International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 5, ISSN No. 2455-2143, Pages 165-172
Published Online September 2022 in IJEAST (http://www.ijeast.com)

MACHINE LEARNING TECHNIQUE FOR EARLY PREDICTION OF HYPERTENSION

Aranuwa, F.O
Department of Computer Science, Adekunle Ajasin University,
Akungba – Akoko, Ondo State, Nigeria

Abstract — **Machine Learning (ML) has proved to be an invaluable tool in medical research today. It is a branch of artificial intelligence that has the ability to learn from complex data, identify patterns and make decisions with minimal or without human intervention. Efforts in this work is focused at developing an intelligent model for early prediction of hypertension because of the complexity associated with the pattern identification of the syndrome. The study acquired data pertaining to the syndrome from selected healthcare centres in the South West region of Nigeria, one of the largest city in Africa. A retrospective learning using ML analyses was conducted on the data. Three different intelligent classification algorithms namely; Gradient Boosting, Random Forest and SVM were used in the modeling. The datasets were trained and tested, and the performance of the algorithms were compared using standard performance metrics of accuracy and throughput. The results from the experiment indicates that the gradient boosting classifier outclassed other algorithms considered in the work with accuracy of 90.17%. The model has proved to be very suitable for early prediction of the syndrome and, will in no small measure assist the medical practitioners in providing quick and more effective way of detecting early the syndrome in patients. This will subsequently lead to timely treatment of the patients, thereby reducing its death rate in the society. Future work is focused at considering complex patterns of the syndrome and larger database.**

Keywords: **Hypertension, Gradient Boosting, HBP, Intelligent Technique, Machine Learning**

I. INTRODUCTION

Today**,** hypertension has been adjudicated one of the major causes of sudden and premature death both in the young and old, especially in the Africa continent**.** The syndrome has a direct relationship with high blood pressure (HBP) and other related health issues that often leads to life-threatening dares such as heart failure, coronary artery diseases, brain and kidney challenges even death if not properly managed (Afeni, Aruleba and Oloyede, 2017; Nimmala, Ramadevi, Sahith and Cheruku, 2018).

By definition, hypertension, is a condition in which the blood vessels have persistently raised pressure (WHO, 2021). Blood pressure is the force that a person's blood exerts against the walls of its blood vessels (Lawes et al., 2000). According to Felman (2019), hypertension arises when blood pressure is too high beyond default measures. Normal adult blood pressure is expected to be between 120 mm Hg (systolic) and 80 mm Hg (diastolic), but when it goes beyond this level, at two consecutives checkups may lead to high risk of hypertension. Globally, several factors that contribute to higher prevalence of hypertension include physical inactivity, inability to detect the syndrome early or, decreased awareness, smoking habit, unhealthy diet, inaccessibility to effective healthcare delivery and cost of medications.

Figure 1 and Figure 2 show blood pressure categories and recommendations respectively, while Figure 3 depicts hypertension complications.

BLOOD PRESSURE CATEGORY	SYSTOLIC mm Hg (upper number)		DIASTOLIC mm Hg (lower number)
NORMAL	LESS THAN 120	and	LESS THAN 80
ELEVATED	120 – 129	and	LESS THAN 80
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 1	130 – 139	or	80 – 89
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 2	140 OR HIGHER	or	90 OR HIGHER
HYPERTENSIVE CRISIS (consult your doctor immediately)	HIGHER THAN 180	and/or	HIGHER THAN 120

Figure 1: Blood Pressure Category
Source: ((American Heart Association, 2022)

Figure 2: Blood Pressure Recommendations
Source: (American Heart Association, 2017).

