Drug Sentiment Analysis using Machine Learning Classifiers

Mohammed Nazim Uddin[1]*, Md. Ferdous Bin Hafiz[2]
Sohrab Hossain[3]
Department of Computer Science and Engineering
School of Science, Engineering and Technology
East Delta University, Chattogram, Bangladesh

Shah Mohammad Mominul Islam[4]
Department of Electrical and Electronic Engineering
School of Science, Engineering and Technology
East Delta University
Chattogram, Bangladesh

Abstract—In recent times, one of the most emerging sub-dimensions of natural language processing is sentiment analysis which refers to analyzing opinion on a particular subject from plain text. Drug sentiment analysis has become very significant in present times as classifying medicines based on their effectiveness through analyzing reviews from users can assist potential future consumers in gaining knowledge and making better decisions about a particular drug. The objective of this proposed research is to measure the effectiveness level of a particular drug. Currently most of the text mining researches are based on unsupervised machine learning methods to cluster data. When supervised learning methods are used for text mining, the usual primary concern is to classify the data into two classes. Lack of technical terms in similar datasets make the categorization even more challenging. The proposed research focuses on finding out the keywords through tokenization and lemmatization so that better accuracy can be achieved for categorizing the drugs based on their effectiveness using different algorithms. Such categorization can be instrumental for treating illness as well as improve one's health and well-being. Four machine learning algorithms have been applied for binary classification and one for multiclass classification on the drug review dataset acquired from the UCI machine learning repository. The machine learning algorithms used for binary classification are naive Bayes classifier, random forest, support vector classifier (SVC), and multilayer perceptron; among these machine learning algorithms, linear SVC was used for multiclass classification. Results obtained from these four classifier algorithms have been analyzed to evaluate their performances. The random forest has been proven to have the best performance among these four algorithms. However, multiclass classification was found to have low performance when applied to natural language processing. On the contrary, the applied linear SVC algorithm performed better for class 2 with AUC 0.82 in this research.

Keywords—*Machine Learning Algorithms; natural language processing; drugs sentiment analysis; text mining*

I. INTRODUCTION

Among many research dimensions in Natural Language Processing (NLP), sentiment analysis has become one of the promising fields of research in the recent century [1] [2]. A wide range of research domains has been covered by sentiment analysis, i.e. economy, polity, and medicine. In the pharmaceutical industry, large volumes of online user's views are evaluated automatically in order to obtain useful information about the efficacy and side effects of pharmaceuticals that could be utilized to enhance pharmacovigilance systems. Throughout the years, Sentiment analysis techniques have grown significantly in the last decade, evolving from basic rules to advanced machine learning techniques like deep learning, which has become a prominent technology in many NLP tasks. This triumph is not lost on sentiment analysis. Besides, several machine learning systems have recently been shown to be better than previous methods. These methods have achieved impactful results on standard sentiment analysis datasets [3] [4].

Various aspects such as 'medical condition', 'treatment procedure', etc., use the medical sentiment as a research study that directly impacts the users' health conditions. Any sort of progress or deterioration can be identified by analyzing patient status periodically. The medical condition can be expressed implicitly or explicitly. Mentioning the symptoms is a part of the implicit sentiment in the medical context. For example, consider the statement: 'I recently started Lexapro 3 days, I'm on extreme weight losses'. The term" weight losses here do not reflect a negative sense; however, it implies the negative medication side effect, where sentiment is defined as negative in the preceding statement. Hence, for making correct interpretations, additional information is required.

On the contrary, analyzing the health conditions is relatively much easier in the case of explicit sentiment. For instance, considering the statement, "I recently started Lexapro 3 days, I'm absolutely lost I feel weak and shaky every day". The words absolutely 'lost', 'weak', and 'shaky' are used to describe symptoms in this statement. Deciding about patients' medical issues is an important aspect, specifically when they learn from other patients' experiences, i.e., choosing a hospital, clinic, and medication [5]. Hospitals gain from this information since it allows them to understand better and address the interests and concerns of their patients. The experience covered with sentiment analysis and passions are being shared by the patients; sentiment analysis is being taught by the power of this type of experiment since this type of study identifies people's sentiment about a topic as well as its characteristics. The medical material available on the internet is completely free. Manually analyzing such a large volume of data is ineffective because of its existence in large volumes. Assessed examinations are denoted as positive or negative, for the most part, based on the pre-programmed acceptance of extreme suppositions. The online and traditional

*Corresponding Author.

review methods are supplanted by notion investigation nowadays, which is led by organizations for finding a broad conclusion regarding their products and service. As a result, their marketing approach and product awareness increase, and user management improve. It is quite imperative to be broken down since a tremendous amount of content is available online. That includes deep comprehensions of standard dialects are included in the programmed examination of this data. In our everyday life, thoughts and sentiments play an essential role. Basic leadership, learning, correspondence, and mindfulness in human circumstances are assisted. Socially produced regional substances are becoming prevalent in online life; hence the importance of dealing with and comprehending vernacular content is growing. Existing materials notwithstanding, such as nearby sayings, Myths, and fables are unearthed, widely disseminated on the internet.

