Lampyrid 2023: Volume 13, 299–305
ISSN: 2041-4900
https://lampyridjournal.com

Machine Learning-Based Performance Analysis for Predicting the Severity of Vitamin D Deficiency

Dr. T. K. S. Rathish Babu[1], M. Nikitha[2], K. Pranavi[3]
[1]Professor, Computer Science & Engineering, Sridevi Women's Engineering College, Hyderabad, India
[2]Computer Science & Engineering, Sridevi Women's Engineering College, B.Tech IV Year, Hyderabad, India
Email id: nikithareddymaram2002@gmail.com
[3]Computer Science & Engineering, Sridevi Women's Engineering College, B.Tech IV Year, Hyderabad, India
Email id: pranavikushana906@gmail.com

ABSTRACT:

There is a squeezing need for harmless strategies to expect the seriousness of vitamin D deficiency (VDD), which is a significant worldwide medical condition. The essential information, which remembered Vitamin D levels for the blood, were assembled from 3044 undergrads between the ages of 18 and 21. VDD was anticipated utilizing age, orientation, level, weight, body mass index (BMI), midriff outline, muscle to fat ratio, bone mass, action, daylight openness, and milk utilization. The target of the review is to look at and examine different ML models for anticipating VDD seriousness. The objectives of our procedure are to estimate utilizing an assortment of refined ML calculations and to evaluate the results utilizing different execution measures, for example, Accuracy, Review, F1-measure, Exactness, and Region under the bend of a beneficiary working trademark a (ROC). The empirical data were checked with the statistical McNemar's test. The last trial results showed that the Arbitrary Timberland Classifier beat the preparation and testing Vitamin D datasets with an exactness of 96%. McNemar's factual tests exhibit that the RF classifier performs better compared to different classifiers.

Keywords: Machine learning algorithms, random forest classifier, the severity of VDD, vitamin D

1. INTRODUCTION

Vitamin D is a fundamental nutrient that essentially affects various body frameworks. Around one billion individuals overall had extreme lack of vitamin D [1]. The absence of vitamin D has been associated with different auto, safe afflictions, including cardiovascular contamination, type 2 diabetes, and chest threatening development [2]. Despite the fact that the clinical field gathers a ton of information consistently, it will be challenging to dissect huge datasets utilizing conventional techniques. Late exploration has shown that utilizing ML models would deliver improved results [5]. New etiologic examples will be found utilizing ML models, which will make it conceivable to take powerful protection general wellbeing measures [6]. Using ML models for VDD determination will set aside cash and lead to better treatment. ML procedures were not used for seriousness expectation in past examination, which rather centered around the results of measurable models. The seriousness of VDD can be anticipated utilizing conventional factual models like LR, yet their exhibition is restricted because of their prescient presentation limit and the enormous number of elements. The appraisal of Vitamin D status is by and by extravagant, and it is settled through biochemical procedures. Consolidating the tedious logical methodology used to identify VDD in patients is expected to fill the distinguished exploration hole. Subsequently, we utilized an assortment of ML models to foresee the seriousness of VDD in patients. On more modest nutrient datasets, measurable models were utilized in past examinations [9]. Execution might endure when standard techniques are applied to bigger datasets. This is, as far as anyone is concerned, the principal study to utilize different ML models to explore the seriousness of VDD. Besides, the etiologic of different geological areas will contrast because of fluctuating environment conditions. Second, to figure out which ML classifier is the best at foreseeing the seriousness of lack of vitamin D by contrasting their outcomes with an assortment of execution measurements like Accuracy, Recall, the F1-measure, and ROC bends [10], as well concerning an assortment of mistake measurements like Cohen's Kappa and connection coefficient. The essential target of the review is to utilize ML classifiers to foresee the seriousness of VDD and to check the observational discoveries with factual importance tests and mistake estimation. A definitive goal is finding the best ML classifiers for foreseeing VDD seriousness.

