






Article

Text Classification in Clinical Practice Guidelines Using Machine-Learning Assisted Pattern-Based Approach

Musarrat Hussain ¹, Jamil Hussain ², Taqdir Ali ¹, Syed Imran Ali ¹, Hafiz Syed Muhammad Bilal ¹,
Sungyoung Lee ^{1,*} and Taechoong Chung ^{1,*}

¹ Department of Computer Science and Engineering, Kyung Hee University, Global Campus, 1732, Deogyong-daero, Giheung-gu, Yongin-si, Gyeonggi-do 17104, Korea; musarrat.hussain@oslab.khu.ac.kr (M.H.); taqdir.ali@oslab.khu.ac.kr (T.A.); imran.ali@oslab.khu.ac.kr (S.I.A.); bilalrizvi@oslab.khu.ac.kr (H.S.M.B.)

² Department of Data Science, Sejong University, Sejong 30019, Korea; jamil@sejong.ac.kr

* Correspondence: sylee@oslab.khu.ac.kr (S.L.); tcchung@khu.ac.kr (T.C.)

Abstract: Clinical Practice Guidelines (CPGs) aim to optimize patient care by assisting physicians during the decision-making process. However, guideline adherence is highly affected by its unstructured format and aggregation of background information with disease-specific information. The objective of our study is to extract disease-specific information from CPG for enhancing its adherence ratio. In this research, we propose a semi-automatic mechanism for extracting disease-specific information from CPGs using pattern-matching techniques. We apply supervised and unsupervised machine-learning algorithms on CPG to extract a list of salient terms contributing to distinguishing recommendation sentences (RS) from non-recommendation sentences (NRS). Simultaneously, a group of experts also analyzes the same CPG and extract the initial patterns “Heuristic Patterns” using a group decision-making method, nominal group technique (NGT). We provide the list of salient terms to the experts and ask them to refine their extracted patterns. The experts refine patterns considering the provided salient terms. The extracted heuristic patterns depend on specific terms and suffer from the specialization problem due to synonymy and polysemy. Therefore, we generalize the heuristic patterns to part-of-speech (POS) patterns and unified medical language system (UMLS) patterns, which make the proposed method generalize for all types of CPGs. We evaluated the initial extracted patterns on asthma, rhinosinusitis, and hypertension guidelines with the accuracy of 76.92%, 84.63%, and 89.16%, respectively. The accuracy increased to 78.89%, 85.32%, and 92.07% with refined machine-learning assistive patterns, respectively. Our system assists physicians by locating disease-specific information in the CPGs, which enhances the physicians’ performance and reduces CPG processing time. Additionally, it is beneficial in CPGs content annotation.

Keywords: recommendation statements identification; guideline processing; pattern extraction; information extraction; clinical text mining



Citation: Hussain, M.; Hussain, J.; Ali, T.; Ali, S.I.; Bilal, H.S.M.; Lee, S.; Chung T. Text Classification in Clinical Practice Guidelines Using Machine-Learning Assisted Pattern-Based Approach. *Appl. Sci.* **2021**, *11*, 3296. <https://doi.org/10.3390/app11083296>

Academic Editor: Elisa Quintarelli

Received: 2 March 2021

Accepted: 2 April 2021

Published: 7 April 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Technological advancements have generated a great boom for the healthcare industry, by enhancing its reach to a wider population pool and augmenting the clinical practices with state-of-the-art research. Clinical Practice Guidelines (CPGs) represent a formalization of the medical intricacies, which would otherwise, greatly hinder the delivery of high quality, healthcare services [1]. CPGs play a pivotal role in standardization and dissemination of medical knowledge, prevention of ad-hoc non-standard practice variations, and providing evidence-based treatments [2,3]. Typically, the contents of a CPG, describe disease-specific process flows, patients’ summaries, medical decisions, content specific alerts, and protocols, which provide the necessary ingredients for dealing with a wide variety of medical situations [4,5]. However, the adherence rate of CPGs, is highly dependent on their nature,

and the applicable clinical scenario, which leads to an effective usage rate between 20% and 100% [6]. Some of the common reasons for non-adherence to these guidelines, include, a lack of awareness for the healthcare practitioners, and the difficulty in understanding the large textual content of the CPGs in a limited time, during the clinical practice [6–8].

One of the possible solutions to this problem is to transform CPGs into a machine-interpretable format and to integrate the knowledge extracted from these, with clinical information systems and Clinical Decision Support Systems (CDSS). This knowledge integration leads to the creation of Guideline-based CDSS, which can provide disease-specific recommendations for optimizing and customizing the patient care. Additionally, machine-interpretable CPGs allow the physicians to save their valuable time, by providing a meaningful abstraction of the contents and disease-specific information, thereby enhancing healthcare delivery.

Based on the importance of the provided information, CPG contents can be categorized into two parts. First, the background information, which includes abstract information related to the background and point of view of the authors. Second, the disease-specific information, which elaborates causes, consequences, and actions related to a disease. For instance, the sentence, “Hypertension remains one of the most important preventable contributors to disease and death.” represents background information, while “In the black hypertensive population, including those with diabetes, a calcium channel blocker or thiazide-type diuretic is recommended as initial therapy.” represents disease-specific information, also known as a recommendation sentence. Therefore, the understandability and classification of CPG contents is an important step, before its transformation to computer interpretable format. Among this information, the recommendation sentences are the main focuses and desired contents that need to be extracted from CPG. These contents assist the domain experts in making evidence-based decisions.

The field of text classification and information extraction has greatly benefited from advances in computing, producing a plethora of algorithms, tools, and applications, based on machine-learning and pattern-based approaches. [9–14]. However, in the clinical domain, most of the natural language processing tasks including, guideline processing and information extraction, are still using pattern-based approaches [15]. Pattern-based approaches perform better than machine-learning models in clinical text classification [16]. The patterns are generally extracted by human experts based on their heuristics [17]. An expert focuses on the sequence of terms used in content for patterns, therefore the terms used in patterns suffer from the problems of polysemy and synonymy [18]. To overcome this problem, we proposed a machine-learning assistive pattern-based approach, which consists of heuristic patterns, part-of-speech (POS) patterns, and unified medical language system (UMLS) patterns for CPGs sentence classification to recommendation sentences (RS) and non-recommendation sentences (NRS). A group of experts extracted the initial heuristic patterns from an annotated guideline based on their heuristics using a group-based decision-making method known as nominal group technique (NGT) [19]. The NGT is selected due to its effectiveness in a group decision-making process. Simultaneously, we apply supervised machine-learning algorithms such as decision tree and rule induction, and unsupervised algorithm including Latent Dirichlet Allocation (LDA) and word2vec [20,21]. We selected the aforementioned algorithms because of their effectiveness and decision transparency. These algorithms provide a list of words, which are mainly contributing to a classification decision. We evaluated and analyzed all contributing words considered by machine-learning and finalized a list of salient terms. We provided the salient terms list to the participating experts to review their extracted patterns by considering those salient terms as well. The experts revised the patterns which increased the sentence classification accuracy. The proposed approach has two-fold benefits. It presents disease-specific information to physicians, which helps in providing standardized clinical services. It can also be used in annotating CPG sentences for computable CPGs generation.

The rest of the article is structured as follows. Section 2 describes related work. Section 3 provides the detail of the proposed solution. Section 4 describes results with discus-

sion, Section 5 evaluates the proposed methodology, and finally, Section 6 concludes the study.

