

A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning



N. Komal Kumar and D. Vigneswari

Abstract The remarkable technological advancements in the health care industry have improved recently for the betterment of patients' life and providing better clinical decisions. Applications of machine learning and data mining can change the available data to valuable information that can be used for recommending appropriate drugs by analyzing symptoms of the disease. In this work, a machine learning approach for multi-disease with drug recommendation is proposed to provide accurate drug recommendations for the patients suffering from various diseases. This approach generates appropriate recommendations for the patients suffering from cardiac, common cold, fever, obesity, optical, and ortho. Supervised machine learning approaches such as Support Vector Machine (SVM), Random Forest, Decision Tree, and K-nearest neighbors were used for generating recommendations for patients. The experimentation and evaluation of the study was carried out on a sample dataset created only for testing purpose and is not obtained from any source (medical practitioner). This experimental evaluation shows that the Random Forest classifier approach yields a very good recommendation accuracy of 96.87% than the other classifiers under comparison. Thus, the proposed approach is considered as a promising tool for reliable recommendations to the patients in the health care industry.

Keywords Drug · Data mining · Machine learning · Prescription · Health care

N. Komal Kumar (✉)

Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Avadi, Chennai, India

e-mail: komalkumarnapa@gmail.com

D. Vigneswari

Department of Information Technology, KCG College of Technology, Karapakkam, Chennai, India

e-mail: vigneswari121192@gmail.com

© Springer Nature Singapore Pte Ltd. 2021

G. S. Hura et al. (eds.), *Advances in Communication and Computational*

Technology, Lecture Notes in Electrical Engineering 668,

https://doi.org/10.1007/978-981-15-5341-7_1

1 Introduction

An Electronic Health Record (EHR) holds the digital version of the patients' medical paper report. They are real-time and instantly available to the users in a secured manner. It contains the medical and treatment histories of the patient right from the entry. This information includes demographics of the patient, medical history, medical diagnosis, medications, treatments, radiology images and test results of disease [1]. The main feature of the EHR is that it can track the medical history of the patient and can be shared among different medical organizations to yield a better diagnosis and treatment of the disease. The data stored in the EHR enable the practitioners to take better decisions on the disease. The Electronic Health Record (EHR) improves the treatment of the patient by reducing medical errors, duplicate treatment approaches, and delay in treatments. With the data stored in the database, a medical practitioner can apply machine learning algorithms to make effective decisions on treatments and medications. Classification approaches intelligently work to provide many accurate recommendations for short-term disease risks. Several reasons have made the authors work on an ensemble classifier to predict disease. Recent work in [2] has proposed an ensemble approach for predicting diabetes mellitus using Random Forest classifiers optimized with a genetic algorithm. Accurate recommendations are needed in the healthcare industry to provide a way better living to the people. Machine learning classifiers can be used in the analysis of predicting lung cancer among patients [3]. Machine learning algorithms with ensemble approaches have yielded good accuracy with success [4]. This paper totally focuses on providing a recommendation of drugs required by the patient to be taken for appropriate medication.

The proposed research is organized in the following sections. Section 2 discussed the related work in machine learning and recommendation systems, Sect. 3 describes the proposed methodology for drug recommendation, Sect. 4 describes the experimental evaluation and findings and finally concluded in Sect. 5.

2 Related Work

This section briefly discusses the recent works of machine learning and data mining approaches in health care.

Recent advances in biotechnology have led to the production of genetic data and medical information which are generated from the Electronic Health Records (EHR). Such data can be converted into valuable information using machine learning and data mining approaches. Kavakiotis et al. [5] investigated the applications of machine learning and data mining approaches for predicting, detecting, complications, and management. In the analysis, Support Vector Machine (SVM) outperformed in predicting and detecting diabetes complications. A technique for predicting the health of a fetal based on maternal medical history using machine learning

