International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Prediction of Blood Pressure Using Machine Learning

Sirsho Das[1], Subir Kumar Roy[2], Soumyajit Chakraborty[3], Prianka Dey[4,] and Dwaipayan Ghosh[5]

Submitted: 05/05/2024 **Revised**: 18/06/2024 **Accepted**: 25/06/2024

Abstract: In the field of digital health today, Blood pressure prediction is very much crucial. Including lots of health conditions besides hypertension, this model in effect can provide practitioners with early warning red flags for events that are coming. In this study, we took the lead in establishing a blood pressure forecast model with three advanced Machine Learning methods i.e. Support Vector Classifier, Random Forest Classifier, and Naive Bayes Classifier. Our study serves the purpose to construct a blood pressure level forecasting tool, based on clinical data. As part of our research, we have gone through the process of collecting a data set and converted it to digital form. It includes important clinical markers that are closely related to blood pressure, including gender, age, body mass index, smoking status, body mass index, and diastolic and systolic blood pressure. We used clean and compare selection techniques to increase the predictive accuracy of our models without increasing their complexity beyond a usable level. Every algorithm was therefore carefully trained and tested against our collected data (2500 training data, 500 testing data). These trials allowed us to tease out different data points on what makes success in prediction such as predicting blood pressure. This research shows the potential of combined Machine Learning methods to better predict long-term outcomes resulting from hypertension. These predictive models will greatly help healthcare professionals in their early detection and targeted intervention for high blood pressure problems.

Keywords: *Support Vector Classifier, Gaussian Naive Bayes Classifier, Random Forest Classifier, Blood pressure, Confusion Matrix, Regression Model, Hypertension, Machine Learning*

Introduction

The health of the globe is currently significantly threatened by hypertension, high blood pressure, and diabetes, with over a billion people affected worldwide. Hypertension is also a major risk factor for heart disease, stroke, and renal illness. Markedly a significant percentage of both deaths from heart disease and those which are attributable to stroke are connected with hypertension. There are breakthroughs in healthcare but in low and middle income nations where access to healthcare is often less available hypertension is frequently under diagnosed and therefore under treated. Hypertension can play a significant role in increasing heart attacks and coronary heart disease [1]. The absence of distinctive drip from the ceiling, small ponds in hollow objects, or rivers and lakes where there ought to be none makes it hard to timely detect and treat hypertension as any other illness.

Non-invasive blood pressure monitoring devices, like sphygmomanometers, are unable to continuously monitor blood pressure due to the impracticality of constantly inflating and deflating the cuff [2]. Many Machine Learning (ML) techniques relevant to blood pressure measurement systems are being employed in medical technology to overcome difficulties such as these, bringing forth innovative This permits the spread of hypocritical care and nursing based on signs of established hypertension. The process of making clinical decisions has been studied through a variety of conventional and sophisticated machine learning methods, including Regression Models. The integration of these data can improve model accuracy and flexibility across different populations as it has been trained through a huge data set [3]. The Support Vector Classifier (Fig1), Gaussian Naive Bayes Classifier (Fig2) and Random Forest Classifier (Fig3) are the three methods we have employed.

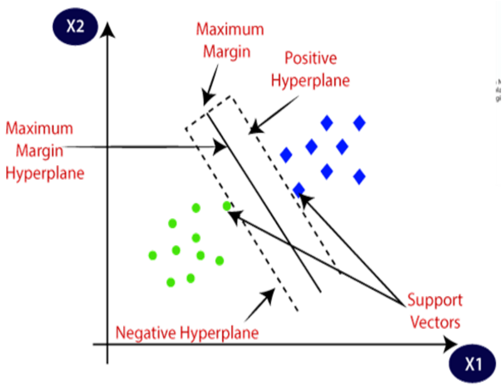

Fig1: An example of Support Vector Classifier

Support Vector Classifier (SVC), is a machine learning algorithm which is used for both classification and regression tasks. It is widely used for classification objectives as it produces significant accuracy with reduced computation power [4]. It employs the principle of separating the classes in a data set using a hyperplane. This machine learning SVC aims to maximize the distance between different types of data points and thus from aims to find a hyperplane that has good generalization exceptional for unseen data. In a word, SVC attempts to

