DETECTING HEALTH MISINFORMATION ONLINE USING DEEP LEARNING METHODS

Dione Morales

Bachelor of Engineering
Computer Engineering Stream

School of Engineering
Macquarie University

November XX, 2018

Supervisor: Associate Professor Adam Dunn

ACKNOWLEDGMENTS

I would like to acknowledge ...

STATEMENT OF CANDIDATE

I, Dione Morales, declare that this report, submitted as part of the requirement for the award of Bachelor of Engineering in the School of Engineering, Macquarie University, is entirely my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualification or assessment an any academic institution.

Student's Name: Dione Morales

Student's Signature:

Date:

ABSTRACT

This is where you write your abstract ...

Contents

Acknowledgments **iii**

Abstract **vii**

Table of Contents **ix**

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**
1.1 Project Overview 1
1.2 Motivation 2
1.3 Aims 2

2 Background and Related Work **5**
2.1 Assessing the Quality and Credibility of Health Information 5
2.1.1 Checklists and Criteria 5
2.1.2 Studies Applying Checklists to Health Information Online 6
2.2 Document Classification Methods 7
2.2.1 Representing Text as Features 7
2.2.2 Machine Learning Methods for Document Classification 8
2.2.3 Deep Learning Methods for Document Classification 10
2.2.4 Classifying Documents with Limited Labelled Examples 13
2.3 Discussion 13
2.3.1 Scaling the Assessment of Quality and Credibility of Health Information 13
2.3.2 Using Deep Learning Methods to Automate Quality and Credibility Assessments 13

3 Proposed Approach **15**
3.1 Rationale 15
3.2 Credibility Criteria 15
3.3 Study Data 15

3.4 System Model 15
3.5 Experiments 15
3.6 Outcome Measures 15

4 Conclusions and Future Work **17**
4.1 Conclusions 17
4.2 Future Work 17

5 Abbreviations **19**

A name of appendix A **21**
A.1 Overview 21
A.2 Name of this section 21

B name of appendix B **23**
B.1 Overview 23
B.2 Name of this section 23

Bibliography **23**

List of Figures

2.1 Determining the optimal hyperplane [5] 9
2.2 (Left) Non-linearly distributed data of two classes in a 2D representation space.
(Right) Linearly separable data of the same classes in a 3D representation space after the application of a kernel [4]. 10
2.3 Layout of a simple RNN [3] 12

List of Tables

2.1 Commonly used classification datasets 11

Chapter 1

Introduction

With the popularity and ubiquity of social media platforms in today's society, the amount and rate at which information propagates online greatly outnumbers the resources available that can evaluate the quality and credibility of the information that gets shared. This is becoming an increasingly growing issue due to the clickbait model that is commonly adopted by social media platforms and the lack of rigour surrounding the publishing of content online [33], causing an increase in the number of articles that contain misinformed content [34] [35]. The lack of resources that attempt to minimise the spread of misinformation can be attributed to the expert-level knowledge required to determine the credibility of information since a variety of factors such as the actual information, information sources, conflicts of interest and writing style of the article must be evaluated to be able to determine its quality and credibility. This issue is further exacerbated in specific domains such as for health-related content, as the spread of misinformation can have a detrimental effect on people and their communities that don't have access to these kinds of limited resources. Thus, by developing a method that is capable of automatically assessing the quality and credibility of an article - various methods that intervene between an article and a user can then be developed to minimise the spread and effects of misinformation as the reliance on expert-level knowledge is no longer present.

1.1 Project Overview

This section details the scope of the project and its associated outcomes outlining the various tasks that must be accomplished to successfully complete the project.

One of the key components required to minimise the propagation of misinformation online is to have the ability of automatically evaluating and quantifying the credibility of articles. However, traditional automated methods - such as machine learning-based techniques, still require the domain knowledge of experts to be able to develop the features required by the model. Thus, this project aims to investigate the performance of Deep Learning-based (DL) techniques in evaluating the credibility of information within domain-specific articles via the classification of set criteria that have deemed to be highly correlated with articles that have low credibility. Specifically, this project will focus on evaluating the credibility of online health articles related to vaccination due to the commonly misinformed and controversial views associated with its effects [10].

1.2 Motivation

Try to answer some of the following questions:

- Why is misinformation and the spread of low credibility information a problem?
- What is vaccine hesitancy, how many people believe that vaccines are harmful? Explain specific examples

The propagation of misinformed content can cause people within specific communities to adopt beliefs and practices that can have harmful effects to themselves and also to the people around them. In the context of health-related content specifically, these beliefs and practices can take form in the misuse or mistreatment of medicine or illnesses. A popular example where this issue has emerged is with the misinformed views by certain communities on the ramifications of vaccinating children.