approaches was presented in [6], the work comprised of gathering dataset from 96 pregnant women and applying the data to the machine learning classifiers. In the predictive analysis Decision Tree achieved higher accuracy than the other supervised and unsupervised classifiers under clinical analysis. Similar work is done in [7], where a health monitoring system is established to monitor the health of pregnant women to avoid future complications in health. The proposed hybrid machine learning technique established relationships between the physical and mental factors to prevent health complications in women during pregnancies. Kuteesa et al. [8] surveyed machine learning techniques in HIV clinical research; the study concluded that there is a steady increase in the production of genetic data which needs machine learning techniques for translating information to knowledge. Kaur [9] proposed a novel and smart healthcare framework using machine learning with advanced security features such as monitoring daily activity, access control, dynamic data encryption standards and validation at endpoints, the proposed was more effective in diagnosing disease. A machine learning technique for detecting type-2 diabetes was presented in [10], where a data-informed framework is proposed using machine learning via feature engineering, the proposed was found effective in identifying medical subjects from an Electronic Health Record (EHR). Darabi et al. [11] applied supervised machine learning algorithms to estimate the risk mortality of patients admitted in Intensive Care Units (ICUs), the model outperformed in predicting the mortality of patients admitted in ICUs. A recommendation system involves presenting the appropriate data or information to the user using data analysis, such work is represented in [12], where a recommendation system was proposed for recommending most appropriate news to the user using a novel news utility model, the proposed model outperformed the existing models in recommending news to the user. Guan [13] proposed a recommendation model using multiple view information for addressing content heterogeneity; the proposed model outperformed using ample experiments and visuals demonstration. A comparison of machine learning tree classifiers for diabetic mellitus is proposed in [14], where the classifiers such as Random Forest, C4.5, Random Tree, REP tree and Logistic Model Tree accuracies were compared in prediction. A multi-model technique for predicting chronic kidney disease was proposed by Kumar [15], which involves two-stage, such as genetic algorithm and machine learning for its disease prediction. A hybrid machine learning classifier was proposed in [16], which involves a multi-swarm optimization and a multi-layer perceptron for the prediction of DENV serotypes.

3 Proposed Methodology

We worked on a sample EHR dataset created only for testing purposes and not obtained from any source (medical practitioner), the dataset contained admission ID, Sex, Age, disease symptoms, and its prescriptions. First, the dataset is subjected to pandas tool for cleaning, which is then subjected to train various machine learning

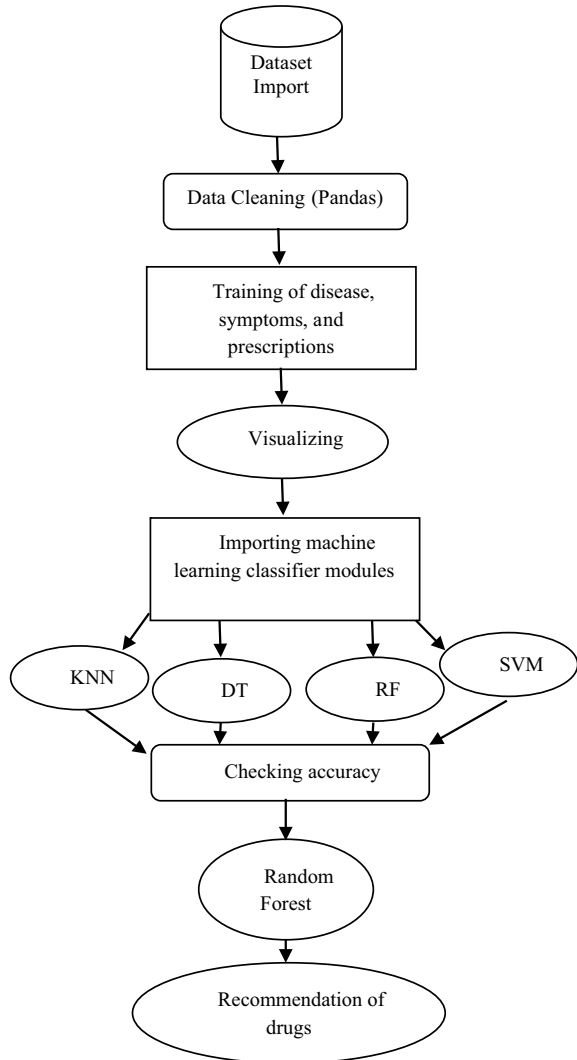
models for prediction, with the help of confusion matrix, the accuracy of the classifiers were calculated, and the classifier with high accuracy were applied to the test dataset for predicting and recommending drugs to the patients. The proposed framework is shown in Fig. 1.

The main steps in providing the recommendation for the disease are as follows:

Step 1: Importing and cleaning Dataset

In this step, the dataset was imported and cleaned using pandas, a python based data analysis tool.

