**Journal of Neonatal Surgery**
***ISSN(Online): 2226-0439***
***Vol. 14, Issue 1s (2025)***
*https://www.jneonatalsurg.com*

OPEN ACCESS

# Developing A Machine Learning Model to Predict Medication Adherence in Chronic Disease Management

**Rashmi Sinha[1], Kishor Kumar Sahu[2]**

[1]Assistant Professor, Department of Pharmacy, Kalinga University, Raipur, India.
[2]Research Scholar, Department of Pharmacy, Kalinga University, Raipur, India.

Cite this paper as: Rashmi Sinha, Kishor Kumar Sahu, (2025) Developing a machine learning model to predict medication adherence in chronic disease management. *Journal of Neonatal Surgery*, 14 (1s), 10-16.

## ABSTRACT

In the healthcare industry, chronic disease prediction is crucial. It is crucial to diagnose the illness early. Large amounts of data are generated in computer science as a result of significant technological advancements. Many medical databases are created as clinical information networks advance. Data mining, the process of managing vast amounts of diverse data and extracting insights from it, has emerged as a crucial area of study. Early illness detection, patient treatment, and community services from huge data creation in the biomedical and healthcare communities are all benefited by the accurate analysis of medical data. Nowadays, one of the main areas of research is the management and extraction of knowledge from vast amounts of diverse data. Accurate processing of health data helps the biomedical and healthcare communities by improving patient care, early illness detection, and community services. However, analytical precision is reduced if medical data is not sufficiently consistent. For the domains of biomedical pattern recognition and master learning, the perception and diagnosis of chronic disease are guaranteed to be consistent. Additionally, the decision-making approach's goal is pushed. The study of high-dimensional, multi-modal biomedical data can be effectively addressed by machine learning. In computer science, chronic disease prediction is crucial. Early detection and prediction of chronic disease is crucial. The dataset for chronic obstructive pulmonary disease is used for analysis by the suggested model. Using supervised machine learning techniques such as Random Forest, Multiplayer Perceptron, Logistic Regression, Stochastic Gradient Descent, and XG boost, we provide a chronic obstructive pulmonary disease prediction system. Next, we examine classification techniques for predicting chronic diseases using a variety of criteria, such as accuracy, precision, sensitivity, ROC, and AUC.

***Keywords:*** *Disease, WHO, decision making.*

## 1. INTRODUCTION

Chronic illnesses are one of the leading causes of death in the modern world. The chronic illness is slowly gaining hold over the patient before eventually taking over. Chronic illness develops gradually and lasts for a long time. Early detection of chronic disease is necessary to prevent it from becoming unmanageable and to enable prompt treatment. According to the World Health Organization (WHO), the primary categories of chronic illnesses include diabetes mellitus, cancer, heart disease, and long-term respiratory conditions such chronic obstructive pulmonary disease (COPD) [13][15]. WHO (2004), Overall, COPD has been responsible for 2.8 fatalities worldwide, with China and India accounting for 65% of these deaths [1]. Comparatively speaking, there aren't many studies focused on respiratory conditions, particularly COPD, which is the subject of our investigation. According to WHO (2020), COPD affected 210 million people worldwide and was the third leading cause of mortality in that year. Emphysema and chronic bronchitis are two of the lung disorders that are referred to be COPD. Breathing becomes difficult due to COPD, a disorder that worsens with time [16]. There are two forms of COPD, which is a prevalent lung condition [11]. Both chronic bronchitis and emphysema [2]. It causes irritation of the alveoli, which are tiny air sacs. The air sacs become rigid with time and stop allowing carbon dioxide to leave your blood and oxygen to enter. Both the big and small airways swell and fill with mucus when someone has chronic bronchitis [4]. The mucus can obstruct the airways, making breathing difficult. Both categories are common in patients with COPD. This illness takes years to develop. Signs can be reduced and the condition can be prevented from worsening with early treatment. Smoking, working in a polluted workplace where we inhale a lot of dust, fumes, smoke, or gases, and hand smoke are the main causes of COPD [3]. COPD symptoms include persistent coughing, wheezing, dyspnea, and tightness in the chest. It must be taken care of, which poses a risk to human activity and increases the risk of lung cancer [9]. Due of their prevalence in causing death and disability globally, the review focuses on conditions like diabetes, cardiovascular disease, and chronic obstructive pulmonary

disease. These days, data science is crucial for analyzing large amounts of data. These days, health prediction is crucial because of the ways that modern life is lived. In light of this COPD scenario, we suggested a COPD prediction model to forecast COPD in its early phases [12].

