
Objective One



- The figure shows that there is a high correlation between the DBP and SBP.
- It also shows that both the DBP and SBP are highest in the morning hours, at around 6:00am, and is lowest at night at around 2:00am, when most individuals are asleep.

Results

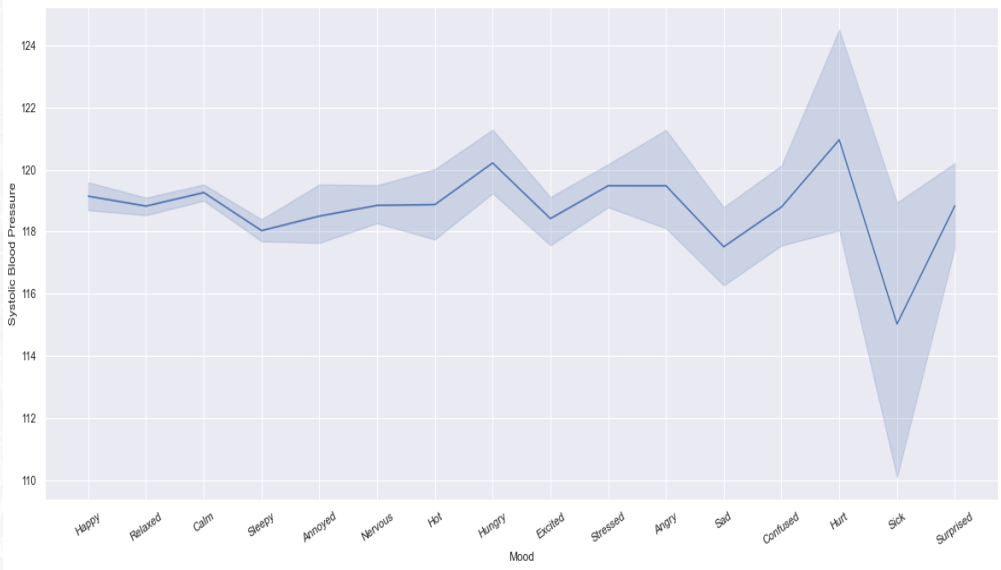
Objective



- The figure shows that HR raises in the morning hours and is low at night, when most individuals are asleep.
- It also reveals that from around 5:00pm, the HR gradually reduces up until around 2:00am, when it is lowest, then it raises.

Results

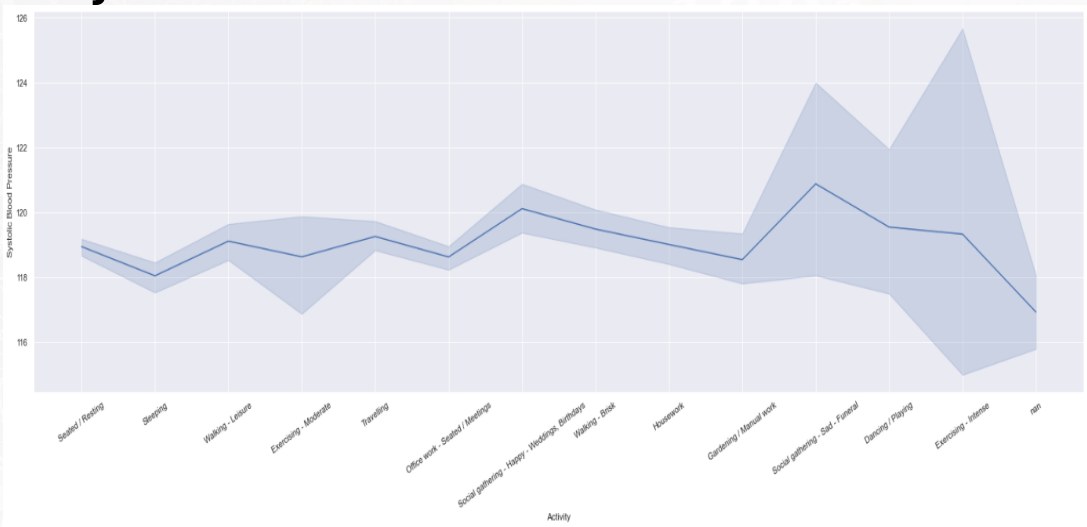
Objective



- This line plot shows that the mood affects the SBP.
- An example would be when an individual is sleepy, the SBP goes low up to about 118, and when an individual is hurt, the SBP spikes up to 124.

