# Predicting Drug Effectiveness Using Deep Learning

Utilizing Deep Neural Networks to Predict Drug Effectiveness

Joshua Qian
Computer Science
Emory University
Atlanta, GA, USA
jyqian2@emory.edu

## ABSTRACT

The pharmaceutical industry plays a critical role in advancing global healthcare, with effective drug use significantly improving patient outcomes. With the increasing volume of patient data, there is a growing need to analyze this data and leverage it to extract meaningful insights that can inform drug development, optimize treatment protocols, and enhance patient care. Predictive modeling, particularly the use of advanced machine learning techniques, has established itself as a powerful tool to evaluate a drug's effectiveness based on real world data. By utilizing these vast amounts of data, we can better understand patient experiences, identify patterns, and make data-driven decisions to improve healthcare outcomes. In this study, we leverage a deep feedforward neural network (DFN) to classify drugs as "good" (effective) or "bad" (ineffective) based on patient reviews and associated metadata. To benchmark its performance, we compare it against other hyperparameter configurations and traditional classification models. Our results demonstrate the superior performance of the DFN model in capturing complex relationships in the data, highlighting its potential in advancing drug classification and prediction tasks. Furthermore, we discuss the limitations of this and existing approaches and propose future directions. This paper underscores the importance of applying deep learning techniques in the healthcare domain to improve patient care and experience.

CCS CONCEPTS

Computing methodologies -> Machine learning -> Neural Networks -> Supervised learning

KEYWORDS

Deep Feedforward Network (DFN), Deep Learning, Neural Networks

**ACM Reference format:**

Joshua Qian, 2024. Predicting Drug Effectiveness Using Deep Learning: Utilizing Deep Neural Networks to Predict Drug Effectiveness. In *Proceedings of CS 485 – Deep Learning. ACM, Atlanta, GA, USA, 6 pages.* https://doi.org/10.1145/1234567890

## 1 INTRODUCTION

The pharmaceutical industry is a cornerstone of global healthcare, driving innovations that help improve patient care. It generated a global revenue of around $1.6 trillion US dollars in 2023, but a large proportion of drugs usage leads to suboptimal outcomes which increases healthcare costs. With the vast amounts of patient-generated data, there is great potential to leverage this information to make advancements in this field. Among the pressing challenges in this domain is the ability to predict the effectiveness of drugs based on drug metadata.

Traditionally, drug evaluations have relied heavily on clinical metrics and expert analysis. However, as patient-generated data becomes more and more accessible, there is an opportunity and a need to incorporate deep learning methods to enhance the prediction of drug effectiveness. The integration of machine learning, particularly Deep Feedforward Networks (DFN), offers a promising avenue to take advantages of these opportunities. By analyzing vast datasets, these models can uncover patterns and insights that would otherwise remain hidden. This paper explores the use of DFNs to predict drug effectiveness, leveraging observational data such as patient-generated reviews and associated metadata. This approach not only aims to improve patient outcomes but also contribute to more informed decision making in the pharmaceutical industry.

## 2 PROBLEM FORMULATION

The task of predicting drug effectiveness from observational data can be formulated as a supervised learning problem. In this context, the goal is to develop a model that is capable of classifying drugs into distinct categories of effectiveness based on patient-generated reviews and other associated metadata. This sections provides a formal definition of the problem, including the input, output, and objectives.

### 2.1 Dataset Overview

The dataset used for this research contains over 11,000 unique drugs and their related data. Each drug in the dataset is represented by the following features:

1. Medicine Name: name of the medication
2. Composition: active ingredients and their concentrations
3. Uses: medical conditions treated by the medication
4. Side effects: adverse effects associated with the medication

5. Image URL: link to an image of the medication packaging
6. Manufacturer: name of company that produces the medication
7. Excellent review %: percentage of users giving an excellent review
8. Average review %: percentage of users giving an average review
9. Poor review %: percentage of users giving a poor review

### 2.2 Dataset Preparation

To prepare the data for this task, we implemented a number of preprocessing steps including:

1. Target Feature Creation:
A target feature, Effectiveness, was derived from the proportion of excellent, average, and poor reviews. If a drug had a poor review percentage greater than 40%, it was classified as having "bad" effectiveness. Otherwise, it was classified as having "good" effectiveness. This transformed the task from a multi-class classification to a binary classification.