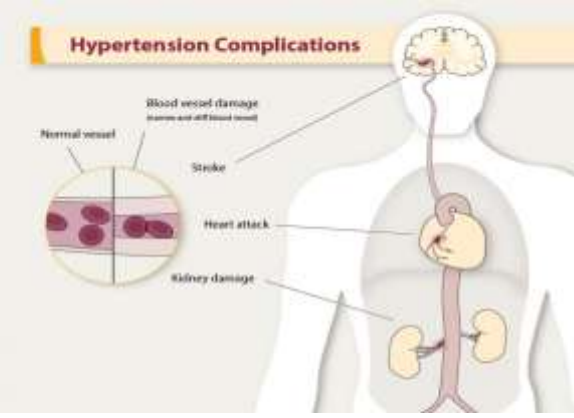

Figure 3: Hypertension Complications
Source: (CDC, 2021)

Normal levels of both systolic and diastolic blood pressure are particularly important for the effective and efficient function of vital body organs such as the heart, brain and kidneys, and for overall health and wellbeing (WHO, 2021). The syndrome can be classified into two main classifications, namely: Primary hypertension**;** also referred to as essential hypertension, which tends to develop gradually over some years (Nimmala et al., (2018). The second class is the secondary hypertension**;** which tends to appear suddenly due to some underlining conditions as highlighted in (Farran et al, 2013). According to Hannedouche and Krummel (2008), the features of primary secondary hypertension include: gradual increase with slow rate of rise in BP, episodic pallor, and dizziness, snoring, hyper somnolence (obstructive sleep apnea), prostatism (chronic kidney disease), muscle cramps, weakness, weight loss, palpitations, heat intolerance (hyperthyroidism), edema, fatigue, frequent urination (kidney disease or failure), facial rounding, and so on. Risk assessment of the disease is significantly more complicated and depends on multitude of factors and transient environmental conditions that can artificially raise blood pressure readings. Commonly identified risk factors include age, gender, body mass index, obesity, stress, triglycerides, uric acid, lipoproteins, cholesterol, smoking habits, and family history of the disease (Shinde & Rajeswari, 2018).
Table 1 shows blood pressure ranges by Age and Gender according to (Digest et al., 2019).

Table 1: Blood pressure according to Age and Gender

Age	Female	Male
1-2	80/34-120/75	83/38-117/76
3	100/59	100/61
4	102/62	101/61
5	104/65	103/66
6	105/68	104/68
7	106/70	106/69
8	107/71	108/71
9	109/72	110/72
10	111/73	112/73
11	113/74	114/74
12	115/74	116/75
13	117/75	117/76
14	120/75	119/77
15	120/76	120/78
16	120/78	120/78
17	120/80	120/78
18	120/80	120/80
19-24	120/79	120/79
25-29	120/80	121/81
30-35	122/81	123/82
36-39	123/82	124/83
40-45	124/83	125/83
46-49	126/84	127/84
50-55	129/85	128/85
56-59	130/86	131/87
60+	134/84	135/88

With aging, there is usually a dramatic increase in the prevalence of syndrome. By age 70 years, most people do have hypertension. It usually develops in man as from the age of 55 years though with a normal blood pressure at te beginning. By 60 years of age, approximately 60% of the population do have hypertension; by 70 years, about 65% of men and about 75% of women do have hypertension. According to Logan (2011), hypertension especially isolated systolic hypertension, is commonly found in older (60–79 years of age) and elderly people (≥80 years of age). Census bureau data and projections reveal that women older than the age of 65 years will make up more than 20% of the population by 2030. Figure 4 shows modifiable and fixed risk factors.

Figure 4: Modifiable and fixed risk factors of hypertension. Source: (Hannedouche & Krummel, 2008).

An estimated 1.28 billion adults aged 30-79 years worldwide have hypertension and 46% of adults with hypertension are unaware that they have the syndrome (WHO 2021). About 7.6 million death have been reportedly linked with this disorder annually. About 54% of stroke and 47% of coronary heart disease are attributable to high Blood Pressure. By year 2025, it has been predicted that an estimated population of 1.56 billion adults will be living with hypertension (WHO, 2021). Meanwhile, researchers and clinicians have used different approaches to discover the pattern of the syndrome, as reported in (Afeni et al., 2017; Nimmala et al., 2018; Farran, Channanath, Behbehani, & Thanaraj, 2013; Zhang, Ren, Huang, Cheng, & Hu, 2019; Zhang, Wei, Ren, Cheng, & Zheng, 2018; Li, He, Shao, Ou, & Lin, 2015; Obe, Balanica, & Neagoe, 2015) The reports from these approaches suffers some inadequacies and inefficiency due to the complexity associated with the syndrome. High blood pressure usually has no warning signs or symptoms, and many people do not know they have it. Hence, the need for an intelligent technique such as machine learning for early prediction/detection of the disease to reduce the harm and mortality caused by it.