[1*,2,3,4]Narula Institute of Technology, 81, Nilgunj Rd, Jagarata Pally, Deshpriya Nagar, Kolkata, West
Bengal- 700109, India, sirsho12341234@gmail.com, roy74480@gmail.com, chakrabortysoumyajit08@gmail.com,priankadey2011@gmail.com
[5]Kalyani Government Engineering College, Kalyani, Nadia, West Bengal-741235, India, dwaipghosh@gmail.com
*Narula Institute of Technology, 81, Nilgunj Rd, Jagarata Pally, Deshpriya Nagar, Kolkata, WestBengal- 700109, India, sirsho12341234

*Corresponding Author: Sirsho Das

maintain a large margin between different classes by finding an optimal hyperplane. A "large-margin" plane (where only a few points are wrongly classifieds) affords better generalization for unseen data than does a "low-margin" plane (where many points are misclassified). Because they directly affect the position and orientation of the hyperplane, the data points that are closest to it are referred to as support vectors. SVC uses kernel methods to transfer the input space into higher-dimensional space such that a separating hyperplane is achievable. This allows SVC to handle both linearly separable datasets and nonlinear separable sets. When input features and labels have a nonlinear connection, this model is highly helpful in solving challenging classification issues. In comparison to ANNs, the SVM classifier exhibits better generalisation and avoids the problem of local minima [5].

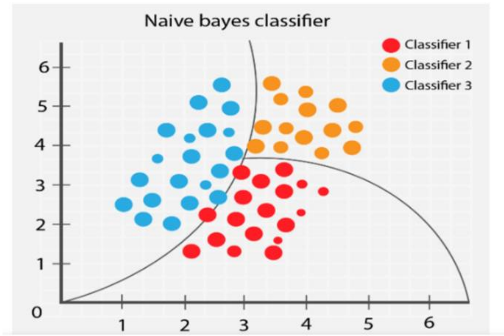

Fig2: An example of Gaussian Naive Bayes Classifier

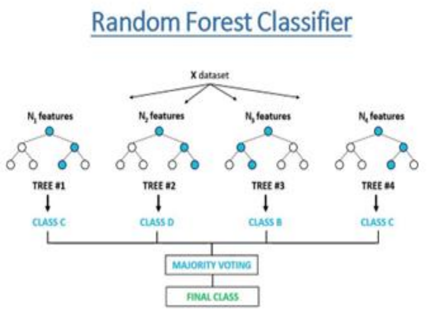

Fig3: An example of Random Forest Classifier

The classification technique under a particular kind of Bayes theorem that presumes the continuous values associated with each feature are normally distributed is the Gaussian Naive Bayes classifier (Fig2). It applies the Bayes theorem under the "naive" assumption of independence between every pair of characteristics, just like all other Naive Bayes classifiers. For most applications, the assumption of class-conditional independence is not true in practice. Even with this reduction, switching between different models can frequently still produce effective outcomes in practice. Thus, Gaussian Naive Bayes performs admirably in a variety of real-world scenarios. The posterior probability of a newly chosen data point under each class is computed. The class with the highest posterior probability is then given the input. Since it explains how continuous features affect class probabilities under the assumption of a normal distribution, the Gaussian model was used for this instance. The Random Forest Classifier, on the other hand, is an ensemble-tree based learning method (Fig3). It is a collection of decision trees drawn at random from a subset of the training set [6]. Ensemble learning is the theory behind it, which combines predictions from numerous machine learning models to get an average value that is more accurate than any one model alone. Random Forest builds multiple decision trees "Merges them together" can get a more stable and accurate prediction. One of Random Forest's primary advantages is its ability to handle overfitting, a common problem in decision trees, by averaging several trees. This approach works particularly well with huge datasets and is quite resilient to noise. It separates nodes using feature randomness and bootstraps to extract data for tree construction. This results in a variety of trees that when combined are found to be a powerful model. Additionally, this aligns with the overarching objective of applying machine learning techniques to provide both targeted and preventive medical treatment benefits as well as a thorough understanding of illnesses like hypertension. Three supervised machine learning methods were used to create hypertension prediction models; these algorithms have not yet been compared. These included the Gaussian Naive Bayes classifier (with cross-validation), the Random Forest classifier, and the Support Vector classifier. Given that our data set has 3000 records, it was only fitting that Random Forest performed the best in our analysis [7].

