As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 10, Issue 04, pages 397-404.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol10issue04.html
DOI: https://doi.org/10.46501/IJMTST1004062

Identifying Vitamin D Deficiency Severity Analysis using Machine Learning Classifiers

Dr. D J Samatha Naidu | K.Venkata Ramya | R.Jayasri

[1]Assistant Professor, Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India.
[2]Department of Computer Science Engineering, Pragati Engineering College, Surampalem , Andhra Pradesh, India.

To Cite this Article
Dr. D J Samatha Naidu, K.Venkata Ramya and R.Jayasri, Identifying Vitamin D Deficiency Severity Analysis using Machine Learning Classifiers, International Journal for Modern Trends in Science and Technology, 2024, 10(04), pages. 397-404. https://doi.org/10.46501/IJMTST1004062

Article Info
Received: 06 April 2024; Accepted: 25 April 2024; Published: 27 April 2024.

ABSTRACT

Vitamin D is an essential vitamin that has powerful influence on several parts of the human body. Vitamin D Deficiency is associated with several auto immune disorders like cardiovascular disease, diabetes mellitus and breast cancer. In existing work, the previous studies the results were compared between the statistical models and they have not used the machine learning algorithms for the severity prediction. The traditional statistical model like LR is used to predict the severity of VDD but its performance is deprived due to its predictive performance limit and many parameters. Currently, the analysis of Vitamin D status is highly expensive, and it is identified using the biochemical methods. The research gap identified urges to condense the cumbersome analytical procedures in identifying VDD among the patients. So, for predicting the severity of VDD among the patients, we have used various machine learning models. The present study throws light on the way of identification and categorization of severity of VDD, Insufficiency and sufficiency. The proposed work, to predict the severity of VDD datasets by using various types of machine learning models like Linear Regression, k-nearest neighbor, Gaussian Naïve Bayes, Decision Tree, Random Forest classifier, Multi-layer Perception, AdaBoost Classifier, Stochastic Gradient Classifier, Boosting Classifier, Linear Discriminate Analysis, Support Vector Machine, and Gradient Boosting classifier. Secondly, to compare the results of machine learning models with various performance measures like Precision, Recall, F1-measure and ROC curves in predicting Vitamin D deficiency severity as well as with different error measures, Cohen's Kappa and correlation coefficient to identify the best machine learning classifier in the prediction severity of VDD.

Keywords: Cardiovascular diseases, Machine learning, Severity, Accuracy, Deficiency, Statistical methods

1. INTRODUCTION

Vitamin D Deficiency (VDD) is one of the most significant global health problem and there is a strong demand for the prediction of its severity using non-invasive methods. The primary data containing serum Vitamin D levels were collected from a total of 3044 college students between 18-21 years of age. The independent parameters like age, sex, weight, height, body mass index (BMI), waist circumference, body fat, bone mass, exercise, sunlight exposure, and milk consumption were used for prediction of VDD. The study aims to compare and evaluate different machine learning models in the prediction of severity in VDD. The objectives of our approach are to apply various powerful machine learning algorithms in prediction and evaluate the results with different performance measures like Precision, Recall, F1-measure, Accuracy, and Area under the curve of receiver operating characteristic (ROC).

The McNemar's test was conducted to validate the empirical results which is a statistical test. The final objective is to identify the best machine learning classifier in the prediction of the severity of VDD. The most popular and powerful machine learning classifiers like K-Nearest Neighbour (KNN), Decision Tree (DT), Random Forest (RF), AdaBoost (AB), Bagging Classifier (BC), ExtraTrees (ET), Stochastic Gradient Descent (SGD), Gradient Boosting (GB), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) were implemented to predict the severity of VDD. The final experimentation results showed that the Random Forest Classifier achieves better accuracy of 96 % and outperforms well on training and testing Vitamin D dataset. The McNemar's statistical test results support that the RF classifier outperforms than the other classifiers.

Machine Learning is one of the fastest emerging recent technologies which is used in various fields due to its high performance and ease in applicability. In recent years, the applications and usage of machine learning in the medical field is very high. The main objective of machine learning is to learn from the input data which is usually called training data and make future predictions with the new data.

