SASTA OLIPHANT SCIENCE AWARDS

SOUTH AUSTRALIAN SCIENCE TEACHERS ASSOCIATION

Prize Winner

Scientific Inquiry

Year 11-12

Shanza Ismail

Wilderness School

Oliphant Science Awards: Science Inquiry

Research Question: Can heartbeat pulses be used as a non-invasive method for estimating blood pressure?

Inquiry Extension: Can heartbeat pulses be used for human identification?

Shanza Ismail, Yr 12, Wilderness School

Introduction

Background Information

Hypertension or high blood pressure (BP) is a serious medical condition and a major cause for numerous cardiovascular diseases including heart attacks, stroke and kidney disease. Whereas low BP may indicate underlying conditions such as heart or endocrine problems, dehydration, severe infection, or even blood loss [1-2]. According to the World Health Organisation (WHO), an estimated 1.28 billion adults aged 30–79 years worldwide have high BP. It estimated that 9.4 million deaths each year are attributable to elevated BP [3]. The detection, diagnosis and monitoring of elevated BP remains a public health challenge, and the development of a user-friendly device to continuously measure BP such as one that is wearable is still an open challenge.

This investigation uses artificial intelligence (AI) and machine learning (ML) methods to investigate whether heart pulse data can be used to estimate BP. This investigation is carried out by developing complex mathematical models to estimate BP. AI/ML-based models were trained using measured heart pulse data and blood pressure measurements. After training these models, the accuracy of the trained models was evaluated using new heart pulse data. The accuracy was evaluated by comparing the estimated blood pressure value generated by the trained model with the actual blood pressure measured by the traditional cuff-based blood pressure monitor. As an extension to this inquiry, the described approach was also adopted to investigate whether heart pulse data can be used for human identification.

Artificial intelligence (AI) is a broad field that focuses on developing mathematical models to mimic natural behaviour and human intelligence to perform complex tasks. Machine learning (ML) is a subset of AI that focuses on enabling systems to learn from data without explicit programming. For a comprehensive list of AI/ML methods/models please refer [4-6].

Investigation Relevance

The main focus of this scientific inquiry is to investigate whether heart pulse information can be used to estimate a person's BP. Currently BP is measured using an inflating cuff that imparts an external pressure to the arm to stop the blood flow. Releasing this external pressure allows determination of the systolic and diastolic BP, as pressures that correspond to stages of initiation and unimpeded flow of blood, respectively. Cuff-based devices often cause discomfort and inconvenience for those who need frequent BP monitoring. Table 1 shows the different categories of BP levels defined by the WHO [7].

Category	Systolic BP (mmHg)	Diastolic BP (mmHg)
Normal	120 ~ 129	80 ~ 84
High-Normal	130 ~ 139	85 ~ 89
Grade I - Hypertension	140 ~ 159	90 ~ 99
Grade II - Hypertension	160 ~ 179	100 ~ 109
Grade III - Hypertension	>180	>110

Table 1: Classification of BP levels

As an extension of this scientific inquiry, we also investigate whether heart pulse information can be used for human identification. Currently common human identification methods include use of passwords and smart cards. However, now these methods are outdated as they can be lost, stolen and/or shared. A biometric system based on heart pulse signals may offer a more robust method for authentication as the heart pulse of an individual cannot be falsified. This is a stronger method as fingerprints can be fooled with fake fingerprints; faces can be extracted using the user's photo; and voice can also be imitated [8].

A literature search on the above two research areas reveals that only a limited amount of work has been conducted in this field so far [8-9]. To the best of the author's knowledge, the existing literature does not sufficiently explore or demonstrate the use of large sets of AI/ML models in investigations within these fields.

Aim & Hypothesis

Aim	To investigate whether heartbeat pulses can be used for estimating blood pressure and human identification.
Hypothesis	Heartbeat pulses contain physiological features that can be analysed to estimate blood pressure and heartbeat pulses are unique to each person.

Table 2: Investigation Aim & Hypothesis

Variables

Independents Variable	The independent variable was heartbeat pulse signals measured using an Arduino pulse sensor. Different samples were collected across time and individuals to provide variety in the data for model learning and was not manually manipulated.
Dependent Variable	The dependent variable was blood pressure in millimetres of mercury (mmHg) measured using an Omron Digital Blood Pressure Monitor. Measurements of the dependent variable were taken at the start and end of each data collection session where the average of both reading was taken as the reference BP value. For human identification, the dependent variable was the individual ID number.

Table 3: Independent & Dependent Variable

Controlled Variable	Method of Control
Pulse Sensor	The same Arduino pulse sensor was used for all trials to reduce measurement variability.
Pulse Sensor Placement	The sensor was secured and placed on the same finger of each participant. Placement of the sensor on different fingers or not securing the sensor to make sure it was not loose could reduce quality of the pulse signal and introduce inconsistency in the data.
Blood Pressure Monitor	The same Omron Digital Blood Pressure Monitor was used throughout the investigation. Using the same device avoided inconsistencies between different monitors.

Posture / Body Position	Data was collected when all participants were calm and resting as well as seated. This was as physical activity or standing could cause fluctuations in heart rate and blood pressure, and this would reduce the accuracy of the model.
Data Collection Time	Each data collection session was approximately 60 seconds.

Table 4: Controlled Variables

Uncontrolled Variable	Why
Intake of Food / Drink	The consumption of food or drink was not monitored or restricted in participants in this investigation. However, these factors could have temporarily influenced cardiovascular function and blood pressure.
Biological Differences	The data of this inquiry was collected from three individuals of differing age, sex and physiological characteristics. These inherent biological differences such as resting heart rate, blood volume and skin thickness could influence the pulse waveform in addition to factors such as fitness level and/or underlying health conditions.

Table 5: Uncontrolled Variables

Materials & Method

The Arduino IDE software is used to program the Arduino UNO board using the laptop and then to read pulse data from a pulse sensor connected to UNO board. Figures 1 and 2 show the connection diagram of the pulse sensor with the UNO and the script used to program the board. Heart pulse data was recorded and stored on the computer using a software called PuTTY. PuTTY is a free and open-source software, primarily used for connecting to external devices via serial ports of the computer to capture data from the connected external device. Figure 3 shows the console of the PuTTY software terminal.

The Pulse Sensor used is a small, lightweight, and low-power heart pulse sensor designed for use with Arduino. The sensor comes with a light emitting diode (LED) and photodetector as shown in Figure 4. The sensor works by shining a green light using the LED onto thin areas of the body, like the fingertip or earlobe, where light can pass through the skin easily. Some of this light is absorbed by the blood, and the rest is reflected to a photodetector. This technique, called Photoplethysmography (PPG), measures changes in blood volume using light. Oxygenated haemoglobin in the blood absorbs more green light, so during a heartbeat, less light is reflected. Between beats, more light is reflected, allowing the sensor to detect pulse patterns based on these fluctuations. To get a stable reading, the pulse sensor was attached to the finger by a Velcro tape as shown in Figure 5.

The recorded data sets were made available to the wider research community by storing them in the publicly accessible GitHub site. These data sets can be accessed from the link given in [10].