Compared to reviews of other products, drug reviews are investigated less. When analyzed, drug reviews are primarily utilized to categorize a particular drug as a positive or a negative one as multi-class classification from text mining can be unyielding. The proposed research facilitates the categorization of drugs not into two categories rather into five classes based on their effectiveness. Such an outcome can be beneficial for both consumers and manufacturers to understand the effectiveness of drugs as well as whether a particular drug has any significant side-effect.

This paper is organized as follows: Section II illustrates literature review for sentiment analysis, Section III explains the algorithms those classify the sentiment of the drugs, Section IV portrays the findings of the study, and Section V demonstrates the contributions of the proposed research.

II. LITERATURE REVIEW

Supervised machine learning methods used by Twitter datasets, such as support bigram, vector machines and unigram, were analyzed by a research study led by Balahur (2013) [6]. Following the applications of these approaches to Twitter data, the results indicated that methods of unigram and bigram support vector machines are outshined. Emotive words, modifiers and unique tags were included in these results, enhancing the performance rating of emotions. Another study conducted by Jianqiang et al. [7] (2018) presented an approach that is word embedded using unsupervised learning as a base. This suggested technique makes use of hidden contextual semantic connections and characterization between words and tweets. The

characteristics of mood polarity and n-gram are combined with the score of the embedded word to structure. A deep convolutional neural network was used to include a collection of emotional characteristics. Facebook, Instagram, and Twitter are a few examples of social media platform that helps to generate data and circulate content quickly. The amount of hate utterances has risen significantly while circulating content related to a particular topic. To filter these sorts of utterances, a research study presented by Schmidt and Wiegand [8] suggested a filtering tool for natural language processing. According to the results, character-level strategies are superior to token-level ones. The authors' methodology demonstrated that using a lexical list of resources to rank them might be beneficial when utilized with others. Based on K-means and cuckoo searching methods, Pandey et al. [9] suggested a unique metaheuristic approach. The best feasible cluster heads are found using this method based on the Twitter dataset's emotional subject material. Wang and Li [10] categorized the changed text algorithms to anticipate motions in image data for the sentiment analysis. Textual and visual features for labelling emotions inside an image are unsuitable for the forecast, according to their technique. The authors conducted experiments on two datasets and found that the recommended technique outperforms current methods in terms of accuracy. Unique research on Hierarchical Deep Fusion (HDF) emotional analysis methodology was studied by Xu et al. [11]. The relationship between the properties of text, images and sentimental content has been analyzed in the proposed model. The authors combined visual content with textual content using three-level Hierarchal Long Short Term Memory (H-LSTM) to investigate the inter-modal association of text and image at various levels. Some of the most widely applied machine learning and deep learning algorithms have been described in Table I.

Most of the above-mentioned researches primarily focus on unsupervised learning method. Compared to other product reviews, number of researches conducted on drug reviews is significantly low. One of the key challenges with dataset similar to the one used in this research is the lack of technical terms. A few researches that utilize supervised learning methods, perform binary classification as multi-class classification using the existing machine learning algorithms have been proven to be challenging. In this research, tokenization and lemmatization identify the key words. Also multi-class classification has been performed unlike the researches mentioned in this section.

TABLE I. EXITING ALGORITHMS ACCURACIES

Algorithm	Description
Neural Network	The neural network approach technique has a very high performance. It is a widely used technique for sentiment analysis and is capable of detecting all possible interactions between attributes. It is effective for dealing with a nonlinear connection between variables that is complex. The main disadvantage is that it takes longer to compute than other algorithms [12].
Naive Bayes	It is a Bayes' theorem based probabilistic classifier. Researchers use this method less commonly to make a prediction. The primary advantage is that it is scalable in comparison to other algorithms [13].
Support Vector machine	It is also a way of supervised machine learning for classification and regression analysis. When dealing with small datasets, the Support Vector Machine is very effective. It is more efficient than other approaches of classification and regression [14].
Decision Tree	Decision trees are simple but extensively used tools for prediction. IF-THEN rules can be simply converted from a decision tree. According to a previous study, prediction and forecast can be done using a decision tree. It can predict drug sentiment with low accuracy [15].
K-Nearest Neighbor	It is a popular pattern recognition method that is less non-parametric. It has the ability to utilize both regression and classification. It provides the finest performance and precision. It is the most basic machine learning algorithm [16].
AdaBoost	AdaBoost combines a number of poor classifiers to create a powerful one by iteratively retraining and weighing the classifiers depending on their accuracy [17].
Logistic regression	The log odds of the dichotomous result can be modelled as a linear combination of the predictor factors using this method [18].
Convolutional neural network	A feed-forward neural network that has been trained to extract key characteristics for the prediction job at hand. Nonlinear functions are used to filter features via convolutions. The dimensionality can then be reduced via pooling [19].
Maximum entropy	The greatest entropy concept is used to create a probabilistic classifier [20].
Conditional random fields	Given an observation series based on a conditional probability distribution across label sequences, this approach for segmenting and labelling structured data can be used [21].