2. Related Work

CPGs have acted as a formal resource for identification and adaptation of best practices in the medical domain, since the late 1970s [22]. However, the transformation of theory, in the form of textual CPGs, into common clinical practice, is only possible, when it is transformed into a machine-interpretable model. Many approaches have been proposed to achieve this transformation, especially focusing on representing and executing knowledge from CPGs over patient-specific clinical data. Some of these approaches include document-centric models, probabilistic models, decision trees, and task network models, which can be used to represent guidelines in an appropriate format [23]. Still, the current solutions for automatic conversion of CPGs into computer interpretable format, have many limitations. The primary hurdle here, is the difficulty in accurately recognizing and extracting, recommendation sentences from the textual content, since most of the recommendation sentences are not written in the *IF < condition > THEN < action >* format. Therefore, a technique is required that can extract recommendation sentences from CPG text. Some of the approaches relevant to the proposed approach are described as follows.

R. Servan et al. [24] developed a methodology for the formalization of CPGs using linguistic patterns and predefined templates. The linguistic templates were generated from the domain ontology. Their proposed methodology performed activities including, pattern extraction, selection of core patterns from extracted patterns for the generation of an executable model, and finally the model evaluation. This methodology produced reusable guidelines blocks/templates for authoring and formalizing CPGs. However, this approach needs a customized domain ontology for mapping the concepts, while generating the template.

R. Wenzina et al. [4] proposed a rule-based method using a combination of linguistic and semantic information of UMLS. The authors hypothesized that each CPG statement has an associated, domain-dependent linguistic and semantic pattern. They proposed a weighting coefficient (relevance rate) to extract the relevant statements by identifying the condition-action combination of terms. In this manner, their proposed model can determine the relevancy of each statement towards the clinical workflow. The authors used one guideline for identifying and extracting 12 “if” and 4 “should” statements. They found that rules of type “if” have a better chance of detection than the type “should”.

H. Hematialam et al. [25] designed supervised learning models using ZeroR, Naive Bayes, J48, and Random Forest for the classification of CPG statements. Using three annotated guidelines (Hypertension, Chapter 4 of asthma, and rhinosinusitis), these models were trained to classify each CPG statement into no condition (NC), condition-action (CA), or condition consequence (CC). Additionally, Part-of-Speech (POS) tagging was used to remove domain dependency constraints and recommendation statements were identified by using modifiers and regular expressions. The most commonly used modifiers were “if”, “in”, “to”, “for”, “when”, and “which”. The identified recommendation statements were then transformed into “if condition then consequences” format for rule generation in later stages. The authors’ used models were one shot models, which require retraining each time when a change occurs in the training dataset.

W. Gad El-Rab et al. [26] proposed a framework for active dissemination of and automatic knowledge extraction from, CPGs. Their proposed framework automates some of the activities to reduce manual efforts. The framework follows a multi-step approach and uses an unstructured information management architecture (UIMA) for identifying medical concepts. The authors achieved this by performing several information processing activities such as XML parsing, text cleansing, medical concept tagging, medical tags disambiguation, clinical context pattern detection, clinical context filtering, and clinical context mapping.

S. Priyanta et al. [27] performed a comparative analysis of sentence subject classification using rule-based and machine-learning models. The authors used opinion patterns for rule generation. They performed sentence subjectivity evaluation on Indonesian news to classify a news sentence into a subject or an objective. This classification was achieved using two machine-learning models, including a Naive-based classifier (NBC) and a multinomial Support Vector Machine (SVM). The evaluation and analysis of the results proved that the rule-based classifier has better performance with 80.36% accuracy, as opposed to SVM with accuracy 74.0% and NBC with accuracy 71%. This difference in accuracy between the rule-based approach and machine-learning algorithms has prevented the usage of the latter, in real solutions [15].

Similar to clinical domain, other domains also extensively use pattern-based approaches for NLP tasks such as opinion mining in various languages such as Persian [28] and Chinese [29]. Dashtipour et al. [28] proposed a hybrid framework of dependency grammar rules and deep neural networks for Persian opinion mining. The framework identifies polarity of the text by applying linguistic rules. However, if there is no rule to trigger for an unseen instance, the framework switches to the neural network model for the classification. Similarly, Qiang et al. [29] uses Chinese language grammar rules as constraints for Bi-LSTM model which improved Chinese sentiment classification compared to other deep-learning models such as RNN and LSTM.

The aforementioned research initiatives, either use patterns, machine-learning models, POS tags, or UMLS mapping for recommendation statement identification. Each approach has its pros and cons. For example, the existing pattern-based identification methods use single patterns (heuristic, POS, or UMLS) which depend on some pre-specified or extracted patterns and face difficulty in producing a generalized solution. In order to mitigate these limitations and to get generalized patterns, we need a mixed-method approach, which combines multiple techniques. Therefore, we proposed a machine-learning-assisted pattern-based approach by combining heuristic patterns, POS patterns, and UMLS patterns. The mixed-method approach increases the chance of accurate detection of recommendation sentences by providing a complete and synergistic use of various individual patterns.

3. Methodology

This research mainly focuses on the accurate extraction of recommendation sentences from CPGs, irrespective of the CPG target disease and format. The process flow of the proposed sentence extraction mechanism is depicted in Figure 1. Our proposed methodology consists of four major steps: document preprocessing, salient terms extraction, the pattern extraction process, and sentence classification. In the Document preprocessing step, we prepare the contents of the CPG according to the required format (sentences in our case). Salient extraction then identifies and extracts sentence decision terms using machine-learning interpretable models. This is followed by Pattern extraction, which provides the steps required for extracting the Heuristic Patterns and their generalizations (POS patterns and UMLS patterns). Finally, Sentence classification applies the extracted patterns and analyzes the CPG sentence characteristics to distinguish between the recommendation and non-recommendation sentences.

3.1. Document Preprocessing

In information processing, Preprocessing is a very important step, which is used to transform raw input data into its cleaner counterpart. This transformation generally influences the data-driven decision modeling pipeline, and it takes 50% to 80% of the computational time [30,31]. The overall objective of preprocessing is to transform input data into a form that is compilable with automated knowledge mining techniques. In this study, the goal of *Document Preprocessing* is to split the CPG documents into sentences. This goal is achieved by three sub-steps. First, the *Document Reader* loads the CPG textual document to computer memory for processing. Second, format alignment is performed by removing all empty lines and replacing multiple spaces with a single space. Finally,

the document is split into sentences by the *Sentence Extractor* using Natural Language Toolkit (NLTK) sentence tokenizer. The extracted sentences are fed to the *Pattern Extraction Process* and *Sentence Classification* components for patterns extraction and to identify the recommendation sentences.

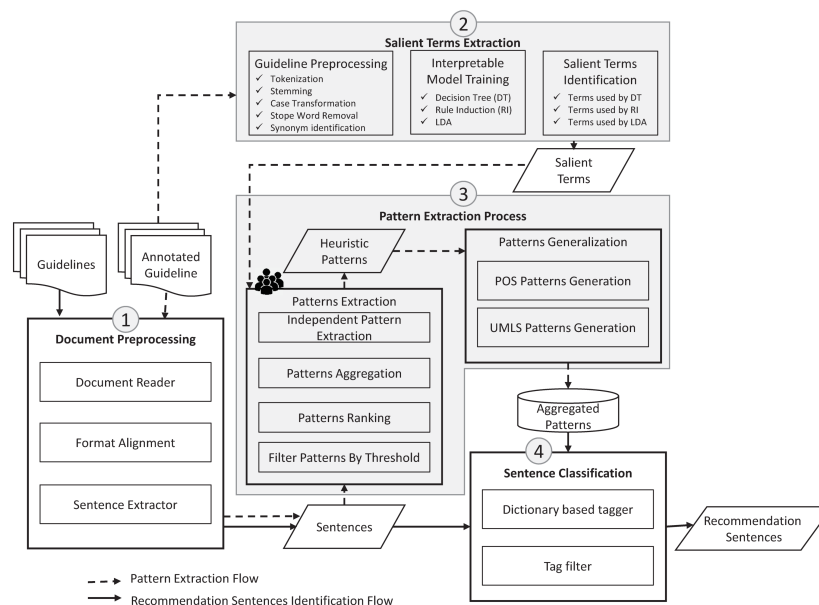


Figure 1. Process flow of the proposed recommendation identification technique.