Machine Learning (ML) has been discovered to be a viable tool for patterns identification in data with vital performance in medicine (Mbbs, 2019). Khanna and Awad (2015), defined machine learning as branch of artificial intelligence that systematically applies algorithms to synthesize underlying relationships among data and information. ML can be categorized into three categories, namely: Supervised Learning, Unsupervised Learning and Reinforcement Learning, (Abdi, 2016). The supervised learning category applies to what has been learned in the past to new data using labeled examples to predict future events. Unsupervised learning are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. While, reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers events. Meanwhile, trial and error search and delayed are the most relevant characteristics of reinforcement learning. Ayodele et al, (2010) and Abdi, (2016), highlights some machine learning algorithms which include: Gradient Boosting, Random Forest, SVM, Linear Regression**,** Logistic Regression**,** Decision Tree, Naive Bayes **,** KNN **,** K-Means. The selected classifiers are type of machine learning boosting techniques for regression and classification problems. According to Brownlee, (2016), Gradient Boosting especially is one of the most powerful machine techniques for building predictive models. Boosting is a method of converting weak learners into strong learners (Grover, 2017; Singh, 2018). It relies on the intuition for the best possible model, and the key idea is to minimize the overall prediction error.

II. METHODOLOGY

This section presents the methodology for the research work. It involves the research design, data collection, data processing, classifications and classifiers evaluation. Figure 5 presented the architecture of the proposed predictive model for the work.

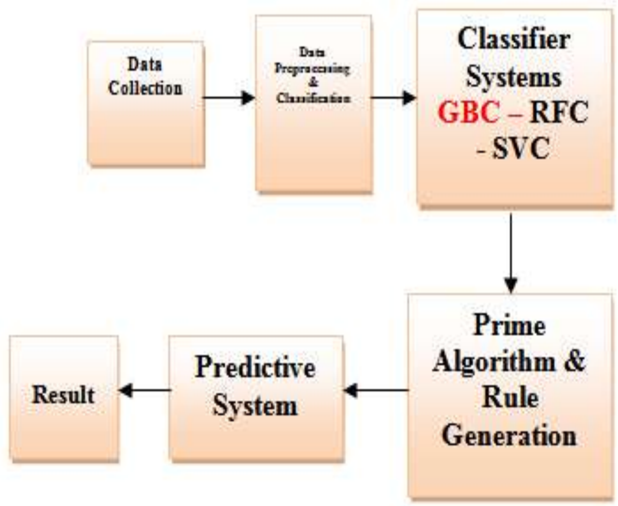

Figure 5: Architecture for the Predictive Model

Data Collection and Processing

A total number of 4,238 hypertension patients' dataset only was acquired due to some professional constraints from some selected health care centres in the Western region of Nigeria. Sixteen major attributes were collated and retrospective learning was conducted using Machine Learning analyses on the data. Table 2 depicted the selected attributes of the data. The extracted data were subjected to cleaning and filtering processes. Table 3 and Table 4 depicts the formatted data and statistical analysis of the data respectively.

TABLE 2: SELECTED DATA ATTRIBUTES

	Age	diabetes	sysBP	diaBP	BMI	G& FH	DAA
0	39	0	106.0	80.0	26.97	0	0
1	46	1	121.0	81.0	28.73	1	1
2	48	0	127.0	89.0	25.34	0	0
3	50	1	150.0	95.0	28.58	1	1
4	46	1	130.0	84.0	23.10	1	1
5	43	0	180.0	110.0	30.30	0	0
6	63	0	138.0	89.0	33.11	1	1
7	45	0	130.0	89.0	21.68	0	0
8	52	1	141.0	89.0	26.36	1	1
9	43	0	162.0	107.0	23.61	1	1
10	50	0	173.0	96.0	22.91	1	0
11	43	1	151.0	88.0	27.64	1	1
12	46	1	162.0	94.0	26.31	0	1
13	41	0	144.0	88.0	31.31	1	1
14	39	0	134.0	80.0	22.35	1	0
15	38	1	140.0	90.0	21.35	1	1
16	48	1	138.0	90.0	22.37	1	1

