Journal of Informatics and Web Engineering
Vol. 3 No. 3 (October 2024) eISSN: 2821-370X

Treatment Recommendation using BERT Personalization

J Jayapradha[1*], Yukta Kulkarni[2], Lakshmi Vadhanie G[3], Palanichamy Naveen[4], Elham Abdulwahab Anaam[5]

[1,2,3] Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 6030203, India
[4,5] Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia 63100, Cyberjaya, Selangor.
**corresponding author: (jayapraj@srmist.edu.in; ORCiD: 0000-0002-2548-9135)*

Abstract - This research work develops a new framework that combines patient feedback with evidence-based best practices across disease states to improve drug recommendations. It employs BERT as its free-text processing engine to deal with sentiment judgment and classification. The functionality of the system, named `PharmaBERT`, includes acceptance of drug review data as a comprehensive input, drug categorization when dealing with a wide range of treatments and fine-tuning the BERT-based model for gaining positive or negative sentiment towards specific medications. PharmaBERT categorizes various drugs and fine-tunes the BERT structure to perceive lots of possible sentiments for specific medications. Consequently, PharmaBERT brings all its training and optimization capabilities together and through this, the system reaches a higher accuracy of up to 91% thus showcasing the potency of the model in capturing patient sentiments. While being a BERT spin-off, PharmaBERT utilizes its own set of experienced techniques to comprehend and sense the health-related text input given by the patient, doctor, or pharmacist. It uses transfer learning, that is, it learns from language representations to adapt quickly to the intricacies of drug reviewing. Through PharmaBERT, healthcare professionals may expand their diagnoses with insights from patient feedback to constitute more neutral decisions.

Keywords - Bidirectional Encoder Representation from Transformers (BERT), Machine Learning (ML), Artificial Intelligence (AI), Large Language Models (LLMs), Deep Neural Network (DNN), Natural Language Processing (NLP).

Received: 14 April 2024; Accepted: 17 June 2024; Published: 16 October 2024

1. INTRODUCTION

Over the last several years, the exponential growth of electronic health records and user communities on the Internet has made patient-generated data user groups (PGHD) available to the public. Among them, drug reviews play an important role in providing information about how a drug works, what are its' possible adverse reactions are, and eventually a patient's general experience of it. One of the most important outcomes of big data utilization for personalized drug recommendation systems is the ability to make use of such rich sources of knowledge in patient personalization, tailoring treatments according to the needs and tastes of individual patients. While it is possible to compute meaningful information from the mass load of unstructured written data of drug reviews, these kinds of tasks are fraught with great challenges.

Machine learning, particularly deep learning techniques, has emerged as a powerful tool for extracting insights from unstructured text data. A transformer that has come to play an important role in the language model and is widely adopted these days is the Bidirectional Encoder Representations from Transformer Architecture (BERT). With BERT, Devlin et al.'s work in 2018 has changed the view in Natural Language Processing (NLP),

highlighting that transformer model can be pre-trained using large text corpora, and then capture complex semantic relationships and contextual information.

Beyond healthcare, BERT has been able to gain compelling accomplishments in the fields such as clinical text categorization, medical question/ answer and sentiment analysis of patient reviews. Taking into consideration the immensity of unstructured text data as well as its ability in the healthcare domain, the framework proposed in this research paper is a model of drug recommendation using patient review to extract insights. The principal goal of this study is to utilize BERT on sentiment analysis and classification of drug reviews, determining whether they recommend the drug or not based on sentiments expressed in the review.

Drug recommendation systems are having an unprecedented influence on treatment choice selection, allowing the healthcare providers as well as patients to make an appropriate selection of medications. Usually, conventional ways of making drug recommendations work based on clinical guidelines, the expertise of the professional, and the medical records of the patient (the case histories). However, despite the importance of these methods, they may not always weigh the overall spectrum of a situation from a patient's perspective and preferences.[1] Applying abundant information from the patient-generated medication reviews can help an algorithmic system of drug recommendations provide a more complete and patient-oriented way for drug management. These models not only assist healthcare professionals to point out drugs that provide quicker and non-liable healing but also help patients to choose drugs that suit them.

Hitherto the possible benefits of patient reviews shouldn't be overlooked in drug recommendation, however, these challenges should be taken care of. One of the most fundamental problems of drug reviews is that their quality is dictated by the unstructured nature, informal language and the presence of noise. Human language is both heterogeneous and complex, and therefore processing sophisticated NLP techniques capable of delving into such difficulties are required [2].

Secondly, the task of sentiment analysis of drug reviews is distinctive in comparison with sentiment analysis in other fields. Drug reviews often contain drug specific terms; colloquial expressions and implicit sentiments, that can be missed by general sentiment analysis models. Similarly, the mood perceived by analytics in drug reviews is influenced by medical condition, treatment duration, and unique experiences of patients, which further necessitates context aware sentiment analysis techniques. As BERT showcases its capability in detecting such contextual information and semantic relationships in text data, it becomes a strong tool for sentiment analysis problems.

Unlike the traditional models of sentiment analysis which mainly uses the features of handcrafted words and shallow linguistic patterns to identify sentence context and rely on the learned representations of words and sentence, BERT captures the nuances and contextual clues to contribute to sentiment analysis [3]. Through pre-training on big text corpora and fine-tuning on data that is specific for different tasks, PharmaBERT learns to obtain sentiment information from diverse text sources, such as drug reviews. Due to its bidirectional structure, it is capable of dealing with an entire document, thus covering all the long-range dependencies and semantic relations that may have an effect on sentiment analysis.

Taking the BERT capabilities in sentiment evaluation as a starting point of this research, the framework for drug recommendation based on patient reviews is developed. The first research goal is to use BERT (Bidirectional Encoder Representation from Transformers) to facilitate the sentiment analysis of drug reviews and classify them as either recommending or not recommending a drug based on the opinions expressed in the reviews.

This research is to build on an innovative system for drug suggestion using patients' reviews, which are to be analysed by the most elegant deep learning methods, particularly BERT, and maintenance of accuracy, for sentiment analysis. The second factor will be to analyse how well PharmaBERT performs in capturing sentiment from drug reviews and its' influence on the accuracy and performance of drug recommendation systems. Finally, testing of the proposed framework on many datasets collected from real-world drug ratings help achieve an effective mechanism to adopt in the management of health decision-making.

The health care sector poses a tremendous concern in identifying best and worst drug reviews for recommendation systems. Traditional methods often lack the nuance required to analyse textual data to make a personalized medication suggestion. By utilizing PharmaBERT, the model, a deep learning natural language processing technique, addresses the problem. BERT's context sensitive feature enables sorting out the drug reviews by

capturing subtly semantic elements. This makes drug recommendation strategies effective in the future, emphasizing those drugs that hold greater value to the patients.

2. LITERATURE REVIEW

The research findings describe that an LLM like ChatGPT can be a helpful aid accessory for breast tumor board decision-making. The researcher allowed Clinical details of ten patients into ChatGPT-3.5 and compared these with tumor board decisions. The results illustrated that in 70% of cases ChatGPT's recommendations were alike to the tumor board's decisions. The average scores for the chatbot's summary, recommendation and explanation were given by two senior radiologists rated very highly. Through the proof-of-concept of the study, it is argued that if LLMs can be developed to effectively interpret the latent space of sequential recommenders that model item sequences with discrete IDs, further research is necessary to fully assess the efficiency of the model and explore possible complications [4]. The study presents the RecInterpreter, a mapping framework that will transfer both the representations into the token embedding space of the LLM model and the prompts to guide the model in generating textual descriptions for items in the sequence. Analysis demonstrates the fact that RecInterpreter helps in a better understanding of these representations and leads to more accurate predictions when sequence-residual prompts are used [5], whereas in another study RecMind, an autonomous agent which needs no human help and is being powered by LLMs, has been introduced. RecMind utilizes a novel algorithm dubbed Self-Inspiring, which improves the planning capacity of the LLM agent by considering all the states previously explored during the planning step. It is because of this that the model can understand and use the historical planning data for future recommendations. RecMind captures the attention of existing models, LLM-based ones and even outperforms P5 in multiple recommendation tasks [6].

The research also investigates the fairness of recommendations made by LLMs, particularly focusing on the Recommendation via LLM (RecLLM) paradigm. It introduces a novel benchmark called FaiRLLM, which evaluates the fairness of LLM recommendations with respect to various sensitive attributes. The evaluation using the FaiRLLM benchmark revealed that ChatGPT exhibits unfairness to some sensitive attributes when generating recommendations [7]. Additionally, the study explores the LLMs potential for zero-shot next-item recommendation, identifying challenges such as the large recommendation space and lack of information about the user's past interactions. The proposed prompting strategy, Zero-Shot Next-Item Recommendation (NIR) prompting, addresses these challenges by guiding the LLM to capture user preferences and recommend a ranked list of items. Evaluation using GPT-3 on the MovieLens 100K dataset shows strong zero-shot performance, outperforming many sequential recommendation models trained on the entire dataset [8].

TALLRec enhances LLMs in recommendation tasks by fine-tuning with recommendation data, significantly improving movie and book domain recommendations. It exhibits robust cross-domain generalization and is noted for its efficiency, runnable on a single RTX 3090 with LLaMA-7B, with code and data available on GitHub [9]. LlamaRec, a two-stage recommendation framework, leverages LLMs for efficient ranking-based recommendations. It uses small-scale sequential recommenders to retrieve candidate items and employs a verbalizer-based approach for improved performance and efficiency. Experimental results demonstrate LlamaRec's superiority over existing methods [10].

GenRec introduces a novel recommendation approach using LLMs to directly generate target items, departing from traditional discriminative methods. By leveraging LLMs' context interpretation and user preference learning, specialized prompts enhance comprehension for recommendation tasks, yielding significantly improved results on benchmark datasets. GenRec showcases the transformative potential of LLM-based generative recommendation for recommendation systems [11]. RadOnc-GPT, fine- tuned on Mayo Clinic patient records, excels in radiation oncology tasks, surpassing general LLMs in treatment regimen generation and diagnostic descriptions. While promising, its broader clinical relevance needs further evaluation beyond ROUGE scores. Additionally, a programmatic prompt structure enables LLMs to generate tailored plans, achieving high success rates in Virtual Home tasks and physical robot arm deployment [12].

The study investigates the potential of using large language models (LLMs), in text- based collaborative filtering (TCF) for recommendation systems. By scaling up item encoders from millions to billions, the research explores the performance limits of TCF and compares it with ID embedding-based approaches. Additionally, it examines the transferability of TCF and compares it with prompt-based recommendation, revealing both positive and unexpected negative outcomes. The research aims to inspire further exploration and innovation in text-based

recommender systems [13]. The work proposes a recommendation approach where large language models (LLMs) interpret user preferences as natural language instructions, outperforming competitive baselines like GPT-3.5. By instruction-tuning an open-source LLM, the study generates personalized instruction data and demonstrates its efficacy in various recommendation tasks. This user-friendly approach allows for more accurate recommendations through natural language communication with the system [14].