3.2. Salient Terms Extraction

The objective of this component is to identify the key terms in CPG contents, using both supervised and unsupervised machine-learning techniques. This objective is achieved in three steps; guideline preprocessing, interpretable model training, and salient terms identification. The *guideline preprocessing* transform CPGs to machine-processable format by tokenization, stemming, case transformation, stop word removal, and synonym identification. We trained a set of supervised machine-learning models comprising of decision tree and rule induction, and unsupervised algorithms LDA, and word2vec to find the key contributing terms in a CPG for taking sentence classification decision. These techniques were selected due to their results transparency and effectiveness in the classification task. We applied various parameter settings for each model to check its classification accuracy and extract the final terms, which are then used for making the classification decision. As an example in the decision tree model, we apply gain ratio, information gain, accuracy, and Gini index splitting criteria. We also evaluate the models' behaviors with and without feature selection. In feature selection, filter-based and wrapper-based techniques were applied to limit the number of features and nodes of the final model by eliminating irrelevant features. However, identifying the correct number of features is still an open research issue, in this study, we used the grid search technique [32] to dynamically set the number of features for a model. The algorithm used for dynamic features selection is given in Algorithm 1. We check the terms considered by the model generated after feature selection to get a valuable list of salient terms considered by the model.

The example of the decision tree model is shown in Figure 2. The decision tree model have considered total 8 unique salient terms including “cosmopolitan”, “reach”, “aged”, “adult”, “channel”, “condition”, “person”, and “bespeak” for distinguishing recommendation sentences from non-recommendation sentences in a CPG. We considered all terms as salient terms, which are extracted by given models with all possible settings. A list of partial salient terms considered by various machine-learning models is given in Table 1. We shared a list of unique salient terms with human experts, hereafter knowledge engineers (KEs), for reconsideration of their extracted patterns which leads to changes in the KEs extracted patterns and causes increase in the final classification accuracy.

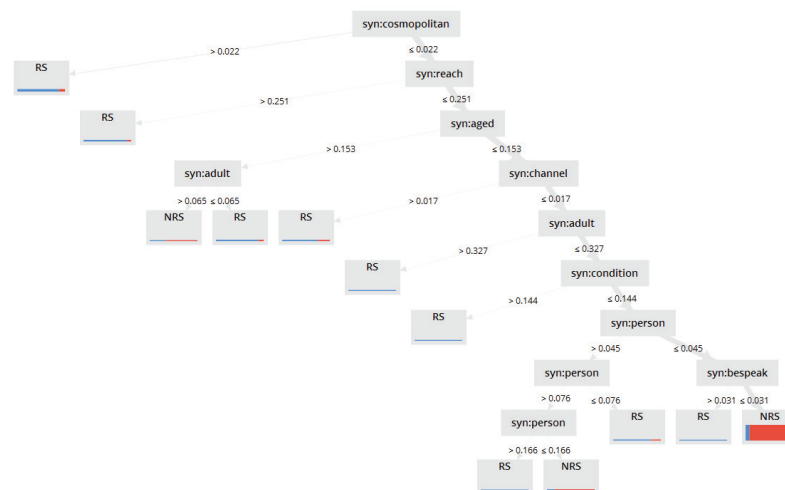
Algorithm 1: Grid search-based dynamic feature selection algorithm.

Input : Guidelines $G = \{g_1, g_2, g_3, \dots, g_n\}$, Models $M = \{m_1, m_2, m_3, \dots, m_n\}$
Result: SelectedFeatures $F' = \{f'_1, f'_2, f'_3, \dots, f'_m\} \mid m \leq n$

```

foreach  $g$  in  $G$  do
  Features  $F \leftarrow \text{getFeatur}(g)$ 
  ModelAccuries  $MA = []$ 
  foreach  $m$  in  $M$  do
    FeaturesAccuries  $FA = []$ 
    for Range  $r$  in  $[5 - \text{lenght}(F)]$  do
      FeatureSubSet  $FS = F.\text{slice}(0, r)$ 
      Model = TrainModel( $m, FS$ )
      accuray = Model.evalute( $g$ )
      FeaturesAccuries.push(accuray)
    end
    ModelAccuries.push(FeaturesAccuries)
  end
  AverageAccuryies  $AV = \text{getAverage}(MA)$ 
  MaxAccuray  $\leftarrow \max(AV)$ 
   $F' = F.\text{slice}(0, \text{MaxAccuray.index})$ 
end

```

**Figure 2.** Example decision tree model.**Table 1.** List of salient terms considered by machine-learning models.

S.No	Decision Tree	Rule Induction	LDA	Word2vec
1	cosmopolitan	cosmopolitan	goal	recommend
2	angiotensin	reach	low	facilitate
3	bespeak	black	population	improve
4	adult	better	treatment	consideration
5	aged	opinion	year	evidence
6	animation	aged	recommendation	assess
7	condition	condition	evidence	condition
8	reach	former	pharmacological	quality
9	black	case	initiate	regardless
10	decrepit	commend	hypertension	referral

3.3. Pattern Extraction Process

In pattern extraction, we applied the NGT process to identify and extract patterns. Five KEs participate in the NGT process. The KEs have more than ten years of experience in biomedical text processing, analysis, and pattern extraction. In the first phase of NGT, we provided the same annotated hypertension guideline [33] to KEs for extracting patterns based on their heuristics. Heuristic-based decisions are premised on the cognitive ability, rule of thumb, intuitive judgment, an educated guess, and common sense of a person. The following five steps were performed in the NGT process for extracting the patterns.

1. Introduce all team members and nominate a leader to cordially handle meetings. The annotated CPG is provided to each member, the leader explained the purpose and process of the study and the voting process.
2. All panel members analyze the provided CPG independently and extract the patterns based on their heuristics that can identify recommendation statements in a CPG.
3. The leader collects all patterns extracted by each member and removes the duplicate patterns. A total of 21 unique patterns were identified by all KEs as shown in Table 2.
4. The panel members discuss each pattern, and the concerned member explains the reason for selecting the corresponding pattern.
5. All five participants rank each pattern from one to five, where one is the lowest and five being the highest rank. The leader aggregate the ranks of each pattern.
6. A threshold value (total rank ≥ 15) is selected with the consensus of all team members, which is the 60% of team members agreement on a pattern.
7. Select those patterns, which have a higher accumulative rank than the threshold value (15). Based on this criterion, 10 patterns are selected as final patterns shown in Table 3.

Table 2. Evaluation matrix for nominal group technique (NGT).

