


# Enhancing Literature Review Efficiency: A Case Study on Using Fine-Tuned BERT for Classifying Focused Ultrasound-Related Articles

Reanna K. Panagides <sup>1,\*</sup>, Sean H. Fu <sup>2,\*</sup>, Skye H. Jung <sup>1</sup>, Abhishek Singh <sup>1</sup>, Rose T. Eluvathingal Muttikkal <sup>1</sup>, R. Michael Broad <sup>2</sup> , Timothy D. Meakem <sup>2</sup> and Rick A. Hamilton <sup>2</sup>

<sup>1</sup> School of Data Science, University of Virginia, Charlottesville, VA 22903, USA; hsj5sn@virginia.edu (S.H.J.); as4hfs@virginia.edu (A.S.); re4mfy@virginia.edu (R.T.E.M.)

<sup>2</sup> Focused Ultrasound Foundation, Charlottesville, VA 22903, USA; mbroad@fusfoundation.org (R.M.B.); tmeakem@fusfoundation.org (T.D.M.); rhamilton@fusfoundation.org (R.A.H.)

\* Correspondence: rkp3em@virginia.edu (R.K.P.); seanhfu@gmail.com (S.H.F.)

**Abstract:** Over the past decade, focused ultrasound (FUS) has emerged as a promising therapeutic modality for various medical conditions. However, the exponential growth in the published literature on FUS therapies has made the literature review process increasingly time-consuming, inefficient, and error-prone. Machine learning approaches offer a promising solution to address these challenges. Therefore, the purpose of our study is to (1) explore and compare machine learning techniques for the text classification of scientific abstracts, and (2) integrate these machine learning techniques into the conventional literature review process. A classified dataset of 3588 scientific abstracts related and unrelated to FUS therapies sourced from the PubMed database was used to train various traditional machine learning and deep learning models. The fine-tuned Bio-ClinicalBERT (Bidirectional Encoder Representations from Transformers) model, which we named FusBERT, had comparatively optimal performance metrics with an accuracy of 0.91, a precision of 0.85, a recall of 0.99, and an F1 of 0.91. FusBERT was then successfully integrated into the literature review process. Ultimately, the integration of this model into the literature review pipeline will reduce the number of irrelevant manuscripts that the clinical team must screen, facilitating efficient access to emerging findings in the field.

**Keywords:** focused ultrasound; machine learning; text classification; BERT



**Citation:** Panagides, R.K.; Fu, S.H.; Jung, S.H.; Singh, A.; Eluvathingal Muttikkal, R.T.; Broad, R.M.; Meakem, T.D.; Hamilton, R.A. Enhancing Literature Review Efficiency: A Case Study on Using Fine-Tuned BERT for Classifying Focused Ultrasound-Related Articles. *AI* **2024**, *5*, 1670–1683. <https://doi.org/10.3390/ai5030081>

Received: 7 August 2024

Revised: 31 August 2024

Accepted: 3 September 2024

Published: 10 September 2024



**Copyright:** © 2024 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

### 1.1. Rationale

Focused ultrasound (FUS), also called high-intensity focused ultrasound (HIFU), is a procedure that uses an acoustic lens to concentrate ultrasound waves on a single focal point deep in the body to heat up and destroy or change tiny patches of human tissue without affecting the surrounding normal tissue. Over the past decade, FUS has emerged as a promising therapeutic modality for the treatment of various medical conditions with extreme precision and accuracy. This technique is especially known for being minimally invasive when accessing these desired anatomical structures [1]. In comparison to the traditional thermal ablation mechanisms used in a clinical practice, such as radiofrequency currents, microwaves, and laser, this incisionless nature leads to reduced postoperative recovery periods allowing patients a quicker return to their daily routines, friends, and family. Presently, there exist nine FDA-approved indications for FUS therapies such as essential tremor, uterine fibroids, Parkinson's disease, prostate cancer, benign prostate hyperplasia, liver tumors, and pain from metastatic lesions to the bone [2]. The growing prominence of FUS therapies makes it imperative for physicians and researchers to stay informed about novel scientific breakthroughs and ongoing developments.

Literature review plays a pivotal role in the advancement of medical knowledge and the adoption of innovative therapies like FUS. However, the traditional approach to literature review, often reliant on manual examination, is labor-intensive, time-consuming, and susceptible to human error. As the volume of published research continues to grow exponentially, the need for more efficient and accurate review processes becomes increasingly critical. This is particularly relevant to the field of focused ultrasound therapies, which has seen an exponential increase in research over the past decade. However, keyword-based database searches often struggle to differentiate between studies related to focused ultrasound therapies and those involving ultrasound for diagnostic purposes. This subtle but important distinction between two similar yet distinct applications of ultrasound technology highlights the need for more advanced article classification methods.

Machine learning offers a promising solution to address these challenges [3]. Algorithms have been previously used for text classification in various applications such as the sentiment analysis [4] of product reviews, the identification of hate speech on social media [5], and the classification of news topics. By utilizing similar advanced algorithms and processes for text classification, as successfully applied in the areas mentioned previously, clinicians and researchers can use machine learning to automate and streamline various stages of the literature review process, enhancing both efficiency and effectiveness. One key area where machine learning can significantly impact literature review is in the screening of articles for inclusion criteria.

Six Steps of the Literature Review Process [6]:

1. Formulating the research question;
2. Searching publication databases for relevant articles;
3. Screening articles for inclusion criteria;
4. Assessing the quality of primary studies;
5. Extracting data;
6. Analyzing data.

Literature review, typically reliant on manual examination, involves painstakingly scanning through manuscripts deemed 'relevant' by Boolean keyword searches in published literature databases [6]. Unfortunately, this current methodology has flaws as search criteria often yield papers containing mere keyword mentions, resulting in many irrelevant results. Machine learning algorithms, on the other hand, can analyze the content of articles in a more nuanced manner, identifying relevant information beyond keyword matches.

By training machine learning models on labeled datasets of relevant and irrelevant articles, researchers can develop classifiers capable of accurately distinguishing between the two categories. These classifiers can then be used to automate the initial screening of articles, flagging those that are likely to meet the inclusion criteria for further review by human experts. This automated screening process could save time and resources and reduce the likelihood of missing important articles. By automating tedious tasks, improving the accuracy of article screening, and enhancing quality assessment, machine learning can help researchers stay up to date on the latest developments and make more informed decisions based on the available evidence.

Given these challenges, there is a knowledge gap in the application of advanced machine learning techniques within the literature review process when researching a specific topic of interest, such as focused ultrasound therapies [7]. While fine-tuned large language models trained on general medical and scientific language exist, there are currently no models specifically designed to classify abstracts related to specific therapies, such as FUS [8,9]. This gap highlights an opportunity to enhance the literature review process by leveraging natural language processing (NLP) to more accurately identify relevant articles while proposing an opportunity for the integration of NLP into the everyday workflow of physicians and researchers. Using FUS as a subject-specific classification target, this project can ultimately serve as a case study, exemplifying the process by which machine learning can be used within the everyday life of a researcher or physician to help them stay updated on relevant research without spending egregious amounts of time finding relevant articles.