2. One-hot Encoding Categorical Data:
Categorical data was one-hot encoded to convert the data into a numerical format that can be understood by machine learning models.

3. Textual Feature Vectorization:
Textual features were vectorized using TF-IDF vectorization. This transformed the text into meaningful representations of integers to be interpreted by machine learning models. These features were then combined into a single input matrix. This allowed the model to benefit from the comprehensive input and leverage potential correlations between features to form more accurate predictions. It also simplified the model input as there were less inputs to provide to the model.

4. Train / Test Split
A train / test split of 80 / 20 was established to train and evaluate model performance.

### 2.3 Inputs

For this study, only *Uses* and *Side Effects* were used as inputs for the model. These features were selected due to their direct relevance to the prediction task. As mentioned earlier, these features were vectorized using TF-IDF Vectorizer and then combined into a single input matrix.
Features such as *Manufacturer* and *Image URL* were removed because they had no direct relevance to the effectiveness of a drug. Incorporating these features could have introduced unnecessary noise to the model which could have adverse effects.

### 2.4 Outputs

The model output a binary classification for the effectiveness of each drug as either "Bad" or "Good". The predictions were based on patterns and correlations learned from the input data during the training of the model.

### 2.5 Objective

The objective of the model is to minimize the binary cross-entropy loss function which is defined as:

$$L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}(Y_i \cdot \log \hat{Y}_i + (1 - Y_i) \cdot \log (1 - \hat{Y}_i))$$

Figure 1: Binary Cross-Entropy Loss Formula [1]
Where $Yhat_i$ is the predicted probability for sample i to belong to the "Good" class.

### 2.6 Preliminaries

This sections outlines the foundational concepts and techniques utilized in the model to predict drug effectiveness.

**TF-IDF Vectorization**
Textual features were converted into numerical representations using TF-IDF Vectorization. The transformations are calculated as follows:

TFIDF

For a term $i$ in document $j$:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{ij}$ = number of occurrences of $i$ in $j$
$df_i$ = number of documents containing $i$
$N$ = total number of documents

Figure 2: TF-IDF formula [2]
This approach captures the importance of term's specific to a drug's description while discounting common terms across multiple drugs.

**Dropout**

To prevent overfitting during model training, dropout regularization was employed. Dropout stochastically sets a specified fraction of the neurons to zero during each training iteration which prevents the network from relying too heavily on specific features.

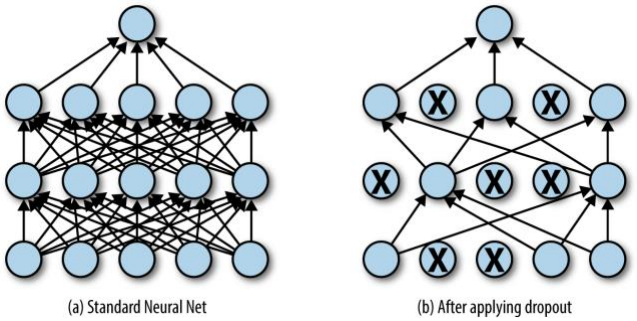



Figure 3: Visualization of dropout regularization [3]

**ReLU Activation Function**

Rectified Linear Unit (ReLU) is an activation function that is often used in deep learning due to its simplicity and efficiency. It introduces non-linearity to the model, enabling it to learn complex relationships between features. The function is defined as:

$$f(x) = \max(0, x)$$

Figure 4: ReLU formula [4]

ReLU only allows positive values to pass through and sets negative values to zero.

### 2.7 Existing Works

Drug effectiveness prediction has been an active area of research. In this section we will examine existing work in this field.

#### 2.7.1 Traditional Machine Learning Techniques

Early works in drug prediction relied on traditional machine learning algorithms such as logistic regression, support vector machines (SVM), and naïve bayes among many others.