The rest of the paper has been arranged as follows. In Section 1 a literature survey is given related to the work, in Section 2 detailed methodology has been provided, in Section 3 the results have been given and the paper concluded with the Section 4.

1. Literature Survey

Integration of machine learning (ML) techniques into healthcare over the past few years, it seems that the major research focus has been on predicting hypertension and blood levels. A review of policies related to artificial information in Chinese hospitals reveals various methodologies, algorithmic frameworks and predictive models. This is exemplified by studies and reviews which together clarify the fact that we are making rapid investments on all aspects within the digital health landscape. Yet how can these differing methods be viewed within the same field? The conclusion is that Machine Learning can not only be utilized to diagnose and predict hypertension levels or blood pressure values. What it gains is usefulness for diversified. The further organization into Dual-Earner Mother-Father Families proposes a modern solution to the struggles of handling complex and qualitative productions. This ripe talent for handling multifaceted healthcare data gives ML new commercial opportunities. Deep Learning (DL), a sub field of ML, provides even more sophisticated-and-contextualizable insights from this mine of health data. The synthesis of findings from studies over the years has provided an insight into ML's effect on hypertension. Similarly, using ML algorithms in hypertension screening and prognosis provides a new way to cure this silent killer while the combined involvement of ML in its model together with such important clinical and physiological indicators as sociodemographic characteristics, BMI, or blood sugar makes predictive models more solid and better able to navigate real clinical usage. It opens up the possibility of combining Machine Learning methods with potential which remains unrealized in either back-testing functions or forecasts for high blood pressure. For several years now this has been called “deep learning + big data” and has already led to a large number of real-time interventions in hypertension medicine. When looking at the literature evidence shows that Machine Learning can greatly improve predictive accuracy, providing earlier detection and targeted intervention strategies. We trained the data using the Support Vector Classifier Gaussian Naive Bayes algorithm and Random Forest, then evaluated the algorithm's accuracy for the two classes independently. This allowed us to confirm the number of true and false cases in the training and test data for the two-class data set. Therefore, finding some unique way of exhaustively exploring and intensifying both Machine Learning based models’ treatment potential for high blood pressure is essential to overcoming this problem. And overall health care stages in which patients are involved. Zheng et. al [8] used Artificial Neural Network (ANN), one of the effective method for the prediction of blood pressure using \$250\$ data sample collected from British Hypertension Society and the American National Standard from the Association for the Advancement of Medical Instrumentation. Levenberg-Marquardt, Bayesian Regularisation, and the Scaled Conjugate Gradient technique have all been employed for training purposes.

Using a database of 498 data samples, the authors of [9] used two different neural network architectures radial basis function networks and back propagation neural network to predict systolic blood pressure. In order to forecast our desired outcome, we have combined machine learning algorithms with medical data, even though our study does not use deep learning. The data set undergoes preprocessing with various supervised and unsupervised algorithms to enhance the accuracy of the solution [10]. In [11] used dynamic recurrent neural network for the prediction of blood pressure. In [12] used machine learning for monitoring and prediction of blood pressure. Another new approach has been proposed by [13][14] for the prediction of blood pressure. For the task, they have improved signal processing and machine learning methods.

So, till now we can say that as blood pressure is an important parameter for the finding the health issues in early stages, so it is very much important to get the prior indication about it. So, we have focused to get a different approach to predict the blood pressure. The detailed methodology has been given in the Section 3.

