The Efficiency of Machine Learning Algorithms in the Prediction of Drug Reactions in Clinical Settings

Christopher Agbonkhese

Department of Digital and Computational Studies/Bates College
Lewiston, ME, 04240, USA
agbonschris@gmail.com

Hettie Abimbola Soriyan

Department of Computer Science and Engineering, Obafemi Awolowo University
Ile-Ife, 7500, Nigeria
hasoriyan@gmail.com

Kolawole Mosaku

Department Mental Health, Obafemi Awolowo University
Ile-Ife, 7500, Nigeria
ckmosaku@oauife.edu.ng

Abstract

In the rapidly evolving landscape of healthcare, the efficient detection of drug reactions is of paramount importance to ensure patient safety and optimize treatment outcomes. This article presents a comprehensive study on the application of machine learning techniques for the early detection of drug reactions through the analysis of drug prescriptions in clinical settings. The study utilized a formulated model with classification and regression tree algorithm, iterative dichotomizer 3, gaussioan, naïve bayes, Bernoulli naïve bayes, multinomial naïve bayes, with adaptive boosting algorithm to extract valuable insights from health records and prescription data and predict the possible occurrence of adverse reactions from prescribed medications. A comparative analysis of the efficiencies of the various algorithms was carried out based on the computational learning theory. Among the myriad models scrutinized, the results showed that an ensemble comprising ID3, MultinomialNB, and AdaBoost emerged as a standout performer, consistently showcasing exceptional performance across multiple metrics.

Keywords: adverse drug reactions, artificial intelligence, machine learning, drug prescription, healthcare, prediction

1. Introduction

In contemporary healthcare, the precise and early identification of drug reactions remains a critical challenge, impacting patient safety and treatment outcomes [1]. Adverse reactions to medications pose substantial risks, often leading to increased morbidity, prolonged hospital stays, and elevated healthcare costs. Adverse Drug Reactions (ADR) are acknowledged factors contributing to morbidity and mortality universally [2][21]. It is estimated that ADRs represent the fourth leading cause of over 100,000 death in the United States alone, behind heart disease, cancer, and stroke [3][22]. Amidst this landscape, harnessing advanced computational methods, particularly through artificial intelligence (AI) and machine learning, has emerged as a promising avenue for transforming the detection and management of drug reactions [4][5]. The intersection of artificial intelligence and healthcare presents a realm of possibilities, particularly in analyzing drug prescriptions in clinical settings [6]. Understanding and predicting potential adverse reactions through data-driven approaches have garnered significant attention due to their potential to revolutionize the landscape of patient care [7][23]. This article focuses on elucidating the utilization of artificial intelligence methodologies to sift through vast datasets of drug prescriptions and patient reactions within clinical settings. Through systematic analysis and predictive modeling, this study endeavors to offer a proactive solution for identifying and managing drug reactions in clinical practice. The implications of this research extend beyond mere detection; it promises to pave the way for a more robust, preemptive healthcare system. Automating the process of detecting potential adverse drug reactions using AI-driven approaches not only enhances patient safety but also empowers healthcare professionals with actionable insights for informed decision-making.

The objective of this manuscript is to provide an in-depth exploration of the application of artificial intelligence in deciphering drug reactions from prescription data in clinical environments based on the computational learning theory [8][24].

