TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Ivan Švaiger 204211IABM

# MULTIMODAL CONVOLUTIONAL NEURAL NETWORKS FOR COVID-19 FAKE NEWS DETECTION

Master's thesis

Supervisor: Nadežda Furs, MBA

Tallinn 2022

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Ivan Švaiger 204211IABM

# MULTIMODAALSED KONVOLUTSIOONILISED NÄRVIVÕRGUD COVID-19 VÕLTSUUDISTE TUVASTAMISEKS

Magistritöö

Juhendaja: Nadežda Furs, MBA

Tallinn 2022

# Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Ivan Švaiger

11.05.2022

# Abstract

Fake news is false or misleading information predominantly created to deliberately misinform or deceive the reader. This kind of information is published in various news sources and thus has the potential to affect millions of people around the world. Recently, a huge stream of misinformation and fake news about the COVID-19 pandemic has undoubtedly negatively affected the field of healthcare. Also this incident had a negative impact on society, the economy and the social sphere. Therefore, the aim of this thesis is to find an approach that could most effectively distinguish reliable sources of information about COVID-19 from unreliable and fake ones.

The dataset used in this thesis includes COVID-19 textual, visual and network modalities about. To achieve the goal of the thesis, the corresponding mono-modal and multimodal architectures were built on the basis of these modalities. In the course of the work, these models are trained using a convolutional neural network, pre-trained word2vec embeddings, and VGG16 image recognition techniques. The data used in this work were cleaned, pre-processed and presented in the appropriate structures required for the experiment. In the final analysis, the experimental results show that the multimodal CNN architecture is able to provide a significant increase in accuracy and efficiency for COVID-19 fake news detection.

Using the multimodal CNN approach, a multimodal network consisting of two textual, visual and meta modalities was trained with an accuracy of 90.55% and a f1-score for fake news class of 81.67%. The multimodal network is able to provide a 6% increase in accuracy and more than 9% increase in f1-score, relative to the best mono-modal CNN model. According to this fact multimodal network is significantly outperforms all other models.

This thesis is in English and contains 64 pages of text, 7 chapters, 28 figures, 10 tables.

# Annotatsioon

## Multimodaalsed konvolutsioonilised närvivõrgud COVID-19 võltsuudiste tuvastamiseks

Võltsuudis on vale või eksitav teave, mis on peamiselt loodud lugeja tahtlikuks eksitamiseks või petmiseks. Taolist teavet avaldatakse mitmesugustes ressurssides, mis tähendab, et sellel on potentsiaalne mõju miljonitele inimestele kogu maailmas. Hiljuti on tervishoiuvaldkonda kahtlemata negatiivselt mõjutanud tohutu eksiteabe ja võltsuudiste hulk COVID-19 pandeemia teemal. Samuti oli antud vahejuhtumil negatiivne mõju ühiskonnale, majandusele ja sotsiaalvaldkonnale. Sellest lähtuvalt on käesoleva lõputöö eesmärk leida lähenemisviis, mis eristaks kõige tõhusamalt usaldusväärseid COVID-19 teabeallikaid ebausaldusväärsetest ja võltsitutest.

Käesolevas lõputöös kasutatav andmestik sisaldab COVID-19 tekstilisi, visuaalseid ja võrgu modaalsusi. Lõputöö eesmärgi saavutamiseks kavandati ning ehitati nende modaalsuste põhjal üles vastavad monomodaalsed ja multimodaalsed arhitektuurid. Töö käigus treenitakse neid mudeleid konvolutsioonilise närvivõrgu, eeltreenitakse word2vec vektorite ja VGG16 pildituvastuse meetodite abil. Töös kasutatud andmeid puhastati, eeltöödeldi ning esitati neid katsele sobival struktuuril. Lõppanalüüsis tehtud katsetulemused näitavad, et multimodaalne CNN-i arhitektuur suudab COVID-19 võltsuudiste tuvastamisel suurendada täpsust ja tõhusust märkimisväärselt.

Multimodaalse CNN-i lähenemisviisi abil treeniti välja kahest tekstilisest, visuaalsest ja metamodaalsusest koosnev multimodaalne võrgustik täpsusega 90,55% ja võltsuudiste klassi F1 skooriga 81,67%. Võrreldes parima monomodaalse CNN-i mudeliga, suudab multimodaalne võrk suurendada täpsust 6% ja F1 skoori rohkem kui 9% võrra. Antud fakti kohaselt edestab multimodaalne võrk märkimisväärselt kõiki teisi mudeleid.

Lõputöö on kirjutatud inglise keeles ning sisaldab 64 teksti leheküljel, 7 peatükki, 28 joonist, 10 tabelit.

# List of abbreviations and terms

| | |
|---|---|
| CNN | Convolutional neural networks |
| COVID-19 | Coronavirus disease 2019 |
| RNN | Recurrent neural network |
| BERT | Bidirectional Encoder Representations from Transformers |
| GloVe | Global Vectors for Word Representation |
| EU | European Union |
| MBFC | Monitor Based Flow Control |
| LSTM | Long short-term memory |
| JPEG | Bitmap graphic format |
| ANN | Artificial neural networks |

# Table of contents

# List of figures

# List of tables

# 1.Introduction

People learn about the world through information published in print, broadcasted on television and radio or published on the Internet. Information can come from virtually anywhere — social media, blogs, personal experiences, books, journal and magazine articles, expert opinions, newspapers, and websites [1]. The advancement of information technologies, the Internet and the rapid implementation of social networking platforms such as Facebook and Twitter, as well as modern methods of monetising information have led to the new sources of information. Consequently, the amount of misinformation has also increased.