III. METHODOLOGY

The dataset that used in this experiment has been collected from the UCI repository [22]. Tokenization and lemmatization were performed on the data after collecting the dataset. Four machine learning algorithms have been applied to the dataset for binary classification. The classes for binary classification are class 0, and class 1, where class 0 represents the effective drugs and class 1 indicates the ineffective drugs. The algorithms used for binary classification are naïve Bayes classifier, random forest (RF), support vector classifier (SVC) and multilayer perceptron (MLP). Linear SVC has also been applied to the dataset for multiclass classification.

Table II represents the classes for multiclass classification along with counts for each class. Class 0 represents highly effective drugs with a count of 1741, whereas class 1 represents considerably effective drugs with a count of 1238. Class 2 represents moderately effective drugs with a count of 529. In addition, class 3 represents marginally effective drugs with a count of 329. Besides, class 4 represents ineffective drugs with a count of 263.

A. Tokenization

In the case of natural language processing, a series of well-defined processes need to be carried out for analyzing the text. One of the primary processes is known as tokenization which plays a crucial role in the efficiency and correctness of the entire analysis. Tokenization refers to splitting the text into meaningful smaller units known as tokens. In most cases, tokens are identified as words or word sequences. Tokens are usually recognized when a white space character is encountered just after scanning a token. Preprocessing of text for punctuation removal and uppercase to lowercase conversion are often involved with the tokenization process [23].

B. Lemmatization

One of the most essential and elementary processes associated with natural language processing (NLP) is lemmatization. The base or dictionary form of a word is called a lemma. The term lemmatization refers to the morphological conversion of a word that exists in the textual form of the dataset to its lemma. The basic idea for this conversion is the removal of the declension from the end part of the word. In the case of a verb, lemma represents the infinitive form; in the case of a noun, lemma represents the singular form, and in the case of adjective or adverb, lemma represents the positive form. For example, the lemma for the word 'better' is 'good'. The lemma for the word 'brought' is 'bring'. Lemmatization can be perceived as a normalization method in which various morphological variants of a word are analyzed as a single item by mapping them into the same underlying lemma. As the aggregate number of specific terms are reduced, the complexity for analyzing the text is significantly decreased and thus, the overall time and resource utilization is improved. Lemmatization is widely applied for preprocessing the text in information retrieval, document clustering, sentiment analysis, etc. [24].

Table III shows tokenization and lemmatization of the feature benefitsReview.

TABLE II. DATA STATISTICS FOR CLASSIFICATION

Class	Effectiveness	Count
0	Highly Effective	1741
1	Considerably Effective	1238
2	Moderately Effective	572
3	Marginally Effective	329
4	Ineffective	263

TABLE III.   DATA SET LEMMATIZATION

effectiveness	benefitsReview	benefitsReview_lemma
1	Slowed the progression of left ventricular dysfunction...	slowed progression left ventricular dysfunction...
1	Although this type of birth control has more cramp...	although type birth control con pro help cramp...
1	I was used to having cramps so badly that they...	used cramp badly would leave balled bed least ...
1	The acid reflux went away for a few months aft...	acid reflux went away month day drug heartburn...
1	I think that the Lyrica was starting to help w...	think Lyrica starting help pain side effects se....

Fig. 1.   WordCloud for Medication Dataset.

Fig. 1 represents the WordCloud for the dataset that we have used in this research. WordCloud is a data visualization technique to highlight the most relevant and the most frequent words. The size indicates the frequency of each word. The most frequent word in the dataset is 'take'.

C. Chi-Squared for Feature Selection

Chi-square scores have been calculated for the most utilized twenty words or terms from the dataset. There exist two possible classes (positive and negative). The Chi-square test can be utilized for evaluating the significance of a given the word to discriminate between the classes [25].

Fig. 2 shows chi-square scores for most utilized words. The word with the highest score is 'none', and the second-highest score is the benefit.

D. Sentiment Classification

After tokenization and lemmatization, we use RF, SVC, MLP, and NB to find the polarity of the drugs.

1) Naïve bayes classifier: Naïve Bayes Classifier is a well-established machine learning algorithm used for the

classification of data. For its ability to work in a time-efficient manner and to disregard noise or irrelevant data, and its simplicity to implementation, Naïve Bayes Classifier is a viral algorithm for test classification in the fields like spam or fake news detection, sentiment analysis, etc. This algorithm is based on a theorem known as Bayes Theorem, first invented by a British scientist named Thomas Bayes [26]. The idea of the Bayes Theorem is to calculate the probability of an event based on any previous knowledge or conditions that have an impact on the event.