2. Related Work

[9] undertook the development of data mining and machine learning models with the primary goal of predicting drug likeness and classifying drugs based on their associated diseases or organ categories. Their dataset, consisting of 762 compounds, was meticulously categorized into two primary groups: drugs (366 compounds) and nondrugs (396 compounds). The compounds were curated from [10], and the DrugBank database [11] was employed to establish the status of approved drugs. To assess the robustness of their prediction model, the compounds were thoughtfully partitioned into a training set (80%, comprising 610 compounds) and a test set (20%, encompassing 152 compounds). A parallel distribution of drugs (73 compounds) and nondrugs (79 compounds) was maintained in the test sets. This partitioning process was executed through independent selection procedures utilizing a representativeness function, as proposed by [12]. The methodology incorporated a simulated annealing optimization strategy to select a subset of objects, namely compounds that best represented the current database from which it was drawn. Subsequently, predictive models were constructed on the training set employing a 10-fold cross-validation approach and seven distinct methods. These models were meticulously tested on the test set. The prediction models were constructed using six diverse machine learning algorithms, encompassing decision trees, random forests (RF), support vector machines (SVM), artificial neural networks (ANN), k-nearest neighbors (k-NN), and logistic regression (LR). In each instance, classification models were established using the training set and subsequently employed to predict the activities, specifically drug status, of the test set compounds to validate the efficacy of the models. The implementation of these models was executed within the Weka framework, with evaluation metrics encompassing accuracy, sensitivity, specificity, and variance serving as performance benchmarks. [13] delved into the intricate realm of causality through the automated extraction of lexical patterns. The study aimed to derive the reliability of extracted lexical patterns in expressing adverse reactions to specific drugs by learning their respective weights. Notably, their method achieved an impressive ADR detection accuracy of 74% on an expansive manually annotated dataset comprising tweets from a social media platform. This dataset encompassed a standardized set of drugs and their associated adverse reactions. Importantly, their model exhibited proficiency in accurately discerning causality between drugs and adverse reaction-related events. However, it is imperative to underscore that while accuracy served as a performance metric, it may not provide a comprehensive assessment of their model's performance, warranting further evaluation using additional metrics. [14] embarked on an exploratory journey into the realm of social media mining for drug safety signal detection. Their pioneering work proposed the utilization of association mining and Proportional Reporting Ratios (PRR) to uncover valuable associations between drugs and adverse reactions. These associations were derived from the rich content contributed by users on social media platforms. In their experimental evaluation, ten drugs and five distinct adverse drug reactions were scrutinized. As a benchmark for assessing their techniques, they turned to the Food and Drug Administration (FDA) alerts. Their findings unveiled promising potential in employing metrics such as leverage, lift, and Proportional Reporting Ratio (PRR) for detecting adverse drug reactions that had been reported to the FDA. Importantly, PRR emerged as the standout performer among these metrics, showcasing its efficacy in identifying these critical drug safety signals. [15] introduced an innovative approach by proposing a Bayesian neural network method for generating signals related to adverse drug reactions (ADRs). Central to their work was the Bayesian Confidence Propagation Neural Network (BCPNN), renowned for its capacity to manage extensive datasets effectively. The BCPNN model exhibited robustness, particularly in handling imbalanced data and complex variables. Drawing from information theory, this tool proved ideal for uncovering drug-ADR combinations that exhibited high associations compared to the overall dataset or specific subsets thereof. Their study yielded compelling results, illustrating the potency of the BCPNN technique in early signal detection, exemplified by instances such as captopril-coughing. Moreover, the model demonstrated its ability to mitigate false positives, notably in cases where common drugs co-occurred with ADRs in the database, exemplified by scenarios involving digoxin and adverse reactions like acne or rash. Furthermore, the study conducted a routine application of the BCPNN to quarterly updates, revealing that a remarkable 1004 suspected drug-ADR combinations reached a confidence level of 97.5% difference from the dataset's generality. Significantly, among these combinations, 307 were identified as potentially serious ADRs, with 53 of them linked to novel drugs. This underscores the invaluable contribution of the BCPNN methodology to signal detection in the realm of adverse drug reactions. [16] introduced a pioneering approach aimed at large-scale prediction of adverse drug reactions (ADRs) by harnessing the collective knowledge of chemical, biological, and phenotypic properties associated with drugs. Their innovative method integrated a diverse range of data, including the phenotypic characteristics of drugs, their indications, known ADRs, chemical structures, biological properties, protein targets, and pathway information. A comprehensive study was conducted, focusing on the prediction of 1385 well-documented ADRs associated with 832 approved drugs. To achieve this, the authors employed five distinct machine-learning algorithms, subsequently comparing their individual performance. The

evaluation of their model was carried out through a rigorous fivefold cross-validation process, ultimately revealing that the support vector machine (SVM) algorithm outperformed its counterparts. Of particular note was the significance of phenotypic data, which emerged as the most informative component for ADR prediction. Furthermore, the augmentation of the baseline chemical information with biological and phenotypic features led to a substantial enhancement in the performance of the ADR prediction model. Notably, the model effectively anticipated ADRs associated with the withdrawal of drugs like rofecoxib and cerivastatin, underscoring its practical utility. These findings collectively emphasized the value of phenotypic drug information in the realm of ADR prediction. Additionally, the study underscored the potential for constructing diverse predictive models by combining chemical, biological, and phenotypic information derived from approved drugs. Such models exhibited the capability to detect clinically significant ADRs across both preclinical and post-marketing phases. [17] introduced a computational model tailored for the prediction of adverse cardiovascular drug reactions (CV ADRs). Their approach relied on a machine learning-based framework that seamlessly integrated various drug features. These encompassed biological aspects, such as drug transporters, targets, and enzymes, as well as chemical attributes like substructure fingerprints. Phenotypic characteristics, which encompassed therapeutic indications and other identified ADRs, were also integrated into the model. To distill the most relevant features, the authors employed the minimum redundancy maximum relevance approach. Additionally, they employed a synthetic minority oversampling technique to address dataset imbalance effectively during model training. Remarkably, the study yielded a total of 504 distinct computational models tailored for predicting 36 unique CV ADRs. Across all models, an impressive accuracy rate of approximately 90% was achieved. Notably, the models incorporating biological and chemical features demonstrated enhanced informativeness compared to those relying solely on chemical properties. These results underscored the high accuracy of the predictive models devised in this study. Furthermore, they emphasized the pivotal role of phenotypic information in drug ADR prediction. The research highlighted the potential for combining various drug properties to construct computational models suitable for early-stage prediction of potential ADRs during drug development. Nevertheless, it is important to note that while accuracy served as a performance metric, a comprehensive assessment of the model's performance would require additional metrics beyond accuracy alone. [18] introduced an intricate methodology for the systematic and automated discovery of adverse drug events (ADRs) from electronic medical records, harnessing the power of the Random Forest algorithm. Their approach involved the utilization of textual information extracted from a vast repository of 9.5 million clinical notes, complemented by prior knowledge regarding drug usages and previously identified ADRs. These data were treated as input variables, subjected to further processing to generate statistics utilized by a discriminative classifier. This classifier was instrumental in assigning probabilities to the association of specific drug-disorder pairs with valid ADRs. The ensuing list of potential ADRs, as identified by the classifier, underwent additional scrutiny through filtration for positive support. This validation process leveraged two independent and complementary data sources, bolstering the reliability of the identified associations. The authors assessed the methodology's effectiveness by evaluating the support for predictions within various curated data sources. Furthermore, they employed a manually curated, time-indexed reference standard, focusing on label change reactions. Remarkably, the classifier exhibited a robust performance, achieving an impressive area under the curve (AUC) score of 0.94 on a held-out test dataset. This classifier was applied to an extensive set of 2,362,950 possible drug-disorder pairs, comprising 1602 unique drugs and 1475 unique disorders, for which relevant data were available. This meticulous analysis yielded a collection of 240 high-confidence, well-supported drug-adverse reaction associations. Notably, 36% of these associations found additional support in external resources that were not accessible to the classifier. This methodology underscored the feasibility of systematic post-marketing surveillance for ADRs, utilizing electronic medical records as a foundational component of the learning healthcare system. [19] presented a cutting-edge model designed for the detection of potential adverse drug reactions (ADRs) using a Deep Neural Network (DNN). Their innovative DNN model was specifically engineered to serve two primary objectives: firstly, to identify potential ADRs associated with existing drugs, and secondly, to predict the possible ADRs that could emerge with the introduction of new drugs. The model's detection performance was enhanced through the incorporation of distributed representations of target drugs in a vector space. This approach was instrumental in capturing intricate relationships among drugs by employing a word-embedding technique to process extensive biomedical literature. Additionally, the authors introduced a mapping function designed to address scenarios where data for new drugs, absent from the existing dataset, needed to be accommodated. The results of their study revealed a robust overall performance, with a mean average precision at the top-10 achieved at 0.523, while the area under the receiver operating characteristic curve (AUC) score reached an impressive 0.844 for ADR prediction on the dataset. Notably, the model demonstrated a high level of effectiveness in identifying potential ADRs associated with existing drugs and predicting the possible ADRs that may emerge with new drug introductions. These findings underscored the model's effectiveness in the crucial task of ADR prediction, reinforcing its potential utility in the field of pharmacovigilance and drug safety monitoring. [20] introduced a robust model for predicting adverse drug reactions (ADRs) using machine learning techniques, with a significant emphasis on the formulation of a mathematical model using a combination of naïve bayes, decision tree, and adaptive boosting algorithms. The model was designed as a decision support tool for