By definition, misinformation or misleading information is false or inaccurate information that is deliberately created and is intentionally or unintentionally propagated [2]. According to different kinds of sources [3],[4] misinformation has a negative impact on people's consciousness and leads to a violation of relations between people, accepted norms and rules in society, values and traditions. Fake news is one of the main driving forces in the spread of various kinds of misinformation.

Fake news is usually defined as false or misleading information, predominantly created to deliberately misinform or deceive the reader [35]. Misleading information through fake news spreads rapidly through social media, where it can impact millions of users. The most striking example of this kind is the „infodemic"[1], announced by the World Health Organization against the background of the policy to control the spread of COVID-19 and the large amount of related misinformation that undermines public consciousness and people's opinion. Given the ease with which disinformation is received and disseminated through social media platforms, it becomes increasingly difficult to know what to trust [5]. Therefore, it is imperative to develop a solution for identification media and news content using the most modern methods and approaches.

[1] https://www.who.int/health-topics/infodemic#tab=tab_1

## 1.1. Aim and research tasks

The aim of the study is to provide a strong and comprehensive comparative research of mono-modal and multimodal approaches for COVID-19 fake news detection, using convolutional neural networks, pre-trained word2vec embeddings, and VGG16 image recognition architecture. The main contribution of this thesis is the development of an approach that could most effectively distinguish reliable sources of information about COVID-19 from unreliable and fake ones. To achieve its goals, the thesis uses four accessible data modalities, consisting of news published by various sources, visual information in the form of images in these news, Twitter posts associated with these news, and general metadata in tweets.

Based on all of the above, the following research tasks were set:

1. Build mono-modal models for each of the available modalities.
2. Combine mono-modal models into logically related multimodal pairs of news and images, as well as tweets and metadata.
3. Combine all four data modalities into one multimodal network and explore the possible benefits of this multimodal network.

## 1.2. Research questions

As part of this thesis, the author gives answers to the following research questions:

1. What mono-modal approach is most effective textual, visual, or metadata based for COVID-19 fake news detection?
2. To what extent are combined multimodal approaches more effective than mono-modal approaches in classifying reliable sources of information from unreliable and fake ones?
3. How justified from the point of view of efficiency is the use of one deep learning architecture of convolutional neural networks to build and train models for all modalities?

## 1.3. Validation methods

Model validation is a foundational technique for evaluating how well built models are going to react to new data [42]. Within this study, the following validation methods will be used: accuracy, precision, recall, f1-score, confusion matrix and model loss and accuracy curves. Accordingly, the author answers all the questions of the study by logically applying the above methods. Validation methods will be described in more detail in the Section 5.

## 1.4. Main tools and datasets

This master's thesis uses the Google Colaboratory tool, Python programming language and its Numpy, Pandas, OpenCV, Keras, Pickle and Tensorflow libraries. And a combination of two datasets:

- A Multimodal COVID-19 Repository[1]. The dataset contains 2029 news articles with visual and textual information, collected from 22 reliable and 38 unreliable websites along with 140820 tweets related to these articles between 01/21/2020 and 05/26/2020.

- MMCoVaR: Multimodal COVID-19 Vaccine Focused Data Repository[2]. The dataset contains 2593 news articles with multimodal data and 24184 tweets between 16/02/2020 and 17/03/2021.

Datasets and main tools will be described in more detail in the Section 4 and Section Y.

## 1.5. Research originality

The originality of the study lies in the use of multimodal approaches to classify reliable sources of information on COVID-19 from fake ones. In addition to the textual and visual modalities of news articles, tweets and metadata are also used. This can improve the accuracy and greatly improve the efficiency of automatic detection systems.

---

[1] https://github.com/apurvamulay/ReCOVery

[2] https://arxiv.org/abs/2109.06416

This study will be useful and beneficial for researchers in this field as it will give: 1) an overview of how misinformation and fake news affects the economy and society; 2) an overview and analysis of mono-modal and multimodal approaches for COVID-19 fake news detection 3) an overview of convolutional neural networks as a common architecture for textual, visual and metadata modalities.

The rest of the thesis is organised as follows. In Section 2, the fake news economical, social and legal overview are discussed. In Section 3, related works are discussed. In Section 4, dataset and used repositories are discussed. In Section 5, methodology and model validation methods are discussed. In Section 6, multimodal architecture and tools are discussed. In Section 7, experiments and results as well as possibilities for further work are discussed.

# 2.Economic, social and legal overview of fake news

This section will describe economic, social and legal overview of fake news and misinformation. Due to the ideological and economic motivations behind the spread of fake news, it is extremely important to look at the problem of fake news from different perspectives. Also, this section describes the legislative initiatives of various states to prevent this problem.

## 2.1. Fake news and society

As Fake News increasingly influences public values, opinions on critical issues and topics, and redefines facts, truths, and beliefs [6], the scale of the problem and the reach that fake news can achieve is hard to imagine.