The Arduino IDE and PuTTY software can be downloaded from [11] and [12] respectively.

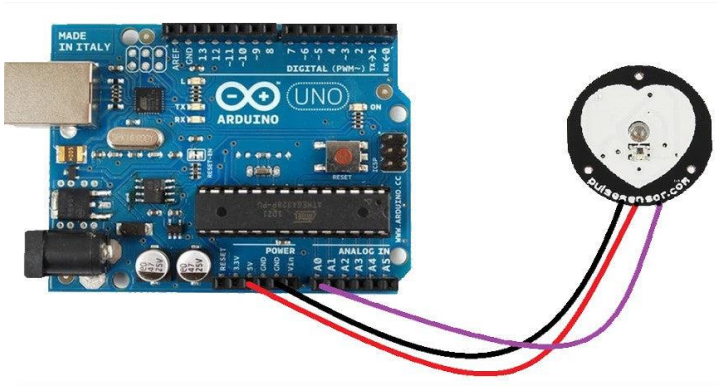

Figure 1: Interfacing Pulse Sensor with Arduino UNO board

read_heartPulse | Arduino 1.8.19

File Edit Sketch Tools Help

read_heartPulse §

```
int const PULSE_SENSOR_PIN = A0;  // 'S' Signal pin connected to A0

int Signal;           // Store incoming ADC data. Value can range from 0-1024
int Threshold = 500;  // Determine which Signal to "count as a beat" and which to ignore.

void setup() {
  pinMode(LED_BUILTIN, OUTPUT);  // Built-in LED will blink to your heartbeat
  Serial.begin(9600);            // Set comm speed for serial plotter window
  //delay(10);
}

void loop() {
  Signal = analogRead(PULSE_SENSOR_PIN);  // Read the sensor value
  //delay(100);

  Serial.println(Signal);  // Send the signal value to serial plotter

  delay(20);
}
```

Figure 2: Script for Programming the Arduino UNO board

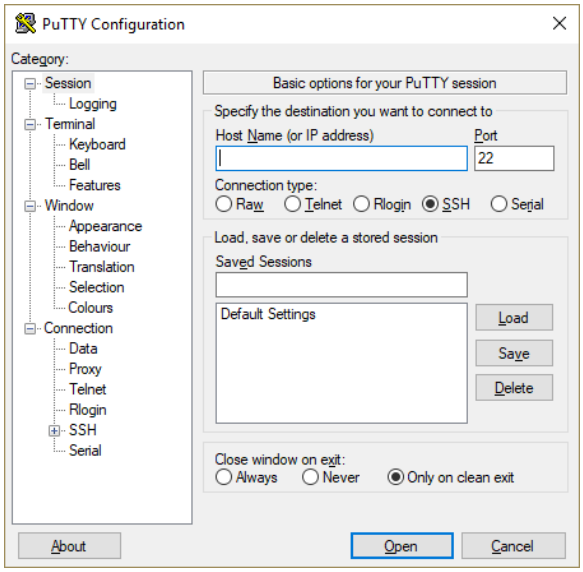

Figure 3: PuTTY Console

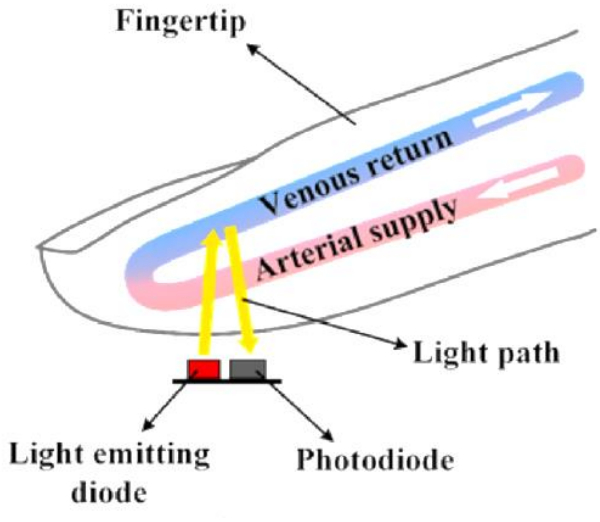

Figure 4: Operation of the Pulse Sensor

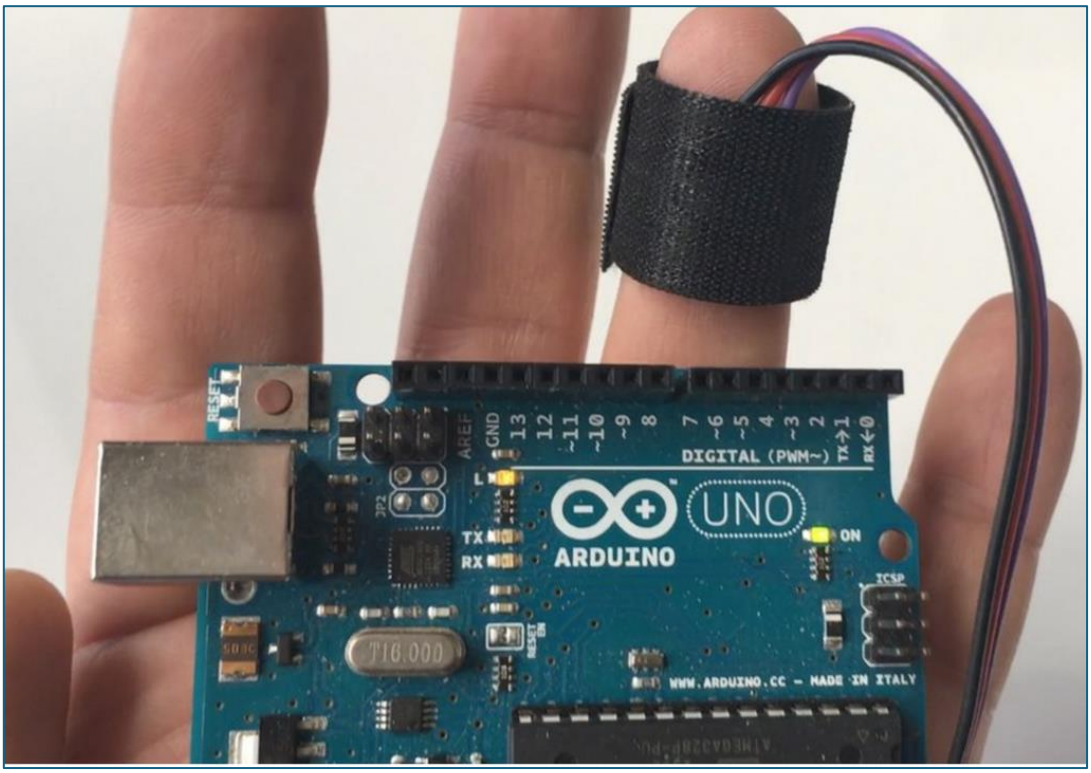

Figure 5: Attaching the Pulse Sensor to the Finger

Materials:

- 1 x Arduino UNO Board
- 1 x USB Type-B Cable
- 1 x PPG Pulse Sensor (Arduino UNO Compatible)
- 1 x Velcro Tape, 2cm
- 1 x Laptop / Computer
- 1 x Omron Blood Pressure Monitor
- Arduino IDE Software
- PuTTY Software
- MATLAB Software [13]
- Microsoft Excel

Method:

Device Setup

1. Pulse sensor was connected to the Arduino UNO board with standard wiring (See Fig.1).
2. Arduino UNO board was interfaced using the USB cable to the computer.
3. Arduino IDE software was opened, and script was uploaded to program Arduino UNO board (See Fig.2).
4. PuTTY was opened on the computer and was configured to capture serial data from the COM port that was connected to the pulse sensor. (See Fig.3).