The previous studies used only the traditional statistical models to predict the severity of deficiency in Vitamin D datasets. The traditional works applied the statistical models on vitamin datasets with a smaller size [9]. Probably if the traditional methods applied to larger datasets then there is a chance of degradation of the performance. To our knowledge, this is the first study to predict the severity of VDD which compared it with various ML models.

2. LITERATURE REVIEW

[1] M. Holick, "Vitamin D Deficiency," New England Journal of Medicine, vol. 357, no. 3, pp. 266-281, 2007.

Once foods in the United States were fortified with vitamin D, rickets appeared to have been conquered, and many considered major health problems from vitamin D deficiency resolved. But vitamin D deficiency is common. This review considers the role of vitamin D in skeletal and nonskeletal health and suggests strategies for the prevention and treatment of vitamin D deficiency.

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

The aim of the present paper was to review the most important mechanisms explaining the possible association of vitamin D deficiency and cardiovascular diseases, focusing on recent experimental and clinical data. Low vitamin D levels favor atherosclerosis enabling vascular inflammation, endothelial dysfunction, formation of foam cells, and proliferation of smooth muscle cells. The antihypertensive properties of vitamin D include suppression of the renin-angiotensin-aldosterone system, renoprotective effects, direct effects on endothelial cells and calcium metabolism, inhibition of growth of vascular smooth muscle cells, prevention of secondary hyperparathyroidism, and beneficial effects on cardiovascular risk factors. Vitamin D is also involved in glycemic control, lipid metabolism, insulin secretion, and sensitivity, explaining the association between vitamin D deficiency and metabolic syndrome. Vitamin D deficit was associated in some studies with the number of affected coronary arteries, postinfarction complications, inflammatory cytokines and cardiac remodeling in patients with myocardial infarction, direct electromechanical effects and inflammation in atrial fibrillation, and neuroprotective effects in stroke. In peripheral arterial disease, vitamin D status was related to the decline of the functional performance, severity, atherosclerosis and inflammatory markers, arterial stiffness, vascular calcifications, and arterial aging. Vitamin D supplementation should further consider additional factors, such as phosphates, parathormone, renin, and fibroblast growth factor 23 levels.

[3] B. Schottker, C. Herder, D. Rothenbacher, L. Perna, H. Muller, 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," European Journal of Epidemiology, vol. 28, no. 3, pp. 267-275, 2013.

Plausible mechanisms of how vitamin D deficiency may contribute to the development of diabetes mellitus have been proposed but longitudinal cohort studies have yielded heterogeneous results. In 7,791 initially diabetes-free participants of a German population-based cohort, aged 50-74 years, adjusted Cox regression models were employed to estimate hazard ratios (HR) with 95 % confidence intervals (CI) for the association of serum 25-hydroxyvitamin D (25(OH)D) quintiles and incident diabetes. Dose-response relationships were assessed with restricted cubic spline curves. Additionally, analyses accounting for the competing risks of diabetes and death were performed. During 8 years of follow-up, 829 study participants developed diabetes. In women, diabetes risk was significantly increased in the lowest 25(OH)D quintile (HR, 1.38; 1.09-1.75) and non-significantly increased in the 2nd quintile (HR, 1.24; 0.98-1.55) compared to women in 25(OH)D quintiles 3-5. The dose-response relationship showed a non-linear inverse association with risk starting to increase at 25(OH)D levels below 70 nmol/L (statistically significant: below 40 nmol/L). In men, 25(OH)D levels were not associated with diabetes incidence. Renal dysfunction was an effect modifier with a more than doubled diabetes risk in 25(OH)D quintile 1 and an about 1.5-fold risk in quintile 2 compared to quintiles 3-5 if subjects had renal dysfunction. The observed associations were not influenced by the competing risk of death. In this large cohort study of older adults, serum 25(OH)D levels were inversely associated with incident diabetes in women but not in men. The association was particularly strong in subjects with renal dysfunction.

[4] Mohr SB et al., "Serum hydroxyvitamin D and prevention of breast cancer: pooled analysis," Anticancer Res, vol. 31, pp. 2939-2948, 2011.