clinical settings by systematically examining the relationships between prescribed medications and observed patient reactions.

3. Methodology

This section presents the research methodology adopted in this work. It begins with a description of the dataset that was used for the study the source of the data. It thereafter describes the various algorithms used for the formulation of the prediction model for drug reaction, the simulation of the model, the tools used, as well as the metrics for the evaluation of the model. Lastly, it presents a description of the architecture of the prediction model for the research.

3.1 Data Acquisition

The first objective of this research was to elicit relevant data on patients, which involved application and obtaining ethical clearance from the Obafemi Awolowo University Teaching Hospital Complex (OAUTHC) Ile Ife, in Osun State. A total of five hundred and eighteen (518) records were extracted from patients' case notes in the Psychiatric Department of the OAUTHC, this was done using case study technique on cases of admitted patients within nine years (that is, between November 2010 and November 2019). The extraction of data regarding cases of diagnosis and corresponding treatment for each admitted patient was carried out because there was no such data available in electronic format in that department as at the time of this study. The variables for which data was collected include age, sex, diagnosis, substance dependence, polypharcacy, prescribed drugs, dose, route of administration and drug reaction. A brief description of these variables is as shown in Table 1

Table 1: Brief Description of Variables

S/N	Feature	Description
1	Sex	A measure of the biological difference between a male and a female
2	Age	The length of time the patient has lived, measured in years
3	Diagnosis	Identification of the nature of illness or reason for admission
4	Polypharmacy	The concurrent use of multiple medications by a patient.
5	Substance dependent	Whether the patient is into any form of drug addiction
6	Prescribed drugs	The drugs that were prescribed for each patient (antipsychotics in this case)
7	Route of drug administration	The path by which a drug is taken into the body
8	Adverse reactions	Whether the patient experienced any adverse reaction in the course of the treatment after the drugs were administered

Some of the records in the case notes were incomplete, while others had patterns that could not be interpreted due to illegible handwritings. Hence the records were cleaned. In order to achieve meaningful prediction, the study required detail contents of the patient's case notes such as the reasons for which the patients were admitted in the hospital (captured in terms of diagnosis for each patient), treatments administered in the course of admission, as well as the outcomes for each treatment.

3.2 Data Pre-processing

The collected data was first pre-processed using data binning, one-hot encoding technique, and data normalization. Data binning was used for the age attribute conversion to categorize it into three bins. This was done firstly by specifying three equally sized bins, the bin array was built from minimum value to maximum value using the bin with the calculated, and labels were thereafter created as "Teenager", "Adult", "Elderly".

3.3 Model Formulation

Here, we formulate an ADR prediction model using stacking technique. The individual classification algorithms were trained based on the complete training set, after which the meta-classifier was be fitted based on the outputs from the base classifiers in order to improve the performance of the prediction model. This involved the use of decision tree algorithm and the naive bayesian (NB) algorithms as base classifiers while the adaptive boosting algorithm was used as a meta-classifier. The use of decision tree was based on their ability to convert large complex datasets into easy-to-understand output and suitability in handling binary data as mostly contained in the ADR dataset. Also, decision tree, through its variant (C5.4) provides for minimizing the error during classification by reducing entropy through the computation of the information gain in the dataset.

