A Model for Drug Discovery on Unstructured Text using Semi-Supervised Learning and Fuzzy Matching

Christine K. Mulunda, Peter W. Wagacha & Lawrence Muchemi

University of Nairobi, Faculty of Science and Technology, Nairobi, KENYA
Department of Computing and Informatics

Abstract

Health related discoveries are mainly published as journal publications and the rate at which they are generated increases as new information and discoveries emerge. Discovery of latent medically-related terms in a document corpus is a challenging task where the researcher is not an expert in that domain and a viable database of medicine related words is not readily available. The objective of the study was to investigate the methodologies and best practices that enable discovery of latent drug terms found in health publications corpus for effective dissemination at county and national levels. Fuzzy matching methodology was considered for its near and exact matching algorithms. DrugBank dataset was chosen as reference for drug terms because of its comprehensive list of drugs, that are frequently updated and freely accessible. Semi-supervised learning was applied in modeling of multi-search medical terms on an hourly basis. Drug-name recognition, sentence categorization and information retrieval are among the features described in the presented model.

Keywords: fuzzy matching, latent drug recognition, classification, information retrieval, dissemination.

1. Introduction

The Kenya National Cancer Control Strategy 2017-2022 [1] aims at ensuring an effective co-ordination of cancer research by developing and maintaining a repository of publications, while one of the proposed flagship programs under the Kenya Health Strategic and Investment plan 2013-2017 [2] is transitioning of research into policy and practical solutions at both the County and National levels. This research therefore, set out research objectives that were aimed at investigating methodologies and best practices for effective dissemination of the results obtained from publications through recognition of latent drug names.

Automatic discovery of latent medically-related terms in a document corpus is a challenging task. The challenges include but are not limited to; the fact that most publications are manually typed and therefore prone to typing errors and spelling mistakes. Secondly, the researcher analyzing the publications to generate latent medical terms from the document corpus are not expert in that domain. Thirdly, finding a viable medically-related database is arduous. Lastly, medical terms are commonly non-English dictionary words, that are written as a mixture of numbers, special characters like (hyphen (-), brackets ()) and non-English dictionary words.

A review of existing research on health dissemination tools was also conducted where [3] utilizes fuzzy logic in automatic topic discovery while Gensim [4] algorithm is on use of Latent Semantic Indexing (LSI) and Latent Document Analysis (LDA). Tools developed for search and retrieval of Medline/PubMed database includes: Ovid medicine [5], PubMed [6], RefMed [7], Medline Ranker [8], MiSearch [9], Semantic Medline [10], MScanner [11], eTBlast [12], PubFocus [13], McSyBi [14], GoPubMed [15], PubNet [16], Babelmesh [17], HubMed [18], Semedico [19]. Google scholar [20] and Scopus [21] web crawler is applicable to various domain. Stanford [22] topic modeling toolbox summarizes text and tracks word usage while Mallet [23] is on document-topic hyper parameter optimization. VisTopic [24] and TopicNets [25] research area is on visual interactive analytics. GibbsLDA++ [26] utilizes LDA and Gibbs sampling in analysis of latent topic structures. Fathom [27] analyses calls for crisis counselling.

As an extension of the research study in [28] and the methodology presented in [31], this study analyzed and classified 3623 PubMed [29] health publications for drug-name Recognition, Sentence Categorization and Information Retrieval by utilizing [30], a fuzzy matching algorithm and semi-supervised learning in the study by [31]. Semi-supervised learning was for modeling multi-search medical terms on an hourly basis. A fuzzy sort ratio is set where a 95% and above match is considered for classification while 85-95% is expert reviewed as illustrated on the conceptual framework in Figure 1.
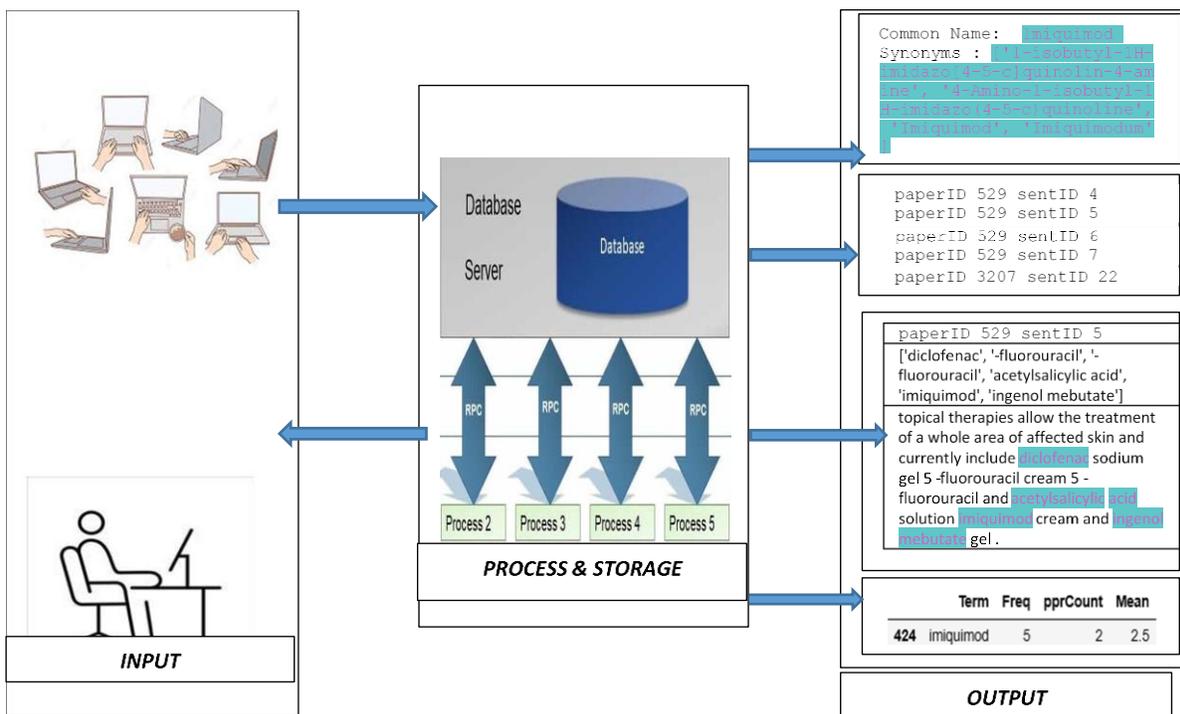