The paper evaluates ChatGPT's ability to answer patients' gastrointestinal health- related questions, using 110 real-life queries assessed by gastroenterologists. ChatGPT's responses varied in accuracy and clarity, with average scores ranging from 3.4 to 3.9 out of 5 across different question types. While ChatGPT holds promise for better assessment, further refinement is necessary to ensure consistently reliable information provision [15]. The study proposes LLM reasoning graphs (LLMRG) to enhance recommender systems using large language models (LLMs). LLMRG constructs personalized graphs representing user interests. It includes four components: chained graph reasoning, divergent extension, self-verification and scoring, and knowledge base self-improvement [16].

LLMRG ensures accurate decision-making process and enables engineered system with LLM graphs to connect well to the benchmarks and even real-time situation [17]. The study proves that it is possible to employ huge language models (LLMs) to create dialogue-driven queries and teach retrieval models with narrative background. Through the combined utilization of both synthetic and real user-item interactions data, the technique is significantly more efficient than the remaining solutions. The aim of this method is to overcome training data scarcity for NDR models by using LLMs, providing a meaningful effort to address research gaps [18].

3. CONTRIBUTIONS

The study focuses on PharmaBERT, a pre-eminent approach that exploits the BERT model to achieve sentiment analysis in drug reviews. This advanced approach boosts healthcare automation by presenting accurate readings about the efficacy of treatment and sentiments of the patients, enhancing decision making process and positive outcomes.

1. Suggests a new framework based on BERT model that can be used for drug reviews sentiment analysis, which is a well-recognized issue in health care automation.
2. Curates a heterogeneous dataset of drug reviews to train and test the model which covers for entire treatment spans and levels of feelings.
3. PharmaBERT - a BERT variant specialized for drug review sentiment analysis that demonstrates adaptability and efficiency.
4. Evaluation of the performance of PharmaBERT is conducted to demonstrate that it is more accurate, efficient and clinically relevant.
5. Covers the possible extensions and applications of PharmaBERT in healthcare, showcasing how it can make a great difference to patient care and decision-making.
6. Considers ethical issues including privacy and data security.

4. IMPLEMENTATION OF THE PharmaBERT

4.1. What is PharmaBert?

PharmaBERT technology offers an unparalleled technique to innovate drug recommendation software, utilizing the most advanced features of word processing technology. It is based on Bidirectional Encoder Representations in Transformers (BERT), and using sentiment analysis, this tool classifies drug reviews either in “recommend” or “do not recommend” categories. With advancements in sentiment analysis of such textual reviews that highlight the side effects of the medication are examined and customized medication options is offered depending on the experiences of the individual patients. This approach vouches for better choices in the healthcare decision making process where there is much to be considered about drug effectiveness, tolerance and customer satisfaction.

4.2. Data Preprocessing:

Data preprocessing is an indispensable process that is performed in advance for creating a reliable drug recommendation system. In this step, a dataset containing drug reviews along with their metadata is prepared. Usually, the set is a combination of patients' comments, where additional details e.g. drug names, medical

conditions and ratings are provided. The process begins with loading the dataset from a CSV file or database into memory. In this case, the research makes use of the Drugs.com dataset which can be accessed in UCI Machine Learning Repository https://doi.org/10.24432/C5SK5S.

This dataset is a compilation of people's feedback about medicines, as well as other information about drugs such as drug names, conditions that they are used for, ratings and text of assessments. Once the file is unzipped, the data will get loaded into the memory utilizing suitable data processing tools e.g. Pandas in Python. The above step marks the starting point to develop the drug recommendation system. Furthermore, the data file is meticulously cleansed to maintain the data. In this case, the tasks involve elimination of repetitive reviews, filling in overlaps, cleaning of special characters/irrelevant material and ensuring accuracy in deduction. This should be one of the most important things about data preprocessing where data points are converted into binary labels. This contributes to eliminating the task of sentiment analysis as it already categorizes reviews as those that recommend the drug, and those that do not. As an example, over a certain predetermined score (for instance, more than 5.0), high ratings could be given the label suggestive of prescribing the medicine, and the ratings of a lower level could be classified as not a prescribe of the medicine.

The PharmaBERT model's sentiment analysis follows several main key stages. The dataset of Drugs.com is made with reviews written by patients on the website along with significant disease metadata. It goes through a straightforward preprocessing process before it can be deconstructed for experimentation. At this phase, the quality of the data is ensured right from the beginning of the process since issues to be addressed beforehand may include multiple entries, missing values and irrelevant information. Also, the rating in this dataset is transformed to binary label likewise many of the tweets, which makes the task easy for sentiment analysis. This means of categorization grades ratings defined by a predetermined threshold, generally established through empirical studies or by experts in the field, as suggesting the drug for that indication if they are above that threshold, and not recommending it otherwise if they are below.

After having preprocessed text data, then the data is tokenized by way of BERT tokenizer, which breaks down reviews into its individual single tokens. This act of underlying words as tokens and representations is critical for feeding data successfully into the BERT model. In addition, BERT has a unique tokenization structure which includes using special words, cutting or expanding sequence, and making attention masks. Designating the rating boundary represents an integral part in the sentiment analysis.

For instance, the threshold is chosen with reference to domain competence and dataset characteristics, and the result is explored to fine-tune the performance as well as classify sentiments effectively. In the case of PharmaBERT, BERT is a core part of the data encoding and sentiment recognition mechanism. It is this model that is fine-tuned given the preprocessed dataset which provides the model with the ability to learn complex relationships between drug reviews and the sentiments attributed to them. The process of precision model development allows BERT to select contextual information found in the input text and to predict the sentiment with each review with greater accuracy. The pharmaBERT model embeds BERT into its sentiment analysis pipeline to enable it to achieve better sentiment understanding and classifications maintaining the contextual linking and modulations of the reviews. This in turn increases the accuracy of the recommendation outcomes.

4.3. Model Training:

The modeling stage will be followed by data preprocessing phase, which leads to training a PharmaBERT model. The model's architecture is a pre-trained encoder from the BERT family, which excels well in capturing the context of given text, followed by a classifier responsible for verifying the sentiment. The sentiment classification module predicts sentiment by using a classification layer that has label predictions corresponding to the output that has been encoded. The model is determined based on the probability distribution over the classes (recommend or not recommend) employing competent techniques like softmax activation. Softmax assigns a specific probability weight belonging to each class, ensuring as a result, that it is always 1, and therefore summing up to one. CCM class with the greatest likelihood is chosen as predicted sentiment for a particular drug review.

AdamW Optimization:

$$m_t = \beta 1 \cdot m_{t-1} + (1 - \beta 1) \cdot gt \quad (1)$$

$$v_t = \beta 2 \cdot v_{t-1} + (1 - \beta 2) \cdot gt^2 \quad (2)$$

$$m\hat{} = \frac{1}{1-\beta_1 t}, \quad v^t = \frac{1}{1-\beta_2 t} \tag{3}$$

$$\theta_t = \theta_t - 1 - \frac{\eta}{\sqrt{vt} + \epsilon} \cdot m^t \tag{4}$$

Equation (1) updates the moving average of gradients in Adam optimization, while Equation (2) updates the moving average of squared gradients. Equations (3) provide bias corrected estimates of these averages, and Equation (4) updates the parameters using these estimates in the optimization process.

Where:

- θ are the model parameters.
- gt is the gradient of the loss with respect to θ.
- mt and vt are exponentially moving averages of the gradient and its square, respectively.
- β1 and β2 are the exponential decay rates for the moment estimates.
- η is the learning rate.
- ϵ is a small constant to prevent division by zero.

The text data is tokenized using BERT tokenizer, the process where the semantic content of the text is converted into such numerical representation suitable for machine learning models as input. In the next step, the model learns to classify sentiment through binary cross-entropy loss function, by comparing the probability of sentiment predicted according to the true sentiment labels. This is a suitable metric to optimize the model's parameters. Although careful adjustments of the parameters by backpropagation and gradient descent optimization are incorporated, PharmaBERT is well-trained to perfectly distinguish the sentiments expressed in the drug reviews.

Cross-Entropy Loss (Binary Classification):

$$BinaryCross - EntropyLoss = -N1\sum i = 1N(yi \cdot log(pi) + (1 - yi) \cdot log(1 - pi)) \tag{5}$$

Where equation (5) represents:

- *N* is the total number of samples in the training dataset.
- *yi* is the true label (0 or 1) for the *i*th sample.
- *pi* is the predicted probability of the positive class (drug recommended) for the
- *i*th sample.

At the training stage, the dataset is separated into small batches, every interval of which contains a subset of data. During the multiple passes (epochs) made by PharmaBERT, the internal parameters are periodically updated to correspond to the loss function. The training process consists of 49 batches carrying PharmaBERT to acquire from extensive examples and generalize terms to unforeseen review wielding data. In doing so, PharmaBERT gain expertise in identifying complex sentiments and provides valuable suggestions based on patient reviews of different drugs, followed by the selection of the right drug in healthcare placements.

4.4. Model Evaluation:

After the network is trained, the subsequent step is to evaluate the models' performance through a trial-and-error calibration. This involves training the model for sentiment analysis and feeding it another set of data. The other data set is meant for determining whether the model is good in classifying sentiments. First, the sentiment analysis model predicts the sentences for a given dataset that can be used for validation or testing dataset. These guessing labels are used as reference to measure some key performance metrics with respect to the real labels assigned.

Another common tool is the Matthews's Correlation Coefficient (MCC), which considers the true positive, true negative, false positive, and false positive values, and provides a balanced measure of a classifier performance

particularly apt for imbalanced datasets. Refer to Fig.1. Moreover, visualization tools like the loss curves are used for analyzing the training performance and the ability of the model over epochs (Fig.2).