The choice for naive bayes was based on the fact that it works very well with binary data in classification task, while its variants provides for the enforcement of the conditional probability at different data points in order to improve the performance of the model. The choice for adaboost for was based on its ability to detect the point where the model misclassified the data and assign weight to those points in order to boost the performance of the final model, which is the meta-model. The following section presents the mathematical formulation of the prediction model:

Let *S* be the ADR dataset,
for the ADR dataset *S*, H(S) measures the amount of the uncertainty in the data as
Entropy

$$\mathrm{H(S)} = \sum_{\mathrm{c}\epsilon\mathrm{C}} -\mathrm{P(C)}\log_2\mathrm{P(C)} \qquad (1)$$

Where,
S: the current ADR dataset for which entropy is being calculated
C: the set of classes in the ADR dataset, *S*
P(C): the proportion of the number of elements in class *C* to the number of elements in set *S*.

Information gain IG(A) measures the difference from before to after the ADR dataset *S* is split on an attribute A. That is, how much uncertainty in ADR dataset *S* was reduced after splitting set *S* on attribute *A*

$$IG(A,S) = H(S) - \sum_{t\epsilon T} p(t)H(t) \quad (2)$$

Where,

$$S = \bigcup_{t=T} t \qquad (3)$$

Where:
H(S) is the entropy of dataset S from equation (2)
T is the subset created from splitting ADR dataset set *S* by attribute *A* such that
P(t) is the proportion of the number of elements in set *S*
H(t) is entropy of set *t*

Also, from Bayesian theorem, dealing with strong independence assumptions between predictors. The theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). The classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. That is, to say that the various attributes that predict the possibility of drug reactions are independent of the others. This assumption is useful when the number of instance, N is high and/or N is small, making (x|c) difficult to estimate. Even if the assumption does not hold, the model classification performance may still be good in practice because the decision boundaries may be insensitive to the specificities of the class-conditional probabilities p(xi|c); that is, variance is reduced because few parameters are required and the biased probability estimates may not matter since the aim is classification rather than accurate posterior class probability estimation.

$$p(c|x)\frac{P(X|C)P(C)}{P(X)} \qquad (4)$$

Where:
P(c|x): the posterior probability of class (target) given predictor (attribute)
P(c): the prior probability of class
P(x|c): the likelihood which is the probability of predictor given class
P(x): the prior probability of predictor

Let D be a training set of tuples and their association class labels. As usual, each tuple is represented by an n-dimensional attribute vector, $X = (x_1, x_2, \ldots, x_n)$, depicting *n* measurement made on the tuple from *n* attributes, respectively, $(A_1, A_2, \ldots, A_n)$.

Suppose there are m classes, $(C_1, C_2, \ldots, C_m)$, given a tuple, X, the classifier will predict that tuple X belongs to the class C_i if and only if

P (C_i|x) > P (C_j|x) for 1 <= j <= m; j ≠ I (5)

Thus, this maximize P(C_i|x). The class C_i for which P (C_i|x) is maximized represents the Maximum Posteriori Hypothesis.

By Bayes' theorem

P (C_i|x) = P ($x|C_i$) P(C_i)/P(x) (6)

As P (x) is constant for all classes, only P ($x|C_i$)P (C_i) need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is;

P (C_1) = P (C_2) = …… = P (C_m) (7)

Maximizing P($x|C_i$). Otherwise, this maximize P($x|C_i$)P(C_1). That the class prior probabilities may be estimated by:

P (C_1) = |C_1, D|/|D| (8)

Where

|C_1, D| is the number of training tuples of class C_1 in D.

Given datasets with many attributes, it would be extremely computationally expensive to compute P ($x|C_i$). In order to reduce computation in evaluating P ($x|C_i$), the naïve Bayes' assumption of class conditional independence is made.

As per the conditional independence assumption of Bayes theorem, the presence or absence of some parameters of a class is independent to the presence or absence of some other parameters, making each parameter's contribution independent to the final result. For instance, for a parameter P(ADR = "Yes") given "polypharmacy" = 'Value from Test Data' is independent of P(ADR = 'No') gives "polypharmacy" = 'Value from Test Data'. In similar way, the probabilities of all the parameters and their individual contribution to the final result in different variables can be calculated. To deal with the condition of zero probability values for some parameter, Laplace Correction will be used. In order to handle the imbalance nature of the sample data, and create a highly accurate prediction model, the study adopts the adaptive boosting algorithm as meta-classifier by focusing on difficult data points which might have been misclassified most by the decision tree algorithm and the Naive Bayesian network classifier, using an optimally weighted majority vote, α_t of these weak classifiers.

Given *m* labeled training examples (x_1, y_1)… (x_m, y_m) where the x_is are in some domain χ, and the labeled y_iЄ {-1, +1}. On each round *t*=1,…,T, a distribution D_t is computed over the *m* training examples, and a given weak learner or weak leaning algorithm will be applied to find a weak hypothesis $h_t : \chi \rightarrow \{-1, +1\}$, where the aim of the weak learner is to find a weak hypothesis with the least weighted error ϵ_t relative to D_t.. The final or combined hypothesis (classifier) H computes the sign of a weighted combination of weak classifies. That is to say that the final hypothesis H or classifier is computed as a weighted majority vote of the weak hypothesis β_t, where each is assigned weight α_t.

$$H(x) = sign(\sum_{i=1}^{n} \alpha_t \beta_t) \quad (9)$$

Where the weak hypothesis β_t are H(s) and P(c/x) gotten from equation (1) and equation (6) respectively. α_t is computed as:

$$\alpha t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \quad (10)$$

Equation (9) then can be written as:

$$H(x) = sign\left(\sum_{i=1}^{n} \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)\left(\left(\sum_{t \epsilon T} p(t)H(t)\right) + \left(\frac{p(x|c)p(c)}{p(x)}\right)\right)\right) \quad (11)$$

Where H(x) represent the final hypothesis, which produces the class prediction for ADR as ADR or a NO ADR.