The effects of vaccination and their relation to causing autism within children is a popular example where the belief of inaccurate facts has had a detrimental effect to specific communities and their surrounding environments. The measles outbreak in Minnesota which occurred in April 2017 is a specific case where the misinformed views on the effects of vaccination caused people within a Somali-American community located in the United States to forego the vaccination of their children causing an outbreak of measles, a disease which was declared to be eliminated from the United States in the year 2000 [16]. Within this specific instance, it was determined that the outbreak of measles was caused by a decline in the coverage of vaccination due to concerns of its effects in relation to the causation of autism. This can be seen in the decreasing trend of the percentage for 24 month-old children that received the measles-mumps-rubella vaccine which fell to a low of 40% in 2014 where it was at a 90% high 10 years prior.

1.3 Aims

With the primary objective of this project being the evaluation on the effectiveness of deep learning models in determining the credibility of online health-related articles. Due to the complexity of this project, a set of activities - divided into main goals and stretch goals, have been defined to ensure that the completion of this project remains feasible in the given time frame. The completion of all activities categorized as main goals will signal the realization of the primary objective and the completion of the project. Stretch goals are activities of interest that have been identified as non-essential to the completion of the primary objective but will be worked on after the completion of the project.

Main Goals

- Implement and evaluate the performance of machine learning methods for assessing the quality and credibility of vaccine-related text.
- Implement a deep neural network method to assess the quality and credibility of vaccine-related text and compare the performance with the previous approaches.

Stretch Goals

- Evaluate the effect of transfer learning methods on the training time and performance of the proposed deep neural network method.

Chapter 2

Background and Related Work

A literature review has been conducted to develop an understanding on the research that has been done in the assessment of the credibility of information, specifically in the context of information related to health and the limitations and capabilities of machine learning techniques and how it differs from deep learning-based methods for the task of document classification.

2.1 Assessing the Quality and Credibility of Health Information

To have the capability of automating the process of evaluating the quality or credibility of online information, a definition that outlines of what is required by an article to be considered as a credible source of information must be developed. While there has been a significant amount of research that has been done on the development of tools and frameworks that aim to assess the credibility of online health information, there is currently no standardized method or benchmark that is universally used. Tools and frameworks that have been identified to be applicable in the context of this project are: DISCERN [1], HealthNewsReview [2] and Quality Index for health-related Media Reports (QIMR) [38].

2.1.1 Checklists and Criteria

DISCERN

DISCERN [1] is a questionnaire designed to assess the reliability of a publication, it consists of 16 questions each with a Likert scale, ranging from 1 (no) to 5 (yes) and is divided into 3 sections. The first section (questions 1 - 8) investigate the reliability of the information and is comprised of questions such as "Are the aims clear?", "Is it balanced and unbiased?" and "Does it refer to areas of uncertainty?". The second section (questions 9 - 15) assesses the quality of information provided by the publication for treatment choices and is composed of questions such as "Does it describe the risks/benefits of each treatment?" and "Does it provide support for shared decision-making?". The final section (question 16) assesses the overall rating of the publication ("Based on the answers to all of the above questions, rate the overall quality of the publication as a source of information about treatment choices").

HealthNewsReview

HealthNewsReview [2] provides a set of 10 criteria designed to act as a framework for evaluating the credibility of health-related media. The criteria is based on the various elements that all health-related media should consist of, the criteria is composed of criterion such as "Does the story compare the new approach with existing alternatives?", "Does the story use independent sources and identify conflicts of interest?" and "Does the story appear to rely solely or largely on a news release?".

QIMR

QIMR [38] is a tool developed to monitor the quality of health research reports presented in the media. The tool considers 5 main factors that have been deemed to be correlated with the quality of research reports based on interviews with health journalists and researchers. The 5 main factors are: background information provided, sources of information used, manner in which results were analysed, context of the research and the validity of their methodology.