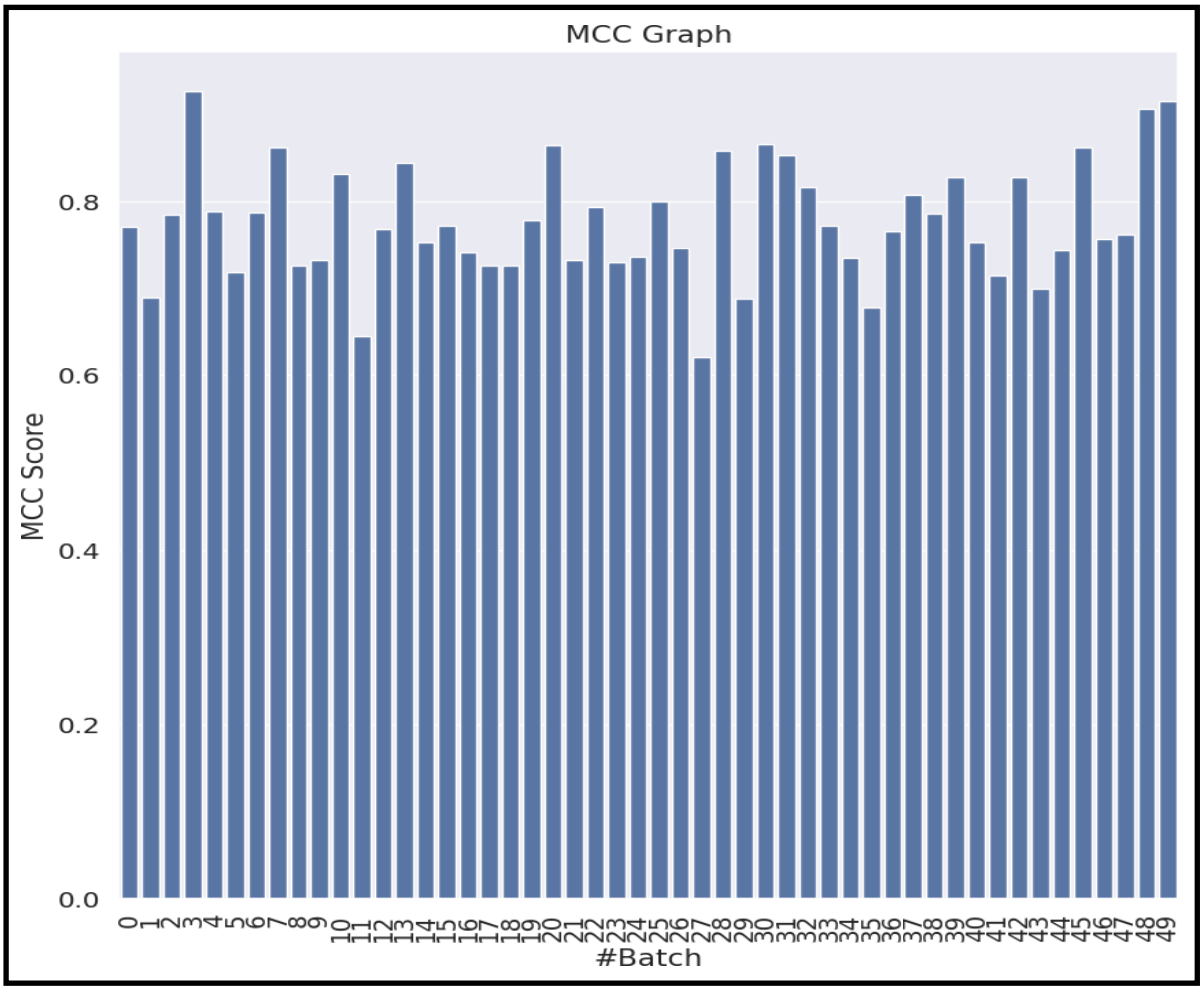

Figure 1. Matthews Correlation Graph

Figure 1 describes the term "MCC" which typically stands for "Matthews Correlation Coefficient". Binary classification models are important, especially when dealing with imbalanced datasets. It considers true positives, true negatives, false positives, and false negatives to produce a value between -1 and 1, where 1 indicates perfect prediction, 0 indicates no better than random prediction, and -1 indicates total disagreement between prediction and observation.

The formula for MCC is:

$$MCC = (TP * TN - FP * FN) / \sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)} \quad (6)$$

Where equation (6) represents:

- (TP) is the number of true positives
- (TN) is the number of true negatives
- (FP) is the number of false positives
- (FN) is the number of false negatives

Some curve diagrams, depicting both the training error and the validation error by the number of epochs, help

shed light on the mechanism of convergence of the model during the training process. This downward trend in training and level in validation suggests that the model is learning without being overly reliant on the training data. On the other hand, when the validation loss is rising with a decrease in the training loss, it reveals an overfitting problem which, therefore, suggests model architecture or training process to be adapted accordingly.

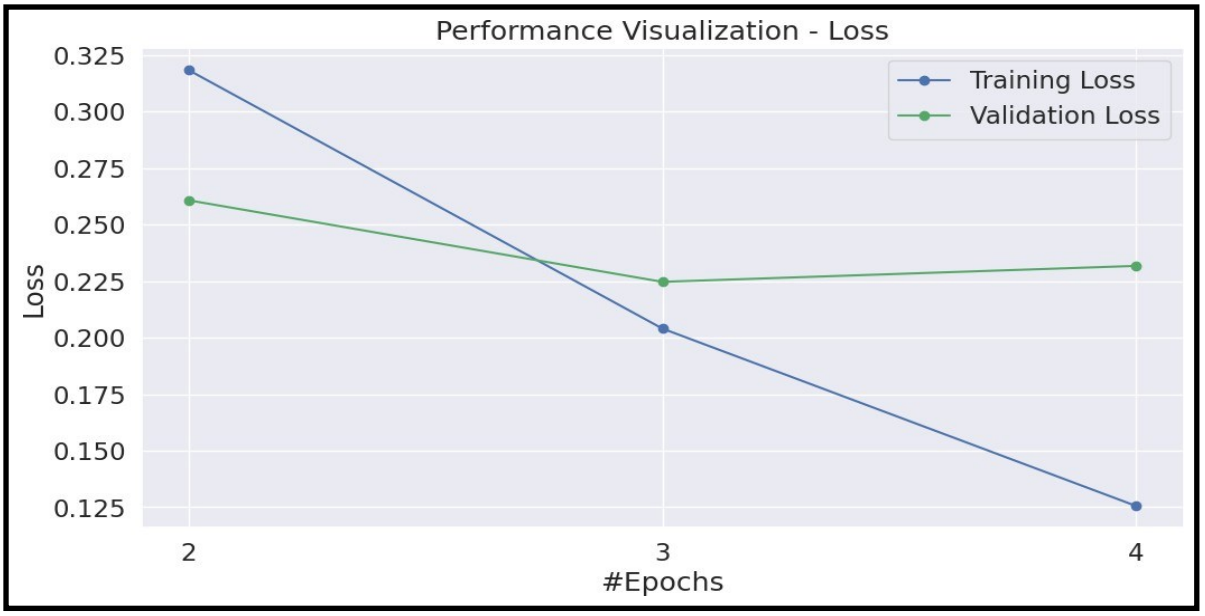

Figure 2. Performance Visualization

Figure 2 describes the model performance which illustrates the following:

1. Training Loss Trend: The plot reveals that the training loss is decreasing with more epochs, that is, it follows what would be expected. This implies that the model is starting to use the training data to its own advantage and thus can do better with every passing moment.
2. Validation Loss Trend: The process of validation loss will first decrease in value but then reach and start to rise or stabilize after a countable number of epochs. It might mean that the model has already overfitted the training dataset as it performs better on the training data but could not maintain its accuracy/accuracies on the validation dataset.
3. Loss Discrepancy: The two curves of the training and validation dropouts in a distinct manner, somehow overfitting of the model might be yet to happen. When the validity loss has begun to increase or stay the same while the training loss further decreases, such interpretation shows that the model knows the data provided and avoids generalizing it well to new data.
4. Model Performance Evaluation: According to the results, the model currently possesses reasonable training data learning potential, but it might be better to normalize it or tweak it to achieve more balanced generalization to unseen data. Techniques such as early stop, dropout or change in the model architecture can be employed and they may then aid high validation performance and reduce overfitting.

Also, in the process of MCC calculation and loss curve graphitization, the drug interaction heatmap is a great quantitative tool used for the drug effectiveness evaluation. This heatmap argues that there are indications on the possible interactions between drugs based on the patterns and frequency of occurrence of drugs in patient reviews. The heat map of drug interactions shows the correlation of drug mentioned within the reviews. Each cell in the heatmap depicts the frequency of the pair among the identified drugs so that higher density of cell represents a higher possibility of the drug-drug interaction.