Age = Age of patient, diabetes =Presence of Diabetes Melitus, SysBP = Systolic Blood Pressure, diaBP = Diastolic Blood Pressure, BMI =Body Mass index, G&FH = Genetic & family History, DAA = Drug –Alcohol Addiction etc

Data Classification

The normalized data was divided into training and testing set using a 66% and 34% split ratio as shown in Table 5. The obtained data was normalized using Z-score normalization in order to make training less sensitive to the scale of features. Z-score converted the data into [0,1] distribution using equation (1).

$$x'_i = \frac{x_i - \mu}{\sigma} \quad \dots \text{ Equation (1)}$$

Where x'_i is the data value, x_i is the data value to be normalized, μ represents the mean of data values in the feature category. The training and testing dataset was done on Anaconda Jupiter Notebook using sklearn library. The training set was used to train the classifiers, the testing set was used to evaluate the predictive model.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 5, ISSN No. 2455-2143, Pages 165-172
Published Online September 2022 in IJEAST (http://www.ijeast.com)

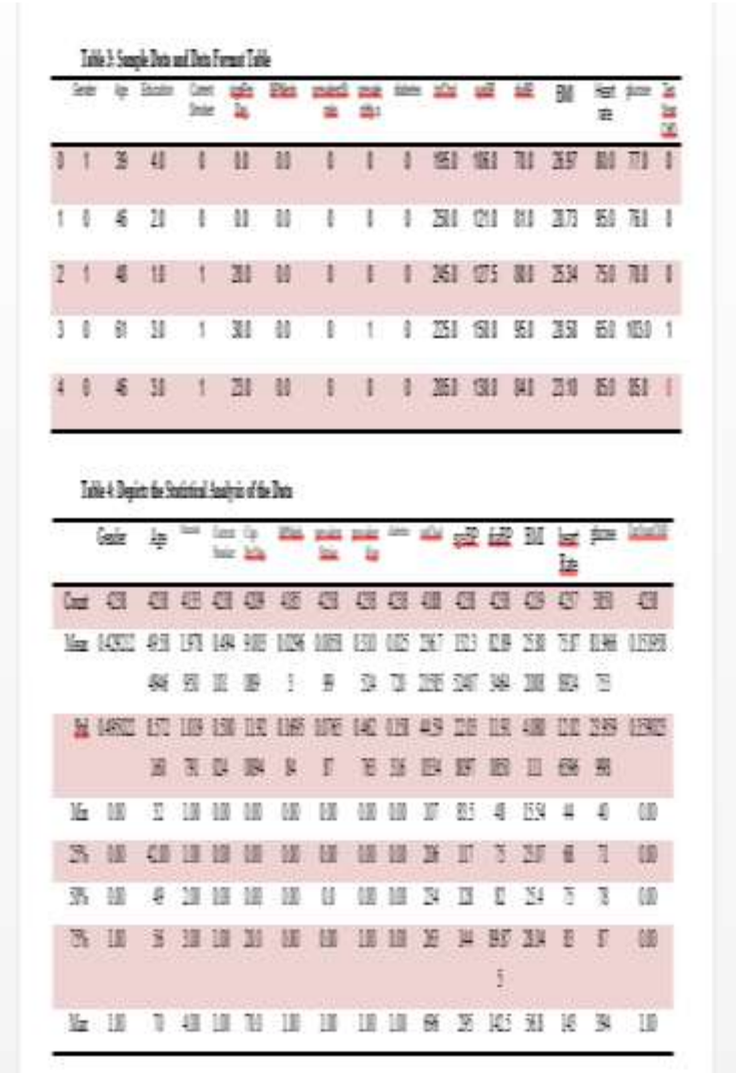
Table 3: Sample Data and Data Format Table

Table 4: Depicts the Statistical Analysis of the Data

Table 5: Data Classification

Class labels	Training set (66%)	Testing set (34%)
Hypertension	1,725	492
Non Hypertension	1,665	356
Total	**3,390**	**848**

III. EXPERIMENTAL DESIGN AND RESULTS

Anaconda platform was used for the experiment. The intelligent algorithm used for the work comprises of Gradient Boosting Classifier, Random Forest Classifier and Support Vector Classifier. Figure 6 shows the overview of Anaconda application window and the data input interface.