2. Methodology

Blood pressure measurement is a crucial parameter for the early detection of any abnormalities. To note this challenge, we have proposed a methodology for the prediction of the blood pressure. A flow chart of the proposed methodology has been shown in the Fig4.

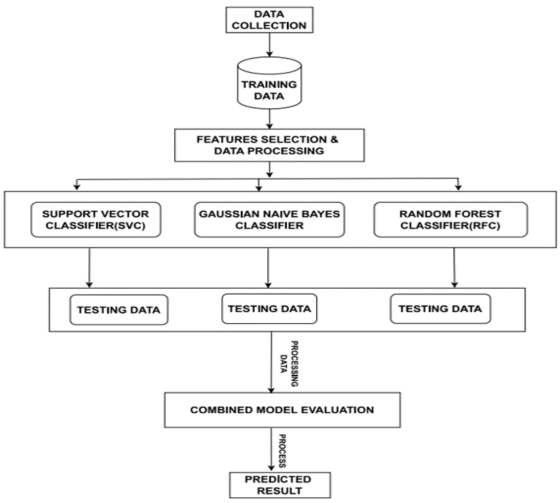

Fig4: Flow chart of the working methodology

2.1 Data Collection

At the very beginning of the machine learning pipeline is the data collection. The necessary data are collected for model training that predicts blood pressure. Encompasses the acquisition of large amounts of data. It is divided into two categories training data (Fig5) and testing data (Fig6) respectively 2500 and 500.

	A	B	C	D	E	F	G	H	I
1	Patient id	Age	Obese	bmi	Smoking Status	Gender	sbp	dbp	Result
2	1	57	1	26.76	1	1	154	87	HIGH
3	2	88	1	16.81	0	0	122	77	NORMAL
4	3	19	0	16.17	1	1	103	63	NORMAL
5	4	68	0	22.52	1	1	176	70	HIGH
6	5	97	0	24.36	0	1	171	74	HIGH
7	6	79	0	26.02	1	1	102	96	HIGH
8	7	87	0	28.87	0	1	151	80	HIGH
9	8	94	1	27.88	1	0	159	61	HIGH
10	9	52	0	28.16	0	0	171	84	HIGH
11	10	38	0	39.98	1	1	134	71	HIGH
	11	40	1	18.56	0	0	142	75	HIGH
13	12	69	0	20.45	0	1	101	74	NORMAL
14	13	73	0	24.17	1	0	155	72	HIGH
15	14	24	0	17.55	0	0	137	79	HIGH
16	15	77	0	24.45	0	0	105	86	HIGH
17	16	66	1	28.36	1	1	146	83	HIGH
18	17	18	0	19.03	1	1	134	96	HIGH
19	18	41	1	16.7	1	1	160	77	HIGH
20	19	20	0	25.92	1	0	118	64	NORMAL
21	20	27	1	27.22	1	1	170	83	HIGH
22	21	69	0	19.51	1	0	171	62	HIGH
23	22	81	1	32	1	1	180	87	HIGH

Fig 5: Sample of Training Dataset

	A	B	C	D	E	F	G	H
1	Patient id	Age	Obese	bmi	Smoking Status	Gender	sbp	dbp
2	1	78	0	15.86	1	0	161	96
3	2	86	0	25.41	0	0	130	72
4	3	86	0	25.83	1	1	108	99
5	4	19	1	24.33	1	1	158	82
6	5	69	1	16.08	1	0	116	95
7	6	45	0	38.29	1	0	150	80
8	7	92	1	16.77	1	1	163	86
9	8	76	1	32.19	0	0	138	62
10	9	21	1	31.41	1	0	130	100
11	10	24	1	16.35	1	0	154	91
12	11	89	0	17.08	0	0	125	71
13	12	86	1	24.99	1	0	166	77
14	13	53	1	38.58	0	0	159	71
15	14	51	1	21.35	1	1	154	72
16	15	73	0	30.28	0	1	128	60
17	16	20	1	29.33	0	1	137	94
18	17	77	0	28.22	1	0	110	93
19	18	30	0	17.39	0	0	137	63
20	19	80	1	36.48	1	1	168	87
21	20	78	1	35.8	1	0	115	66
22	21	51	0	31.04	0	1	102	90
23	22	19	1	38.4	0	0	139	97