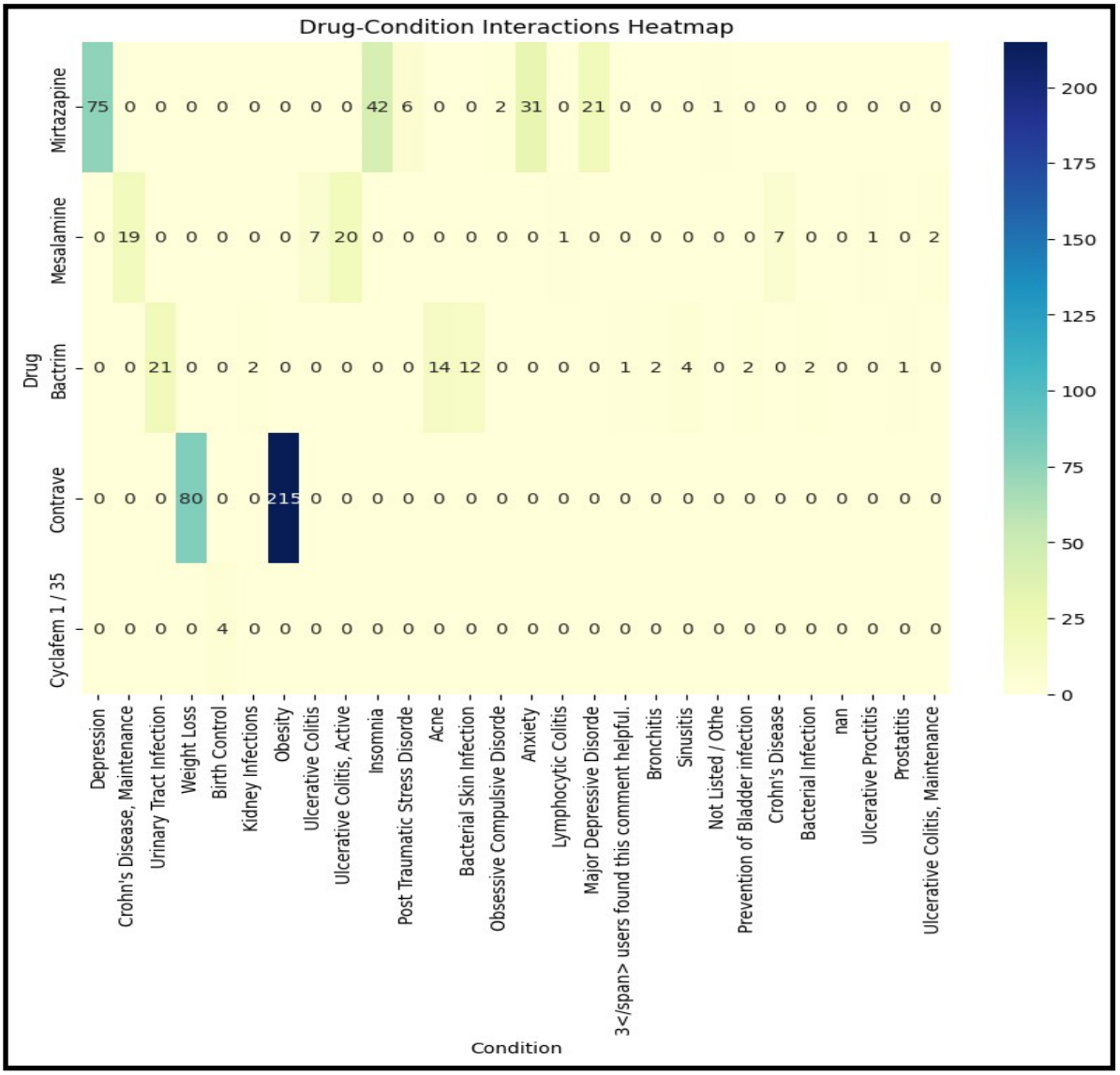

Figure 3. Drug Condition Interaction Heatmap

Figure 3 presents the drug-disease interaction heatmap in which these underlying conditions are linked with the specific elements of the dataset. The top row lists drugs and the columns under it are the names of medical conditions. Cells within the heat map are colored in varying numbers to indicate the frequency or count of reviews received by the drug-disease combination. More numbers indicate a larger number of reviews that contained a respective drug-condition mention, which means that the interactions between the drug and the specific medical condition might have been stronger. To easily identify the effectiveness and safety of drugs across different medical conditions, it is important to examine the of drug use among patients.

4.5. Recommendation Generation:

The system moves on to recommendation creation stage using a well-calibrated sentiment analysis model. The new drug reviews are classified and personalized recommendations are created by sentiment analysis sharpening the model. PharmBERT uses BERT tokenization process to break up the text data into words before classifying new drug reviews and constructing personalized recommendations. These tokenized inputs are fed into the pre-trained PharmBERT model, which produce contextual embeddings of these reviews. After this, these representations are sent to the classification layer that classifies the reviews either recommending or not recommending the drugs.

These personalized recommendations are, in turn, informed by the individual patient's experience and preferences revealed by the reviews entered. To obtain probability distribution over sentiment classes, the SoftMax activation

is one such method that allows the model to make well informed recommendations based on sentiment predictions. The SoftMax activation and binary classification techniques are employed by PharmBERT to label whether a new review has a sentiment of recommended or not recommended a drug respectively.

The SoftMax activation gives a probability to each sentiment class which is helpful for the mode to find the chances of having positive or negative reviews about drugs. Similarly, binary classification separates the review groups by setting a fixed threshold according to the probabilities that can be deemed as a good dividing line between the most probable reviews and low probable ones. The model can simply tell whether a review is recommending or non-recommending of a drug by applying its threshold on the predicted probabilities resulting in informed decision making in medication management. The equation (7) gives the formula for SoftMax activation.

SoftMax Activation:

$$P(yi = 1 \mid x) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z}_{j}} \quad (7)$$

Where:

- $P(yi=1|x)$ is the probability that the sentiment is positive (or recommending)
- z_i is the raw score for class i
- K is the total number of classes.

Medical related information can be communicated and accessed in different features like graphical charts or tables in various formats by medical professionals and other users. Entailing the growing ability of any medicine to deal with every individual's specific problem. Such an approach where everything is tailored to a patient's personal life may turn out to be an instance of better decision-making in healthcare and as a result, positive patient outcomes. In this comprehensive framework, the BERT-based sentiment analysis foundation is created which forms the basis for development of drug recommendation system.

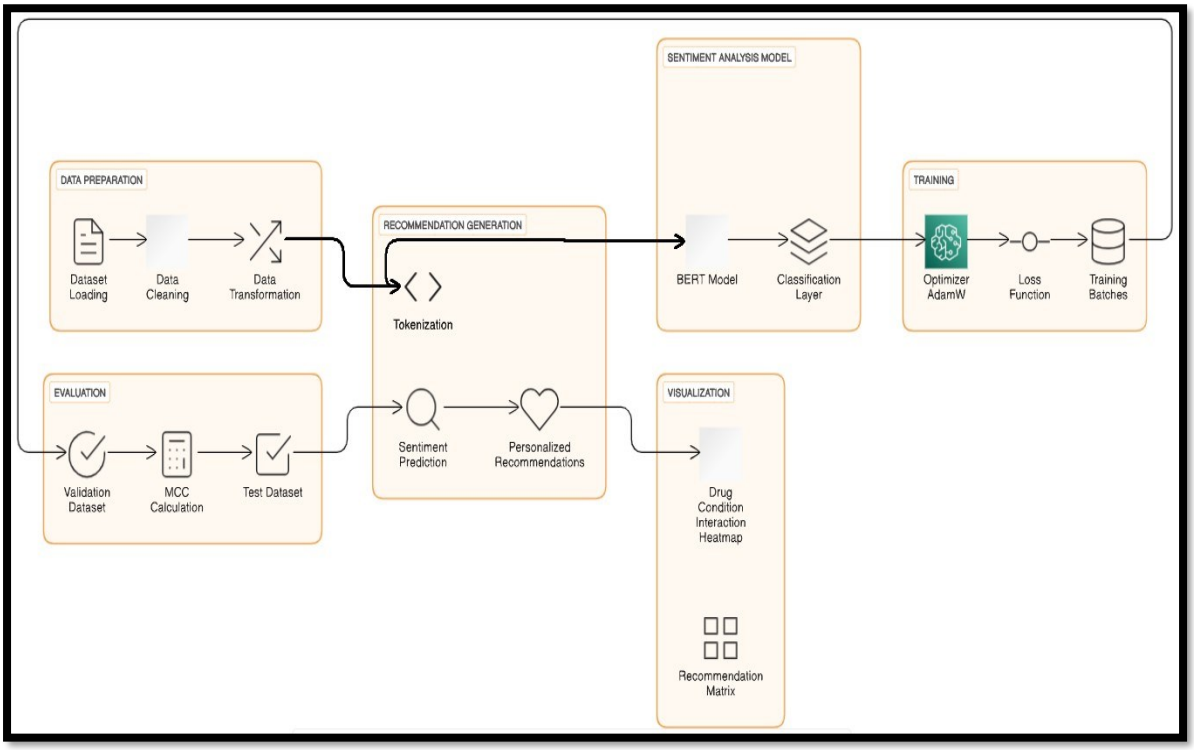

Figure 4. Architecture Diagram

The PharmaBERT architecture diagram in Figure 4 illustrates the structure starting with data loading, followed by tokenization while focusing on the embedding phase with attention masks. These steps are made within the

BERT tokenizer using ‘Bert-base-uncased’ model. Next, data is preprocessed, and the model is designed to address the specific problem being tackled while fitting the model with the 'BertForSequenceClassification' class which is binary classifier. The model is thought to be the 'Bert-base-uncased' pre-trained model, which belonged to the 'Bert' model series. The dataset is subsequently divided into training, validation, and test sets, employing random sampling for the training set and sequential sampling for the validation set. This partitioning ensures a consistent and uniform distribution of the data throughout the process.

During the training phase, the BERT model is iteratively executed using the training data. The algorithm performs forward and backward passes, measures loss, backpropagates, and applies gradient clipping to adjust model weights. The optimizer and scheduler are used to update model weights in each step. After training, the model is evaluated on the test dataset, and the output logits are extracted and stored. Accuracy is then calculated from the results.

The architecture concludes with data analysis, focusing on condition distribution and recommendation tendencies for specific drug-condition combinations. The analysis is presented through pie charts, weight matrix, drug interaction heatmap and tracking data on positive and negative recommendations.

4.6. Algorithm of PharmaBERT:

Step 1: #Preprocess Dataset
1. Initialize an empty list input_ids and an empty list attention_mask.
2. For each sentence s in sentences, do:
  - Encode s using the BERT tokenizer with add_special_tokens=True, max_length=64, and return_attention_mask=True.
  - Append the encoded input to input_ids.
  - Append the attention mask to attention_mask.

Step 2: #Define BERT Model Architecture
1. Initialize the BERT tokenizer with BertTokenizer.from_pretrained('bert-base-uncased', do_lower_case=True).
2. Initialize the BERT model for sequence classification with BertForSequenceClassification.from_pretrained("bert-base-uncased", num_labels=2, output_attentions=False, output_hidden_states=False).
3. Move the model to the desired device (dev).

Step 3: #Split Dataset
1. Split the dataset into training and validation sets (train_data, val_data) using random sampling.

Step 4: #Define Optimizer and Scheduler
1. Initialize the AdamW optimizer with AdamW(mod.parameters(), lr=3e-5, eps=1e-8).
2. Create a linear scheduler with warmup using get_linear_schedule_with_warmup with num_warmup_steps=0 and num_training_steps=tot_s.

Step 5: #Train BERT Model
1. For each batch b in the training data (train_dataloader), do:
  - Perform a forward pass to get logits and loss from the model.
  - Compute the loss and perform a backward pass to backpropagate the loss and update the model weights using the optimizer and scheduler.

Step 6: #Preprocess Test Data
1. Load the test data (sentences, labels, drug names, conditions).
2. Convert labels to binary (1 if >=5.0, else 0).

Step 7: #Evaluate Model
1. For each batch in the test data, do:
   - Perform a forward pass to get logits from the model and append the logits to a list preds.
2. Calculate the accuracy using calculate_accuracy(preds, true_labels).

Step 8: #Analyze Drug-Condition Relationships
1. For each unique drug in the dataset, do:
   - Filter the data for the drug to get condition_counts.
   - Create a pie chart to visualize the distribution of conditions.

Step 9: #Analyze Drug-Condition Recommendations
1. For each unique drug in the dataset, do:
   - Initialize empty lists yes and no.
   - For each condition in the dataset, do:
     - Filter the data to get positive (y) and negative (n) ratings.
     - Append the counts to yes and no.

PharmaBERT algorithm suggests a holistic approach to building drug recommendation systems that leverage advanced NLP techniques such as BERT-based sentiment analysis. Firstly, there is data preparation whereby the dataset from Drugs.com is preprocessed by tokenization and label encoding. Then there is definition of model architecture where BERT tokenizer and model are initialized for sequence classification. Next, datasets were split into three sets of training data, validation data and testing data accordingly to facilitate model development and evaluation. Fine-tuning BERT model techniques were employed to get optimal results. Thus, these evaluation approaches provide the degree of precision that can be ascertained based on these data. However, through detailed analysis of drug – condition interactions, PharmaBert shares with the reader the knowledge necessary to understand these interactions and prescription pattern analysis for distinct medical problems as well. The goal of PharmaBERT is to change the way the data is processed to achieve a better result than previously existing systems.