The standard formula for the theorem is:

$$P(X|Y) = \frac{P(Y|X)P(Y)}{P(X)} \tag{1}$$

In equation 1,

$P(X|Y)$=Posterior probability

$P(Y|X)$=Likelihood

$P(Y)$=Prior probability

$P(X)$=Marginal probability

As there are multiple aspects to be considered to classify the data, multiplication rules need to be applied, and the equation becomes:

$$P(X|Y_1, Y_2, \ldots Y_n) = \frac{P(X)*P(Y_1|X)P(Y_2|X)\ldots P(Y_n|X)}{P(Y_1)P(Y_2)\ldots P(Y_n)} \tag{2}$$

In equation 2, X is a class variable, and Y is a dependent feature vector $(Y_1, Y_2, \ldots Y_n)$

As the denominator of the right-hand side of the equation will remain the same for all data in a particular dataset, it can be written as:

$$P(X|Y_1, Y_2, \ldots Y_n) \propto P(X) * P(Y_1|X)P(Y_2|X)\ldots P(Y_n|X) \tag{3}$$

Fig. 2.   Chi-square Scores for the Most Frequent Words.

After simplifying equation 3, we get the following probabilistic model:

$$P(X|Y_1, Y_2, \dots Y_n) \propto P(X) * \prod_{i=1}^{n} P(Y_i|X) \qquad (4)$$

Equation 4 exhibits the probabilistic model, which is obtained after simplifying equation 3.

From this probabilistic model, a classifier model is generated by calculating the probability of all given inputs for the possible values of the class variable, and the maximum value is identified to determine the specific value of the class. It can be expressed as follows:

$$X = argmax \, P(X) * \prod_{i=1}^{n} P(Yi|X) \qquad (5)$$

2) Random forest: The high variance problem of the decision tree classifier, where a minor change in the training data set can produce a very different tree, makes the decision tree classifier unstable. To eradicate this issue, the concept of the random forest was proposed which is an ensemble of decision trees. Random forest is a classifier with various classification methods or a single method with various parameters from the dataset. Assume a learning data set D=((x₁,y₁), (x₂,y₂), …., (xₙ,yₙ)) that consists of n vectors, where x ∈ X (X is a set of numerical observations) and y ∈ Y (Y is a set of class labels). For a classification instance, a classifier maps X -> Y. Each tree of the forest is responsible for classifying a new input vector. Random forest combines the idea of bootstrapping data from a learning dataset to form training data set and selecting parameters randomly to construct decision trees. Bootstrap refers to selecting three-fourths of the learning dataset (sometimes two-thirds of the learning dataset) and replacing the rest of the data with some of the selected samples. While constructing a decision tree, features and their positions as nodes in a particular tree are chosen randomly. Thus random forest classifier, h, can be defined as:

$$h = \{h1(X), h2(X), \dots, hn(X)\} \qquad (6)$$

In equation 6, h_k is a decision tree having parameters θ_k, which is a subset of features chosen randomly [27].

3) Support vector classifier: Support Vector Classifier is a supervised learning algorithm to analyze data for classification. One of the distinctive properties of this algorithm is the ability to reduce empirical classification error and expand the geometric margin simultaneously. SVC provides high accuracy in text categorization, image classification, hand-written digits recognition, data classification, sentiment analysis, etc. SVC constructs a maximal separating hyperplane. Two parallel hyperplanes, which represent two different classes, are built on both sides of this separating hyperplane. By using SVC, an input vector is mapped to any of these two parallel hyperplanes. The purpose of the separating hyperplane is to maximize the space between the two parallel hyperplanes. It is assumed that less classification error can be achieved with a higher distance or margin between the parallel hyperplanes.

Consider a training dataset ((x₁,y₁), (x₂,y₂), …., (xₙ,yₙ)) where each x_n is a p dimensional vector that maps to corresponding y_n, which specifies the class. The value of y_n can be either -1 or +1. The equation of the separating hyperplane can be written as:

$$w.x - b = 0 \qquad (7)$$

In equation 7, w is a one-dimensional vector and b is a scalar. The equations for the two parallel hyperplanes can be written in equation 8 and equation 9 as follows:

$$w.x + b = 1 \qquad (8)$$

$$w.x + b = -1 \qquad (9)$$

The distance between these two hyperplanes is $\frac{b}{|w|}$ which implies that by minimizing |w|, we can maximize the distance between the two hyperplanes and thus achieve high performance.

In the case of hard margin where no misclassification is allowed from the training dataset, the problem can be stated as:

Minimize |w| for

$$y_i(w^T x_i - b) \geq 1 \, for \, i = 1,2,..,n \qquad (10)$$

In equation 10, the classifier is determined by solving w and b for the above problem statement.