4. Simulation and Evaluation of the ADR Prediction Model

This section presents a discussion on the simulate and evaluation of the performance of the ADR prediction model. This was done in view of understanding the behaviour of the ADR prediction model under various conditions. The process involves the adjustment of the input data ratios for training and testing of the ADR prediction model, using the hold-out cross-validation technique in order to avoid over-fitting by dividing the entire ADR dataset into two parts viz- training data and testing data. The model was trained using the training data and then evaluate on the testing dataset. The ratios of the training set to testing set of the ADR prediction model was on 70:30, 75:25 and 80:20 respectively. This allows the data to be first shuffled before performing the

training and testing at different data points. The evaluation was carried out using accuracy, precision, sensitivity, specificity and error rate as performance metrics, with special attention given to the precision and sensitivity of the model because the system under study contains imbalance data which was biased towards the negative cases of ADR, hence the quest for a solution that maximizes the precision and sensitivity of the ADR prediction model. The various metrics present different explanation of the behaviour of the ADR prediction model.

4.1: Framework of the Proposed ADR Prediction Model

The Architecture of the proposed model for predicting ADR is as shown in Figure 1. The patients' vital signs as well as diagnosis and corresponding drug prescriptions data were used to train the base classifiers. (Decision Tree and Naïve Bayes). The process considered the two variants of Decision Tree algorithm (Classification and Regression Tree (CART) and Iterative Dichotomiser 3 (ID3)), as well as the three variants of the Naïve Bayes (Gaussian Naïve Bayes (GaussianNB), Bernoulli Naïve Bayes (BernoulliNB) and the Multinomial Naïve Bayes (MultinoialNB)). An instance of the ADR prediction model takes one variant of the Decision Tree algorithm and one variant of the Naïve Bayes algorithm as base classifiers and use docking technique to fuse their outputs as inputs to the AdaBoost algorithm. The AdaBoost algorithm then assigns weights to the various data points, and higher weight to the misclassified points that were identified with higher prediction error. This was done for all possible combinations of the selected variants of the Decision Tree algorithm and all the selected variants of the Naïve Bayes algorithm, with their various outputs fused with AdaBoost algorithm. The final output from the AdaBoost algorithm at any one instance, predicts the occurrence or non-occurrence of the ADR result as depicted in Figure 2. The training and testing of each of the instances of the ADR prediction model was done on the data ratios of 70:30, 75:25, and 80:20 respectively, in order to have an in-depth understanding of the performance behaviour of the ADR prediction model.

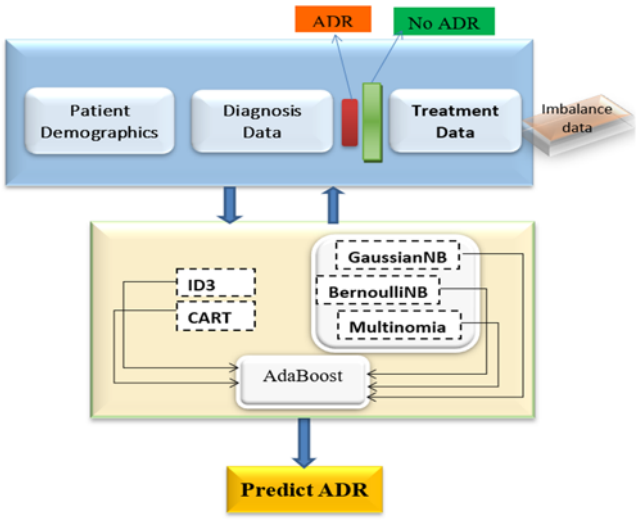

Figure 1: Proposed Framework for ADR Prediction

5. Results and Discussion

Computational learning theory is mainly concerned with how to precisely formulate and address questions regarding the performance of different learning algorithms so that careful comparisons of both the predictive power and the computational efficiency of alternative learning algorithms can be made [8]. This research utilised different machine learning algorithms to formulate a mathematical model to precisely predict the occurrence of adverse drug reactions in the midst of imbalance data. The study also simulated the model while also considering the different variants of the algorithms which attempt to explore the possibility of building a predictive model that would produce the best output among the variants of the selected machine learning algorithms. In order to have a meaningful ground for comparing the predictive power of each instance of the predictive models, the classifiers were first used individually to build the prediction model which were simulated at varying degrees of input for training and testing of the learning models. The careful comparison of the various models was based on their performance metrics (accuracy, precision, sensitivity, specificity and error rate) which were computed from the various confusion matrixes.