2. LITERATURE SURVEY

"On the Scalability of Machine-Learning Algorithms for Breast Cancer Prediction in Big Data Context"

A surge of information has come about because of late mechanical headway, introducing another time of enormous information. Unfortunately, the original properties of large information render conventional ML calculations unequipped for taking care of them. Large information and bosom malignant growth expectation are the focal point of this review. We analysed two sorts of information: DNA methylation (DM) and gene expression (GE). Utilizing each dataset independently and by and large, this review means to increase arrangement ML calculations. Therefore, we picked Apache Flash as our foundation. Nine models that can anticipate bosom malignant growth were made involving three unmistakable arrangement techniques in this review: random forest, decision tree, and support vector machine (SVM). We did an extensive correlation study utilizing three situations utilizing GE, DM, and GE and DM consolidated to learn which of the three information types would convey the most elevated precision and mistake rate. Besides, we drove a preliminary connection of two phases (Flash and Weka) to display their approach to acting while simultaneously overseeing the enormous proportions of data. Utilizing the GE dataset, the exploratory outcomes exhibited that the scaled SVM classifier in the Flash climate beats different classifiers as far as exactness and blunder rate.

"A novel approach for prediction of vitamin d status using support vector regression"

Foundation: Various ongoing sicknesses are connected to lack of vitamin D, as per epidemiological examination. Nonetheless, the recognized biomarker of vitamin D status, blood 25-hydroxyvitamin D focus, may not be imaginable to gauge in huge epidemiological examinations straightforwardly. An alternate methodology is to utilize a forecast model in view of the qualities of the poll information to survey vitamin D status. Multiple linear regression (MLR) models just made sense of a little piece of the variety in past examinations, and projected values were just marginally corresponded with noticed values.

“Performance of Statistical Models to Predict Vitamin D Levels”

Vitamin D testing has become exceptionally well known as of late. Both the patient expenses and general wellbeing uses have expanded because of this inclination. Patients could quit paying for blood tests by utilizing the vitamin D expectation choice. We investigated various measurable techniques for essentially anticipating vitamin D levels in light old enough, orientation, and organic elements. The members were a gathering of hospitalized patients from various divisions at the College Medical clinic Focus of Oujda who had legitimate biochemical and vitamin D qualities. There were 17 variables and 124 patients with ages going from 9-87 (mean = 45.19, median = 49 Various physiological markers, including calcium and glucose, had frail associations with vitamin D, as per experimental outcomes. On account of a little data set, the SVR model performed better compared to irregular woods and MARS in these tests. The people most at danger for lack of vitamin D might be related to the help of this expectation.

“An assessment of the risk factors for vitamin D deficiency using a decision tree model”

The Targets and foundation: Vitamin D (25-hydroxyvitamin D or 25OHD) expects a critical part in the improvement of different consistent sicknesses. Lack of vitamin D is a worldwide general medical condition that influences many individuals. We endeavoured to survey the gamble factors related with lack of vitamin D by utilizing a choice tree approach. Techniques: 988 high school young ladies between the ages of 12 and 18 were remembered for the review. All blood count boundaries, serum biochemical factors, socioeconomics, and minor components like zinc, copper, calcium, and Turf were checked out. Vitamin D insufficiency was portrayed as serum levels under 20ng/ml. A decision tree development preparing dataset contained 70% of these females (618 examples). The choice tree's exhibition was assessed utilizing the excess 30% (285 occurrences) as the testing dataset. There are 14 info factors in this model: age, father's scholastic standing, midriff outline, hip to midsection proportion, zinc, copper, calcium, SOD, FBG, HDL-C, RBC, MCV, MCHC, and HCT The improvement of a receiver operating characteristic (ROC) bend was utilized to survey the legitimacy of the model.