Fig 6: Sample of Testing Dataset

2.2 Dataset

Once the data has been gathered, it is compiled to form a training data set. This set serves as a representation of the problem space and entails what will be employed to teach the machine learning models how to make predictions.

2.3 Features Selection and Encoding

The procedure of feature selection and encoding equips the system so that it would identify the data attributes that are most crucial in predicting blood pressure and change these into formats that could be processed relatively fast using learning machines. Considering that X and Y are the variables of which x_i and y_i are the samples of the variables. to extract the features of the variables we have used co relation coefficient r which is given by Eq:1.

$$r = \frac{\sum(x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum(x_i - \bar{x}_i)^2 \sum(y_i - \bar{y}_i)^2}} \qquad (1)$$

2.4 Data Processing

The next step is crucial for a more defined data set and enhancing it prior to model training and validation. Furthermore, this step may involve measuring the scaling features, normalization of the data, and eventual dimensionality reduction to enhance the performance of the model. To meet this goal, we have to scale the data between 0 and 1 by the Eq:2.

$$X_{norm} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \qquad (2)$$

To standardize the data, we have used the Z-score normalization function which is denoted by Eq:3.

$$X_{std} = \frac{(X - \alpha)}{\beta} \qquad (3)$$

where α is the mean of the sample and β is the standard deviation of the sample variables.

2.5 Model Training and Techniques used

Three individual models Support Vector Classification (SVC), Gaussian Naive Bayes, and Random Forest Classifier are initialized and trained on the pre-processed training data set. First, we have used the training data set (Fig5) for training using the Support Vector Classification algorithm given by the Algorithm 1.

Algorithm 1: Support Vector Classification Training Algorithm

Data: Training set ***D*** = *(a, b_1), (a_2, b_2),…, (a_n, b_n)*

Result: Hyperplane parameters ω and h.

Initialize ω and h **repeat**

for $i=1$ to n do

if $y_i(\omega.a_i + h) < 1$ then

$\omega \leftarrow \omega + \eta(b_i a_i - 2\lambda\omega)$;

$b \leftarrow h + \eta b_i$;

else

$\omega \leftarrow \omega - 2\eta\lambda\omega;$

until **convergence**;

After the application of the SVC, we have used Gaussian Naive Bayes classifier on the same training data set. First, we have trained the data set using the Algorithm 2 and then by using the Algorithm 3 we have classified the high blood pressure patient and normal blood pressure patient. The training algorithm provides the prior probability of the sample and using that probability the classification algorithm classifies the attributes using the Gaussian probability density function.

Algorithm 2: Naive Bayes Training Algorithm

Data: Training set $\boldsymbol{D} = (a, b_1), (a_2, b_2), \ldots, (a_n, b_n)$

Result: Class conditional probabilities $P(a|b)$ and class priors $P(b)$.

Calculate class priors : $P(b) = \frac{\text{count}(b)}{n}$ for each class b;

for each feature a_i, **do**

Calculate mean $\alpha_{b,i}$ and standard deviation $\beta_{b,i}$ of a_i of class b;

Algorithm 3: Naive Bayes Classification Algorithm

Data: Instance to classify: $\boldsymbol{A} = (a_1, a_2, \ldots, a_d)$

Result: Predicted class Y.

for each class Y **do**

Calculate $P(Y|a) = P(Y)\prod_{i=1}^{d} P(a_i|Y)$ using Gaussian probability density function with $\alpha_{Y,i}$ and $\beta_{Y,i}$;

$$Y = \arg\max_{Y} P\,(Y|A);$$

Then we have used the random forest classification technique for doing the same using the Algorithm 4 and Algorithm 5 for training and classification respectively.