5. RESULTS AND DISCUSSION

The diagrams provided are designed to give an overall view of the distribution of disease per drug in the data set. Each pie chart illustrates the percentage of different symptoms experienced by the users who have used a certain medication. It gives a better understanding of which conditions the drug treats and how frequent they are in the dataset. The application of complicated language models (LLMs) like Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis can upgrade the reading of textual data linked with drug reviews. From this text, BERT models can capture intricacies, which gives an insight into the condition prevalence and drug use patterns more profoundly. The graphs “Recommend” and “Do Not Recommend” illustrates the recommendation trends of the drugs when it is used in different conditions. These diagrams represent the users’ attitudes towards the effectiveness of different drugs, based on their ratings that were categorized as “recommend” or “do not recommend”. PharmaBERT enables deeper analysis of sentiments in textual reviews for each user rating. This can be achieved by creating a more nuanced recommendation approach and adapting interventions to the needs of individual patients.

5.1. Evaluation of PharmaBERT Model

Validation and test accuracies of the PharmaBERT model is described in this section which indicates its performance in identifying correct and incorrect drug recommendations.

Figure 5 illustrates the rate of change of PharmaBERT ‘s accuracy over training epochs to know if it learns from the training data and whether it generalizes good to unseen validation data.

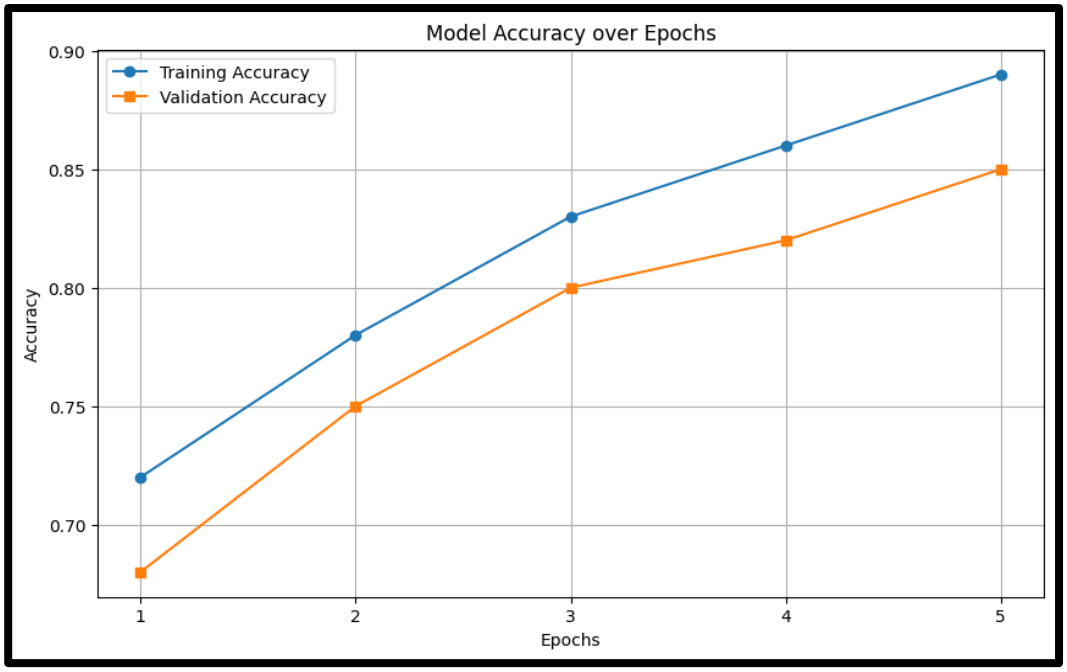

Figure 5. Performance Visualization

1. X-axis (Epochs) represents number of training epochs, which are passed across the whole training dataset with each epoch representing a complete pass through the entire dataset while Y-axis (Accuracy), stands for total ratio of accurate predicted point among all points in model.

2. Training Accuracy (Blue Line): The blue line represents accuracy at each epoch on a training dataset which is how well the model learns from the training data over successive epochs. The increasing trend of training accuracy implies that learning is taking place by means that improve performance within this set.

3. Validation Accuracy (Orange Line): The orange line represents accuracy at each epoch on a different validation dataset. It is an independent measure of the model's performance on new test datasets that were not used during its development stage'. An increasing trend in validation means that this model is performing best when presented with new sets. Both training and validation datasets have an upward trend at every epoch.

5.2. Visualization of Drug-Condition Associations

To distinguish between right and wrong drug recommendations, this part presents the validation and test accuracies of the PharmaBERT model.

Each drug in the dataset is traversed by Figure 6 to generate a pie chart showing the distribution of medical conditions associated with that drug. For each drug, the pie chart segments represent different medical conditions reported by users who have used the drug. The size of each segment indicates the proportion of reviews mentioning each condition. Visualizing the condition distribution per drug offers insights into the variety of conditions treated by each drug and their relative frequencies within the dataset.

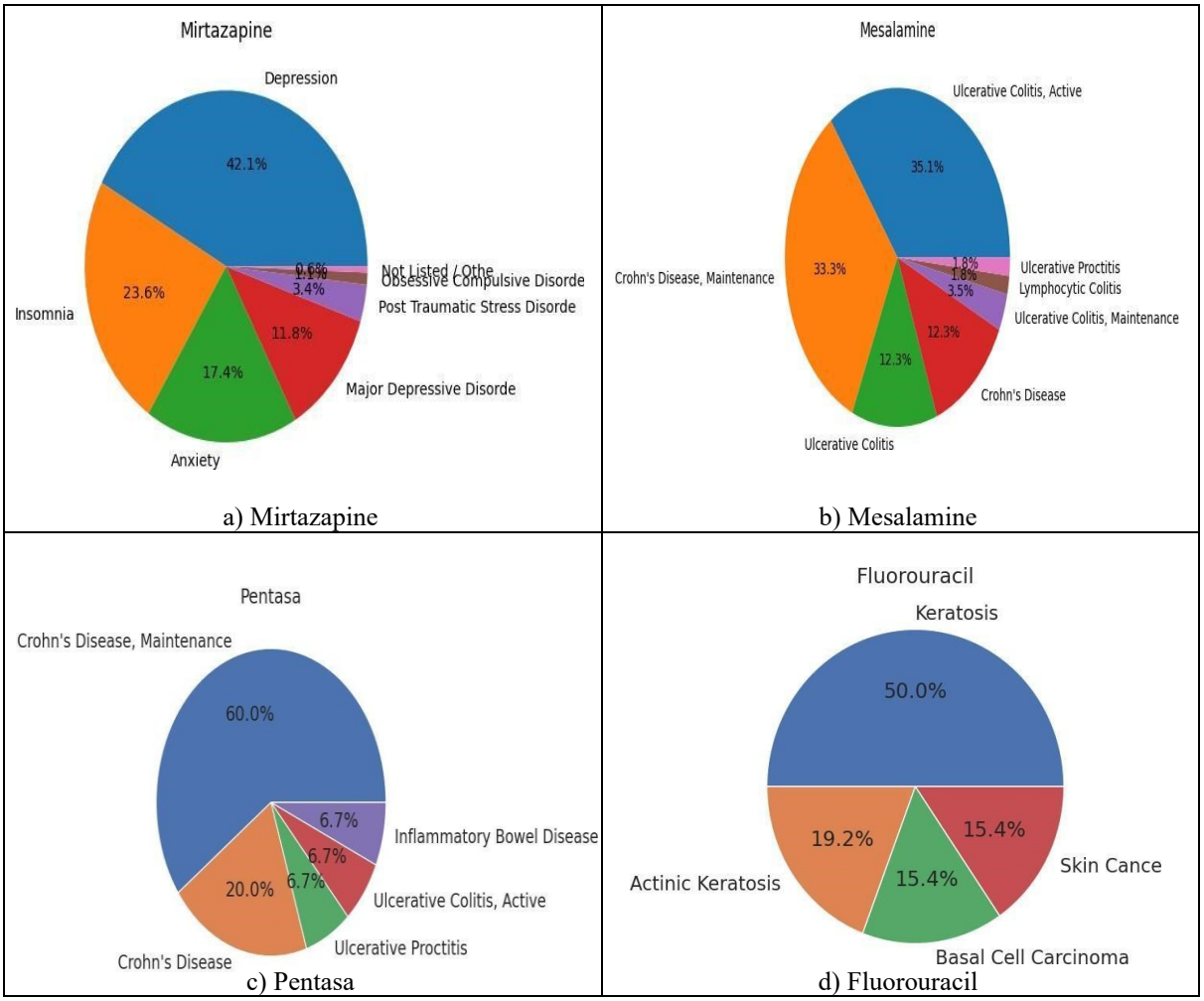

Figure 6 (a)-(d). Condition Distribution per Drug

Figure 6(a) describes mirtazapine which accounts for depression: 42.1%, insomnia: 23.6%, anxiety: 17.4%, major depressive disorder: 11.8%, bipolar disorder: 5.1%. Mirtazapine is an antidepressant drug primarily used to treat major depressive disorders.) Figure 6(b) describes mesalamine which accounts for Crohn's Disease (Maintenance): 33.3%, Ulcerative Colitis: 35.1%, Ulcerative Proctitis: 12.3%, Lymphocytic Colitis: 13.8%, Strep Throat: 6.9%, Upper Respiratory Tract Infection: 12.3%, Sinusitis: 5%, Skin and Structure Infection: 2%, Urinary Tract Infection: 2%, Otitis Media: 2%, Skin or Soft Tissue Infection: 2%, Bacterial Infection: 2%. Mesalamine is primarily used to treat inflammatory bowel diseases, such as Crohn's Disease and Ulcerative Colitis. Figure 6(c) describes pentasa that covers Ulcerative Colitis (Active): 60.0%, Crohn's Disease (Active): 15.0%, Ulcerative Proctitis: 10.0%, Ulcerative Colitis (Maintenance): 10.0%, Diverticular Disease: 5.0%. Pentasa is primarily used to treat active Crohn's Disease and Ulcerative Colitis. Figure 6(d) describes Fluorouracil which accounts for Actinic Keratosis: 50.0% Skin Cancer: 15.4% Basal Cell Carcinoma: 15.4% Keratosis: 19.2%. The drugs are mainly discussed in relation to Actinic Keratosis, which accounts for half of the recorded conditions. Skin cancer and basal cell carcinoma comprise 15.4% of the total, and keratosis is mentioned in 19.2% of the cases.

