Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 2, pp. 249–260
DOI : 10.47974/JSMS-1251

Predictive analytics for patient adherence : A machine learning approach in medication management

A. A. Patil [‡]
Department of Pharmacology
Krishna Institute of Pharmacy
Krishna Vishwa Vidyapeeth (Deemed to be University)
Karad
Maharashtra
India

Vaishali Niranjane [†]
Department of Electronics and Telecommunication
Yeshwantrao Chavan College of Engineering
Nagpur
Maharashtra
India

S. B. Dhadde [§]
Department of Pharmacology
Krishna Institute of Pharmacy
Krishna Vishwa Vidyapeeth (Deemed to be University)
Karad
Maharashtra
India

Anil Kumar *
Department of Artificial Intelligence & Data Science
Poornima Institute of Engineering &Technology
Jaipur
Rajasthan
India

[‡] *E-mail:* **anuppatil.pharma@gmail.com**
[†] *E-mail:* **niranjane_vaishali@yahoo.com**
[§] *E-mail:* **sbdhadde@hotmail.com**
* *E-mail:* **anil.vashu@gmail.com (Corresponding Author)**

© Taru Publications

Santosh Borde [@]
Department of Computer Engineering
JSPM's Rajarshi Shahu College of Engineering
Pune
Maharashtra
India

Pankaj Chandre [#]
MIT School of Computing
MIT Art Design And Technology University
Maharashtra
India

Abstract

In this paper, we investigate the use of predictive analytics and machine learning to improve patients' adherence to prescribed medications. By analyzing past patient data, we are able to create a model that may foresee adherence behaviors based on characteristics like age, gender, health status, and income. In order to analyze and identify major drivers of non-adherence and enable personalized therapies, our method incorporates cutting-edge machine learning techniques. The methodology takes into account each individual's situation while making suggestions, leading to a more proactive and specific strategy for improving drug adherence. Predictive analytics allows healthcare providers to foresee adherence issues, intervene at the right time, and boost patient outcomes. This research contributes to the developing landscape of healthcare by utilizing the potential of machine learning for precision medicine in drug management, addressing a vital part of patient care and public health

Subject Classification: *68Q32.*

Keywords: *Machine learning, Predictive model, Patient adherence, Medication management.*

1. Introduction

Optimizing healthcare results depends in large part on patients following their prescribed drug schedules. Medication non-adherence is a major problem that can lead to poor health and higher healthcare costs. In this work, we look at how predictive analytics and machine learning can be used in the field of drug management to help solve this problem [1]. Patient care that is both individualized and highly efficient will require increasingly sophisticated technological tools in the future. Adherence to

[@] *E-mail:* **santoshborde@yahoo.com**
[#] *E-mail:* **pankajchandre30@gmail.com**

medication regimens is a multifaceted phenomenon that is affected by a wide range of variables. When it comes to meeting the unique requirements of each patient, standard medical care often falls short [2]. To close this knowledge gap, this research will use machine learning algorithms to analyze past patient data to create a predictive analytics model. Patterns and predictors of non-adherence are identified by the model, allowing healthcare providers to engage proactively and tailor their approach [11].

By helping healthcare [3] providers anticipate and respond to adherence difficulties before they increase, this study has the potential to radically improve drug management. This patient-centered, machine learning-driven strategy aims to improve patient outcomes and develop healthcare practices by increasing the precision and efficacy of interventions. This research [4] adds to existing literature on the use of technology and data-driven insights to improve medication adherence and patient care in today's dynamic healthcare system. Effective healthcare requires an appreciation for the importance of medication adherence. Complexities [5] in drug administration, such as forgetfulness, adverse effects, and cost, call for creative approaches to treatment. The research examines past interventions and assesses traditional tactics, highlighting the limitations of each in the context of the dynamic nature of patient care today [12]. This section delves into predictive analytics and describes how data-driven methods might radically alter healthcare by enabling accurate forecasting of patient compliance. This enables preemptive interventions, encouraging a personalized approach to patient care, signifying a paradigm change in healthcare delivery [6].

2. Proposed Method

2.1 *Data Collection*

2.2.1 Source of Patient Data

It is critical to locate a wide variety of thorough patient data sources. Possible data sources include electronic health records (EHRs), pharmacy records, and data from wearable medical devices [7]. Here, we take a look at the many data sets that might be mined for insights into patients' propensity to take their medications as prescribed.

