BIOPRESS JOURNAL OF COMPUTATIONAL LIFE SCIENCES, VOL. 1, ISSUE. 1 (2025)

The Significance of Text Mining in Research: A Comprehensive Review

Badhe Pravin[1,]

[1]Swalife Biotech Ltd Unit 3D North Point House, North Point Business Park, Ireland

Corresponding author: Badhe Pravin,
E-mail: drbadhepravin@gmail.com

Doi: 10.31219/osf.io/yqf2e
Received: 15.03.2025
Accepted: 17.04.2025

Abstract

Text mining has emerged as a pivotal tool in various domains of research, revolutionizing the way scholars and scientists extract valuable insights from vast volumes of textual data. This comprehensive review explores the significance of text mining across different disciplines, focusing on its role in uncovering hidden knowledge and trends. Text mining involves the automated extraction, analysis, and interpretation of information from diverse text sources, including articles, books, websites, and social media. This paper discusses the applications of text mining, such as natural language processing, information retrieval, and data mining techniques, to discover patterns, relationships, and trends within textual data.

The review delves into the role of text mining in fields such as finance, healthcare, web content mining, business, accounting, and scientific literature analysis. In finance, text mining aids stock market prediction by analyzing news articles and textual data for trading decisions. In healthcare, it identifies trends in research and scientific literature, enabling researchers to discover novel patterns. Web content mining assists in extracting useful information from web documents, while in the business domain, it enhances decision-making by analyzing customer feedback and market trends. Text mining also streamlines accounting processes and audit automation.

Furthermore, this review highlights the application of text mining in predicting drug interactions, molecular targets, and drug repurposing in the context of breast cancer treatment. It showcases how text mining techniques, coupled with machine learning and deep learning algorithms, enable the identification of potential drug-target interactions, biomarkers, and therapeutic avenues. In breast cancer research, text mining aids in the discovery and validation of biomarkers, improving individualized treatment and prognosis.

The paper concludes by emphasizing the importance of text mining in facilitating drug safety analysis, particularly in pharmacovigilance. It explores text mining's capabilities in extracting potential adverse drug events from various sources, including medical case reports, social media platforms, clinical records, and electronic health records. Text mining's role in supporting drug safety professionals and early adverse event detection is underscored, highlighting its contribution to pharmaceutical decision-making and patient safety.

In summary, text mining is a transformative tool that transcends disciplinary boundaries, providing researchers and professionals with a powerful means to extract valuable insights from textual data and unlock hidden knowledge. Its versatility and applications make it an indispensable tool in the modern research landscape.

1.0 Introduction

Text mining plays a pivotal role in modern research across various disciplines, revolutionizing the way scholars and scientists extract valuable insights from vast amounts of textual data. This process involves the automated extraction, analysis, and interpretation of information from text sources, such as articles, books, websites, and social media. Its applications span from natural language processing to information retrieval, providing researchers with a powerful tool for uncovering hidden knowledge and trends.

1.1 Role of text mining in research

Text mining, also known as text data mining, is a research area that focuses on analyzing unstructured text data to extract valuable information and knowledge[1]. It involves the application of various techniques and algorithms from fields such as data mining, natural language processing (NLP), machine learning, and information retrieval [2]. The goal of text mining is to discover patterns, relationships, and trends within textual data [3].One of the most common techniques used in text mining is the Vector Space Model (VSM), which represents text as a bag of words This model has been widely used in traditional text mining approaches. However, recent research has explored the use of graph-based text representation models, which capture semantic relationships between words and concepts.These models have shown promise in improving the accuracy and effectiveness of text mining tasks [2].

Text mining has a wide range of applications across various domains. In the field of finance, text mining techniques have been used for stock market prediction. By analyzing news articles and other textual data, text mining can provide valuable insights for trading decisions[4]. In the healthcare domain, text mining has been used to identify trends in research and analyze scientific literature. It can help researchers discover novel patterns and relationships in the field of oropharyngeal dysphagia [3].

Text mining also plays a crucial role in web content mining, where it is used to extract useful information from web documents. With the rapid growth of the internet and the availability of large amounts of web data, text mining techniques are essential for analyzing and understanding web content. Web content mining can be categorized into different approaches, including unstructured text mining, structured mining, semi-structured text mining, and multimedia mining.

In the business domain, text mining has been applied to analyze vast amounts of text documents and extract valuable insights for decision-making.It can help organizations analyze customer feedback, market trends, and competitor information. By analyzing text data, businesses can improve their product quality, introduce new products to the market, and gain a competitive advantage [5].

Text mining also has applications in the field of accounting, where it can be used for accounting automation and audit automation. By automatically processing large amounts of text data, text mining techniques can extract and structure content, discover hidden data, and make automated decisions. This can help organizations streamline their accounting processes and improve efficiency [6].

Despite the advancements in text mining techniques, there are still challenges and open research questions in the field. Some of these challenges include the need for interdisciplinary coordination and cooperation, standardization of procedures, and addressing issues related to intellectual property rights [7]. Additionally, the effective integration of text mining with other data mining techniques and the development of personalized prescriptive systems are areas that require further research

1.2 Unlocking Insights: Text Mining for Extracting Knowledge from Vast Scientific Literature

Text mining techniques have gained significant popularity in academia and industry as a means to extract relevant information from a vast amount of scientific literature. The use of text mining in scientific literature has primarily focused on abstracts due to their availability. However, there is a growing interest in analyzing full-text articles to gain a more comprehensive understanding of the research. Text mining involves the application of natural language processing (NLP) and data mining techniques to extract facts, relationships, and knowledge from unstructured or semi-structured textual data. It aims to transform unstructured text into structured information that can be used to annotate specialized databases, transfer knowledge between domains, and support decision-making processes.