Figure 1: Conceptual Framework

The document corpus is automatically and sequentially analyzed, pre-processed and content fuzzy matched with DrugBank [32] datasets that was pre-downloaded and locally stored. Fuzzy matching was also applied to multi-searched terms. The output of the model includes: list of publications and sentences categorized by drug names; synonyms of drug name; aggregated frequency of drug names in a document and collection of documents; and number of documents containing similar drug names. DrugBank database was selected as reference for drug terms because of its comprehensive list of drugs, free accessibility and frequency of drug name updates in their databases.

2. Related works

The reviewed publications were on health dissemination tools, techniques and algorithms that are applicable to both short and full-text documents. The aim was to understand their application to topics discovery, classification and information retrieval.

Jain, Seeja and Jindal [33] applied Latent Semantic Analysis (LSA) and fuzzy similarity measure in their methodology, for concepts discovery and correlations identification. In topic modeling approach by Rashid, et.al [34] utilized fuzzy k-means latent semantic analysis (FKLSA) on medical and health text corpora. Multiple kernel fuzzy topic modeling technique by J. Rashid, et.al [35] was proposed for biomedical text mining. Fuzzy LSA-W and FLSA-V [36] was applied to topic embedding in text classification. In implementation of fuzzy keyword search over encrypted data in cloud computing [37] used edit distance to quantify keywords similarity. [3] presented fuzzy latent semantic analysis topic model as applied to health and medical corpora for automatic discovery of topics while Gensim [4] was developed for automatic extraction of semantic topics from documents. Google scholar [20] and Scopus [21] are applicable to multiple domain and its research areas includes algorithms on search optimization, information retrieval, ranking on documents.

In search, information retrieval and relevance ranking in health-related journals, Ovid Medline [5], PubMed [6], Medline ranker [8], MScanner [11], eTBlast [12], PubFocus [13] and BabelMeSH [17] tools application is on Medline database while RefMed [7], MiSearch [9], Semantic Medline [10], McSyBi [14], GoPubMed [15], PubNet [16], HubMed [18], Semedico [19] application is on PubMed Journal.

3. Methodology

Data mining of content-wise information from publications is vital in document classification for a dissemination tool. This requires a priori data to inform the extraction. In this study the a priori data is from DrugBank datasets and used for purposes of fuzzy extraction of content-wise data from health publications and normalizing multi-search inputs.

The study applies the semi-supervised topic model [31] approach to update the most searched N words from the specified topics. User search terms must have a fuzzy threshold ratio of .95 with the DrugBank dataset and also must exist in the publications. Results to users are documents categorized by drug names and sentences with highlighted drug-names.

In reference to the conceptual framework in figure 1, the study methodology of the proposed health dissemination tool is in three parts; input process, document process and output process.

3.1 Input process

The system allows for multi search medical terms as inputs. These inputs are processed to remove joining characters like (+/\&#). Fuzzy matching algorithm for exact and near matches are applied to these inputs and matched with DrugBank datasets. A fuzzy match threshold of .95 is set where all inputs including their timestamps are analyzed. If they are above threshold they are automatically modeled using semi-supervised learning methodology discussed in [31], and this is done on an hourly basis. The terms below the threshold are separately stored for expert review.

The adapted method in [31] is utilized as follows: - The initial population N is the extracted medical terms from the document corpus. These terms are then ranked based of frequency of their occurrence in documents. The top searched terms on an hourly basis are introduced into the population and a fitness function applied. This fitness function is the average of the aggregated terms from document corpus divided by the number of documents containing the terms. The least fit in the population based on the fitness function is removed to maintain the initial population size of N. This is done on a continuous basis and in real-time as the search operation is performed.

3.2 Document process

Document processing is sequential and automatic and is applied to full-text publications. The document corpus is a collection of type *.pdf* health related publications. The documents are converted to text and pre-processing done. Fuzzy matching approach is the technique used to extract exact and near match drug names from the documents. For each document ID text is extracted and sentence tokenization is performed while assigning a sentence ID to each tokenized sentence. A fuzzy threshold of .95 is set for extracting drug-names from the sentences. If the extracted term is below the threshold, it is saved separately for expert review. This is process is done automatically and sequentially for all documents processed in the repository.