3. METHODOLOGY

Tamune utilized an irregular timberland classifier to foresee lack of vitamin B in individuals with mental side effects. Awareness, particularity, and an area under the curve (AUC) were utilized to assess the model. Carlberg and co expressed that VDD was deduced utilizing both regulated and solo ML methods. The irregular backwoods tree was generally utilized for arranging steady information. The random forest classifier incorporates data about the natural framework being scrutinized. Multivariate and time series investigation were utilized to quantify patients' vitamin D levels. Natural elements incorporate age, calcium, chlorine, LDL cholesterol, and HDL cholesterol. It is possible to take measures that work. The utilization of ML models to analyse vitamin D and C insufficiency will be savvy for further developed treatment. The appraisal of Vitamin D status is by and by over the top expensive, and it is settled through biochemical strategies. The recognized exploration hole calls for shortening the tedious logical methods used to recognize patients lacking Vitamin D and C. To foresee the seriousness of VDD in patients, we utilized an assortment of AI models. Strategies for deciding the seriousness of vitamin D and L-ascorbic acid insufficiency, inadequacy, and adequacy are currently better perceived.

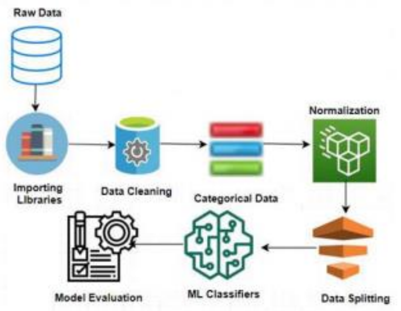


Fig 1: Architecture

IMPLEMENTATION

The following algorithms were utilized in this implementation;

AdaBoost Classifier: An ensemble arrangement for the machine learning popular as the Ada Boosting Classifier connects feeble beginner models to produce a forceful trainee. A base beginner is a machine learning plan that sustains weights from preparation data. Using sklearn, we established the AdaBoost Classifier. The development set will be chosen incidentally from the beginning, and the model will arrange it in addition previously. In the after-redundancy step, misclassified notes are likely more burden and a taller tendency. Until the preparation set data sets in the model outside the mistake, this plan will happen again.

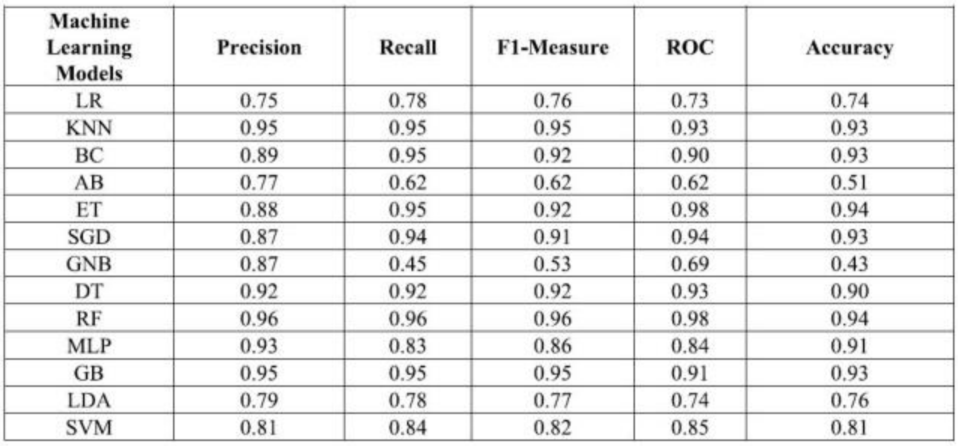
Extra Trees Classifier: The Extra Trees Classier (ET) is a gathering ML plan that utilizes a randomized meta-assessor on a different example of the readiness dataset to control overfitting and advance visualization veracity. The Extra Trees more tasteful is executed via sklearn. Ensemble. The start operating system will be set to fake by default in consideration of produce numerous trees. Instead of utilizing dossier bootstrapping, the unpredictability of the preparation datasets will be got by carelessly dividing bureaucracy into subsets.