One area where text mining has been extensively applied is in the field of biomedical research. Biomedical text mining focuses on extracting information about biological entities such as genes, proteins, phenotypes, and biological pathways[8]. It has been used to extract protein-protein associations, disease-gene associations, and protein subcellular associations. For example, the popular STRING database utilizes automated text mining of the scientific literature to integrate known molecular interactions[9].

Text mining techniques have also been employed in materials science research. By applying natural language processing and information extraction techniques, researchers have been able to consume and codify scientific literature to generate rich datasets for data science and machine learning [10]. This has enabled data-driven materials research, allowing for the discovery of patterns and relationships in large-scale datasets [11].

In the field of design research, text mining techniques have been used to identify major academic branches and detect research trends. By analyzing scientific literature in design research, clustering and bibliometric analysis can be used to shape academic branches and summarize their characteristics [12]. This approach provides valuable insights into the evolution and trends within the field.

Text mining techniques have also been applied to fields such as human health risk assessment and electromagnetics. In assessing the human health risk of electromagnetic fields, text mining techniques have been used to extract information such as purpose statements, research categories, and sources of EMF [13]. This enables a systematic analysis of scientific literature to inform risk assessment processes.

The use of text mining in scientific literature has been facilitated by advancements in big data frameworks and machine learning algorithms. These technologies enable the processing and analysis of large volumes of text data, allowing for the extraction of valuable information. For example, SparkText is a biomedical text mining framework that leverages big data technologies to analyze large numbers of abstracts and full-text articles [11].

Despite the benefits of text mining in scientific literature, challenges still exist. One challenge is the extraction of information from tables within scientific publications. Traditional text mining techniques are primarily focused on unstructured text and may struggle to extract information from tables [14] However, efforts are being made to develop automated software tools that can extract information directly from tables and figures within scientific literature [15].

2.0 Text Mining in Breast Cancer drug discovery

Text mining is a valuable tool in drug discovery for breast cancer. It involves the process of extracting and analyzing information from a large number of unstructured texts to identify essential genes and potential drugs. By using text mining and bioinformatics analysis, researchers can search for candidate genes and drugs associated with breast cancer [15]. This approach allows for the identification of new drug targets and biomarkers for cancer diagnosis [16].

One study by focused on searching for essential genes and drug discovery in breast cancer using text mining and bioinformatics analysis. The researchers aimed to identify candidate drugs targeted to these genes for expanding the potential drug indications [15]The study found that some candidate genes and drugs were correlated with breast cancer through text mining [17].

Another study by explored the use of semantic text mining in early drug discovery for type 2 diabetes. Text mining was described as the process of discovering and capturing knowledge from a large number of unstructured texts. This technique has been applied to various problems in drug discovery, including the identification of new drug targets and the extraction of valuable information from literature [18].

Text mining can also be combined with other methods, such as microarray data analysis, to facilitate drug discovery. used text mining and microarray data analysis to identify potential drugs for canine pyometra disease [19]. The study analyzed genes and protein-protein interaction networks to select important genes for gene-drug interaction analysis. Similarly, text mining and microarray data analysis have been applied to the screening of disease biomarkers, the identification of signaling pathways, and the discovery of new drugs [19].

2.1 Text Mining for Breast Cancer Biomarker Discovery and Validation

Breast cancer is a complex disease with various subtypes and clinical outcomes. The identification and validation of biomarkers associated with breast cancer are crucial for improving individualized treatment and prognosis [20]. Text mining, a branch of data mining, can be a valuable tool in this process by extracting relevant information from biomedical literature and generating hypotheses about potential biomarkers[21].

One study by used text mining to identify molecular gene signatures associated with aggressive breast cancer. They found that elevated expression of cyclin B2, a cell cycle regulator, was associated with unfavorable clinical outcomes in invasive breast carcinoma [20]. This highlights the potential of text mining in identifying biomarkers that can provide prognostic information for breast cancer patients.

Another study by demonstrated the integration of text mining with network analysis to investigate breast cancer trends. They used text mining to analyze a large-scale collection of biomedical abstracts and generate hypotheses about breast cancer biomarkers [21]. This approach can help researchers identify novel biomarkers and understand the underlying biological mechanisms of breast cancer.

In addition to identifying biomarkers, text mining can also aid in the validation of biomarkers. For example, conducted a study to discover and validate salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. They used text mining to analyze the literature and identify potential biomarkers, which were then validated in preclinical samples. The study showed that transcriptomic and proteomic signatures in saliva can accurately detect breast cancer[22].

Text mining can also be used to retrieve and analyze hormone receptor expression characteristics in metastatic breast cancer. developed a novel tool that uses data mining on pathology reports to retrieve information about hormone receptor expression in metastatic breast cancer. This tool can help researchers and clinicians understand the hormone receptor status of metastatic breast cancer, which is crucial for treatment decisions and prognosis [23].