3.3 Output process

The output of the extracted drugs is a statistical analysis of drug frequencies contained in a document, aggregated frequency of these drugs from the document corpus and the number of documents containing these drugs. This is irrespective of it being a common drug name or its synonym. Output from search inputs returns the common name of the drug searched and its synonyms. Thereafter, it returns the paperID, SentenceID and sentences where the common name or search name is found.

This also returns other drug names found in the document and their common name and synonyms. Drugs found in the sentences extracted are also highlighted

4. Results

In the study, 6.58GB of health publications from the PubMed Central database was downloaded, this was equivalent to 3,623 files from 177 Folders. Figure 2 shows the highest frequency of drug related words from the document corpus.
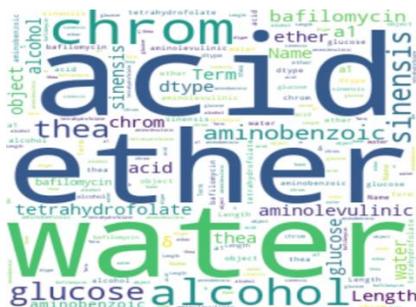

Figure 2. WordCloud

_____

Table 1 in Appendix shows the results from processing the multi-search inputs. The *drug searched* column is a representation of how each drug name that has been fuzzy matched is individually analyzed to give its *common name* and *synonyms* before extracting documents and sentences that contain either of these drug names. Further, analysis of other drugs found in the sentence is done to extract them and thereafter the sentence column shows the highlighted drug names.

5. Conclusion

Dissemination of information from publications is crucial for quick decision making by policy makers. Health related discoveries are mainly published as journal publications and the rate increases as new information and discoveries emerge. It is therefore important for health researchers to have tools and algorithms that allows fast information retrieval and categorization of related information. For precision retrieval it's crucial that these techniques are domain specific and application is located where the corpus is stored for sequential analysis of the continuously updated publications. Monitoring frequency of searched terms acts as a guide to policy makers on trends.

A prototype for automatic discovery of latent drug terms from corpus was developed. This effectively deviates from the norm of disseminating full-text publications that are tedious to manually analyze for specific content. The summary obtained as a classification of publication and sentences that only have the searched results *common* drug name or its *synonyms* allows for quick evaluation of content importance. Challenge is that some English dictionary words have fuzzy sort ratio score that are same or near same as the drug names which requires another layer to distinguish them. This process requires a dedicated expert reviewer to distinguish the drug names from English dictionary words.

As a developing country effective dissemination of health data in summarized format allows for rapid research growth and fast decision making by policy makers at both county and national levels. Future works proposes to implement the solution on a health repository in Kenya in line with [1,2] and the Kenya National Cancer Control Strategy and Kenya Health Sector Strategic and Investment Plan.

Data Availability

The data that supports the findings of this study are available upon request to the Corresponding Author at https://github.com/cmulunda/LatentDrugRecognition

DrugBank datasets restrictions apply to the availability of these data and were used under Academic Research Agreement. The data is available from [32] at https://go.drugbank.com/ .

References

[1] Anon.: National Cancer Control Strategy (2017 - 2022), Ministry of Public Health and Sanitation & Ministry of Medical Services, Kenya. https://repository.kippra.or.ke/handle/123456789/2802/.

[2] Anon.: Health Sector Strategic and Investment Plan (2013 - 2017), Ministry of Health, Kenya. http://guidelines.health.go.ke:8000/media/Kenya_Health_Sector_Strategic_Investment_Plan_2013_to_2017.pdf.

[3] A. Karami, A. Gangopadhyay, B. Zhou, and H. Kharrazi, "Fuzzy Approach Topic Discovery in Health and Medical Corpora," *International Journal of Fuzzy Systems*, vol. 20, pp. 1334-1345, 2018. https://doi.org/10.1007/s40815-017-0327-9

[4] R. Rehrek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," In *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*, pp. 46-50, Valletta, Malta, 2010.

[5] Anon., n.d: Ovid Medline, [Online] https://ovidsp.ovid.com/.

[6] Anon., n.d: PubMed, [Online] https://www.ncbi.nlm.nih.gov/pubmed/.

[7] H. Yu, T. Kim, J. Oh, S. Kim, "RefMed: relevance feedback retrieval system for PubMed," In *Proceedings of the 18th ACM conference on Information and knowledge management (CIKM '09)*. Association for Computing Machinery, pp. 2099-2100, New York, NY, USA, 2009. https://doi.org/10.1145/1645953.1646322

[8] J. F. Fontaine, A. Barbosa-Silva, M. Schaefer, M. R. Huska, E. M. Muro, M. A. Andrade-Navarro, "MedlineRanker: flexible ranking of biomedical literature," *Nucleic Acids Research*, vol. 37, pp. 141-146, 2009. https://doi.org/10.1093/nar/gkp353

[9] D. J. States, A. S. Ade, Z. C. Wright, A. V. Bookvich, B. D. Athey, "MiSearch adaptive pubMed search tool," *Bioinformatics*, vol. 25(7), pp. 974-976, 2009. https://doi.org/10.1093/bioinformatics/btn033

[10] T. C. Rindflesch, H. Kilicoglu, M. Fiszman, G. Rosemblat, D. Shin, "Semantic MEDLINE: An advanced information management application for biomedicine," *Information Services and Use*, vol. 31(1-2), pp. 15-21, 2011.