Algorithm 4: Random Forest Training Algorithm

Data: Training set **D** = $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$, Number of trees **T**, Number of features **F**.

Result: *Random Forest model*.

for $t = 1$ to T

Randomly sample F features without replacement from the set of all features;

Construct a decision tree T_t using the sampled features;

Algorithm 5: Random Forest Classification Algorithm

Data: Instance to classify: **A**, Random Forest model.

Result: Predicted class $\hat{Y}$.

for each tree T_t in the forest **do**

$\hat{Y}_t \leftarrow$ Classification result of A using T_t

$\hat{Y} \leftarrow$ Majority among $\hat{Y}_t$.

After each application of the classifications we need to find the error of the model and for this purpose we have calculated the mean square error (MSE) of the samples for the regression model using the Eq:4

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y_i}) \qquad (4)$$

where N is the total number of samples, y_i is the original value of the sample and $\overline{y_l}$ is the predicted value of i-th sample. Later to this we have augmented all three classifications to get the final accuracy of the prediction.

2.6 Model Evaluation

Each model makes predictions regarding the outcomes on the testing data set using the confusion matrix. The statistical parameters Accuracy and Recall extracted from the confusion matrix helps for the model evaluation for each case. The accuracy has calculated by the Eq:5.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (5)$$

where TP (True Positive) is the correctly predicted true values, TN (True Negative) is the correctly predicted false values, FP (False Positive) is number of wrong true prediction and FN (False Negative) is the number of wrong false prediction. The accuracy of the outcomes can be determined by comparing the predictions made to the true outcomes of the testing data; this way, the ability of the models to accurately predict blood pressure classifications can be quantified. To find the ability of a model to get all relevant cases (all actual positives) Recall has been calculated by the Eq:6

$$recall = \frac{TP}{TP + FN} \qquad (6)$$

2.7 Combined Model Prediction

To check the correctness of the algorithms we have combined SVM, Gaussian Naive Bayes' Classifier and Random forest classifier. All algorithms forecaster outcomes independently for every case in the validation data. Their predictions were then combined by taking the most common prediction for each example. This approach intends to leverage the combined knowledge of the models to generate predictions anticipated to have heightened precision. The accuracy of the combined model is evaluated, and a confusion matrix is created to negotiate the performance of the model in predicting between 'High' and 'Normal' blood pressure occurrences. The testing data is modified using the final predictions.

3. Results and Discussions

A total of 3000 data was collected for our study which shows patients with High and normal both type of blood pressure level. We divided the data into two categories one is our training data and another is our testing data. We have a number of 2500 data for training and 500 data for testing purpose. We have total 2500 instances and 8 gets trained but in the set of 500 testing data we also have 8 attributes given and our model predicts the result this time after getting trained. So, the "Result" is our target attribute here. The other attributes are shown in the Table1.

Table1: Attributes of the data set

Patient Id	Numeric Value
Age	The age of the patient
Obese	If *Yes* then numeric value is 1 if *No* then numeric value is 0.
Smoking status	If *Yes* then 1, if *No* then 0.
Gender	For *female* value is 1, for *male* value is 0.
SBP	Systolic Blood Pressure (Higher than 120 is high).
DBP	Diastolic Blood Pressure (Higher than 80 is high).

In our training data set we have 2005 number of High Blood pressure and 495 number of Normal Blood pressure values. Support Vector Classifier (SVC), Gaussian Naive Bayes, and Random Forest Classifier models were trained and tested using a training data set of 2500 observations and a testing data set of 500 observations. The Support Vector Classifier (SVC) was implemented with the probability estimate enabled, so it could be suitable for both multinomial and binary classification. After that we have combined all three models to check the compatibility of the models and in that case, we get a positive result in terms of accuracy for the prediction. Gaussian Naive Bayes is strong and its good performance in all sorts of applications despite the simplified assumptions. The Random Forest Classifier, whose robustness to noise and good performance on large datasets means it is widely used, was also applied for this purpose. A specific number is used in the random number generation to ensure reproducibility. A detailed result has been shown in the Table 2. The graph for the training data set for final result has been given in the Fig7.