2.1.2 Studies Applying Checklists to Health Information Online

There has been an extensive amount of research conducted which aimed to assess the quality of content for various health-related domains using one of the aforementioned tools [12] [13] [20] [32]. A factor shared by these commonly performed studies however, is the use of experts to leverage the tool in assessing the quality and credibility of the information, indicating that commonly used tools such as DISCERN is designed to be used by domain experts in order to produce reliable and consistent results. This sentiment is shared by Batchelor et al. [9] who evaluated performance of the DISCERN tool when used by health professionals and patients. Due to the resource intensive nature of expertly performing this task, the number of articles evaluated, among the various studies examined have been limited ranging between 10 - 300 information sources. This highlights the issue that these tools are not designed to be capable of evaluating the quality and credibility of articles at the scale and speed required to match the pace of online activity and thus must be adapted for the automation of this process. While there has been work in adapting criteria to better encompass the factors that are correlated to the quality and credibility of information within a specific context, such as by Matsoukas et al. [24], who expanded the DISCERN tool to improve its capability in assessing the quality of online information, there has not been any published work found that aims to adapt these tools to provide the capability of autonomously assessing the quality and credibility of resources.

2.2 Document Classification Methods

The task of document classification is a heavily researched topic due to its wide number of applications in various domains. Formally, document classification is defined to be the task of finding a classifier $f : D \to L$ where $D = \{d_1, d_2, \ldots, d_n\}$ is a collection of documents and $L = \{l_1, l_2, \ldots, l_k\}$ is the set of possible labels that a document d can be classified as.

$$f : D \to L \text{ where } f(d) = l \tag{2.1}$$

Common applications of document classification algorithms are: organization and filtering of news articles, document retrieval, opinion mining, email classification and spam filtering [5].

Traditional machine learning-based approaches for document classification, such as Naive Bayes, Support Vector Machines (SVM) and Random Forests, require the manual extraction of features [5] [7] [22] [28] as they are incapable of performing feature learning [8]. These features are typically either hand-crafted and domain specific, requiring expert-level knowledge for the task domain [28] or are simple and general but prone to the loss of information (e.g. being unable to account for context) since these features, such as term frequency-inverse document frequency (TF-IDF) scores or bag of words (BoW) models, are unable to represent and account for the sequential nature of text.

2.2.1 Representing Text as Features

Due to the unstructured nature of natural text, it is common practice for natural language processing (NLP) tasks to transform the text into a structured representation that minimises the loss of information stored within the original text whilst allowing the model to learn and analyse the features associated with the classification task. However, despite traditional representation methods such as count-based approaches (e.g. TF-IDF) or probabilistic-based approaches (e.g. N-Grams) showing respectable performances in the classification of texts [39] [40], these representations are typically sparsely distributed and have a high dimensionality that scales with the vocabulary size, has difficulty in generalising over unfamiliar information or writing styles [29] as the texts are represented using simple mechanisms that are unable to represent high-level information such as context or the semantic relationship of words. Due to the high dimensionality and sparseness of these features, they are computationally intensive to calculate and can require a relatively larger amount of space for storage.

Word embeddings are a more sophisticated language model that overcome these limitations as they are created by learning word vectors through the optimisation for a different task e.g. the prediction of a word based on its context. Through this, the learned embeddings are able to store information such as the similarity of different contexts, syntactic and semantic information within the text and analogies. These embeddings are relatively more efficient and require less time to compute due to their smaller dimensionality. Whilst these embeddings are able to be created using shallow neural networks, deep learning-based models for NLP tasks utilise these embeddings which can also be at the character, phrase or sentence levels [37]. Word embeddings have produced state-of-the-art results in not only classification tasks [18] but also for a wide range of NLP tasks such as image annotation [36] and sentiment analysis [11], which used the popular word2vec embedding proposed by Mikolov et al. [26]

2.2.2 Machine Learning Methods for Document Classification

Naive Bayes Classifier

Naive Bayes is a probabilistic classifier that functions on an assumption in how the data (i.e. the words in a document) is generated. The assumption that these Bayesian classifiers are based on is that the distribution of different words within a corpus are independent from each other. Despite this assumption clearly being wrong when considering the distribution of words within a document (due to the sequential nature of text), Naive Bayes classifiers are still able to perform well in document classification tasks such as the filtering of spam emails [30].

Naive Bayes classifiers utilize Bayes' theorem, which attempts to find the probability of an event B occurring given some prior condition A i.e. $P(B|A)$. In the context of document classification, Naive Bayes classifiers classifies a document d by calculating the probabilities that the document belongs for all labels $l_i \in L$ and then selecting the highest probability [7]:

$$P(L = l_i|d) = \frac{P(l_i)P(d|l_i)}{P(d)} \tag{2.2}$$

According to Aggarwal et al. [5], there are two main types of Naive Bayes classifiers that are commonly used, they are the multivariate Bernoulli model and the multinomial model. The main discriminating factor between these two models is that the Bernoulli model does not take into account the frequency of words as it represents a document using a vector of binary features which signify the presence or absence of words, based on some vocabulary, for a given document. This is in contrast with the multinomial model, which accounts for the frequency of words as the document is represented using a BoW model. Deciding on which model to use for document classification largely depends on the size of the vocabulary as, according to McCallum et al. [25], the multinomial models almost always outperforms the Bernoulli models if the size of the vocabulary is large (> 1000) or even if the size has been optimally chosen for both models.