1. Logistic regression: assumes linear relationship between input features and log-odds of the output class

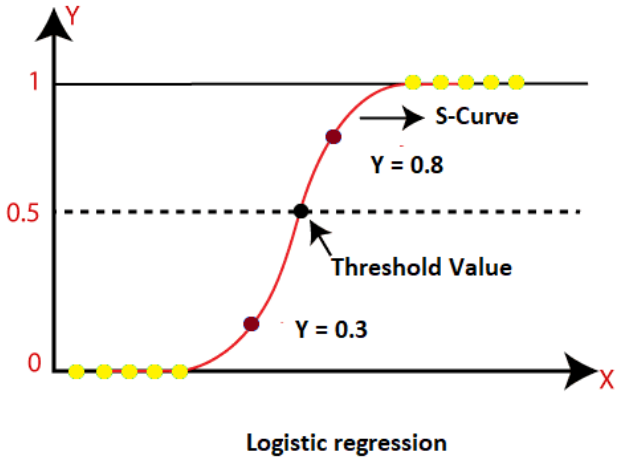



Figure 8: Logistic regression visualization [8]

While logistic regression is very popular for its simplicity and interpretability, its linear decision boundary makes it unsuitable for datasets with complex relationships.

2. Support vector machines (SVM): SVMs find a hyperplane that best separates data into classes

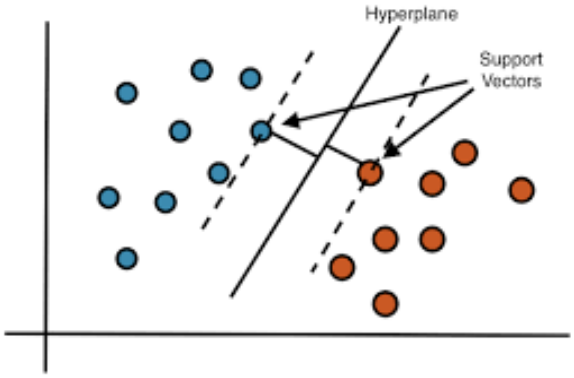



Figure 9: SVM visualization [9]

While SVMs are effective for both linear and non-linear classification problems, SVMs are computationally expensive which may make it infeasible for large datasets.

3. Naïve Bayes: probabilistic classifier that assumes feature independence

Naïve Bayes' assumption of feature independence limits its practical use in real-world datasets as it oversimplifies the dataset.

#### 2.7.2 Graph Neural Networks (GNN)

Graph Neural Networks have emerged as a powerful class of machine learning models for tasks involving structured data. Characteristics of GNNs include:

1. Graph representation: entities are represented as nodes in a graph and their relationships are represented as edges. This allows GNNs to capture relationships and dependencies between entities.

2. Node feature encoding: node features represent characters of each entity. This helps GNNs learning meaningful patterns within the graph.
3. Graph aggregation: GNNs use message passing mechanisms to aggregate information from neighboring nodes and edges. This allows the model to learn both local and global patterns.
4. Classification: After several rounds of message passing, GNNs summarize node features into a graph-level representation. This graph-level embedding is passed through fully connected layers to predict the target class.

While GNNs have demonstrated promise in the field of classification tasks, they have their limitations. For one, they are computationally expensive, which can lead to memory bottlenecks and slower training, making it infeasible for larger tasks. GNNs can also suffer from over-smoothing. In regard to drug classification, it can mask the distinctiveness of features like side effects, leading to poor performance. Lastly, GNNs struggle to capture non-linear interactions.

