ZKG INTERNATIONAL ISSN: 2366-1313

Machine Learning Algorithms Based Proficient Data-Driven for Hypertension Risk Forecast

[1] Sai manikanta kunasani, [2] M. Rama Bhadra Rao,

[1] MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

saimanikantakunasani0@gmail.com

[2,] Assistant Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280,

Abstract: *Hypertension is a continual condition characterized by means of high stress in the arteries of the human body. As an end result, the coronary heart is compelled to work extra intensively for the normal circulate of blood in the frame. It is one of the maximum crucial risk elements for future deadly and non-cardiovascular illnesses, stroke and kidney failure. In this article, Machine Learning (ML) is used to design powerful models for predicting the lengthy-time period danger of older individuals (over 50 years old) being recognized with hypertension. Our purpose is to teach fashions with high sensitivity in identifying subjects at risk to keep away from the future improvement and occurrence of high blood pressure following the proper interventions. In the context of the followed technique, one-of-a-kind elegance balancing techniques are taken into consideration, underneath which functions ranking is applied, and two ML models (namely, Decision tree and Naive Bayes) are compared based totally on Precision, Recall, F-Measure, Accuracy and Area under Curve (AUC).*

Keywords— Hypertension, blood pressure, prediction, ma- chine learning

I. INTRODUCTION

Blood stress is the strain exerted via the blood on the walls of the arteries and depends at the pulse volume (i.e. how much blood our coronary heart expels in each contraction) and vascular resistance blood glide. Blood pressure is measured with the aid of two numerical signs, one is the systolic pressure, and the alternative is the diastolic stress. Systolic pressure indicates the pressure-tension exerted by the blood at the walls of blood vessels whilst it leaves the heart, at the same time as diastolic pressure expresses the stress exerted by using the blood on the walls of blood vessels when the coronary heart dilates to refill with blood. The devices of pressure are millimetres of mercury (mmHg) [1].

According to the World Health

ISSN: 2366-1313

Organization, the everyday blood stress value of a grownup should be much less than one hundred forty/ninety mmHg. Specifically, the systolic stress need to no longer exceed 140mmHg, and the diastolic should not exceed 90mmHg. These values are the boundaries for the definition of high blood pressure [2].

Hypertension is a sickness of the heart and blood vessels and is without exaggeration an endemic of contemporary society. Its exacerbation is due, on the only hand, to the getting older of the population and, however, to the modern-day behaviour and dispositions of humans [3].

More in particular, according to studies, many elements con- tribute to high blood pressure. A sedentary way of life land lock of bodily exercising result in weight problems. In addition, consuming healthy meals wealthy in salt and fat is a chance thing. Consumption of caffeine and alcoholic liquids will increase the risk of hyper- anxiety. Finally, smoking and stress worsen the circumstance [4]–[6].

In 95% of sufferers, high blood pressure is characterized as idiopathic, because it cannot be attributed to a known pathological cause [7]. When there may be a motive of high blood pressure (illnesses of the kidneys, blood vessels, heart, thyroid, adrenal glands), then we discuss with secondary hypertension [8].

Early diagnosis of high blood pressure is crucial to save you coronary heart assault or stroke as well as harm to organs consisting of the heart, brain, and kidneys [9]. In this path, the technology of medication collaborates with data science. The techniques of artificial intelligence and device studying have a widespread contribution to the development of most desirable prediction models

[10] For numerous illnesses, which include type 2 diabetes [11]–[14]), ldl cholesterol [15],[16],sleep problems[17], CVDs[18],COPD [19], stroke [20] and Covid-19 [21], and so on. Besides preceding illnesses, much research had been performed for hypertension, as a way to be the issue of hobby on this examine.