Furthermore, text mining can be integrated with other data mining techniques, such as network analysis, to identify genetic trends and interactions in breast cancer. applied text mining to extract information from biomedical literature and identify interactions between disease-associated biological units [24]. This approach can help uncover novel biomarkers and understand the genetic basis of breast cancer.

Overall, text mining can play a significant role in the identification and validation of biomarkers associated with breast cancer. By analyzing biomedical literature, text mining can generate hypotheses, identify potential biomarkers, and provide insights into the underlying biology of breast cancer. Integrating text mining with other data mining techniques can further enhance the discovery and understanding of breast cancer biomarkers.

2.2 Text mining in identifying potential biomarkers for diagnosis, prognosis, and treatment response prediction in breast cancer

Breast cancer is a significant global health burden, with over 1.3 million cases diagnosed and approximately 0.5 million associated mortalities annually worldwide [25]. The identification of precise biomarkers for the diagnosis, prognosis, and treatment of breast cancer is crucial.

One potential biomarker for breast cancer is MUC1, a glycoprotein that has been implicated in the metastasis and invasion of multiple tumors, including breast cancer. Overexpression of MUC1 has been associated with a poor prognosis in breast cancer patients. It has also been found to be positively correlated with CREB3L4 and may serve as a potential prognostic factor and therapy target [26].

Another potential biomarker is SLC50A1, a novel sugar transporter that has been identified as a potential serum-based diagnostic and prognostic biomarker for breast cancer. High expression of SLC50A1 has been associated with estrogen receptor-positive and HER2-positive breast cancer, as well as unfavorable outcomes in high-grade breast cancer. SLC50A1 has also been found to be associated with estrogen receptor status and HER2 status [27].

PPP1CA and PPP4C, members of the PPPCs family of phosphoprotein phosphatases, are other potential biomarkers for breast cancer. These proteins have clinical significance in breast cancer and may serve as potential biomarkers for diagnosis, prognosis, and treatment [28]. ADH1A and IGSF10 are two genes that have been identified as potential biomarkers for breast cancer. These genes have been associated with breast cancer prognosis and have shown significant down-regulation in breast cancer tissues[29].

FLAD1 is another potential biomarker that has been found to have diagnostic and prognostic value in breast cancer. Its expression has been explored across pan-cancer analysis and has shown promise as a biomarker for breast cancer [30].

MicroRNAs have also been investigated as potential biomarkers for breast cancer. Several studies have identified microRNAs that are associated with breast cancer prognosis and may serve as potential biomarkers [29,30]. These microRNAs have consistently shown down-regulation in breast cancer tissues [29].

Radiogenomics approaches using low-dose perfusion computed tomography (CT) have been explored to predict prognostic biomarkers and molecular subtypes in breast cancer. Machine learning approaches have been applied to radiogenomics to identify potential biomarkers and improve the prediction of prognosis and treatment response in breast cancer [31].

It is important to note that while there have been advancements in identifying potential biomarkers for breast cancer, no validated biomarker currently exists for routine clinical practice [32]. Further research and validation are needed to translate these potential biomarkers into clinical applications.

In summary, various potential biomarkers have been identified for the diagnosis, prognosis, and treatment response prediction in breast cancer. These biomarkers include MUC1, SLC50A1, PPP1CA, PPP4C, ADH1A, IGSF10, FLAD1, and microRNAs. Radiogenomics and machine learning approaches have also been explored to identify potential biomarkers in breast cancer. However, further research and validation are needed to determine the clinical utility of these biomarkers.

3.0 Text mining to predict interactions between drugs and molecular targets

Text mining has emerged as a valuable tool in predicting interactions between drugs and molecular targets. By analyzing large volumes of text data from biomedical literature and databases, researchers can uncover relationships between drugs and their target molecules. Several studies have utilized text mining techniques to predict drug-target interactions and facilitate drug discovery.

One study by proposed a deep-learning based model called DeepDTA for predicting drug-target binding affinities. The model utilized sequence information of both targets and drugs to predict the interaction binding affinities. The authors used convolutional neural networks (CNNs) to model protein sequences and compound 1D representations, achieving effective predictions of drug-target binding affinities [33].

Another study by introduced SimBoost, a read-across approach for predicting drug-target binding affinities using gradient boosting machines [34]. The study focused on the computational prediction of drug-target interactions and achieved accurate predictions for various binary drug-target benchmark datasets.

Text mining has also been applied to identify potential drug targets and biomarkers for specific diseases. For example, in a study on ovarian cancer, identified low expression of KCNN3 as a potential biomarker affecting drug resistance [35].

The authors used text mining and protein-gene interaction analyses to establish the association between KCNN3 and drug resistance.

In the field of computational biology, text mining has been combined with other analytical tools to discover new relationships between genes, diseases, and drugs. utilized text mining and biomedical databases to identify genes associated with endometriosis-induced infertility and potential drugs for treatment [36]. The authors analyzed gene enrichment and protein-protein interaction networks to select candidate drugs for further investigation.

Furthermore, text mining has been employed in drug repurposing efforts. conducted a study on canine pyometra disease, using text mining and microarray data analysis to identify potential drug targets and discover new drugs [37]. The authors analyzed gene-drug interactions and protein-protein interaction networks to provide evidence for potential drug discovery.