Data Collection

5. Blood pressure was monitored using Omron BP monitor at beginning of data capture session.
6. Velcro was used to attach sensor to fingertip to avoid movement of the sensor (See Fig.5).
7. Heart pulse signal was recorded for approximately 60 seconds (for each session).
8. Blood pressure was monitored using the Omron BP monitor at the end of the data capture session.
9. Blood pressure monitored with the Omron BP monitor before and after the data capture was averaged and associated with the corresponding heart pulse data.

*Steps 4-9 were repeated for all participants and multiple sessions were conducted at different times of the day over a period of 7 days to ensure variability in the data.

Data Pre-processing

10. Raw heart pulse data captured using PuTTY was opened in Microsoft Excel.
11. Each set of continuous data points was divided in segments of 150 points per row. This analysis used 242 segments; this corresponds to a total of 36,300 heart pulse data points. (See Fig.6, 7)
12. The systolic and diastolic BP readings were recorded in the adjacent columns for each row. This was the average of the two BP readings taken in before and after the data capture sessions.
13. A final column was added to identify the person's ID (eg, 1 = Dad, 2 = Myself, 3 = Sister)

Model Development

14. The fully labelled data table was saved as a [.csv] file named 'Pulse_BP_Data.csv' for use in MATLAB software [13].
15. The 242 segments were randomly shuffled in Excel using a [RAND()] column and sorting function.
16. The file was imported in MATLAB and data was divided into two sets with the first 200 rows used as training data whilst the remaining 43 rows for testing. The training and testing of the models were carried using input data as shown in Figure 8.
17. Using the 'Import Data' tool in each app on MATLAB, the CSV was imported into the 'Regression Learner App' for BP estimation and the 'Classification Learner App' for human identification. (See Fig.9)
18. A range of models was trained automatically using built in presets in each app (eg. Squared GPR, Matern GPR, Cubic SVM). Description of the types of AI/ML models available in the MATLAB Apps is given in [14].

Model Evaluation

19. After training the AI/ML models, the test data set was used to evaluate the performance of each model. It should be noted that the test data was not used at part of the training data.
20. For the models trained to estimate the BP values, the following four metrics were generated as part of the model evaluation process.
    - **Mean absolute error (MAE):** A lower MAE indicates better model performance.
    - **Mean squared error (MSE):** A lower MSE suggests a more accurate model.
    - **Root mean squared error (RMSE):** A lower RMSE suggests a more accurate model

- **Coefficient of determination (R^2):** R^2 ranges from 0 to 1 (though it can be negative for very poor models). $R^2 = 1$: Perfect prediction; Higher R^2 means better fit.

21. The accuracy of models trained for human identification was estimated by calculating the percentage of correct identifications out of the total number of tests.
22. A test confusion matrix was generated to further evaluate model performance. This is a table that compares predicted labels with actual (true) labels, providing a clear visualisation of how accurately the classification model is performing.

Risk	Why	Control Method
Electric Shock	If the Arduino UNO board or computer came in contact with any liquids, it could cause a mild electric shock.	The devices were placed away from any liquids and the participants made sure they had dry hands before data collection.
Skin Irritation	The Velcro could cause mild skin irritation especially with repeated data collected sessions.	The data collections were short, and participants were monitored for skin irritation.
Device Overheating	The Arduino UNO board or computer could start to overheat if it was used for long periods of time.	Devices had time to cool in between sessions, and they were monitored.

Table 6: Risks and Methods of Controls

Segment (Row-wise)	1	2	3	4	.....	150	Systolic BP	Diastolic BP	Individual ID

Figure 6: Format of each Data Segment

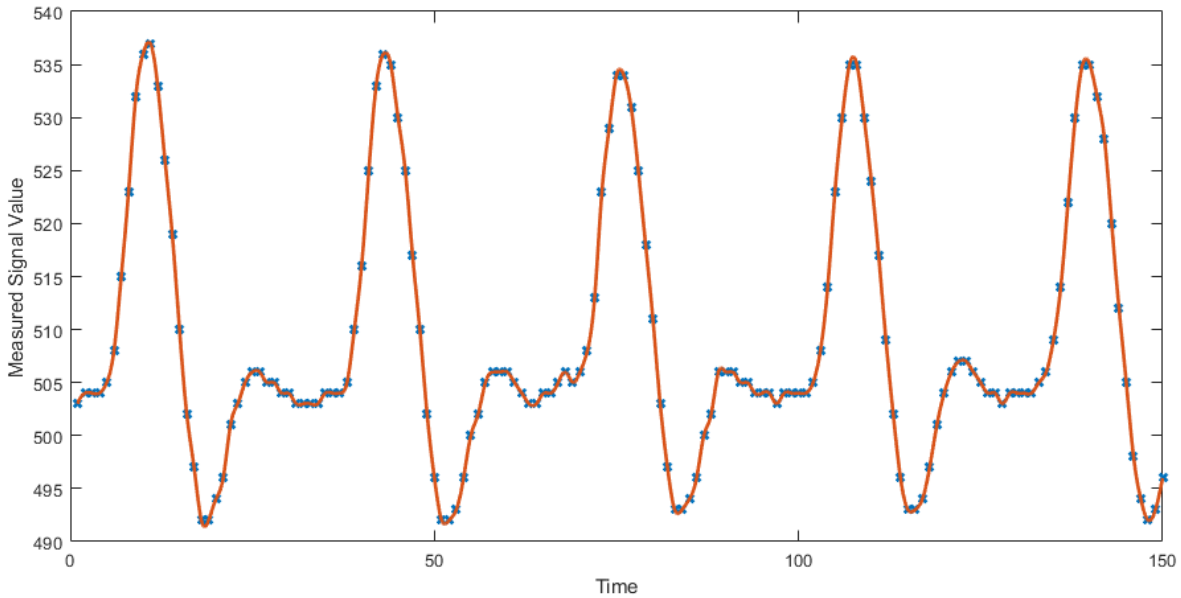

Figure 7: Example: 150 Heart Pulse Data Points in one Segment [Systolic-124; Diastolic-78; ID1-Dad]

