Interpretable One-Class Classification Framework for Prescription Error Detection Using BERT Embeddings and Dimensionality Reduction

Yassine OUZAR[a,*], Faiza AJMI[b], Sarah BEN OTHMAN[c], Chloé Rousselière[d], Bertrand DECAUDIN[e], Pascal ODOU[e], Slim HAMMADI[f]

[a] *UMR 9189 CRISTAL, Univ. Lille, CNRS, F-59000 Lille, France*
[b] *ICL, JUNIA, Université Catholique de Lille, LITL, F-59000 Lille, France*
[c] *Polytech Lille, UMR 9189 CRISTAL, Univ. Lille, CNRS, F-59000 Lille, France*
[d] *CHU Lille, Institut de Pharmacie, 59000 Lille, France*
[e] *Univ. Lille, CHU Lille, ULR 7365 - GRITA, F-59000 Lille, France*
[f] *Centrale Lille, UMR 9189 CRISTAL, Univ. Lille, CNRS, F-59000 Lille, France*

Abstract

Ensuring accurate prescriptions and proper medication administration is critical for patient safety and effective clinical outcomes. Identifying and preventing prescription errors can significantly reduce healthcare costs and adverse health effects. Current solutions range from rule-based systems, which rely on predefined rules and clinical expertise but lack adaptability for unexpected errors, to supervised machine learning approaches, which are hindered by limited labeled error data and opaque algorithmic processes To overcome these limitations, we propose a prescription error detection method based on one-class classification approach. Leveraging the publicly available MIMIC database, advanced language modeling and dimensionality reduction techniques, our framework autonomously learns meaningful representations of medication prescriptions without requiring on explicit error labels. Addi-

[*] Yassine OUZAR

Email addresses: yassine.ouzar@univ-lille.fr (Yassine OUZAR), faiza.ajmi@univ-catholille.fr (Faiza AJMI), sarah.ben-othman@polytech-lille.fr (Sarah BEN OTHMAN), chloe.rousseliere@chu-lille.fr (Chloé Rousselière), bertrand.decaudin@univ-lille.fr (Bertrand DECAUDIN), pascal.odou@univ-lille.fr (Pascal ODOU), slim.hammadi@centralelille.fr (Slim HAMMADI)

tionally, we apply the SHAP method to explain the model's predictions, providing clinicians with interpretable insights into the decision-making process and enhancing trust in the model's reliability. Three experiments were conducted to evaluate the effectiveness of our approach. The results reveal that leveraging BERT embeddings in conjunction with PCA for dimensionality reduction and Local Outlier Factor-based one-class classification achieves the highest performance, with : *precision* = 81.71% ; *Recall* = 87.32% ; *F1-score* = 86.84%. These results highlight our method's effectiveness in detecting potential prescription errors without the need for labeled error data.

Keywords: medication iatrogeny, one-class classification, dimensionality reduction, BERT, explainable AI

1. Introduction

Pharma 4.0, a new era powered by cutting-edge technologies, is revolutionizing the pharmaceutical industry. By leveraging artificial intelligence, Internet of Things (IoT), big data, and advanced automation, this digital revolution aims to optimize every stage of the medication lifecycle from discovery and development to production, distribution, and ultimately, patient care [1]. This interconnected and intelligent ecosystem promises significant improvements in medication efficacy, safety, and accessibility. However, despite these remarkable technological advancements, medication-related harm remains a significant public health concern. This underscores the need for innovative solutions aimed at minimizing the risks associated with medication use.

Medication-related iatrogeny is a prevalent concern in the healthcare sector due to its impact on health and economic aspects [2]. While medications are essential for treating various health problems, they also carry potential risks. Factors such as prescription errors, allergic reactions, drug interactions, and incorrect dosages can pose significant threats to patient safety [3, 4]. Studies show a significant percentage of hospitalized patients experience medication-related side effects, some of which are preventable [5].

Preventing inappropriate medication prescriptions is crucial in healthcare to reduce medication-related issues and to safeguard patient well-being [6]. Medication errors, although possible at any treatment stage, most commonly arise during the prescription and administration phases [7]. Potentially inappropriate prescriptions often result from omissions or errors in the prescribed medication's duration, dosage form, route, frequency, or quantity [3]. According to Nanji et al.'s research, 42.9% of medication prescribing errors in outpatient settings involved administering the medication via the wrong route, 28.6% were caused by prescribing the incorrect dosage frequency, and 17.1% were due to prescribing an excessive dose [8]. Research conducted by Al-khani et al. revealed that medication prescribing errors in outpatient settings were most prevalent in wrong dose (53%), followed by wrong medication (10%) and wrong frequency (9%) [4]. A study by Shrestha and Prajapati found that medication prescribing errors in outpatient settings were most common in the form of potential drug-drug interactions (10.2%), followed by prescribing the wrong strength (0.5%) and the wrong medication name or dosage form (0.2%) [9]. Kaushal et al. found that most potential adverse drug events arose during the medication ordering process (79%) and most medication errors primarily due to inaccurate drug dosages (28%), followed by route of administration (18%), medication administration record transcription and documentation (14%), date (12%), and frequency of administration errors (9.4%) [10]. The most prevalent medication errors in neonates involve wrong dose (28%) and incorrect route (29%), as reported by Eslami et al. [11].

Studies agree that systematic analysis of prescriptions by a pharmacist plays a crucial role in significantly mitigating the risk of medication errors [12, 13]. However, conducting a comprehensive and high-quality analysis of prescriptions is challenging due to time constraints and staffing shortages [14]. In response to these challenges, the healthcare community has actively sought solutions by developing a range of tools designed to optimize point-of-care prescription processes. Among these tools are Computerized Physician Order Entry (CPOE) systems and Computerized Clinical Decision Support (CCDS) alert systems [15]. While these technologies hold promise, they come with their own set of issues. CPOE, for instance, can inadvertently introduce new prescribing errors, increasing complexity. On the other hand, CCDS systems often identify only a limited number of errors and may generate excessive false alerts, which can lead to "alert fatigue" and disrupt workflow efficiency [5].

Existing prescription errors detection systems can be categorized into two types: knowledge-based and non-knowledge-based [16, 17].

1.1. Knowledge-based systems

Knowledge-based systems rely on clinical knowledge and employ IF-THEN rules, incorporating literature, practice, or patient data to generate recommendations and alerts [16]. These systems are easily automated and can be applied to electronic patient data [18, 19]. However, they come with several drawbacks. The formulation of rules requires high clinical prior knowledge from human experts, which is both laborious and time-consuming process. Furthermore, these systems can only identify errors explicitly programmed in their rules [20]. This rigidity means they often miss novel or unseen errors, potentially overlooking critical issues [21]. Their rule-based nature also leads to excessive alerts, causing 'alert fatigue' among healthcare professionals [20].

1.2. Non-knowledge-based systems

In contrast to rule-based systems that rely on predefined explicit rules and expert knowledge, non-knowledge-based systems leverage machine learning algorithms to autonomously derive patterns and make decisions from data [16, 22]. These data-driven approaches are gaining popularity due to their adaptability and ability to detect novel unseen errors. However, some of these systems are commercial software that do not disclose any information about the algorithms they employ [23], while others are trained on private electronic health data. This lack of transparency and reliance on proprietary data hinder the reproducibility and evaluation of the algorithms.

Within this framework, this paper introduces a machine learning-based approach for detecting valid and invalid prescriptions. Unlike previous studies that relied on private databases, our method leverages the MIMIC database, a comprehensive and publicly available large-scale dataset. To the best of our knowledge, this is the first attempt to use a public dataset for medication error detection, promoting broader adoption and transparency in this field. Furthermore, we consider this task as a one-class classification problem, given the limitation of having only valid prescriptions available for training.