[11] G. L. Poulter, L. D Rubin, R. B Altman and C. Seoighe, "MScanner: a classifier for retrieving Medline citations," *BMC Bioinformatics*, vol. 9(108), 2008. https://doi.org/10.1186/1471-2105-9-108

[12] M. Errami, J. D. Wren, J. M. Hicks. and H. R. Garner, "eTBLAST: a web server to identify expert reviewers, appropriate journals and similar publications," *Nucleic Acids Research*, vol. 35, 2007. https://doi.org/10.1093/nar/gkm221

[13] M. V. Plikus, Z. Zhang, and C. Chuong, "PubFocus: semantic MEDLINE/PubMed citations analytics through integration of controlled biomedical dictionaries and ranking algorithm," *BMC Bioinformatics*, 2007, vol. 7(424).

_____

[14] Y. Yamamoto, and T. Takagi, "Biomedical knowledge navigation by literature clustering," *Journal of Biomedical Informatics*, vol. 40, pp. 114-130, 2007. https://doi.org/10.1016/j.jbi.2006.07.004

[15] A. Doms and M. Schroeder, "GoPubMed: exploring PubMed with the Gene Ontology," *Nucleic Acids Research*, vol. 33, pp. 783-786, 2005. https://doi.org/10.1093/nar/gki470

[16] S. M. Douglas, G. T. Montelione and M. Gerstein, "PubNet: a flexible system for visualizing literature derived networks," *Genome Biology*, vol. 6(9), 2005. https://doi.org/10.1186/gb-2005-6-9-r80

[17] F. Liu, M. Ackerman and P. Fontelo, "BabelMeSH: Development of a Cross-Language Tool for MEDLINE/PubMed," *AMIA Annu Symp Proc.*, 2006.

[18] A. D. Eaton, "HubMed: a web-based biomedical literature search interface," *Nucleic Acids Research*, vol. 34, pp. 745-747, 2006. https://doi.org/10.1093/nar/gkl037

[19] E. Faessler, and U. Hahn, "Semedico: A Comprehensive Semantic Search Engine for the Life Sciences," *Proceedings of ACL'17, System Demonstrations*, Vancouver, Canada, 2017.

[20] Anon., n.d.: Google Scholar, [Online] https://en.wikipedia.org/wiki/Google_Scholar

[21] Anon., n.d.: Scopus, [Online] https://www.scopus.com/home.uri.

[22] D. Ramage, and E. Rosen, "Stanford Topic Modeling Toolbox, 2009. https://downloads.cs.stanford.edu/nlp/software/tmt/tmt-0.2/.

[23] A. K. McCallum, "MALLET: A Machine Learning for Language Toolkit, http://mallet.cs.umass.edu

[24] Y. Yang, Q. Yao and H. Qu, "VISTopic: A visual analytics system for making sense of large document collections using hierarchical topic modeling," *Visual Informatics*, vol. 1(1), pp. 40-47, 2017. https://doi.org/10.1016/j.visinf.2017.01.005

[25] B. Gretarsson, J. O'Donovan, S. Bostandjiev, T. Hollerer, A. Asuncion, D. Newman, P. Smyth, "TopicNets: Visual Analysis of Large Text Corpora with Topic Modeling," *ACM Transactions on Intelligent Systems and Technology*, vol. 3(2), pp. 1-26, 2012. https://dl.acm.org/doi/10.1145/2089094.2089099

[26] X. H. Phan and C. T. Nguyen, "GibbsLDA++: A C/C++ Implementation of Latent Dirichlet Allocation," http://gibbslda.sourceforge.net.

[27] K. Dinakar, J. Chen, H. Lieberman, R. Picard, and R. Filbin, "Mixed-Initiative Real-Time Topic Modeling & Visualization for Crisis Counseling," *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pp. 417-426, 2015. https://doi.org/10.1145/2678025.27013

[28] C. K. Mulunda, P. W. Waiganjo and L. Muchemi, "Towards Implementation of an Information Dissemination Tool for Health Publications: Case of a Developing Country," *IST-Africa Conference (IST-Africa),* Kampala, Uganda, pp. 1-11, 2020.

[29] https://www.ncbi.nlm.nih.gov/

[30] gandersen101 / spaczz

[31] C. K. Mulunda, P. W. Wagacha, and L. Muchemi, "Semi-Supervised Topic Model for Sequential Data: A Genetic Algorithm Approach," *6th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pp. 90-94, Johannesburg, South Africa, 2019.

[32] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, "DrugBank5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Res*, 2018, https://doi.org/10.1093/nar/gkx1037

[33] S. Jain, K. R. Seeja, R. Jindal, "A New Methodology for Computing Semantic Relatedness: Modified Latent Semantic Analysis by Fuzzy Formal Concept Analysis," *Procedia Computer Science*, vol. 167, pp. 1102-1109, 2020. https://doi.org/10.1016/j.procs.2020.03.412

_____

[34] J. Rashid, S. S. Adnan, and A. Irtaza, "A novel fuzzy k-means latent semantic analysis (FKLSA) approach for topic modeling over medical and health text corpora," *Journal of Intelligent & Fuzzy Systems*, vol. 37(5), pp. 6573–6588, 2019. http://dx.doi.org/10.3233/JIFS-182776

[35] J. Rashid, J. Kim, A. Hussain, U. Naseem, and S. Juneja, "A novel multiple kernel fuzzy topic modeling technique for biomedical data," *BMC Bioinformatics*, vol. 23(275), 2022. https://doi.org/10.1186/s12859-022-04780-1