Text mining has proven to be a valuable tool in predicting interactions between drugs and molecular targets. By analyzing biomedical literature and databases, researchers can uncover new relationships, identify potential drug targets, and facilitate drug discovery efforts. The combination of text mining with other computational techniques, such as deep learning and network analysis, has shown promising results in predicting drug-target interactions and identifying potential therapeutic targets and biomarkers.

3.1 Text mining to predict interactions between drugs and molecular targets in the context of breast cancer.

To predict interactions between drugs and molecular targets in the context of breast cancer, text mining techniques can be employed to analyze biomedical literature and databases. Several studies have utilized text mining and machine learning approaches to predict drug-target interactions and facilitate drug discovery.

One study by developed a computational framework called DeepDDI that uses deep neural networks to accurately generate human-readable sentences describing drug-drug or drug-food constituent interactions [38]. The authors applied DeepDDI to predict drug-drug interactions (DDIs) and drug-food constituent interactions (DFIs) and showcased its applications in understanding DDI mechanisms causing adverse drug events (ADEs), suggesting alternative drug members, and predicting the effects of food constituents on interacting drugs [38].

In another study by , QSAR modeling was used to predict drug-drug interactions. The authors developed and validated QSAR models for the prediction of DDIs, representing interacting drug pairs as binary chemical mixtures and using quantitative neighborhoods of atoms (QNA) descriptors to predict the likelihood of DDIs. The models showed balanced accuracy for external test sets, demonstrating the potential of QSAR modeling in predicting DDIs [38].

Furthermore, machine learning techniques have been applied to predict complexation performance between cyclodextrins and guest molecules, which can be relevant in drug formulation and delivery. These studies utilized molecular descriptors and experimental conditions as inputs to predict complexation free energy and solubilizing capacity of cyclodextrin systems [39]

It is important to note that while text mining and machine learning approaches have shown promise in predicting drug-target interactions, further validation and experimental studies are necessary to confirm the predicted interactions and their clinical relevance. Additionally, the integration of other data sources, such as omics data and clinical data, can enhance the accuracy and applicability of these predictions.

Text mining and machine learning techniques have been utilized to predict interactions between drugs and molecular targets in the context of breast cancer. These approaches have shown potential in understanding drug interactions, suggesting alternative drug members, and predicting the effects of food constituents on drug efficacy. However, further research and validation are needed to translate these predictions into clinical applications.

4.1 Text mining can facilitate the repurposing of existing drugs

Text mining has been recognized as a valuable approach to facilitate the repurposing of existing drugs. By analyzing biomedical literature and databases, text mining techniques can uncover relationships between drugs, targets, and clinical outcomes, providing a platform for computer-aided drug repurposing (CADR). One study by proposed an integrated network approach for CADR, associating drugs, targets, and clinical outcomes. The authors utilized bio- and cheminformatics tools, intensive data mining, and curation to apply CADR. By linking drug-target information through semantic frameworks, such as Chem2Bio2RDF, computer-aided drug repurposing can be achieved [40].

In the context of cancer, developed an open-access database of licensed cancer drugs, which can serve as a valuable resource for literature mining and drug repurposing efforts. The database provides curated datasets of drugs licensed for therapeutic antineoplastic purposes, enabling researchers to explore potential repurposing candidates [41].

Machine learning algorithms and natural language processing techniques have also been employed in drug repurposing research. utilized database tables of drug repurposing candidates, molecular targets, and pathways to develop a literature-based methodology for drug repurposing in oncology [41]. The availability of curated datasets and the mining of published scientific literature contribute to the identification of potential drug repurposing opportunities.

In the field of neuroimmunological investigations, developed a drug repurposing workflow named Adera2.0, which utilizes deep neural networks for text mining and drug repurposing The workflow combines deep learning networks to embed text into matrices and repurpose drug candidates, shortening the drug development cycle [42].

Graph-based methods have also been employed in drug repurposing, including the search for new therapeutic avenues. These methods can mine drug characteristics, such as usages, indications, actions, and targets, to identify potential repurposing opportunities [41].

Overall, text mining plays a crucial role in facilitating the repurposing of existing drugs. By analyzing biomedical literature, databases, and curated datasets, text mining techniques can uncover relationships between drugs, targets, and clinical outcomes, providing valuable insights for drug repurposing efforts. The integration of machine learning algorithms, natural language processing, and graph-based methods further enhances the identification of potential repurposing candidates and therapeutic avenues.

4.2 Repurposing of existing drugs for breast cancer treatment

The Repurposing Drugs in Oncology (ReDO) project has developed a database called ReDO_DB, which includes data from 268 drugs that have been repurposed for various indications, including breast cancer. The database provides a comprehensive resource for researchers and clinicians interested in exploring repurposed drugs for breast cancer treatment [43].

In a review article by , the authors provide a comprehensive overview of the current scenario of drug repurposing for breast cancer treatment. They discuss the strategies and challenges associated with drug repurposing and provide several examples of repurposed drugs that have been used in breast cancer treatment, including alkylating agents, anthracyclines, antimetabolites, CDK4/6 inhibitors, aromatase inhibitors, mTOR inhibitors, and mitotic inhibitors [44].