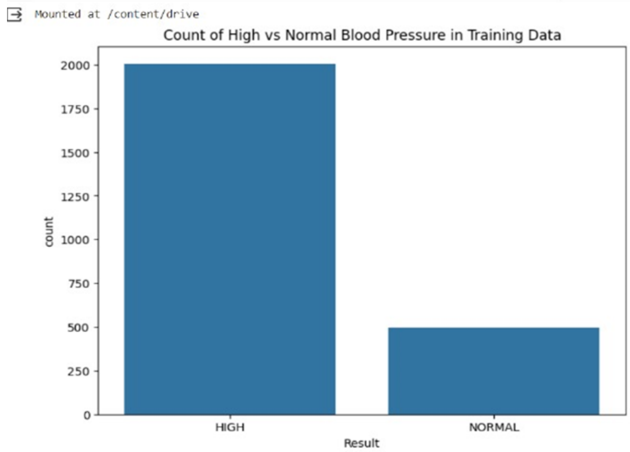

Fig7: Count of High vs Low Blood Pressure for the Training Data

Sl. No.	Model Name	TP	FP	TN	FN	TPR	FPR	Acc	F_1 Score
1	SVC	420	11	6	63	0.87	0.65	0.85%	0.92

2	GNBC	426	7	8	59	0.88	0.47	0.87%	0.93
3	RFC	429	5	12	54	0.88	0.29	0.88%	0.93
4	Combined	419	4	14	50	0.89	0.25	0.86%	0.94

Table 2: Results for all individual models and combined model

```
     Patient id  Age  Obese    bmi  Smoking Status  Gender  sbp  dbp  \
0             1   78      0  15.86               1       0  161   96
1             2   86      0  25.41               0       0  130   72
2             3   86      0  25.83               1       1  108   99
3             4   19      1  24.33               1       1  158   82
4             5   69      1  16.08               1       0  116   95
..          ...  ...    ...    ...             ...     ...  ...  ...
495         496   42      1  15.79               0       0  134   95
496         497   87      1  34.55               1       1  155   77
497         498   75      1  39.41               1       1  104   72
498         499   62      1  36.46               1       1  154  100
499         500   86      0  16.01               1       1  153   90

    SVC_prediction Gaussian_NB_prediction Random_Forest_prediction  \
0             High                   High                     High
1             High                   High                     High
2             High                   High                     High
3             High                   High                     High
4             High                   High                     High
..             ...                    ...                      ...
495           High                   High                     High
496           High                   High                     High
497         Normal                 Normal                   Normal
498           High                   High                     High
499           High                   High                     High
```

Fig8: Classifier Based Prediction

The classifier based prediction is given in the Fig8. The final prediction of the combined model has been shown in the Fig9.

```
    final_prediction
0               High
1               High
2               High
3               High
4               High
..               ...
495             High
496             High
497           Normal
498             High
499             High

[500 rows x 1 columns]
```