In the case of soft margin where a few misclassifications are allowed from the training data set to achieve better accuracy for testing dataset, the problem can be stated as:

Minimize |w| for

$$[\frac{1}{n} \sum_{i=1}^{n} max \, (0,1 - y_i(w^T x_i - b))] + \lambda|w|^2 \qquad (11)$$

In equation 11, the trade-off between putting x_i in the right hyperplane and maximizing the distance between the parallel hyperplanes is determined by λ [28].

4) Multilayer perceptron: Inspired by the functioning procedure of human nervous systems, the concept of artificial neural networks has been developed and applied to design mathematical models to solve complex classification or regression problems. The building blocks of the artificial neural network are 'artificial neurons' or 'neurons'. Frequently these neurons are referred to as nodes. In a multilayer perceptron, which is a feedforward artificial neural network, these neurons are organized in layers and completely interconnected with each other via edges to construct a directed graph. The term 'feedforward' refers that this graph as acyclic. Each of these edges is associated with a real number which is called the weight of the edge. The layers of multilayer perceptron neural networks are the input layer, a number of hidden layers, and the output layer. For each neuron, there exists a summation function and an activation function. The summation function can be written as:

$$S_j = \sum_{i=1}^{n} w_{ij} I_i + \beta_j \qquad (12)$$

In equation 12,

$w_{ij} = $ *The connection weight from neuron* I_i *to* I_j

$\beta_j = $ *Bias weight*

$n = $ *The total number of neuron inputs*

The output of this summation function becomes an input of the activation function. There are various types of activation functions. One of the most applied activation functions is a nonlinear 'S' shaped curved sigmoid activation that can be expressed as:

$$f(x) = \frac{1}{1+e^{-x}} \tag{13}$$

Applying this activation function from equation 13, the output of the neuron I_j can be expressed as equation 14, which is shown below:

$$y_j = f_j(\sum_{i=1}^{n} w_{ij} I_i + \beta_j) \tag{14}$$

Once the neural network is constructed, the set of weights are tuned to estimate the required result [29].

IV. RESULTS

A confusion matrix has been developed as a binary prediction for each algorithm utilized to evaluate the performance system. One of the widely utilized methods for calculating predictions is a binary prediction which consists of the most significant building blocks of a ROC curve [30]. Each classification problem contains two classes. There exist two sets of positive and negative ((**P**) and (**N**)) labels of class for every instance. There are four possible categories for a classifier instance. True positive (TP) refers to the number of positive instances being classified appropriately. Similarly, true negative (TN) represents the number of negative instances being classified without any error. Opposite to that, if a positive instance is classified as a negative instance, it is considered as false positive (FP). Likewise, if a negative example is classified as a positive example, it is labelled as a false negative. While applying the algorithms to the dataset, in each instance, 80 percent of the dataset has been utilized as training data and the rest 20 percent has been used for testing.

We found accuracy, precision (P), recall (R), and F1 score by using the following equations:

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \tag{15}$$

In equation 15, accuracy is measured as the total number of correctly identified cases divided by the total number of test cases.

$$\text{Precision (P)} = TP/(TP + FP) \tag{16}$$

In equation 16, precision is measured as number of true positive cases divided by number of all predicted positive cases.

$$\text{Recall (R)} = TP/(TP + FN) \tag{17}$$

In equation 16, recall is measured as the number of true positive cases divided by all actual positive cases.

$$\text{F1 Score} = 2 * (R * P) / (R + P) \tag{18}$$

In equation 18, F-1 score is measured from the calculated precision and recall values.

Table IV shows the confusion matrix for the proposed machine learning classifiers for all features. Random forest works as the best classifier with 94% accuracy among the four algorithms. The accuracies for MLP, SVC, and NB are almost the same. Later we calculate the accuracy by changing the number of features ranging from 5000 to 9000.

Table V shows the accuracy for the proposed machine learning classifiers for features ranging from 5000 to 9000. The proposed model provides consistent results for different features ranging from 5000 to 9000.

Fig. 3 shows ROC curves for individual classes and all classes for linear SVC. Fig. 3(a) shows that the area under the ROC curve is 0.56, which indicates that linear SVC is not very efficient in the case of identifying class 0 (highly effective drugs). Fig. 3(b) shows that the area under the ROC curve is 0.64, which indicates that linear SVC is slightly more efficient in the case of identifying class 1(considerably effective drugs) than class 0. Fig. 3(c) shows that the area under the ROC curve is 0.82, which indicates that linear SVC is very efficient in the case of identifying class 2 (moderately effective drugs). Fig. 3(d) shows that the area under the ROC curve is 0.65, which indicates that linear SVC is approximately as efficient in the case of identifying class 3 (marginally effective) as class 1. Fig. 3(e) shows that the area under the ROC curve is 0.59, which indicates that linear SVC is slightly less efficient in the case of identifying class 4 (ineffective) than class 3. Fig. 3(f) plots precision against recall for all the classes. After analyzing the ROC curve, it is conspicuous that linear SVC has a significant positive performance in identifying class 2 drugs.