Conducted a study on tissue-of-origin-specific gene repositioning in breast and prostate cancer. They identified genes that reposition in breast cancer tissues and found tissue-specific organization of the genome in normal breast tissues. By testing prostate gene positioning in breast tissue, they identified additional genes that reposition in breast cancer [45].

These studies highlight the potential of repurposing existing drugs for breast cancer treatment. The ReDO_DB database provides a valuable resource for identifying repurposed drugs, and ongoing research continues to explore new opportunities for drug repurposing in breast cancer [43].

Breast cancer is a significant global health problem and remains a leading cause of cancer deaths in women. The development of new drugs for breast cancer treatment is a time-consuming and expensive process. Therefore, repurposing existing drugs has emerged as a promising strategy to expedite the drug development process and provide alternative treatment options [46].

Drug repurposing, also known as drug repositioning, involves identifying new therapeutic uses for drugs that were originally developed for other indications. This approach takes advantage of the well-established pharmacokinetics, pharmacodynamics, and toxicological profiles of existing drugs, making it easier to gain regulatory approval for new. Additionally, repurposing drugs can potentially reduce costs and shorten the timeline for drug development [46].

Several drugs have been successfully repurposed for breast cancer treatment. For example, alkylating agents, anthracyclines, antimetabolites, CDK4/6 inhibitors, aromatase inhibitors, mTOR inhibitors, and mitotic inhibitors have been repositioned for breast cancer therapy. These drugs have shown efficacy in treating breast cancer and have been used in combination with other therapies [46].

The process of drug repurposing involves various strategies, including computational approaches, high-throughput screening, and identification of new targets. Advances in genomics and proteomics have facilitated the identification of potential drug targets and the discovery of new uses for existing drugs [46]. For example, the use of kinases as potential targets for drug repurposing in oncology has gained attention [47].

In addition to identifying new therapeutic uses for existing drugs, drug repurposing also involves exploring drug combinations. Combining repurposed drugs with existing antineoplastic agents has shown promise in breast cancer therapy. For example, combinations of antimalarial drugs with 5-FU, DOX, and PTX have demonstrated efficacy in breast cancer treatment [48]. The choice of the antineoplastic drug influences the success of the drug combination, highlighting the importance of selecting appropriate drug pairs [48].

Despite the potential benefits of drug repurposing, there are challenges associated with this approach. One challenge is selecting the appropriate patient population for clinical trials of repurposed drugs. Physiological responses to repurposed drugs may differ from the original indications, necessitating careful consideration of inclusion and exclusion. Additionally, the repurposed drugs for breast cancer often work best in combination therapy rather than monotherapy [46].

In conclusion, drug repurposing has emerged as a promising strategy for breast cancer treatment. It offers the potential to expedite the drug development process, reduce costs, and provide alternative treatment options for patients. Several drugs have already been successfully repurposed for breast cancer therapy, and the identification of new drug combinations holds promise for improving treatment outcomes. However, careful consideration of patient selection and the choice of drug combinations is necessary for the success of drug repurposing efforts in breast cancer treatment.

5.1 Text Mining for Breast Cancer Drug Safety Analysis

Text mining techniques are widely used to analyze adverse events and safety concerns in the field of pharmacovigilance. These techniques involve the extraction of meaningful information from large amounts of unstructured text, such as medical case reports, clinical records, social internet message boards, and electronic health records [49,50,51,52,53,54,55,56].

One approach to text mining for adverse events is the extraction of potential adverse drug events from medical case reports. developed a strategy for identifying and extracting information about potential adverse drug events from free-text resources [49]. They used computational methods to automatically collect potential adverse drug events communicated as free-text data, providing a scalable auto-assistance platform for drug safety professionals. This approach allows for the efficient identification and extraction of information from medical case reports, supporting pharmacovigilance research and pharmaceutical decision making[49].

Another text mining technique involves mining social internet message boards to identify potential adverse drug events. reported on the extraction of drug-event pairs using co-occurrence of terms within a window. This method can be used to identify potential adverse drug events from user-generated content on social media platforms, providing valuable insights into safety concerns [49].

Text mining can also be applied to clinical records and electronic health records to flag adverse events early and enable drug safety surveillance. Lependu et al. demonstrated that analyzing large volumes of free-text clinical notes enables drug safety surveillance using untapped data sources [53]. Electronic trigger tools, which mine vast amounts of clinical and administrative data, offer a promising method for detecting patient safety events and adverse events. These tools are more efficient and effective in detecting adverse events compared to voluntary reporting or the use of patient safety indicators [52].

In addition to extracting information from text, text mining techniques can be used for text classification and keyword extraction. proposed a deep learning-based text classification model for adverse nursing events, which showed promising results compared to existing methods [57]. demonstrated the feasibility of using an automated approach to detect multiple adverse events in various data sources, and suggested the use of natural language processing techniques for monitoring adverse events in hospitals [54].

Overall, text mining techniques provide valuable tools for analyzing adverse events and safety concerns. They enable the efficient extraction of information from unstructured text sources, such as medical case reports, clinical records, social internet message boards, and electronic health records. These techniques can support pharmacovigilance research, pharmaceutical decision making, and drug safety surveillance. However, further research is needed to address technical challenges and determine the relative contribution of each textual source to improving pharmacovigilance [50].