In [22], the authors present a neural network version so that it will predict hyper tension and attain this with an accuracy of eighty two%. Moreover, in [23], a sequence of Machine Learning prediction fashions concerning AUC, Sensitivity and Specificity are applied, and the stacking ensemble is chosen because the great performer. Besides, the authors in [24] compare k- nearest acquaintances(k-NN),aid vector system(SVM)with radial foundation kernel feature, linear and quadratic discriminate evaluation (LDA), decision bushes (DT), and naive Bayes (NB) classifiers for the arterial high blood

ISSN: 2366-1313

pressure analysis. The LDA achieves the highest classification accuracy. Finally, in [25]. The authors used four classification algorithms (SVM, DT applied by using C4.5 algorithm, random wooded area (RF), and intense gradient boosting) to predict if a participant has high blood pressure or not. The excessive gradient boosting has the best prediction performance with accuracy, F1, and AUC equal to ninety four.36%, 0.875, and zero.927, respectively.

The contemporary paintings analyze the danger factors of hypertension and afford the primary steps of the followed method. Specifically, information sampling techniques (random beneath sampling and oversampling primarily based on Synthetic Minority Oversampling Technique (SMOTE)) are exploited in this examine for balancing class distribution. Decision Trees and Naive Bayes are applied with exclusive overall performance measures to evaluate their predictive capacity. A public dataset has been exploited to validate the fashions' overall performance. In parallel, the same fashions will be assessed as part of the GATEKEEPER [26] mission with pilot statistics.

II THE GATEKEEPER SYSTEM

The important goal of GATEKEEPER is to permit the development of a clever virtual platform that connects health- care companies, companies and elderly residents with a view to sell healthier impartial lives for the ageing population. For this motive, superior Information and Communications Technologies (ICTs) are mixed and carried out. A factor of the gadget is the assessment and incorporation of algorithms from the area of AI and ML to be used as part of its interventions. The development of predictive models targets to early expect a customised risk based totally on facts from the pilots of the undertaking. Hypertension is most of the situations with the intention to be investigated within the GATEKEEPER.

III DATA SET DESCRIPTION

The present studies work exploits a dataset derived from the Kaggle internet site. More especially, the quantity of individuals is 848. Each instance of the statistics set is defined through 13 attributes, which are fed into ML models, and 1 characteristic that represents the target class. The attributes are analyzed as follows:

• Age (years): This variable captures the participant's age concentrated on individuals who are older than 50 years.

◄Gender: This variable indicates the player's gender. The percent of men and women is 51.4% and 48.6%, respectively.

◄BMI (Kg/m2)[27]:This variable denotes the participant's body mass index.

◄Smoking [5]: This variable captures the

ISSN: 2366-1313

smoking behaviour of a player (smoker, non-smoker). 52.7% of individuals are people who smoke.

-Daily steps: This variable captures the number of common each day steps taken by the participant.

-Daily alcohol (ml): This variable captures the participant's common day by day alcohol consumption.

-Daily salt (gr): This variable indicates the participant's average daily salt consumption.

-Stress Level [4]: This variable indicates the player's pressure level, that's captured into three classes (excessive 35.9%, medium 32.4% and occasional 31.7%).

TABLEI
CLASS DISTRIBUTION PER SAMPLING METHOD

	No Sampling	Under sampling	Oversampling
Hyp	394	424	457
Non-Hyp	454	424	454
Total	848	848	911

•CKD [28]: This variable suggests if the player suffers from continual kidney disease or no longer. The CKD occurrence within the dataset is 47.2%.

-Hb (mg/dl) [29]: This variable captures the level of haemoglobin (a protein in red blood cells).

-Adrenal and thyroid disorders (ATD) [30]: This vary in a position suggests if the participator suffers from adrenal and thyroiddisorders.ATD'sprevalenceinthedat asetis43%.

-SBP (mmHg) [31]: It is the variable that captures the systolic blood strain.

-DBP (mmHg) [31]: This variable captures the diastolic blood stress.