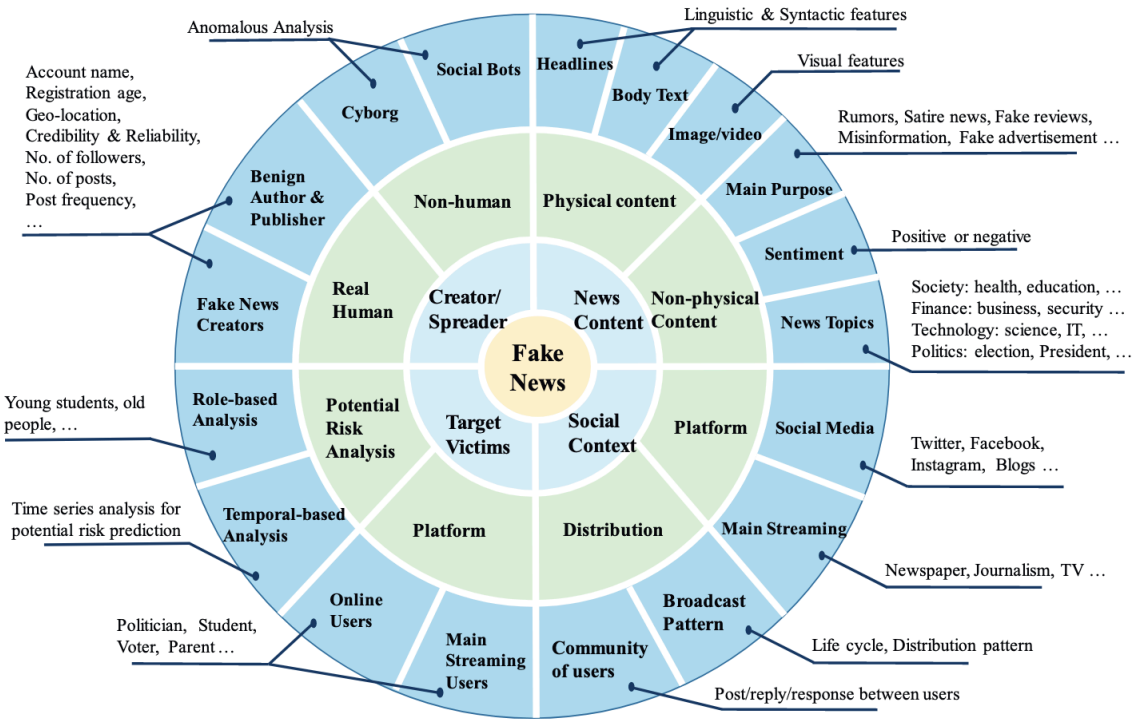


Figure 1. Fake news and its basis.[1]

According to statistics, social networks are the least reliable source of news worldwide. The spread of fabricated or fake news circulated via social media or fake news is

[1] https://www.researchgate.net/publication/331913505_An_overview_of_online_fake_news_Characterization_detection_and_discussion

difficult to control and can distort the perception and knowledge of audiences on issues ranging from healthcare and business to the political system. The first main motivation underlie the production and development of fake news: ideological. Fake news providers produce false news stories to promote particular ideas or people that they favour, often by discrediting others [7].

Although, in general, there is a clear trend to trust classic media channels more than Internet channels on these issues, global media trust has fallen by 8% from 2020 to 2021. And nearly 80% of respondents in the US have encountered fake news about the COVID-19 pandemic in their news feed. And according to a 2018 survey[1] across all 28 European member states, 36 percent come across fake news every day, or almost every day.

## 2.2. Fake news and economy

In addition to the social sphere and society, fake news also affects the economy. The second motivation underlie the production and development of fake news is financial. Often fake news go viral simply because they are outrageous and provide content producers with clicks that are convertible to advertising revenue [7].

It was estimated that, in 2021, websites that repeatedly published fake news generated 2.6 billion U.S. in advertising revenue worldwide. In the United States only, the figure stood at 1.62 billion U.S. dollars [8]. The revenue comes from programmatic advertising, whose buying is an automatised process, and advertisers have little influence on where and around what kind of content their ads are placed [8]. Fake news not only brings profit to those who write it, but also brings losses to various companies. According to analysts at the University of Baltimore, shareholders are losing $39 billion annually due to false news globally, while the damage to the global economy is $78 billion a year [9].

[1] https://www.valitsus.ee/en/news/estonian-population-least-critical-dangers-fake-news-eu

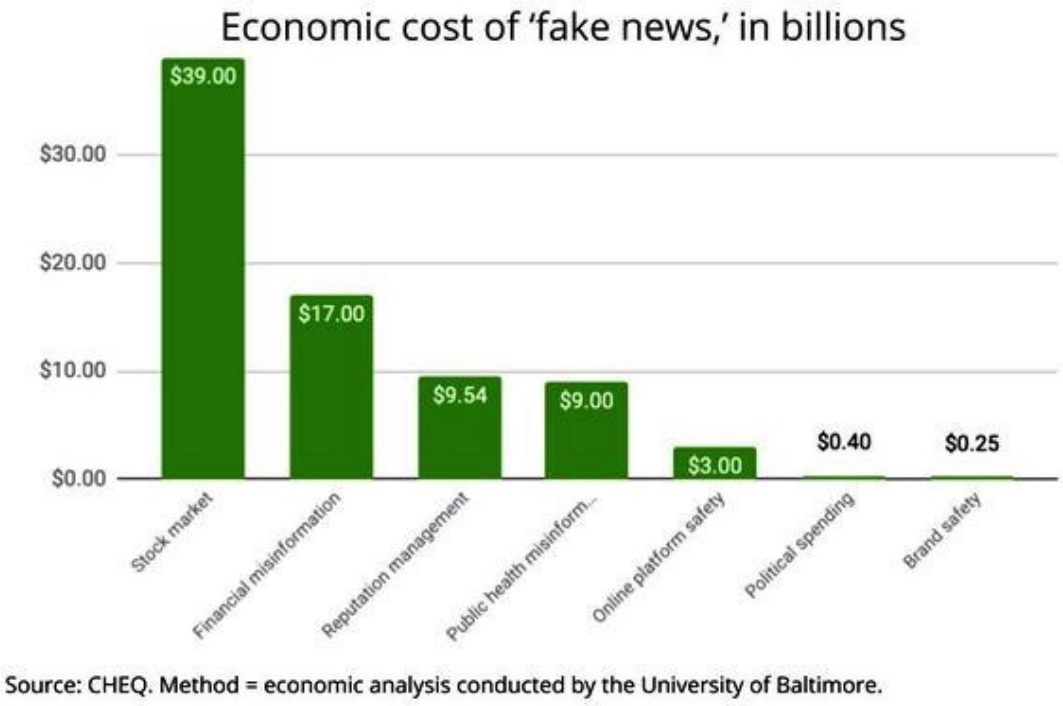


Figure 2. Economic cost of fake news.[1]

Cybersecurity company CHEQ conducted a study with the University of Baltimore (2019) and the results show that companies will lose about $9 billion annually due to healthcare misinformation, $17 billion due to financial misinformation, $9 billion due to reputation management. $3 billion due to platform security efforts and $400 billion due to bogus political ads. Brands lose about $235 million a year due to displaying ads alongside fake news [9].