S.No	Extracted Patterns	KE-1	KE-2	KE-3	KE-4	KE-5	Total Score
1	.*lead(s)? to.*	3	3	2	1	2	11
2	.*treatment (should with to).*	2	3	4	4	3	16
3	.*initiat(.*?)treatment.*	2	3	2	3	4	14
4	.*to improve.*	4	4	1	3	5	17
5	.*evidence(.*?)support.*	1	4	2	1	3	11
6	.*(patient(s)?)? with (disease).*	3	3	4	5	4	19
7	.*should (include continue).*	5	3	3	2	5	18
8	.*appli(es ed)ed (to)?.*	2	1	3	2	3	11
9	.*can be used.*	3	2	2	1	2	10
10	.*(add remove)(.*?) drug.*	4	4	3	5	5	21
11	.*(panel)(.*?)(recommend(ed)? conclude(ed)? include(d)?).*	4	2	2	1	3	12
12	.*less effective.*	1	3	4	2	3	13
13	.*treatment (does not)? need.*	2	3	2	3	3	13
14	.*regardless of.*	3	4	3	3	2	15
15	.*meet.*goal.*	2	1	2	2	3	10
16	.*(increase decrease).*dose.*	5	4	5	5	4	23
17	.*(recommend(ed)?) treatment.*	3	3	4	3	3	16
18	.*(improve(ment)? high quality).*dose.*	2	2	3	3	2	12
19	.*(Recommendation /d+ /s+;.*	5	5	5	5	5	25
20	.*expert(s)?.*opinion.*	3	3	2	3	2	13
21	.*(dis)?continu(e ed ing ation).*	4	3	3	2	4	16

In the first phase of the NGT, the KEs were unaware of the extracted salient terms while extracting these patterns so that they can extract the patterns based on their heuristics without any external bias. In the second phase of the NGT, we provided a list of salient terms to all KEs and asked them to reevaluate their extracted patterns. The aforementioned steps of NGT were performed again to reexamine the patterns with consideration of salient terms. The KEs modified the extracted patterns based on the salient terms and the final agreed-upon heuristics patterns list is given in Table 4. The patterns became more general compared to patterns without considering salient terms. Most of the selected patterns included some of the salient terms to boarder its scope. As an example the pattern “.*(recommend(ed)?) treatment.*” became “.*(recommend(ed)? | better) treatment.*” after reflecting salient term “better” in the pattern.

Table 3. Extracted heuristics patterns without salient terms.

S.No	Patterns without Salient Terms
1	.*(add remove) (.*) drug.*
2	.*(recommend(ed)?) treatment.*
3	.*to improve.*
4	.*(increase decrease) .*dose.* *
5	.*treatment (should with to).*
6	.*Recommendation /d+/s+.*
7	.*should (include continue).*
8	.*(dis)?continu(e ed ing ation).*
9	.*regardless of.*
10	.*(patient(s)?)?with (disease).*

The key advantage of this approach is its ease of use and comprehensibility for human beings without detailed domain knowledge. However, this approach highly depends on the terms and terminologies of a specific guideline. Therefore, the extracted patterns may not well-perform for all guidelines. To overcome this drawback, we generalized the extracted patterns with the incorporation of two other techniques POS and UMLS patterns for getting a generic solution.

Table 4. Extracted heuristics patterns with salient terms.

S.No	Patterns with Salient Terms
1	.*(give add remove) (.*) drug.*
2	.*([I i]n) (black general) (.*) population.*
3	.*(recommend(ed)? better) treatment.*
4	.*(increase decrease) .*dose.*
5	.* ((public)? opinion) .* treatment (should with to).*
6	.*Recommendation /d+/s+.*
7	.*should (include continue).*
8	.*(dis)?continu(e ed ing ation) reach .* goal .*
9	.* (regardless of) (having age).*
10	.*(patient(s)? adult (population group))?with (disease).*

The general purpose of POS tagger is to briefly characterize and disambiguate the grammatical category of words in a specific context. It helps to find the similarity and distinction between words. In the proposed method, the POS-based classification is used to generalize the solution for avoiding domain dependency. In this study, the application of the POS tag produced inferior results. Therefore, we used the semi-POS method, which is the combination of POS tags along with clue words. For example, in “.* VB .* drug .*” “VB” is a POS that represents a verb while “drug” is a clue word. The list of POS tags, used in the study, is described in Table 5.

Table 5. List of used POS tags.

Tag	Description	Tag	Description
CD	Cardinal number	IN	Preposition/sub-conj
MD	modal	NN	Noun, sign. or mass
JJ	Adjective	TO	‘to’
JJR	Adjective, comparative	VBG	Verb present participle
VB	Verb base form	-	-

The extracted heuristic patterns shown in Table 4 is transformed into POS patterns as shown in Table 6. We employed the Stanford CoreNLP parser [34] to parse the input

sentences to their POS categories. The input sentences were assessed by matching with the POS tags listed in Table 5. The sentences matched with one or more patterns were tagged as RS and NRS, otherwise. Finally, all NRS sentences were filtered out, and RS sentences were left for further processing. The POS-based filter reduced domain dependency and increased the accuracy of our proposed system. Here, the most significant POS tags used for the identification are “Nouns” and “Verbs”.

Table 6. List of extracted POS patterns.

S.No	Patterns
1	. * VB . * drug . *
2	. * IN . * JJ . * population . *
3	. * (VB JJR) . * treatment . *
4	. * NN . * dose . *
5	. * (JJ)? NN . * treatment (MD IN TO) . *
6	NN(/s+)?:(/s+)?CD . *
7	. * VB (include continue) . *
8	. * (VB+) . * goal . *
9	. * (regardless of) VBG age . *
10	. * (JJ NN) IN disease . *

The heuristic patterns displayed in Table 4, are also transformed into UMLS-based patterns to achieve further generalization. The UMLS-based patterns, also known as semantic patterns, cover a wide range of recommendation sentences. This process, additionally improves the accuracy of the system by identifying the semantics of words and phrases in a sentence to clarify its contextual meaning.

The UMLS is a knowledge source, which contains medical vocabularies, maintained by the US National Library of Medicine [35]. It provides an interface for retrieving biomedical concepts and semantic relations, by integrating a plethora of services, and assisting in biomedical information processing and retrieval. RS mostly contains some of the biomedical phrases, which can help to distinguish RS from NRS. Using this heuristic, first, we identify the UMLS phrases using a tool called MetaMap [36] which can identify the UMLS concepts behind medical text. Using this information, we map phrases of each sentence with its corresponding biomedical concept. We then extract UMLS patterns by analyzing the tagged sentences, identifiers, and their sequence. The example for one of the extracted patterns is shown in Figure 3. A list of UMLS patterns used in our study is shown in Table 7. The matched sentences with one or more of the UMLS patterns are finally tagged as RS, and NRS otherwise. The NRS tagged sentences are then filtered out, and RS sentences are stored for further processing.

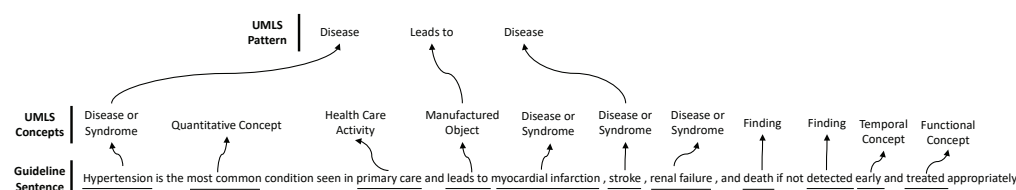


Figure 3. Example of UMLS-based pattern extraction.

3.4. Sentence Classification

The extracted patterns (Heuristic, POS, UMLS) shown in Tables 4, 6, and 7, respectively, are used to classify a CPG sentence as RS or NRS. We combine the sentences labeled as RS by heuristic patterns, POS patterns, and UMLS patterns, removing duplicates and storing the RS tagged sentences in the recommendation sentence repository.

Table 7. List of extracted UMLS Patterns.