## 2. PURPOSE AND OBJECTIVE OF THE STUDY

Early COPD identification is crucial to preventing the condition from getting worse. Early diagnosis of COPD is crucial for doctors to treat patients appropriately and prevent its progression. In medical care, individuals' preexisting problems can be managed and prevented from contributing to the development of COPD. In this regard, a number of models are put up to forecast COPD in its early stages. According to the evaluation, the author focused more on this chronic illness than others. We are developing a model that can forecast patients at an early stage of COPD. The suggested models are effective enough to forecast both healthy (Normal) and COPD [5]. Our method is based on a machine learning model that is more effective and has a greater accuracy rate in predicting COPD and normal. Our goal is to investigate how the suggested machine learning classifier is affected by both discrete and hybrid features. Our approach uses the ROI (Region of Interest) feature from the lung mask to predict the COPD patient. We guarantee that our approach will assist medical professionals in determining whether a patient has COPD or not. The system's goal is to assist medical professionals in predicting patients and referring them for additional treatment. The suggested solution aims to lessen the workload for practitioners by predicting and examining the return on investment of lung masks. For healthcare professionals, we suggested work to create a deserving system that could produce precise statistical reports and make flawless predictions. The common symptoms of COPD include limited airflow and ongoing respiratory difficulties. Chronic COPD is a major cause of death and depression globally, placing a financial and social burden on society [17]. According to a survey study, air flow is consistently restricted. Every year, patients in India's tertiary care facilities pass away as a result of inadequate modern monitoring. Because of chronic pulmonary obstructive disease (COPD), 17% of patients die. With the aid of a report produced by the system, the suggested model informs patients whether or not they are in the primary stage of COPD [6]. We believe that medical professionals should be able to tell patients whether they have mild COPD or not. Since we started the system with characteristics generated from ROI of lung mask CT images, the results are beneficial in many ways, allowing patients to be informed of the condition of lung mask infection. The suggested model's ability to provide predictions effectively aids medical professionals in evaluating patients based on system parameter reports and administering the proper care [8]. In the event of a lung infection, our algorithm can also detect a patient's health status early enough to alert them before they reach the latter stages of COPD. Our approach aims to prevent patients from reaching the worst phases of chronic diseases, which lessens the strain on the healthcare industry in a time when everyone on the planet is afflicted with them [7]. Our suggested model analyzes the effectiveness of the suggested machine learning technique on a number of metrics and assesses the impact of feature selection from ROI of CT images of lung masks using machine learning classifiers. We calculate the derived features of the COPD machine learning dataset's CT pictures. We use the suggested machine learning approach to illustrate the model with hybrid features and discrete feature selection. Our algorithm analyzes the COPD machine learning dataset and uses a subset of the dataset's attributes to determine whether a person has COPD or not [14].

- Examine the current body of knowledge based on machine learning applications.
- Choose a dataset sample that contains data on chronic obstructive pulmonary disease.
- Apply feature pre-processing to the dataset, which includes feature reduction and feature selection.
- Use supervised machine learning techniques to create predictive models.
- Examine models and choose the most accurate prediction model based on the average classification accuracy for chronic disease prediction.

## 3. PROPOSED FRAMEWORK

Although there are other machine learning models, supervised and unattended models are the most widely used. If the input and output values are saved in the training data, supervised learning can take place. Each collection of data that contains the inputs and the anticipated outcome is given a tracking signal name. As the inputs are loaded into the model, the training process is carried out based on the variation from the processed outcome. A COPD patient is represented as 1 and a normal (healthy) person is represented as 0. The proposed machine learning model was trained using COPD Machine Learning Datasets, 2020, and the training COPD dataset was DLCST [8]. The Frederikshavn testing dataset, which was created as a distinct testing dataset, was used to evaluate the model. GSS and KDEI feature datasets are included in the training dataset. These characteristics are assessed using the ROI of patient CT scans. Previously, we looked into GSS features, which are ROI intensity value histograms. Four scales, eight filters, and histograms are used to create the resulting GSS characteristics. We conducted tests using a mixed feature set of both features as well as individual GSS and KDEI. The machine learning model is fed the suggested model training cases and labels, and the model with the optimal hyperparameter is chosen. The learning algorithm is optimized using the grid search method, which looks for the optimal parameter and retrained model based on grids. To validate the trained model, the training dataset is divided into validation sets. If validation is necessary,

the Grid search method is used to identify the best model [10].