Background: Low serum levels of 25-hydroxyvitamin D [25(OH)D] have been associated with a high risk of breast cancer. Since publication of the most current meta-analysis of 25(OH)D and breast cancer risk, two new nested case-control studies have emerged.

Materials and methods: A PubMed search for all case-control studies on risk of breast cancer by 25(OH)D concentration identified 11 eligible studies. Data from all 11 studies were combined in order to calculate the pooled odds ratio of the highest vs. lowest quantile of 25(OH)D across all studies.

Results: The overall Peto odds ratio summarizing the estimated risk in the highest compared to the lowest quantile across all 11 studies was 0.61 (95% confidence interval 0.47, 0.80).

Conclusion: This study supports the hypothesis that higher serum 25(OH)D levels reduce the risk of breast cancer. According to the review of observational studies, a serum 25(OH)D level of 47 ng/ml was associated with a 50% lower risk of breast cancer.

[5] Y. Lee, R.-M. Ragguett, R. B. Mansur, J. J. Boutilier, J. D.Rosenblat, A. Trevizol, E. Brietzke, K. Lin, Z. Pan, M. Subramaniapillai, T. C. Chan, D. Fus, 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," Journal of Affective Disorders, vol. 241, pp. 519–532, 2018.

Psychological distress is a significant and growing issue in society. In particular, depression and anxiety are leading causes of disability that often go undetected or late-diagnosed. Automatic detection, assessment, and analysis of behavioural markers of psychological distress can help improve identification and support prevention and early intervention efforts. Compared to modalities such as face, head, and vocal, research investigating the use of the body modality for these tasks is relatively sparse, which is partly due to the limited available datasets and difficulty in automatically extracting useful body features. To enable our research, we have collected and analyzed a new dataset containing full body videos for interviews and self-reported distress labels. We propose a novel

approach to automatically detect self-adaptors and fidgeting, a subset of self-adaptors that has been shown to correlate with psychological distress. We perform analysis on statistical body gestures and fidgeting features to explore how distress levels affect behaviors. We then propose a multi-modal approach that combines different feature representations using Multi-modal Deep Denoising Auto-Encoders and Improved Fisher Vector Encoding. We demonstrate that our proposed model, combining audio-visual features with detected fidgeting behavioral cues, can successfully predict depression and anxiety in the dataset.

6. S. Alghunaim and H. H. Al-Baity, "On the Scalability of MachineLearning Algorithms for Breast Cancer Prediction in Big Data Context," IEEE Access, vol. 7, pp. 91535-91546, 2019.

Recent advances in information technology have induced an explosive growth of data, creating a new era of big data. Unfortunately, traditional machine-learning algorithms cannot cope with the new characteristics of big data. In this paper, we address the problem of breast cancer prediction in the big data context. We considered two varieties of data, namely, gene expression (GE) and DNA methylation (DM). The objective of this paper is to scale up the machine-learning algorithms that are used for classification by applying each dataset separately and jointly. For this purpose, we chose Apache Spark as a platform. In this paper, we selected three different classification algorithms, namely, support vector machine (SVM), decision tree, and random forest, to create nine models that help in predicting breast cancer. We conducted a comprehensive comparative study using three scenarios with the GE, DM, and GE and DM combined, in order to show which of the three types of data would produce the best result in terms of accuracy and error rate. Moreover, we performed an experimental comparison between two platforms (Spark and Weka) in order to show their behavior when dealing with large sets of data. The experimental results showed that the scaled SVM classifier in the Spark environment outperforms the other classifiers, as it achieved the highest accuracy and the lowest error rate with the GE dataset.

3. PROPOSED ALGORITHMS/TECHNIQUES

a) Logistic Regression

Logistic Regression (LR) is a statistical model that comes under a supervised ML technique which uses the logistic function to solve multi-class classification problems. We implemented LR using the multiclass parameters [29], [30] by using sklearn.linear_model.LogisticRegression. Let us consider an example $x \in R\ m$, then the score represented using y= MRx+f where matrix $V \in S\ C*M$ and vector $f \in S\ C$ are variables learned from the datasets. The Vitamin D deficiency severity of each class is given by the sigmoid of each individual class score z(yc = 1) = $\sigma(yc\ ) = \sigma(VC\ Tx + f)$.We have used One-vs-rest classifier for training sets and testing datasets which trains single classifier for each deficiency severity class.