Support Vector Machines

SVMs are a type of supervised machine learning-based binary classifiers that are extensively used for document classification due to their capability of handling the high dimensional and sparse nature of the common techniques used to represent text documents [19]. SVMs are able to classify a document by using a hyperplane to separate the different classes into separate regions. The hyperplane used is determined by choosing the one that maximises the margin of separation (i.e. The Euclidean distance between the hyperplane and all data points in the representation space) between the two classes [5].

Consider the illustration shown in figure 2.1 that presents three two-dimensional hyperplanes, A, B and C, which separates the classes 'x' and 'o'. Visually, we can determine that A is the hyperplane that maximises the margin of separation, thus, A will be used as a decision rule to classify any new document based on its location with respect to A. Mathematically, determining which hyperplane to use as the decision rule is an optimisation problem that attempts to maximise the margin represented as shown in Equation 2.3:

$$\text{maximise: } \frac{1}{2}||\frac{1}{w}||^2 \tag{2.3}$$

This optimisation problem is often re-framed to the minimisation of Equation 2.4:

$$\text{minimise: } \frac{1}{2}||w||^2 \tag{2.4}$$

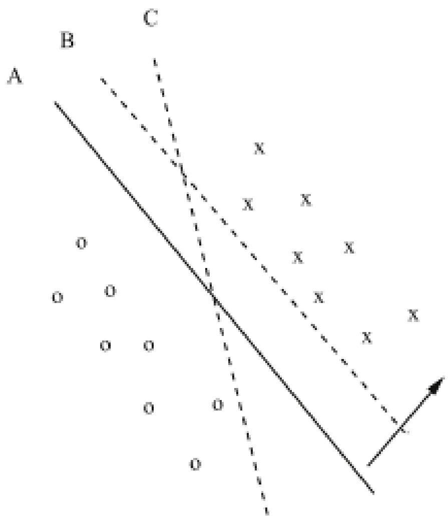

Figure 2.1: Determining the optimal hyperplane [5]

This technique of determining the hyperplane requires that the two classes are linearly separable. In situations where this isn't true, the kernel trick [6] is applied which is essentially a function that maps data in a particular representation space to another representation space of a different dimension. This change in dimensionality can allow the classes to become linearly separable and thus a hyperplane can then be constructed that separates each class as illustrated in Figure 2.2.

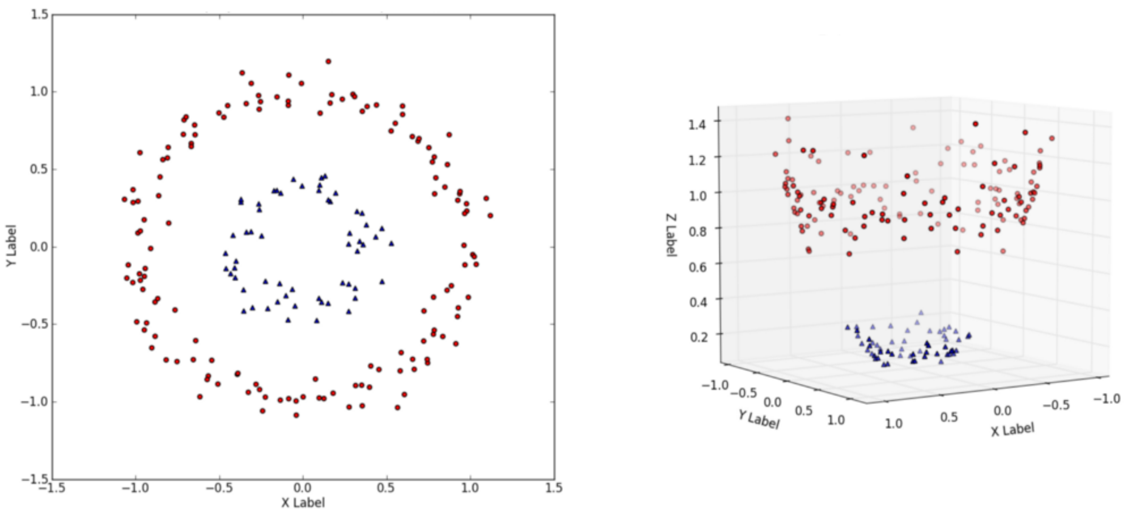

Figure 2.2: (Left) Non-linearly distributed data of two classes in a 2D representation space.
(Right) Linearly separable data of the same classes in a 3D representation space after the application of a kernel [4].