[36] E. Rijcken, F. Scheepers, P. Mosteiro, K. Zervanou, M. Spruit and U. Kaymak, "A Comparative Study of Fuzzy Topic Models and LDA in terms of Interpretability," *IEEE Symposium Series on Computational Intelligence (SSCI)*, Orlando, FL, USA, pp. 1-8, 2021. https://doi.org/10.1109/SSCI50451.2021.9660139

[37] N. Shekokar, K. Sampat, C. Chandawalla, J. Shah, "Implementation of Fuzzy Keyword Search over Encrypted Data in Cloud Computing," *Procedia Computer Science*, vol. 45, pp. 499-505, 2015. https://doi.org/10.1016/j.procs.2015.03.089

C O A S

Table I: Shows the transformation of results from the inputted multi-search medical terms to output

Manually Typed Search Terms				*Inputs: acetylsalicylic acid+Ticagrelor+imiquimod*
Processed Inputs				*Processed Inputs: ['acetylsalicylic acid', 'ticagrelor', 'imiquimod']*
Drug Searched	**CommonName and Synonym**	**List of papers and Sentences containing either CommonName or Synonym**	**Other Drugs found in the extracted Sentences**	**Sentence Extracted**
acetylsalicylic acid	Common Name: Acetylsalicylic acid Synonyms : ['2-Acetoxybenzenecarboxylic acid', '2-Acetoxybenzoic acid', 'Acetylsalicylate', 'Acetylsalicylsäure', 'acide 2-(acétyloxy)benzoïque', 'Acide acétylsalicylique', 'ácido acetilsalicilico', 'Acidum acetylsalicylicum', 'ASA', 'Aspirin', 'Aspirina', 'Azetylsalizylsäure', 'o-acetoxybenzoic acid', 'O-acetylsalicylic acid', 'o-carboxyphenyl acetate', 'Polopiryna', 'Salicylic acid acetate']	paperID 48 sentID 19	['aspirin', 'clopidogrel', 'atorvastatin', 'rosuvastatin', 'rosuvastatin', 'aspirin', 'clopidogrel']	patients will take aspirin clopidogrel and high-intensity statins (atorvastatin or rosuvastatin) within 24 hours of stroke onset followed by aspirin or clopidogrel alone from day 22 .
		paperID 529 sentID 5	['diclofenac', '-fluorouracil', '-fluorouracil', 'acetylsalicylic acid', 'imiquimod', 'ingenol mebutate']	topical therapies allow the treatment of a whole area of affected skin and currently include diclofenac sodium gel 5 - fluorouracil cream 5 -fluorouracil and acetylsalicylic acid solution imiquimod cream and ingenol mebutate gel .
		paperID 745 sentID 36	['alcohol', 'aspirin']	(12) the clinical backgrounds and comor- bidities such as age sex smoking alcohol consumption anthra- noid use aspirin use diabetes mellitus inflammatory bowel disease parkinson's disease and hypothyroidism related to colorectal neoplasms or mc were investigated using electronic medical records .
		paperID 1002 sentID 10	['acetylsalicylic acid', 'ticagrelor']	after anterior myocardial infarction was diagnosed 300 mg acetylsalicylic acid and 180 mg ticagrelor loading doses were given .
		paperID 1002 sentID 28	['acetylsalicylic acid', 'acetylsalicylic acid']	loading doses of 300 mg acetylsalicylic acid and 180 mg ticagrelor were given .
		paperID 1088 sentID 3	['naproxen', 'aspirin', 'ketoprofen']	this study was aimed at conducting a network meta-analysis to assess the efficacy and safety of 5 otcas – naproxen ibuprofendiclofenac aspirin and ketoprofen – in patients with primary dysmenorrhea .
		paperID 1088 sentID 8	['aspirin', 'naproxen', 'ibuprofen', 'diclofenac', 'ketoprofen']	as for efficacy outcomes all the included analgesics except aspirin were more effective than placebo in treating dysmenorrhea [naproxen (or 3 . 99 95 % ci 2 . 18 –7 . 30) ibuprofen (or 10 . 08 95 % ci 3 . 29 –30 . 85) diclofenac (or 11 . 82 95 % ci 2 . 66 –52 . 48) and ketoprofen (or 5 . 12 95 % ci 1 . 57 –16 . 69) .
		paperID 1088 sentID 10	['aspirin', 'ibuprofen', 'diclofenac']	aspirin was less effective than ibuprofen (or 0 . 17 95 % ci 0 . 04 –0 . 73) and diclofenac (or 1 . 17 95 % ci 0 . 02 –0 . 85) .
		paperID 1088 sentID 11	['diclofenac', 'ibuprofen', 'ketoprofen', 'naproxen', 'aspirin']	the sucra curves showed that diclofenac and ibuprofen were the most and second most effective (85 . 1 % and 83 . 8 % respectively) followed by ketoprofen naproxen and aspirin .
		paperID 1088 sentID 13	['diclofenac', 'ibuprofen', 'ketoprofen', 'diclofenac', 'ketoprofen', 'aspirin']	diclofenac versus ibuprofen (or 4 . 31 95 % ci 1 . 18 – 15 . 67) ketoprofen versus diclofenac (or 0 . 18 95 % ci 0 . 04 –0 . 78) and ketoprofen versus aspirin (or 0 . 41 95 % ci 0 . 18 –0 . 97) presented statistically significant differences .
		paperID 1088 sentID 14	['ketoprofen', 'ibuprofen', 'naproxen', 'aspirin', 'diclofenac']	ketoprofen and ibuprofen were ranked the best (sucra 90 . 6 % and 79 . 6 %) followed by naproxen aspirin and diclofenac .
		paperID 1217 sentID 14	['heparins', 'bivalirudin', 'aspirin']	guidelines recommend the use of adjunctive therapies during ppci including anticoagulants (heparins and bivalirudin) and antiplatelet activation drugs (aspirin and p2 y12 inhibitors) .
		paperID 1319 sentID 2	['acide acétylsalicylique', 'apixaban']	2 les antithrombotiques ne sont pas recommandés pour la thromboprophylaxie chez les patients atteints de covid-19 légère non hospitalisés dans un essai randomisé et contrôlé cessé prématurément en raison de taux d'événements moindres que prévu ni l'acide acétylsalicylique (aas) ni l'apixaban (à des doses prophylactiques ou thérapeutiques) n'ont réduit les taux d'événements cardiovasculaires ou pulmonaires majeurs comparativement au placebo chez des patients atteints de covid-19 légère non hospitalisés23 .
		paperID 1319 sentID 6	['acide acétylsalicylique']	5 l'acide acétylsalicylique n'améliore pas l'issue chez les patients hospitalisés non sélectionnés l'aas n'est pas recommandé pour prévenir la mortalité ou le recours au soutien des fonctions vitales chez les patients hospitalisés pour la covid-19 non sélectionnés36 .