b) K-Nearest Neighbor

The K-Nearest Neighbor (KNN) algorithm is a supervised machine learning algorithm [31], [32] which is used for classification and regression problems. The KNN algorithm looks for close proximity to the datasets. Initially, it calculate the mathematical values of the nearest data points and the nearest neighbor contributes more than the distant ones.The test data will be compared with the data points of the training sets then it finds the probability of similar data points. The algorithm classifies the output based on the points which have the highest probability. The irrelevant features may affect the performance of the KNN, so the relevant features are considered from the datasets. We have implemented KNN using import sklearn.neighbors. KNeighbors Classifier for the experiment.

c) Gaussian Naive Bayes Gaussian

Naïve Bayes is a special type supervised machine learning classifier used for classification problem which follows the probabilistic method. We have implemented GNB using sklearn.naive_bayes import GaussianNB for the experiment. It follows the preceding and future probability of the different severity classes in the vitamin D training and testing dataset respectively. $p(x = v\,|\,Sk\ ) = 1\ \sqrt{2\pi\sigma k}\ 2\ e\ -\ (v-\mu k)\ 2\ 2\sigma k\ 2$ (2) where, S-Different Severity measures;µk – Mean of the different variables; v- probability distribution; σk 2 – Variance of the values

d) AdaBoost Classifier

AdaBoosting Classifier is an ensemble machine learning approach which makes itself a strong learner by combining the weak learner model. A machine learning algorithm will act as a base learner when it accepts the weights from the training data. We have implemented AdaBoost Classifier by using sklearn.ensemble import AdaBoostClassifier. Initially, the training set will be selected randomly, and the model trains it iteratively. The misclassified observations are assigned with higher weight and it will get a higher probability in the next process of iteration. This procedure will continue until the training set data fits in the model without any fault.

e) Decision Tree

Decision Tree Classifier is an eminent supervised ML tool that is used for solving classification problems and it has a tree-like model or graphs. The DT can capture the decision-making knowledge from the given data. We have used DT using sklearn.tree import DecisionTreeClassifier for the experiment. In DT that every branch indicates the output of the test set and every leaf node represents the particular label. The classification rules are represented by the path from the root node to the leaf node. For our vitamin D deficiency severity modeling, each node in the tree predicts the deficiency severity and each branch indicates the states of the variable. The Vitamin D dataset has the four deficiency severity types as the outcome and has many independent variables,(c,T) =($c1$, $c2$, $c3$, $c4$…,T), where, T is the deficiencyseverity variable and the vector c is comprised of several independent variables like $c1$, $c2$, $c3$, $c4$ … cnused for classification.

f) Random Forest Classifier

Random forest classifier (RF) as proposed by Breiman [34], is an ensemble machine learning method used for solving classification problems. RF constitutes many decision trees randomly from the training set and then it aggregates the values from different decision trees and predicts final severity deficiency as the outcome. The parameters considered for RF are n_estimators (n=10), criterion, minimum samples split (split=2) and minimum sample leaf (leaf=1) in the dataset. The previous studies showed that the RF classifier outperforms well with the other classifiers [41]. We have used RF using sklearn.ensemble import RandomForestClassifier for the experiment.

g) Multi-Layer Perceptron Classifier

Multi-Layer perceptron (MLP) is a feed-forward artificial neural network and it is inspired by the biological brain, that tries to mathematically express the real brain that maps the set of inputs to the corresponding output. The MLP has three layers namely input, hidden and output layer. It has one input and output layer, but it has one or more hidden layers. The perceptron has inputs{$x1$, $x2$, $x3$, … xn) and each input has corresponding weights {$w1$, $w2$, $w3$, … wn).It has a summation processor with function ($g(x) = \sum wi\ .\ xi\ n\ i{=}0$) and it has an activation function. It has parameters like 4. iterations, learning rate, input/output layers, different 5. deficiency severity classes, and activation function. The training data and labels will be provided to the classifier for training. We have used MLP using sklearn.neural_network import MLPClassifier for the experiment. The MLP models use the log loss function to predict the accuracy and try to minimize the values where the good model has a log loss of 0. If the log loss increases, then there is a divergence in the prediction with the actual label.