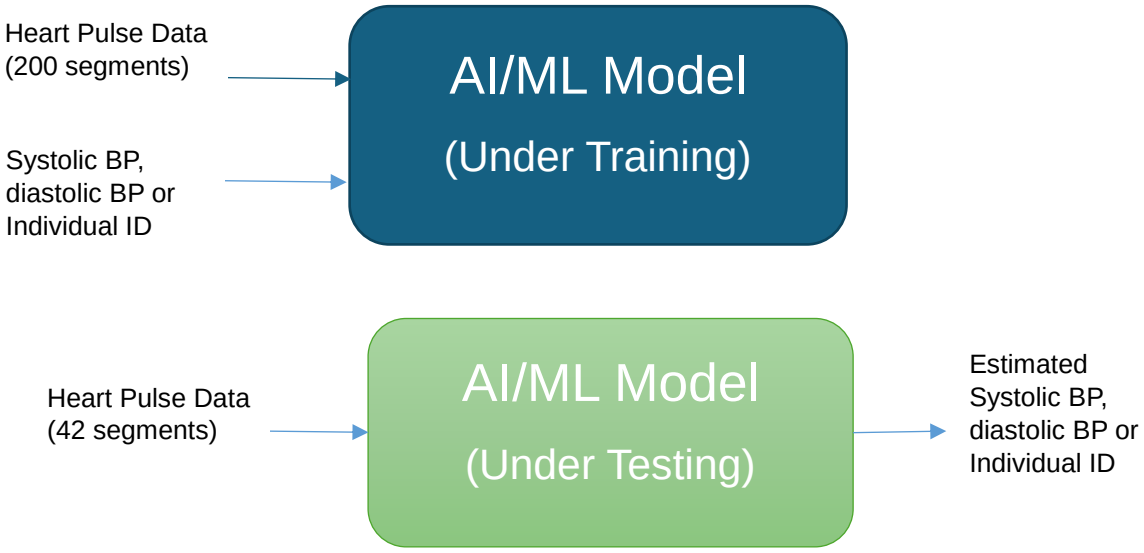

Figure 8: Training and Testing AI/ML Models

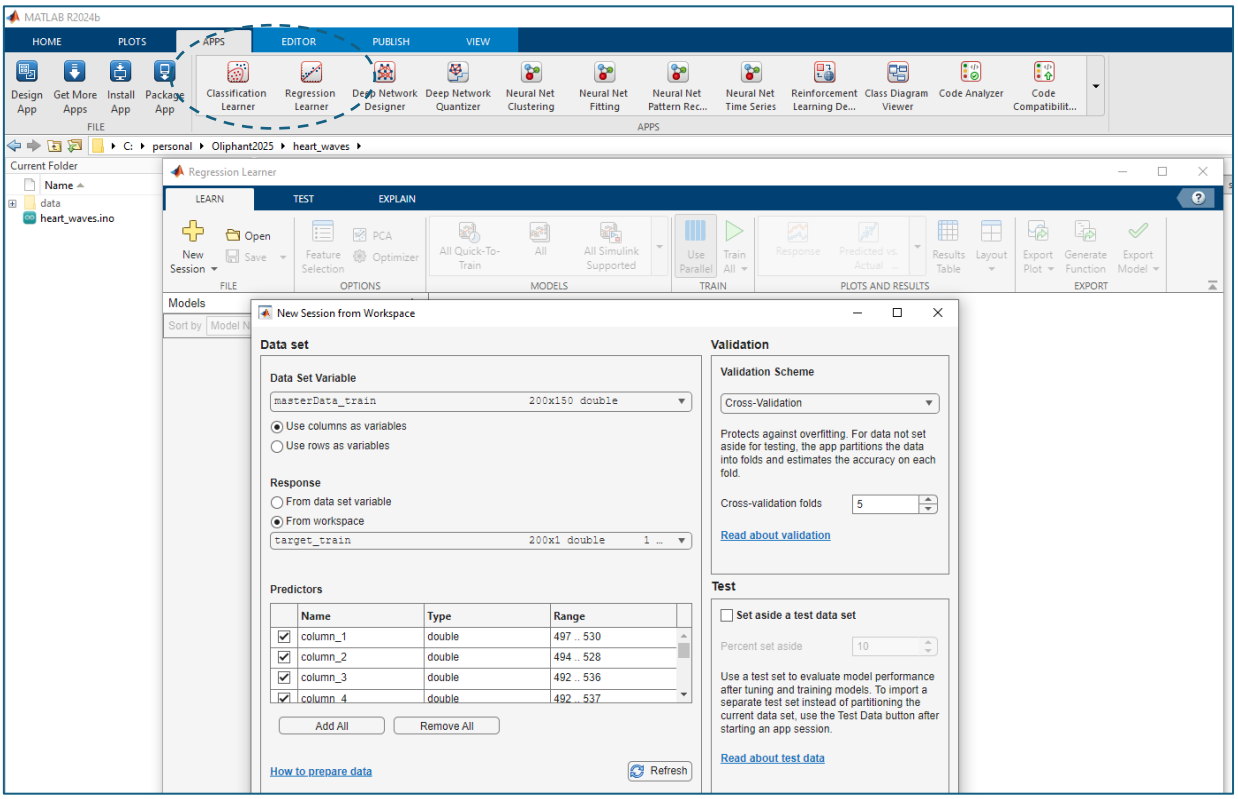

Figure 9: MATLAB Software with AI/ML Apps [13]

Results & Discussion

Prediction of Systolic BP

The test performance metrics generated for the trained models to estimate systolic BP are tabulated in Table 7.

Model Number	Model Type	MAE	MSE	RMSE	RSquared
2.1	Linear Regression	28.0	1142.1	33.8	-3.1
2.3	Linear Regression (Robust)	27.7	1134.7	33.7	-3.0
2.5	Tree (Fine)	10.0	303.2	17.4	-0.1
2.6	Tree (Medium)	10.5	277.0	16.6	0.0
2.7	Tree (Coarse)	12.6	244.3	15.6	0.1
2.8	SVM (Linear)	17.9	434.6	20.8	-0.5
2.9	SVM (Quadratic)	21.9	2248.4	47.4	-7.0
2.10	SVM (Cubic)	32.7	7752.0	88.0	-26.6
2.11	SVM (Fine Gaussian)	9.5	118.0	10.9	0.6
2.12	SVM (Medium Gaussian)	11.3	175.4	13.2	0.4
2.13	SVM (Coarse Gaussian)	13.8	274.6	16.6	0.0
2.14	Efficient Linear (Least Squares)	15.0	283.8	16.8	0.0
2.15	Efficient Linear (SVM)	14.3	302.3	17.4	-0.1
2.16	Ensemble (Boosted Trees)	8.8	135.6	11.6	0.5
2.17	Ensemble (Bagged Trees)	7.8	94.0	9.7	0.7
2.18	Gaussian Process Regression (Squared)	7.4	76.6	8.8	0.7
2.19	Gaussian Process Regression (Matern)	7.4	77.3	8.8	0.7
2.20	Gaussian Process Regression (Exponential)	7.7	86.5	9.3	0.7
2.21	Gaussian Process Regression (Quadratic)	7.4	76.6	8.8	0.7
2.22	Neural Network (Narrow)	22.7	1237.6	35.2	-3.4
2.23	Neural Network (Medium)	20.7	1141.4	33.8	-3.1
2.24	Neural Network (Wide)	23.6	1207.1	34.7	-3.3
2.25	Neural Network (Bilayered)	21.1	1122.2	33.5	-3.0
2.26	Neural Network (Trilayered)	14.0	305.4	17.5	-0.1
2.27	Kernel (SVM)	12.7	191.7	13.8	0.3
2.28	Kernel (Regression)	8.6	105.6	10.3	0.6

Table 7: Test Performance Metrics for Systolic BP Models

The test RMSE metrics for the trained models are shown in Figure 10.