## 1.2. Purpose

The primary research question guiding this study is as follows: Can machine learning models, specifically those that utilize deep learning methods such as transformers, be applied to automate the classification of scientific abstracts related to focused ultrasound therapies, thereby potentially improving the efficiency and accuracy of the literature review process? Therefore, the aim of this study is twofold. Our primary objective is to explore and compare various machine learning techniques for the binary classification of abstracts pertaining to FUS therapies, distinguishing those related from those unrelated. Secondly, we aim to integrate these machine-learning techniques into the literature review process. This integration is highlighted in a workflow diagram depicting the application of these techniques to augment the efficiency of screening articles based on inclusion criteria specific to FUS therapies.

It is important to note that although we propose a workflow for the easiest integration of this classification model within the literature review process, this preliminary study will not include a further thematic analysis of included articles as one does in a typical systematic literature review but rather propose a more efficient way to exclude articles that are irrelevant to a specific topic. Our chosen subject matter, focused ultrasound therapies, can serve as an example of how these types of deep learning models could be trained to perform such a task.

## 1.3. Background-Text Classification Pipeline

The study of text classification as a subset of natural language processing dates back to the 1960s [10]. While various models for text classification have evolved over time, the general pipeline using machine learning has remained relatively consistent. This pipeline typically comprises six stages: (1) obtaining data, (2) preprocessing the data, (3) feature extraction, (4) model selection and training, (5) model evaluation, and (6) making predictions on unseen data [10].

1. **Obtaining Data:** The initial step in the pipeline, obtaining the data, involves accessing data from reputable sources and ensuring that the data are labeled into binary classes based on their relation to FUS. For this study, we are solely interested in the supervised learning of pre-labeled data although it is possible to use unsupervised learning for text classification [10].
2. **Preprocessing:** Preprocessing the data are crucial to clean and transform them into a suitable format for model input. This may entail tasks such as tokenization, the removal of English stop words, normalization, text standardization/stemming, and lemmatization. Preprocessing enhances the quality of the input text, thereby improving the performance of the model [11].
3. **Feature Extraction:** The third step, feature extraction, is used in traditional machine learning and deep learning approaches to text classification to convert words into numerical vector representation. Examples of feature extraction methods include Bag-of-Words (BoW), Term Frequency-Inverse, Document Frequency (TF-IDF), and word embedding (Word2Vec, BERT, etc.) [10,12]. The choice of feature extraction methods depends on the model chosen for the text classification task [12].
4. **Model Selection and Training:** Model selection and training, the fourth step, involves choosing a suitable model for training on the input data. Various models can be utilized for text classification, broadly categorized into traditional machine learning and deep learning models (Table 1). Traditional machine learning methods include Naive Bayes, Support Vector Machines (SVMs), Decision Trees (DTs), Random Forests (RFs), Logistic Regression, and K-Nearest Neighbor (KNN) [13]. Deep learning methods include neural networks, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers [14]. It is important to note that although all mentioned methods have been used for text classification, the invention of transformers, such as Bidirectional Encoder Representations from Transformers (BERT) [15], have been shown to significantly outperform all other types of traditional

and deep learning methods for this NLP task [16]. Although fine-tuning BERT models for text classification requires more computational power than training traditional models, we are interested in seeing how the performance of fine-tuned BERT models compare. We decided to only investigate the performance of transformers, over CNNs and RNNs, due to the literature supporting the use of transformers for text classification tasks over other deep learning methods. Similarly, a few traditional methods were also explored to assess whether the time and effort required to train a more complex model like BERT was justified, or if a simpler model could achieve comparable results.

5. **Model Evaluation:** In the fifth step, model evaluation involves selecting appropriate metrics to quantify performance. Certain metrics can be optimized depending on their relevancy and appropriateness to the problem addressed through the fine-tuning of hyperparameters. Common metrics include accuracy, precision, recall, and F1-score [17].

**Table 1.** Materials and Methods.

Traditional Methods	Deep Learning Methods
Naive Bayes	CNNs
Support Vector Machines	RNNs
Decision Trees	Transformers (i.e., BERT)
Random Forests	
Logistic Regression	
K-Nearest Neighbor	

## 2. Materials and Methods

The study was conducted in two distinct phases: (1) exploration of machine learning (ML) methods for text classification, and (2) integration of ML methods into the scientific literature review process. In the initial phase, we investigated various traditional and deep learning approaches to text classification to identify the most effective one. Subsequently, in the second phase, we adapted the existing literature review pipeline to include ML automation, aiming to enhance its efficiency.

### 2.1. Obtaining Data

The data consist of a curated collection of scientific abstracts sourced from PubMed, a search engine maintained by the National Center for Biotechnology Information (NCBI). The Focused Ultrasound Foundation (FUSF) provided Excel files from their monthly literature review process, covering February to August 2023. These literature searches yielded a variety of articles both related to FUS and unrelated to FUS, and used search parameters selected from an expert in the field, outlined below:

Articles had to include at least one of the following keywords:

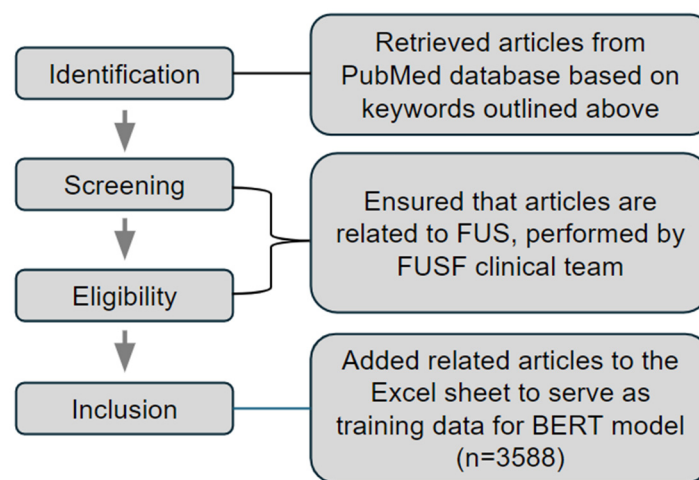
1. Focused Ultrasound;
2. HIFU (High-Intensity Focused Ultrasound);
3. MRgFUS (MR-Guided Focused Ultrasound);
4. LIFU (Low-Intensity Focused Ultrasound);
5. Ultrasound Imaging;
6. Transducer;
7. Ultrasound Ablation;
8. High Intensity Focused Ultrasound Ablation;
9. Diagnostic Ultrasound.