References

1. Torre, C., Sánchez, D., Blanco, I., & Martin-Bautista, M. (2018). Text mining: techniques, applications, and challenges. International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, 26(04), 553-582. https://doi.org/10.1142/s0218488518500265
2. Chang, J. and Kim, I. (2013). Analysis and evaluation of current graph-based text mining researches.. https://doi.org/10.14257/astl.2013.42.23
3. Krishnamurthy, R. and Balasubramanium, R. (2019). Using text mining to identify trends in oropharyngeal dysphagia research: a proof of concept. Communication Sciences & Disorders, 24(1), 234-243. https://doi.org/10.12963/csd.19579
4. Islam, M., Alshaikhli, I., Nor, R., & Vijayakumar, V. (2018). Technical approach in text mining for stock market prediction: a systematic review. Indonesian Journal of Electrical Engineering and Computer Science, 10(2), 770. https://doi.org/10.11591/ijeecs.v10.i2.pp770-777
5. Chhajer, S. and Satpathy, R. (2022). Adaptation of text mining as text data mining, similar to text analytics for the process of driving high-quality information from text. Technoaretetransactions on Intelligent Data Mining and Knowledge Discovery, 1(1), 1-5. https://doi.org/10.36647/ttidmkd/01.01.a001
6. ÖZYİĞİT, H. (2022). Muhasebe alanina güncel yaklaşimlar: metin madenciliği. Muhasebe Ve Vergi Uygulamaları Dergisi, 15(3), 637-663. https://doi.org/10.29067/muvu.1104525
7. Senseney, M. and Koehl, E. (2018). Text data mining beyond the open data paradigm: perspectives at the intersection of intellectual property and ethics. Proceedings of the Association for Information Science and Technology, 55(1), 890-891. https://doi.org/10.1002/pra2.2018.14505501162
8. Westergaard, D., Stærfeldt, H., Tønsberg, C., Jensen, L., & Brunak, S. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. Plos Computational Biology, 14(2), e1005962. https://doi.org/10.1371/journal.pcbi.1005962
9. Srivastava, P., Bej, S., Yordanova, K., & Wolkenhauer, O. (2021). Self-attention-based models for the extraction of molecular interactions from biological texts. Biomolecules, 11(11), 1591. https://doi.org/10.3390/biom11111591
10. Olivetti, E., Cole, J., Kim, E., Kononova, O., Ceder, G., Han, T., … & Hiszpanski, A. (2020). Data-driven materials research enabled by natural language processing and information extraction. Applied Physics Reviews, 7(4). https://doi.org/10.1063/5.0021106
11. Ye, Z., Tafti, A., He, K., Wang, K., & He, M. (2016). Sparktext: biomedical text mining on big data framework. Plos One, 11(9), e0162721. https://doi.org/10.1371/journal.pone.0162721
12. Nie, B. and Sun, S. (2017). Using text mining techniques to identify research trends: a case study of design research. Applied Sciences, 7(4), 401. https://doi.org/10.3390/app7040401
13. Lee, S., Kwon, J., Lee, B., & Kim, E. (2020). Scientific literature information extraction using text mining techniques for human health risk assessment of electromagnetic fields. Sensors and Materials, 32(1), 149. https://doi.org/10.18494/sam.2020.2572
14. Singh, G., Kuzniar, A., Mulligen, E., Gavai, A., Bachem, C., Visser, R., … & Finkers, R. (2018). Qtltableminer++: semantic mining of qtl tables in scientific articles. BMC Bioinformatics, 19(1). https://doi.org/10.1186/s12859-018-2165-7
15. Luo, L., Zheng, W., Chen, C., & Sun, S. (2021). Searching for essential genes and drug discovery in breast cancer and periodontitis via text mining and bioinformatics analysis. Anti-Cancer Drugs, 32(10), 1038-1045. https://doi.org/10.1097/cad.0000000000001108
16. Feng, Z., Shen, Z., Li, H., & Li, S. (2022). E-tsn: an interactive visual exploration platform for target–disease knowledge mapping from literature. Briefings in Bioinformatics, 23(6). https://doi.org/10.1093/bib/bbac465
17. Wang, Y., Zhao, B., & Yuan, X. (2022). Drug discovery and potential gene and pathway associated with polycystic ovary syndrome through text mining and biomedical databases. Reproductive and Developmental Medicine, 7(1), 44-49. https://doi.org/10.1097/rd9.0000000000000035
18. Hansson, L., Hansen, R., Pletscher-Frankild, S., Berzins, R., Hansen, D., Madsen, D., … & Jensen, J. (2020). Semantic text mining in early drug discovery for type 2 diabetes. Plos One, 15(6), e0233956. https://doi.org/10.1371/journal.pone.0233956

19. Wang, X. and Yu, G. (2023). Drug discovery in canine pyometra disease identified by text mining and microarray data analysis. Biomed Research International, 2023, 1-11. https://doi.org/10.1155/2023/7839568

20. Shubbar, E., Kovács, A., Hajizadeh, S., Parris, T., Nemes, S., Gunnarsdóttir, K., … & Helou, K. (2013). Elevated cyclin b2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. BMC Cancer, 13(1). https://doi.org/10.1186/1471-2407-13-1

21. Conceição, S. and Couto, F. (2021). Text mining for building biomedical networks using cancer as a case study. Biomolecules, 11(10), 1430. https://doi.org/10.3390/biom11101430.