Table 2: Summary of Results from Decision Tree and Naïve Bayes

	Ratio	ID3	CART	GNB	BNB	MNB
Accuracy (%)	60/40	79.6	82.0	74.1	77.1	80.8
	70/30	84.2	85.1	84.3	84.3	85.1
	75/25	88.8	85.1	78.2	84.3	81.1
Precision (%)	60/40	56.2	62.5	71.9	68.7	75.0
	70/30	58.3	62.5	75.0	70.8	75.0
	75/25	65.0	62.5	70.0	70.8	75.0
Sensitivity (%)	60/40	48.6	54.0	41.1	44.9	51.1
	70/30	60.8	62.5	58.1	58.6	75.0
	75/25	72.2	62.5	46.7	58.6	55.5
Specificity (%)	60/40	85.0	86.9	74.6	79.2	82.3
	70/30	90.7	90.7	86.6	87.6	87.6
	75/25	93.8	90.9	80.2	87.6	82.7
Error Rate	60/40	0.203	0.179	0.259	0.220	0.191
	70/30	0.157	0.148	0.157	0.157	0.149
	75/25	0.118	0.148	0.128	0.157	0.157

In the face of imbalance data, the predictive power of the algorithms are with emphasis on results that maximizes the precision and sensitivity of the classifiers and at the same time minimizes the error rate of the classifier. Classification and Regression Tree (CART), Iterative Dichotomizer 3 (ID3), Gaussian Naïve Bayes (GaussianNB), Bernoulli Naïve Bayes (BernoulliNB) and the Multinomial Naïve Bayes (MultinomialNB) classifiers were used individually. The simulation results as shown in Table 2

From the results, The ID3 algorithm demonstrated a consistent trend of enhancement across all measured metrics as the training set increased. Notably, its precision and sensitivity achieved their apex at the 75:25 ratio, suggesting an advantageous utilization of a larger portion of the dataset for training purposes. Conversely, the CART model showcased an optimal performance plateauing at the 70:30 ratio, implying a potential saturation point where increased data beyond this threshold may not yield substantial performance gains.

The Naïve Bayes classifiers (GaussianNB, BernoulliNB, MultinomialNB) exhibited diverse patterns in response to changing training ratios. GaussianNB revealed erratic fluctuations across precision, sensitivity, and specificity, showcasing a delicate balance affected by training data distribution. BernoulliNB displayed a stable and optimal performance at the 70:30 ratio, maintaining consistent precision, sensitivity, and specificity levels, indicative of its reliability within that specific training range. MultinomialNB, while outperforming other classifiers overall, faced a slight decline in performance metrics as the training set expanded to 75:25, suggesting a potential limitation in accommodating larger training proportions.

Upon thorough analysis, it becomes evident that the MultinomialNB classifier emerged as the most robust model, showcasing superior performance across multiple metrics, particularly at the 70:30 training ratio. However, this finding does not discount the significance of choosing the appropriate model and training ratio based on the specific requirements of the task at hand. While larger training datasets generally benefited model performance, the saturation point observed in some models at certain ratios warrants careful consideration.

The analysis of the individual results showed that MultinomialNB classifier exhibited commendable performance, achieving the best balance of precision and sensitivity at the 70:30 ratio, the choice of an ideal model and training ratio should be contingent upon the application's unique demands. This comprehensive analysis underscores the importance of judiciously selecting both model types and training ratios to optimize predictive performance while considering the nuances of dataset characteristics and desired performance metrics.

However, in order to handle the imbalance data, boosting method was applied using an ensemble of the variants of Decision Tree and the variants of Naïve Bayes algorithms as base classifiers and AdaBoost algorithm as meta-classifier. The idea of boosting originated from the computational learning theory, upon which this research was based. The quest here is to see that the error of the combined classifiers on the training data approaches zero very quickly as more and more iterations are performed. The simulations integrating various Decision Tree variants

(such as ID3 and CART) and Naïve Bayes classifiers (like MultinomialNB, GaussianNB, and BernoulliNB) with AdaBoost present valuable insights into ensemble learning's impact on predictive modeling in healthcare, particularly in adverse drug reaction (ADR) prediction.

The experiments consistently show that integrating AdaBoost with different algorithms leads to enhanced predictive accuracy as the training set grows. This trend suggests that a larger volume of training data contributes positively to overall model accuracy.

Across the combinations, there was a notable interplay between precision and sensitivity concerning the training ratio. Some combinations exhibit peak precision and sensitivity at specific training ratios, indicating the importance of finding the optimal balance between correctly predicting positive instances (sensitivity) and avoiding false positives (precision). The specificity of the models, representing their ability to accurately predict negative instances, demonstrates an increasing trend with more extensive training sets. Simultaneously, the error rates consistently reduce, indicating improved overall model performance and reliability. Different combinations display diverse behaviors concerning their accuracy, precision, sensitivity, and specificity. While some combinations exhibit clear patterns of improvement with increasing training data, others show fluctuations or plateauing in certain metrics, highlighting the significance of selecting the appropriate combination for optimal performance. Among the combinations tested, the ensemble of ID3, MultinomialNB, and AdaBoost stands out with its remarkable accuracy of 95.0%. This combination showcases balanced precision and sensitivity, making it highly suitable for handling imbalanced data in clinical settings. Healthcare Decision Support These findings have significant implications for healthcare decision-making. The ensemble models' accuracy and ability to handle imbalanced data suggest their potential in assisting clinicians in predicting adverse drug reactions more accurately, thereby improving patient care.