The main contributions of this study are summarized as follows:

1. We propose a novel transformer-based approach for valid/non-valid prescription detection.

2. By leveraging transformers, our model can effectively capture contextual information across various segments of the prescription text. This capability greatly aids in identifying inaccuracies in names, values, or the order of features within sequences.

3. We are the first to leverage the MIMIC database for identifying invalid prescriptions.

4. We integrate AI explainability techniques to help clinicians grasp the reasons behind the predictions and assess the model's reliability.

5. We evaluated our approach using both traditional one-class classification algorithms and cutting-edge deep learning models. The best results were achieved by combining the Local Outlier Factor (LOF) algorithm with transformer-based text embeddings and dimensionality reduction.

2. Related work

2.1. Prescription errors detection

Over the past decades, deep learning models have achieved significant breakthroughs across various domains, including healthcare. However, their integration into clinical settings has been relatively slow and limited. This is mainly due to several factors, such as the regulatory and ethical considerations, data availability, accessibility and quality issues, as well as the opacity of model decisions [24, 25].For these reasons, previous research on detecting invalid prescriptions has predominantly utilized rule-based systems, which do not require large scale datasets for training and provide inherently interpretable results. These systems leverage clinical expertise and use IF-THEN rules, drawing on medical literature, established practices, or patient data to generate alerts and recommendations [16, 26].

While these rule-based approaches are relatively simple to implement and integrate into clinical workflows [18, 26], they come with significant limitations. The formulation of rules requires high clinical prior knowledge from human experts, a process that is both time-consuming and labor-intensive. Moreover, such systems are confined to detecting only those errors that have been explicitly encoded into their rules [20]. This rigidity often results in their inability to catch new or unexpected errors, potentially missing critical issues [21].

To overcome the limitations of rule-based systems, some studies have turned to machine learning models that make decisions without relying on predefined rules [16]. These models leverage their capacity to learn from data patterns, providing greater adaptability and the potential to identify previously unseen errors. However, many of these systems are trained on private electronic health data, hindering reproducibility and evaluation, while others are proprietary software, with no transparency about the algorithms they use [23].

2.2. One-class classification for anomaly detection

To address the challenge of having only valid (normal) data available, one-class classification (OCC) algorithms were specifically designed to handle such imbalanced datasets. Unlike traditional supervised classification, OCC adopts an unsupervised approach to avoid classifying all data into a single category. The goal of OCC is to identify whether a given example belongs to the same distribution as the training data. By modeling the characteristics of the normal class, OCC algorithms can effectively distinguish between anomalies/outliers and the normal/inlier data that is accessible during training, even in the absence of labeled examples of these outlier cases [27].

Various OCC algorithms have been proposed which can be devided into classical and deep learning methods. Classical techniques rely on the hypothesis of generating a score function based on the characteristics of the single known class. These functions can be categorized as either likelihood-based (normal/seen score) by indicating how closely a new instance resembles the known class, or dissimilarity-based (anomaly/unseen score) by quantifying how different a new instance is from the known class. Ultimately, a threshold value is determined as a decision boundary to classify data into the known (valid) class or other class [28].

Tax et al. introduced support vector data description (SVDD) [29], which defines a hypersphere around the data distribution. Points inside this hypersphere are classified as normal/inlier, while those outside are considered outlier or unseen. In contrast, the one-class support vector machine (OCSVM) proposed by Schölkopf et al. aims to construct a hyperplane that separates the data from the origin with maximum margin [30]. Points on the same side of the hyperplane as the data are classified as normal/inlier, while those on the opposite side are considered outlier or unseen. Despite their different geometric interpretations, SVDD and OCSVM become equivalent when a Gaussian kernel is applied [31]. This kernel transforms the data into a high-dimensional feature space where the decision boundary adopts a spherical form. In this space, the hyperplane used by OCSVM looks like a sphere in the original data, making it functionally equivalent to SVDD. Local Outlier

Factor (LOF), proposed by Breunig et al., measures the local density of each data point compared to its neighbors [32]. LOF assumes outliers have significantly lower local densities than their neighbors, making it effective for detecting anomalies in datasets with varying densities. In contrast, Isolation Forest, developed by Liu et al., uses a random tree structure for outlier detection [33]. It assumes that outlier samples become isolated within the tree structure, while normal samples are grouped together. This approach differs from density-based methods by focusing on the isolation of data points rather than their proximity to neighbors.

Traditional one-class classification methods mentioned above, often struggle when dealing with scenarios involving high-dimensional data and large datasets due to computational inefficiency and the curse of dimensionality. These methods usually require significant feature engineering to perform well. Deep learning, on the other hand, can automatically learn important and complex features from high-dimensional data, outperforming traditional approaches. Deep one-class classification methods typically leverage deep neural networks to learn and extract latent representations of normal data, which can then be used to identify anomalies or unseen categories based on how much a new data point diverges from this learned representation [34].

Several neural network architectures are commonly used in deep one-class-classification (DOCC). Classically, DOCC used autoencoders that are trained to compress and reconstruct input data. The underlying assumption is that normal data and outliers will generate distinct latent representations, and the differences in their reconstruction errors can be leveraged to distinguish between these two categories of samples [35, 36, 37]. More recently, another set of methods has utilized Generative Adversarial Networks (GANs) to model the density distribution of the training data, enabling the detection of outliers when a sample demonstrates low density [38, 39, 40]. Finally, the third category includes approaches such as DeepSVDD [34] and DROCC [41] which focus on optimizing a one-class loss function specifically designed for deep learning architectures [42].

3. Materials and Methods

3.1. Dataset

The availability of large datasets and advanced neural architectures has played a crucial role in the success of deep learning, particularly in computer vision and natural language processing. However, the absence of large-scale