## 2.3. Fake news and legislation

In order to counteract the flow of disinformation and fake news mainly in social media and thereby reduce their negative impact on the social sphere, society and the economy, decisive steps must be taken at the state level. The European Commission's 2018 report on disinformation „A multi-dimensional approach do disinformation", offers a collaborative approach to countering disinformation around the world. The general recommendations presented in the report include transparency so that citizens have clear information about news sources and funding; diversity of information both online and

[1] https://priorityconsultants.com/blog/fake-news-and-its-impact-on-the-economy/

offline because this fuels critical judgment; credibility of information must be obvious to citizens; and inclusivity [10].

Recently, the International Press Institute has documented at least 19 countries around the world that adopted fake news regulations just during the COVID-19 pandemic. All the actions of governments of various countries against misinformation will be presented further in Figure 3.

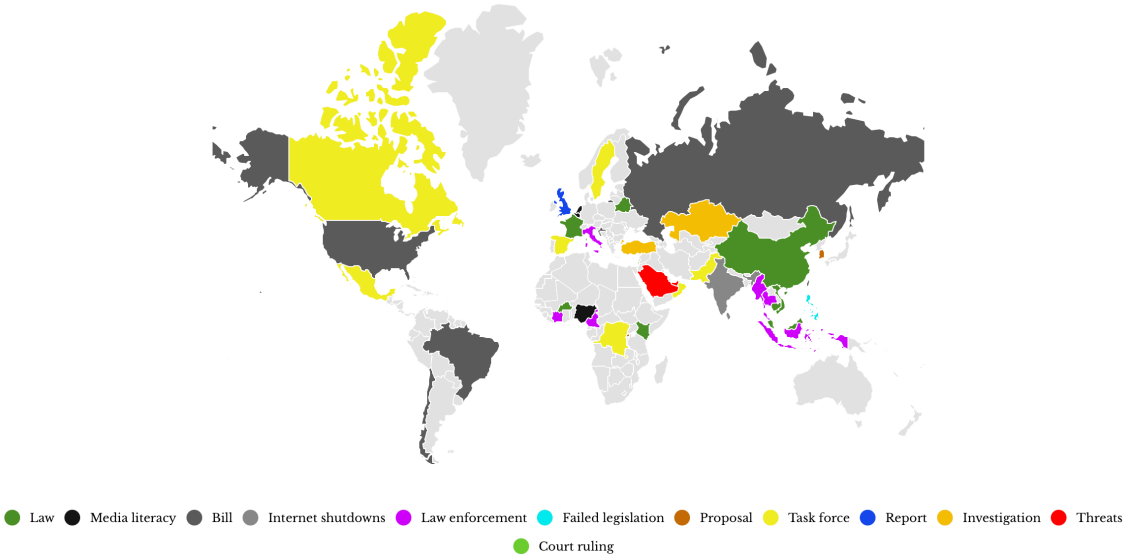


Figure 3. Governments actions against online misinformation[1]

At the moment, there is no general legislative framework for the regulation of disinformation in social networks. At the beginning of 2021, out of 27 states that are members of the European Union, only nine: Belgium, Croatia, Denmark, France, Germany, Holland, Spain, Italy and Sweden have various regulations related to disinformation. The most stringent legislation in Germany. The law, adopted in June 2017, titled NetzDG, forces online platforms with more than 2 million members to remove "obviously illegal" posts within 24 hours or risk fines of up to €50 million [11]. In the US, disinformation actions are patchy and state-specific, and the first bills began to appear only after the 2016 presidential election.

As of April 2022, an agreement has been reached between the 27 EU Member States and EU legislators on a digital services law aimed at combating hate speech,

[1] https://www.poynter.org/ifcn/anti-misinformation-actions/

disinformation and harmful online content [12]. The law is expected to come into force as early as 2024.

# 3.Related works

This section provides a review of various works related to the topic of fake news and a summary of each article. These are academic papers submitted over the past 5 years focusing on mono-modal and multimodal approaches and various deep learning architectures.

## 3.1. Mono-modal fake news articles

Answini Thota, Priyanka Tilak, Simeratjeet Ahluwalia and Nibhrat Lohia in their article [13] propose three variations of dense neural network (DNN) deep learning models using the Fake News Challenge (FNC-1) dataset for rumour debunking. The dataset includes the body of the news article, the headline of the news article, and the label attribute. The authors perform stop words and punctuation removal. In addition, the authors used stemming and other data pre-processing techniques to represent words using three different methods:: word2vec, Tf-Idf and Bag of Words. In all cases, a model with batch size of 64, dropout rate of 0.2, and Softmax as final layer activation function is used. The Tf-Idf DNN model showed the best result with an overall accuracy of 94.31%, which outperforms existing model architectures built on the same dataset by 2%.