During these literature searches, conditions were added to ensure that neither retracted publications nor preprints were included in the search. We also adhered to all privacy and text mining policies, and confirmed that all publications were open source-accessible. After compiling the articles into an Excel file, the Focused Ultrasound Foundation clinical team performed a manual review to validate that the scraped publications in the Excel

sheet were indeed related to FUS, totaling 489 articles. This initial dataset was bolstered by 90 labeled abstracts of scientific articles used in prior work by FUSF. Furthermore, a dataset of FUS-related scientific abstracts was available in a publicly accessible Zotero database on the FUSF website [18]. We used all abstracts within this database, except those categorized as veterinary-related, amounting to 1960 articles.

There was an overlap of keywords in articles on FUS technology and those studying the use of ultrasound for diagnostic purposes. To help our model identify articles using ultrasound for diagnosis as non-FUS related, we made sure to incorporate additional articles that fell under the “Ultrasound Imaging” or “Diagnostic Ultrasound” categories of the PubMed literature searches.

All collected abstracts of scientific articles, along with their corresponding labels, were consolidated into a CSV file. Duplicate abstracts and null values were removed from the dataset. The final dataset consisted of 1794 abstracts related to FUS and an equivalent number of non-FUS abstracts for a total of 3588 abstracts (Figure 1).



**Figure 1.** A simple PRISMA Flowchart to illustrate our process of data acquisition.

## 2.2. Data Preprocessing and Feature Extraction

The process of preparing the data for our analysis began with the classification of scientific abstracts, sourced from the FUS therapy field. The classification was conducted by the Chief Medical Officer (CMO) of the Foundation, with each abstract marked with a binary indicator to denote relevance to the FUS domain. After compiling our collected data, we were able to move on to the next step.

In the preprocessing phase, each abstract was transformed to standardize the text to make it more conducive to the machine learning analysis. The first step in this process was to convert all the abstracts to string format and lowercase, which is a common practice in text processing that helps reduce the complexity of the text data by consolidating variations of the same word [11]. Following this, we used tokenization to break down the text into individual words or tokens. This step is crucial for understanding the structure of sentences and for subsequent feature extraction techniques. We specifically used a tokenization process compatible with the BERT model. This involved using the AutoTokenizer function with specific parameters made to prepare the text for modeling. The tokenizer was configured to ensure that there was uniform sequence length across the dataset through padding, where shorter sequences were extended to match the longest sequence in the batch using a designated padding token. Also, sequences exceeding the BERT model’s maximum input size of 512 tokens were truncated to this limit. Lastly, the tokenized output was converted into PyTorch tensors to align with the computational framework used for model training.

After the text tokenization step, we used several pre-trained BERT models for feature extraction, translating the scientific abstracts into numerical data that reflected both the meaning and structure of the language in the text. BERT does this by assigning each token

to vectors (also known as token embeddings) that capture the meanings of the words, using positional embedding, to maintain word order, and segment embeddings, to manage different parts of the text [19]. BERT's transformer architecture allows for the model to understand each word in the context of the whole text, which is enhanced through its pre-training on large datasets [20]. Pre-training helps BERT better understand complex patterns in language, making it able to identify relevant nuances in FUS literature.

For the analysis using Logistic Regression, Support Vector Machine (SVM), and Naive Bayes models, we used the TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer to convert text data into a matrix of TF-IDF features. This method is particularly effective for these models as it not only considers the frequency of words but also how unique the words are across the entire dataset, which enhances the performance of these traditional machine learning models.

### 2.3. Model Selection and Training

To address the primary aim of our study, we conducted a comparative analysis between traditional machine learning and deep learning methodologies for text classification. As for traditional machine learning methods, Logistic Regression [21], Naive Bayes [22], and Support Vector Machine [23] (SVM) models were employed for classification, chosen over some of the other models for their specialization in binary classification (SVM and Logistic Regression) and working with text data (Naive Bayes). These models used features extracted from a TF-IDF [3] tokenizer as parameters for binary classification, with no additional optimization performed.

In contrast, for the deep learning approach, we selected and trained several BERT (Bidirectional Encoder Representations from Transformers) models. These transformer models fit the scope of our project well and are the current state-of-the-art approach for text classification. The BERT models we trained included TinyBERT [24], SciBERT [25], DistilBERT [26], and Bio-ClinicalBERT [27] for performance evaluation. Each BERT model had been pre-trained on a distinct corpus of text. For instance, Bio-ClinicalBERT [27] was pre-trained on biomedical and clinical text data, tailored for healthcare and biomedicine tasks, while DistilBERT [26] and TinyBERT [24] were scaled-down versions of the original BERT model designed for improved efficiency while maintaining performance. SciBERT [25], on the other hand, was pre-trained on scientific text, particularly enhancing performance within scientific domains (Table 2).

**Table 2.** Description of BERT model variants.

Model	Total Parameters	Dataset
TinyBERT [24]	14,350,874	Original BERT corpus
SciBERT [25]	109,920,002	Semantic Scholar corpus
DistilBERT [26]	66,955,010	Original BERT corpus
Bio-Clinical BERT [27]	108,311,810	MIMIC-III corpus

We selected these classification methods and used 2906 abstracts to train traditional machine learning models and fine-tune the pre-trained BERT models.

We fine-tuned each pre-trained BERT model on our corpus of FUS-related abstracts for more precise classification specific to FUS literature and coined our new model, 'FusBERT'. It is important to note that we decided to only train FusBERT on scientific abstracts for efficiency and easier integration into the literature review pipeline, which involves screening abstracts before screening the entire article.

A hyperparameter optimization of BERT models was conducted by applying a systematic grid search approach using Ray Tune over a predefined set of hyperparameters (number of epochs, training batch size, and learning rate) [28]. The chosen values for grid search were based on the hyperparameter search space recommended by BERT authors [16]. Additionally, a seed was set for reproducibility.

- Epochs: 2, 3, 4;
- Train batch size: 8, 16, 32;
- Learning rate:  $2 \times 10^{-5}$ ,  $3 \times 10^{-5}$ ,  $5 \times 10^{-5}$ ;
- Seed value: 6013.

Once the optimal hyperparameters for our models were determined, the best-performing model was selected for integration into the literature review pipeline. The hyperparameter optimizations were evaluated using a validation dataset of 323 abstracts. The BERT hyperparameters chosen after grid search are summarized below (Table 3).

**Table 3.** BERT model hyperparameter summary.

Model	Train Batch Size	Learning Rate	Epochs
TinyBERT	16	$5 \times 10^{-5}$	3
SciBERT	32	$5 \times 10^{-5}$	3
DistilBERT	16	$3 \times 10^{-5}$	3
Bio-ClinicalBERT	8	$3 \times 10^{-5}$	4

In pursuit of our secondary aim, the chosen model was integrated into the literature review process. The initial steps of the pipeline, involving article retrieval and compilation, were automated using Python scripts to scrape articles from PubMed and preprocess the abstracts. Subsequently, our FusBERT model generated predictions regarding the relatedness of abstracts to FUS therapies, aiding researchers in efficiently screening articles. The workflow for integrating FusBERT underwent iterative refinement through collaboration with the clinical and data management team at the Focused Ultrasound Foundation, with the final workflow presented in the subsequent Results section.