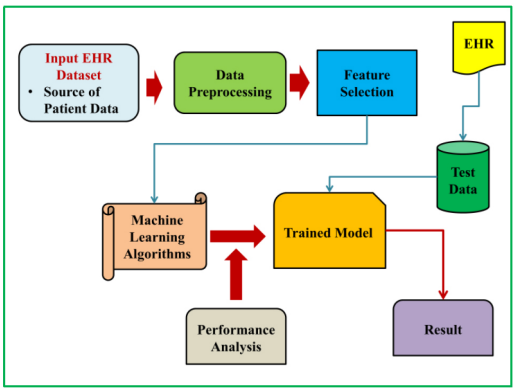

Figure 1

Proposed system architecture

2.2.2 Data Preprocessing

Raw data must be preprocessed as shown in Figure 1 to ensure quality and uniformity before being analyzed. Taking care of missing data, standardizing file formats, and dealing with anomalies are all part of this process. For the sake of openness, it is crucial that these preprocessing processes be discussed.

2.2 *Feature Selection*

Understanding how demographic factors like age, gender, and ethnicity affect adherence trends and feeding them into the predictive model is essential. By looking at a patient's entire medical history, from childhood diseases to the present, we can learn a lot about their likelihood of sticking to their treatment plan [8]. Socioeconomic factors, including income, education level, and access to healthcare, have a critical influence in influencing adherence. Inclusion and deep processing of these factors in the model enhance its applicability across varied patient populations, ensuring a comprehensive understanding of the multiple influences on medication adherence.

2.3 *Machine Learning Algorithms*

2.3.1 Selection Criteria

Choosing appropriate algorithms involves considering the nature of the problem and the characteristics of the data [9]. This section outlines

the criteria used for selecting machine learning algorithms, such as accuracy, interpretability, and scalability.

2.3.2 Model Training and Validation

Detailing the process of training the selected algorithms on the dataset and validating their performance is critical [13]. Discussing the use of training and validation sets, cross-validation techniques, and model evaluation metrics ensures the reliability and generalizability of the predictive model.

3. Predictive Model for Patient Adherence

3.1 *Model Architecture*

Detailing the architecture of the prediction model is critical for transparency and replicability. This segment describes the model's internal organization, including its various levels and interconnections [10]. Explaining the structure of the model used to predict patient adherence is helpful whether the model is a support vector machine (SVM), a decision tree, or a Naive Bayes classifier.

a. Support Vector Machine:

For classification and regression, a popular supervised machine learning approach is support vector machines (SVM). In the context of predictive analytics for patient adherence, below is a step-by-step mathematical model for a support vector machine:

1. Data Preparation:

   Let D be the dataset with n instances and m features. Each instance i is represented as xi with corresponding binary labels *yi*

2. Objective Function:

   The SVM aims to find a hyperplane with weights w and bias b that separates the data with the largest margin while minimizing classification errors. The objective function is given by:

$$\text{Minimize}\frac{1}{2}\|w\|^2 + C\Sigma i = 1^n \max(0,1 - yi(w \cdot xi - b)) \quad (1)$$

   Where, $\|w\|$ is the Euclidean norm of the weight vector, and C is the regularization parameter controlling the trade-off between maximizing the margin and minimizing classification errors.

3. Decision Function:

   The decision function for a new instance x is given by:

$$f(x) = w \cdot x - b \tag{2}$$

4. Constraints:

   The constraints ensure that each instance is on the correct side of the decision boundary with a margin of at least 1:

$$yi(w \cdot xi - b) \geq 1 \tag{3}$$

5. Hinge Loss:

   The term $\max(0, 1 - yi(w \cdot xi - b))$ represents the hinge loss, penalizing instances that are misclassified or close to the decision boundary.