h) Support Vector Machine

Support Vector Machine (SVM) is best known supervised machine learning classifier as proposed by Cortes &Vapnik [36] used to solve classification and regression problem The purpose of SVM is to find the hyperplane with the number of given independent variables and it distinctly categorizes the data points. The implementation of SVM is done by using sklearn.svm import SVC for our experiment. The distance between the classes should be maximized and we have use linear SVM. Hyperplane are used to classify the data points which is considered as decision boundaries. The points which falling both sides of the hyperplane are considered as different classes. The hyperplane dimension will depend on the number of features. The data points which nearer to the hyperplane are called support vectors. By using a support vector, the margin of the classifier is maximized and as well as it changes the hyperplane position.

4. RESEARCH METHODOLOGY SYSTEM ARCHITECTURE

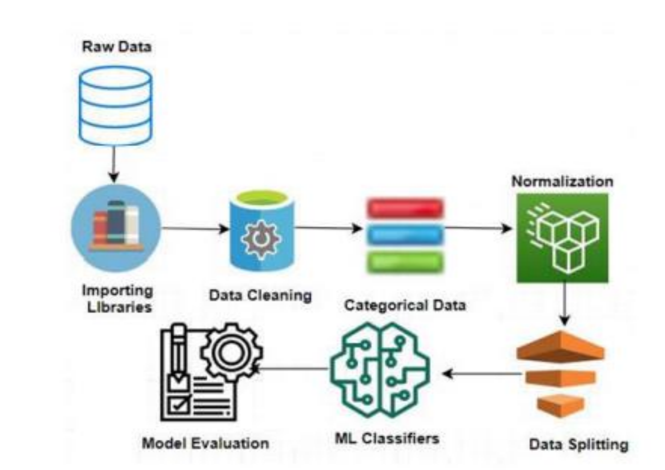


Figure 1: Data processing framework for VDD

5. PROPOSED MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

- Login,
- Browse and Train & Test Data Sets
- View Trained and Tested Accuracy in Bar Chart
- View Trained and Tested Accuracy Results
- View Vitamin Prediction
- View Vitamin Prediction Type Ratio
- Download Predicted Data Sets
- View All Remote Users
- Logout

Registration Module

In this module, the new remote user can register by entering he/his details i.e.,

- Username
- Password
- Email
- Country
- Signup

Remote User Module

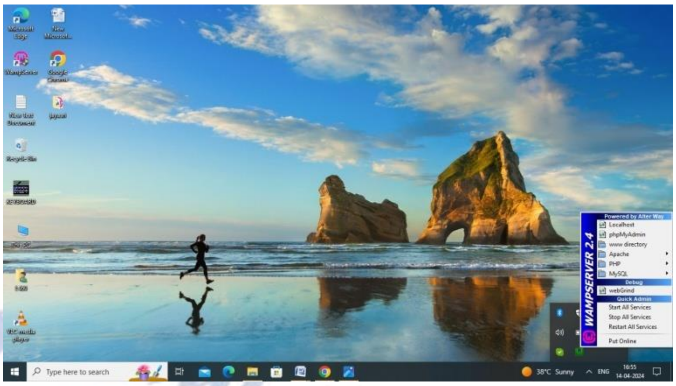
In this module, there are n numbers of users are present. User should register before doing any operations. Once Login is successful user will do some operations like

- Register and login
- Browse and Train & Test Data Sets
- View Trained and Tested Accuracy in Bar Chart
- View Trained and Tested Accuracy Results
- View Vitamin Prediction
- View Vitamin Prediction Type Ratio
- Download Predicted Data Sets

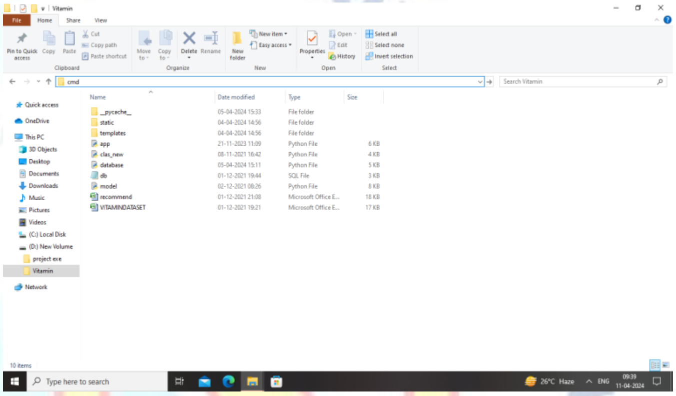
6. RESULT ANALYSIS

In this section, we discuss about the performance measures, statistical test and error measures with the comparison of accuracy of training versus testing datasets and different error measures for different machine learning models that we applied in our of Vitamin D dataset [40].