S.No	Patterns
1	.*(give add remove) .* Pharmacologic Substance .*
2	.*([I i]n) .* Population Group .*
3	.*(Health Care Activity Qualitative Concept).* Functional Concept.*
4	.*Functional Concept .* Pharmacologic Substance.*
5	.*Qualitative Concept .*Functional Concept (should with to).*
6	.*Idea or Concept /d+/s+:.*
7	.* should .* (Functional Concept Idea or Concept).*
8	.*Idea or Concept .* Intellectual Product .*
9	.* regardless of Organism Attribute.*
10	.* Population Group .* with .* (Disease or Syndrome) .*

4. Results and Discussion

We evaluated the proposed methodology based on the system's accuracy in correctly identifying the RS sentences. We extracted patterns as well as salient terms from a published hypertension guideline annotated by a physician [33]. The guideline consists of 78 recommendation sentences out of 278 sentences. The guideline sentences were annotated as Condition-Action (CA), Condition Consequences (CC), Action (A), and Not Applicable (NA). However, we considered CA, CC, and A tagged as recommendation sentences while NA tagged sentences as NRS. For method evaluation, we used 70% of sentences for pattern extraction and 30% for the testing. Furthermore, we evaluated the extracted patterns on Rhinosinusitis [37] and chapter 4 of asthma guideline [38] to check the generalization and accuracy of the extracted patterns. The details of datasets are given in Table 8. The evaluation detail of each method is described in the following subsections.

Table 8. Details of dataset.

Guideline	Total Sentence	Recommendation Sentences	Non-Recommendation Sentences
Hypertension	278	78 (28.06%)	200 (71.94%)
Rhinosinusitis	761	151 (19.84%)	610 (80.16%)
Asthma	171	53 (30.99%)	118 (69.01%)

4.1. Results: Preprocessing

The preprocessing steps required for KEs were simple, and the only requirement was to split the CPG documents into sentences. However, the preprocessing steps required for machine-learning models are more impactful in terms of the final model accuracy and the number of salient terms. We compared the models with applying feature selection techniques and without feature selection. We used the information gain ratio to assign a weight to features and selected top k features. As mentioned earlier, the value of k highly affects the model accuracy and the salient terms considered by the model. Therefore, we tested the model on different values of k by apply Algorithm 1. The detail of k values and their effects on the accuracy of the decision tree model is shown in Figure 4. As shown in Figure 4, initially the accuracy was increasing gradually with an increment of k value. From $k = 40$ to $k = 79$ the accuracy remained stable with maximum value, while accuracy started to decrease as the value of k increased from 79. The average number accuracy of the decision tree model in maximum at $k = 40$. Therefore, we selected top 40 features for model training i.e., $k = 40$. The accuracy starts decreasing due to less relevant terms consideration as k approaches beyond 79.

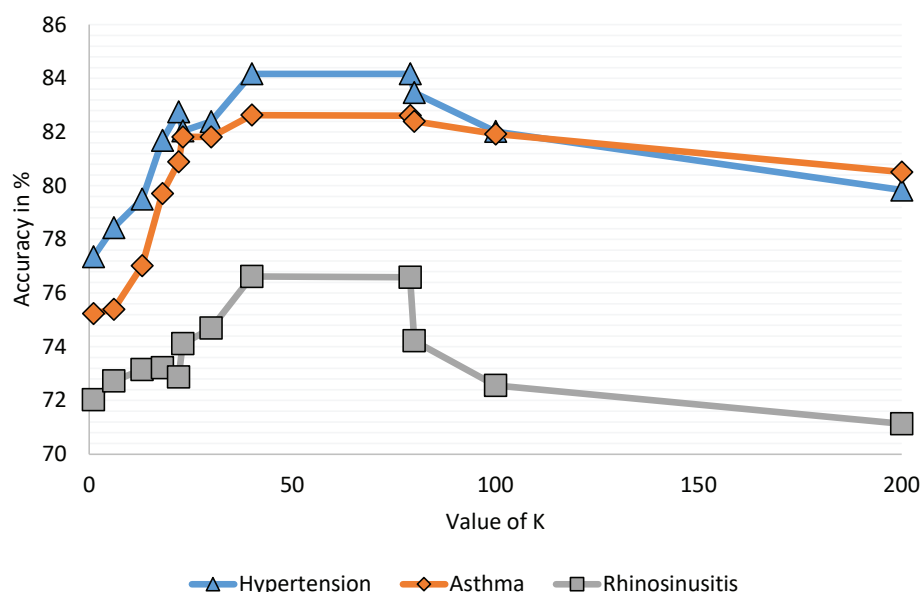


Figure 4. Top k features and the model accuracy.

4.2. Results: Salient Terms Extraction

We evaluated our trained models: decision tree, rule induction, and gradient boosted tree with and without feature selection on the hypertension [33], rhinosinusitis [37] and chapter 4 of asthma guideline [38]. The models achieved classification accuracy as given in Figure 5. Where graph (a) represents model accuracies when features selection was not performed and (b) represents accuracies with features selection. Based on the results shown in Figure 5, the accuracy of the model increases with feature selection. Also, the final generated model changes the extracted salient terms.

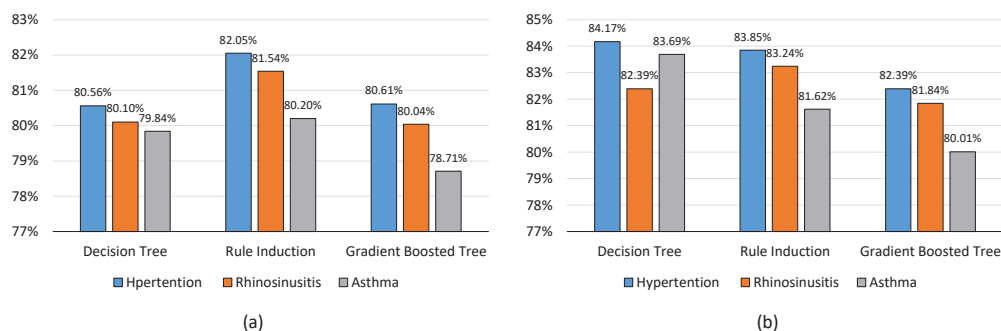


Figure 5. Model accuracy without and with features selection (a) Models accuracy without features selection (b) Models accuracy with features selection.

4.3. Results: Pattern Extraction

We have three types of patterns: heuristics patterns, POS-based patterns, and UMLS-based patterns. The CPGs sentence classification accuracy of each approach is given in the subsequent subsections.

4.3.1. Heuristic Patterns

The heuristic pattern-based method without considering the salient terms list gives 84.93% accuracy on the test dataset (30% of the hypertension guideline). The results showed that the extracted patterns work well on the test dataset. The extracted patterns, given in Table 3, were also applied on Rhinosinusitis [37] and chapter 4 of asthma [38] guidelines to evaluate the accuracy of the extracted patterns. Our proposed method achieved an accuracy of 71.93%, 75.56%, and 84.93% on asthma [38], Rhinosinusitis [37],

and Hypertension [33], guidelines, respectively, as depicted in Figure 6a. When the patterns were reevaluated by considering machine-learning extracted salient terms, KEs updated the pattern as shown in Table 4 that result increase in accuracy to 73.29%, 74.37%, and 86.04% in asthma [38], Rhinosinusitis [37], and Hypertension [33], guidelines, respectively as shown in Figure 6b.