22. Zhang, L., Xiao, H., Karlan, S., Zhou, H., Gross, J., Elashoff, D., … & Wong, D. (2010). Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. Plos One, 5(12), e15573. https://doi.org/10.1371/journal.pone.0015573.

23. Chang, K., Wang, J., Chang, C., & Chu, Y. (2020). Development of a novel tool for the retrieval and analysis of hormone receptor expression characteristics in metastatic breast cancer via data mining on pathology reports. Biomed Research International, 2020, 1-7. https://doi.org/10.1155/2020/2654815

24. Wang, C., Jin, J., Chang, J., Hayakawa, M., Kitazawa, A., Tsai, J., … & Sheu, P. (2020). Identification of most influential co-occurring gene suites for gastrointestinal cancer using biomedical literature mining and graph- based influence maximization. BMC Medical Informatics and Decision Making, 20(1). https://doi.org/10.1186/s12911-020-01227-6.

25. Park, E., Lee, K., Seo, B., Cho, K., Woo, O., Son, G., … & Chang, Y. (2019). Machine learning approaches to radiogenomics of breast cancer using low-dose perfusion computed tomography: predicting prognostic biomarkers and molecular subtypes. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-54371-z

26. Jing, X., Liang, H., Hao, C., Yang, X., & Cui, X. (2018). Overexpression of muc1 predicts poor prognosis in patients with breast cancer. Oncology Reports. https://doi.org/10.3892/or.2018.6887

27. Wang, Y., Shu, Y., Gu, C., & Fan, Y. (2019). The novel sugar transporter slc50a1 as a potential serum-based diagnostic and prognostic biomarker for breast cancer. Cancer Management and Research, Volume 11, 865-876. https://doi.org/10.2147/cmar.s190591

28. Xie, W., Zeng, Y., Hu, L., Zhi, J., Ling, H., Zheng, X., … & Gao, M. (2021). Comprehensive analysis of pppcs family reveals the clinical significance of ppp1ca and ppp4c in breast cancer. Bioengineered, 13(1), 190-205. https://doi.org/10.1080/21655979.2021.2012316

29. Wu, M., Li, Q., & Wang, H. (2021). Identification of novel biomarkers associated with the prognosis and potential pathogenesis of breast cancer via integrated bioinformatics analysis. Technology in Cancer Research & Treatment, 20, 153303382199208. https://doi.org/10.1177/1533033821992081

30. Mei, M., Song, W., Wang, Y., & Zhang, M. (2021). Significant diagnostic and prognostic value of flad1 and related micrornas in breast cancer after a pan-cancer analysis. Disease Markers, 2021, 1-21. https://doi.org/10.1155/2021/6962526

31. Park, J., Shin, Y., Kim, T., Kim, D., & Lee, A. (2019). Urinary metabolites as biomarkers for diagnosis of breast cancer: a preliminary study. Journal of Breast Disease, 7(2), 44-51. https://doi.org/10.14449/jbd.2019.7.2.44

32. Pultz, B., Seimon, R., Faria, P., Oliveira, A., Araújo, R., & Silva, M. (2014). Far beyond the usual biomarkers in breast cancer: a review. Journal of Cancer, 5(7), 559-571. https://doi.org/10.7150/jca.8925

33. Öztürk, H., Özgür, A., & Ozkirimli, E. (2018). Deepdta: deep drug–target binding affinity prediction. Bioinformatics, 34(17), i821-i829. https://doi.org/10.1093/bioinformatics/bty593

34. He, T., Heidemeyer, M., Ban, F., Cherkasov, A., & Ester, M. (2017). Simboost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. Journal of Cheminformatics, 9(1). https://doi.org/10.1186/s13321-017-0209-z

35. Liu, X., Wei, L., Zhao, B., Cai, X., Dong, C., & Yin, F. (2018). Low expression of kcnn3 may affect drug resistance in ovarian cancer. Molecular Medicine Reports. https://doi.org/10.3892/mmr.2018.9107

36. Lin, Y., Chen, Z., Huang, J., Fang, J., Qi, Y., Xie, L., … & Chen, L. (2022). Computational drug discovery in patients with endometriosis-induced infertility via text mining and biomedical databases. Clinical and Experimental Obstetrics & Gynecology, 49(8), 168. https://doi.org/10.31083/j.ceog4908168

37. Wang, Y., Shu, Y., Gu, C., & Fan, Y. (2019). The novel sugar transporter slc50a1 as a potential serum-based

diagnostic and prognostic biomarker for breast cancer. Cancer Management and Research, Volume 11, 865-876. https://doi.org/10.2147/cmar.s190591