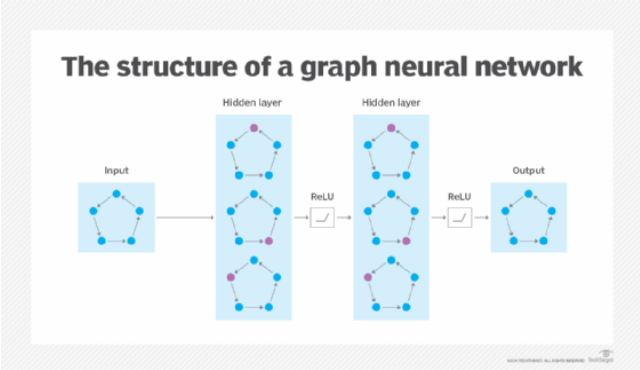



Figure 10: GNN structure visualization [10]

2.7.3 MEDICASCY

MEDICASCY is a multilabel-based boosted random forest machine learning method. MEDICASCY differs from other approaches in that it requires data regarding the chemical structure of molecules. Experimental data may be unavailable in some situation and MEDICASCY leverages a boosted random forest structure to predict the effectiveness of a drug. This model stands out for its ability to make robust predictions with minimal input requirements. This makes it particularly useful in the early stages of drug development.

While MEDICASCY demonstrates significant advancements in predicting drug effectiveness, it has its own limitations. First of all, it needs further experimental validation. Though it has been validated for certain applications, it is not exhaustive and requires further experimentation before conclusions can be drawn. For example, experimentation on highly complex molecules can still be explored. Further, its dependency on chemical structure simplifies its input but may fail to capture important information not included in the chemical structure.

## 3 TECHNICAL DESIGN

This section provides an overview of the architecture and methodology employed in the study.

### 3.1 Model Architecture

The primary model used here is a Deep Feedforward Neural Network (DFN). The model consists of the following components:

1. Input layer: Takes the raw input data and transforms it using methods described above into an appropriate format
2. Hidden layers: Fully connected layers, each followed by ReLU activation functions to introduce non-linearity and enable the model to learn complex feature interactions
3. Dropout regularization: Applied after each hidden layer to prevent overfitting
4. Output layer: Fully connected layer with a softmax activation function to output the prediction.

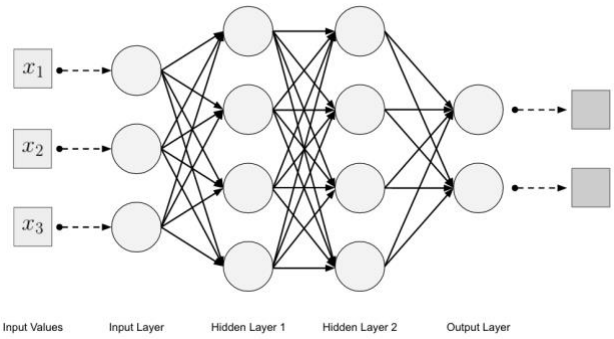



Figure 5: Sample architecture diagram [5]

### 3.2 Training Process

The dataset was split into training and testing sets with a ratio of 80:20. This was determined as a good balance between providing sufficient training data and testing data. The training process involves iteratively minimizing the loss function over 50 epochs with a batch size of 32. Validation performance was also monitored to ensure the model generalizes well to unseen data.

### 3.3 Hyperparameters

The following hyperparameters were considered for the design of this model:

1. Number of hidden layers: 4 hidden layers were included to balance model complexity and computational efficiency
2. Learning rate: learning rate of 0.001 was selected for balanced convergence
3. Number of epochs: 50 epochs allowed the model enough time to converge without overfitting
4. Dropout rate: dropout rate of 0.5 applied after each hidden layer to prevent overfitting

5. Hidden layer sizes: utilized a decreasing hidden layer structure to extract high-level features first and refine them later.

### 3.4 Evaluation Metrics

To evaluate the performance of the model and compare it to traditional models, we used the following evaluation metrics:

1. Accuracy: measures proportion of correctly classified samples out of total samples
2. Precision: measures how many of the predicted "good" drugs were actually correct
3. Recall: measures how many of the "good" drugs in the dataset were correctly classified
4. F1 Score: harmonic mean of precision and recall

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Figure 6: Precision and Recall formulas [6]

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Figure 7: F1 score formula [7]