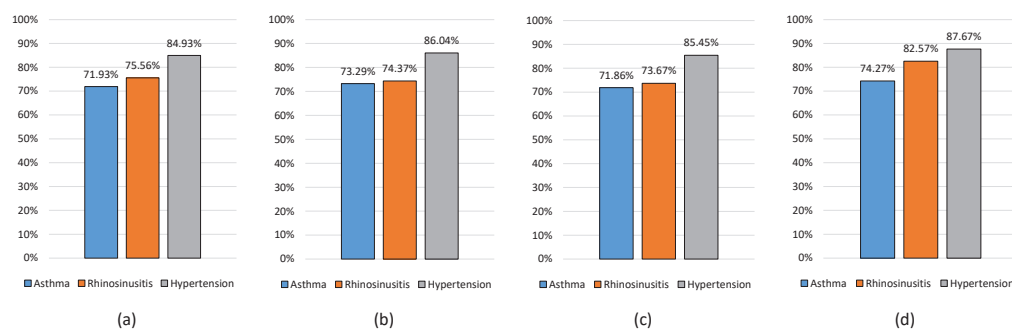


Figure 6. Extracted patterns accuracy (a) Heuristic patterns without salient terms accuracy (b) Heuristic patterns with Salient terms accuracy (c) POS based patterns accuracy (d) UMLS based pattern accuracy.

The heuristic patterns performed well on the testing part (remaining 30%) of the hypertension guideline [33]. However, the accuracy decreased by 12.75% on the other two guidelines i.e., asthma and rhinosinusitis. The primary reason for this low accuracy was the diverse format of the guidelines. One CPG uses different words and their sequence for representing the same concepts as the others. Therefore, to overcome this issue and to maintain accuracy, we added the POS-based patterns into the proposed technique.

4.3.2. POS Patterns

In the POS pattern technique, we combined the POS tags with clue words of the RS sentences. Because the combination of POS tags and the clue words increased the system accuracy. To evaluate the accuracy of the technique, all three guidelines (asthma, rhinosinusitis, and hypertension) were used in the experiment, and we achieved an accuracy of 71.86%, 73.67%, and 85.45%, respectively, as shown in Figure 6c.

The results of Figure 6c depicts that the POS-based pattern did not perform well than the heuristic patterns. However, POS patterns are applicable on all CPGs irrespective of the CPG format. We achieved better accuracy than the POS without clue words, the primary reason was the generalization of the patterns along with clue words. However, some of the clue words may not be used in different guidelines. Therefore, a complete and generic solution is required to resolve the aforementioned problem. To remove this deficiency, we merged UMLS-based patterns into the proposed technique, which increased the system accuracy. The detailed results of the UMLS pattern are described in the following subsection.

4.3.3. UMLS Patterns

The UMLS patterns, given in Table 7, classified recommendation sentences with the accuracy of 74.27%, 82.57%, and 87.67% for asthma, rhinosinusitis, and hypertension guidelines, respectively, as shown in Figure 6d. The reason for the improvement of accuracy was the UMLS concepts used in the recommendation sentences. Mostly, the recommendation sentences use tags of “Population Group”, and “Pharmacologic Substance”; therefore, UMLS-based patterns can easily recognize these sentences and increase the accuracy of the systems’ classification.

After individual evaluation, we combined all three techniques and evaluated asthma, Rhinosinusitis, and Hypertension guidelines before providing salient terms and after providing salient terms. Before using salient terms the extracted patterns achieved the accuracy of 76.92%, 84.63%, and 89.16%, respectively, as shown in Figure 7a. However, after using salient terms the pattern accuracy increased to 78.89%, 85.32%, and 92.07%,

respectively, as shown in Figure 7b. Here each sentence was evaluated by the three patterns and tagged independently. A sentence tagged by one or more techniques was finally considered to be an RS sentence otherwise NRS.

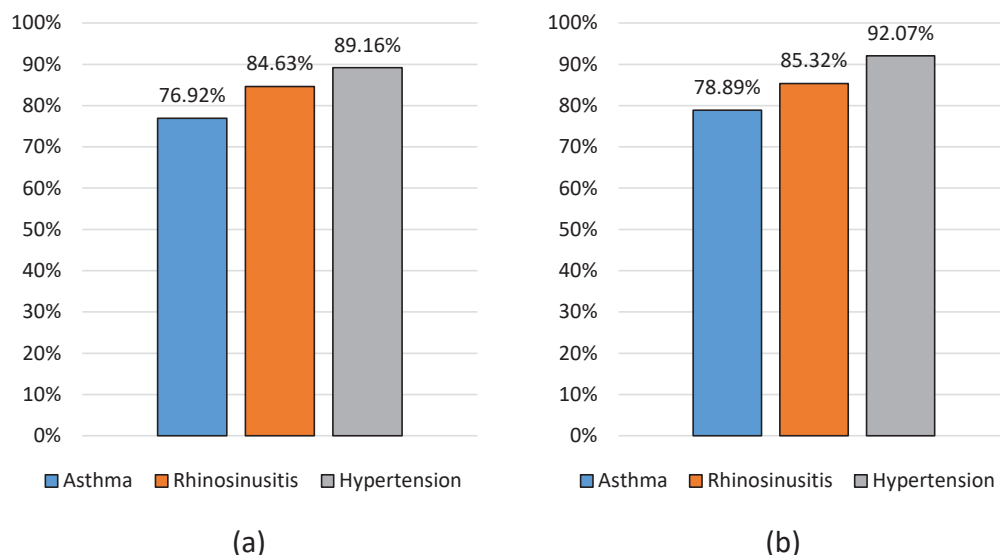


Figure 7. Combined patterns accuracy (a) without salient terms (b) with salient Terms.

As shown in Figures 5–7 the feature selection, salient terms, and combined patterns increased the classification accuracy, respectively. However, we performed a non-parametric p -value test to check the significance of the improvements [39]. The improvement shown in Figure 5 via feature selection (hereafter Model FS) compared to without feature selection (hereafter Model WFS) is evaluated with a threshold value of 0.05 under the following hypothesis.

- Null hypothesis H_0 : Model FS is not better than Model WFS
- Alternate hypothesis H_1 : FS is better than WFS

The calculated p -value for the above hypothesis is 0.035, which is less than the threshold value of 0.05. Therefore, it rejects the null hypothesis H_0 and conclude that model FS is better than WFS. Similarly, we calculated the p -value for other two cases, with and without salient terms Figure 6, and combined vs individual patterns Figure 7 with resulted value of 0.038 and 0.040, respectively. Hence the p -values showed the improvement caused by feature selection, salient terms, and combination of heuristics, POS, and UMLS patterns are statistically significant.

5. System Evaluation

The proposed technique is evaluated and compared with existing classical and advanced machine-learning models. In classical models, we targeted zeroR, Naive Bayes, J48, and Random Forest as shown in Figure 8a, while in advanced models, our focused algorithms are neural network (CNN), long short-term memory (LSTM) and Bi-directional LSTM (Bi-LSTM) as shown in Figure 8b. In classical models, ZeroR achieved 69%, Naive Bayes 69%, J48 67%, and Random Forest achieved an accuracy of 67% on asthma guideline; however, the proposed approach achieved higher accuracy of 78.89%. Similarly, the accuracies of these algorithms on Rhinosinusitis guideline were, 80%, 80%, 81%, 84%, respectively, while the proposed technique performed better with accuracy of 85.32%. Likewise, the proposed algorithm correctly classified Hypertension CPG sentence with an accuracy of 90.07%, which is higher than all classical models as depicted in Figure 8a. The improved results of the proposed methodology are mainly due to the relevant patterns execution, by combining expert heuristics with machine-learning techniques, and the generalization of the patterns through POS, and UMLS techniques.