Decision Tree: Utilizing a tree-like model or diagrams, the Decision Tree Classifier is a recognizable coordinated ML plan for classification issues. The DT power salvage dynamic dossier from the current news. We handed down DT going with the sklearn.shrub importance Decision Tree Classifier for the investigation. The course from the root place to the leaf community tends to the depiction rules. In our wellspring of sustenance D deficiency centrality affecting, all middle in the wood calculates the deficiency risk, and each arm tends to the evolving's states. The Vitamin D dataset has four deficiency risk types as the outcome and bountiful autonomous determinants, (c, T) D (c_1,c_2,c_3,c_4….T), place T is the deficiency peril alterable and the heading c is added up to of any free determinants took advantage of for classification, e.g., c_1,c_2,c_3….$c_{n.}$

Random Forest Classifier: An ensemble machine learning resolution for handling classification issues is Breiman's hint of random forest classier (RF). RF produces miscellaneous choice seedlings inarticulately from the development set, before, before, consolidates the values from the unique choice saplings and estimates the consequence as final danger deficiency. Standards, least example split (splitD2), and least example leaf (leafD1) are as far as possible decided for RF. The RF classifier outflanks the extra classifiers, as prior represent. We second-hand RF accompanying the sklearn.ensemble significance Random Forest Classifier for the experiment.

4. RESULTS AND DISCUSSION

For an assortment of ML models, McNamar's test is a matched speculative test that is applied to both the training and testing sets. The exact consequence of the RF classifier (accuracyD96.40) and the p-esteem (p0.035) were affirmed utilizing a measurable speculation test. Consequently, the empirical findings were validated using the statistical test. The exploratory revelations and quantifiable hypothesis test show that the RF is the best classifier for expecting reality in VDD. The measurable speculation test supports the correlation and determination of the best classifiers.



Machine Learning Models	Precision	Recall	F1-Measure	ROC	Accuracy
LR	0.75	0.78	0.76	0.73	0.74
KNN	0.95	0.95	0.95	0.93	0.93
BC	0.89	0.95	0.92	0.90	0.93
AB	0.77	0.62	0.62	0.62	0.51
ET	0.88	0.95	0.92	0.98	0.94
SGD	0.87	0.94	0.91	0.94	0.93
GNB	0.87	0.45	0.53	0.69	0.43
DT	0.92	0.92	0.92	0.93	0.90
RF	0.96	0.96	0.96	0.98	0.94
MLP	0.93	0.83	0.86	0.84	0.91
GB	0.95	0.95	0.95	0.91	0.93
LDA	0.79	0.78	0.77	0.74	0.76
SVM	0.81	0.84	0.82	0.85	0.81

Fig 2: accuracy, precision, recall and F1 score values for machine learning algorithms

It is utilized to decide if the consequences of different AI models are genuinely substantial. It is feasible to decipher the speculative test's p-esteem as p>alpha, which demonstrates that there is no distinction and doesn't dismiss H0. P alpha, then again, rejects H0, demonstrating a massive distinction. Using McNamar's test, the VDD reality conjecture is shown to be really not exactly equivalent to the ground truth when the p-regard is under 0.05. Utilizing a continency table with section upsides of Yes/No and negative/Indeed, the McNamar's model tracks down classifier mistakes. The consequences of this test will presumably show on the off chance that there is a tremendous contrast to the quantity of these cells. Accordingly, the invalid speculation isn't dismissed assuming the counts are something very similar and the two classifiers have a similar blunder rate. Since classifiers have a comparative level of mistakes on the test set and an alternate rate, we can characterize the consequences of this measurable test. We can sort the outcomes of this verifiable test since, assuming that the cell count isn't inside an identical degree of misstep, the invalid hypothesis is excused.

5. CONCLUSION

The best machine learning (ML) the model for anticipating VDD seriousness is the essential goal of this review. The forecast's exactness was determined and diverged from the preparation and testing sets. With the end goal of this examination, we utilized eleven ML models and execution pointers like accuracy, recall, the F1-measure, and precision. 11 boundaries and the RFE technique were utilized to choose highlights for the seriousness expectation. McNamar's measurable importance test affirms the observational discoveries. According to McNamar's test, RF clearly beats various models in conjecture, and Pearson's association coefficient and botch assessments support this end. It is feasible to substitute techniques in light of ML for the powerful and exact expectation of VDD seriousness. This review's discoveries showed that the arbitrary wood's classifier and other ML models, specifically, had the option to foresee the seriousness of vitamin D and L-ascorbic acid insufficiency precisely. The Random Forest classifier, specifically, beat different classifiers and accomplished the most elevated accuracy (96%). This ML classifier will have a more grounded plausibility as a general rule clinical district, helping specialists in rapidly choosing the reality of VDD. The essential advantage of this study is that it appropriately assessed the results of ML models utilizing different execution markers among teens and explored a clever procedure for foreseeing VDD seriousness utilizing the Random Forest model. Thus, the review affirms that the Random Forest model is preferred capable over different models to foresee the seriousness of VDD. The model will be endorsed with various types of Vitamin D and C datasets from all age bundles from this point forward.