Manually Typed Search Terms				_Inputs: acetylsalicylic acid+Ticagrelor+imiquimod_
Processed Inputs				_Processed Inputs: ['acetylsalicylic acid', 'ticagrelor', 'imiquimod']_
Drug Searched	**CommonName and Synonym**	**List of papers and Sentences containing either CommonName or Synonym**	**Other Drugs found in the extracted Sentences**	**Sentence Extracted**
		paperID 1352 sentID 1	['nattokinase', 'nattokinase', 'aspirin']	221 lab anim res 2013 : 29 (4) 221 -225 http://dx . doi . org/10 . 5625 /lar . 2013 . 29 . 4 . 221 nattokinase improves blood flow by inhibiting platelet aggregation and thrombus formation ja-young jang1 # tae-su kim2 # jingmei cai1 jihyun kim1 youngeun kim1 kyungha shin1 kwang sei kim1 sung kyeong park3 sung-pyo lee2 ehn-kyoung choi1 man hee rhee4 * yun-bae kim1 * 1 college of veterinary medicine chungbuk national university cheongju korea 2 misuba rtech co . ltd . asan korea 3 daejeon health sciences college daejeon korea 4 college of veterinary medicine kyungpook national university daegu korea the effects of nattokinase on the in vitro platelet aggregation and in vivo thrombosis were investigated in comparison with aspirin .
		paperID 1352 sentID 8	['nattokinase', 'aspirin']	in addition a high dose (500 mg/kg) of nattokinase fully prevented the occlusion as achieved with aspirin (30 mg/kg) .
		paperID 1615 sentID 8	['aspirin']	aspirin may also be effective for patients with ischemic stroke who have contraindications to anticoagulants although direct comparisons with anticoagulants are not available .
		paperID 1764 sentID 15	['aspirin']	in 1971 vane[5] demonstrated that aspirin mdomethacin and other non-steroid anti-inflammatory drugs inhibited the synthesis of prostaglandins and he postulated that these drugs owed their anti-inflammatory anti-pyretjc ancj analgesic actions to their ability to inhibit pg biosynthesis .
		paperID 2435 sentID 9	['aspirin']	interventions: the patient was administered g-globulin 1 . 0 g/kg·d for 2 consecutive days and oral aspirin 5 mg/kg·d .
		paperID 2435 sentID 11	['aspirin']	the dose of aspirin was reduced to 3 mg/kg·d after 2 months and to 1 . 5 mg/kg·d after 3 months .
		paperID 2703 sentID 13	['aspirin']	to prevent thrombus formation or attempt to resolve an existing thrombus low molecular weight heparin (schwarzwald et al . 2002 ; feige et al . 2003) acetylsalicylic acid (schoster 2017) and recombinant tissue plasminogen activator (bäumer et al . 2013) have been proposed as systemic treatments .
		paperID 2791 sentID 15	['aspirin']	she received dual antiplatelet therapy for 1 year followed by lifelong aspirin in addition to high-intensity statin and a beta-blocker .
		paperID 2799 sentID 26	['aspirin']	similar effect modification was observed for aspirin (or 1 . 90 ; 95 %ci 0 . 77 to 4 . 70 for non-users and or 1 . 02 ; 95 %ci 0 . 34 to 3 . 08 for users) and β-blockers . 2 these findings raise questions about the role of coronary circulation in cold-related pathogenesis of scd .
		paperID 2939 sentID 17	['aspirin', 'cilostazol', 'prednisolone']	since the pci she has been taking aspirin 100 mg; clopidogrel 75 mg and cilostazol 100 mg for coronary artery disease treatment and prednisolone 5 mg and hydroxyquinolone for lupus treatment .
		paperID 2989 sentID 15	['aspirin']	• aspirin is not recommended for patients with congestive heart failure (chf) .
		paperID 3036 sentID 23	['aspirin', 'atenolol', 'furosemide', 'furosemide', 'budesonide', 'albuterol', 'umeclidinium', 'umeclidinium']	his outpatient medication includes aspirin atenolol furosemide budesonide nebuliser albuterol and umeclidinium inhaler .
		paperID 3241 sentID 5	['aspirin']	we correlate the influence of chronic statin therapy (at least 6 months) with factors including age up to 50 years helicobacter pylori infection smoking and drinking habits ulcer history gastrotoxic drug consumption (low-dose aspirin [asa] anticoagulants) and comorbidities .
		paperID 3241 sentID 12	['aspirin', 'nitric oxide', 'pge2', 'pgi2', 'pgi2']	abbreviations: asa = low-dose aspirin ci = confidence interval hf = heart failure no = nitric oxide nsaids = non-steroidal anti-inflammatory drugs ors = odds ratio pge2 = prostaglandin e2 pgi2 = prostaglandin i2 ppi = proton pomp inhibitors sat = systemic atherosclerosis .
		paperID 3241 sentID 13	['aspirin']	keywords: comorbidities endoscopic lesions helicobacter pylori low-dose aspirin statin 1 .
		paperID 3382 sentID 16	['cascades', 'arachidonic acid', 'adenosine diphosphate',	generally platelet activation is triggered by several intracellular signaling cascades stimulated by different adhesive proteins and abbreviations: aa arachidonic acid; adp adenosine diphosphate; ac adenylyl cyclase; camp cyclic adenosine