TABLE IV.   ACCURACY FOR 5000, 6000, 7000, 8000 AND 9000 FEATURES

RF		MLP		SVC		NB	
938	16	829	49	872	6	857	21
46	36	75	83	113	45	96	62

TABLE V.   ACCURACY FOR 5000, 6000, 7000, 8000 AND 9000 FEATURES

features	RF	MLP	SVC	NB
5000	94.21	88.71	88.80	88.71
6000	94.02	87.55	88.90	88.42
7000	93.92	84.75	88.51	88.71
8000	94.02	84.75	88.51	88.42
9000	94.11	88.32	88.42	88.61

TABLE VI.   CLASSIFICATION REPORT FOR RF

	precision	recall	f1-score	support
Class 0	0.69	0.44	054	82
Class 1	0.95	0.98	0.97	954
accuracy	0.94			1036
macro avg	0.82	0.71	0.75	1036
weighted avg	0.93	0.94	0.93	1036

(a) ROC curve for class 0.

(b) ROC curve for class 1.

(c) ROC curve for class 2.

(d) ROC curve for class 3.

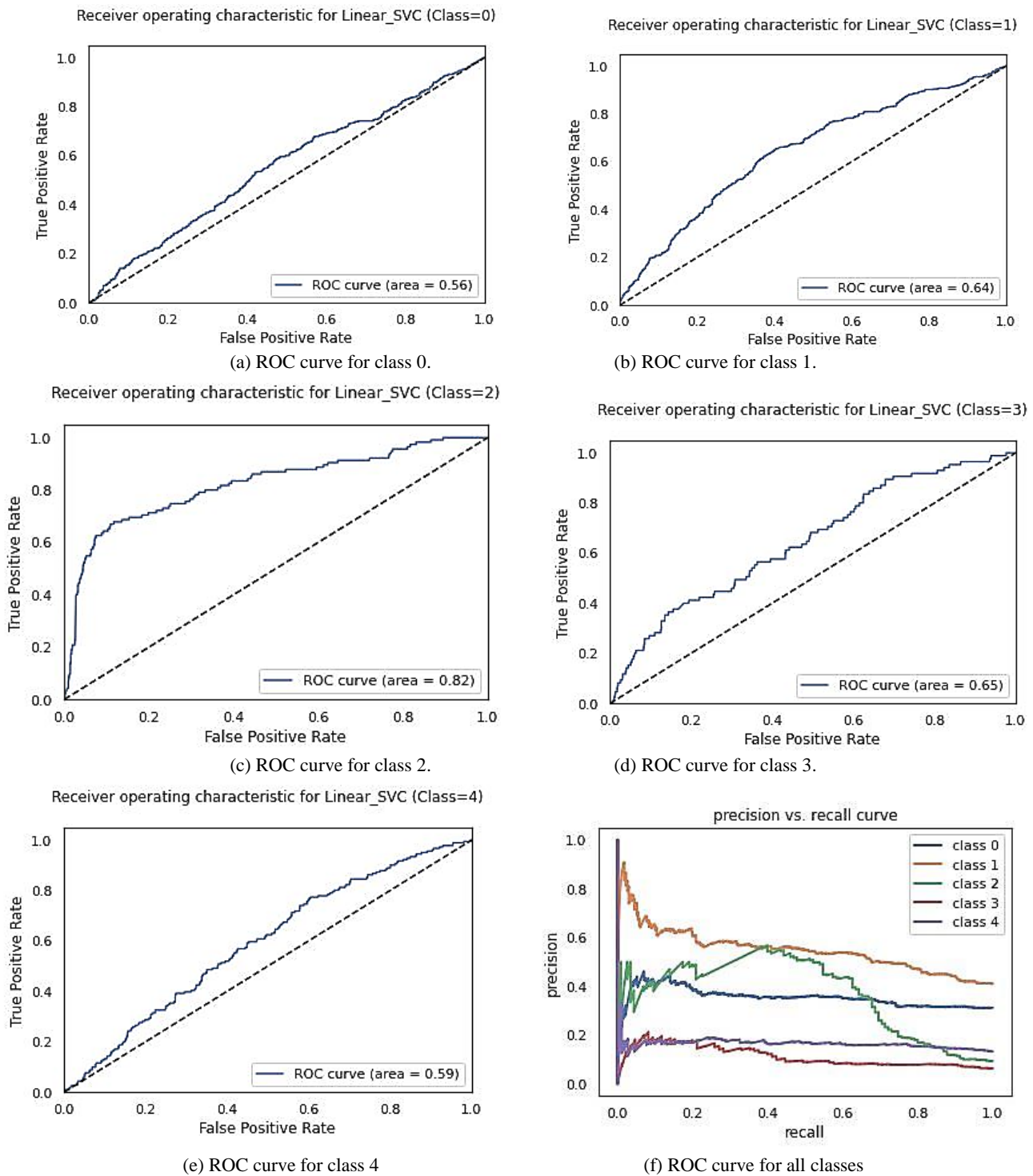(e) ROC curve for class 4

(f) ROC curve for all classes

Fig. 3.   ROC Curve for Linear SVC.