SCREEN 1:WAMP SERVER PAGE

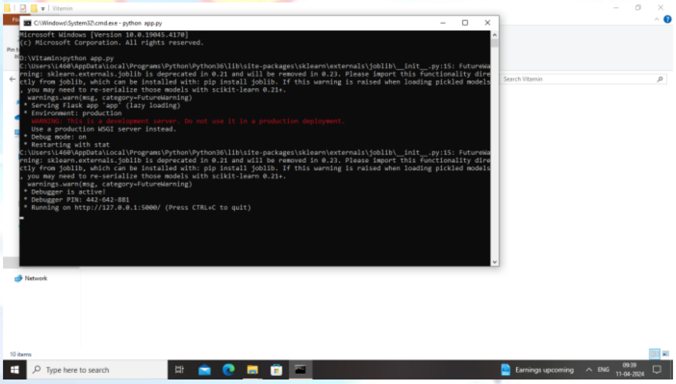


SCREEN 2:PROJECT FOLDER PAGE



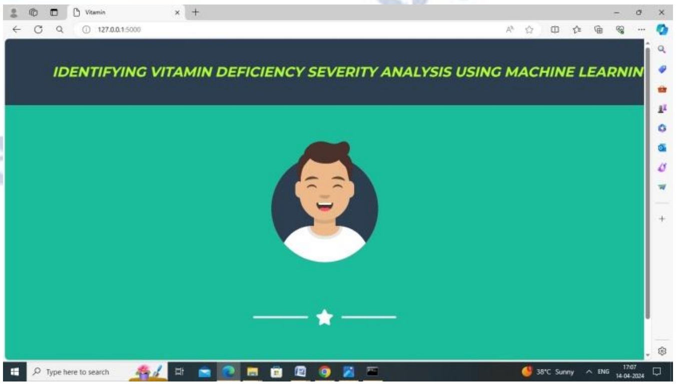
Name	Date modified	Type	Size
__pycache__	05-04-2024 15:33	File folder	
static	04-04-2024 14:56	File folder	
templates	04-04-2024 14:56	File folder	
app	21-11-2023 11:09	Python File	6 KB
clas_new	08-11-2021 16:42	Python File	4 KB
database	05-04-2024 15:11	Python File	5 KB
db	01-12-2021 19:44	SQL File	3 KB
model	02-12-2021 08:26	Python File	8 KB
recommend	01-12-2021 21:08	Microsoft Office E...	18 KB
VITAMINDATASET	01-12-2021 19:21	Microsoft Office E...	17 KB