6. Lagrange Multipliers:

   Introduce Lagrange multipliers (αi) to convert the inequality constraints into equality constraints:

$$L(w, b, \alpha) = \frac{1}{2} \| w \|^2 + C \Sigma i = 1^n \max(0, 1 - yi(w \cdot xi - b)) - \Sigma i = 1^n \alpha i [yi(w \cdot xi - b) - 1] \tag{4}$$

7. Dual Form:

   Derive the dual form of the optimization problem by finding the saddle point of the Lagrangian. The dual form involves maximizing:

$$W(\alpha) = \sum_{i=1}^{n} \alpha i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha i \alpha j yi\, yj (xi \cdot xj) \tag{5}$$

   subject to $\sum_{i=1}^{n} \alpha i\, yi = 0$ and $0 \leq \alpha i \leq C$ for all i.

b. Decision Tree:

1. Objective Function:

   The objective is to find the optimal splits that minimize impurity, commonly represented as:

$$\text{Minimize} \sum_{t=1}^{T} \text{Impurity}(Dt)$$

   Where, T is the number of terminal nodes (leaves), and Dt is the dataset in the t-th leaf.

2. Splitting Criteria:

   Choose a feature f and a threshold t to split the dataset into subsets left_Dt and right_Dt:

$$\text{Impurity}_{Gain(D,f,t)} = \text{Impurity}(D) - \left[\frac{|D_{left}|}{|D|} X \,\text{Impurity}(D_{left})\right] - \left[\frac{|D_{left}|}{|D|} X \,\text{Impurity}(D_{right})\right]$$

3. Recursive Splitting:

   Repeat the splitting process for each subset until a stopping criterion is met.

4. Decision Rules:

   Formulate decision rules for each leaf node based on the majority class:

$$\text{Decision_Rule}(Dt) = \text{Majority_Class}(Dt)$$

c. Naïve Bayes:

1. Data Preparation:

   Prepare a dataset D with n instances and m features. Each instance i is represented as xi with corresponding binary labels yi.

2. Prior Probability:

   Calculate the prior probability of adherence (P(Adherent)) and non-adherence (P(Non-Adherent)):

$$P(\text{Adherent}) = \frac{\text{Number of Adherent Instances}}{n}$$

$$P(\text{Non - Adherent}) = \frac{\text{Number of Non}-\text{Adherent Instances}}{n}$$

3. Conditional Probability:

   For each feature xj given the class, calculate the conditional probability using the assumption of feature independence:

$$P(xj \mid \text{Adherent}) = \frac{\text{Number of instances with } xj}{\text{Number of Adherent instances}}$$

$$P(xj \mid \text{Non - Adherent}) = \frac{\text{Number of Non-Adherent instances with } xj}{\text{Number of Non-Adherent instances}}$$

4. Likelihood Calculation:

   Calculate the likelihood for a given instance's feature values under each class:

$$\text{Likelihood}(\text{Adherent}) = P(\text{Adherent}) \times \prod_{j=1}^{m} P(xj \mid \text{Adherent})$$

$$\text{Likelihood}(\text{NonAdherent}) = P(\text{NonAdherent}) \times \prod_{j=1}^{m} P(xj \mid \text{Non Adherent})$$

5. Posterior Probability:

   Calculate the posterior probability for each class:

$$P(\text{Adherent} \mid x1, x2, \ldots, xm) \propto \text{Likelihood (Adherent)}$$

$$P(\text{Non - Adherent} \mid x1, x2, \ldots, xm) \propto \text{Likelihood}(\text{Non - Adherent})$$

6. Prediction:

   Classify the instance as Adherent or Non-Adherent based on the class with the higher posterior probability.

4. Results and Discussion

Predictive findings from three machine learning models Support Vector Machine (SVM), Decision Tree (DT), and Naive Bayes (NB) are summarized in Table 1. Key performance parameters such as Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC) were used to assess these models, yielding insightful information on their usefulness in a specific predictive analytics setting.

The Support Vector Machine (SVM) model showed excellent predictive abilities, with an Accuracy of 89.33%. The model's ability to accurately detect positive instances while minimizing false positives is demonstrated by its Precision score of 87.52%.