Fig9: Combined Model Prediction

4. Conclusion and Discussions

As a result of our study, were able to test the effectiveness of different machine learning (ML) algorithms in predicting blood pressure levels. We paid particular attention to Support Vector Machine (SVM), Random Forest Classifier (RFC), and Naive Bayes (NB) classifiers. Our method of incorporating important clinical indicators improved the prediction accuracy for those models. We kept an appropriate balance between complexity and ease of use. Our results show that the combination of these ML techniques has more solid ability to predict results of hypertension. In high blood pressure cases, whether early intervention or aided diagnosis can occur depends upon the predictability of individual SVM, RFC, and NB classifiers. This performance makes ML an easily available option for early detection and intervention going forward in high blood pressure cases, marking out new territory in digital health domain. Integration of more predictors can be found in other works. For instance, future work could explore integrating lifestyle attributes, environmental factors and stress indicators with predictive factors in that combinations way. This approach may help the model become more accurate and expand its range of application into different population groups. Future research could explore how to use deep learning and other advanced machine-learning techniques to improve the predictive capabilities of these models. AI could also make it helpful for healthcare professionals be able to understand these predictions better and combine physiological data with these algorithms. We could revolutionize the management of hypertension if future models can predict real time blood-pressure from wearable technology. Our analysis makes it possible that in future a machine based method may enable us to predict blood pressure levels more accurately. By using ML and meanwhile continually refining our models according to the future direction research points, detection of hopeful hypertension cases can be enhanced (too much and ultimately) managed. This represents an advantage for the future better health of all mankind.

5. Compliance with Ethical Standards

- **Conflict of Interest**

The authors have no conflicts of interest to declare that are relevant to the content of this article.

- **Financial Support**

The authors did not receive financial support from any organization for the submitted work. This survey paper does not involve any animals or human participants.

- **Authors' Contribution**

Every author has contributed equally in this research.

References

[1] Martinez-Ríos, E., Montesinos, L., Alfaro-Ponce, M., Pecchia, L.: A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. Biomedical Signal Processing and Control 68, 102813 (2021).

[2] González, S., Hsieh, W.-T., Chen, T.P.-C.: A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram. Scientific Data 10(1), 149 (2023).

[3] Qin, K., Huang, W., Zhang, T., Tang, S.: Machine learning and deep learning for blood pressure prediction: a methodological review from multiple perspectives. Artificial Intelligence Review 56(8), 8095–8196 (2023).

[4] Olson, R.S., Cava, W.L., Mustahsan, Z., Varik, A., Moore, J.H.: Data-driven advice for applying machine learning to bioinformatics problems. In: Pacific Symposium on Biocomputing 2018: Proceedings of the Pacific Symposium, pp. 192–203 (2018).

[5] Jackins, V., Vimal, S., Kaliappan, M., Lee, M.Y.: Ai-based smart prediction of clinical disease using random forest classifier and naive bayes. The Journal of Supercomputing 77(5), 5198–5219 (2021).

[6] Singh, M., Martins, L.M., Joanis, P., Mago, V.K.: Building a cardiovascular disease predictive model using structural equation model & fuzzy cognitive map. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1377–1382 (2016).

[7] AlKaabi, L.A., Ahmed, L.S., Al Attiyah, M.F., Abdel-Rahman, M.E.: Predicting hypertension using machine learning: Findings from Qatar Biobank Study. Plos one 15(10), 0240370 (2020). IEEE.

[8] Zheng, J., Yu, Z.: A novel machine learning-based systolic blood pressure predicting

[9] model. Journal of Nanomaterials 2021, 1–8 (2021).

[10] Wu, T.H., Pang, G.K.-H., Kwong, E.W.-Y.: Predicting systolic blood pressure using machine learning. In: 7th International Conference on Information and Automation for Sustainability, pp. 1–6 (2014). IEEE.

[11] Suresh, A., Udendhran, R. and Vimal, S. eds., 2020. Deep neural networks for multimodal imaging and biomedical applications. IGI Global.

[12] Iqbal, T., Elahi, A., Wijns, W., Shahzad, A.: Exploring unsupervised machine learning classification methods for physiological stress detection. Frontiers in Medical Technology 4, 782756 (2022).

[13] Senturk, U., Polat, K., Yucedag, I.: A non-invasive continuous cuffless blood pressure estimation using dynamic recurrent neural networks. Applied Acoustics 170, 107534 (2020).

[14] Arunachalam, S.: Cardiovascular disease prediction model using machine learning algorithms. Int. J. Res. Appl. Sci. Eng. Technol 8, 1006–1019 (2020)1.

[15] Martinez Rios, E.A., et al.: Model for the detection of elevated blood pressure based on machine learning and signal processing technique.