In the article "The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors» [14], the authors test the power of deep learning on new news topics and generalisations to detect fake news in novel subjects only from language patterns. The work uses various data collections for fake and real news from the Kaggle repository, real news samples from The New York Times and The Guardian published before, after and during the 2016 United States Presidential Election. To represent words, the authors use pre-trained word2vec 1,000-dimensional representation embedding, as well as convolutional neural network with 128 filters of input, and fully connected output with Softmax function. The authors track units of the pooling layers from top-20 units, which caused the activation of the max-pooling. The accuracy of the text convolutional network presented by the authors using language patterns with the

word "Trump" as a hold-out topic is on average 87.7%. Using as a test sample 4000 randomly selected articles, the accuracy of the model improves to 93.5%. This may indicate overfitting and bias of the model regarding new topics.

Hybrid approaches are also used to solve fake news detection problem. Oluwaseun Ajao, Deepayan Bhowmik and Shahrzad Zargari proposed in their paper [15] a framework based on a hybrid approach of convolutional neural networks (CNN) and long-short term recurrent neural networks (RNN) for Twitter posts fake messages detection and classification. The authors implemented three deep neural networks: RNN, RNN with dropout regularisation and hybrid RNN and CNN, on a dataset consisting of 5800 tweets centred on five rumour stories. Using only the text modality the best result in terms of precision, recall and F-measure showed the plain vanilla LSTM model with an accuracy of 82.3%. The hybrid implementation of RNN and CNN showed the worst result with an accuracy of 80.4%. The authors attribute this to the fact that hybrid models require much larger amounts of data for efficient training.

As for COVID-19 fake news detection, Apurva Wani, Isha Joshi, Snehal Khandve, Vedangi Wagh, and Raviraj Joshi [16] proposed two different approaches: transformer-based BERT and DistilBERT convolutional neural network models, as well as LSMT, Bi-LSTM + Attention and HAN sequential models. For sequential models, the authors use two types of word embeddings 100-dimensional pre-trained Glove and 300 dimensional FastText. The dataset used is The Contraint@AAAI 2021 Covid-19 Fake news detection dataset, consisting of 10700 tweets. The authors used pre-training with COVID-19 corpus and fine-tuning the transformer-based models. Transformer-based models showed the best accuracy. The results showed that transformer-based models outperform other models with a difference of 3-4% and the best result was shown by the pre-trained BERT model with an accuracy of 98.41%.

In another article [17] on ensemble learning for COVID-19 fake news detection, the authors propose approaches using the transformer-based BERT, RoBERTa and COVID-Twitter-BERT ensemble models. The authors use the dataset contains 10700 manually annotated social media posts with 37505 number of unique words (vocabulary size). The following steps were taken for the dataset pre-processing: removing and tokenising

hashtags, URLs, emoji and mentions. Using Linear SVC model results with bag of words with F1-Score equal to 88.39% as baseline, the authors using fine tuning and ensemble models were able to improve the accuracy of the model by more than 10% and the experimental results showed that CT-BERT solution achieved 98.69% of the weighted F1-Score on test data. To achieve this result, the authors trained 3 models and then used hard-voting to ensemble predictions together.

## 3.1. Multimodal fake news articles

In addition to mono-modal approaches, where in the vast majority of cases researchers use only one text modality for the detection and classification of fake news, there are also multimodal approaches.

To improve the model accuracy of the automatic anti-vaccine message detector in the article "Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning» [19], the authors propose a deep learning network that uses both visual and textual information. The proposed model consists of three branches: text modality, hashtags, visual modality; and can generate complex combined functions for predictions. According to the authors, this approach should be used, since the existing systems are not enough to detect anti-vaccine messages with heavy visual components. The dataset was collected 31282 Instagram posts from January 2016 to October 2019 of which 50% are anti-vaccine posts. Based on the experiments, the following important conclusions were drawn: a multimodal network works better than models that take into account only unimodal information; and text based models perform better than image based models. On average, various variations of multimodal networks proposed by the authors are 10% more accurate than single-modal text networks and 13-14% more accurate than single-modal visual networks. Best multimodal model with the proposed ensemble method achieved 96.1% for the average precision and 95.8% recall values separately.

The proposed multi-modal structure or framework Coupled ConvNet [20], according to the authors, is superior to various modern methods for detecting fake news, combines text and visual data modules and effectively classifies online news depending on their

content. Model architecture consists of Text-CNN module embedding words into vectors for text classification and Image-CNN implemented 8 different CNN architectures — AlexNet, Xception, VGG16, VGG19, ResNet50, MobileNetV2, IncepctionV3 and DenseNet for visual component classification. The ConvNet presented in this paper is a hybrid two-stream convolutional architecture based on Convolution Neural Networks and tested on three datasets of two multimodal ones: Ti-CNN and Emergent, as well as Image-only MICC-F220,. An interesting fact is that the work presents various combinations of fusion weights that were obtained experimentally based on two datasets TI-CNN and EMERGENT and, for example, for VGG16 and text modality, this proportion is 0.5 to 0.5. The results of the experiments are as follows: the Text-CNN module showed an excellent accuracy of 96.26% for Ti-CNN and 93.56 for Emergent. On the other hand, Image-CNN module with VGG16 showed the best accuracy with 82.72% for Ti-CNN and Image-CNN Xception and Resnet50 with 51.26% accuracy each. For the Image-only MICC-F2200 dataset, the Image-CNN model with Xception pre-trained model showed the best accuracy. Multimodal Text-CNN models with VGG16 improved mono-modal scores to 98.93% for Ti-CNN and 94.05% for Emergent, respectively.