Fig. 1 Proposed framework



Step 2: Training and visualizing of data.

In this step, the dataset is trained with its associated parameters and visualized using mathematical plotting functions to check the feasibility of data values.

A total of 891 testing records were visualized. Data visualizations help the data scientist to check the feasibility of the dataset attributes and its values. Figure 2 shows the relationship between males and female patients under analysis. The disease relationship between males and females is shown in Fig. 3. Symptom relationship between male and female is shown in Figs. 4 and 5, a prescription relationship between male and female is shown in Figs. 6 and 7. The scatter diagram for age and disease is shown in Fig. 8 and finally, the relationship between prescription and disease is shown in Fig. 9 respectively.

Step 3: Applying supervised machine learning classifiers

In this step, the cleansed data are subjected to machine learning classifiers such as Support Vector Machine (SVM), Random Forest, Decision Tree, and K-nearest neighbor.

Fig. 2 Male versus female

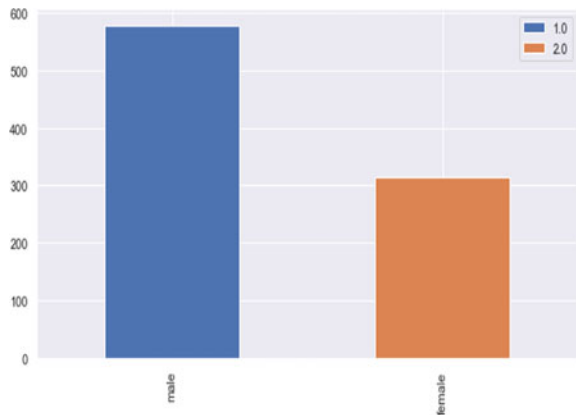


Fig. 3 Disease: male versus female

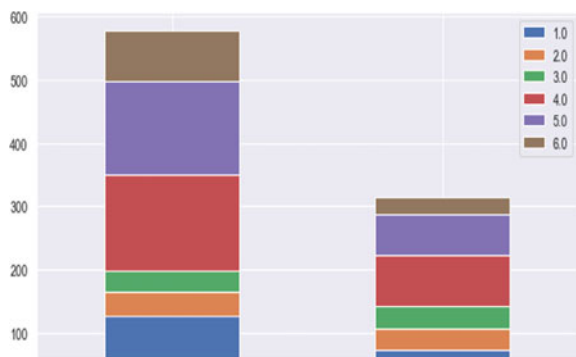


Fig. 4 Symptom 1: male versus female

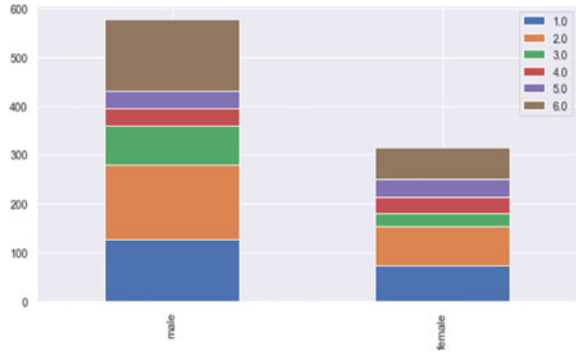


Fig. 5 Symptom 2: male versus female

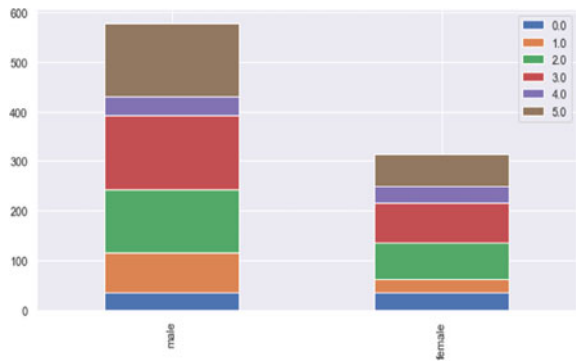
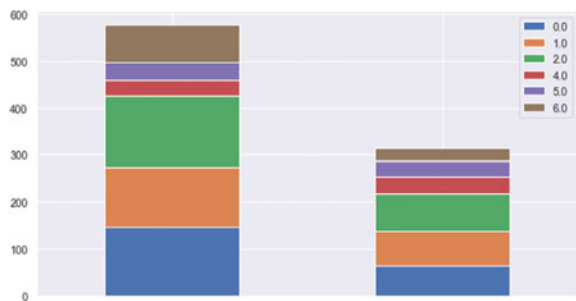


Fig. 6 Prescription 1: male versus female



Step 4: Accuracy measurement

In this step, the classifiers were tested for its accuracy in recommending prescriptions for certain medical symptoms.