Table 1

Result for predictive analysis using Machine learning model

Model	Accuracy	Precision	Recall	F1 Score	AUC
SVM	89.33	87.52	91.55	90.89	90.23
DT	91.23	90.44	92.33	89.56	89.22
NB	90.56	92.45	90.12	90.54	88.45

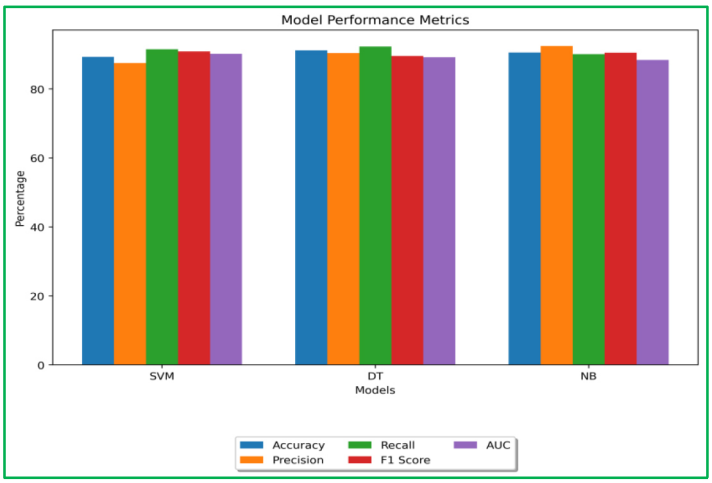

Figure 2

Representation of Model performance

In addition, SVM showed its competence in catching a sizable fraction of genuine positive events with a Recall of 91.55%. The model's strong performance is reflected in its F1 Score of 90.89%, which is a balance of Precision and Recall. The Area Under the Curve (AUC) value of 90.23% implies a high discriminatory power, making SVM well-suited for the prediction task. The Accuracy of the Decision Tree (DT) model was even higher, at 91.23 percent as shown in Figure 2.

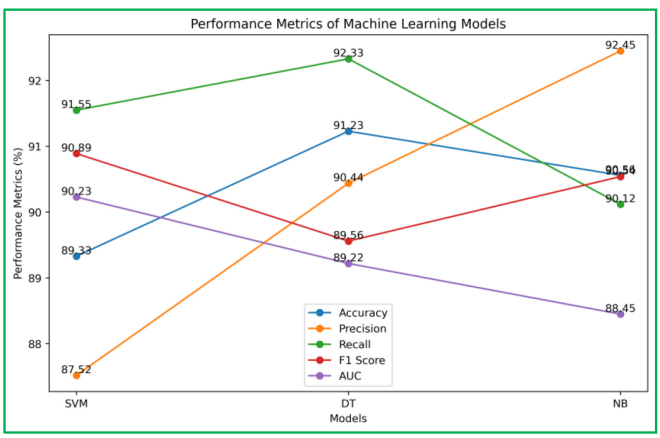

Figure 3

Representation of evaluation parameter comparison

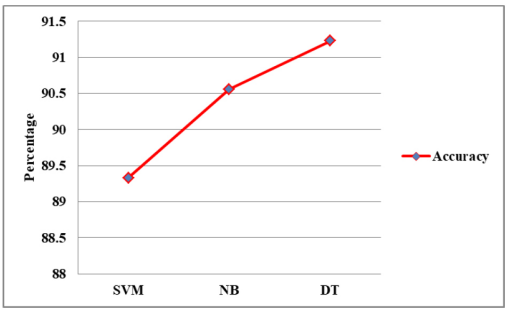

Figure 4

Comparison of Accuracy of proposed method

The model's accuracy in classifying positive examples (90.44%) and its efficiency in catching true positives (92.33%) are highlighted by the Precision and Recall metrics, respectively. Nonetheless, as compared to SVM, the F1 Score of 89.56% indicates a somewhat less optimal balance between Precision and Recall. The AUC of 89.22% highlights the DT model's prediction capability and reveals its great discriminatory potential. The Naive Bayes (NB) model was able to obtain an Accuracy of 90.56%, proving its proficiency in producing accurate predictions. With a Precision of 92.45% and a Recall of 90.12%, it appears that positive instances are being accurately identified. The F1 Score of 90.54% strikes a well-balanced performance between Precision and Recall. However, as compared to SVM and DT, the AUC of 88.45% indicates a somewhat weaker discriminatory power as shown in Figure 3, 4 and 5.