Table 3. Result summary from ADR prediction model using DT and NB with AdaBoost algorithms

Performance Metrics	Data Ratio	Adaboost ID3 GNB	Adaboost ID3 BNB	Adaboost ID3 NB	Adaboost CART GNB	Adaboost CART BNB	Adaboost CART MNB
	60/40	87.0	87.0	92.6	85.8	88.8	90.7
Accuracy	70/30	88.4	88.4	94.2	86.7	80.7	92.6
(%)	75/25	91.1	91.1	95.0	89.1	89.1	94.1
Precision	60/40	81.2	81.2	84.4	75.0	75.0	81.3
(%)	70/30	79.1	79.1	87.5	75.0	75.0	83.3
	75/25	75.0	75.0	85.0	75.0	75.0	80.0
Sensitivity	60/40	63.3	63.3	79.4	61.6	61.5	76.5
(%)	70/30	67.9	67.9	84.0	64.3	64.3	80.0
	75/25	82.6	82.6	89.5	71.4	71.4	88.0
Specificity	60/40	88.5	88.5	94.6	88.5	83.5	93.1
(%)	70/30	90.7	90.7	95.9	89.9	89.9	94.8
	75/25	95.0	75.1	97.5	92.6	92.6	97.5
Error Rate	60/40	0.129	0.129	0.074	0.142	0.142	0.093
	70/30	0.016	0.116	0.058	0.132	0.132	0.074
	75/25	0.089	0.089	0.049	0.109	0.098	0.059

The precision of models, representing the proportion of true ADRs among those predicted as ADRs, showed fluctuations across models and training ratios. Similarly, sensitivity, which measures the proportion of actual ADRs correctly predicted, varied, indicating the models' differing abilities to detect true ADR occurrences. Furthermore, specific combinations like CART, MultinomialNB, and AdaBoost displayed noteworthy precision and sensitivity at a 70:30 ratio, signifying a balanced performance with a moderate training dataset.

Generally, while certain models and combinations demonstrated consistent improvement in performance metrics with increased training data, others showed fluctuating trends. The effectiveness of these models in predicting ADRs represent the complexity of handling imbalanced datasets, where a careful balance between precision, sensitivity, and accuracy is pivotal. The remarkable performance of the ID3, MultinomialNB, and AdaBoost

ensemble suggests its potential for practical application in clinical settings for ADR prediction. However, the varying performance across different models also emphasizes the importance of considering the context, dataset characteristics, and the trade-offs between different evaluation metrics when selecting the most appropriate model for real-world implementation.

Conclusion

The research embarked on a meticulous exploration of diverse machine learning models to predict Adverse Drug Reactions (ADRs) within imbalanced datasets, employing rigorous evaluation metrics encompassing accuracy, precision, sensitivity, specificity, and error rates. This comprehensive analysis aimed to discern the strengths and weaknesses of each model in handling the intricacies of imbalanced data, specifically in the realm of ADR prediction. Among the myriad models scrutinized, an ensemble comprising ID3, MultinomialNB, and AdaBoost emerged as a standout performer, consistently showcasing exceptional performance across multiple metrics. Notably, this ensemble demonstrated an outstanding accuracy rate of 95.0% when trained with a 75:25 ratio of data, exhibiting a remarkable ability to discern and predict ADR occurrences accurately. This finding is promising for its potential application within clinical settings, signifying the ensemble's robustness in identifying adverse reactions to pharmaceutical interventions, thus enabling proactive management and intervention strategies Furthermore, the research elucidated a critical facet in model performance: the influence of training dataset sizes on predictive capabilities. While some models exhibited steady and incremental improvements in accuracy, precision, sensitivity, and specificity metrics with increased training data, others displayed varying trends or less pronounced patterns. This observation shows the importance of optimizing models through sufficient and balanced training datasets, an aspect particularly relevant in the context of imbalanced datasets characteristic of ADR prediction scenarios. The study therefore advocated for the adoption of diverse techniques and ensemble approaches to effectively address the challenges posed by imbalanced datasets in ADR prediction. Essentially, it emphasized the efficacy of employing resampling methods within ensemble classifiers, highlighting their potential to mitigate data imbalance issues and enhance the predictive accuracy of models in clinical applications. Furthermore, his research significantly contributes to the burgeoning field of machine learning in healthcare by not only identifying a superior model ensemble for ADR prediction but also by emphasizing the nuanced interplay between dataset characteristics and model performance. It serves as a foundational guide for practitioners and researchers navigating the complexities of predictive analytics in clinical settings, and offering insights into optimal model selection and the importance of dataset balance and diversity.