The evaluation metric used to represent the multiple class classification is the ROC curve which plots True Positive Rate (TPR) against False Positive Rate (FPR) at different threshold settings. Though the ROC curve is applicable for binary classification, there is an alternative way to integrate it for multiple-class classification. This approach is known as 'one vs rest' where a multiple-class problem is treated as a binary classification problem. In a ROC curve, a higher value in the X-axis indicates a greater false positive rate than the true negative rate A higher value in the Y-axis indicates a greater false-negative rate than the true positive rate. The efficient way to discriminate between accurate and inaccurate classification is to measure the area under the ROC curve (AUC). This area accepts values between 0 and 1, where 0 indicates the completely inaccurate classification of the class and 1 indicates perfectly accurate classification. Although

multiclass sentiment classification is extremely challenging for textual data, but Fig. 3(c) shows very promising accuracy in this research.

V. CONCLUSION

From the experimental result, the calculated average accuracy for Radom Forest, Multilayer Perceptron, Support Vector Classifier and Naïve Bayes Classifier is 94.06%, 86.82%, 88.63% and 88.57%, respectively. It is found that the Random Forest algorithm has generated the best accuracy among the four algorithms. In the case of Random Forest, higher precision, recall, and f1-score have been achieved for effective drugs compared to those measurements of ineffective drugs. The reason behind calculating the f-1 score is to get accuracy measurement from a different perspective as the f-1 score delivers the balance between precision and recall. Although multiclass classification is a challenging task for sentiment analysis, linear SVC shows the promising result for class 2 (moderately effective drugs). In this research, we have applied five machine learning algorithms. Unlike most of the similar researches in NLP when text mining is used for clustering the data, supervised learning methods have been implemented in this research to gain a better understanding of a drug by measuring its level of effectiveness. It can play an important role for curing diseases. In future, we intend to apply deep learning algorithms like Long Short Term Memory Networks (LSTMs), Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs). In addition to that, we would like to implement multi-language sentiment analysis using data-driven approaches.

REFERENCES

[1] M. Levis, C. L. Westgate, J. Gui, B. V. Watts, and B. Shiner, "Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models," (in English), Psychological Medicine, Article vol. 51, no. 8, pp. 1382-1391, Jun 2021, Art. no. Pii s0033291720000173.

[2] D. M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study," (in English), Journal of Medical Internet Research, Article vol. 22, no. 10, p. 16, Oct 2020, Art. no. e22635.

[3] M. Lewis, R. Li, M. Booth, M. Latymer, C. Borlenghi, and M. Kissner, "Sentiment analysis of Altmetric data using a human-augmented, artificial intelligence (AI) approach," (in English), Current Medical Research and Opinion, Meeting Abstract vol. 37, pp. 16-16, Apr 2021.

[4] T. H. McCoy, V. M. Castro, A. Cagan, A. M. Roberson, I. S. Kohane, and R. H. Perlis, "Sentiment Measured in Hospital Discharge Notes Is Associated with Readmission and Mortality Risk: An Electronic Health Record Study," (in English), Plos One, Article vol. 10, no. 8, p. 10, Aug 2015, Art. no. e0136341.

[5] A. Sansone et al., "The Sentiment Analysis of Tweets as a New Tool to Measure Public Perception of Male Erectile and Ejaculatory Dysfunctions," (in English), Sexual Medicine, Article vol. 7, no. 4, pp. 464-471, Dec 2019.

[6] A. Balahur, "Sentiment analysis," in 4th workshop on computational approaches to subjectivity, sentiment and social media analysis, 2013, pp. 120–128.

[7] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," IEEE Access, vol. 6, pp. 23253–23260, 2018.

[8] A. Schmidt and M. Wiegand, "A Survey on Hate Speech Detection using Natural Language Processing," in Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, 2017, pp. 1–10.

[9] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Data clustering using hybrid improved cuckoo search method," in 2016 Ninth International Conference on Contemporary Computing (IC3), 2016, pp. 1–6.

[10] Y. Wang and B. Li, "Sentiment Analysis for Social Media Images," in 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 2015, pp. 1584–1591.