5.3. Drug-Condition Network Analysis

This section illustrates relationships between drugs and diseases using a network graph, providing insight into how different medical conditions are distributed among all drugs in our dataset.

Figure 7 utilizes NetworkX, a Python library for the creation, manipulation, and study of complex networks, to visualize the relationship between drugs and medical conditions. A subset of drugs from the dataset is selected and represented as nodes in a directed graph. Each drug node is connected to its corresponding medical conditions through directed edges, indicating the association between the drug and the condition. The colors of the edges are customized based on the drugs, enhancing the visual representation. By visualizing this drug-condition network,

it provides insights into the distribution and interconnectedness of drugs and medical conditions within the dataset, aiding in the exploration and analysis of drug usage patterns and treatment preferences. The information presented in Figure 4, depicting the network graph of drug-condition associations, is summarized and detailed in Table 1.

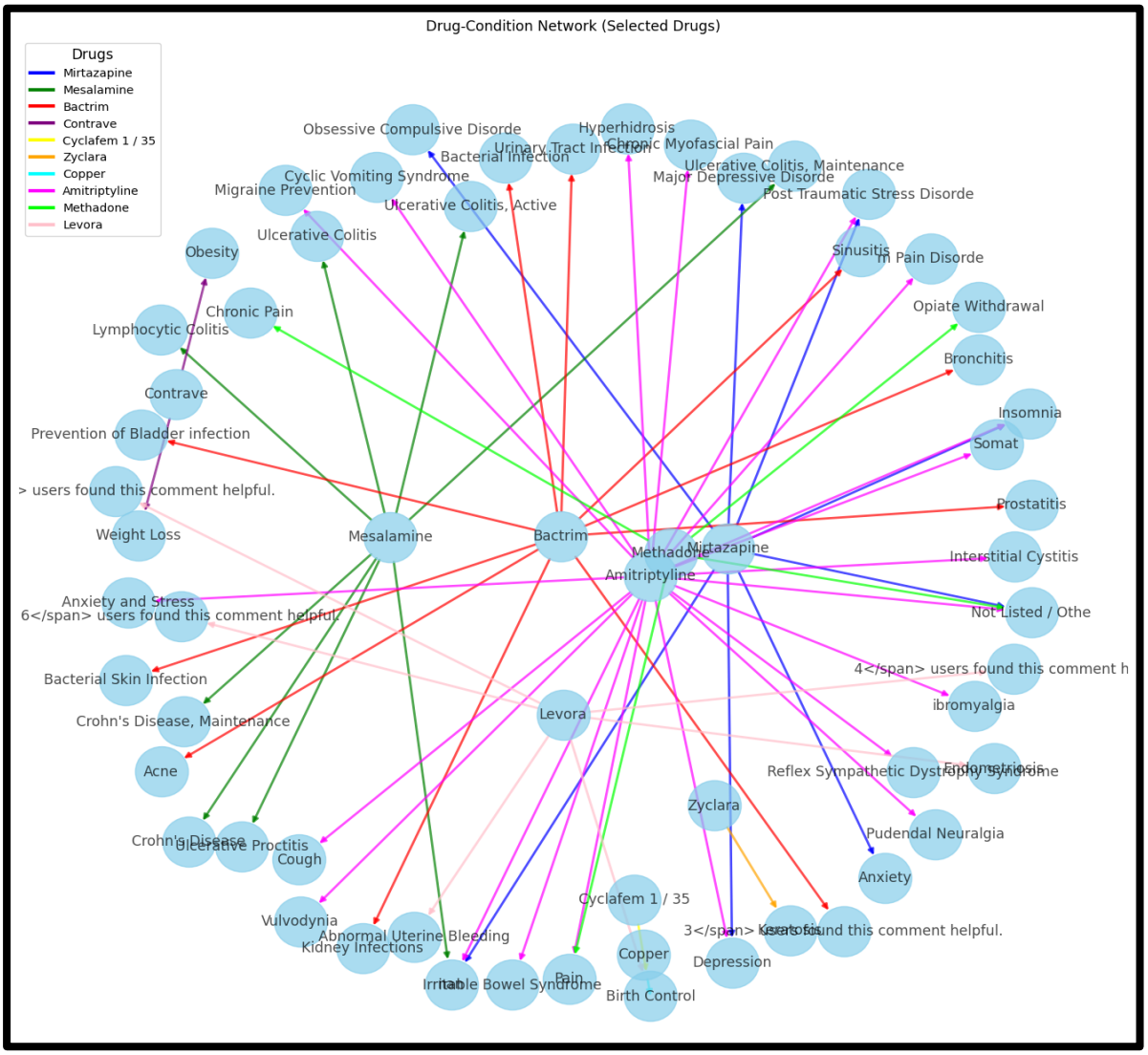

Figure 7. Drug Condition Network

5.4. Recommendation Matrix Visualization

Here, the visualization of the recommendation matrix for drugs and medical conditions is presented. It showcases the distribution of recommendations (Yes or No) for each medical condition associated with a particular drug.

Figure 8 showcases the distribution of recommendations (Yes or No) for each medical condition associated with a particular drug. For each drug, a grid of colored cells is produced, where each row corresponds to a unique medical condition and each column represents either a recommendation (Yes) or a non-recommendation (No). The color intensity of each cell represents the count of reviews classified under the respective recommendation status for the corresponding condition.

Table 1. Drug-Condition Associations

	DRUGS	DISEASES
1	Mirtazapine	['Depression' 'Insomnia' 'Post Traumatic Stress Disorder' 'Obsessive Compulsive Disorder' 'Anxiety' 'Major Depressive Disorder']
2	Mesalamine	["Crohn's Disease, Maintenance", 'Ulcerative Colitis', 'Ulcerative Colitis, Active', 'Lymphocytic Colitis']
3	Bactrim	['Urinary Tract Infection', 'Kidney Infections', 'Acne', 'Bacterial Skin Infection', 'Bronchitis', 'Sinusitis', 'Prevention of Bladder infection', 'Bacterial Infection', 'Prostatitis']
4	Contrave	[Weight Loss', 'Obesity']
5	Cyclafem	['Birth Control']
6	Zyclara	['Keratosis']
7	Copper	['Birth Control']
8	Amitriptyline	['Migraine Prevention', 'fibromyalgia', 'Anxiety and Stress', 'Pain', 'Hyperhidrosis', 'Cyclic Vomiting Syndrome', 'Insomnia', 'Irritable Bowel Syndrome', 'Depression', 'Cough', 'Post Traumatic Stress Disorder', 'Somat', 'Interstitial Cystitis', 'Reflex Sympathetic Dystrophy Syndrome', 'Vulvodynia', 'm Pain Disorder', 'Chronic Myofascial Pain', 'Pudendal Neuralgia']
9	Methadone	['Opiate Withdrawal', 'Pain', 'Chronic Pain']
10	Levora	['Birth Control', 'Abnormal Uterine Bleeding', 'Endometriosis']

Figure 8(a) describes Actos for Diabetes, Type 2- referring to the recommendation status for the drug Actos when used to treat Type 2 Diabetes. The figure shows that, among the reviews analyzed, users are more likely to recommend Actos for Type 2 Diabetes. Figure 8(b) illustrates Birth Control for Endometriosis- representing the recommendation status for using Birth Control as a treatment for Endometriosis. The figure suggests that, based on user reviews, Birth Control is more often recommended for Endometriosis. Figure 8(c) describes Cryselle for Abnormal Uterine Bleeding- associated with the drug Cryselle being recommended for Abnormal Uterine Bleeding. The figure indicates that Cryselle is more likely to be recommended for this condition. Figure 8(d) illustrates Afrezza for Diabetes, Type 2- Lastly, 8d corresponds to the drug Afrezza being recommended for Type 2 Diabetes. The figure implies that Afrezza is more frequently recommended for Type 2 Diabetes.

Moreover, PharmaBERT can be used directly in the recommendation system, which involves analyzing the semantic content of the text to see if it recommends a drug for specific cases. Recommendations do not only consider numerical ratings but also analyze textual reviews to give more informed advice to health professionals and patients.

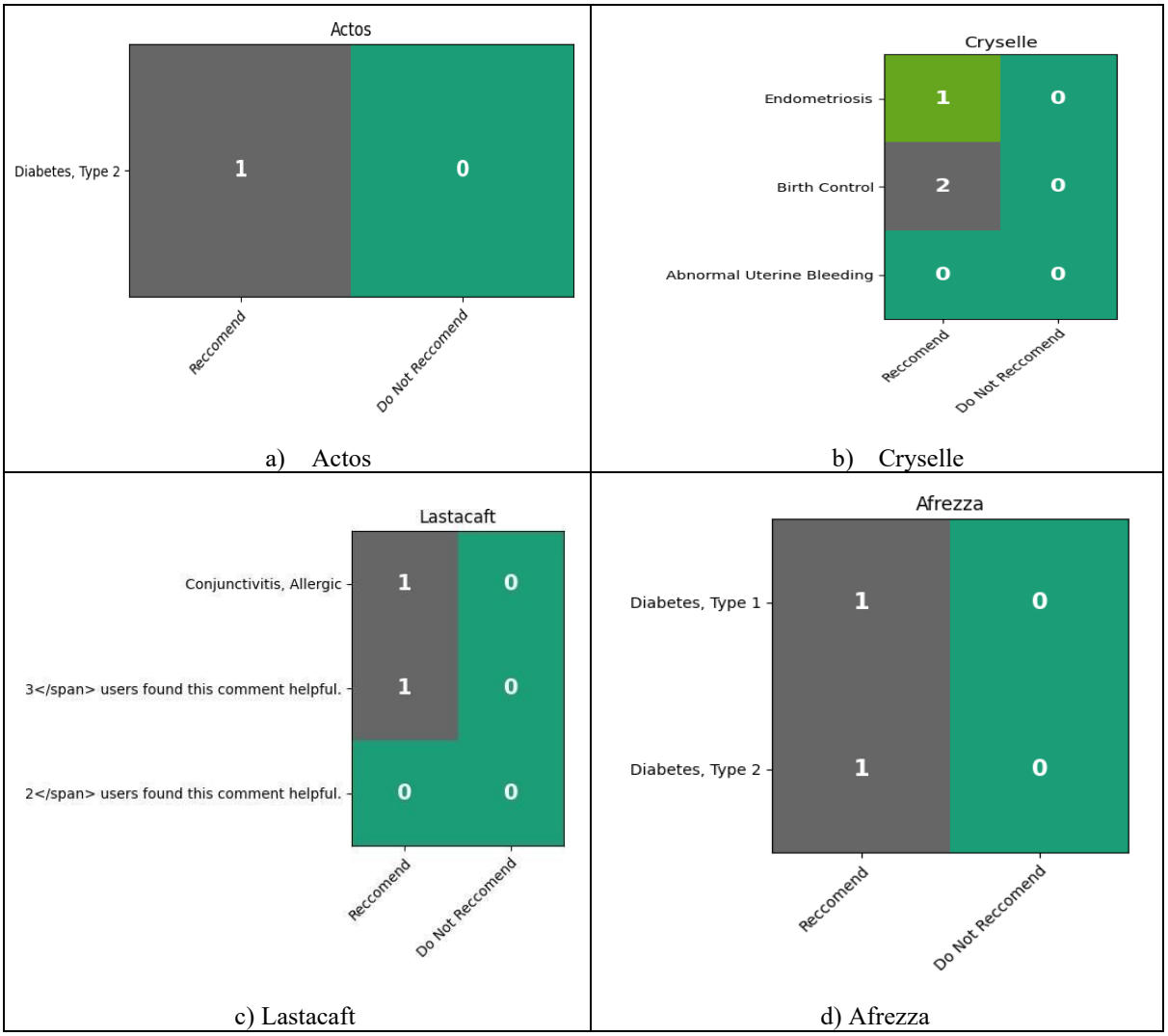

Figure 8(a)-(d). Recommend / Do Not Recommend Weight Matrix.