Manually Typed Search Terms				Inputs: acetylsalicylic acid+Ticagrelor+imiquimod
Processed Inputs				*Processed Inputs: ['acetylsalicylic acid', 'ticagrelor', 'imiquimod']*
Drug Searched	**CommonName and Synonym**	**List of papers and Sentences containing either CommonName or Synonym**	**Other Drugs found in the extracted Sentences**	**Sentence Extracted**
			'adenosine', 'adenosine', 'camp', 'cyclic adenosine monophosphate', 'adenosine monophosphate', 'adenosine', 'adenosine', 'acetylsalicylic acid', 'adenosine triphosphate', 'adenosine', 'adenosine', 'calcium', 'inositol-', 'platelet-activating factor', 'prothrombin', 'tyrosine']	monophosphate; asa acetylsalicylic acid; atp adenosine triphosphate; akt protein kinase b; [ca2 þ]i intracellular calcium ion; cox cyclooxygenase; crp collagen-related peptide; csf crude saponin fraction; erk extracellular signalregulated kinase; gpvi glycoprotein vi; ic50 half maximal (50 %) inhibitory concentration; ip3 inositol-145 -triphosphate; jnk c-jun n-terminal kinase; mapk mitogen-activated protein kinase; mkk4 mitogen-activated protein kinase 4 ; mlc myosin light chain; paf platelet-activating factor; par proteinase-activated receptor; pi3 k phosphatidylinositol 3 -kinase; pka protein kinase a; pkc protein kinase c; pkg protein kinase g; cpla2 a cytosolic phospholipase a2 a; pla2 phospholipase a2 ; plcg2 phospholipase c gamma-2 ; aptt activated partial thromboplastin time; pt prothrombin time; ppd protopanaxadiol; ppt protopanaxatriol; rock rho-associated protein kinase; sfk src family kinase; syk spleen tyrosine kinase; ts total saponin; txa2 thromboxane a2 ; txb2 thromboxane b2 ; txas thromboxane-a synthase; txr thromboxane receptor; vasp vasodilator-stimulated phosphoprotein; vwf von willebrand factor .
		paperID 3407 sentID 23	['aspirin']	during workup patient complained of transient right sided face and arm tingling diagnosed with tia and started on aspirin .
		paperID 3423 sentID 13	['vonoprazan', 'simvastatin', 'simvastatin', 'cholesterol', 'simvastatin', 'simvastatin', 'simvastatin', 'verapamil', 'apatinib', 'vonoprazan', 'potassium', 'aspirin']	aphy-tandem mass spectrometry rat liver microsomes introduction as one of the drugs with wide clinical application simvastatin is mainly used to treat dyslipidemia and it can reduce the concentrations of total cholesterol triglyceride and low-density lipoprotein (ldl) in plasma accomplished by increasing the concentration of high-density lipoprotein (hdl) . 1 it has been reported that simvastatin is the substrates for cytochrome p450 (cyp) enzymes p-glycoprotein (p-gp) and organic anion-transporting polypeptide transporter (oatp1 b1) and it is mainly metabolized via cyp3 a4 /5 with cyp2 c8 /9 /19 and cyp2 d6 also involved in partial metabolism . 2 recently drug– drug interactions between simvastatin and other drugs were reported such as simvastatin could inhibit the metabolism of verapamil and apatinib and increase their bioavailability by inhibiting cyp3 a and p-glycoprotein (p-gp) efflux pumps . 34 vonoprazan a potassium-competitive acid blocker (p-cab) is a novel reversible proton pump inhibitor (ppi) that is used for the treatment or the prevention of gastric acid-related clinical conditions such as erosive esophagitis gastroesophageal reflux gastroduodenal ulcer aspirin or nsaid-induced peptic ulcer and it can also be used for the eradication of helicobacter drug design development and therapy 2022 :16 1779 –1789 1779 © 2022 hong et al .
		paperID 3491 sentID 13	['warfarin', 'aspirin', 'warfarin']	two patients changed their treatment from warfarin to aspirin; others remained on warfarin .
		paperID 3503 sentID 13	['aspirin', 'clopidogrel', 'metoprolol']	she was started on aspirin clopidogrel and metoprolol .
ticagrelor	Common Name: Ticagrelor Synonyms : ['(1S-2S-3R-5S)-3-(7-((1R-2S)-2-(3-4-Difluorophenyl)cyclopropylamino)-5-(propylthio)-3H-(1-2-3)triazolo(4-5-d)pyrimidin-3-yl)-5-(2-hydroxyethoxy)cyclopentane-1-2-diol', 'Ticagrelor']	paperID 1002 sentID 1	['ticagrelor']	60 case reports diffuse alveolar hemorrhage associated with ticagrelor therapy after percutaneous coronary intervention samet yılmaz oğuz kılıç yalın tolga yaylalı department of cardiology pamukkale university hospitals; denizli-turkey introduction after acute coronary syndrome (acs) various antiplatelet agents are used to prevent stent thrombosis death and recurrent myocardial infarction .
		paperID 1002 sentID 2	['clopidogrel', 'prasugrel', 'ticagrelor']	the currently available antiplatelet agents are clopidogrel prasugrel and ticagrelor .
		paperID 1002 sentID 3	['ticagrelor']	the european society of cardiology guidelines recommend ticagrelor as a first choice antiplatelet regimen after acs (1) .
		paperID 1002 sentID 6	['ticagrelor', 'ticagrelor']	although dyspnea is frequently reported in patients receiving ticagrelor dah due to ticagrelor use is very rare in the literature (2 3) .
		paperID 1002 sentID 7	['ticagrelor']	thus we present two cases of antiplatelet-related dah occurring after ticagrelor therapy .
		paperID 1002 sentID 10	['acetylsalicylic acid', 'ticagrelor']	after anterior myocardial infarction was diagnosed 300 mg acetylsalicylic acid and 180 mg ticagrelor loading doses were given .
		paperID 1002 sentID 20	['ticagrelor']	based on radiological findings dah due to ticagrelor emerged as the most obvious reason for hemoptysis .