#### 2.4. Evaluation Metrics

Each model was carefully evaluated through several key performance metrics: accuracy, precision, recall, and F1 score [16,28]. These metrics each provide unique insights into the model's effectiveness and suitability for the literature review process by examining the relationships between true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Accuracy measures the proportion of true results (both true positives and true negatives) among the total number of cases examined. In the context of our study, high accuracy indicates that our model correctly identifies abstracts related to FUS therapies, as well as those that are unrelated, showcasing the model's overall effectiveness in filtering the relevant literature from the vast array of publications [29].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision (or positive predictive value) measures the proportion of true positive results in all positive predictions made by the model. It reflects the model's ability to return relevant abstracts while minimizing false alarms—abstracts incorrectly classified as relevant. High precision is crucial for ensuring that the literature review process is not only efficient but also effective by focusing on truly relevant studies [29].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall (or sensitivity) measures the proportion of true positive results among all actual positives. This metric assesses the model's ability to identify all relevant abstracts from the dataset. In our study, a high recall value is indicative of the model's capacity to capture a

comprehensive range of studies pertinent to FUS, ensuring that no significant research is overlooked during the review [29].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1 Score is the harmonic mean of precision and recall, providing a single metric to balance the trade-off between the two. Since precision and recall are often inversely related, the F1 score serves as an indicator of the model's balanced performance. An optimal literature review model would achieve a high F1 score, demonstrating both relevance and comprehensiveness in its classification of FUS literature [29].

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In the evaluation of our text classification models, these metrics collectively inform the selection of the most suitable model for integrating into the literature review process. The chosen model should maximize efficiency in identifying relevant studies, thus facilitating a streamlined and effective literature review in the rapidly evolving field of FUS therapies.

Given the prevalence of non-FUS articles within the collection of literature that needs to be reviewed, our model must prioritize minimizing the number of false negatives over false positives. This approach ensures that all FUS-related articles are identified and classified accurately. While the model may classify some non-FUS articles as relevant to FUS, this is a more manageable scenario for the reviewers. It is considerably more convenient for the team to exclude a few misplaced non-FUS articles from their selection than to sift through an extensive number of articles to find missed FUS-related studies. Therefore, the emphasis on reducing false negatives aligns with our objective to enhance the efficiency and effectiveness of the literature review process, ensuring that no FUS-related article is overlooked. For this reason, recall will be the primary focus of our evaluation metrics, followed by accuracy.

### 3. Results

We used a web scraper to retrieve scientific articles related to FUS from PubMed based on the keyword search parameters listed previously. After a manual filtration and validation, we used those articles to train various machine learning and deep learning models, and compared the results to discover which model had the best metrics for classifying the FUS-related literature.

#### 3.1. FusBERT Model–Fine-Tuned Bio-ClinicalBERT Model

Overall, the deep learning models outperformed traditional machine learning methods used in this study. The deep learning models achieved higher accuracy, recall, and F1 scores. However, the traditional machine learning methods achieved higher precision scores. These metrics are based on model performance on a test dataset of 359 abstracts (Table 4).

**Table 4.** Performance comparison between different models.

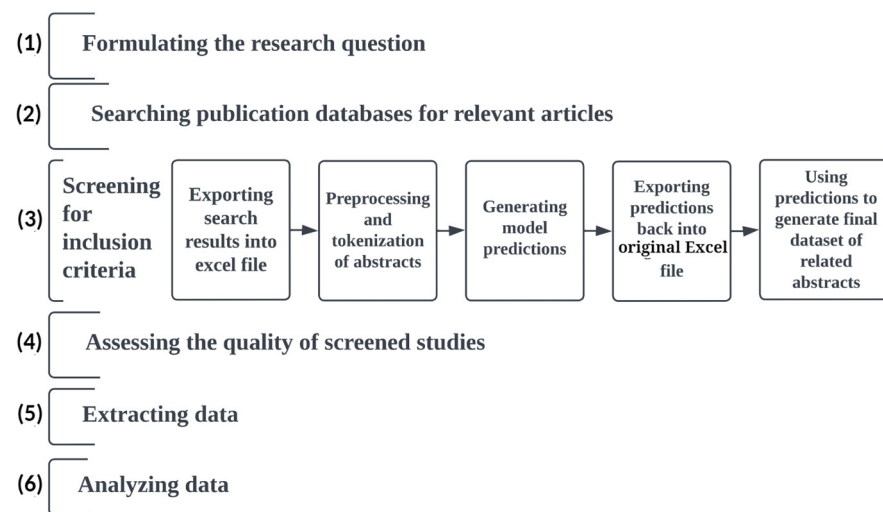
Model	Accuracy	Precision	Recall	F1
Naive Bayes	0.87	0.89	0.87	0.87
SVM	0.89	0.90	0.89	0.89
Logistic Regression	0.90	0.91	0.90	0.90
TinyBERT	0.88	0.82	0.98	0.89
SciBERT	0.89	0.89	0.99	0.90
DistilBERT	0.90	0.84	0.99	0.91
FusBERT	0.91	0.85	0.99	0.91



Based on our aim to prioritize high recall over accuracy, precision, and F1, while also ensuring competitive performance across all metrics, we concluded that the fine-tuned FusBERT model, which leveraged the pre-trained Bio-ClinicalBERT architecture, yielded the most favorable results. The best-performing fine-tuned FusBERT model exhibited the following performance metrics on our test data: accuracy 0.91, precision 0.85, recall 0.99, and F1 0.91 (refer to Supplementary Materials for more details).

### 3.2. Development of ML-Assisted Literature Review Workflow

As mentioned previously, the literature review process consists of six distinct steps. Step 3, which involves screening for inclusion criteria, emerges as a prime candidate for leveraging machine learning techniques to enhance efficiency (Figure 2).



**Figure 2.** Integration of ML in the literature review process.

The integration of ML into the relevant publication screening step of the entire process can be further delineated into five stages. Upon conducting searches in publication databases to identify relevant articles, researchers first export the search results into an Excel file. Subsequently, the abstracts undergo preprocessing and tokenization to prepare them as inputs for our model. The FusBERT model receives these abstracts as inputs, generates predictions, and exports them back into the original Excel file. These predictions are then utilized to compile a final dataset comprising pertinent abstracts. This curated dataset is subsequently subject to quality assessment.

### 3.3. Example Use Case of ML-Assisted Literature Review Workflow

In this section, we will run through what these ML-assisted steps within the literature review process may look like in practice with an example use case for this type of workflow. A physician may be interested in keeping up to date on the state of science within the field of focused ultrasound. Each month, they search publication databases for new articles about focused ultrasound research. They begin by inputting their search terms, such as “focused ultrasound”, within the search field of their desired database and include articles published within the previous month. Although the search produces many relevant articles, yet some of them are about diagnostic ultrasound rather than focused ultrasound research. The physician does not have time to sift through all these articles for relevancy, so they decide to use FusBERT to further identify article relevancy.