[11] J. Xu et al., "Sentiment analysis of social images via hierarchical deep fusion of content and links," Appl. Soft Comput., vol. 80, pp. 387–399, Jul. 2019.

[12] I. S. Alimova and E. V. Tutubalina, "Entity-Level Classification of Adverse Drug Reaction: A Comparative Analysis of Neural Network Models," Programming and Computer Software, vol. 45, no. 8, pp. 439-447, Dec 2019.

[13] M. Devaraj, R. Piryani, and V. K. Singh, "Lexicon Ensemble and Lexicon Pooling for Sentiment Polarity Detection," Iete Technical Review, vol. 33, no. 3, pp. 332-340, May-Jun 2016.

[14] S. S. Liu and I. Lee, "Email Sentiment Analysis Through k-Means Labeling and Support Vector Machine Classification," Cybernetics and Systems, vol. 49, no. 3, pp. 181-199, 2018.

[15] S. Amin, M. I. Uddin, D. H. AlSaeed, A. Khan, and M. Adnan, "Early Detection of Seasonal Outbreaks from Twitter Data Using Machine Learning Approaches," Complexity, vol. 2021, Mar 2021, Art. no. 5520366.

[16] X. Y. Wang, D. Li, M. Y. Jiang, Z. L. Pei, and L. Xu, "K Nearest Neighbor Algorithm Coupled with Metabonomics to Study the Therapeutic Mechanism of Sendeng-4 in Adjuvant-Induced Rheumatoid Arthritis Rat," Evidence-Based Complementary and Alternative Medicine, vol. 2018, 2018, Art. no. 2484912.

[17] A. Tharwat, T. Gaber, A. E. Hassanien, and M. Elhoseny, "Automated toxicity test model based on a bio-inspired technique and AdaBoost classifier," Computers & Electrical Engineering, vol. 71, pp. 346-358, Oct 2018.

[18] Y. Choi and Y. Boo, "Comparing Logistic Regression Models with Alternative Machine Learning Methods to Predict the Risk of Drug Intoxication Mortality," International Journal of Environmental Research and Public Health, vol. 17, no. 3, Feb 2020, Art. no. 897.

[19] K. Yanagisawa et al., "Convolutional Neural Network Can Recognize Drug Resistance of Single Cancer Cells," International Journal of Molecular Sciences, vol. 21, no. 9, May 2020, Art. no. 3166.

[20] K. Yanagisawa et al., "Convolutional Neural Network Can Recognize Drug Resistance of Single Cancer Cells," International Journal of Molecular Sciences, vol. 21, no. 9, May 2020, Art. no. 3166.

[21] K. Y. Chang, T. P. Lin, L. Y. Shih, and C. K. Wang, "Analysis and Prediction of the Critical Regions of Antimicrobial Peptides Based on Conditional Random Fields," Plos One, vol. 10, no. 3, Mar 2015, Art. no. e0119490.

[22] Archive.ics.uci.edu. S. Kallumadi and F. Gräßer 2018. UCI Machine Learning Repository: Drug Review Dataset (Druglib.com) Data Set. [online] Available at: <https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Druglib.com%29> [Accessed 24 October 2021].

[23] Mullen et al., (2018). Fast, Consistent Tokenization of Natural Language Text. Journal of Open Source Software, 3(23), 655, https://doi.org/10.21105/joss.00655

[24] A. Ozcift, K. Akarsu, F. Yumuk, and C. Soylemez, "Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): an empirical case study for Turkish," Automatika, vol. 62, no. 2, pp. 226-238, Apr 2021.

[25] F. H. Khan, U. Qamar, and S. Bashir, "Senti-CS: Building a lexical resource for sentiment analysis using subjective feature selection and normalized Chi-Square-based feature weight generation," Expert Systems, vol. 33, no. 5, pp. 489-500, Oct 2016.

[26] M. Devaraj, R. Piryani, and V. K. Singh, "Lexicon Ensemble and Lexicon Pooling for Sentiment Polarity Detection," Iete Technical Review, vol. 33, no. 3, pp. 332-340, May-Jun 2016.

[27] Azar, Ahmad Taher et al. "A random forest classifier for lymph diseases." Computer methods and programs in biomedicine vol. 113,2 (2014): 465-73. doi:10.1016/j.cmpb.2013.11.004

[28] M. H. Afif and A. Hedar, "Data classification using Support Vector Machine integrated with scatter search method," 2012 Japan-Egypt Conference on Electronics, Communications and Computers, 2012, pp. 168-172, doi: 10.1109/JEC-ECC.2012.6186977.

[29] Basheer I, Hajmeer M (2000) Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods 43(1):3–31

[30] F. Kunneman, M. Lambooij, A. Wong, A. van den Bosch, and L. Mollema, "Monitoring stance towards vaccination in twitter messages," Bmc Medical Informatics and Decision Making, vol. 20, no. 1, Feb 2020, Art. no. 33.