6. REFERENCES

[1] M. Holick, ''Vitamin D deficiency,'' New England J. Med., vol. 357, no. 3, pp. 266–281, 2007.

[2] I. R. Reid and M. J. Bolland, ''Role of vitamin D deficiency in cardiovascular disease,'' Heart, vol. 98, no. 8, pp. 609–614, Apr. 2012.

[3] B. Schöttker, C. Herder, D. Rothenbacher, L. Perna, H. Müller, and H. Brenner, ''Serum 25-hydroxyvitamin D levels and incident diabetes mellitus type 2: A competing risk analysis in a large population-based cohort of older adults,'' Eur. J. Epidemiol., vol. 28, no. 3, pp. 267–275, Mar. 2013.

[4] S. B. Mohr, E. D. Gorham, J. E. Alcaraz, C. J. Kane, C. A. Macera, J. K. Parsons, D. L. Wingard, and C. F. Garland, ''Serum 25-hydroxyvitamin D and prevention of breast cancer: Pooled analysis,'' Anticancer Res., vol. 31, no. 9, pp. 2939–2948, 2011.

[5] Y. Lee, R.-M. Ragguett, R. B. Mansur, J. J. Boutilier, Z. Pan, D. Fus, J. D. Rosenblat, A. Trevizol, E. Brietzke, K. Lin, M. Subramaniapillai, T. C. Y. Chan, C. Park, N. Musial, H. Zuckerman, V. C.-H. Chen, R. Ho, C. Rong, and R. S. McIntyre, ''Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review,'' J. Affect. Disorders, vol. 241, pp. 519–532, Dec. 2018.

[6] S. Alghunaim and H. H. Al-Baity, ''On the scalability of machine-learning algorithms for breast cancer prediction in big data context,'' IEEE Access, vol. 7, pp. 91535–91546, 2019.

[7] S. Guo, R. Lucas, and A. Ponsonby, ''A novel approach for prediction of vitamin D status using support vector regression,'' PLoS ONE, vol. 8, no. 11, Nov. 2013, Art. no. e79970.

[8] S. Bechrouri, A. Monir, H. Mraoui, E. H. Sebbar, E. Saalaoui, and M. Choukri, ''Performance of statistical models to predict vitamin D levels,'' in Proc. New Challenges Data Sci., Acts 2nd Conf. Moroccan Classification Soc. ZZZ (SMC), New York, NY, USA, 2019, pp. 1–4.

[9] K. Gonoodi, M. Tayefi, M. Saberi-Karimian, A. A. Zadeh, S. Darroudi, S. K. Farahmand, Z. Abasalti, A. Moslem, M. Nematy, G. A. Ferns, S. Eslami, and M. G. Mobarhan, ''An assessment of the risk factors for vitamin D deficiency using a decision tree model,'' Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 13, no. 3, pp. 1773–1777, May 2019.

[10] J.-J. Beunza, E. Puertas, E. García-Ovejero, G. Villalba, E. Condes, G. Koleva, C. Hurtado, and M. F. Landecho, ‘‘Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease),’’ J. Biomed. Informat., vol. 97, Sep. 2019, Art. no. 103257.

[11] T.K.S Rathish Babu et al., “MLPNN-RF: Software Fault Prediction based on Robust weight optimization and Jacobian Adaptive Neural Network”, ‘Concurrency and Computation Practice and Experience”,DOI:10.1002/cpe.7122ISNN:1532-0634(2021).