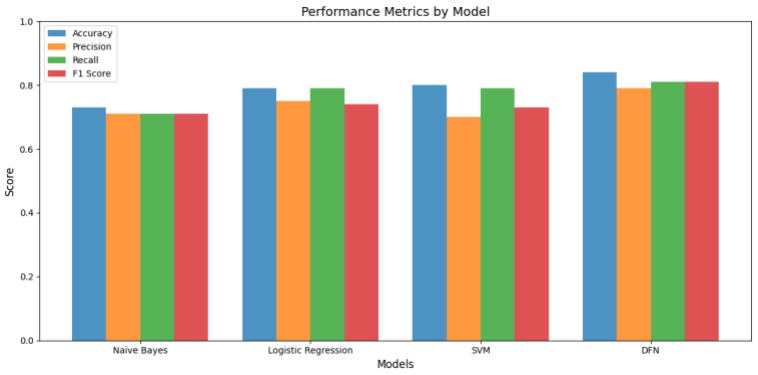
## 4 RESULTS AND ANALYSIS

This section discusses the results of the study and analyzes them.

### 4.1 Model Performance

The table below shows the evaluation metrics (accuracy, precision, recall, f1 score) for the benchmark, traditional models (naïve bayes, logistic regression, and SVM) as well as the Deep Feedforward Network (DFN).

| | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| Naïve Bayes | 0.73 | 0.71 | 0.71 | 0.71 |
| Logistic Regression | 0.79 | 0.75 | 0.79 | 0.74 |
| SVM | 0.80 | 0.70 | 0.79 | 0.73 |
| DFN | 0.84 | 0.79 | 0.81 | 0.81 |




The DFN outperformed the baseline models across the board. The improvement in recall suggests that it was able to effectively identify both "good" and "bad" drugs. While the SVM showed competitive accuracy and recall scores, it lacked in precision and F1 score. This tells us that it had challenges classifying ambiguous cases.

### 4.2 Interpretations

The results from our experiments demonstrate the effectiveness of different machine learning models in predicting the effectiveness of drugs. Below are some key observations:

1. Naïve Bayes: Achieved the lowest performance across all metrics. This can be attributed to the fact that it assumes feature independence which likely does not hold in this dataset.
2. Logistic Regression: Performance was in the middle of the models. As a simpler model, it benefits from the dataset's well separated classes but struggles to capture more complex features.
3. SVMs: Performance was also in the middle of the models. It's ability to find an optimal hyperplane contributed to its strong classification ability but appears to be sensitive to class imbalance as the precision score was lower than other metrics.
4. DFN: Demonstrated the best performance. It's ability to capture non-linear feature relationships and leverage dropout regularization likely contributed to its robustness and superior performance.

### 4.3 Challenges and Limitations

Despite achieving promising results, this study faced several challenges and limitations:

1. Dataset Diversity: This dataset may not comprehensively represent diversity among different drug types and their effectiveness across various populations which can limit the generalizability of the results
2. Binary Classification Constraint: Reducing the task to a binary classification task may be oversimplifying the problem.

3. Computational Costs: The DFN model required more computational resources and training time compared to traditional models.

### 4.4 Future Directions

This research represents a significant step towards leveraging deep learning models to predict drug effectiveness. However, there are numerous opportunities for future work to expand and improve upon this study:

1. Enhancing Feature Engineering: Finding additional data and incorporating more features could help to uncover more relationships that this study did not. Further, incorporating temporal features such as changes in review proportions could provide additional context for predictions.
2. Expanding Classification Scope: Extending the model to handle multi-class classification could provide more actionable insights for researchers.
3. Exploring Ensemble Methods: Ensemble methods could be used to improve the robustness and accuracy of predictions. Techniques such as bagging or boosting can combine the strengths of multiple models to enhance performance.

## 5 CONCLUSION

This study explored the application of machine learning techniques, including a DFN to predict drug effectiveness based on textual and categorical features from a publicly available dataset. The DFN model outperformed traditional machine learning approaches like Naïve Bayes, SVM, and Logistic Regression using accuracy, precision, recall, and F1 score as evaluation metrics. This demonstrates the potential of deep learning model in capturing meaningful patterns in data, especially with complex relationships. However, challenges such as limited feature engineering, high computational price, and a limited scope highlight areas for improvement.