Step 5: Termination

Terminate the process when the machine learning classifier model is built for a recommendation.

Fig. 7 Prescription 2: male versus female

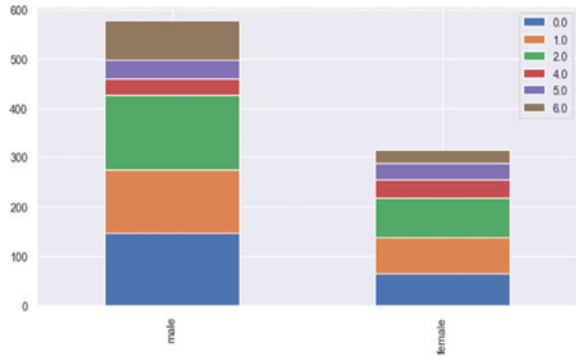


Fig. 8 Age versus disease

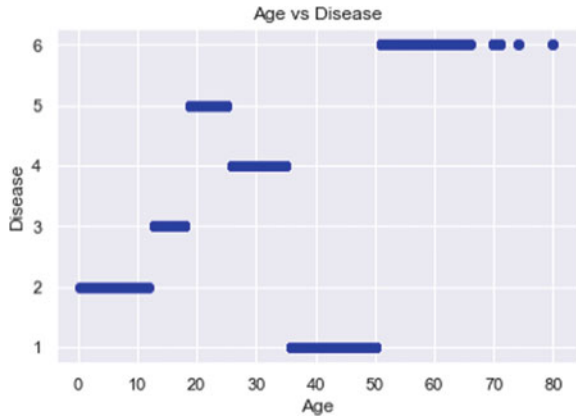
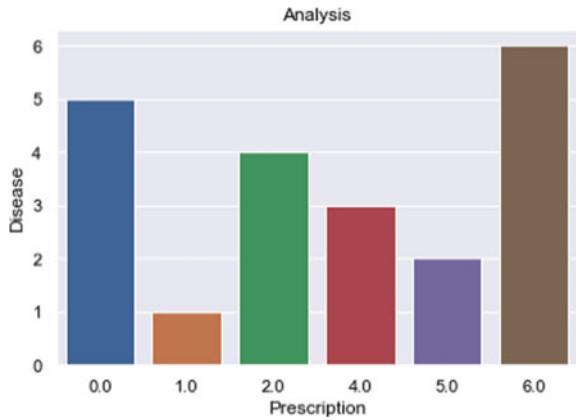


Fig. 9 Prescription versus disease



Algorithm 1: Training

```

1  Input: Dataset D
2  Output: Training Dataset for disease  $D_{T1}$ , symptom  $D_{T2}$ ,  $D_{T3}$ 
   and prescription  $P_1$  and  $P_2$ 
3  begin
4   $D_1 = \text{read}(\text{patientdata})$ 
5  Check if  $D_1.isnull()$ 
6  ExecuteMap(Sex)=  $D_1['Sex'].map(\{'male':1,'female':2\})$ 
7   $D_{T1} = \text{train}(\text{Disease}, D_1).map(\text{Cardiac}=1, \text{cold}=2, \text{fever}=3, \text{obesity}$ 
    $=4, \text{optical}=5, \text{Ortho}=6)$ 
8   $D_{T2} = \text{train}(\text{sympt1}, D_1).map(\text{ChestDiscomfort}=1, \text{fat}=2, \text{pain in}$ 
    $\text{bone}=3, \text{Running nose}=4, \text{Temperature increase}=5, \text{Vision}$ 
    $\text{problems}:6)$ 
9   $D_{T3} = \text{train}(\text{sympt1}, D_1).map(\text{decrease in bone density}=1, \text{heart}$ 
    $\text{burn}=2, \text{lack of physical activity}=3, \text{mild}$ 
    $\text{fever}=4, \text{myopia}=5, \text{nil}=0)$ 
10  $P_1 = \text{train}(\text{pres1}, D_1).map(\text{Aceinhibitors}=1, \text{Physicalactivity}=2,$ 
    $\text{Contactlens}=3, \text{Paracetamol}=4, \text{ibuprofen}=5, \text{Antidepressants}=6)$ 
11  $P_2 = \text{train}(\text{pres2}, D_1).map(\text{Aldosterone}=1, \text{bupropionnaltrexone} =$ 
    $2, \text{contact lens}=3, \text{ibuproifn}=4, \text{acetaminophen}=5,$ 
    $\text{Corticosteroids} =6)$ 
12 end