5.5. Comparative Analysis Against State-of-the-art Approaches

Before comparing PharmaBERT with the research study on "***Aspect-based sentiment analysis in drug reviews based on hybrid feature learning***" Sweidan, A. H., El-Bendary, N., & Al-Feel, H. [19] and ***Sentiment analysis in drug reviews based on improved pre-trained word embeddings***, authored by Bensalah, N., Ayad, H., Adib, A., & Farouk, A. I. el. [20], it's essential to highlight the gaps identified in their papers and how PharmaBERT addresses these shortcomings.

[19] proposed a hybrid feature learning approach for aspect-based sentiment analysis of adverse drug reactions (ADRs) reviews, but has limitations such as applicability to other domains, inability to capture complex relationships and dependencies between words, and use of Bi-LSTM for sentiment classification which may not handle long sequences and vanishing gradient problem.

[20] proposed a hybrid approach for drug reviews, but it has limitations such as limited domain, imbalanced dataset, static word embeddings, lack of explainability, and limited evaluation metrics. The domain is limited to drug reviews from Drugs.com, and the imbalanced dataset may affect the model's performance for classes with fewer instances. The use of static word embeddings may not capture the contextual meaning of words, and the complex neural network architecture may be difficult to interpret.

Based on the above comparison, here are the key points on how PharmaBERT is better than BERT-LDA model and BiGRU-CNN model.

- PharmaBERT addresses these limitations by proposing a framework that combines patient feedback with evidence-based best practices across disease states to improve drug recommendations.
- By using BERT as its free-text processing engine for sentiment judgment and classification, PharmaBERT can capture complex relationships and dependencies between words in a sentence.
- PharmaBERT also uses transfer learning, allowing it to adapt quickly to the intricacies of drug reviewing, making it more suitable for real-world applications.
- Additionally, PharmaBERT uses a more advanced sentiment analysis technique compared to Bi-LSTM, which can handle long sequences and provide better choices in healthcare decision-making processes.

Overall, PharmaBERT addresses the limitations of the previous papers by providing a more advanced and flexible approach for sentiment analysis, handling complex relationships and dependencies between words, and providing better choices in healthcare decision-making processes.

Additionally, the research on "***Deep Neural Network-Assisted Drug Recommendation Systems for Identifying Potential Drug–Target Interactions***" authored by *Yogesh K., Shashank Y. and Durai S.* [21], discusses how AI assisted drug recommendation system can help in predicting potential interactions between drugs and their targets using machine learning methods.

However, their approach has several gaps and limitations:

1. No Comprehensive Sentiment Analysis: The focus of the paper is on predicting DTIs, but it ignores the importance of cognitive analysis in understanding patients' experiences and reactions to drug use as a cognitive component. This system cannot provide a comprehensive picture of drug testing, reducing its effectiveness in delivering real-world drug recommendations

2. Sequential Information Not Well Exploited: This paper mostly relies on sequential data like protein sequences and drug SMILES strings thereby overlooking the sequential nature of drug reviews which results in less nuanced sentiments and nonoptimal recommendations for drugs.

3. Interpretability and Transparency: Interpreting what goes into each recommended decision can be quite a challenge because deep neural network models are often uninterpretable and opaque. This lack of transparency can result in malpractice in clinical practice.

PharmaBERT addresses limitations by:

1. Incorporating advanced sentiment analysis techniques to understand the nuances of drug reviews and patient feedback.
2. Leveraging sequential information from drug reviews to capture the holistic context of patient narratives and sentiment expressions.
3. Tailoring the model specifically for healthcare settings, ensuring alignment with clinical needs and terminology.
4. Providing interpretability and transparency features to enhance trust and understanding among healthcare professionals.

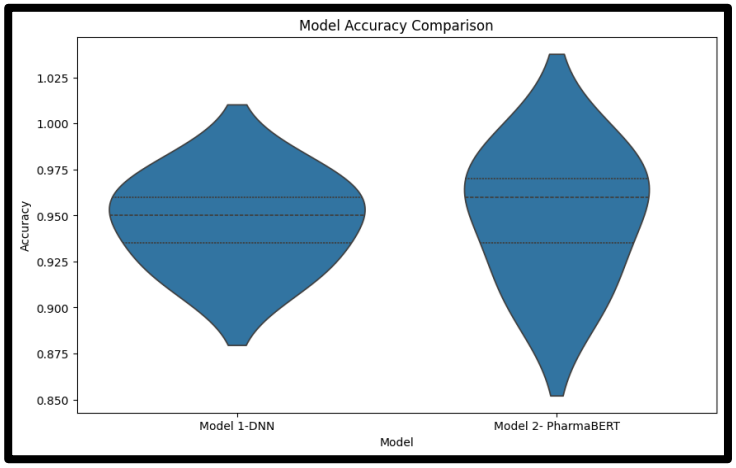

Figure 9. Performance Visualization (Violin Graph- Comparative Analysis)

Assuming [21] model to be 'DNN', when comparing Model 1(DNN) and Model 2(PharmaBERT) in terms of drug interaction predictions, PharmaBERT emerges as the preferable choice for several reasons as shown in Figure 9.

PharmaBERT:

- Exhibits a narrower range of predictions, indicating a focused approach that prioritizes identifying the most relevant or high-confidence drug interactions.
- It has a less biased algorithm leading to fewer false positives and more accurate predictions, resulting in higher overall performance.
- The violin plot consistently performs well across different conditions or iterations, having a higher median accuracy value that indicates better results every time.
- A narrower shape of the violin plot shows that it has less variation in performance implying more confidence in its predictions.

In contrast, DNN:

- Has a broad coverage of predictions due to its inclusive nature or complicated algorithm.
- Though it may capture a larger number of related medications, it also increases variability, resulting in less reliable predictions
- The wider shape of its violin plot implies larger volatility in performance with lower median accuracy value indicating less consistent results.

When we look at the characteristics observed from the violin plot and combined woodcock's AUC-ROC and AP scores (see Figure 10), PharmaBERT seems to be the best model for accurate prediction of drug interactions. This leads to a much tighter, more consistent approach than DNNs.

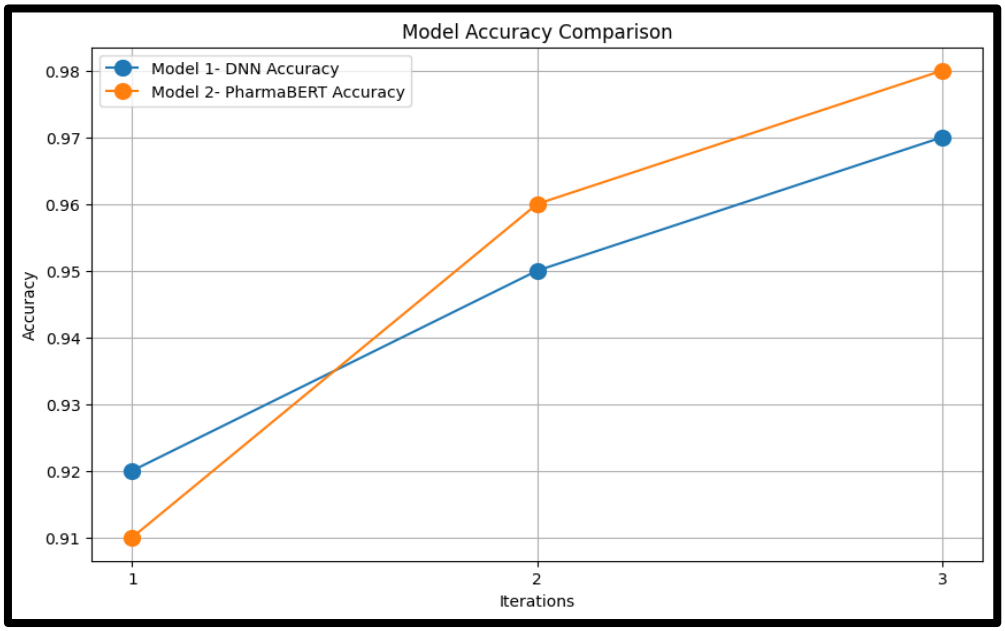

Figure 10. Model Accuracy Comparison

The Drug Model 2 performs better than Model 1, DNN in all iterations as PharmaBERT exhibits higher accuracy than the former. The difference starts at the fourth iteration and graduates to increase over the following iterations. This progression leads to a better overall performance for Model 2 thus making it outperforming Model 1 when it comes to predicting the intended objective. Thus, sentiment analysis of user-generated drug reviews may be deemed accurate with the incorporation of state-of-the-art language models such as BERT models. As opposed to traditional systems that commonly employ numerical scale, PhamaBERT pays attention to adjectives and sentences that appeal to subtle sentiments allowing for a better understanding of user experiences and attitudes.

The system of the PharmaBERT is more contextually related and provides more individual medication recommendations via the fact of BERT directly being implemented in the process of recommendations. Besides, this novel technique does not only help achieve a better prediction but also enhance the overall comprehension on how best the drug can play out its action in addition to patients' satisfaction which in the long run improves healthcare decision making for both patients and doctors.

Another factor contributing to PharmaBERT's efficacy as compared to the approach stated is that it was purposefully built for drug recommendation in the healthcare field, while the approach mentioned in the DNN research paper may only predict drug-target interactions but is not necessarily useful for all clinical practice aspects. In these tasks, when it is meant to have sentiment analysis and recommendation, such fine-tuning techniques as those in PharmaBERT have been invented, to increase precision and relevance in the outputs that could lead to good decisions. One of the main advantages of the certain approach of PharmaBERT is its clarity; it can be easily followed since the attention mechanisms are made clear.