Future directions include sourcing additional data and incorporating more features, expanding the classification scope, and exploring ensemble methods to further refine the model's performance and applicability. By addressing these challenges, the predictive power of deep learning in pharmaceutical applications can be greatly improved, leading to more effective patient care.

## REFERENCES


[1] Matej Mikulic. 2023. Global pharmaceutical industry - statistics & facts | Statista. *Statista*. Retrieved December 8, 2024 from https://www.statista.com/topics/1764/global-pharmaceutical-industry/

[2] Sheel Saket. 2020. Count Vectorizer vs TFIDF Vectorizer | Natural Language Processing. *LinkedIn*. Retrieved December 8, 2024 from https://www.linkedin.com/pulse/count-vectorizers-vs-tfidf-natural-language-processing-sheel-saket

[3] 4. Fully Connected Deep Networks - TensorFlow for Deep Learning . *O'Reilly*. Retrieved December 8, 2024 from https://www.oreilly.com/library/view/tensorflow-for-deep/9781491980446/ch04.html

[4] Nagababu Molleti. 2023. RELU & GELU Activation Functions in Neural Networks. *LinkedIn*. Retrieved December 8, 2024 from https://www.linkedin.com/pulse/relu-gelu-activation-functions-neural-networks-nagababu-molleti-nu8tc

[5] 4. Major Architectures of Deep Networks - Deep Learning [Book]. *O'Reilly*. Retrieved December 8, 2024 from https://www.oreilly.com/library/view/deep-learning/9781491924570/ch04.html

[6] Barak Or. 2021. Is F1 the appropriate criterion to use? What about F2, F3,…, F beta . *Medium*. Retrieved December 8, 2024 from https://towardsdatascience.com/is-f1-the-appropriate-criterion-to-use-what-about-f2-f3-f-beta-4bd8ef17e285

[7] f1 Score Definition | Encord. *Encord*. Retrieved December 8, 2024 from https://encord.com/glossary/f1-score-definition/

[8] Logistic Regression in Machine Learning | tutorialforbeginner.com. *TutorialForBeginner*. Retrieved December 8, 2024 from https://tutorialforbeginner.com/logistic-regression-in-machine-learning

[9] Vatsal. 2021. Support Vector Machine (SVM) Explained - Towards Data Science. *Towards Data Science*. Retrieved December 8, 2024 from https://towardsdatascience.com/support-vector-machine-svm-explained-58e59708cae3

[10] Alexander Gillis. What is Gen AI? Generative AI Explained | TechTarget. *TechTarget*. Retrieved December 8, 2024 from https://www.techtarget.com/searchenterpriseai/definition/generative-AI

[11] Das, Shuvendu and Badhon, Afroj Jahan and Jalal, Maddassar, Predicting Effectiveness of Drug from Patient's Review 2022. Proceedings of the Advancement in Electronics & Communication Engineering 2022, http://dx.doi.org/10.2139/ssrn.4157245

[12] Hongyi Zhou, Hongnan Cao, Lilya Matyunina, Madelyn Shelby, Lauren Cassels, John F. McDonald, and Jeffrey Skolnick. 2020. MEDICASCY: A Machine Learning Approach for Predicting Small-Molecule Drug Side Effects, Indications, Efficacy, and Modes of Action. *Molecular Pharmaceutics* 17, 5 (2020), 1558-1574. DOI:https://doi.org/10.1021/acs.molpharmaceut.9b01248

[13] [1]Jannik P. Roth and Jürgen Bajorath. 2024. Machine learning models with distinct Shapley value explanations decouple feature attribution and interpretation for chemical compound predictions. Cell Reports Physical Science 5, 8 (2024), 102110. DOI:https://doi.org/https://doi.org/10.1016/j.xcrp.2024.102110

[14] Bara A. Badwan, Gerry Liaropoulos, Efthymios Kyrodimos, Dimitrios Skaltsas, Aristotelis Tsirigos, and Vassilis G. Gorgoulis. 2023. Machine learning approaches to predict drug efficacy and toxicity in oncology. *Cell Reports Methods* 3, 2 (2023), 100413. DOI:https://doi.org/10.1016/j.crmeth.2023.100413