Results

Objective



- This line plot shows that the activity affects the SBP.
- An example would be when an individual is sleeping, the SBP goes low up to about 118, and when an individual is in a social gathering that is relatively happy, the SBP spikes up to 121.

Results

Objective Two

- Using machine learning libraries like Scikit-learn, the researcher was able to test several models in order to find one that works well with the data collected. These were the models that were identified, trained and tested:

i. Multiple Linear Regression; the MSE score was 0.975 and R^2 score was 0.966.

ii. Lasso; the MSE score of 1.157 and R^2 score of 0.957.

iii. Elastic Net; the MSE score of 0.927 and R^2 score of 0.961.

iv. K-Nearest Neighbours; the MSE score of 0.773 and R^2 score of 0.971.

v. Decision Tree Regressor; the MSE score of 0.219 and R^2 score of 0.992.

vi. Gradient Boosting Regressor; the MSE score of 0.182 and R^2 score of 0.992.

Results

Objective

Out[34]:

	Prediction	Test Data	Difference
0	118.186517	118	-0.186517
1	114.195993	114	-0.195993
2	115.098294	115	-0.098294
3	119.853309	120	0.146691
4	116.238001	116	-0.238001
5	124.926891	125	0.073109
6	114.347829	115	0.652171
7	110.343997	110	-0.343997
8	137.079042	137	-0.079042
9	119.895421	119	-0.895421

This is the prediction of SBP. The difference between the predicted data and the test data is very small.

Results

Objective Two

Out[20]:

	Prediction	Test Data	Difference
0	76.931912	77	0.068088
1	73.975209	74	0.024791
2	74.489936	75	0.510064
3	77.975605	78	0.024395
4	75.298633	75	-0.298633
5	81.035571	81	-0.035571
6	74.620231	74	-0.620231
7	71.006671	71	-0.006671
8	88.999732	89	0.000268
9	77.247206	78	0.752794

This is the prediction of DBP. The difference between the predicted data and the test data is very small.

Results

Objective Three

- Mean square error and R2 have been used to validate this predictive model.
- The MSE was 0.182 and the R2 score was 0.99.
- To further validate the model, new data was collected from the participants and fed into the ML model for prediction purposes.

Results

Objective Three

ACTUAL PREDICTIONS FOR INDIVIDUALS

```
In [25]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_01 = model.predict([[74, 61, 0, 23, 6, 21, 10, 38]])
         print(individualBP_01)

         [114.01360554]

In [26]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_02 = model.predict([[80, 86, 1, 30, 6, 33, 21, 45]])
         print(individualBP_02)

         [123.14378541]

In [27]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_03 = model.predict([[77, 77, 1, 18, 5, 3, 20, 41]])
         print(individualBP_03)

         [118.82677256]

In [28]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_04 = model.predict([[75, 66, 0, 27, 11, 9, 15, 46]])
         print(individualBP_04)

         [115.58930192]

In [29]: #'Dbp', 'Hr', 'Hyper_hist', 'BMI', 'Activity', 'User', 'Hour', 'Age'

         individualBP_03 = model.predict([[83, 92, 1, 18, 5, 3, 6, 41]])
         print(individualBP_03)

         [127.56852753]
```

- The image shows the variables the researcher used to predict an individual's BP.
- These variables include the DBP, SBP, the participant's hypertension history, their BMI, the activity, mood, the hour the BP was taken and their age.

Conclusions

- Almost all the participants wanted to either remain with the watches, or buy them. They liked the functionality of the watch and the Smarthealth application. The limitation encountered in this section was the time it took for the development of the Smarthealth application.
- The analyses revealed strong relationship between some variables, examples being the BMI and sleep pattern, the BMI and age and between the medication and age of an individual to name a few. The limitation encountered in this section was the lack of complete data.
- The performance of the model was validated using the mean square error and R Squared (R^2). Both scoring methods gave good results of 0.18 and 0.99 respectively. The scores showed that the model achieved its main aim of predicting an individual's future BP.

Recommendations

- For objective one, it is recommended to use the application developed with a better smartwatch in terms of Bluetooth connectivity, which will then aid in collection of more data from the individuals.
- For the second objective, further studies could include the use of other models that can predict panel data.
- In terms of validation of the predictive model, which was the third objective, further studies could use alternative evaluation methods to further check the correctness of the models.
- The methods of this research can be applied in making custom informative notifications through a phone application, website, email or Short Message Service (SMS), for each individual in order to prevent high BP or even lower BP in case of a hypertensive patient.

THANK YOU!