```

Algorithm 2: Classifier modeling

```

1 Input: Training Dataset for disease  $D_{T1}$ , symptom  $D_{T2}$ ,  $D_{T3}$  and
  prescription  $P_1$  and  $P_2$ , cross validation  $K, K$  nearest neighbour  $Kn$ 
2 Output: Classifier models  $C_1, C_2, C_3,$  and  $C_4$ 
3 begin
   $C_1 = KNeighborsClassifier(n\_neighbors = 13)$ 
4 Calculate score = cross_val_score(clf, train, target, cv=k_fold,
  n_jobs=1, scoring=scoring)
   $C_2 = SVC()$ 
6 Calculate score = cross_val_score(clf, train, target, cv=k_fold,
  n_jobs=1, scoring=scoring)
   $C_3 = RandomForestClassifier(n\_estimators=13)$ 
7 Calculatescore = cross_val_score(clf, train, target, cv=k_fold,
  n_jobs=1, scoring=scoring)
   $C_4 = DecisionTreeClassifier()$ 
8 Calculate score = cross_val_score(clf, train, target, cv=k_fold,
  n_jobs=1, scoring=scoring)
9 end

```

Algorithm 3: Predicting and recommendation

```

1 Input: RandomForestClassifier(n_estimators=13)
2 Output: Prediction  $P$ , Prescription  $P_r$ 
3 begin
4  $C_3 = RandomForestClassifier(n\_estimators=13)$ 
5  $C_3.fit(train, target)$ 
6 prediction =  $C_3.predict(test)$ 
7 out (Prediction  $P$ , Prescription  $P_{r1}$  or  $P_{r2}$ )
8 end

```

The training algorithm is given in Algorithm 1 deals with training dataset which will then be given to the classifiers for generating appropriate drug recommendation, classifiers modeling algorithm in Algorithm 2 deals with producing classifier models for predicting disease, and finally, prediction and recommendation in Algorithm 3 deals with predicting disease and recommending suitable medication to the patient (Fig. 10).

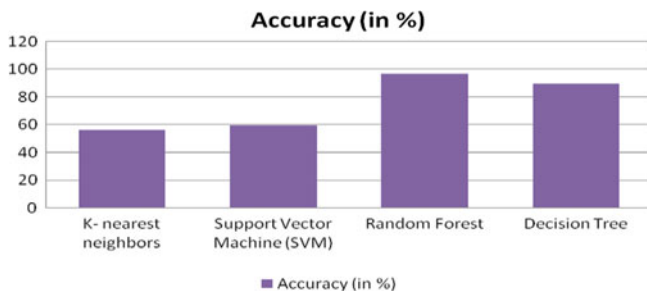


Fig. 10 Accuracy of the classifiers

Table 1 Accuracy of classifiers

Classifier	Accuracy (in %)
K-nearest neighbors	56.52
Support Vector Machine (SVM)	59.65
Random Forest	96.87
Decision Tree	89.68

4 Experimental Analysis and Findings

This section contains the experimental results and analysis of a recommendation system for multi-disease. A dual-core i3 system with 4 GB of RAM, pandas, NumPy, Ipython, SciKit_Learn, SciPy, StatsModels, and Matplotlib libraries was used in the experimental analysis for generating recommendations. Machine learning classifier algorithms were evaluated for its accuracy in recommending appropriate drug recommendations to overcome certain medical issues. The experimental analysis takes place in two steps. In the first step, the dataset is imported and cleansed using pandas tool, the second step involves in training machine learning models for the recommendation system where the test data is subjected to various classifiers such as SVM, Random Forest, Decision Tree, and K-NN and accuracy is measured, then the classifier with high accuracy is used in predicting disease with various symptoms taken into consideration, with the trained symptoms and prescriptions, the Random Forest classifier provides recommendations to the patients (Table 1).