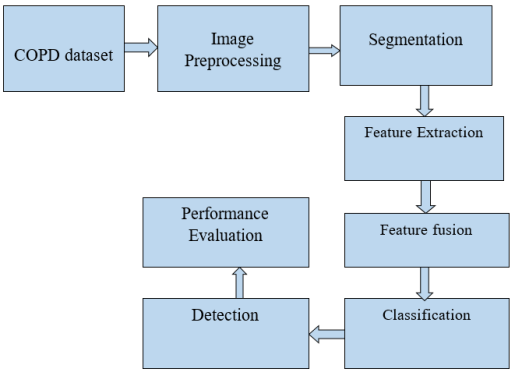


**Figure 1: Proposed architecture**

Features from CT scans of patients at the National Jewish Center in Denver, Colorado, are obtained from the COPD Machine Learning Datasets (2020). This COPD dataset includes a collection of features that were obtained from the ROI of a chest CT scan. The DLCST training dataset and the Frederikshavn testing dataset are included in this COPD dataset. Two features, gss and kdei, are derived (extracted) from the provided dataset. We represent KDEI for KDEI and GSS for GSS. It is a technique for turning unprocessed data into functions that more accurately capture the fundamental issue with forecasting models, improving prediction performance with fresh data. Feature engineering primarily aims to achieve two things:

- Creating a suitable dataset that complies with the requirements of the machine learning algorithm.
- A machine learning model with improved machine learning efficiency.

From the disease dataset represented by "n" qualities, the feature selection process seeks to extract the fewest optimal features. The feature selection approach for the analysis of chronic disease data is shown in Figure 3.1 below. The chronic disease dataset is provided as input to the filter-based feature selection methods, as shown in the image. Based on the features' merits, the filter-based feature selection chooses the feature. The proposed method will use the notion of probabilistic discrepancy to calculate the merit of the characteristics. Therefore, using potential statistical metrics like the correlation coefficient, mutual information, probability density, fisher score, standard error, and z score, the feature's relevance is assessed. By preventing the chance of repetition, this leads to the creation of an ideal subset of the features, which will have a direct effect on the prediction outcome. Because filter approaches don't require the models to be trained beforehand, they are substantially faster than wrapper techniques. The accuracy of the machine learning classifier is used by the wrapper technique to determine the feature significance. The wrapper approach is typically more accurate than the filter approach. However, the wrapper method is computationally demanding because it requires fitting the machine learning model with these feature examples before the relevance of the features can be determined. While filter techniques may not always be able to discover the best subset of features, wrapper methods will always find the important features by using the statistical model to comprehend the underlying assumptions and potential hypotheses in the data.

Label Set L(i) = {COPD(1), NORMAL(0)} is the dataset labeled for the proposed model, which was trained using Feature Set F(i) = {GSS(i), KDEI(i)}. Gathering and preparing feature and label data from the raw COPD machine learning dataset values is the first initialization step. The model's behavior is controlled by hyperparameters during the hyperparameter tuning procedure. The hyperparameter, which masterfully controls the model's nature, improves the algorithm while it is being prepared. Initialization provides the training model with these effective parameters. The cost of the specified hyper-parameter controls the machine learning training model's learning. When the model is initialized, hyperparameters are passed inside the contentions. The problem of selecting a collection of optimal hyperparameters for a learning computation is known as hyperparameter tuning or advancement in artificial intelligence. A boundary whose value is used to control the preparatory interaction could be called a hyperparameter. However, the estimates of a few others, such as the bounds of hub loads, are typically learned. It is also possible to anticipate that different constraints, loads, or rates will result in different information architectures inside an identical serene AI model. This kind of boundary is thought of as a hyperparameter, and by adjusting the model with this boundary, the critical thinking skills were effectively enhanced. Due to the hyperparameter setting's typical independence, the grid-search method is both embarrassingly parallel and suffering a dimensionality current. The grid search approach is arguably the most fundamental and basic way to tune parameters. The Grid search approach is used to generate models with the optimal hyperparameter combinations, examine each model, and select the best outcome system.

## 4. EVALUATION

The training phase and the testing phase are the two main components of the suggested paradigm. As we covered, the three feature section techniques are a crucial component of feature preprocessing. Training and testing are included on the proposed system's base page. Feature selection and feature reduction provide the foundation of the suggested model. Figure 2 displays all of the suggested model's significant possibilities. The COPD dataset under discussion is used as the train dataset, which is fed for training purposes before being tested and metrics from the model are estimated.