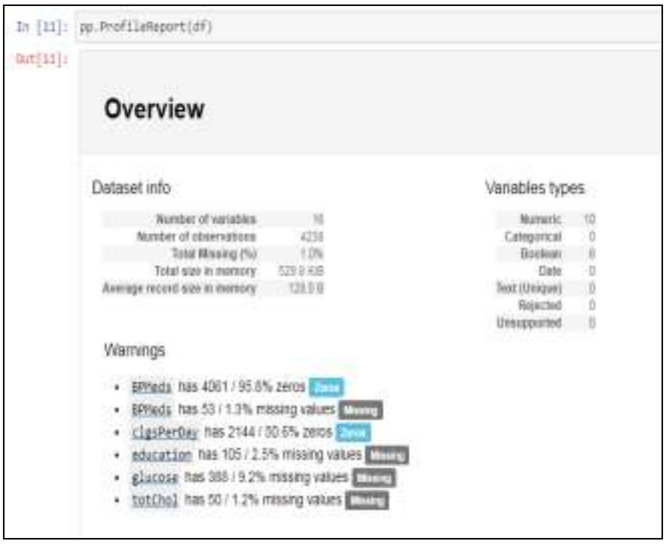

Figure 6: Overview of Anaconda Application Window

During the experiment, the dataset was fed into the Anaconda platform and Figure 7 shows the dataset information with the number of variables, number of observations and warnings.

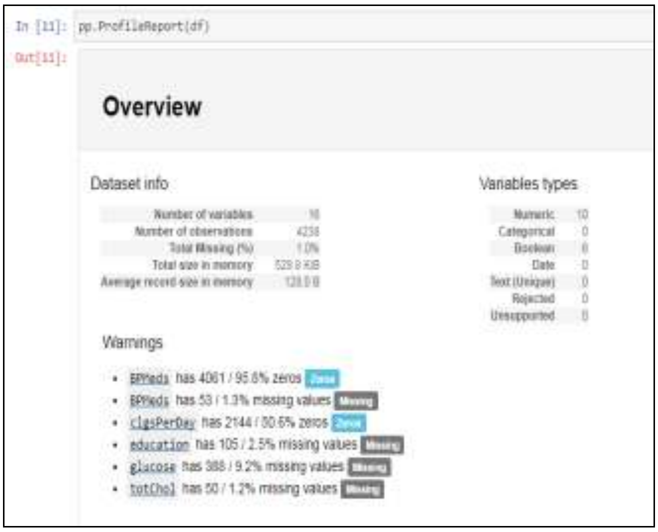

Figure 7: showing the data input interface.

Figure 8 and Figure 9 show the samples of attributes classification and correlation output respectively.