-Hypertension: This variable suggests whether or not a participator is hypertensive or now not. In the subsequent, the notation Hyp will discuss with the high blood pressure elegance. 46.4% of members have high blood pressure.

All variables are numeric aside from Gender, Smoking, Stress Level, CKD, ATD and Hypertension that are nominal.

IV METHODOLOGY

The followed technique consists of the following steps:

- Class Balancing
- Features Ranking
- Design of Classification Framework
- Models Evaluation

These ranges might be analytically offered inside the approaching sections.

A. Class Balancing

There are various strategies to tackle the trouble of non- uniform magnificence distribution. In this study, we are able to

ISSN: 2366-1313

cognizance on well-known sampling tactics [32].

First, random below sampling is applied to target the general public elegance by way of randomly putting off times until to obtain the favoured balance (or the instances are identical) in each lessons. Also, in this take a look at, SMOTE [33] turned into accomplished to growth the statistics of the minority elegance via sixteen%. It is an oversampling technique that increases the facts through growing synthetic data [34] on minority magnificence using 5-NN classifier on the equal features. It is used to alternate-off between precision and don't forget or increase bear in mind on the price of precision. Class distribution is recorded in Table I, assuming no, below and over sampling cases. Also, the effect of sampling methods at the members' distribution in step with age organization and gender is depicted in Figures 1 and a pair of.

B. **Features Importance**

Features ranking will assist us correctly constitute every file, that specialize in the ones so that it will supply greater records approximately

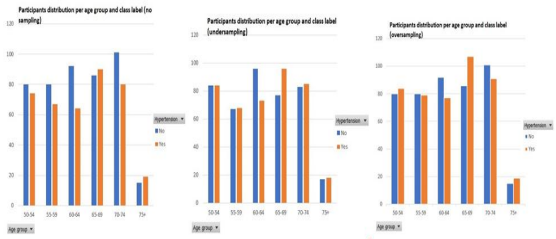

Fig.1.Participants distributionperclassandagegroup (nosampling,undersampling,oversampling).

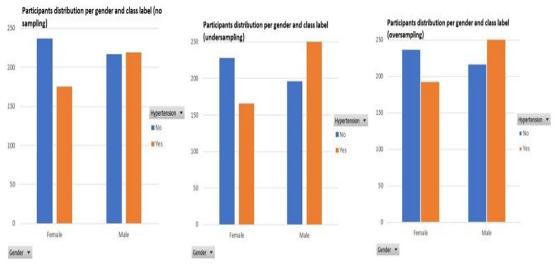

Fig.2.Participantsdistributionperclassandgender (nosampling, undersampling,oversampling).

The target class. For this purpose, Random Forest [35] changed into used to apply feature ranking. This technique measures a function's rank based on Gini impurity [36]. Table II shows the function's significance in three one of a kind sampling cases.

In the original facts, the rank reassesses crucial capabilities the SBP, DBP, CKD, ATD and Hb. Smoking and strain ranges are ranked with 0 significance. The rest capabilities are of negative significance. These out comes suggest that these functions don't make contributions to the goal magnificence. However, salt and alcohol intake are vital threat factors for the incidence of Hypertension. Actually, a food regimen with excessive salt and alcohol use impacts blood stress and, hence, they're crucial inside the control of hypertension (remedy and manipulate)

ISSN: 2366-1313

[37].The capabilities' importance is differentiated between the sampling methods and the no sampling case. More in particular, after underneath sampling the preliminary dataset, day by day salt intake is fourth so as, and not one of the function shads a bad ranking. In over sampling case, the salt significance is fine but lower than the one in the under sampling case. Also, best pressure and alcohol are of poor importance. At this point, we would love to observe that each one capability could be considered to teach and test the ML models.