SCREEN 3:GENERATING LINK IN CMD



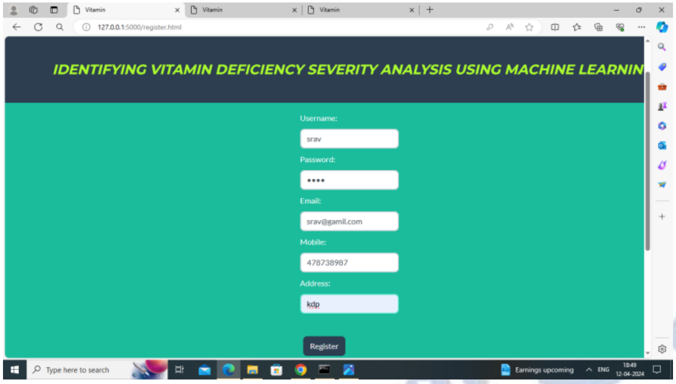
```
Microsoft Windows [Version 10.0.19045.4170]
(c) Microsoft Corporation. All rights reserved.

D:\Vitamin>python app.py
C:\Users\L460\AppData\Local\Programs\Python\Python36\lib\site-packages\sklearn\externals\joblib\__init__.py:15: FutureWa
rning: sklearn.externals.joblib is deprecated in 0.21 and will be removed in 0.23. Please import this functionality dire
ctly from joblib, which can be installed with: pip install joblib. If this warning is raised when loading pickled models
, you may need to re-serialize those models with scikit-learn 0.21+.
  warnings.warn(msg, category=FutureWarning)
 * Serving Flask app 'app' (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: on
 * Restarting with stat
C:\Users\L460\AppData\Local\Programs\Python\Python36\lib\site-packages\sklearn\externals\joblib\__init__.py:15: FutureWa
rning: sklearn.externals.joblib is deprecated in 0.21 and will be removed in 0.23. Please import this functionality dire
ctly from joblib, which can be installed with: pip install joblib. If this warning is raised when loading pickled models
, you may need to re-serialize those models with scikit-learn 0.21+.
  warnings.warn(msg, category=FutureWarning)
 * Debugger is active!
 * Debugger PIN: 442-642-881
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

SCREEN 4:HOME PAGE



IDENTIFYING VITAMIN DEFICIENCY SEVERITY ANALYSIS USING MACHINE LEARNIN

SCREEN 5:REGISTRATION PAGE



IDENTIFYING VITAMIN DEFICIENCY SEVERITY ANALYSIS USING MACHINE LEARNIN

Username: srav

Password: ••••

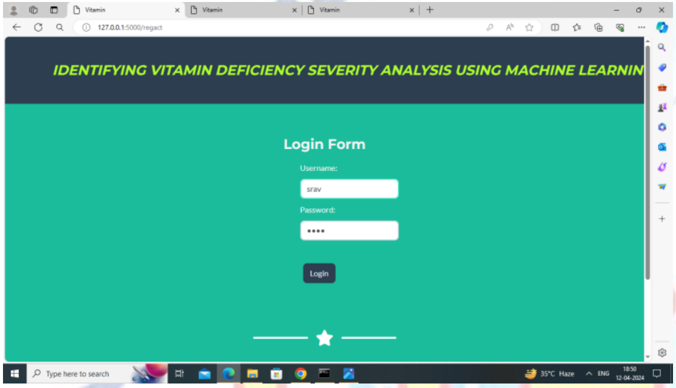
Email: srav@gamil.com

Mobile: 478738987

Address: kdp

Register

SCREEN 6:LOGIN PAGE



IDENTIFYING VITAMIN DEFICIENCY SEVERITY ANALYSIS USING MACHINE LEARNIN

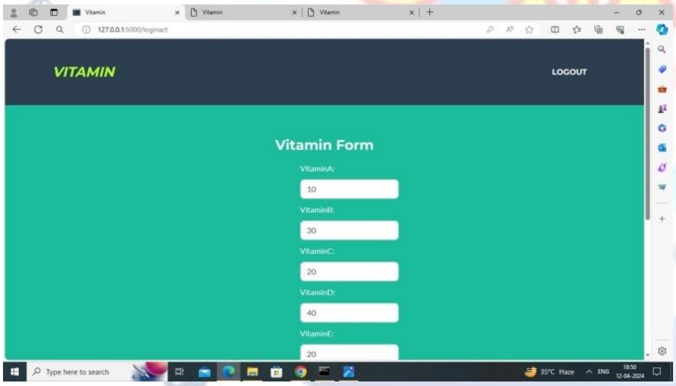
Login Form

Username: srav

Password: ••••

Login

SCREEN 7:TRAINING AND TESTINNG PAGE



VITAMIN

LOGOUT

Vitamin Form

VitaminA: 10

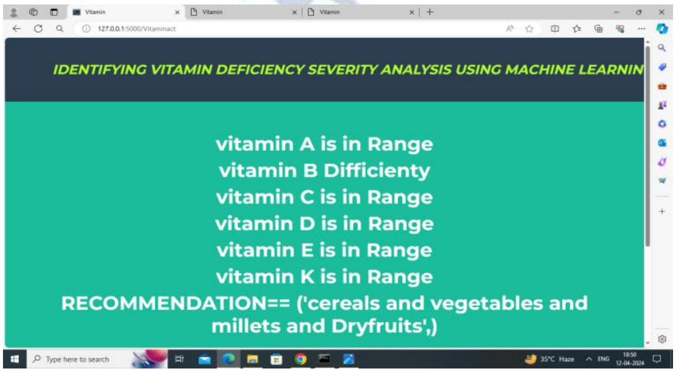
VitaminB: 30

VitaminC: 20

VitaminD: 40

VitaminE: 20

SCREEN 8:RESULT PAGE



IDENTIFYING VITAMIN DEFICIENCY SEVERITY ANALYSIS USING MACHINE LEARNIN

vitamin A is in Range
vitamin B Difficiency
vitamin C is in Range
vitamin D is in Range
vitamin E is in Range
vitamin K is in Range
RECOMMENDATION== ('cereals and vegetables and millets and Dryfruits',)

7. CONCLLUSION

The main objective of this study is to identify the best machine learning model in the prediction of VDD severity. The prediction accuracy was calculated and compared with the training and testing set. For this study, we have used 11 machine learning models and performance measures like precision, recall, F1-measure, and accuracy. We have used 11 parameters in the severity prediction and RFE [28] technique is used for feature selection. McNemar's statistical significance test is used to validate the empirical results. From McNemar's test, it is undoubtedly RF scores high in prediction when compared to different models and the Pearson's correlation coefficient and error measures result concluded the same.