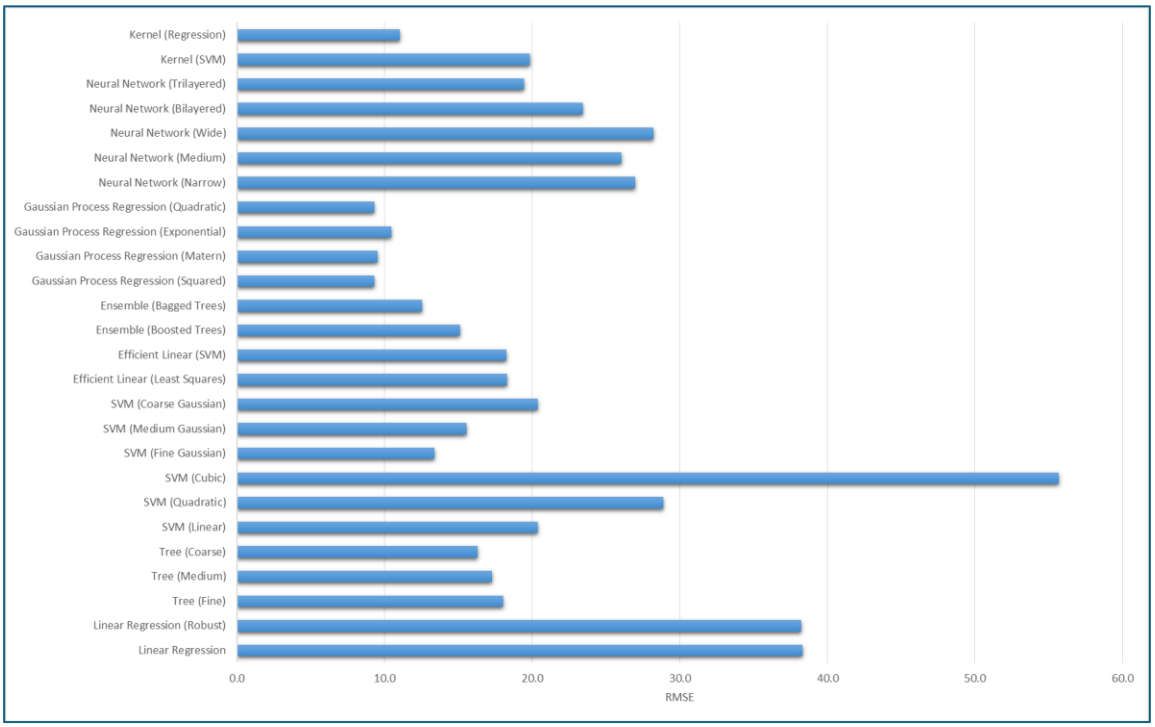

Figure 10: Test RMSE metrics for the trained Systolic AI/ML Models

From the results presented above, it was evident that the worst performing model was Cubic SVM whilst the best performing models for estimating systolic BP are the Squared Gaussian Process Regression (GPR), Matern GPR and Quadratic GPR. All these models give an RMSE of 8.8. By further analysing the test results, it was observed that these models are able to predict the systolic BP within +/-10 mmHg 81%, 83%, 81% of the test cases respectively. Out of these three models, the best performing model is therefore Matern GPR. Considering the high prediction accuracy (83%), it can be concluded that, heart pulse data can be used to estimate systolic BP.

Prediction of Diastolic BP

The test performance metrics generated for the trained models to estimate diastolic BP are tabulated in Table 8. The test RMSE metrics for the trained models are shown in Figure 11.

Model Number	Model Type	MAE	MSE	RMSE	RSquared
2.1	Linear Regression	31.8	1806.5	42.5	-8.6
2.3	Linear Regression (Robust)	31.8	1814.4	42.6	-8.6
2.5	Tree (Fine)	6.5	97.0	9.9	0.5
2.6	Tree (Medium)	9.4	172.6	13.1	0.1
2.7	Tree (Coarse)	10.6	149.4	12.2	0.2
2.8	SVM (Linear)	18.0	563.5	23.7	-2.0
2.9	SVM (Quadratic)	27.0	3323.9	57.7	-16.6
2.10	SVM (Cubic)	29.2	3254.7	57.1	-16.3
2.11	SVM (Fine Gaussian)	10.9	136.2	11.7	0.3
2.12	SVM (Medium Gaussian)	8.3	104.9	10.2	0.4
2.13	SVM (Coarse Gaussian)	12.5	236.5	15.4	-0.3
2.14	Efficient Linear (Least Squares)	13.4	191.7	13.8	0.0
2.15	Efficient Linear (SVM)	13.8	301.1	17.4	-0.6
2.16	Ensemble (Boosted Trees)	8.6	104.8	10.2	0.4
2.17	Ensemble (Bagged Trees)	9.1	113.1	10.6	0.4
2.18	Gaussian Process Regression (Squared)	8.2	96.0	9.8	0.5
2.19	Gaussian Process Regression (Matern)	7.9	88.1	9.4	0.5
2.20	Gaussian Process Regression (Exponential)	7.7	77.5	8.8	0.6
2.21	Gaussian Process Regression (Quadratic)	8.2	96.0	9.8	0.5
2.22	Neural Network (Narrow)	22.2	1145.2	33.8	-5.1
2.23	Neural Network (Medium)	22.3	1134.7	33.7	-5.0
2.24	Neural Network (Wide)	22.3	1145.4	33.8	-5.1
2.25	Neural Network (Bilayered)	19.1	852.1	29.2	-3.5
2.26	Neural Network (Trilayered)	19.4	1236.9	35.2	-5.6
2.27	Kernel (SVM)	11.6	153.7	12.4	0.2
2.28	Kernel (Regression)	9.4	109.2	10.5	0.4