The physician first exports the search results into an Excel file that may look something like Figure 3 below.

PubMed ID	Article Title	Author List	Publication Type	Abstract	Link
34921489	Treatment of superficial benign vascular tumors by high intensity focused ultrasound: Observations in two illustrative cases	Jacek Calik, Tomasz	Journal Article	Background: Existing therapies. Methods: Two volunteer subjects. Results: During treatment, blood flow increased. Conclusion: High intensity focused ultrasound is a promising treatment for vascular tumors. Keywords: HIFU; cherry angioma	<a href="https://pubmed.ncbi.nlm.nih.gov/34921489/">https://pubmed.ncbi.nlm.nih.gov/34921489/</a>
36444908	Low-Intensity Focused Ultrasound Ameliorates Ischemic Heart Failure Related to the Cholinergic Anti-Inflammatory Pathway	Shenrong Zhong, Qianyun Cai, Longhe Zhang	Journal Article	Objectives: This study aims to evaluate the effects of low-intensity focused ultrasound (LIFU) on ischemic heart failure (IHF) related to the cholinergic anti-inflammatory pathway. Methods: Sprague-Dawley rats with IHF were treated with LIFU. Results: LIFU reduced heart failure severity. Conclusions: LIFU ameliorates IHF via the cholinergic anti-inflammatory pathway. Keywords: AchRα7; HMGB1	<a href="https://pubmed.ncbi.nlm.nih.gov/36444908/">https://pubmed.ncbi.nlm.nih.gov/36444908/</a>
36541473	Autonomic modulation by low-intensity focused ultrasound stimulation of the vagus nerve	Ning Ji, Yuanheng Li, Longhe Zhang	Journal Article	Objective: Our previous study showed that low-intensity focused ultrasound (LIFU) stimulation of the vagus nerve (VNS) can modulate autonomic function. Methods: We investigated the effects of LIFU-VNS on autonomic modulation. Results: LIFU-VNS modulated autonomic function. Conclusions: LIFU-VNS is a potential non-invasive method for autonomic modulation. Keywords: autonomic modulation; LIFU	<a href="https://pubmed.ncbi.nlm.nih.gov/36541473/">https://pubmed.ncbi.nlm.nih.gov/36541473/</a>
35396078	American Association of Clinical Endocrinology Disease State Clinical Review: The Clinical Utility of Minimally Invasive Interventional Procedures in the Management of Benign and Malignant Thyroid Lesions	Sina Jasim, Kepal N. Jayaraman	Journal Article	Objective: The objective of this review is to evaluate the clinical utility of minimally invasive interventional procedures in the management of benign and malignant thyroid lesions. Methods: A literature search was conducted. Results: Minimally invasive procedures are effective for thyroid lesions. Conclusion: Surgery and radiofrequency ablation are common treatments. Keywords: ethanol; radiofrequency ablation	<a href="https://pubmed.ncbi.nlm.nih.gov/35396078/">https://pubmed.ncbi.nlm.nih.gov/35396078/</a>
35326687	Applications of Focused Ultrasound in the Treatment of Genitourinary Cancers	John Panzone, Timol	Journal Article	Traditional cancer treatments have been associated with substantial morbidity for patients. Focused ultrasound offers a novel modality for the treatment of various forms of genitourinary cancers.	<a href="https://pubmed.ncbi.nlm.nih.gov/35326687/">https://pubmed.ncbi.nlm.nih.gov/35326687/</a>
25753877	Application of ultrasound technology in the diagnosis and treatment of digestive tract diseases	W-X Fu, Q Wang, Y-C Wang	Journal Article	Ultrasound is commonly used in the diagnosis and treatment of digestive tract diseases. This article reviews the current status of ultrasound technology in the diagnosis and treatment of digestive tract diseases.	<a href="https://pubmed.ncbi.nlm.nih.gov/25753877/">https://pubmed.ncbi.nlm.nih.gov/25753877/</a>
37682185	The ambiguous "internal carotid artery"-Ultrasound diagnosis of congenital absence of the internal carotid artery: A case report and review of the literature	Jing Ning, Gang Zhou	Journal Article	Rationale: Congenital absence of the internal carotid artery (ICA) is a rare anomaly. Patient concerns: A 61-year-old patient was diagnosed with ICA absence. Diagnoses: The patient was diagnosed with ICA absence. Interventions: After relevant literature review, the patient received surgery. Outcomes: The patient recovered well. Lessons: Congenital absence of the ICA is a rare anomaly that should be considered in the differential diagnosis of carotid artery anomalies.	<a href="https://pubmed.ncbi.nlm.nih.gov/37682185/">https://pubmed.ncbi.nlm.nih.gov/37682185/</a>
22608626	Bipolar disorder	Daniel J Smith, Elizabeth	Journal Article	Bipolar disorder is a serious mental health condition. This article reviews the current status of bipolar disorder diagnosis and treatment.	<a href="https://pubmed.ncbi.nlm.nih.gov/22608626/">https://pubmed.ncbi.nlm.nih.gov/22608626/</a>
28748549	Ultrasound Curricula in Undergraduate Medical Education: A Scoping Review	Usman Tarique, Bran	Journal Article	The clinical applications of ultrasound in medical education are increasing. This scoping review evaluates the current status of ultrasound curricula in undergraduate medical education.	<a href="https://pubmed.ncbi.nlm.nih.gov/28748549/">https://pubmed.ncbi.nlm.nih.gov/28748549/</a>
3519082	Ultrasound instrumentation	R J Bartrum Jr	Journal Article	This article begins by reviewing the history of ultrasound instrumentation and its applications in medicine.	<a href="https://pubmed.ncbi.nlm.nih.gov/3519082/">https://pubmed.ncbi.nlm.nih.gov/3519082/</a>

Figure 3. Exported Excel file from PubMed initial search results.

They then use this Excel file as data input for the Python script that will run the FusBERT algorithm on the 'Abstract' column. This script automatically formats the data via the preprocessing and tokenization of the abstracts and generates model predictions. A new column is created in the Excel file that identifies which articles are related to focused ultrasound (coded as 1) and which ones are not (coded as 0), which can be seen in Figure 4. The physician can then sort the dataset by the predictions column, to then continue with the rest of their literature review, assessing only relevant articles for full-text article analysis and knowledge synthesis.