Another framework [21] called SpotFake proposed for fake news detection detects fake news without taking into account any other subtasks using textual and visual features of data. The framework architecture consists of a text module based on BERT encoder and a visual module based on VGG-19 pre-trained on ImageNet dataset. Next, the two modalities are concatenated together into one common model, and then the training on two datasets TwitterMediaEval model, consisting of 17000 tweets and associated images, as well as a Weibo dataset in Chinese, collected from news sources of China. Interesting from a practical point of view is the proposed detailed experimental setup of the SpotFake framework, and specifically the method of concatenating two models, visual and text, into one. In both cases, the text and visual models use a Dense layer equal to 32 for concatenation, which concatenates into a Hidden layer equal to 35 and a layer with sigmoid activation. The SpotFake framework outperforms the previously best MVAE model obtained on this dataset by more than 3% with an average accuracy of 77.77% for the TwitterMediaEval dataset. For the Weibo dataset, this framework

outperforms the MVAE model by almost 7% with a score of 89.23% versus 82.40% for MVAE, respectively.

In master's thesis entitled "Automated Identification of Information Disorder in Social Media from Multimodal Data» [22], author Armin Kirchknopf uses a four-modality approach similar to that which will be used in this thesis. Based on the Fakeddit dataset, consisting of more than a million samples of information disorder obtained from the title of posts on the Reddit social network, images for them, comments and metadata about users corresponding to four modalities - two textual, visual and metadata, various mono-modal and multimodal models. The author uses BERT, ResNet50v2 and MLP architectures to build these mono-modal models as well as a multimodal network consisting of all four modalities combined. The best result among all mono-modal models was shown by title modality with an accuracy of 88.23%, and the worst result by meta modality with a result of 77.34%, using only best single features. As for multimodal models from two modalities, the best result was shown by Title-Visual modality with an accuracy of 91.0%, and the worst result was shown by Title-Meta with an accuracy of 78.2%. The multimodal network of all four modalities outperforms all of these models with an accuracy of 95.54%. Based on these results, it can be concluded that the multimodal models for the Fakeddit dataset outperform the mono-modal ones and significantly improve their accuracy.

# 4.Data

This section provides basic information about the repositories and datasets used, as well as the process of combining the two COVID-19 repositories into one common dataset, which will be used later in the thesis. A set of attributes will also be defined and a brief description of them will be given.

## 4.1. COVID-19 repositories

The data in this thesis uses a combination of two multimodal repositories: ReCOVery [40] and MMCoVaR [41], dedicated to COVID-19 fake news detection. The repositories provide multimodal information of news articles on coronavirus, including textual, visual, temporal, and network information. They have a similar architecture and therefore can be combined into one common dataset.

The ReCOVery repository contains 2029 news articles on SARS-CoV-2, COVID-19 and Coronavirus visual and textual information, collected from 22 reliable and 38 unreliable websites along with 140820 tweets related to these articles between 01/21/2020 and 05/26 /2020.

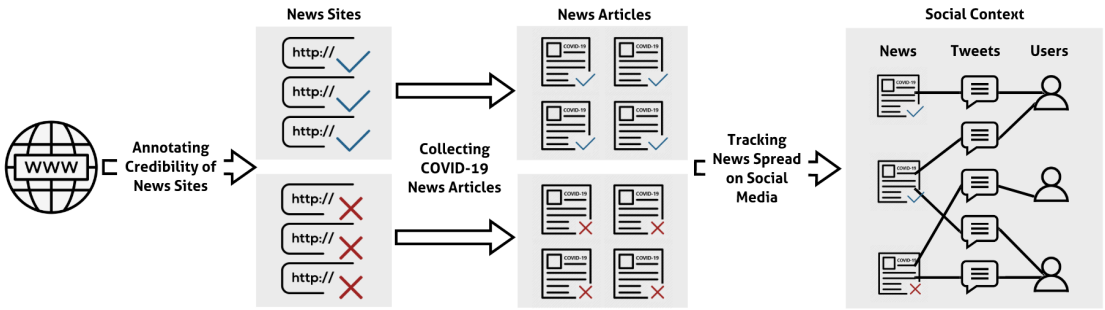


Figure 4. Data collection process for ReCOVery repository[1].

This dataset is imbalanced in news class – the proportion of reliable and unreliable sources of new articles is approximately 2:1. For the reliability determination process, NewsGuard and Media Bias/Face Check (MBFC) technologies are used. Each news article in the repository has 12 components: unique news id, news url, name of

[1] https://github.com/apurvamulay/ReCOVery

publisher, publication date, news article author, news title and body text (the main textual information), news image URL (visual information), publication country, political bias and NewsGuard and MBFC reliability score (0 or 1). The authors also used the Twitter Search API to track news articles by news url on Twitter and collect detailed information about these tweets: their IDs, text, languages of text, times of being created, statistics on retweeted/replied/liked, and so on.

The MMCoVaR repository contains 2593 news articles covering all topics related to COVID-19 from 80 publishers with multimodal consisting of images, text and temporal information data and 24184 tweets between 16/02/2020 and 17/03/2021.