Table 8: Test Performance Metrics for Diastolic BP Models

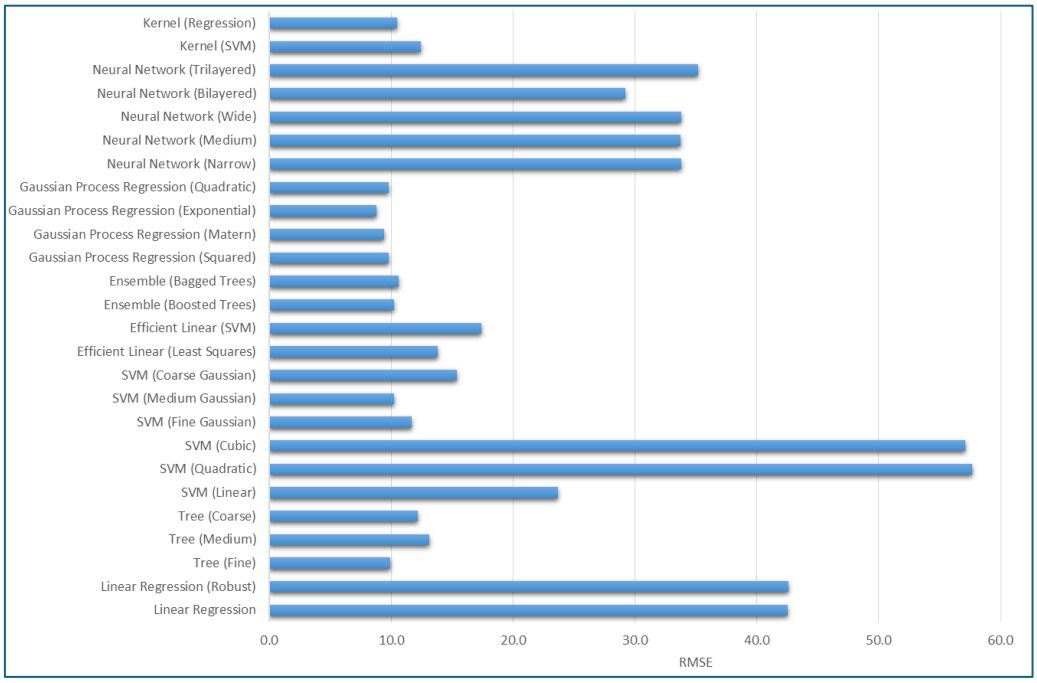

Figure 11: Test RMSE metrics for the trained Diastolic AI/ML Models

The generated evaluation results show the best model for estimating diastolic BP is Exponential Gaussian Process Regression (E-GPR). This model gives a test MAE, MSE, RMSE and R^2 values of 7.7, 77.5, 8.8 and 0.6 respectively. The test results also show that the E-GPR was able to predict the BP within +/-10 mmHg 67% of the test cases. Compared to the results obtained for systolic BP estimation, the prediction accuracy for diastolic BP prediction is lower. However, considering a prediction accuracy of 67%, it can be concluded that there is a relationship between heart pulse data and systolic BP.

Human Identification

The test results generated from the trained models developed for human identification are shown in Table 9, Figure 12 and Figure 13.

Model Number	Model Type	Accuracy %
2.1	Tree (Fine)	73.8
2.2	Tree (Medium)	73.8
2.3	Tree (Coarse)	64.3
2.4	Discriminant (Linear)	38.1
2.6	Efficient Logistic Regression	54.8
2.7	Efficient Linear SVM	54.8
2.8	Naive Bayes (Gaussian)	47.6
2.9	Naive Bayes (Kernel)	71.4
2.10	SVM (Linear)	59.5
2.11	SVM (Quadratic)	90.5
2.12	SVM (Cubic)	**95.2**
2.13	SVM (Fine Gaussian)	66.7
2.14	SVM (Medium Gaussian)	85.7
2.15	SVM (Coarse Gaussian)	54.8
2.16	KNN (Fine)	90.5
2.17	KNN (Medium)	71.4
2.18	KNN (Coarse)	28.6
2.19	KNN (Cosine)	73.8
2.20	KNN (Cubic)	71.4
2.21	KNN (Weighted)	81.0
2.22	Ensemble (Boosted Trees)	76.2
2.23	Ensemble (Bagged Trees)	78.6
2.24	Ensemble (Subspace Discriminant)	35.7
2.25	Ensemble (Subspace KNN)	**90.5**
2.26	Ensemble (RUSBoosted Trees)	76.2
2.27	Neural Network (Narrow)	83.3
2.28	Neural Network (Medium)	81.0
2.29	Neural Network (Wide)	88.1
2.30	Neural Network (Bilayered)	78.6
2.31	Neural Network (Trilayered)	83.3
2.32	Kernel (SVM)	83.3
2.33	Kernel (Regression)	83.3

Table 9: Accuracy of the trained AI/ML Model for Human Identification

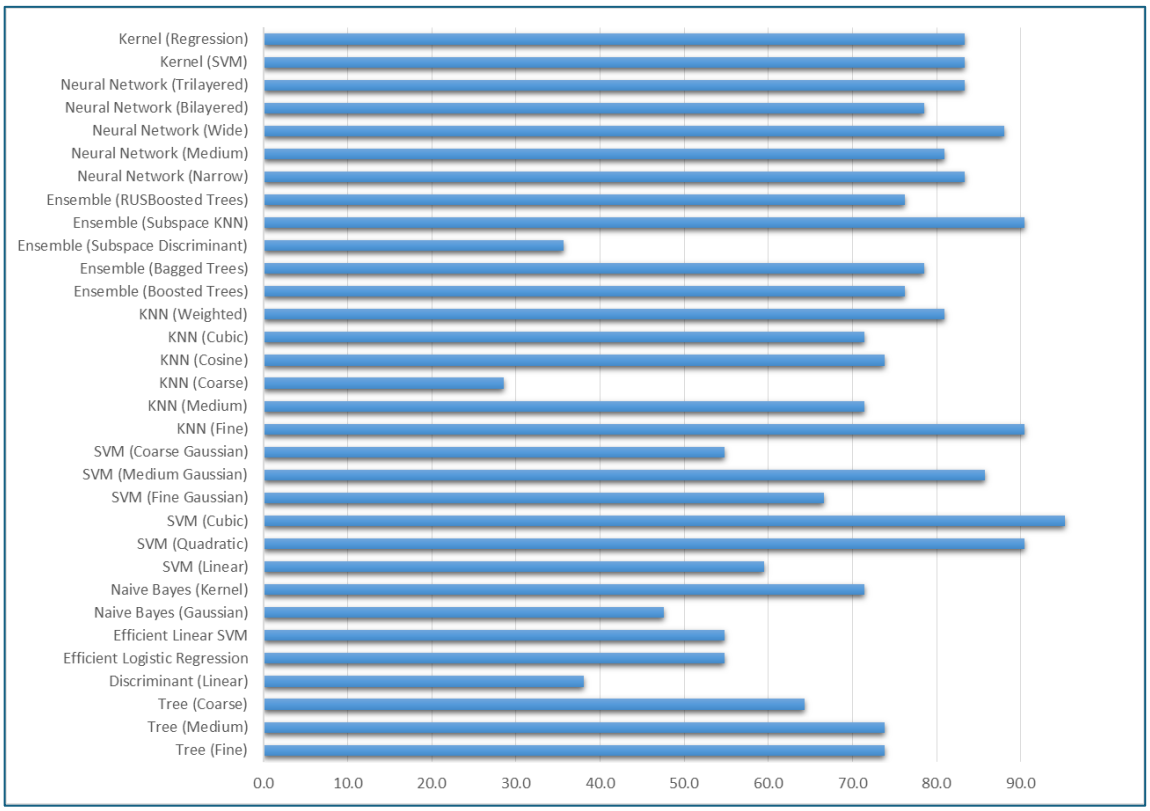

Figure 12: Test Accuracy (%) for the trained Human Identification AI/ML Models

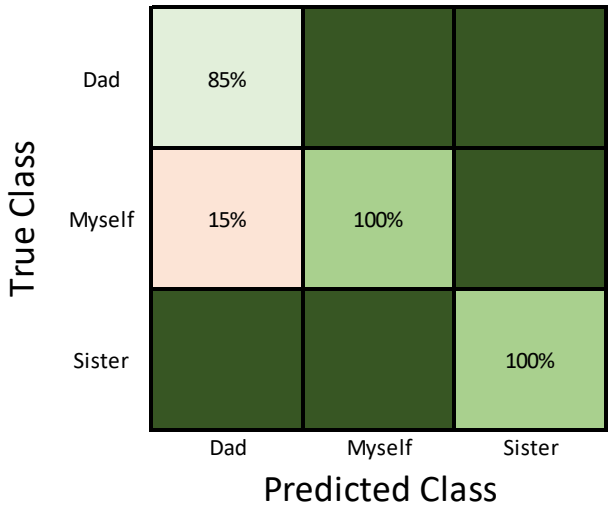

Figure 13: Test Confusion Matrix for Model 2.12 (Cubic-SVM)