5 Conclusion and Future Directions

In this work, a drug recommendation system for multi-disease using machine learning for healthcare was developed using a sample dataset which was created only for testing purposes and not obtained from any source (medical practitioner). The proposed method using Random Forest machine learning classifier showed that it

can be an effective drug recommendation tool in healthcare. The machine learning classifiers used in the analysis include K-nearest neighbors, Support Vector Machine (SVM), Random Forest, and Decision Tree to achieve accuracy and provide drug recommendations for the patients suffering from short-term disease. The experimental results showed that the proposed method using the Random Forest classifier yielded a higher predictive performance compared with the other classifiers under analysis. Based on the experimental results obtained, the proposed method is found to be effective in improving the quality of drug recommendation, whereby improving the healthcare industry. In the future, we are interested in applying ensemble machine learning algorithms for predicting and generating drug prescriptions for multi-disease.

References

1. <https://www.healthit.gov/faq/what-electronic-health-record-ehr>
2. Komal Kumar N, Vigneswari D, Vamsi Krishna M, Phanindra Reddy V (2019) An optimized random forest classifier for diabetes mellitus. In: Abraham A, Dutta P, Mandal J, Bhattacharya A, Dutta S (eds) Emerging technologies in data mining and information security. Advances in intelligent systems and computing, vol 813. Springer, Singapore. https://doi.org/10.1007/978-981-13-1498-8_67
3. Komal Kumar N, Vigneswari D, Kavya M, Ramya K, Lakshmi Druthi T (2018) Predicting non-small cell lung cancer: a machine learning paradigm. *J Comput Theor Nanosci* 15(6/7):2055–2058. <https://doi.org/10.1166/jctn.2018.7406>
4. Das R, Turkoglu I, Sengur A (2009) Effective diagnosis of heart disease through neural networks ensembles. *Expert Syst Appl* 36:7675–7680
5. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I (2017) Machine learning and data mining methods in diabetes research. *Comput Struct Biotechnol J* 15:104–116
6. Akbulut A, Ertugrul E, Topcu V (2018) Fetal health status prediction based on maternal clinical history using machine learning techniques. *Comput Methods Programs Biomed* 163:87–100
7. Lakshmi BN, Indumathi TS, Ravi N (2016) An hybrid approach for prediction based health monitoring in pregnant women. *Procedia Technol* 24:1635–1642
8. Bisaso KR, Anguzu GT, Karungi SA, Kiragga A, Castelnuovo B (2017) A survey of machine learning applications in HIV clinical research and care. *Comput Biol Med* 91:366–371
9. Kaur P, Sharma M, Mittal M (2018) Big data machine learning based secure healthcare framework. *Procedia Comput Sci* 132:1049–1059
10. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, Yang G, Chen Y (2017) A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int J Med Inf* 97:120–127
11. Darabi HR, Tsinis D, Zecchini K, Whitcomb WF, Liss A (2018) Forecasting mortality risk for patients admitted to intensive care units using machine learning. *Procedia Comput Sci* 140:306–313
12. Zihayat M, Ayanso A, Zhao X, Davoudi H, An A (2019) A utility-based news recommendation system. *Decis Support Syst* 117:14–27
13. Guan Y, Wei Q, Chen G (2019) Deep learning based personalized recommendation with multi-view information integration. *Decis Support Syst* 118:58–69
14. Vigneswari D, Komal Kumar N, Ganesh Raj V, Gagan A, Vikash SR (2019) Machine learning tree classifiers in predicting diabetes mellitus. In: IEEE-2019 5th international conference on advanced computing and communication systems (ICACCS), pp 84–87. <https://doi.org/10.1109/icaccs.2019.8728388>

15. Komal Kumar N, Lakshmi Tulasi R, Vigneswari D (2019) An ensemble multi-model technique for predicting chronic kidney disease. *Int J Electr Comput Eng* 9(2):1321–1326
16. Komal Kumar N, Roopa VD, Devi BAS (2018) MSO—MLP diagnostic approach for detecting DENV serotypes. *Int J Pure Appl Math* 118(5):1–6