PubMed ID	Article Title	Author List	Publication Type	Abstract	Link	Prediction	Confidence	Logits
34921489	Treatment of superficial benign vascular tumors by high intensity focused ultrasound: Observations in two illustrative cases	Jacek Calik, Tomasz	Journal Article	Background: Existing therapies. Methods: Two volunteer subjects. Results: During treatment, blood flow increased. Conclusion: High intensity focused ultrasound is a promising treatment for vascular tumors. Keywords: HIFU; cherry angioma	<a href="https://pubmed.ncbi.nlm.nih.gov/34921489/">https://pubmed.ncbi.nlm.nih.gov/34921489/</a>	1	97.51520753	[[-1.6755231618881226, 1.994294285774231]]
36444908	Low-Intensity Focused Ultrasound Ameliorates Ischemic Heart Failure Related to the Cholinergic Anti-Inflammatory Pathway	Shenrong Zhong, Qianyun Cai, Longhe Zhang	Journal Article	Objectives: This study aims to evaluate the effects of low-intensity focused ultrasound (LIFU) on ischemic heart failure (IHF) related to the cholinergic anti-inflammatory pathway. Methods: Sprague-Dawley rats with IHF were treated with LIFU. Results: LIFU reduced heart failure severity. Conclusions: LIFU ameliorates IHF via the cholinergic anti-inflammatory pathway. Keywords: AchRα7; HMGB1	<a href="https://pubmed.ncbi.nlm.nih.gov/36444908/">https://pubmed.ncbi.nlm.nih.gov/36444908/</a>	1	97.46402502	[[-1.6663931608200073, 1.9825135469436646]]
36541473	Autonomic modulation by low-intensity focused ultrasound stimulation of the vagus nerve	Ning Ji, Yuanheng Li, Longhe Zhang	Journal Article	Objective: Our previous study showed that low-intensity focused ultrasound (LIFU) stimulation of the vagus nerve (VNS) can modulate autonomic function. Methods: We investigated the effects of LIFU-VNS on autonomic modulation. Results: LIFU-VNS modulated autonomic function. Conclusions: LIFU-VNS is a potential non-invasive method for autonomic modulation. Keywords: autonomic modulation; LIFU	<a href="https://pubmed.ncbi.nlm.nih.gov/36541473/">https://pubmed.ncbi.nlm.nih.gov/36541473/</a>	1	97.73917198	[[-1.7370295524597168, 2.0295417308807373]]
35396078	American Association of Clinical Endocrinology Disease State Clinical Review: The Clinical Utility of Minimally Invasive Interventional Procedures in the Management of Benign and Malignant Thyroid Lesions	Sina Jasim, Kepal N. Jayaraman	Journal Article	Objective: The objective of this review is to evaluate the clinical utility of minimally invasive interventional procedures in the management of benign and malignant thyroid lesions. Methods: A literature search was conducted. Results: Minimally invasive procedures are effective for thyroid lesions. Conclusion: Surgery and radiofrequency ablation are common treatments. Keywords: ethanol; radiofrequency ablation	<a href="https://pubmed.ncbi.nlm.nih.gov/35396078/">https://pubmed.ncbi.nlm.nih.gov/35396078/</a>	1	97.89092541	[[-1.7765820026397705, 2.061021089553833]]
35326687	Applications of Focused Ultrasound in the Treatment of Genitourinary Cancers	John Panzone, Timol	Journal Article	Traditional cancer treatments have been associated with substantial morbidity for patients. Focused ultrasound offers a novel modality for the treatment of various forms of genitourinary cancers.	<a href="https://pubmed.ncbi.nlm.nih.gov/35326687/">https://pubmed.ncbi.nlm.nih.gov/35326687/</a>	1	97.96125293	[[-1.7842141389846802, 2.0880227088928223]]
25753877	Application of ultrasound technology in the diagnosis and treatment of digestive tract diseases	W-X Fu, Q Wang, Y-C Wang	Journal Article	Ultrasound is commonly used in the diagnosis and treatment of digestive tract diseases. This article reviews the current status of ultrasound technology in the diagnosis and treatment of digestive tract diseases.	<a href="https://pubmed.ncbi.nlm.nih.gov/25753877/">https://pubmed.ncbi.nlm.nih.gov/25753877/</a>	0	99.07332659	[[2.402134895324707, -2.2698824405670166]]
37682185	The ambiguous "internal carotid artery"-Ultrasound diagnosis of congenital absence of the internal carotid artery: A case report and review of the literature	Jing Ning, Gang Zhou	Journal Article	Rationale: Congenital absence of the internal carotid artery (ICA) is a rare anomaly. Patient concerns: A 61-year-old patient was diagnosed with ICA absence. Diagnoses: The patient was diagnosed with ICA absence. Interventions: After relevant literature review, the patient received surgery. Outcomes: The patient recovered well. Lessons: Congenital absence of the ICA is a rare anomaly that should be considered in the differential diagnosis of carotid artery anomalies.	<a href="https://pubmed.ncbi.nlm.nih.gov/37682185/">https://pubmed.ncbi.nlm.nih.gov/37682185/</a>	0	99.40914512	[[2.568174362182617, -2.557249069213867]]
22608626	Bipolar disorder	Daniel J Smith, Elizabeth	Journal Article	Bipolar disorder is a serious mental health condition. This article reviews the current status of bipolar disorder diagnosis and treatment.	<a href="https://pubmed.ncbi.nlm.nih.gov/22608626/">https://pubmed.ncbi.nlm.nih.gov/22608626/</a>	0	99.82409477	[[3.3475353717803955, -2.9937052726745605]]
28748549	Ultrasound Curricula in Undergraduate Medical Education: A Scoping Review	Usman Tarique, Bran	Journal Article	The clinical applications of ultrasound in medical education are increasing. This scoping review evaluates the current status of ultrasound curricula in undergraduate medical education.	<a href="https://pubmed.ncbi.nlm.nih.gov/28748549/">https://pubmed.ncbi.nlm.nih.gov/28748549/</a>	0	99.76135492	[[3.210164785385132, -2.8253657817840576]]
3519082	Ultrasound instrumentation	R J Bartrum Jr	Journal Article	This article begins by reviewing the history of ultrasound instrumentation and its applications in medicine.	<a href="https://pubmed.ncbi.nlm.nih.gov/3519082/">https://pubmed.ncbi.nlm.nih.gov/3519082/</a>	0	99.5505929	[[2.7637088298797607, -2.636791467666626]]

Figure 4. New columns added once FusBERT has processed abstracts.