2.2.3 Deep Learning Methods for Document Classification

Deep learning models are a class of machine learning models that have the capability of automatically learning a hierarchical representation of data [8]. These hierarchical representations are constructed through the use of artificial neural networks, the main underlying mechanism of deep learning models. The most commonly used models for document classification are Recurrent Neural Networks (RNNs), Long Short-Term Memory Networks (LSTMs) - a variant of RNNs and Convolutional Neural Networks (CNNs).

Common datasets for text classification can be found under the 5 papers of interest doc.

For DL language models, a common one is wikitext103 but there might be more in the paper by Jeremy Howard

For the task of text classification, there are commonly used, open datasets for specific classification tasks that aim to provide a benchmark in which the performance of different implementations can be compared and to facilitate the data needs of deep learning models. Table 2.1 contains the commonly used datasets for various text classification tasks.

Recurrent Neural Networks

RNNs and LSTMs are commonly used for document classification, or NLP tasks in general, due to their ability to create language models that are able to capture the context

Dataset	Description	Type	Objective
SST	Movie reviews from Stanford Sentiment Treebank	Sentence	Sentiment
IMDB	Internet Movie Database	Document	Sentiment
MDSD	Product Reviews	Document	Sentiment
RN	Reuters Newswire topics classification	Document	Topics
QC	Question Classification	Sentence	Question Types

Table 2.1: Commonly used classification datasets

and relationships of words within documents over long distances and represent this information at a much more sophisticated level when compared to traditional language models such as TF-IDF or n-grams [37].

RNNs have had widespread use in solving NLP-related tasks due to their ability in capturing the sequential nature of text at the character, word or sentence level. Consequently, RNNs are capable of creating language models that account for the semantic meaning of words based on the previously occurring words in the sentence, allowing models to be capable of understanding the difference between similar words or phrases (e.g. that the word "dog" is likely referring to the animal whereas "hot dog" would be more likely to refer to food). Accounting for RNNs' capability in handling variable length inputs (e.g. long sentences, paragraphs or documents), they have shown to produce state-of-the-art results in classification tasks such as sentiment, question and topic classification [18].

The simplest form of an RNN, as illustrated in Figure 2.3, consists of three layers: the input layer x_t, hidden state s_t and the output layer o_t, where t represents the current timestep. The input layer is typically represented as a one-hot encoding or embedding, the output layer is the resulting output which can take many forms, most commonly, it is the output of the softmax function and the hidden state is essentially the network's memory, as it captures and incorporates the information from previous timesteps into the current one. The hidden state is calculated by evaluating Equation 2.5:

$$s_t = f(Ux_t + Ws_{t-1}) \tag{2.5}$$

Where f is a nonlinear function e.g. Rectifier Linear Unit (ReLU), U, V and W are weight matrices that are shared across timesteps [37].

LSTMs [17] are a variant of RNNs that improves on it by introducing a 'forget' gate which regulates a cell's state within the network as it can allow information from different cells to be removed or added. This addition allows LSTMs to overcome the vanishing and exploding gradient problem [31] that feedforward networks i.e. RNNs are prone to.

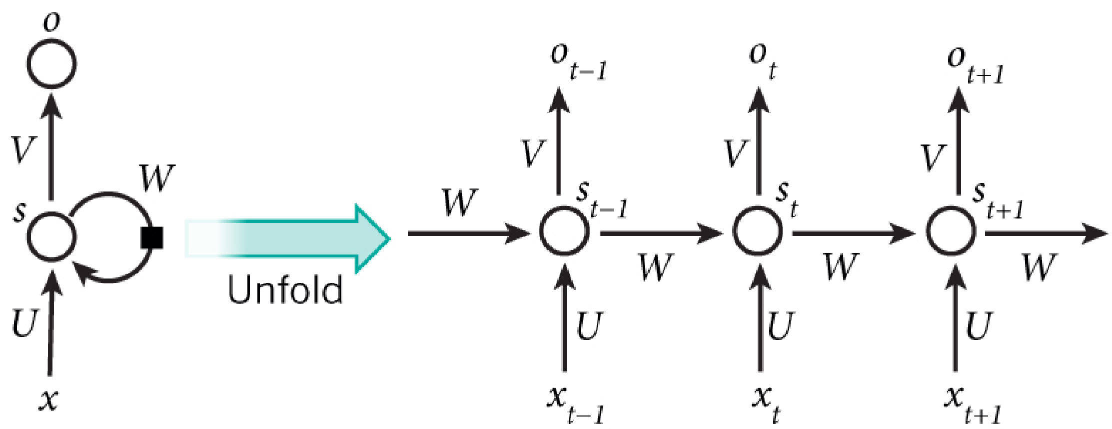

Figure 2.3: Layout of a simple RNN [3]