References

[1] White, R. W., et al. (2016). Early identification of adverse drug reactions from search log data. J. Biomed. Inform., 59, 42-48. https://doi.org/10.1016/j.jbi.2015.11.005
[2] Hallas, J., Harvard, B., & Gram, L. F. (1990). Drug related hospital admissions: the role of definitions and intensity of data collection and possibility of prevention. Journal of Internal Medicine, 228(2), 83-90.
[3] Null, G., Dean, C., Feldman, M., & Rasio, D. (2005). Death by medicine. J. Orthomolecular Med., 20(1), 182-196.
[4] Morrow, E., et al. (2023). Artificial intelligence technologies and compassion in healthcare: A systematic scoping review. Front. Psychol., 13, 971044. https://doi.org/10.3389/fpsyg.2022.971044
[5] Agbonkhese, C., & Soriyan, H. A. (2019). A Machine Learning Approach for Predicting Drug Reactions from Patients' Case Notes. Proceedings of International Conference of the Application of Information Communication Technologies to Teaching, research and Administration (AICTTRA), 45–48.
[6] Agbonkhese, C., Soriyan, H. A., & Oyelami, O. (2016). Fuzzy-based Dosage Model for Aqueous decoction of Adansonia Digitata for the Management of Sickle Cell Anaemia Patients in African Traditional Medicine. Nigerian Journal of Natural Products and Medicine (NJNPM), 20.
[7] Stoumpos, A. I., Kitsios, F., & Talias, M. A. (2023). Digital Transformation in Healthcare: Technology Acceptance and Its Applications. Int J Environ Res Public Health, 20(4), 3407. https://doi.org/10.3390/ijerph20043407
[8] Anthony, M., & Biggs, N. (1992). Computational learning theory, Cambridge tracts in theoretical computer science (30). Cambridge University Press.
[9] Yosipof, A., Guedes, R. C., & García-Sosa, A. T. (2018). Data Mining and Machine Learning Models for Predicting Drug Likeness and Their Disease or Organ Category. Frontiers in Chemistry, 6(162), 1-16
[10] García-Sosa, A. T., & Maran, U. (2013). Drugs, non-drugs, and disease category specificity: organ effects by ligand pharmacology. SAR QSAR Environ. Res., 24(4), 319-331.
[11] Wishart, D. S., et al. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res., 34(12), 1.
[12] Yosipof, A., & Senderowitz, H. (2014). Optimization of molecular representativeness. J. Chem. Inf. Model., 54(9), 67–77.
[13] Bollegala, D., et al. (2018). Causality Patterns for Detecting Adverse Drug Reactions from Social Media: Text Mining Approach. JMIR Public Health and Surveillance, 4(2). https://doi.org/10.2196/publichealth.8214
[14] Yang, C. C., et al. (2012). Social media mining for drug safety signal detection. In International Workshop on Smart Health and Wellbeing (pp. 33-40).
[15] Bate, A., et al. (1998). A Bayesian neural network method for adverse drug reaction: Signal generation. European Journal of Clinical Pharmacology, 54(4), 315-321.
[16] Liu, M., et al. (2012). Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. J. Am. Med. Inform. Assoc., 19(1), 28-35.
[17] Jamal, S., et al. (2019). Computational models for the prediction of adverse cardiovascular drug reactions. J. Transl. Med., 17(1), 171.
[18] Wang, G., et al. (2015). A method for systematic discovery of adverse drug events from clinical notes. Med Inform Assoc, 22(11), 93–

102.
[19] Wang, C. S., et al. (2019). Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model. J. Med. Internet Res., 21(2), 2196-11016.
[20] Agbonkhese, C., et al. (2023). Formulation of A Computational Model for Predicting Drug Reactions Using Machine Learning. International Journal of Computer, 49(1).
[21] Shepherd, G., et al. (2012). Adverse drug reaction deaths reported in United States vital statistics. The Annals of Pharmacotherapy, 46(2), 169–175.
[22] Pirmohamed, M., et al. (2004). Adverse drug reactions as cause of admission to hospital: prospective analysis of 18,820 patients. Brit. Med. J., 329(7456), 15-19.
[23] Zheng, T., et al. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform, 97, 120-127.
[24] Peng, Y. J., et al. (2014). Ensemble-based hybrid probabilistic sampling for imbalanced data learning in lung nodule CAD. Computerized Medical Imaging and Graphics, 38(3), 137–150.

Conflict of Interest: The authors have no conflict of interest

Authors Profile

Dr. Christopher Agbonkhese has a Bachelor of Science (BSc) Degree in Computer Science, Master of Science (MSc) in Computer, and Doctor of Philosophy (PhD) degree in Information. He is currently a Visiting Assistant Professor in the Department of Digital and Computational Studies, Bates College, Lewiston, USA. His research encompasses Health informatics and data analytics, where he employs techniques like machine learning and data mining to construct models for addressing critical challenges especially in the health domain.

Professor Hettie Abimbola Soriyan is a distinguished scholar with a comprehensive academic background, holding BSc, MSc, and PhD degrees in Computer Science. This extensive educational foundation underpins her expertise and authority in the field. As the current Head of the Department at Kings University, Professor Soriyan showcases a blend of academic prowess and practical insight, specializing in Information Systems, specifically within the domain of Health Information Systems. Her dedicated focus and extensive qualifications mark her as a notable figure in both academia and the advancement of healthcare technologies.

Professor Kolawole Mosaku is a distinguished figure in the field of psychiatry, with a career steeped in both academia and clinical practice. His journey began with a foundation in medicine, earning his Bachelor of Medicine and Bachelor of Surgery (MBBS) before delving deeper into the intricacies of mental health. Dr. Kolawole pursued further studies, becoming a Fellow of the Faculty of Psychiatry at the National Postgraduate Medical College of Nigeria (FMCPsych). His thirst for knowledge led him overseas, where he acquired a Master of Public Health (MPH) from the University of Liverpool in the UK, broadening his perspective on global mental health dynamics. His pursuit of excellence culminated in the attainment of a Doctor of Medicine (MD) from the National Postgraduate Medical College of Nigeria. Currently positioned as a Professor of General Adult Psychiatry and serving as a Consultant Psychiatrist at the esteemed Obafemi Awolowo University Teaching Hospital in Ile-Ife, Osun State, Nigeria, Dr. Kolawole embodies a dedication to both the academic and practical aspects of psychiatry. Within the hallowed halls of academia, he imparts his wealth of knowledge to aspiring medical minds, nurturing the next generation of mental health professionals.