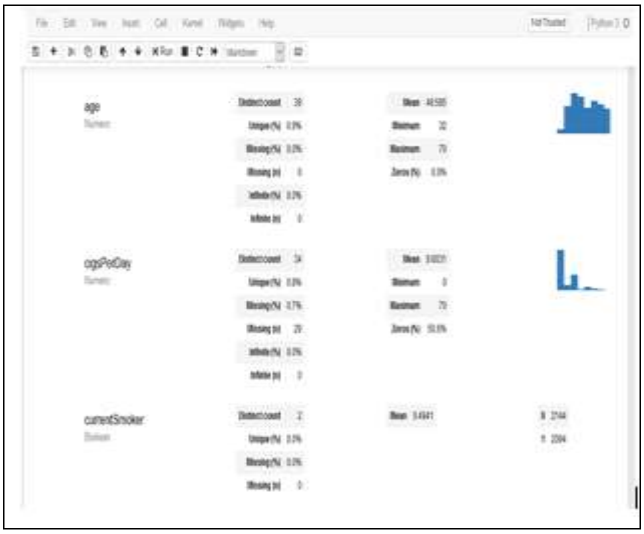

Figure 8: Samples of Attributes Classification

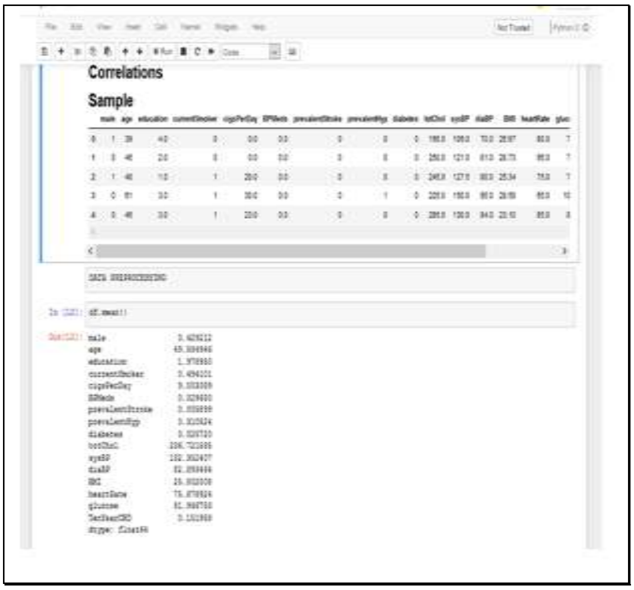

Figure 9: Sampled Correlation Output

The performance of the algorithms in the experiment in terms of accuracy, precision and F1 score metrics are presented in Table 6. Figures 10 depicts the comparison of the algorithm's performance.

Table 6: Classifiers Performance

	RFC	GBC	SVC
ACCURACY	87.97	90.17	89.3
PRECISION	78	85	78
F1 SCORE	80.3	84.47	82.1

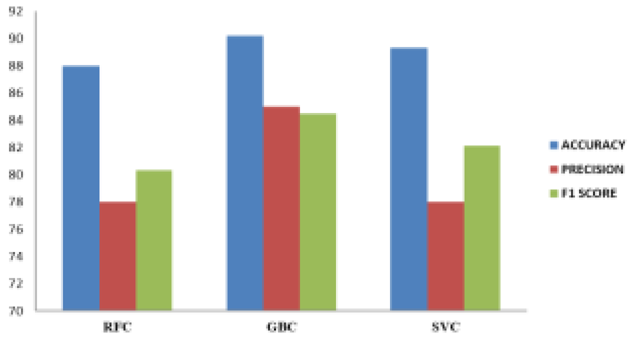

Figure 10: Comparison of the Algorithm's Performance

From the experimental results, the gradient boosting algorithm outperformed all the other algorithm tested in term of accuracy, precision and F1 score metrics with 90.17%, 85.0%. 85.0% respectively. The algorithm successfully learn and classified the input and predict correctly the level of the syndrome in each input during the testing mode. Commonly identified risk factors include age, gender, body mass index, obesity, stress, uric acid, lipoproteins, cholesterol, smoking habits, and family history of the disease.

IV CONCLUSION

Efforts in this work is focused at developing an intelligent model for early prediction of hypertension using a machine learning technique. The objective of the study was critically observed and achieved. The dataset collected was trained and tested with the intelligent classifiers developed for the work. The performance of the algorithms considered were compared using standard metrics of accuracy and throughput (precision and F1 scores). The results from the experiment show that the gradient boosting classifier outclassed other algorithms with an accuracy of 90.17%. The model has proved to be suitable for early prediction of the syndrome. The deployment of this model in our health care delivery system is believe will help to reduce the mortality rate of the syndrome in our society through timely detection and treatment of the syndrome in patients.

V REFERENCES

[1]. Abdi, A. (2016). Three types of Machine Learning Algorithms List of Common Machine Learning Algorithms. Retrieved from https://doi.org/10.13140/RG.2.2.26209.10088.

[2]. Afeni, B., Aruleba, T., and Oloyede, I. (2017). Hypertension Prediction System Using Naive Bayes Classifier. Journal of Advances in Mathematics and Computer Science, 24(2), 1–11. https://doi.org/10.9734/jamcs/2017/35610.

[3]. American Heart Association (2022). Healthy and unhealthy blood pressure ranges Retrieved online on 28/6/2022 **at** https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings**.**

[4]. American Heart Association (2017). Understanding Blood Pressure Readings. Retrieved 15/04/2021 **at** https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings**.**

[5]. Ayodele, T. O. (2010). Types of Machine Learning Algorithms. New Advances in Machine Learning. Retrived 28/6/2022 from www.intechopen.com.