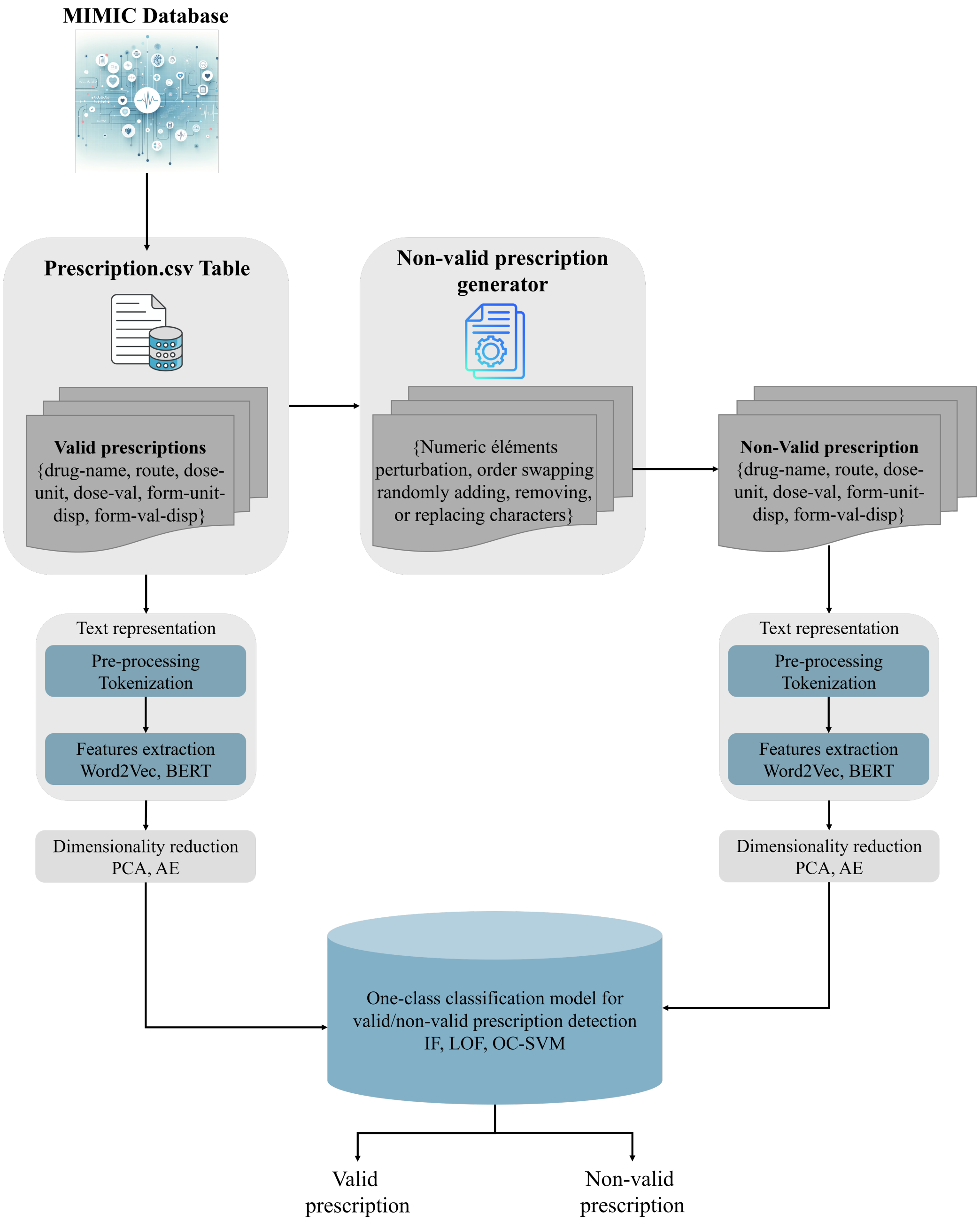
MIMIC Database
Prescription.csv Table
Valid prescriptions
{drug-name, route, dose-unit, dose-val, form-unit-disp, form-val-disp}
Non-valid prescription generator
{Numeric éléments perturbation, order swapping randomly adding, removing, or replacing characters}
Non-Valid prescription
{drug-name, route, dose-unit, dose-val, form-unit-disp, form-val-disp}
Text representation
Pre-processing Tokenization
Features extraction Word2Vec, BERT
Dimensionality reduction PCA, AE
Text representation
Pre-processing Tokenization
Features extraction Word2Vec, BERT
Dimensionality reduction PCA, AE
One-class classification model for valid/non-valid prescription detection IF, LOF, OC-SVM
Valid prescription
Non-valid prescription

Figure 1: Overview of the proposed framework for valid/non-valid prescription detection. Initially, synthetic non-valid prescriptions are generated from the input prescription data sourced from the MIMIC database by introducing various perturbations, such as numeric element modifications, order swapping, and character manipulation. Subsequently, both valid and non-valid prescriptions undergo text preprocessing and tokenization before being transformed into feature vectors using Word2Vec or BERT embeddings. Dimensionality reduction techniques, such as PCA and AE, are applied to reduce the dimensionality of the feature vectors. Finally, one-class classification models, including Isolation Forest (IF), Local Outlier Factor (LOF), and One-Class SVM (OC-SVM), are trained on the reduced feature representations to distinguish between valid and non-valid prescriptions.

public datasets specifically for prescription error detection has hindered the application of deep learning models in this area.

In this study, we leverage the Medical Information Mart for Intensive Care (MIMIC) database, a comprehensive and freely accessible critical care dataset created by the Massachusetts Institute of Technology. Now in its fourth version, MIMIC contains anonymized data from more than 40,000 patients who were treated in the critical care units at Beth Israel Deaconess Medical Center. This database supports a wide range of applications, including academic and industrial research, quality improvement projects, and educational programs [43].

3.2. Proposed framework

The overarching framework for valid/non-valid prescription detection is illustrated in Figure 1. We treat this task as a one-class classification problem. First, a pre-processing step is carried out on the input prescription data acquired from the MIMIC database to handle missing values and ensure proper data formatting. Concurrently, non-valid prescription generation is performed as the dataset used in our experiment consists solely of valid prescriptions. Next, features are extracted using text embedding techniques. Subsequent to text vectorization and features extraction, dimensionality reduction techniques are applied. Finally, the reduced feature set is inputted into one-class classification algorithms for the classification of valid/non-valid prescriptions.

3.2.1. Non-valid prescriptions generation

The MIMIC prescription dataset exclusively contains valid prescriptions. However, having non-valid prescriptions is essential to evaluate the performance of machine learning models. To achieve this, we generate synthetic erroneous prescriptions by introducing plausible errors to the raw valid prescriptions from the MIMIC dataset.

This process involves systematically modifying key prescription elements, such as medication names, dosages, units, and routes of administration, to create realistic yet incorrect entries that simulate common clinical errors. These synthetic errors include dosage errors, unit inconsistencies, medication name confusions, inappropriate routes of administration, and frequency inaccuracies.

For text elements, such as drug_name, dose_unit_rx, form_unit_disp, and route, incorrect names were generated by randomly adding, removing, or replacing characters in one or multiple input fields. Numeric elements, like dose_val_rx and form_val_disp, were perturbed to simulate over-dosing and under-dosing scenarios. Additionally, order swapping was applied, which placed correct elements in incorrect fields, leading to inconsistencies. Lastly, we randomly swapped medication names among prescriptions where other elements differed from the valid entries.

By generating this synthetic incorrect prescriptions, we aimed to rigorously test our model's error detection capabilities across various error types and severities, enabling a comprehensive performance evaluation. However, while this synthetic errors generation may not be the optimal way for effectively training and validating a machine learning model, it can still serve as a viable solution in the absence or lack of data.

3.2.2. pre-processing

MIMIC is structured into 26 tables that encompass a variety of data types, including demographic information, vital signs, lab results, procedures, medications, clinical notes, and more. Among the various tables, we have focused specifically on the prescriptions table which is particularly relevant for our research. Each row corresponds to a unique prescription and includes detailed information about the medication prescribed. While the table contains 21 columns of prescription-related information, we focused on six key elements: drug_name, dose_val_rx, dose_unit_rx, form_val_disp, form_unit_disp, and route. Previous studies have identified these features as the most common types of prescription errors [3, 4].

Considering the large volume of data available in MIMIC-IV, we performed a data pre-processing and cleaning process using the NLTK library. The process involved converting text to lowercase, removing special characters, splitting prescription elements into words, eliminating common stop words, and removing rows with missing information to ensure that all prescription data entries are correct. Additionally, we merged the columns of prescription information into a unified text format to leverage natural language processing techniques. This helps identify potential errors by analyzing linguistic patterns and contextual relationships within the prescription text.

3.2.3. Prescription encoding

Following the pre-processing stage, which converts key prescription elements into a consistent text format, we apply advanced text encoding techniques to transform these prescription texts into high-dimensional numerical vectors. This process aims to capture and learn the detailed structure of prescription data. For this task, we leverage two state-of-the-art natural language processing methods—BERT and Word2Vec—to capture both semantic relationships and the sequential order of prescription elements. BERT belongs to the category of contextual methods. This model generates context-aware embeddings, providing multiple representations of a word based on its surrounding context [44]. Word2Vec, on the other hand, is a static text embeddingss technique that converts words into dense vector representations. This vector captures semantic relationships based on its co-occurrence with other words [45]. By applying such techniques, the analysis moves beyond basic word-to-vector conversion, embracing a richer representation that enhances the detection of nuanced errors and unusual patterns, especially those stemming from variations in the sequence of prescription information.