An abstract that our model correctly predicts as being related to FUS is as follows: *“Traditional cancer treatments have been associated with substantial morbidity for patients. Focused ultrasound offers a novel modality for the treatment of various forms of*

*cancer which may offer effective oncological control and low morbidity. We performed a review of PubMed articles assessing the current applications of focused ultrasound in the treatment of genitourinary cancers, including prostate, kidney, bladder, penile, and testicular cancer. Current research indicates that high-intensity focused ultrasound (HIFU) focal therapy offers effective short-term oncologic control of localized prostate and kidney cancer with lower associated morbidity than radical surgery. In addition, studies in mice have demonstrated that focused ultrasound treatment increases the accuracy of chemotherapeutic drug delivery, the efficacy of drug uptake, and cytotoxic effects within targeted cancer cells. Ultrasound-based therapy shows promise for the treatment of genitourinary cancers. Further research should continue to investigate focused ultrasound as an alternative cancer treatment option or as a complement to increase the efficacy of conventional treatments such as chemotherapy and radiotherapy. Keywords: cancer; review; treatment; ultrasound [30]."*

The types of language seen within this article can be compared to an article related to ultrasound but not focused ultrasound. Our model correctly classifies the following abstract as being not related to focused ultrasound:

*"Ultrasound is commonly used in clinical examination, which is economic, non-invasive and convenient. Ultrasound can be used for the examination of solid organs and hollow organs. Due to the presence of air, routine ultrasound examination of the digestive tract is not very appropriate. Because of the development of endosonography and its related technology, diagnosis of gastrointestinal diseases have been improved which is valuable in clinic. This review focused on the application of ultrasound technology in the diagnosis of digestive tract diseases [31]."*

The integration of FusBERT within the physician's review greatly reduces the time it takes to screen abstracts for subject-specific relevancy. For example, within seconds, FusBERT can classify a month's worth of new medical abstracts, while it may take a human 1 min to classify only one abstract. Although a human reviewer may be able to more accurately classify abstracts, it takes them a lot more time. This use case highlights how leveraging deep learning methods such as fine-tuned BERT models, for literature review has the potential to significantly boost efficiency, especially with large datasets.

#### **4. Discussion and Conclusions**

Ultimately, our FusBERT model satisfied our condition to optimize the recall value, demonstrating its efficacy in minimizing the number of false negatives over false positives, while maintaining relatively high levels of accuracy, precision, and F1 score. This best-performing model was then successfully integrated into the conventional literature review process. We then presented examples of correct classifications of scientific literature by FusBERT despite cases where the topics are very closely related.

Our findings demonstrate that BERT models can efficiently automate the classification of the scientific literature with high accuracy in the fields of and relating to FUS. Whether such findings can be expanded to other domains necessitates further studies and experimentation. However, the adaptability of BERT models, characterized by their understanding of context and nuance in text, showcase their position as valuable and flexible tools for researchers. Thus, this project also highlights the significant potential for integrating machine learning techniques, particularly BERT models, into the literature review process across various fields beyond FUS therapies.

Looking ahead, the potential for expanding the use of BERT models in literature review processes is vast. As various fields related to biomedical and health research grow, the amount of potential training data for BERT models also increases. As long as a wealth of data already exists in a given area, we can retrieve those related articles by adjusting the keyword parameters on the scraper, giving BERT models much promise to adapt and thrive. Given that our study was limited to the binary classification of abstracts, a promising direction for future research could be the development of BERT models capable

of multi-class classification. This would allow for the incorporation of multiple inclusion and exclusion criteria in the article screening process [32]. This advancement would enable the models to categorize the literature into multiple predefined categories, further refining the review process. This capability would significantly enhance the precision of literature reviews, making it easier for researchers to locate studies relevant with greater granularity [32].

However, the application of BERT models, particularly for multi-class classification, comes with certain challenges. A notable limitation of using BERT for this task is the requirement for large amounts of data to train these models effectively. BERT models rely on extensive data to understand and interpret the nuances of language accurately. As the complexity of classification increases from binary to multi-class, the demand for more diverse and expansive datasets grows accordingly. These limitations do not necessarily affect the generalizability of our findings but rather make it difficult for individual researchers to obtain the amount of data required to train a model for their specific research purposes.

Furthermore, the computational resources required to train and run BERT models, especially for large datasets and complex classification tasks, present another challenge. In this study, all models were trained using an academic research center's server which contains larger computational power than the average person's personal computer. This specific challenge makes it difficult for physicians or researchers who do not have access to additional higher-powered servers to leverage the full potential of using transformers for their specific research needs, ultimately limiting the widespread adoption of these tools within specialized fields of research.

Finally, given that the precision requirements for medical care are very high, one limitation of this approach is the question of whether the accuracy ranges of 0.85–0.99 are sufficient. While this may be good for carrying out academic data reviews, it may not be high enough to conduct a data search that relates to clinical care.

In conclusion, the incorporation of machine learning, and BERT models in particular, into literature review processes is promising for enhancing research efficiency. While limitations such as data requirements, computational demands, and a need for near-perfect accuracy are still present, the potential benefits in terms of time efficiency and the ability to handle large volumes of literature are compelling. As the ability to collect and process data advances, the potential of machine learning in literature review processes also grows.

**Supplementary Materials:** For further information on the code and model, please see our GitHub page at [https://github.com/rteb8/MSDS\\_FUSCapstone23/](https://github.com/rteb8/MSDS_FUSCapstone23/).

**Author Contributions:** Conceptualization, R.A.H., R.M.B., T.D.M. and R.K.P.; methodology, R.K.P., S.H.F., S.H.J. and R.T.E.M.; software, R.K.P., S.H.F., S.H.J., A.S. and R.T.E.M.; validation, R.A.H., R.M.B., T.D.M., R.K.P., S.H.F. and S.H.J.; formal analysis, R.K.P., S.H.F., S.H.J., A.S. and R.T.E.M.; investigation, R.K.P., S.H.F. and S.H.J.; resources, R.A.H.; data curation, R.K.P., S.H.J. and R.T.E.M.; writing—original draft preparation, R.K.P., S.H.J., A.S. and R.T.E.M.; writing—review and editing, S.H.F. and R.A.H.; supervision, R.A.H., R.M.B. and T.D.M.; project administration, R.A.H., R.M.B. and T.D.M.; funding acquisition, R.A.H., R.M.B. and T.D.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Focused Ultrasound Foundation, Charlottesville, Virginia.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article/supplementary materials, and further inquiries can be directed to the corresponding authors.