38. Zakharov, A., Varlamova, E., Lagunin, A., Dmitriev, A., Muratov, E., Fourches, D., … & Nicklaus, M. (2016). Qsar modeling and prediction of drug–drug interactions. Molecular Pharmaceutics, 13(2), 545-556. https://doi.org/10.1021/acs.molpharmaceut.5b00762
39. Zhao, Q., Ye, Z., Su, Y., & Ouyang, D. (2019). Predicting complexation performance between cyclodextrins and guest molecules by integrated machine learning and molecular modeling techniques. Acta Pharmaceutica Sinica B, 9(6), 1241-1252. https://doi.org/10.1016/j.apsb.2019.04.004
40. Oprea, T., Nielsen, S., Ursu, O., Yang, J., Taboureau, O., Mathias, S., … & Bologa, C. (2011). Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. Molecular Informatics, 30(2-3), 100-111. https://doi.org/10.1002/minf.201100023
41. Pantziarka, P., Capistrano, R., Potter, A., Vandeborne, L., & Bouche, G. (2021). An open access database of licensed cancer drugs. Frontiers in Pharmacology, 12. https://doi.org/10.3389/fphar.2021.627574
42. Lazarczyk, M., Duda, K., Bhaumik, S., AK, O., Paszkiewicz, J., Kowalczyk, A., … & Sacharczuk, M. (2022). Adera2.0: a drug repurposing workflow for neuroimmunological investigations using neural networks. Molecules, 27(19), 6453. https://doi.org/10.3390/molecules27196453
43. Pantziarka, P., Verbaanderd, C., Sukhatme, V., Capistrano, I., Crispino, S., Gyawali, B., … & Bouche, G. (2018). Redo_db: the repurposing drugs in oncology database. Ecancermedicalscience, 12. https://doi.org/10.3332/ecancer.2018.886
44. Aggarwal, S., Verma, S., Aggarwal, S., & Gupta, S. (2021). Drug repurposing for breast cancer therapy: old weapon for new battle. Seminars in Cancer Biology, 68, 8-20. https://doi.org/10.1016/j.semcancer.2019.09.012
45. Meaburn, K., Agunloye, O., Devine, M., Leshner, M., Roloff, G., True, L., … & Misteli, T. (2016). Tissue-of-origin-specific gene repositioning in breast and prostate cancer. Histochemistry and Cell Biology, 145(4), 433-
446. https://doi.org/10.1007/s00418-015-1401-8
46. Aggarwal, S., Verma, S., Aggarwal, S., & Gupta, S. (2021). Drug repurposing for breast cancer therapy: old weapon for new battle. Seminars in Cancer Biology, 68, 8-20. https://doi.org/10.1016/j.semcancer.2019.09.012
47. Islam, S., Wang, S., Bowden, N., Martin, J., & Head, R. (2021). Repurposing existing therapeutics, its importance in oncology drug development: kinases as a potential target. British Journal of Clinical Pharmacology, 88(1), 64-74. https://doi.org/10.1111/bcp.14964
48. Duarte, D., Rêma, A., Amorim, I., & Vale, N. (2022). Drug combinations: a new strategy to extend drug repurposing and epithelial-mesenchymal transition in breast and colon cancer cells. Biomolecules, 12(2), 190. https://doi.org/10.3390/biom12020190
49. Gurulingappa, H., Mateen-Rajput, A., & Toldo, L. (2012). Extraction of potential adverse drug events from medical case reports. Journal of Biomedical Semantics, 3(1). https://doi.org/10.1186/2041-1480-3-15
50. Harpaz, R., Callahan, A., Tamang, S., Low, Y., Odgers, D., Finlayson, S., … & Shah, N. (2014). Text mining for adverse drug events: the promise, challenges, and state of the art. Drug Safety, 37(10), 777-790. https://doi.org/10.1007/s40264-014-0218-z
51. LePendu, P., Iyer, S., Bauer-Mehren, A., Harpaz, R., Mortensen, J., Podchiyska, T., … & Shah, N. (2013). Pharmacovigilance using clinical notes. Clinical Pharmacology & Therapeutics, 93(6), 547-555. https://doi.org/10.1038/clpt.2013.47
52. Murphy, D., Meyer, A., Sittig, D., Meeks, D., Thomas, E., & Singh, H. (2018). Application of electronic trigger tools to identify targets for improving diagnostic safety. BMJ Quality & Safety, 28(2), 151-159. https://doi.org/10.1136/bmjqs-2018-008086
53. LePendu, P., Iyer, S., Fairon, C., & Shah, N. (2012). Annotation analysis for testing drug safety signals using unstructured clinical notes. Journal of Biomedical Semantics, 3(S1). https://doi.org/10.1186/2041-1480-3-s1-s5
54. Piscitelli, A., Bevilacqua, L., Labella, B., Parravicini, E., & Auxilia, F. (2020). A keyword approach to identify adverse events within narrative documents from 4 italian institutions. Journal of Patient Safety, 18(1), e362- e367. https://doi.org/10.1097/pts.0000000000000783

55. Eskildsen, N., Eriksson, R., Christensen, S., Aghassipour, T., Bygsø, M., Brunak, S., … & Hansen, S. (2020).Implementation and comparison of two text mining methods with a standard pharmacovigilance method for signal detection of medication errors. BMC Medical Informatics and Decision Making, 20(1). https://doi.org/10.1186/s12911-020-1097-0

56. Yu, Y., Ruddy, K., Mansfield, A., Zong, N., Wen, A., Tsuji, S., … & Jiang, G. (2020). Detecting and filtering immune-related adverse events signal based on text mining and observational health data sciences and informatics common data model: framework development study. Jmir Medical Informatics, 8(6), e17353. https://doi.org/10.2196/17353

57. Lu, W., Jiang, W., Zhang, N., & Xue, F. (2021). A deep learning-based text classification of adverse nursing events. Journal of Healthcare Engineering, 2021, 1-10. https://doi.org/10.1155/2021/9800114