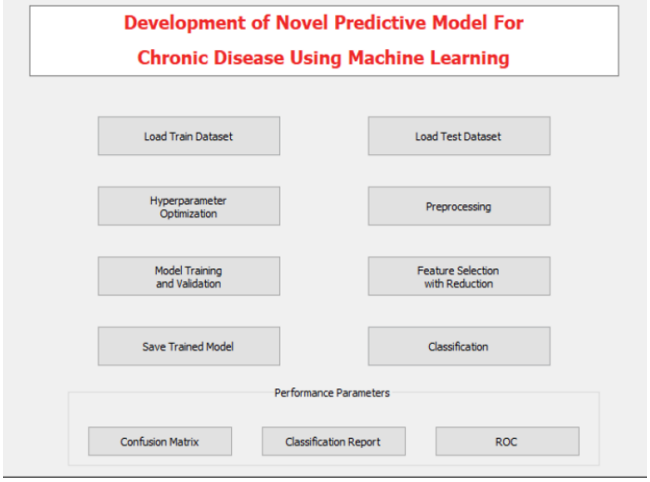


**Figure 2: Snapshot for base page of the proposed COPD prediction model**

The weights of several qualities produced by the entropy technique without taking into account the preferences of CR vehicular nodes are shown graphically in Figure 3. When the sum of these is 1, they are referred to as objective weights. Figure 4 displays the network selection ranking for spectrum handoff using these weights, excluding CR vehicular node preferences.

```
Train COPD GSS Features
   0.0001204725122480122  ...      0.13
0               0.000418  ...  0.086049
1               0.000100  ...  0.000000
2               0.000428  ...  0.014205
3               0.000628  ...  0.030217
4               0.001523  ...  0.000000

[5 rows x 320 columns]
(29999, 320)

Train COPD KDEI Features
   1.5193864395801615e-11  ...  0.0006449872441201412 8
0           2.385245e-17  ...                0.000519
1           4.755253e-12  ...                0.000570
2           8.673617e-19  ...                0.000352
3           8.105495e-16  ...                0.000444
4           2.007676e-08  ...                0.000707

[5 rows x 256 columns]
(29999, 256)
```

**Figure 3: Snapshot of loading of training dataset (COPD)**

In order to lower the dimensionality of the data, Figure 4 highlights key processes of principal component analysis on scaled features. Several PCA sets were created using the input features.

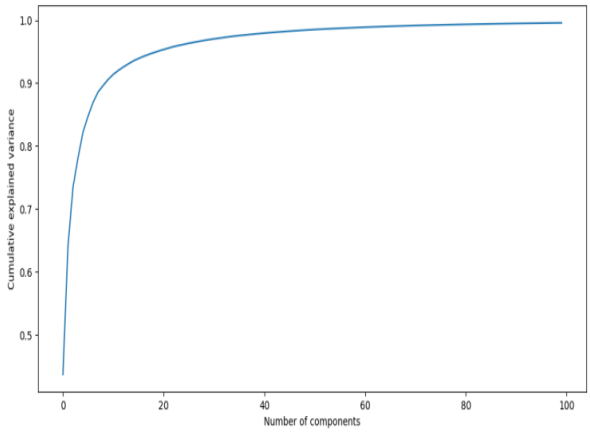


**Figure 4: Principal component analysis graph**

Figure 5 displays the ROC curve for the chosen classifier, which indicates performance at various thresholds using the classifier's TPR and FPR. In order to display the performance of every classifier, this option additionally generates the AUC graph of the chosen algorithm and the combined ROC graph for the chosen features.

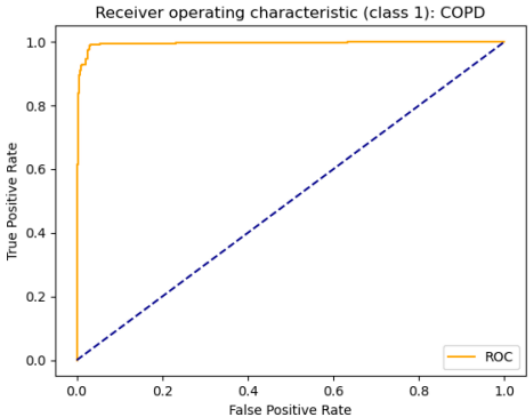


**Figure 5: Snapshot of ROC curve of selected classifier for COPD prediction**

the comparison between the suggested predictive model and the performance of the prior model. AUC and accuracy, two crucial categorization metrics, are used to compare prediction models. Our machine learning approach for predicting COPD looks at the outcomes of earlier models.