Convolutional Neural Networks

Despite CNNs being initially developed for object recognition tasks [23], they are commonly employed for text classification tasks such as sentence classification and sentiment analysis as they have shown to produce competitive results when compared to RNNs and LSTMs [14] [21] [27].

In the context of text classification, a typical CNN is composed of the following main elements [37]:

- Kernels - Convolutional filters that acts as a sliding window function through an embedding matrix. CNNs typically employ hundreds of kernels each of which learns to extract for a specific n-gram pattern.
- Pooling layer - Commonly either a max or average pooling layer, which maps the input to a fixed dimension in order to reduce the dimensionality of the output and ensuring that the most salient n-gram features of a sentence is kept.
- Nonlinear activation function - Such as a ReLU function applied to the results to output a prediction.

By stacking the kernels and pooling layers, deep CNNs can be constructed which can automatically capture an abstract and rich representation of the information [37]. These representations are considered to be efficient when compared to traditional representation methods (e.g. n-grams) as they don't require the storage of the entire vocabulary and is not as computationally intensive. CNNs are also more computationally efficient when compared to RNNs and LSTMs, CNNs typically require larger amounts of data in order to produce competitive results against its RNN and LSTM counterparts due to the higher number of trainable parameters that CNNs have. Another limitation of CNNs is their inability to model long-distance relationships and preserving the sequential nature of text within its representations.

2.2.4 Classifying Documents with Limited Labelled Examples

Typically, large amounts of training data is required to train a deep learning model in learning the language model for state of the art results, in the task of document classification for instance, the size of commonly used non-domain specific datasets range from hundreds of thousands of training examples to millions [15] [39]. Due to these constraints, it is not feasible to procure a dataset for the domain specific task of this project due to the aforementioned knowledge expertise and time requirements to manually label the articles required.

Transfer Learning

Talk about transfer learning and how it works and how it is applicable to this project.

N-Shot Learning

Talk about zero/few/etc-shot learning and how it works and how it is applicable to this project.

2.3 Discussion

Summarize lit review and describe why DL-based approaches should be preferred over ML-based for this type of problem. Also talk about Transfer/N-Shot learning and describe which one will be feasible given the project's time constraints

2.3.1 Scaling the Assessment of Quality and Credibility of Health Information

Summarise the idea that the previous studies applying instruments/tools to assess the quality of online health information have been relatively small because they rely on experts. It would be useful to be able to develop a way to automate the assessment of nline health information because it could lead to new ways of dealing with the spread of misinformation online, and that is hasn't been done properly before.

2.3.2 Using Deep Learning Methods to Automate Quality and Credibility Assessments

Explain that deep learning is a good choice for building an automated assessment tool because it should outperform traditional methods of document classification, but because DL methos usually require large datasets of labelled exampled than it makes sense to investigate transfer learning/n-shot lerning.

Chapter 3

Proposed Approach

3.1 Rationale

Introduce and discuss the factors that led to me choosing the proposed approach

3.2 Credibility Criteria

Introduce and discuss the 7 criteria that will be classified and describe how the criteria was determined

3.3 Study Data

Talk about the data I'll be using, how we got it, its characteristics etc.

3.4 System Model

Describe the architecture of the model

3.5 Experiments

Describe the experiments that I'm planning to do (in such a way that they are easily reproducible)

3.6 Outcome Measures

Talk about the type of analyses that I'll be doing to determine the performance of my proposed model

Chapter 4

Conclusions and Future Work

4.1 Conclusions

The end

4.2 Future Work

Chapter 5

Abbreviations

AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BS	Base Station
CSI	Channel State Information
CSIR	Channel State Information at Receiver
CSIT	Channel State Information at Transmitter
dB	Decibels
DPC	Dirty Paper Coding
GS	Gram-Schmidt
RVQ	Random Vector Quantisation
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SINR	Signal to Interference plus Noise Ratio
MISO	Multiple Input Single Output
SIMO	Single Input Multiple Output
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
MRC	Maximum Ratio Combining
QoS	Quality of Service
TDD	Time Division Duplex
FDD	Frequency Division Duplex
ZF	Zero-Forcing
ZFBF	Zero-Forcing Beamforming
ZMCSCG	Zero Mean Circularly Symmetric Complex Gaussian

Appendix A

name of appendix A

A.1 Overview

here is the Overview of appendix A ...