TABLEII
FEATURES IMPORTANCE BASED ON RANDOM FOREST AS RANKING METHOD

No Sampling		Under sampling		Oversampling	
Features	**Rank**	**Features**	**Rank**	**Features**	**Rank**
SBP	0.3148	SBP	0.3847	SBP	0.3532
DBP	0.2394	DBP	0.3106	DBP	0.2804
CKD	0.1804	Hb	0.2708	CKD	0.2206
ATD	0.1285	Daily salt	0.2528	ATD	0.1526
Hb	0.0771	Daily steps	0.2481	Hb	0.1115
Stress	0	CKD	0.1969	Gender	0.0483
Smoking	0	ATD	0.1580	Age	0.03535
Gender	-0.0035	Alcohol	0.1392	Smoking	0.03293
Age	-0.0071	Gender	0.0743	Daily steps	0.0206
Daily steps	-0.0230	BMI	0.0666	Daily salt	0.0143
Daily salt	-0.0307	Age	0.0271	BMI	0.0038
BMI	-0.0443	Stress	0.0212	Stress	-0.0018
Alcohol	-0.0856	Smoking	0.0165	Alcohol	-0.0555

AUC to ensure that new subjects can be correctly classified. For this purpose, we evaluate the prediction performance of Decisions Trees (DT) [38] and Naive Bayes classification methods.

It must be stated that J48 and Naive Bayes have the identical recall of ninety.2%, but the advantage in Naive Bayes is higher than J48whencomparedtonosampling.Also, there call of J48 inside the Hyp magnificence has been favoured through the oversampled information without worsening the bear in mind of Non-Hyp subjects and the precision of each. In Naive Bayes, there is a 4.4% boom inside the bear in mind of the Hyp magnificence accompanied by a 0.Eight% growth within the precision of the equal elegance. Beside precision and do not forget, a combinatory metric, F-Measure, has been recorded. This metric shows that each model are greater green if they're skilled with below sampled data.

Moving on to the ROC values, the AUC of J48 is the identical in both lessons. A comparable trend is discovered in Naive Bayes, which yields a better AUC. Under sampling is the winner technique. In either case, both models reap values very close to 1. AUC reveals the models' functionality to discriminate the hyper tensile from non hypertensive subjects. It is apparent that, within the underneath sampling case, Naive Bayes achieves this with a excessive chance of 96.7%. Comparable overall performance is succeeded by using J48, with an AUC of ninety two.8%.

The accuracy of J48 and Naive Bayes is illustrated in Figure

TABLEIX
RECALL OF NAÏVE BAYES

Naïve Bayes	Recall		
	No Sampling	**Under sampling**	**Oversampling**
Non-Hyp	0.941	0.943	0.938
Hyp	0.858	0.887	0.902
	0.902	0.915	0.934

ISSN: 2366-1313

TABLEX
F-MEASURE OF NAÏVE BAYES

NaiveBayes	F-Measure		
	No Sampling	Under sampling	Oversampling
Non-Hyp	0.911	0.917	0.908
Hyp	0.891	0.913	0.903
Average	0.902	0.915	0.906

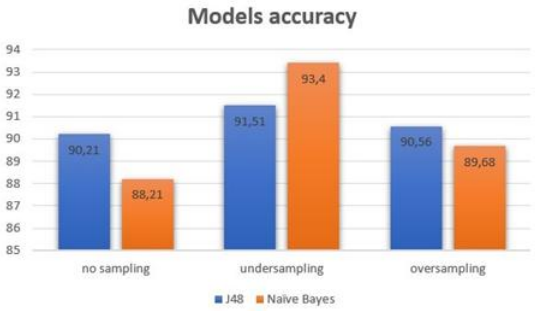

Fig.3.J48andNaiveBayesaccuracyintermsofthesamplingmethod

3. It is another metric that captures the total classification performance in both classes. Finally, taking into account the accuracy of the models with the above metrics, we conclude that in the specific data, the under sampling is more suitable for the design of the desired models.