**Acknowledgments:** We extend our gratitude to our sponsor, the Focused Ultrasound Foundation, for their support throughout the duration of our project. Their contribution has been invaluable in driving our research forward. We also wish to thank the School of Data Science Capstone Teaching team for their guidance and constructive feedback. Special thanks to Heman Shekari whose mentorship helped shape our research direction.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Izadifar, Z.; Izadifar, Z.; Chapman, D.; Babyn, P. An Introduction to High Intensity Focused Ultrasound: Systematic Review on Principles, Devices, and Clinical Applications. *J. Clin. Med.* **2020**, *9*, 460. [CrossRef] [PubMed]
2. Bachu, V.S.; Kedda, J.; Suk, I.; Green, J.J.; Tyler, B. High-Intensity Focused Ultrasound: A Review of Mechanisms and Clinical Applications. *Ann. Biomed. Eng.* **2021**, *49*, 1975–1991. [CrossRef] [PubMed]
3. Dogra, V.; Verma, S.; Kavita; Chatterjee, P.; Shafi, J.; Choi, J.; Ijaz, M.F. A Complete Process of Text Classification System Using State-of-the-Art NLP Models. *Comput. Intell. Neurosci.* **2022**, *2022*, 1883698. [CrossRef] [PubMed]
4. Majumder, N.; Poria, S.; Gelbukh, A.; Cambria, E. Deep Learning-Based Document Modeling for Personality Detection from Text. *IEEE Intell. Syst.* **2017**, *32*, 74–79. [CrossRef]
5. Cambria, E.; Das, D.; Bandyopadhyay, S.; Feraco, A. (Eds.) *Affective Computing and Sentiment Analysis*. In *A Practical Guide to Sentiment Analysis*; Springer International Publishing: Cham, Switzerland, 2017; pp. 1–10. [CrossRef]
6. Paré, G.; Kitsiou, S. Chapter 9 Methods for Literature Reviews. In *Handbook of eHealth Evaluation: An Evidence-Based Approach [Internet]*; University of Victoria: Victoria, BC, Canada, 2017. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK481583/> (accessed on 26 July 2024).
7. Masoumi, S.; Amirkhani, H.; Sadeghian, N.; Shahraz, S. Natural language processing (NLP) to facilitate abstract review in medical research: The application of BioBERT to exploring the 20-year use of NLP in medical research. *Syst. Rev.* **2024**, *13*, 107. [CrossRef] [PubMed]
8. Li, F.; Jin, Y.; Liu, W.; Rawat, B.P.S.; Cai, P.; Yu, H. Fine-tuning bidirectional encoder representations from transformers (BERT)-based models on large-scale electronic health record notes: An empirical study. *JMIR Med. Inform.* **2019**, *7*, e14830. Available online: <https://medinform.jmir.org/2019/3/e14830> (accessed on 26 August 2024). [CrossRef] [PubMed]
9. Rasmy, L.; Xiang, Y.; Xie, Z.; Tao, C.; Zhi, D.M.B. Pretrained Contextualized Embeddings on Large-Scale Structured Electronic Health Records for Disease Prediction | NPJ Digital Medicine. *NPJ Digit. Med.* **2021**, *4*, 1–13. Available online: <https://www.nature.com/articles/s41746-021-00455-y> (accessed on 26 August 2024). [CrossRef] [PubMed]
10. Li, Q.; Peng, H.; Li, J.; Xia, C.; Yang, R.; Sun, L.; Yu, P.S.; He, L. A Survey on Text Classification: From Shallow to Deep Learning. *arXiv* **2021**, arXiv:2008.00364. [CrossRef]
11. Silva, M.D. Preprocessing Steps for Natural Language Processing (NLP): A Beginner's Guide. Medium. Available online: <https://medium.com/@maleeshadesilva21/preprocessing-steps-for-natural-language-processing-nlp-a-beginners-guide-d6d9bf7689c9> (accessed on 26 August 2024).
12. Liang, H.; Sun, X.; Sun, Y.; Gao, Y. Text feature extraction based on deep learning: A review. *EURASIP J. Wirel. Commun. Netw.* **2017**, *2017*, 211. [CrossRef] [PubMed]
13. Ishankulov, T.; Danilov, G.; Kotik, K.; Orlov, Y.; Shifrin, M.; Potapov, A. The Classification of Scientific Abstracts Using Text Statistical Features. In *MEDINFO 2021: One World, One Health—Global Partnership for Digital Innovation*; IOS Press: Amsterdam, The Netherlands, 2022; pp. 263–267. [CrossRef]
14. Wan, Z. Text Classification: A Perspective of Deep Learning Methods. *arXiv* **2023**, arXiv:2309.13761. [CrossRef]
15. List of keywords. In Proceedings of the IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings, Winnipeg, MB, Canada, 15–16 May 1995.
16. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805. [CrossRef]
17. Paliwal, S.; Bharti, V.; Mishra, A.K. Chapter 9—Changing the outlook of security and privacy with approaches to deep learning. In *Trends in Deep Learning Methodologies*; Piuri, V., Raj, S., Genovese, A., Srivastava, R., Eds.; Hybrid Computational Intelligence for Pattern Analysis; Academic Press: New York, NY, USA, 2021; pp. 207–226. [CrossRef]
18. Publications. Focused Ultrasound Foundation. Available online: <https://www.fusfoundation.org/publications/> (accessed on 26 July 2024).
19. BERT Embeddings. Available online: <https://tinkerd.net/blog/machine-learning/bert-embeddings/> (accessed on 26 July 2024).
20. An Explanatory Guide to BERT Tokenizer—Analytics Vidhya. Available online: <https://www.analyticsvidhya.com/blog/2021/09/an-explanatory-guide-to-bert-tokenizer/> (accessed on 26 July 2024).
21. Short Text Classification with Machine Learning in the Social Sciences: The Case of Climate Change on Twitter—PMC. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10540966/> (accessed on 26 August 2024).
22. Zhang, W.; Gao, F. An Improvement to Naive Bayes for Text Classification. *Procedia Eng.* **2011**, *15*, 2160–2164. [CrossRef]
23. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
24. Jiao, X.; Yin, Y.; Shang, L.; Jiang, X.; Chen, X.; Li, L.; Wang, F.; Liu, Q. TinyBERT: Distilling BERT for Natural Language Understanding. *arXiv* **2020**, arXiv:1909.10351. [CrossRef]
25. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text. *arXiv* **2019**, arXiv:1903.10676. [CrossRef]
26. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv* **2020**, arXiv:1910.01108. [CrossRef]
27. Alsentzer, E. EmilyAlsentzer/clinicalBERT. Python. 25 July 2024. Available online: <https://github.com/EmilyAlsentzer/clinicalBERT> (accessed on 26 July 2024).

28. Kamsetty, A. Hyperparameter Optimization for Transformers: A Guide. Distributed Computing with Ray. Available online: <https://medium.com/distributed-computing-with-ray/hyperparameter-optimization-for-transformers-a-guide-c4e32c6c989b> (accessed on 26 July 2024).
29. Blagec, K.; Dorffner, G.; Moradi, M.; Samwald, M. A critical analysis of metrics used for measuring progress in artificial intelligence. *arXiv* **2021**, arXiv:2008.02577. [[CrossRef](#)]
30. Panzone, J.; Byler, T.; Bratslavsky, G.; Goldberg, H. Applications of Focused Ultrasound in the Treatment of Genitourinary Cancers. *Cancers* **2022**, *14*, 1536. [[CrossRef](#)] [[PubMed](#)]
31. Yu, X.; Yang, Y.; Li, J. Application of ultrasound in the diagnosis of gastrointestinal tumors. *Eur. J. Inflamm.* **2020**, *18*, 2058739220961194. [[CrossRef](#)]
32. Khadhraoui, M.; Bellaaj, H.; Ammar, M.B.; Hamam, H.; Jmaiel, M. Survey of BERT-Base Models for Scientific Text Classification: COVID-19 Case Study. *Appl. Sci.* **2022**, *12*, 2891. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.