A.2 Name of this section

here is the content of this section ...

Appendix B

name of appendix B

B.1 Overview

here is the Overview of appendix B ...

B.2 Name of this section

here is the content of this section ...

Bibliography

[1] "DISCERN - The DISCERN Instrument." [Online]. Available: http://www.discern.org.uk/discern{_}instrument.php

[2] "Our Review Criteria - HealthNewsReview.org." [Online]. Available: https://www.healthnewsreview.org/about-us/review-criteria/

[3] "Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs – WildML." [Online]. Available: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

[4] "Understanding the kernel trick. – Towards Data Science." [Online]. Available: https://towardsdatascience.com/understanding-the-kernel-trick-e0bc6112ef78

[5] C. C. Aggarwal and C. Zhai, "A SURVEY OF TEXT CLASSIFICATION ALGORITHMS," 2012. [Online]. Available: http://alias-i.com/lingpipe/

[6] M. A. Aizerman, E. A. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol. 25, pp. 821–837, June 1964.

[7] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," Tech. Rep., 2017. [Online]. Available: http://en.wikipedia.org/wiki/Statistics

[8] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," Tech. Rep., 2000. [Online]. Available: www.elsevier.com/locate/jmicmeth

[9] J. M. Batchelor and Y. Ohya, "Use of the DISCERN Instrument by Patients and Health Professionals to Assess Information Resources on Treatments for Asthma and Atopic Dermatitis," Tech. Rep., 2009. [Online]. Available: www.jsaweb.jp!

[10] D. C. Burgess, M. A. Burgess, and J. Leask, "The MMR vaccination and autism controversy in United Kingdom 1998-2005: Inevitable community outrage or a failure of risk communication?" *Vaccine*, vol. 24, pp. 3921–3928, 2006. [Online]. Available: https://ac.els-cdn.com/S0264410X06002076/1-s2.0-S0264410X06002076-main.

pdf?{_}tid=46d1dda6-f576-4f5e-ad53-d550f1cd9990{&}acdnat=1534726962{_}1b237371d8bb916694f34f0f951c84bc

[11] E. Cambria, S. Poria, A. Gelbukh, I. P. Nacional, and M. Thelwall, "AFFECTIVE COMPUTING AND SENTIMENT ANALYSIS Sentiment Analysis Is a Big Suitcase," Tech. Rep., 2017. [Online]. Available: www.computer.org/intelligent

[12] N. Cantey Banasiak and M. Meadows-Oliver, "Journal of Asthma and Allergy Dovepress Evaluating asthma websites using the Brief DISCERN instrument," *Journal of Asthma and Allergy*, pp. 10–191, 2017. [Online]. Available: http://dx.doi.org/10.2147/JAA.S133536

[13] C. Cipriani, J. Pepe, S. Minisola, and ·. E. M. Lewiecki, "Adverse effects of media reports on the treatment of osteoporosis," *Journal of Endocrinological Investigation*. [Online]. Available: https://doi.org/10.1007/s40618-018-0898-9

[14] R. Collobert, J. Weston, J. Com, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural Language Processing (Almost) from Scratch," *Journal of Machine Learning Research*, vol. 12, pp. 2493–2537, 2011. [Online]. Available: http://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf

[15] A. Conneau, H. Schwenk, Y. Le Cun, and L. Loıc Barrault, "Very Deep Convolutional Networks for Text Classification," 2017. [Online]. Available: https://arxiv.org/pdf/1606.01781.pdf

[16] V. Hall, E. Banerjee, C. Kenyon, A. Strain, J. Griffith, K. Como-Sabetti, J. Heath, L. Bahta, K. Martin, M. McMahon, D. Johnson, M. Roddy, D. Dunn, and K. Ehresmann, "Measles Outbreak — Minnesota April–May 2017," *MMWR. Morbidity and Mortality Weekly Report*, vol. 66, no. 27, pp. 713–717, jul 2017. [Online]. Available: http://www.cdc.gov/mmwr/volumes/66/wr/mm6627a1.htm

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[18] J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification," Tech. Rep., 2018. [Online]. Available: http://nlp.fast.ai/ulmfit.