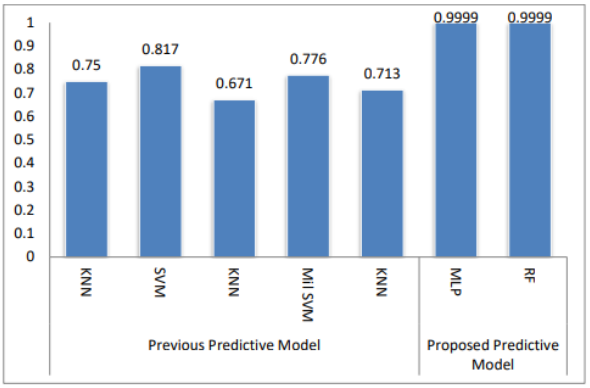


**Figure 6: Performance comparison (AUC) graph of the proposed model and previous model using hybrid feature selection.**

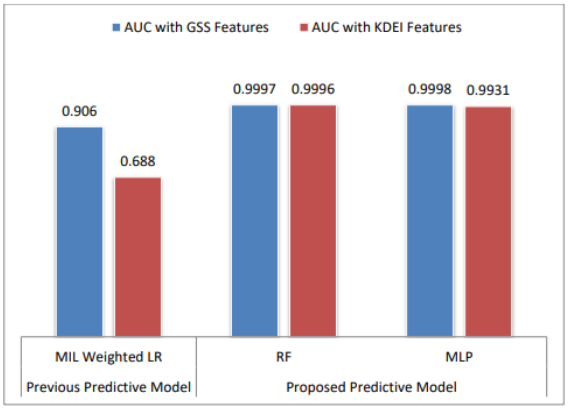


**Figure 7: Performance comparison (AUC) of the proposed model and previous model using discrete feature selection GSS and KDEI on the same COPD dataset.**

According to the AUC comparison, our suggested prediction is more accurate than the prior model, as seen in Figure 5-83. The MLP and RF models made the best predictions, according to the suggested machine learning algorithm's high accuracy and AUC.

## 5. CONCLUSION

We obtained the highest accuracy and AUC in the hybrid feature selection process using a machine learning classifier. In order to assess model performance, we also looked at the crucial ROC metric. We discovered that the suggested model performs exceptionally well across the board, particularly for Random Forest and Multilayer Perceptrons, which are close to 1.0. Our model's ability to correctly identify positive and negative classes is demonstrated via ROC curves. This demonstrates how well our machine learning model works with the suggested COPD dataset. Accuracy and AUC, two of the most crucial predictive model metrics, were missing from the prior model but were added to the suggested COPD prediction model. We effectively examined the impact of feature selection and obtained the best AUC and accuracy from the suggested machine learning classifier, making it suitable for diagnosing COPD datasets. Our COPD prediction model was examined using a hybrid and discrete feature selection method by the suggested machine learning classifier's effective performance. The suggested machine learning model can accurately forecast data related to COPD. Our strategy outperforms the previously published method on the same dataset in terms of accuracy and AUC. We tested the model using 7199 patient samples, and our machine learning model performs exceptionally well in identifying relevant examples and input feature categories. Both individual feature selection and hybrid feature selection are demonstrated by the variations in the samples, such as GSS and KDEI, with machine learning classifiers. The comparative study made it abundantly evident that our COPD prediction model is more effective and capable.

## REFERENCES

[1] Wang L, Fan R, Zhang C, Hong L, Zhang T, Chen Y, Liu K, Wang Z, Zhong J. Applying machine learning models to predict medication nonadherence in Crohn's disease maintenance therapy. Patient preference and adherence. 2020 Jun 3:917-26. https://doi.org/10.2147/PPA.S253732

[2] Menon PA, Gunasundari R. Deep Feature Extraction and Classification of Alzheimer's Disease: A Novel Fusion of Vision Transformer-DenseNet Approach with Visualization. https://doi.org/10.58346/JISIS.2024.I4.029

[3] Zullig LL, Jazowski SA, Wang TY, Hellkamp A, Wojdyla D, Thomas L, Egbuonu-Davis L, Beal A, Bosworth HB. Novel application of approaches to predicting medication adherence using medical claims data. Health services research. 2019 Dec;54(6):1255-62. https://doi.org/10.1111/1475-6773.13200

[4] Claycomb WR, Huth CL, Flynn L, McIntire DM, Lewellen TB, Center CI. Chronological examination of insider threat sabotage: Preliminary observations. J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.. 2012 Dec;3(4):4-20.