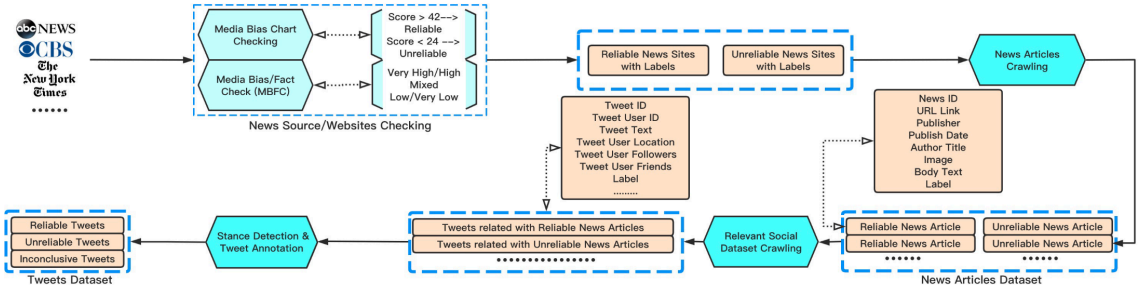


Figure 5. Data collection process for MMCoVaR repository[1]

The dataset is less imbalanced in news class - the proportion of reliable and unreliable news articles is approximately 3:2. The reliability process uses NewsGuard and Media Bias/Face Check (MBFC) technologies, as well as MB Chart. Each news article in the repository has 12 components: unique news id, news url, name of publisher, publication date, news article author, news title and body text (the main textual information), news image URL (visual information), publication country, and NewsGuard and MBFC/MB Chart reliability score (0 or 1). There is no political bias in this repository compared to ReCOVery. Regarding the collection of information on tweets, in this case, the Twitter Search API is also used, as in the ReCOVery repository. And exactly the same approach.

## 4.2. Merging two repositories

Combining two repositories into one dataset, it is necessary to take into account the fact that at least one tweet and at least one image must correspond to one news article. Since

[1] https://arxiv.org/pdf/2109.06416.pdf

not all information about pictures and tweets for April 2022 is available, the final size of the ReCOVery repository corresponds to 1723 news articles, 1214 of which are reliable and 509 unreliable and 24659 tweets. In turn, the size of the MMCoVaR repository is equal to 1761 records, 1119 of which are reliable and 642 unreliable and 7526 tweets.

The next step was to remove all images that cannot be resized for the VGG16 architecture (gif files, empty images, etc.). All tweet information obtained using the Hydrator[1] tool has been anonymised in accordance with the Twitter's Terms of Service. After that all tweets related to a particular news article are combined into single tweet, while for metadata only average values are used.

The final COVID-19 multimodal dataset used in this thesis consists of 3014 news articles, tweets, visual information and meta information. Of which 2194 are reliable news sources and 820 are unreliable news sources. As part of this thesis, it is worth considering reliable news sources as not fake, but unreliable as fake. All dataset attributes used to build mono-modal and multimodal models and their descriptions are presented in Table 1.

Table 1. Multimodal COVID-19 dataset attributes and its description.

| **Attribute** | **Description** |
|---|---|
| body_text | Text of news articles |
| twitter_text | Text of merged tweets |
| image | URL link to visual information |
| user_verified_rate | Share of twitter accounts with verified status (0 to 1) |
| retweet_count_avg | Average number of tweets retweeted |
| user_followers_count_avg | Average number of followers for twitter accounts |
| user_favourites_count | Average number of favorites |
| user_friends_count_avg | Average number of friends |
| user_listed_count_avg | Average number of users who added accounts to the list |
| reliability | Reliability of data (1 if reliable or not fake, otherwise 0) |

[1] https://github.com/DocNow/hydrator

The body_text and twitter_text attributes represent textual information, image represents visual information. User_verified_rate, retweet_count_avg, user_followers_count_avg, user_favourites_count, user_friends_count_avg and user_listed_count_avg represent meta information. Reliability is a class attribute or class label used for prediction purposes.

# 5.Methodology

This section provides information on methodology and validation methods. The methodology include: data pre-processing, model training using Google Colab environment, comparison and selection model evaluation. In addition, this section details the technologies used: convolutional neural network, word2vec and VGG16.

## 5.1. Data pre-processing

Since the multimodal COVID-19 dataset used in this thesis, described in detail in Section 4, uses different types of data, it is necessary to perform pre-processing for each type of data separately in parallel. Data pre-processing includes data cleaning, data transformation and data splitting.

### 5.1.1. Data cleaning

Data cleaning for textual data news and tweets is the same and includes textual data cleaning involved getting rid of punctuation, lowering each word, removal of non-alphabetic words and stop words, as well as lemmatisation. Unlike textual data for visual data and meta preliminary data cleaning is not required, since it was produced at the stage of combining two repositories into one common dataset.

### 5.1.2. Data transformation

To use text, visual, and meta data in CNN deep learning models, a preliminary data transformation is required.

Transformation for text data includes tokenisation or splitting text into individual words, transform each text to a sequence of integers and also transform the resulting list of sequences into tensor or multi-dimensional array. After applying the above procedures, the shape of train tensor for news articles is (3014, 54759) and (3014, 44419), where 3014 is the number of records in the dataset, and 54759 and 44419 are the maximum sequence lengths for news articles and tweets, respectively.

Transformation for image data includes decode the JPEG content to RGB grids of pixels with size (224, 224, 3), convert these into floating-point tensors for input to neural nets, rescale the pixel values (between 0 and 255) and extract the mean [103.939, 116.779, 123.68] for VGG16 pre-process input. To transform the metadata, scaling for all numerical meta data features to the interval [0,1] was carried out.

### 5.1.3. Data splitting

The dataset was separated into two parts, one part is used as training dataset to produce the prediction model, and the other part is used as test dataset to test the accuracy of our model. Since empirical studies show that the best results are obtained if we use 20-30% of the data for testing, and the remaining 70-80% of the data for training; then, in this study, we will split our dataset into 80:20 ratio where 80% for training and 20% for testing. For the COVID-19 multimodal dataset, this means that 2411 records are used as training data and 603 records are used as test data.

## 5.2. Feature engineering

Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data [23]. This thesis uses such feature extraction techniques as word2vec and VGG16, which will be described in more detail below.