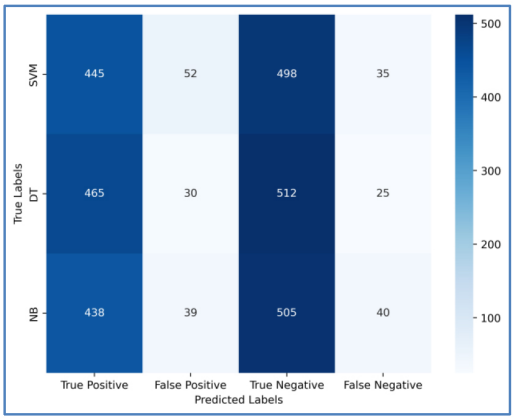

Figure 5

Confusion matrix for ML Model

Overall, SVM performed very well, and the trade-off between Precision and Recall was reasonable. The Decision Tree model had high Accuracy and Recall, making it useful in situations where it was important to capture positive examples. When reducing the number of false positives is crucial, Naive Bayes optimized for high Precision performs well.

5. Conclusion

This research presents a predictive analytics strategy for improving medication adherence by utilizing individual patients' demographic, health, and socioeconomic information. The models' in-depth examination of these factors yields individualized insights that help healthcare providers anticipate risks of non-adherence and design interventions to address them. Beyond simple prediction, advanced analytics in healthcare are helping to usher in a new era of preventative and individualized health care. Models have shown promise in predicting patients' compliance, but further development and validation are necessary before they can be used in practice. Furthermore, continuous monitoring is required due to ethical concerns around data protection and model interpretability. Still, the findings highlight the enormous potential of machine learning-driven predictive analytics in improving drug management, which should lead to better, patient-centered healthcare for everyone. To fully realize their potential in improving patient outcomes, future research should concentrate on making these models scalable, interoperable, and easily integrated into preexisting healthcare systems.

References

[1] A. Kalatzis, B. Mortazavi and M. Pourhomayoun, "Interactive Dimensionality Reduction for Improving Patient Adherence in Remote Health Monitoring," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, pp. 748-751 (2018).

[2] Y. Gu, A. Zalkikar, L. Kelly, K. Daly and T. E. Ward, "Predicting Injectable Medication Adherence via a Smart Sharps Bin and Machine Learning," 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, pp. 1-7 (2020), doi: 10.1109/WF-IoT48130.2020.9221014.

[3] A. Gupta and S. Joshi, "Emerging Trends and Scope of Healthcare Analytics," 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC), Bengaluru, India, pp. 93-101 (2022).

[4] Z. E. Rasjid, "Predictive Analytics in Healthcare: The Use of Machine Learning for Diagnoses," 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, pp. 1-6 (2021).

[5] C. K. Leung, T. H. Daniel Mai, N. D. Thong Tran and C. Y. Zhang, "Predictive Analytics to Support Health Informatics on COVID-19 Data," 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE), Kragujevac, Serbia, pp. 1-9 (2021).

[6] A. R. Reddy and P. S. Kumar, "Predictive Big Data Analytics in Healthcare," 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, pp. 623-626 (2016).

[7] N. D. Thong Tran, C. K. Leung, E. W. R. Madill and P. T. Binh, "A Deep Learning Based Predictive Model for Healthcare Analytics," 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), Rochester, MN, USA, pp. 547-549 (2022).

[8] M. Bende, M. Khandelwal, D. Borgaonkar and P. Khobragade, "VISMA: A Machine Learning Approach to Image Manipulation," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India (2023).

[9] M. M. Rahman et al., "Mining weighted frequent sequences in uncertain databases", *Information Sciences*, vol. 479, pp. 76-100 (2019).

[10] K. K. Roy et al., "Mining sequential patterns in uncertain databases using hierarchical index structure", PAKDD, pp. 29-41 (2021).

[11] M. Alnemari and N. Bagherzadeh, "Efficient deep neural networks for edge computing", IEEE EDGE, pp. 1-7 (2019).

[12] F. Khennou, C. Fahim, H. Chaoui and N. E. H. Chaoui, "A machine learning approach: Using predictive analytics to identify and analyze high risks patients with heart disease", *International Journal of Machine Learning and Computing*, vol. 9, no. 6, pp. 762-767 (2019).

[13] N. Hasan, "A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model", Internet of Things, vol. 11, pp. 100228, Sep. (2020).