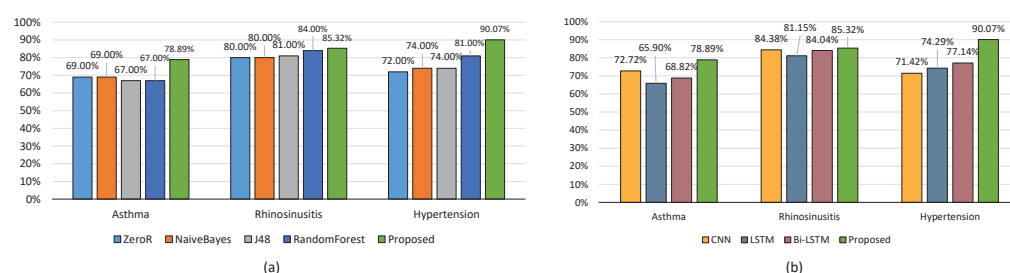


Figure 8. Evaluation of proposed method on small datasets (a) with classical models (b) with advanced models.

In advanced models, the accuracy of CNN is 72.72%, LSTM is 65.90%, Bi-LSTM is 68.82%, and the proposed system is 78.89% on asthma guideline. On Rhinosinusitis CPG, the accuracies were 84.38%, 81.15%, 84.04%, and 85.32%, respectively. However, in the Hypertension guideline, our proposed approach showed better results than the advance machine-learning models, which is 90.07% higher than 71.42%, 74.29, and 77.14% as shown in Figure 8b. The results obtained from the deep-learning models surpassed the classical models in terms of accuracy. However, the proposed technique performed better than deep-learning models. This is mainly because deep-learning models are data-hungry models and required a large training data than the provided one.

The datasets used in the study have a small number of sentences, and the distribution between recommendation and non-recommendation sentences is also very biased towards non-recommendation. Therefore, data-hungry models such as deep-learning models did not perform well as shown in Figure 8b. To overcome this deficiency, we checked the applications of these advanced models with a large dataset by bootstrapping our dataset. Three different experiments using bootstrapping and data balancing techniques were performed and the results obtained are shown in Figure 9.

Initially, we merged all three datasets given in Table 8 resulted in a comparatively large and an imbalanced dataset of 1210 sentences with 282 recommendation and 928 non-recommendation sentences. We named the generated dataset as “Merged Data”. The application of classical and advanced machine-learning models on this dataset is shown in Figure 9a,b, respectively. Among the classical model, decision tree (J48) model performed the best at an accuracy of 77.19%, but still below the proposed technique which stands at 81.63%. In deep-learning models CNN achieved 77.69%, LSTM 76.86%, and Bi-LSTM surpassed the proposed technique by 0.39%. The merged dataset is more inclined toward non-recommendation sentences, therefore, the trained models are also biased toward the non-recommendation sentence. We overcome dataset biases by duplicating the number of RS sentences, and swap their tokens, repeatedly. The resultant dataset referred to as “Swap Data” in Figure 9 consist of 846 RS and 929 NRS of 1775 sentences. The evaluation results of classical and deep-learning models on Swap Data are reflected in Figure 9, where the Naive Bayes achieved the highest accuracy of 76.95% in classical models while Bi-LSTM achieved highest accuracy of 79.88% in deep-learning model compared to 77.61% accuracy of the proposed technique.

Duplicating instances and swapping tokens may not be an efficient approach for trained a generalized model. Therefore, we balanced and enlarge the dataset by data augmentation [40], where we generated various RS sentences from the existing RS sentences by replacing word tokens with their synonyms. The resultant dataset referred to as “Augmented Data” in Figure 9 consists of 846 RS, 929 NRS sentences. The application of classical and deep-learning models on the augmented data is shown in Figure 9 where the naive-based remains at top; however, its accuracy dropped to 73.03%, while the proposed method accuracy dropped to 74.97% highest in the classical models. Similar to the previous cases, Bi-LSTM remains at top by achieving an accuracy of 83.05%, 8.08% higher than the proposed technique. Despite better performance of deep-learning models, the tree-based and pattern-based approaches are preferred in real clinical practices. Because the

pattern-based approaches perform well on small datasets compared to deep-learning models as observed from results in Figure 8b. Additionally, clinical decision-making needs transparent solutions to enhance the physician satisfaction. However, the pattern-based decision-making is traceable instead of deep-learning models.

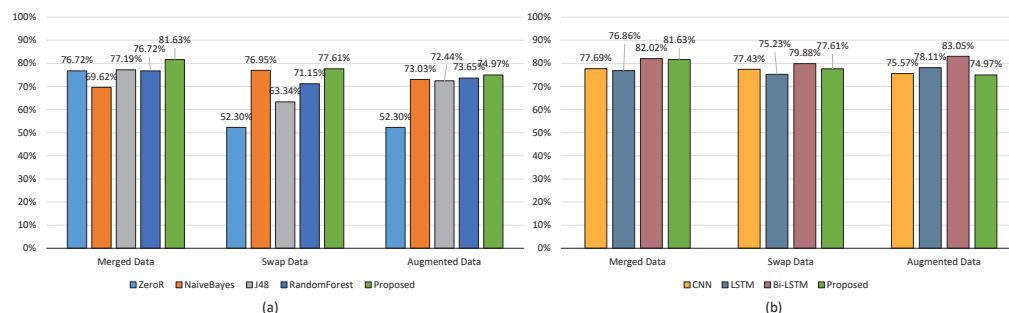


Figure 9. Evaluation of proposed method on large datasets (a) with classical models (b) with advanced models.

6. Conclusions

Clinical practice guidelines assist the domain experts in decision-making for diagnosis, management, and treatment. Healthcare providers face difficulties in CPG use. The effectiveness of CPGs can be increased by locating disease-specific information in a real-time manner. The primary contribution of this study is the set of patterns identified from the guidelines with and without machine-learning assistance and proposed the hybrid technique with a combination of heuristic, POS-based, and UMLS-based patterns for recommendation statement identification in guidelines. The extracted patterns identified recommendation sentences with 78.89%, 85.32%, and 92.07% accuracy in asthma, rhinosinusitis, and hypertension guidelines, respectively. These patterns can provide two-fold benefits. First, it can be used to identify specific information in a lengthy guideline. It increases the effectiveness of guidelines, their use, improves healthcare quality, helps in providing evidence-based practice, and reduces processing time for identifying disease-specific information. Second, it can be used for recommendation sentence annotation in CPG-related applications. In the future, we will extend this research work for guideline-based knowledge acquisition for assisting clinical decisions.

Author Contributions: Conceptualization, M.H.; Funding acquisition, S.L.; Methodology, M.H. and J.H.; Project administration, S.L. and T.C.; Software, M.H.; Supervision, S.L. and T.C.; Validation, T.A., S.I.A. and H.S.M.B.; Visualization, M.H., S.I.A. and H.S.M.B.; Writing—original draft, M.H.; Writing—review & editing, J.H., T.A., S.I.A. and H.S.M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-0-01629) supervised by the IITP (Institute for Information & communications Technology Promotion), by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00655), by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2020-0-01489) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation) and NRF-2019R1A2C2090504.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Field, M.J.; Lohr, K.N. (Eds.). *Clinical Practice Guidelines: Directions for a New Program*; National Academies Press: Washington, DC, USA, 1990.
- Davis, D.A.; Taylor-Vaisey, A. Translating guidelines into practice: A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *CMAJ* **1997**, *157*, 408–416.