3.2.4. One-class classification algorithms

One-class classification algorithms, also known as anomaly detection, are designed to identify instances that significantly deviate from the expected behavior or patterns within a dataset. In contrast to conventional classification, which aims to differentiate between numerous classes, one-class classification focuses on modeling the characteristics of the normal class and identifying instances that differ from this established norm, labeling them as anomalies or outliers [27].

Common techniques for one-class classification include:

Isolation Forests :

The Isolation Forest algorithm is an an unsupervised anomaly detection technique used to identify rare and unusual instances within a dataset [33]. This algorithm relies on randomly isolating features to identify potential anomalies. It constructs binary trees and calculates the number of splits needed to isolate data points. Outliers are expected to require fewer splits, indicating that are isolated more quickly than normal instances.

One-Class Support Vector Machine :

One-Class Support Vector Machines (One-Class SVMs) are a special type of traditional SVM designed for unsupervised one-class classification. In contrast to isolation forests, which employ a tree-based methodology, One-Class SVMs construct a hyperplane rather than trees to encapsulate the normal or majority of the data within a higher-dimensional space. Instances lying on the other side of the hyperplane are considered anomalies [30].

Similar to traditional SVMs, One-Class SVMs can leverage different kernel functions to map the original data into a higher-dimensional space, potentially allowing for effective separation between normal and anomalous data by enabling the algorithm to find optimal hyperplanes.

Commonly used kernels include the linear kernel, which assumes a linear decision boundary, and the radial basis function (RBF) kernel, which is effective in capturing non-linear patterns. Linear kernels are suitable when the data can be effectively separated by a straight line or hyperplane, while RBF kernels are versatile and suitable for non-linear data separation. They can handle complex relationships between data points by mapping them into a high-dimensional space, often leading to improved separation for non-linearly separable data. Moreover, polynomial kernels can be used to capture non-linear relationships in the data. They transform the input space into a higher-dimensional space using polynomial functions.

The choice of kernel in One-Class SVMs involves a trade-off between model complexity and its ability to accurately capture intricate patterns in the training data. This underscores the importance of experimenting with different kernels to fine-tune the algorithm and achieve optimal performance.

Local Outlier Factor :

Unlike Isolation Forest, which isolates data points to detect anomalies, the Local Outlier Factor (LOF) evaluates the local density of each data point compared to its neighbors, rather than considering the entire dataset [32]. LOF first identifies the k-nearest neighbors of each data point in the dataset, where the number of neighbors is chosen based on the data and desired sensitivity. The local reachability density is then calculated for each data point. Finally, the LOF score is calculated for each data point. This score represents the ratio of the average local density of k-nearest neighbors and its own local density. It depends on number of neighbors and the proportion of outliers in the data set (contamination). Instances with significantly lower LOF scores than to their neighbors are considered potential anomalies. This is because they have a lower local density compared their neighbors, indicating a deviation from the typical data distribution.

4. Experiments and Results

4.1. Experiment 1 : Evaluation of text embeddings-based one-class classification model performance for valid/non-valid prescription detection

Valid/non-valid prescription detection is considered as a one-class classification task. This experiment aims to evaluate the effectiveness of combining text embeddings (Word2Vec and BERT) with one-class classification algorithms (OC-SVM, LOF, and Isolation Forest) for detecting prescription errors within a dataset of prescriptions.

We used prescription data from MIMIC database as valid data, while non-valid prescription were synthetically generated as detailed in subsection 3.2.1. Six distinct features, namely: $drug\_name$, $dose\_val\_rx$, $dose\_unit\_rx$, $form\_val\_disp$, $form\_unit\_disp$, and $route$ were utilized. These features were combined to create a unified text representation. Subsequently, text encoding techniques, specifically Word2Vec and BERT embeddings, were employed to transform prescription elements into feature vectors. Finally, the resulting feature vectors were used to train one-class classification algorithms for valid/non-valid prescription detection task.

Tables 1 and 2 present the performance outcomes achieved using Word2Vec and BERT, respectively, coupled with one-class classification algorithms. These tables evaluate the models' performance in detecting valid and non-valid prescriptions. The assessed metrics include precision, recall, F1-score, and ROC AUC. To ensure robustness and generalizability on unseen data, a 5-fold cross-validation strategy was employed. The dataset was divided into five folds, with 80% of the data used for training in each fold and the remaining 20% used for testing. The average of each metric across all five folds is reported in Tables 1 and 2.

In terms of text embeddings performance, the BERT-based model significantly outperformed the Word2Vec-based embeddings across most evaluation metrics, demonstrating a substantial margin of improvement. Specifically, all one-class classification algorithms combined with BERT achieved substantially higher precision and greater recall, leading to enhanced F1 score that reflect a better balance between identifying errors and minimizing false positives. This marked superiority underscores BERT's effectiveness in capturing both semantic relationships and sequential order of prescription elements. Unlike Word2Vec, which generates static embeddings for each prescription element regardless of context and position, BERT provides a more nuanced and context-aware representation of prescription data. BERT's contextual-

Table 1: Valid vs Non-valid prescription classification results using Word2Vec-based word embedding and OCC algorithms.

Method	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
Isolation Forest	50.82	50.48	49.90
LOF	51.22	50.09	46.32
OC-SVM (Poly)	50.65	50.02	44.76
OC-SVM (Linear)	48.08	48.04	47.11
OC-SVM (RBF)	48.94	49.32	47.77

ized embeddings dynamically adjust based on the surrounding words, allowing it to better capture the complex relationships between medication names, dosages, administration routes, and other prescription details. Furthermore, BERT's positional embeddings encode the relative position of each prescription element, enabling the model to understand and leverage the sequential order of components within a prescription.

Among the one-class classification algorithms evaluated, the local outlier factor (LOF) exhibited superior performance, outperforming other algorithms across all three key metrics : precision, recall, and F1-score, with a significant margin of improvement. This remarkable performance highlights LOF's effectiveness in accurately identifying outliers and distinguishing valid prescriptions from non-valid ones.

LOF's effectiveness can be attributed to its density-based methodology, which excels at identifying local anomalies by comparing the density of each data point to that of its surrounding neighbors. This adaptability to varying density patterns within prescription data, along with its robustness against noise, enables LOF to effectively distinguish valid prescription from invalid ones. By focusing on local neighborhoods, LOF is especially capable of spotting contextual anomalies that global one-class classification methods might overlook.

The combination of BERT and LOF enhances the model's ability to accurately identify invalid prescriptions by capturing contextual and positional embeddings while detecting subtle inconsistencies and deviations from valid prescriptions.

Table 2: Valid vs Non-valid prescription classification results using Bert-based word embeddings and OCC algorithms.

Method	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
Isolation Forest	71.89	63.09	66.71
LOF	**79.93**	**82.62**	**80.76**
OC-SVM (Poly)	70.59	61.60	65.37
OC-SVM (Linear)	70.55	61.56	65.33
OC-SVM (RBF)	71.67	62.84	66.48

4.2. Experiment 2 : Evaluation of dimensionality reduction techniques on BERT embeddings-based one-class classification model performance for valid/non-valid prescription detection

In this experiment, we use BERT-based text embeddings, which demonstrated superior performance compared to Word2Vec, and integrate it with two prominent dimensionality reduction techniques: Auto-Encoder (AE) and Principal Component Analysis (PCA). Our objective is to evaluate the impact of dimensionality reduction on the effectiveness of valid and invalid prescription detection when combined with one-class classification algorithms.