[6]. Ben-david, S and Shalev-Shwartz, S (2014). Understanding Machine Learning : From Theory to Algorithms. Cambridge University Press Retrieved on 25/04/2021from: https://books.google.com.ng/books?id=ttJkAwAAQBAJ&source=gbs_navlinks_s.

[7]. Brownlee, J. (2016). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning. Machine Learning Mastery Retrieved on 28/04/2021 from https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/.

[8]. CDC (2021). High Blood Pressure symptoms and causes centres for disease control and prevention retrieved online at: https://www.cdc.gov/bloodpressure/about.htm# on 28/6/2022.

[9]. Digest, M., Does, W., Pressure, B., and& To, R. (2019). Blood Pressure Chart : Low , Normal , High Reading by Age.

[10]. Farran, B., Channanath, A. M., Behbehani, K., & Thanaraj, T. A. (2013). Predictive models to assess risk of type 2 diabetes , hypertension and comorbidity : machine-learning algorithms and validation using national health data from Kuwait — a cohort study. https://doi.org/10.1136/bmjopen-2012-002457.

[11]. Felman, A. (2019). What to know about high blood pressure. (February) Retrieved on 19/09/2022 at: https://www.medicalnewstoday.com/articles/150109

[12]. Grover, P. (2017). Gradient Boosting from scratch – ML Review – Medium. Medium. Retrieved on 28/04/2021 from https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d.

[13]. Hannedouche, T., & Krummel, T. (2008). [High blood pressure in adults]. La Revue Du Praticien, 58(10), 1117–1129. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18652414.

[14]. Khanna, R., & Awad, M. (2015). Efficient Learning Machines : Theories , Concepts , and Applications for Engineers and System Designers. Retrieved on 25/05/2021from https://doi.org/10.1007/978-1-4302-5990-9

[15]. Lawes, C. M. M., Hoorn, S. Vander, Law, M. R., Elliott, P., Macmahon, S., & Rodgers, A. (2000).

High blood pressure- World Health Organization. Retrieved from https://www.who.int/publications/cra/chapters/volume1/0281-0390.pdf on 28/04/2021.

[16]. Li, G., He, Z., Shao, F., Ou, A., & Lin, X. (2015). Patient classification of hypertension in Traditional Chinese Medicine using multi-label learning techniques. 8(3), 4–9). retieved from http://www.biomedcentral.com/1755-8794/8/S3/S4.

[17]. Logan, AG (2011). Hypertension in Aging Patients. Expert Rev Cardiovasc Ther. 2011;9(1):113-120. Reported in Medacpe on 28/6/2022. https://www.medscape.com/viewarticle/734880_1#:~:text=Hypertension%2C%20especially%20isolated%20systolic%20hypertension,treatment%20greatly%20reduces%20cardiovascular%20events.

[18]. Mbbs, P. S. (2019). Machine learning and blood pressure. Retrieved on 24/4/2021). https://doi.org/10.1111/jch.13700

[19]. Nimmala, S., Ramadevi, Y., Sahith, R., &. Cheruku, R. (2018). High blood pressure prediction based on AAA++ using machine-learning algorithms. Cogent Engineering, 5(1), 1–12. Retrieved from https://doi.org/10.1080/23311916.2018.1497114.

[20]. Obe, O. O., Balanica, V., & Neagoe, E. (2015). An ANN-Based Predictive Model for Diagnosis and Forecasting of Hypertension. 9(4), 942–945.

[21]. Scagliarini, L. (2019). What is Machine Learning ? A definition. 2019. Retrieved 12/06/2021 from /https://www.expert.ai/blog/machine-learning-definition/

[22]. Shinde, S. A., & Rajeswari, P. R. (2018). Intelligent health risk prediction systems using machine learning : a review. 7(3), 1019–1023. https://doi.org/10.14419/ijet.v7i3.12654.

[23]. Singh, H. (2018). Understanding Gradien Boosting {Machines}. Medium, 5. Retrieved on 26/05/2021from https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab.

[24]. WHO (2021). What is Hypertension Retrieved on 28/6/2022 at: https://www.who.int/news-room/fact-sheets/detail/hypertension