The test results show that the Cubic SVM performs best for human identification. The Cubic SVM model gives an accuracy of 95%; indicating 40 segments out of 42 segments (each with 150 heart pulse data points) produced correct identification (i.e. matched with the individual ID associated with the respective segment). The confusion matrix shown in Figure 9 indicates that the identification accuracy for myself (ID2) and my sister (ID3) were correct for all test cases, while for my dad (ID1) only 85% of the test cases were correctly identified. Nevertheless, it can be concluded overall that human identification can be recognised with high accuracy using only heart pulse data.

Conclusion

This scientific inquiry investigated the use of heart pulse data to predict blood pressure. As an extension of this work, the efficacy of using heart pulse data for human identification was also explored. The recorded datasets showed that heart pulse data could predict systolic and diastolic blood pressure within ±10 mmHg with accuracies of 83% and 67%, respectively. These findings support the hypothesis and indicate that it is reasonably feasible to estimate blood pressure using heart pulse data. They also show that systolic blood pressure can be predicted more accurately than diastolic pressure using pulse readings. However, a limitation of this investigation was the small sample size which only included three individuals from the same household. This limited the diversity of the dataset and may have affected the model's ability to generalise to a broader population. To improve this investigation, further research is needed to utilise larger datasets, optimise model performance, and incorporate additional features such as age, weight, gender, and health conditions. This could help uncover deeper relationships between heart pulse data and blood pressure, potentially improving the model's accuracy and generalisability. The development of accurate and reliable blood pressure estimation models using heart pulse data has the potential to revolutionise hypertension management and contribute to improved cardiovascular health. The proposed method is also expected to provide a non-invasive, continuous, and real-time blood pressure monitoring solution.

The models developed for human identification showed that the Cubic-SVM produced an accuracy of more than 95% for the test cases. The high accuracy of the Cubic-SVM clearly proves that the AI/ML method for human identification is a viable option. However, there are many important issues that require further investigation before a robust system is developed for human identity recognition. This includes the stability of the heart pulse data over long periods of time (e.g. years), robustness to variation in mental and emotional state, and scalability to larger populations.

Word Count: 2126

References

[1] Sharma, S., M. F. Hashmi, and P. T. Bhattacharya. "Hypotension." StatPearls, StatPearls Publishing, 2025, https://www.ncbi.nlm.nih.gov/books/NBK499961/.

[2] Noh, Su A, et al. "History and Evolution of Blood Pressure Measurement." Clinical Hypertension, vol. 30, no. 1, 2024, p. 9. https://doi.org/10.1186/s40885-024-00268-7.

[3] World Health Organization. Hypertension. 16 Mar. 2023, World Health Organization, https://www.who.int/news-room/fact-sheets/detail/hypertension.

[4] Xu, Yongjun, et al. "Artificial Intelligence: A Powerful Paradigm for Scientific Research." The Innovation, vol. 2, no. 4, 2021, p. 100179. https://doi.org/10.1016/j.xinn.2021.100179.

[5] Russell, Stuart J., and Peter Norvig. Artificial Intelligence: A Modern Approach. 4th ed., Pearson, 2020.

[6] Brzeziński, Dariusz, et al., editors. Algorithms, Artificial Intelligence and Beyond: Theorising Society and Culture of the 21st Century. Routledge, 2025. https://doi.org/10.4324/9781032646930.

[7] Ma, Yinji, et al. "Relation between Blood Pressure and Pulse Wave Velocity for Human Arteries." Proceedings of the National Academy of Sciences, vol. 115, no. 44, 2018, pp. 11144–11149. https://doi.org/10.1073/pnas.1814392115.

[8] Ramli, D. A., M. Y. Hooi, and K. J. Chee. "Development of Heartbeat Detection Kit for Biometric Authentication System." Procedia Computer Science, vol. 96, 2016, pp. 305–314. https://doi.org/10.1016/j.procs.2016.08.143.

[9] Mohamed, M. "Measuring of Systolic Blood Pressure Based on Heart Rate." Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1, 2006, pp. 431–434.

[10] Ismail, S., "Measured Heart Pulse Time Series ad Blood Pressure", June 2025, GitHub: https://github.com/Shanza0/Heart-Pulse-Time-Series/blob/main/pulseData.csv

[11] Arduino Software 2025, https://www.arduino.cc/en/software/

[12] PuTTY 2025, Download PuTTY - a free SSH and telnet client for Windows, www.putty.org.

[13] MATLAB Software: https://au.mathworks.com/products/matlab.html

[14] MATLAB: Types of Machine Learning Models: https://au.mathworks.com/discovery/machine-learning-models.html

Appendix:

LOGBOOK

20/03/25

Today I was playing tennis and noticed my smartwatch displayed my heart rate in beats per minute. I wondered how it was possible for a watch to calculate this and whether it could also estimate blood pressure. I was already familiar with the cuff-based Oscillo-metric method to measure blood pressure as my dad regularly checked his, using a monitor. I questioned whether a wearable device like a smartwatch a more convenient alternative would be as the monitor was bulky and wired.

21/03/25

I was thinking about how my smartwatch was able to measure heart rates and wondered whether the same technology could be extended to estimate blood pressure. I began researching sensors used in smartwatches and came across a technique called photoplethysmography. Photoplethysmography is a technique which uses light to detect blood volume changes in the skin. If heartbeats create a measurable pulse wave, could that pulse signal contain enough information to non-invasively estimate blood pressure?

22/03/25

I wanted to explore if pulse waves could be used to estimate blood pressure and tried to finalise a research question for this inquiry. My initial question I created was ‘can pulse waves be used to predict blood pressure?’, this question however was vague and presented the topic rather than a detailed research question. To be more specific I changed the question to ‘can heartbeat pulses be used to estimate blood pressure?’. I thought a necessary change was changing ‘predict’ to ‘estimate’ as ‘estimate’ was a more accurate and precise word when referring to calculating blood pressure from indirect measurements. This new question was a significant improvement from my initial question however I changed it one final time to ‘can heartbeat pulses be used as a non-invasive method for estimating blood pressure?’. This final change was made as I believed it included the necessary details and was a scientific concise question that clearly presented what my inquiry was going to investigate.