Auto-Encoders (AEs) are powerful tools for unsupervised feature learning and dimensionality reduction [46, 47]. By training an AE to reconstruct input embeddings while learning a compact representation, we aim to extract the most salient features for accurate prescription representation.

Similarly, Principal Component Analysis (PCA) provides a linear transformation to reduce the dimensionality of high-dimensional data while preserving its essential variance [48, 49]. By projecting the BERT embeddings onto a lower-dimensional subspace, PCA allows for the extraction of the most informative features for classification.

Following the same 5-fold cross-validation methodology performed in Experiment 1 (Section 4.1), we evaluated the performance of models incorporating dimensionality reduction techniques. Specifically, we applied either an AE or Principal PCA to reduce the BERT embeddings before classification. After text vectorization using BERT, the resulting high-dimensional embeddings were compressed to a lower-dimensional space using either AE or PCA. These reduced embeddings were subsequently fed into one-class classification algorithms for valid/non-valid prescription detection. The average performance metrics across all folds are presented in Tables 3 and 4.

Table 3: Valid vs Non-valid prescription classification results using AE-Bert-based word embeddings and OCC algorithms.

Method	**Precision(%)**	**Recall(%)**	**F1 Score(%)**
AE-Isolation Forest	71.99	75.52	74.00
AE-LOF	**81.01**	**86.09**	**85.03**
AE-OC-SVM (Poly)	72.13	63.37	66.96
AE-OC-SVM (Linear)	72.08	63.30	66.90
AE-OC-SVM (RBF)	70.93	68.73	69.79

The performance comparison between PCA and AE-based dimensionality reduction methods indicates that both methods improved the performance of one-class classification algorithms. While there were slight differences between the two methods, both consistently outperformed the baseline model without dimensionality reduction. Overall, PCA-LOF achieved the highest performance metrics with a precision of 81.71%, recall of 87.32%, and F1 score of 86.84%, slightly outperforming AE-LOF, which also performed well with a precision of 81.01%, recall of 86.09%, and F1 score of 85.03%. For Isolation Forest, AE outperformed PCA by achieving higher recall (75.52% vs. 65.40%) and F1 score (74.00% vs. 68.79%), indicating its effectiveness in capturing anomalies with this algorithm. When comparing OC-SVM models, PCA generally delivered better precision than AE for polynomial and linear kernels. For instance, PCA-OC-SVM (Polynomial Kernel) recorded higher precision (74.54% vs. 72.13%) and F1 score (69.41% vs. 66.96%) than AE-OC-SVM (Polynomial Kernel). However, AE demonstrated better results with the RBF kernel, achieving a higher recall (68.73% vs. 56.64%) and F1 score (69.79% vs. 61.99%) compared to PCA.

Table 4: Valid vs Non-valid prescription classification results using PCA-Bert-based word embeddings and OCC algorithms.

Method	Precision(%)	Recall(%)	F1 Score(%)
PCA-Isolation Forest	73.91	65.40	68.79
PCA-LOF	**81.71**	**87.32**	**86.84**
PCA-OC-SVM (Poly)	74.54	66.07	69.41
PCA-OC-SVM (Linear)	75.26	56.91	62.35
PCA-OC-SVM (RBF)	72.06	56.64	61.99

The results demonstrate that incorporating dimensionality reduction techniques has led to improved performance across all evaluation metrics when compared to the baseline model without dimensionality reduction. Notably, PCA exhibited marginally better performance than the AE approach. This slight advantage of PCA can be primarily attributed to the complexity involved in optimizing AE's performance, which requires careful fine-tuning of its hyper-parameters.

This experiment underscores the importance of incorporating dimensionality reduction techniques into the model pipeline, as it enhances the efficiency and effectiveness of prescription detection models based on BERT embeddings. Additionally, it highlights the potential of PCA as a straightforward and effective method for dimensionality reduction in this context, while also emphasizing the need for careful hyper-parameter tuning when employing more complex techniques like auto-encoder.

4.3. Experiment 3 : Performance comparison across prescription error types

The aim of this experiment is to identify and assess the most discriminative types of error in prescriptions, upon which the model relies to distinguish between valid and non-valid prescriptions. By thoroughly analyzing these errors, we aim to gain deeper insights into the underlying factors influencing the model's decision-making process and classification outcomes.

To accomplish this, we levergae the model that yielded the best performance in the previous experiments and evaluate it on three categories of non-valid prescription data : prescriptions with incorrect information, prescriptions indicating overdoses or underdoses, as well as prescriptions containing order errors. Performance of LOF model coupled with BERT embeddings and PCA-based dimensionality reduction across prescription error types is shown in Table 5.

Table 5: Performance comparison of LOF model coupled with BERT embeddings and PCA for dimensionality reduction across prescription error types.

Method	**Precision(%)**	**Recall(%)**	**F1 Score(%)**
Incorrect information	**84.62**	**91.65**	**89.07**
Overdoses/Underdoses	82.39	86.47	88.48
Order errors	78.12	83.84	82.97

The obtained results show differences in performance across different types of prescription errors. Incorrect information related to medication name, route of administration, dosage form, and dosage unit were the most discriminating types of prescription errors, followed by over/under-dosage. Conversely, order errors exhibited the lowest performance, albeit with a marginal difference compared to other error types. Incorrect prescriptions are easily discriminated, because they are generally unseen in the training data (valid prescription). However, order errors are more challenging due to the complexity of capturing such anomalies within text data.

Overall, our model demonstrates strong performance across all types of prescription errors, showcasing its ability to effectively capture contextual information such as inaccuracies in names, values, and order within the prescription text.

4.4. Experiment 4 : XAI techniques for model interpretability

We further investigated explainable AI (XAI) techniques to provide both global and local feature explanations for enhancing the transparency and reliability of our model. Specifically, we leveraged two commonly used XAI methods: LIME (Local Interpretable Model-agnostic Explanations) [50] and SHAP (SHapley Additive exPlanations) [51]. LIME generates locally faithful explanations by approximating the model with an interpretable surrogate model around a specific prediction. This approach focuses on individual predictions, making it particularly valuable for analyzing the model's behavior on a case-by-case basis. In contrast, SHAP explains the output of machine learning models by computing the contribution of each feature, offering deeper insights into the overall behavior of the model. These methods provide a deeper understanding of the key factors driving the model's predictions for potential prescription errors, helping to overcome the opacity of the machine learning model. This enables clinicians to assess the reliability of the model's predictions, thereby enhancing their confidence in the decision-making process.

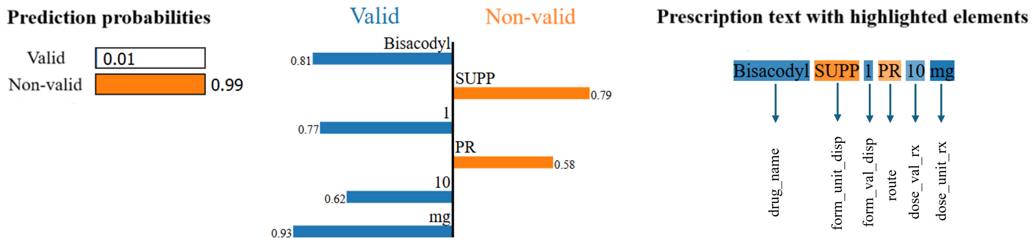

Figure 2: LIME explanation for a prescription sample featuring the medication "Bisacodyl", highlighting prediction probabilities and showcasing the contribution of individual elements in identifying a potentially invalid prescription.