The machine learning methods could be used as a substitute for the efficient prediction of severity of VDD with high accuracy. The results of this research work proved that the machine learning models especially the random forest classifier accurately predict the severity of Vitamin D deficiency. In particular, the Random forest classifier achieved the highest accuracy (96%) and outperforms well than other classifiers. This machine learning classifier will have a greater opportunity in the real-world medical domain which would assist experts to efficiently identify the severity of VDD. The major advantage of this study is that it has explored a new approach for the prediction of VDD severity using the Random Forest model and it has evaluated the results of the machine learning models using various performance measures accurately among the adolescents. So, the study claims that the Random forest model can be used to predict the severity of VDD with high accuracy than the other models. The future direction of our research is to validate the model with a different type of Vitamin D datasets of all age groups

ACKNOWLEDGEMENT

We thankful to all the referred journal authors for their elaborative study helps me to write this paper.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] M. Holick, “Vitamin D Deficiency,” New England Journal of Medicine, 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, 2012.

[3] B. Schottker, C. Herder, D. Rothenbacher, L. Perna, H. Muller, 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,” European Journal of Epidemiology, vol. 28, no. 3, pp. 267-275, 2013.

[4] Mohr SB et al., “Serum hydroxyvitamin D and prevention of breast cancer: pooled analysis,” Anticancer Res, vol. 31, pp. 2939-2948, 2011.

[5] Y. Lee, R.-M. Ragguett, R. B. Mansur, J. J. Boutilier, J. D.Rosenblat, A. Trevizol, E. Brietzke, K. Lin, Z. Pan, M. Subramaniapillai, T. C. Chan, D. Fus, 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 metaanalysis and systematic review,” Journal of Affective Disorders, vol. 241, pp. 519–532, 2018.

[6] S. Alghunaim and H. H. Al-Baity, “On the Scalability of MachineLearning 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, 2013.

[8] Souad Bechrouri, Abdelilah Monir, Hamid Mraoui, El houcine Sebbar, Ennouamane Saalaoui, Mohamed Choukri, “Performance of Statistical Models to Predict Vitamin D Levels,” SMC ’19: Proceedings of the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society, New York, NY, USA,2019, pp. 1-4.

[9] Gonoodi, K., Tayefi, M., Saberi-Karimian, M., Amirabadi zadeh, A., Darroudi, S., Farahmand, S., Abasalti, Z., Moslem, A., Nematy, M., Ferns, G., Eslami, S. and Mobarhan, M., “An Assessment Of The Risk Factors For Vitamin D Deficiency Using A Decision Tree Model,” Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 13, pp. 1773-1777, 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),” Journal of Biomedical Informatics, vol. 97, pp. 103257, 2019.