Manually Typed Search Terms		*Inputs: acetylsalicylic acid+Ticagrelor+imiquimod*			
Processed Inputs		*Processed Inputs: ['acetylsalicylic acid', 'ticagrelor', 'imiquimod']*			
Drug Searched	**CommonName and Synonym**	**List of papers and Sentences containing either CommonName or Synonym**	**Other Drugs found in the extracted Sentences**	**Sentence Extracted**	
		paperID 1002 sentID 21	['ticagrelor', 'oxygen']	ticagrelor was stopped and continuous oxygen inhalation was started at 2 l/min .	
		paperID 1002 sentID 24	['clopidogrel', 'ticagrelor']	clopidogrel instead of ticagrelor was started as the second antiaggregant therapy and the patient was discharged from the hospital on the 7 th day .	
		paperID 1768 sentID 24	['ticagrelor', 'clopidogrel']	ticagrelor was used in 27 patients and clopidogrel in 3 as the dual antiplatelet .	
imiquimod	Common Name: Imiquimod Synonyms : ['1-isobutyl-1H-imidazo[4-5-c]quinolin-4-amine', '4-Amino-1-isobutyl-1H-imidazo(4-5-c)quinoline', 'Imiquimod', 'Imiquimodum']	paperID 529 sentID 4	['diclofenac', 'ingenol mebutate', 'imiquimod']	with an approach focused on the photo-damaged skin instead of single lesions topical therapies have the advantage over surgical or ablative therapies of treating subclinical lesions as well with recent evidence showing the latter also have the ability to degenerate into scc . 2 cost-efficacy analysis of 3 % diclofenac sodium ingenol mebutate and 3 . 75 % imiquimod in the treatment of actinic keratosis steven nisticò1 ester del duca1 valeria torchia2 micaela gliozzi2 ugo bottoni1 and carolina muscoli3 abstract actinic keratosis (ak) is a clinical condition characterized by keratinocytic dysplastic lesions of the epidermis affecting individuals chronically exposed to sunlight .	
		paperID 529 sentID 5	['diclofenac', '-fluorouracil', '-fluorouracil', 'acetylsalicylic acid', 'imiquimod', 'ingenol mebutate']	topical therapies allow the treatment of a whole area of affected skin and currently include diclofenac sodium gel 5 -fluorouracil cream 5 -fluorouracil and acetylsalicylic acid solution imiquimod cream and ingenol mebutate gel .	
		paperID 529 sentID 6	['diclofenac', 'ingenol mebutate', 'imiquimod']	due to the comparable efficacy of 3 % diclofenac ingenol mebutate and 3 . 75 % imiquimod in treating ak multiple lesions a pharmacoeconomic evaluation of cost-effectiveness of the three treatments was needed .	
		paperID 529 sentID 7	['diclofenac', 'ingenol mebutate', 'imiquimod']	a cost-efficacy analysis comparing 3 % diclofenac sodium with ingenol mebutate and 3 . 75 % imiquimod was performed .	
		paperID 3207 sentID 22	['fluorouracil', 'imiquimod', 'ingenol mebutate', 'diclofenac']	with multiple lesions the second-line therapy is the use of topical treatments including fluorouracil imiquimod ingenol mebutate and diclofenac [5] .	