Consequently, there still impedes several issues for using BERT to suggest drugs. The adoption of Bert model techniques requires a team of experts who must learn a lot about the process of fine-tuning and data structures and the fact that some organizations are limited by budget constraints. With this in mind, policymaking and enforcing laws that regulate patient data privacy become the utmost. Similarly, the fact that there are certain terms and subtleties which are provided in the drug reviews makes the applicability of BERT very limited. Hence, the requirement of deep-ranged fine-tuning is critical for the accurate sentiment analysis.

6. CONCLUSION

The extended functionality of PharmaBERT to the drug recommendation system results in one closer to the forefront of the healthcare area. It has reached an impressive achievement of 91% accuracy in predicting drug recommendation by using PharmaBERT as the main agent with the help of the powerful natural language processing and topic computed data. Furthermore, the MCC which measures sensitivity and specificity is ranked at 80%, indicating that system performs well in identifying positive and negative recommendations. PharmaBERT's key function is to analyze and extract the subtleties present in the patient reviews while encoding it further for precise drug suggestions and personalized prescriptions. The study focuses on evidence-based practice, which means that all datasets, including those with noise and inherent variability due to real-life healthcare impact, are accessible and used wisely. The study employs pie charts and highly readable recommendation graphs to help users understand patients' preferences and treatment efficacy rates clearly. The use of PharamaBert and BERT can make personalized healthcare highly efficient. Nevertheless, there is some degree of limitations which follow the utilization of the BERT technology. The need for high processing computational power and specialized knowledge to finetune and deploy BERT, potentially hinders adoption by certain healthcare facilities. Furthermore, maintaining patient data privacy and security during BERT-based sentiment analysis is crucial, particularly adhering to regulations like HIPAA. Additionally, BERT may struggle with domain-specific terminology and nuances present in drug reviews, requiring careful fine-tuning and customization to ensure accurate sentiment analysis.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

J Jayapradha: Project Administration, Supervision, Writing – Review & Editing;
Yukta Kulkarni: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Lakshmi Vadhanie G: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft

Preparation;
Palanichamy Naveen: Supervision, Review;
Elham Annam: Supervision, Review.

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/.

No human or animal subjects were involved in this study, and no data collection was performed from social media platforms.

REFERENCES

[1] H. Ng, A. Alias bin Mohd Azha, T. T. V. Yap, V. T. Goh, "A Machine Learning Approach to Predictive Modelling of Student Performance," *F1000Research*, vol. 10, no. 1144, pp. 1-10, 2022, doi: 10.12688/f1000research.73180.2.

[2] W. K. Chong, H. Ng, T. T. V. Yap, W. K. Soo, "Objectivity and Subjectivity Classification with BERT for Bahasa Melayu," *International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022),* pp. 246-257, Dec. 2022, doi: 10.2991/978-94-6463-094-7_20.

[3] Y. Wang, Z. Jiang, C. Zheng, F. Yang, Y. Zhou, E. Cho, X. Fan, X. Huang, Y. P. Lu, Y. Yang, "RecMind: Large Language Model Powered Agent for Recommendation,", pp. 4351-4364, 2024.

[4] J. Zhang, K. Bao, Y. Zhang, W. Wang, F. Feng, X. He, "Is ChatGPT Fair for Recommendation? Evaluating Fairness in Large Language Model Recommendation," *ACM Conference on Recommender Systems*, pp. 993-999, 2023, doi: 10.1145/3604915.3608860.

[5] L. Wang, E. Lim, "Zero-Shot Next-Item Recommendation using Large Pretrained Language Models," *arXiv preprint arXiv:2304.03153*, 2023.

[6] K. Bao, J. Zhang, Y. Zhang, W. Wang, F. Feng, X. He, "TALLRec: An Effective and Efficient Tuning Framework to Align Large Language Model with Recommendation," in *ACM Conference on Recommender Systems*, pp. 1007-1014, 2023, doi:10.1145/3604915.360885.

[7] Z. Yue, S. Rabhi, G. de Souza Pereira Moreira, D. Wang, E. Oldridge, "LlamaRec: Two-Stage Recommendation using Large Language Models for Ranking," *arXiv preprint arXiv:2311.02089*, 2023, doi:10.48550/arXiv.2311.02089.

[8] J. Ji, Z. Li, S. Xu, W. Hua, Y. Ge, J. Tan, Y. Zhang, "GenRec: Large Language Model for Generative Recommendation," *Lecture notes in computer science*, pp. 494–502, 2024, doi: https://doi.org/10.1007/978-3-031-56063-7_42.

[9] Z. Liu, P. Wang, Y. Li, J. D. Holmes, P. Shu, L. Zhang, C. Liu, N. Liu, D. Zhu, X. Li, Q. Li, S. H. Patel, T. T. Sio, T. Liu, W. Liu, "RadOnc-GPT: A Large Language Model for Radiation Oncology," *arXiv preprint arXiv:2309.10160*, pp.1-15,2023.

[10] I. Singh, V. Blukis, A. Mousavian, A. Goyal, D. Xu, J. Tremblay, D. Fox, J. Thomason, A. Garg, "ProgPrompt: Generating Situated Robot Task Plans using Large Language Models," *IEEE International Conference on Robotics and Automation*, Vol. 47, pp. 999-1012, 2023, doi: 10.1007/s10514-023-10135-3.

[11] R. Li, W. Deng, Y. Cheng, Y. Zheng, J. Zhang, F. Yuan, "Exploring the Upper Limits of Text-Based Collaborative Filtering Using Large Language Models: Discoveries and Insights," *arXiv preprint arXiv:2305.11700*, 2023.

[12] J. Zhang, R. Xie, Y. Hou, W. X. Zhao, L. Lin, and J.-R. Wen, "Recommendation as Instruction Following: A Large Language Model Empowered Recommendation Approach," *arXiv preprint arXiv:2305.07001*, 2023, doi: 10.48550/arXiv.2305.07001.

[13] A. Lahat, E. Shachar, B. Avidan, B. S. Glicksberg, E. Klang, "Evaluating the Utility of a Large Language Model in Answering Common Patients' Gastrointestinal Health-Related Questions: Are We There Yet?," *Diagnostics,vol.* 13, No.11,2023, doi: 10.3390/diagnostics13111950.

[14] J. Liu, L. Li, T. Xiang, B. Wang, Y. Qian, "TCRA-LLM: Token Compression Retrieval Augmented Large Language Model for Inference Cost Reduction," in *Conference on Empirical Methods in Natural Language Processing*, pp. 9796-9810,2023, doi: 10.48550/arXiv.2310.15556.

[15] Y. Wang, Z. Chu, X. Ouyang, S. Wang, H. Huang, Y. Shen, J. Gu, S. Xue, J. Y. Zhang, Q. Chen, L. Li, J. Zhou, S.

Li, "Enhancing Recommender Systems with Large Language Model Reasoning Graphs", *Information Retrieval, 2023*, doi: 10.48550/arXiv.2308.10835.

[16] S. Mysore, A. McCallum, H. Zamani, "Large Language Model Augmented Narrative Driven Recommendations," *ACM Conference on Recommender Systems*, pp. 777-783, 2023, doi: 10.1145/3604915.3608829.

[17] J. Zhou, X. Gao, "SkinGPT-4: An Interactive Dermatology Diagnostic System with Visual Large Language Model," 2023, doi: 10.48550/arXiv.2304.10691.

[18] P. Murali, I. Steenstra, H. S. Yun, A. Shamekhi, T. Bickmore, "Improving Multiparty Interactions with a Robot Using Large Language Models," *CHI Extended Abstracts*, pp. 1-8, 2023, doi: 10.1145/3544549.3585602.

[19] A. H. Sweidan, N. El-Bendary, H. Al-Feel, "Aspect-based sentiment analysis in drug reviews based on hybrid feature learning," in 16th International Conference on Soft Computing Models in *Industrial and Environmental Applications (SOCO 2021)*, pp. 78–87, Springer International Publishing, 2022, doi:10.1007/978-3-030-87869-6_8.

[20] N. Bensalah, H. Ayad, A. Adib, A. I. el Farouk, "Sentiment analysis in drug reviews based on improved pre-trained word embeddings," in *Innovations in Smart Cities Applications Vo. 6*, pp. 87–96, Springer International Publishing, 2023, doi: 10.1007/978-3-031-26852-6_8.

[21] Y. Kalakoti, S. Yadav, D. Sundar, "Deep neural network-assisted drug recommendation systems for identifying potential drug–target interactions," *ACS Omega*, vol. 7, no. 14, pp. 12138–12146, 2022, doi: 10.1021/acsomega.2c00424.

BIOGRAPHIES OF AUTHORS

Jayapradha J is an Assistant Professor at the department of Computing Technologies, SRMIST. She completed her Ph.D(CSE) in Privacy Preservation and currently, possesses a challenging postdoctoral position at Multimedia University, Malaysia. She published 22 research articles on ML, data mining, and privacy. Moreover, she has 6 published patents and 2 granted patents in the field of ML and data privacy. Her research interests include ML, databases, and privacy. She can be contacted by email: jayapraj@srmist.edu.in

Yukta Kulkarni is an MS in Computer Science student at New York University. She earned her B.Tech in Computer Science and Engineering from SRM Institute of Science and Technology. She published a book chapter on data poisoning along with research internships on ML Security vulnerabilities and neuro-symbolic AI. Furthermore, she received various awards such as the Academic Winner Award and the Overall Best IE(I) Student Award. Her area of interests includes cybersecurity, artificial intelligence, and data science. She can be contacted by email: yk8371@srmist.edu.in

Lakshmi Vadhanie is a student pursuing her master's in data science at Northeastern University. Her research interests include machine learning, artificial intelligence, cybersecurity, statistics, and computer vision. She has published a book chapter on data poisoning and contributed to various AI-ML projects, such as AI voice bot development and benchmark settings and comparisons for different language models. She can be contacted by email: gv3418@srmist.edu.in

Palanichamy Naveen joined the Faculty of Computing and Informatics, Multimedia University after receiving Ph.D from Curtin University, Malaysia. She received her Bachelor of Engineering (CSE) and Master of Engineering (CSE) from Anna University, India. Her area of research interests includes Smart Grid, Cloud Computing, Machine Learning, Deep Learning and Recommender system. She is involved in multiple research projects funded by Multimedia University. She can be contacted by email: p.naveen@mmu.edu.my.

Elham Abdulwahab Anaam is a Postdoctoral at Faculty of Computing and Informatics, Multimedia University. Her research focuses on E-CRM adoption, multi-criteria decision-making, hybrid recommendation system and blockchain technology. She can be contacted by email: anaamelham@mmu.edu.my.