[19] F. Informatik, "UNIVERSIT AT D ORTMUND Text Categorization with Support Vector Machines: Learning with Many Relevant F eatures Thorsten Joachims," Tech. Rep., 1997. [Online]. Available: https://www.cs.cornell.edu/people/tj/publications/joachims{_}97b.pdf

[20] J. Kaicker, V. Borg Debono, W. Dang, N. Buckley, and L. Thabane, "Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument," Tech. Rep., 2010. [Online]. Available: http://www.biomedcentral.com/1741-7015/8/59

[21] Y. Kim, "Convolutional Neural Networks for Sentence Classification," Tech. Rep. [Online]. Available: http://nlp.stanford.edu/sentiment/

[22] V. Korde and C. N. Mahender, "TEXT CLASSIFICATION AND CLASSIFIERS: A SURVEY," *International Journal of Artificial Intelligence & Applications (IJAIA)*, vol. 3, no. 2, 2012. [Online]. Available: http://aircconline.com/ijaia/V3N2/3212ijaia08.pdf

[23] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object Recognition with Gradient-Based Learning." Springer, Berlin, Heidelberg, 1999, pp. 319–345. [Online]. Available: http://link.springer.com/10.1007/3-540-46805-6{_}19

[24] K. Matsoukas, S. Hyun, L. Currie, M. P. Joyce, J. Oliver, S. Patel, O. Velez, P.-Y. Yen, and S. Bakken, "Expanding DISCERN to create a tool for assessing the quality of Web-based health information resources," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, p. 1048, 2008.

[25] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," Tech. Rep., 1998. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.9324{&}rep=rep1{&}type=pdf

[26] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Tech. Rep. [Online]. Available: https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

[27] C. Nogueira, D. Santos, and M. Gatti, "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts," Tech. Rep. [Online]. Available: http://www.aclweb.org/anthology/C14-1008

[28] K. Pasupa, "A Comparison between Shallow and Deep Architecture Classifiers on Small Dataset," 2016.

[29] R. Rosenfeld, "TWO DECADES OF STATISTICAL LANGUAGE MODELING: WHERE DO WE GO FROM HERE?" Tech. Rep., 2000. [Online]. Available: https://www.cs.cmu.edu/{~}roni/papers/survey-slm-IEEE-PROC-0004.pdf

[30] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approach to filtering junk e-mail," Tech. Rep. Cohen, 1998. [Online]. Available: http://research.microsoft.com/en-us/um/people/horvitz/junkfilter.htm

[31] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," Tech. Rep. [Online]. Available: http://nlp.stanford.edu/

[32] R. Som and N. P. Gunawardana, "Internet chemotherapy information is of good quality: assessment with the DISCERN tool," *British Journal of Cancer*, vol. 107, no. 2, pp. 403–403, jul 2012. [Online]. Available: http://www.nature.com/articles/bjc2012223

[33] S. Sommariva, C. Vamos, L. U.-L. Mantzarlis, Alexios Dao, and D. Martinez Tyson, "Spreading the (Fake) News: Exploring Health Messages on Social Media and the Implications for Health Professionals Using a Case Study," *American Journal of Health Education*, vol. 49, no. 4, pp. 246–255, jul 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/19325037.2018.1473178

[34] "Germany investigating unprecedented spread of fake news online — World news — The Guardian," 2017. [Online]. Available: https://www.theguardian.com/world/2017/jan/09/germany-investigating-spread-fake-news-online-russia-election

[35] S. Vosoughi, D. Roy, and S. Aral, "SI:The spread of true and false news online," Tech. Rep. 6380, 2018. [Online]. Available: http://science.sciencemag.org/content/sci/359/6380/1146.full.pdf{%}0Ahttp://www.sciencemag.org/lookup/doi/10.1126/science.aap9559

[36] J. Weston, S. Bengio, and N. Usunier, "WSABIE: Scaling Up To Large Vocabulary Image Annotation," Tech. Rep. [Online]. Available: http://torch5.sourceforge.net/

[37] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," Tech. Rep. [Online]. Available: http://veredshwartz.blogspot.sg.

[38] D. Zeraatkar, M. Obeda, J. S. Ginsberg, and J. Hirsh, "The development and validation of an instrument to measure the quality of health research reports in the lay media," *BMC Public Health*, vol. 17, no. 1, p. 343, dec 2017. [Online]. Available: http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-017-4259-y

[39] X. Zhang, J. Zhao, and Y. Lecun, "Character-level Convolutional Networks for Text Classification," 2015. [Online]. Available: https://arxiv.org/pdf/1509.01626.pdf

[40] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM Neural Network for Text Classification," 2015. [Online]. Available: https://arxiv.org/pdf/1511.08630.pdf