3. Kaiser, K.; Miksch, S.; Tu, S.W. *Computer-Based Support for Clinical Guidelines and Protocols: Proceedings of the Symposium on Computerized Guidelines and Protocols (CGP 2004)*; IOS Press: Amsterdam, The Netherlands, 2004.
4. Wenzina, R.; Kaiser, K. Identifying condition-action sentences using a heuristic-based information extraction method. In *Process Support and Knowledge Representation in Health Care*; Springer: Berlin, Germany, 2013; pp. 26–38.
5. Fox, J.; Patkar, V.; Chronakis, I.; Begent, R. From practice guidelines to clinical decision support: Closing the loop. *J. R. Soc. Med.* **2009**, *102*, 464–473. [[CrossRef](#)] [[PubMed](#)]
6. Rello, J.; Lorente, C.; Bodí, M.; Diaz, E.; Ricart, M.; Kollef, M.H. Why do physicians not follow evidence-based guidelines for preventing ventilator-associated pneumonia?: A survey based on the opinions of an international panel of intensivists. *Chest* **2002**, *122*, 656–661. [[CrossRef](#)] [[PubMed](#)]
7. Kilsdonk, E.; Peute, L.W.; Riezebos, R.J.; Kremer, L.C.; Jaspers, M.W. From an expert-driven paper guideline to a user-centred decision support system: A usability comparison study. *Artif. Intell. Med.* **2013**, *59*, 5–13. [[CrossRef](#)] [[PubMed](#)]
8. Davis, D.A.; Thomson, M.A.; Oxman, A.D.; Haynes, R.B. Evidence for the effectiveness of CME: A review of 50 randomized controlled trials. *JAMA* **1992**, *268*, 1111–1117. [[CrossRef](#)] [[PubMed](#)]
9. Jang, B.; Kim, M.; Harerimana, G.; Kang, S.u.; Kim, J.W. Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism. *Appl. Sci.* **2020**, *10*, 5841. [[CrossRef](#)]
10. Thangaraj, M.; Sivakami, M. Text classification techniques: A literature review. *Interdiscip. J. Inf. Knowl. Manag.* **2018**, *13*, 117–135. [[CrossRef](#)]
11. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150. [[CrossRef](#)]
12. Jiang, M.; Liang, Y.; Feng, X.; Fan, X.; Pei, Z.; Xue, Y.; Guan, R. Text classification based on deep belief network and softmax regression. *Neural Comput. Appl.* **2018**, *29*, 61–70. [[CrossRef](#)]
13. Xu, S. Bayesian Naïve Bayes classifiers to text classification. *J. Inf. Sci.* **2018**, *44*, 48–59. [[CrossRef](#)]
14. Cai, D.; Garg, N.; Dobrzynski, M.; Guo, W.Q.; Khanna, A.; Xu, N. Content Pattern Based Automatic Document Classification. US Patent App. 15/713,445, 2019-03-28.
15. Fu, S.; Chen, D.; He, H.; Liu, S.; Moon, S.; Peterson, K.J.; Shen, F.; Wang, L.; Wang, Y.; Wen, A.; others. Clinical concept extraction: A methodology review. *J. Biomed. Informatics* **2020**, *109*, 103526. [[CrossRef](#)] [[PubMed](#)]
16. Yao, L.; Mao, C.; Luo, Y. Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 71. [[CrossRef](#)] [[PubMed](#)]
17. Bui, D.D.A.; Zeng-Treitler, Q. Learning regular expressions for clinical text classification. *J. Am. Med Inform. Assoc.* **2014**, *21*, 850–857. [[CrossRef](#)] [[PubMed](#)]
18. Zhong, N.; Li, Y.; Wu, S.T. Effective pattern discovery for text mining. *IEEE Trans. Knowl. Data Eng.* **2010**, *24*, 30–44. [[CrossRef](#)]
19. Gallagher, M.; Hares, T.; Spencer, J.; Bradshaw, C.; Webb, I. The nominal group technique: A research tool for general practice? *Fam. Pract.* **1993**, *10*, 76–81. [[CrossRef](#)] [[PubMed](#)]
20. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
21. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J.; Sutskever, L.; Zweig, G. word2vec. 2013. Available online: <https://code.google.com/p/word2vec/> (accessed on 13 January 2021).
22. Jacobsen, P.B. Clinical practice guidelines for the psychosocial care of cancer survivors: Current status and future prospects. *Cancer* **2009**, *115*, 4419–4429. [[CrossRef](#)]
23. Peleg, M. Computer-interpretable clinical guidelines: A methodological review. *J. Biomed. Inform.* **2013**, *46*, 744–763. [[CrossRef](#)]
24. Serban, R.; ten Teije, A.; van Harmelen, F.; Marcos, M.; Polo-Conde, C. Extraction and use of linguistic patterns for modelling medical guidelines. *Artif. Intell. Med.* **2007**, *39*, 137–149. [[CrossRef](#)]
25. Hematiam, H.; Zadrozny, W. Identifying condition-action statements in medical guidelines using domain-independent features. *arXiv* **2017**, arXiv:1706.04206.
26. Gad El-Rab, W.; Zaiane, O.R.; El-Hajj, M. Formalizing clinical practice guideline for clinical decision support systems. *Health Inform. J.* **2017**, *23*, 146–156. [[CrossRef](#)] [[PubMed](#)]
27. Priyanta, S.; Hartati, S.; Harjoko, A.; Wardoyo, R. Comparison of sentence subjectivity classification methods in Indonesian News. *Int. J. Comput. Sci. Inf. Secur.* **2016**, *14*, 407.
28. Dashtipour, K.; Gogate, M.; Li, J.; Jiang, F.; Kong, B.; Hussain, A. A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing* **2020**, *380*, 1–10. [[CrossRef](#)]
29. Lu, Q.; Zhu, Z.; Xu, F.; Guo, Q. Chinese Sentiment Classification Method with Bi-LSTM and Grammar Rules. *Data Anal. Knowl. Discov.* **2019**, *3*, 99–107.
30. HaCohen-Kerner, Y.; Miller, D.; Yigal, Y. The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE* **2020**, *15*, e0232525. [[CrossRef](#)] [[PubMed](#)]
31. Srividhya, V.; Anitha, R. Evaluating preprocessing techniques in text categorization. *Int. J. Comput. Sci. Appl.* **2010**, *47*, 49–51.
32. Shekar, B.; Dagnew, G. Grid search-based hyperparameter tuning and classification of microarray cancer data. In Proceedings of the IEEE 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 25–28 February 2019; pp. 1–8.

33. James, P.A.; Oparil, S.; Carter, B.L.; Cushman, W.C.; Dennison-Himmelfarb, C.; Handler, J.; Lackland, D.T.; LeFevre, M.L.; MacKenzie, T.D.; Ogedegbe, O.; et al. 2014 evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA* **2014**, *311*, 507–520. [CrossRef]
34. Manning, C.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S.; McClosky, D. The Stanford CoreNLP natural language processing toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA, 23–24 June 2014; pp. 55–60.
35. Bodenreider, O. The unified medical language system (UMLS): Integrating biomedical terminology. *Nucleic Acids Res.* **2004**, *32*, D267–D270. [CrossRef]
36. Aronson, A.R.; Lang, F.M. An overview of MetaMap: Historical perspective and recent advances. *J. Am. Med Inform. Assoc.* **2010**, *17*, 229–236. [CrossRef]
37. Chow, A.W.; Benninger, M.S.; Brook, I.; Brozek, J.L.; Goldstein, E.J.; Hicks, L.A.; Pankey, G.A.; Seleznick, M.; Volturo, G.; Wald, E.R.; et al. IDSA clinical practice guideline for acute bacterial rhinosinusitis in children and adults. *Clin. Infect. Dis.* **2012**, *54*, e72–e112. [CrossRef]
38. Society, B.T. Scottish Intercollegiate Guidelines Network. *Br. Guidel. Manag. Asthma. Thorax* **2003**, *58*, i1–94.
39. Jurafsky, D. Speech and Language Processing. Available online: https://web.stanford.edu/~jurafsky/slp3/slides/4_NB_Jan_10_2021.pdf (accessed on 19 March 2021).
40. Wei, J.; Zou, K. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv* **2019**, arXiv:1901.11196.