In this experiment, we utilize LIME and SHAP to interpret the model employed in Experiment 3, which integrates BERT-based text embeddings with a Local Outlier Factor (LOF) one-class classification algorithm.

The figure 2 illustrates a LIME explanation of our one-classification model applied to a prescription sample (of the medication "Bisacodyl"), showcasing how our model identifies prescription errors. The Prediction Probabilities panel on the left reveals that the model assigns a high probability of 0.99 to classifying the prescription input as an anomaly, with only a 0.01 probability for a valid prescription, indicating strong confidence in its anomaly detection. The central bar chart reveals the contribution of individual prescription elements to this classification decision. Blue bars indicate elements contributing to a valid classification, while orange bars highlight words contributing to the anomaly classification. For instance, terms like "Bisacodyl" and "mg" positively influence a valid classification, whereas words such as "SUPP" and "PR" strongly indicate an Anomaly. This visual representation highlights the parts of the prescription that played a key role in the model's decision.

On the right, the text with highlighted words panel provides a direct view of the prescription text, with color-coding to indicate each word's impact:

orange highlights elements contributing to anomaly detection, while blue marks those associated with a valid prescription. This visualization allows clinicians and researchers to quickly identify irregularities in the medication prescription, clearly indicating which elements led to the model's decision.

To further delve into the model's decision-making process, Figure 3 presents the Force Plot representation for the same prescription sample featuring the medication "Bisacodyl". This visualization showcases the contributions of individual prescription elements and their overall impact on the model's prediction. The Force Plot displays the contributions of each feature (prescription element), where values pushing the prediction towards a higher anomaly score are highlighted in red, while those supporting a lower anomaly score are in blue. In this example, terms like "PO" and "10" have the highest positive SHAP values, significantly contributing towards an Anomaly classification, as seen in their red coloring. Conversely, terms such as "Bisacodyl," "2," "TAB," and "mg" contribute towards a more Valid interpretation, represented by the blue segments. The combined impact of these features shifts the base value of the model's prediction to a final value of 2.55, providing a clear understanding of how the model arrives at its decision and differentiates between valid and non-valid prescriptions.

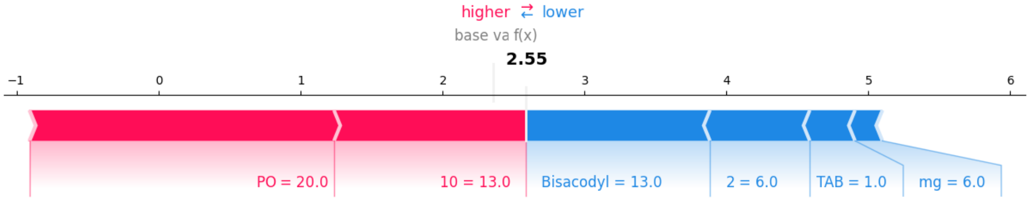

Figure 3: SHAP Force Plot for a prescription sample featuring the medication "Bisacodyl" showcasing the contributions of individual prescription elements and their overall impact on the model's prediction.

In this plot, features that push the prediction towards a higher anomaly score are highlighted in red, while those that support a lower anomaly score are shown in blue. In this example, terms like "PO" and "10" have the highest positive SHAP values, significantly contributing to a non-valid prescription, as indicated by their red coloring. On the other hand, terms such as "Bisacodyl", "2", "TAB", and "mg" contribute towards a more valid prescription, represented by the blue segments. The base value of the model's prediction is adjusted to a final value of 2.55 due to the combined influences of these features. This visualization provides deeper insights into the model's behavior and how it distinguishes between valid and non-valid prescriptions.

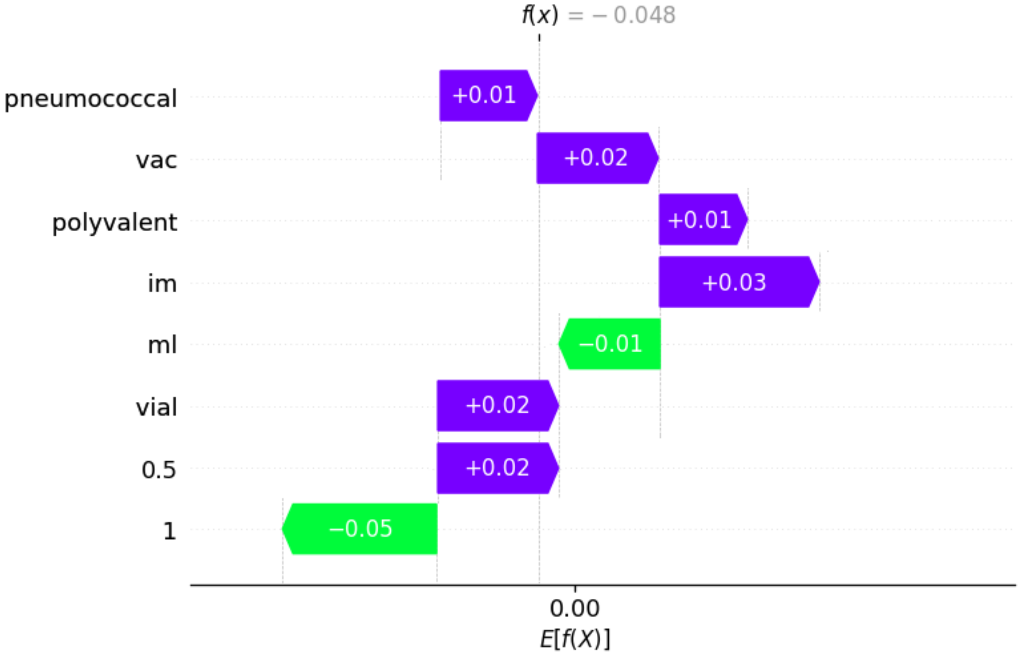

Figure 4: SHAP Summary Plot for a prescription sample featuring the medication "Bisacodyl", highlighting the impact of individual prescription elements on the model's prediction.

The SHAP summary plot presented in the figure 4, visualizes the impact of each element within a prescription sample featuring the medication "Pneumococcal vac polyvalent" on the model's final prediction. Positive SHAP values indicate elements that contribute to a valid prescription, while negative SHAP values highlight prescription elements that push the prediction towards a non-valid prescription. For instance, elements like "pneumococcal", "im" and "vial" have positive SHAP values (+0.01, +0.03 and +0.02), indicating their contribution to the prescription being classified as valid. These terms are highlighted in purple, denoting their positive impact. In contrast, words like "ml" and "1" show negative contributions (-0.01 and -0.05), pushing the classification towards a non-valid prescription, likely due to irregular usage or context in this prescription. The total SHAP contributions result in a final prediction score of f(x) = -0.048, reflecting the model's decision-making process for this instance. This SHAP graph provides an intuitive understanding of how individual words affect the overall classification, helping practitioners comprehend why a prescription is deemed valid or potentially invalid.

5. Discussion

In this study, we conducted a series of experiments to evaluate the performance of a one-class classification models for detecting valid and non-valid prescriptions. The first experiment focused on assessing the effectiveness of combining text embeddings, specifically Word2Vec and BERT, with various one-class classification algorithms, including OC-SVM, Local Outlier Factor (LOF), and Isolation Forest. We utilized prescription data from the MIMIC database as valid examples while synthetically generating non-valid prescriptions. The model was trained on six distinct features: drug name, dose value, dose unit, form value, form unit, and route. The results showed that BERT embeddings significantly outperformed Word2Vec in most performance metrics, such as precision, recall, and F1-score, indicating that BERT's contextualized embeddings were more effective at capturing complex relationships within prescription data. Among the one-class classification algorithms, Local Outlier Factor emerged as the most effective in identifying outliers and distinguishing valid from non-valid prescriptions. LOF's ability to detect local anomalies and its robustness against noise made it highly suitable for identifying prescription errors, especially in capturing subtle inconsistencies.