24/03/25

I made some notes and a mind map to outline how I could investigate if pulse waves could be used to measure blood pressure. I realised I would need a way to collect heart pulse data and compare it to actual blood pressure readings to check the accuracy. Since I already had some experience with Arduino boards from previous projects, I decided to use an Arduino UNO along with a compatible heart pulse sensor to record the data. I started searching online for components, tutorials, and sample codes that could help me set up a working system.

25/03/25

I started reading more scientific papers and online resources about non-invasive blood pressure monitoring and how machine learning has been applied to similar biomedical problems. These articles have given me the idea to use AI/ML techniques to train a model on heart pulse signals and estimate blood pressure values. The more papers I read, and new information learnt made me more excited to build and test my own model.

01/04/25

I made a list of things I would need for my inquiry for data collection and noted that I would need a pulse sensor. I also decided to purchase a new Arduino UNO board as I could not find my old one, so I searched

through eBay and ordered an Arduino UNO board and a Heart Pulse Sensor that was compatible with the board.

05/04/25
My Arduino board and pulse sensor have not arrived yet however I wanted to work on my inquiry. I searched for an Arduino code that could read heart pulse data from the pulse sensor. I found a script online that I thought would work so I modified it. I also set up a framework for my data collection and recording.

10/04/25
My Arduino UNO board and Pulse Sensor arrived today via AustPost. I ran some basic tests however I encountered a problem. I found that Arduino IDE didn't support saving data. I researched some solutions for this problem and solved it by downloading and configuring PuTTY to capture heart pulse data from the Arduino serial port.

11/04/25
I conducted a literature review and investigated how AI/ML is used in health diagnostics. I even read about how blood pressure changes with health conditions and how pulse sensors work. I read an interesting article where I discovered that blood pressure is also connected to dementia and regulating blood pressure could decrease the risk of having dementia.

14/04/25
I printed out the risk assessment form and filled it in.

15/04/25 – 28/04/25
Over these few days I collected pulse data from myself, my father, and my sister. I used the blood pressure monitor I had to measure blood pressure before and after each pulse recording. This monitor was an Omron BP monitor. I recorded 60-second segments of data over different times of the day for 7 days.

29/04/25
I looked through all the data I had recorded and trimmed the ends to remove erroneous reading from when initialling and then I saved the final segments in Excel. I divided the final data into two segments which were training datasets and test datasets.

05/05/25
I needed to use the software MATLAB for AI/ML model development and testing so I watched some online tutorials and asked my dad for a training session on the basics of using this software. This training session was helpful, and I used the skills I learned from it as well as tutorials I found on YouTube to try using built-in AI apps to test regression models for BP estimation.

10/05/25
I decided to extend my inquiry and explore another idea I had which was using heart pulse data for identifying individuals. This was as the same data I was measuring could be used to answer this interesting secondary question that was not leaving my thoughts alone.

11/05/25
I carried out more literature searches. I researched biometric systems and how they use ECG and PPG for authentication. I even read about support vector machines (SVM) and classification models.

15/05/25
Today, I started training AI models in MATLAB using the data I captured. The results however were not good, and I was a bit disappointed with these results that were generated for the blood pressure prediction. However, I found that the human identification models showed high accuracy from the start. This made me quite happy.

20/05/25
I found out that the unexpected results were due to incorrect selection of the column/row variables by MATLAB. Now, the results look OK. It looks like more than 80% of test cases generate a BP estimate within +/-10 mmHg of the reading measured the Omron BP monitor. These results made me happy as it showed that heartbeat data can be used to estimate blood pressure. At this stage the results don't look that great, but they are promising. There are several variables that we cannot have a good handle on. For example, even the Omron BP monitor also does not give the same reading every time. Also, the pulse sensor can also give erroneous reading when the finger is slightly moved.

05/06/25
I repeated the human identification model using Cubic SVM. This was very good and was very happy as I got 95% accuracy. The results impressed me with how accurate it was.

06/06/25

I uploaded the data I measured onto GitHub so that other people could access it and be able to conduct further investigations.

07/06/25
I created an outline for the report with the introduction, method, results, discussion, and conclusion. I added to these sections with ideas of things I wanted to include.

10/06/25
I started adding my results into the results and discussion section. I included graphs from MATLAB and model performance results and discussed these results.

15/06/25
Today, I finished the introduction section. I added background information about the tools I used as well as information about blood pressure and why it needs to be measured. I also included my aim, hypothesis and variables.

17/06/25:
My materials and method section was completed. I had written my method in large paragraphs explaining what I did however I found that this did not make the method look that clear. To make it clearer and simpler so that another person could easily replicate it, I added labelled figures and added numbered steps in my method instead of large chunks of writing.

20/06/25
I completed the first draft of my report and reviewed everything to make sure it logically made sense.

28/06/25
I looked through the rules on the Oliphant Science Awards website and found that it clearly said the word count was 2000 words however my word count was slightly over by 500 words. I read through my inquiry report again and made my sentences more concise and reduced the word count to be 2194 words.

OSA RISK ASSESSMENT FORM

for all entries in (✓) ☐ Models & Inventions and ☑ Scientific Inquiry

This must be included with your report, log book or entry. One form per entry.

STUDENT(S) NAME: Shanza Ismail ID: 25047

SCHOOL: Wilderness School

Activity: Give a brief outline of what you are planning to do.

Topic: Can heartbeat pulses be used to estimate blood pressure?

This inquiry investigates the use of heart beat pulse data to estimate blood pressure.

As part of this investigation, a mathmatical model (using AI learning algorithms) will be developed

to predict blood pressure for a given set of pulse data values.

Are there possible risks? Consider the following:

- Chemical risks: Are you using chemicals? If so, check with your teacher that any chemicals to be used are on the approved list for schools. Check the safety requirements for their use, such as eye protection and eyewash facilities, availability of running water, use of gloves, a well-ventilated area or fume cupboard.
- Thermal risks: Are you heating things? Could you be burnt?
- Biological risks: Are you working with micro-organisms such as mould and bacteria?
- Sharps risks: Are you cutting things, and is there a risk of injury from sharp objects?
- Electrical risks: Are you using mains (240 volt) electricity? How will you make sure that this is safe? Could you use a battery instead?
- Radiation risks: Does your entry use potentially harmful radiation such as UV or lasers?
- Other hazards.

Also, if you are using other people as subjects in an investigation you must get them to sign a note consenting to be part of your experiment.

Risks	How I will control/manage the risk
This inquiry will use commonly used instruments (such as blood pressure monitor and heart pulse measuring devices to capture heart beat and blood pressrure data. A software will be used to develop the AI model. No risks identfied.	N/A

(Attach another sheet if needed.)

Risk Assessment indicates that this activity can be safely carried out

RISK ASSESSMENT COMPLETED BY (student name(s)): Shanza Ismail

SIGNATURE(S): Shanza

☑ By ticking this box, I/we state that my/our project adheres to the listed criteria for this Category.

TEACHER'S NAME: Sam Bartram

SIGNATURE: DATE: 24/6/25