[5] Robinson L, Arden MA, Dawson S, Walters SJ, Wildman MJ, Stevenson M. A machine-learning assisted review of the use of habit formation in medication adherence interventions for long-term conditions. Health Psychology Review. 2024 Jan 2;18(1):1-23. https://doi.org/10.1080/17437199.2022.2034516

[6] tülay Çağatay I, Özbaş M, Yılmaz HE, Ali N. Determination of antibacterial effect of Nannochloropsis oculata against some rainbow trout pathogens. Natural and Engineering Sciences. 2021 Jul 1;6(2):87-95. http://doi.org/10.28978/nesciences.970543

[7] Lo-Ciganic WH, Donohue JM, Thorpe JM, Perera S, Thorpe CT, Marcum ZA, Gellad WF. Using machine learning to examine medication adherence thresholds and risk of hospitalization. Medical care. 2015 Aug 1;53(8):720-8. https://doi.org/10.1097/MLR.0000000000000394

[8] Biljana V, Mile V, Ljubica F, Dragan A, Evica J. Analysis of Occupational Injuries in an Iron Ore Mine in Bosnia and Herzegovina in the Period from 2002 to 2021. Archives for Technical Sciences/Arhiv za Tehnicke Nauke. 2024 Jan 1(30). https://doi.org/10.59456/afts.2024.1630.033V

[9] Ramakrishnan J, Ravi Sankar G, Thavamani K. Publication Growth and Research in India on Lung Cancer Literature: A Bibliometric Study. Indian Journal of Information Sources and Services. 2019;9(1):44-7. https://doi.org/10.51983/ijiss.2019.9.S1.566

[10] Kanyongo W, Ezugwu AE. Machine learning approaches to medication adherence amongst NCD patients: A systematic literature review. Informatics in Medicine Unlocked. 2023 Jan 1;38:101210. https://doi.org/10.1016/j.imu.2023.101210

[11] Hartigan P. Diabetic Diet Essentials for Preventing and Managing Chronic Diseases. Clinical Journal for Medicine, Health and Pharmacy. 2023 Oct 9;1(1):16-31.

[12] Burgess-Hull AJ, Brooks C, Epstein DH, Gandhi D, Oviedo E. Using machine learning to predict treatment adherence in patients on medication for opioid use disorder. Journal of Addiction Medicine. 2023 Jan 1;17(1):28-34. https://doi.org/10.1097/ADM.0000000000001019

[13] Nejad ND. Diagnosis of heart disease and hyperacidity of stomach through iridology based on the neural network introduction. International Academic Journal of Science and Engineering. 2015;2(6):17-25.

[14] Kanyongo W, Ezugwu AE. Feature selection and importance of predictors of non-communicable diseases medication adherence from machine learning research perspectives. Informatics in Medicine Unlocked. 2023 Jan 1;38:101232. https://doi.org/10.1016/j.imu.2023.101232

[15] Nithyalakshmi V, Sivakumar R, Sivaramakrishnan A. Automatic detection and classification of diabetes using artificial intelligence. Int Acad J Innov Res. 2021;8(1):1-5. https://doi.org/10.9756/IAJIR/V8I1/IAJIR0801

[16] Kanyongo W, Ezugwu AE. Machine learning approaches to medication adherence amongst NCD patients: A systematic literature review. Informatics in Medicine Unlocked. 2023 Jan 1;38:101210. https://doi.org/10.1016/j.imu.2023.101210

[17] Xu J, Zhao X, Li F, Xiao Y, Li K. Prediction Models of Medication Adherence in Chronic Disease Patients: Systematic Review and Critical Appraisal. Journal of Clinical Nursing. 2024. https://doi.org/10.1111/jocn.17577

[18] Wang L, Fan R, Zhang C, Hong L, Zhang T, Chen Y, Liu K, Wang Z, Zhong J. Applying machine learning models to predict medication nonadherence in Crohn's disease maintenance therapy. Patient preference and adherence. 2020 Jun 3:917-26.