The second experiment investigated the impact of dimensionality reduction techniques, specifically Auto-Encoder and Principal Component Analysis, on the performance of BERT-based embeddings in conjunction with one-class classification algorithms. As BERT generates high dimensional space embeddings of prescription text, using dimensionality reduction helps to address computational efficiency, improve model performance, and enhance interpretability by minimizing computational costs, removing redundancy, reducing the risk of overfitting, and facilitating the visualization by projecting embeddings into 2D or 3D space. Both dimensionality reduction (AE and PCA) techniques improved the performance of the model compared to the baseline, with PCA showing slightly better results overall. The highest performance was achieved by PCA-LOF, which demonstrated the best precision, recall, and F1-score. However, AE performed better with the Isolation Forest algorithm, particularly in terms of recall and F1-score. This experiment highlights the importance of dimensionality reduction in improving the efficiency and performance of prescription error detection models, with PCA being a more straightforward method for this context.

Experiment 3 focused on assessing the model's performance across various types of prescription errors to identify which errors were most discriminative in distinguishing valid from non-valid prescriptions. The results revealed that the model was most effective in detecting errors related to incorrect medication information, such as name, route of administration, dosage form, and dosage unit. Overdoses and underdoses were also accurately detected but presented slightly more complexity. However, order errors, while still detectable, proved more challenging due to their complexity and nuanced nature within the prescription data. Overall, our findings demonstrate that the model effectively captures contextual information related to inaccuracies within prescription texts, showcasing its potential for improving medication safety by accurately identifying prescription errors.

Unlike traditional rule-based systems, which are inherently interpretable, the adoption of machine learning models raises concerns about transparency, as these models often operate as "black boxes," making their decision-making processes difficult to decipher. To address the opacity of these models, the final experiment focuses on improving the interpretability of our framework by applying two XAI techniques, namely LIME and shap. LIME provides local, case-specific explanations, highlighting how individual prescription elements influence the model's decision, while SHAP offers both local and global insights into feature contributions across the dataset. The visualizations, such as LIME's bar charts and SHAP's Force and Summary plots, reveal how specific elements in prescriptions (e.g., medication names, dosage) contribute to classifying a prescription as valid or invalid. This level of transparency is critical in clinical settings, ensuring that practitioners understand not only when a prescription is flagged as non-valid but also why the model made that decision. Such interpretability is crucial for building trust in AI-driven solutions, especially in sensitive areas like healthcare.

6. Conclusion and Futur Works

In this study, we developed and evaluated an interpretable one-class classification model for detecting anomalies in prescription data, with a focus on identifying potential medication errors. Our approach integrated advanced natural language processing techniques, dimensionality reduction methods, and one-class classification algorithms to effectively distinguish valid from non-valid prescriptions.

Through extensive experiments, we introduced a novel framework that integrates BERT embeddings for prescription text representation with PCA for dimensionality reduction and Local Outlier Factor for one-class classification. The use of BERT embeddings proved superior to Word2Vec, effectively capturing complex relationships within prescription data due to its contextual embeddings and positional encoding. Additionally, incorporating dimensionality reduction techniques like PCA and AE further enhanced model performance by removing redundancy and improving computational efficiency and interpretability. PCA emerged as a slightly more effective method overall, particularly when combined with LOF, due to its straightforward application and ability to preserve essential data variance. The analysis of various prescription error types highlighted the model's strength in detecting inaccuracies related to medication details, although order errors posed more challenges. Finally, the application of LIME and SHAP for model interpretability provided valuable insights into the decision-making process, fostering transparency and trust in the AI-driven system.

These findings underscore the potential of machine learning-based solutions for improving medication safety and reducing the risk of adverse drug events. Future research could explore the integration of additional features, such as patient demographics and clinical context, to further refine the model's accuracy and applicability in real-world clinical settings. Additionally, we envisage investigating self-supervised learning to tackle the challenge of having only valid prescription data available. By designing suitable pretext tasks, the model can learn meaningful and robust feature representations from the valid prescriptions without requiring explicit labeled examples of prescription errors. These learned representations can then be leveraged to identify errors as deviations from the learned distribution of "valid" prescriptions.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT-3.5 in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] D. Sharma, P. Patel, M. Shah, A comprehensive study on industry 4.0 in the pharmaceutical industry for sustainable development, Environmental Science and Pollution Research (2023) 1–11.

[2] A. Hodkinson, N. Tyler, D. M. Ashcroft, R. N. Keers, K. Khan, D. Phipps, A. Abuzour, P. Bower, A. Avery, S. Campbell, et al., Preventable medication harm across health care settings: a systematic review and meta-analysis, BMC medicine 18 (2020) 1–13.

[3] M. Anzan, M. Alwhaibi, M. Almetwazi, T. M. Alhawassi, Prescribing errors and associated factors in discharge prescriptions in the emergency department: A prospective cross-sectional study, PLoS One 16 (2021) e0245321.

[4] S. Al-Khani, A. Moharram, H. Aljadhey, Factors contributing to the identification and prevention of incorrect drug prescribing errors in outpatient setting, Saudi Pharmaceutical Journal 22 (2014) 429–432.

[5] J. Corny, A. Rajkumar, O. Martin, X. Dode, J.-P. Lajonchère, O. Billuart, Y. Bézie, A. Buronfosse, A machine learning–based clinical decision support system to identify prescriptions with a high risk of medication error, Journal of the American Medical Informatics Association 27 (2020) 1688–1694.

[6] A. Desnoyer, B. Guignard, P.-O. Lang, J. Desmeules, N. Vogt-Ferrier, P. Bonnabry, Prescriptions médicamenteuses potentiellement inappropriées en gériatrie: quels outils utiliser pour les détecter?, La Presse Médicale 45 (2016) 957–970.

[7] D. O'Mahony, D. O'Sullivan, S. Byrne, M. N. O'Connor, C. Ryan, P. Gallagher, Stopp/start criteria for potentially inappropriate prescribing in older people: version 2, Age and ageing 44 (2014) 213–218.

[8] K. C. Nanji, J. M. Rothschild, C. Salzberg, C. A. Keohane, K. Zigmont, J. Devita, T. K. Gandhi, A. K. Dalal, D. W. Bates, E. G. Poon, Errors

associated with outpatient computerized prescribing systems, Journal of the American Medical Informatics Association 18 (2011) 767–773.

[9] R. Shrestha, S. Prajapati, Assessment of prescription pattern and prescription error in outpatient department at tertiary care district hospital, central nepal, Journal of pharmaceutical policy and practice 12 (2019) 1–9.

[10] R. Kaushal, D. W. Bates, C. Landrigan, K. J. McKenna, M. D. Clapp, F. Federico, D. A. Goldmann, Medication errors and adverse drug events in pediatric inpatients, Jama 285 (2001) 2114–2120.

[11] K. Eslami, F. Aletayeb, S. M. H. Aletayeb, L. Kouti, A. K. Hardani, Identifying medication errors in neonatal intensive care units: a two-center study, BMC pediatrics 19 (2019) 1–7.

[12] H. Khalili, S. Farsaei, H. Rezaee, S. Dashti-Khavidaki, Role of clinical pharmacists' interventions in detection and prevention of medication errors in a medical ward, International journal of clinical pharmacy 33 (2011) 281–284.