V CONCLUSION

In conclusion, a publicly available dataset became employed to research the significance of diverse threat elements for Hyper- anxiety. Then, these elements were considered to quantify the chance of hypertension incidence, focused on people older than50 years. We consciousness on this age organization and in particular those living at home to improve their quality of existence via AI gear for personalised interventions. A framework with records-pushed methods is suggested, and the position of sophistication balancing (particularly, random beneath sampling and oversampling) in function ranking and ML techniques overall performance became investigated. The effects of both techniques provided excessive take into account and AUC, which show that the fashions have high discrimination capability in identifying new topics with Hypertension. Also, the accuracy and F-Measure screen the general class performance of the fashions. In destiny work, we purpose to experiment with greater fashions as unmarried classifiers (like SVM, Random Forest, Logistic Regression, and Neural Networks) or observe ensemble studying strategies which include Stacking. Finally, our cause is to emphasise class balancing and observe hybrid sampling techniques earlier than the assessment of the ML fashions.

REFERENCES

1. M.Brunstro¨mandB.Carlberg,“Asso ciationofbloodpressureloweringwit h mortality and cardiovascular disease across blood pressure levels: a systematic review and meta-analysis,” *JAMA internal medicine*, vol.178, no. 1, pp. 28–36, 2018.
2. “World health organization: Hypertension,” https://www.who.int/news-

room/fact-sheets/detail/hypertension,(accessedon1stJune2022).

3. N. M. Kaplan, *Kaplan's clinical hypertension*. Lippincott Williams &Wilkins, 2010.
4. S. Kulkarni, I. O'Farrell, M. Erasi, and M. Kochar, "Stress and hypertension." *WMJ: official publication of the State Medical Society of Wisconsin in*, vol. 97, no. 11, pp. 34–38, 1998.
5. A.Virdis, C.Giannarelli, M.Fritsch Neves, S.Taddei,and L.Ghiadoni, "Cigarette smoking and hypertension," *Current pharmaceutical design*, vol. 16, no. 23, pp. 2518–2525, 2010.
6. E.S.FordandR.S.Cooper,"Riskfactorsforhypertensioninanationalcohort study." *Hypertension*, vol. 18, no. 5, pp. 598–606, 1991.
7. M. Wall, "Idiopathic intracranial hypertension," *Neurologic clinics*, vol. 28, no. 3, pp. 593–617, 2010.
8. Prasadu Peddi (2016), Comparative study on cloud optimized resource and prediction using machine learning algorithm, ISSN: 2455-6300, volume 1, issue 3, pp: 88-94.
9. J. R. Chiong, W. S. Aronow, I. A. Khan, C. K. Nair, K. Vijayaraghavan, R. A. Dart, T. R. Behren beck, and S. A. Geraci, "Secondary hypertension: current diagnosis and treatment," *International journal of cardiology*, vol. 124, no. 1, pp. 6–21, 2008.
10. S. Gulec, "Early diagnosis saves lives: focus on patients with hypertension," *Kidney international supplements*, vol. 3, no. 4, pp. 332–334, 2013.
10 M.Nilashi,O.binIbrahim,H.Ahmadi, andL.Shahmoradi,"Ananalyt-ical method for diseases prediction using machine learning techniques, *"Computers & Chemical Engineering*, vol. 106, pp. 212–223, 2017.
11. S. Alexiou, E. Dritsas, O. Kocsis, K. Moustakas, and N. Fakotakis, "Anapproachforpersonalizedcontinuousglucosepredictionwithregressiontrees,"in*20216thSouth-EastEuropeDesignAutomation,ComputerEn-gineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*.IEEE, 2021, pp. 1–6.
12. E. Dritsas, S. Alexiou, I. Konstantoulas, and K. Moustakas, "Short term glucose prediction based on oral glucose to clearance test values, "in *International Joint Conference on Biomedical*

ISSN: 2366-1313

Engineering Systems and Technologies - HEALTHINF, vol. 5, 2022, pp. 249–255.

13. Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.