[13] A. Barbier, C. Rousselière, L. Robert, E. Cousein, B. Décaudin, Development of a methodological guide on the implementation of a pharmaceutical decision support system: feedback from a french university hospital, in: Annales Pharmaceutiques Francaises, volume 81, 2022, pp. 163–172.

[14] H. Palisson, R. Darwich, A. Olivo, M.-C. Chaumais, É. Paoli, É. Adnet, S. Drouot, Détection des patients à risque d'événement indésirable médicamenteux à l'aide du logiciel pharmaclass®, Journal de Pharmacie Clinique 41 (2022) 110–112.

[15] L. Robert, E. Cuvelier, C. Rousselière, S. Gautier, P. Odou, J.-B. Beuscart, B. Décaudin, Detection of drug-related problems through a clinical decision support system used by a clinical pharmacy team, in: Healthcare, volume 11, MDPI, 2023, p. 827.

[16] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, K. I. Kroeker, An overview of clinical decision support systems: benefits, risks, and strategies for success, NPJ digital medicine 3 (2020) 17.

[17] J. Graafsma, R. M. Murphy, E. M. van de Garde, F. Karapinar-Çarkit, H. J. Derijks, R. H. Hoge, J. E. Klopotowska, P. M. van den Bemt, The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: a scoping review, Journal of the American Medical Informatics Association 31 (2024) 1411–1422.

[18] R. Rozenblum, R. Rodriguez-Monguio, L. A. Volk, K. J. Forsythe, S. Myers, M. McGurrin, D. H. Williams, D. W. Bates, G. Schiff, E. Seoane-Vazquez, Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation, The Joint Commission Journal on Quality and Patient Safety 46 (2020) 3–10.

[19] L. Robert, C. Rousseliere, J.-B. Beuscart, S. Gautier, E. Chazard, B. Décaudin, P. Odou, Integration of explicit criteria in a clinical decision support system through evaluation of acute kidney injury events. (2021).

[20] G. Segal, A. Segev, A. Brom, Y. Lifshitz, Y. Wasserstrum, E. Zimlichman, Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting, Journal of the American Medical Informatics Association 26 (2019) 1560–1565.

[21] H. Van Der Sijs, J. Aarts, T. Van Gelder, M. Berg, A. Vulto, Turning off frequently overridden drug alerts: limited opportunities for doing it safely, Journal of the American Medical Informatics Association 15 (2008) 439–448.

[22] S. Ben Othman, B. Décaudin, P. Odou, C. Rousselière, E. Cousein, S. Hammadi, Pharmaceutical decision support system using machine learning to analyze and limit drug-related problems in hospitals, in: MEDINFO 2023—The Future Is Accessible, IOS Press, 2024, pp. 1593–1597.

[23] G. D. Schiff, L. A. Volk, M. Volodarskaya, D. H. Williams, L. Walsh, S. G. Myers, D. W. Bates, R. Rozenblum, Screening for medication errors using an outlier detection system, Journal of the American Medical Informatics Association 24 (2017) 281–287.

[24] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, L. Cilar, Interpretability of machine learning-based prediction models in healthcare, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10 (2020) e1379.

[25] S. Maleki Varnosfaderani, M. Forouzanfar, The role of ai in hospitals and clinics: transforming healthcare in the 21st century, Bioengineering 11 (2024) 337.

[26] E. Cuvelier, L. Robert, E. Musy, C. Rousseliere, R. Marcilly, S. Gautier, P. Odou, J.-B. Beuscart, B. Décaudin, The clinical pharmacist's role in enhancing the relevance of a clinical decision support system, International journal of medical informatics 155 (2021) 104568.

[27] N. Seliya, A. Abdollah Zadeh, T. M. Khoshgoftaar, A literature review on one-class classification and its potential applications in big data, Journal of Big Data 8 (2021) 1–31.

[28] T. Hayashi, D. Cimr, H. Fujita, R. Cimler, Critical review for one-class classification: recent advances and the reality behind them, arXiv preprint arXiv:2404.17931 (2024).

[29] D. M. Tax, R. P. Duin, Support vector domain description, Pattern recognition letters 20 (1999) 1191–1199.

[30] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, R. C. Williamson, Estimating the support of a high-dimensional distribution, Neural computation 13 (2001) 1443–1471.

[31] D. M. Tax, R. P. Duin, Support vector data description, Machine learning 54 (2004) 45–66.

[32] M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander, Lof: identifying density-based local outliers, in: Proceedings of the 2000 ACM SIGMOD international conference on Management of data, 2000, pp. 93–104.

[33] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation-based anomaly detection, ACM Transactions on Knowledge Discovery from Data (TKDD) 6 (2012) 1–39.

[34] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, M. Kloft, Deep one-class classification, in: International conference on machine learning, PMLR, 2018, pp. 4393–4402.

[35] M. Sakurada, T. Yairi, Anomaly detection using autoencoders with nonlinear dimensionality reduction, in: Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis, 2014, pp. 4–11.

[36] C. Zhou, R. C. Paffenroth, Anomaly detection with robust deep autoencoders, in: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017, pp. 665–674.

[37] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, H. Chen, Deep autoencoding gaussian mixture model for unsupervised anomaly detection, in: International conference on learning representations, 2018.

[38] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, G. Langs, Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in: International conference on information processing in medical imaging, Springer, 2017, pp. 146–157.

[39] D. T. Nguyen, Z. Lou, M. Klar, T. Brox, Anomaly detection with multiple-hypotheses predictions, in: International Conference on Machine Learning, PMLR, 2019, pp. 4800–4809.

[40] B. Tian, Q. Su, J. Yin, Anomaly detection by leveraging incomplete anomalous knowledge with anomaly-aware bidirectional gans, arXiv preprint arXiv:2204.13335 (2022).

[41] S. Goyal, A. Raghunathan, M. Jain, H. V. Simhadri, P. Jain, Drocc: Deep robust one-class classification, in: International conference on machine learning, PMLR, 2020, pp. 3711–3721.

[42] S. Dhar, B. Gonzalez-Torres, Doc 3: deep one class classification using contradictions, Machine Learning 113 (2024) 5109–5150.

[43] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R. G. Mark,

Mimic-iii, a freely accessible critical care database, Scientific data 3 (2016) 1–9.

[44] C. Hadiwinoto, H. T. Ng, W. C. Gan, Improved word sense disambiguation using pre-trained contextualized word representations, arXiv preprint arXiv:1910.00194 (2019).

[45] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).

[46] M. T. R. Laskar, C. Chen, J. Johnston, X.-Y. Fu, S. Bhushan TN, S. Corston-Oliver, An auto encoder-based dimensionality reduction technique for efficient entity linking in business phone conversations, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 3363–3367.

[47] M. Ramamurthy, Y. H. Robinson, S. Vimal, A. Suresh, Auto encoder based dimensionality reduction and classification using convolutional neural networks for hyperspectral images, Microprocessors and Microsystems 79 (2020) 103280.

[48] K. N. Singh, S. D. Devi, H. M. Devi, A. K. Mahanta, A novel approach for dimension reduction using word embedding: An enhanced text classification approach, International Journal of Information Management Data Insights 2 (2022) 100061.

[49] B. M. S. Hasan, A. M. Abdulazeez, A review of principal component analysis algorithm for dimensionality reduction, Journal of Soft Computing and Data Mining 2 (2021) 20–30.

[50] M. T. Ribeiro, S. Singh, C. Guestrin, "why should I trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, 2016, pp. 1135–1144.

[51] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances

in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.