### 5.2.1. Word2vec

Word2vec is a two-layer neural network that processes text to obtain linguistic context using vectors. It takes a text corpus as input, constructs a vocabulary of words from the text and produces the vocabulary of word vectors as output [24]. The purpose and usefulness of word2vec is to group the vectors of similar words together in vector space [25]. Using word vectors as discrete states, word2vec predict target context and analyses the probability that words can occur simultaneously in a text.

The advantages of using word2vec technology include good performance with the help of pre-trained google-news vectors [24], using which separate embedding matrices were built for both news and tweets, with total embedded 31169 and 16050 common words.

These embedding matrices are in turn used as weights in Embedding layers in CNN text models. Also, due to its simpler architecture, word2vec is faster with comparable efficiency than its counterparts in the form of BERT and GloVe.

### 5.2.2. VGG16

VGG16 is a special CNN model whose main task is to classify images with maximum accuracy. VGG16 consists of 16 convolutional layers and 138 million parameters. VGG16 can classify images into 1000 object categories, including keyboard, animals, pencil, mouse, etc [26]. Additionally, the model has an image input size of 224-by-224. According to Opengenius[1] VGG16 is currently the most preferred choice in the community for extracting features from images. Like word2vec, VGG16 has a pre-trained weight configuration based on the ImageNet dataset, which provides better accuracy and faster model compilation.

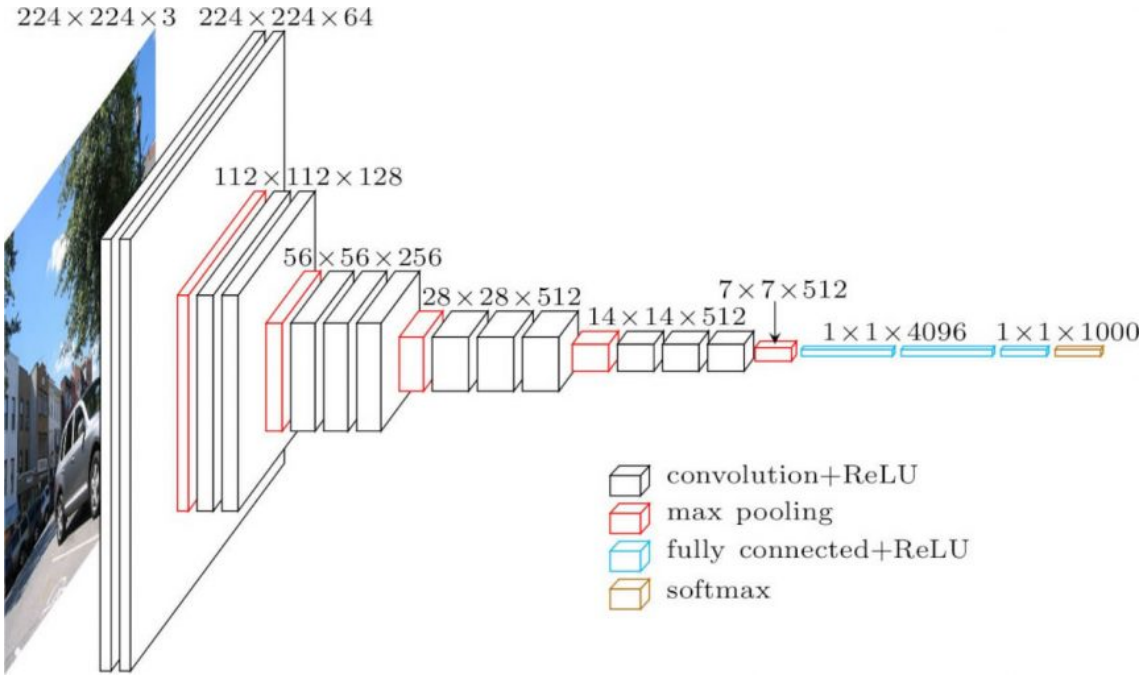


Figure 6. VGG16 Architecture[2].

The model input is 224 x 224 RGB image. The image is passed through a stack of convolutional layers, where the filter size is 3x3 and the kernel size is gradually reduced from 224x224 to 14x14, followed by max pooling layers of 2x2. All convolutional

[1] https://iq.opengenus.org

[2] https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/

layers have relu activation function. Three fully-connected layers follow a stack of convolutional: the first two have 4096 channels each, the third performs 1000-way classification and thus contains 1000 channels (one for each class) [27]. The final fully connected layer has softmax activation with 1000 nodes.

## 5.3. Models training using deep learning techniques

The model training uses a deep neural network approach such as convolutional neural networks, as well as transfer learning and a multimodal approach, which will be described later in this Section.

### 5.3.1. Convolutional neural networks

Convolutional neural network or CNN is part of deep learning technologies and is a special architecture of artificial neural networks. Although convolutional neural networks (CNNs), originally invented for computer vision, have been shown to achieve strong performance on text classification tasks as well as other traditional Natural Language Processing (NLP) tasks, even when considering relatively simple one-layer models [28]. The difference between the CNN architecture and other deep learning approaches is that instead of neurones and weights, filtering layers are used, such as convolutional layers, pooling layers and fully connected layers to analyse the input data.

The innovation of convolutional neural networks is the ability to automatically learn a large number of filters in parallel specific to a training dataset under the constraints of a specific predictive modelling problem [29]. Among the advantages of CNNs over other neural networks is the computational efficiency due to convolution and pooling operations, thereby transforming data into a form that is easier to process, without losing features that are critical for making a good prediction. For this reason, the CNN architecture outperforms other deep learning approaches such as ANN and RNN in terms of the speed of building various models, which is extremely important in this thesis, where complex multimodal models are compiled. In addition, CNN has high accuracy in image and text recognition tasks, and automatic detection of the important features. Also, CNN does not process data in the forward direction, but accesses the same data several times.

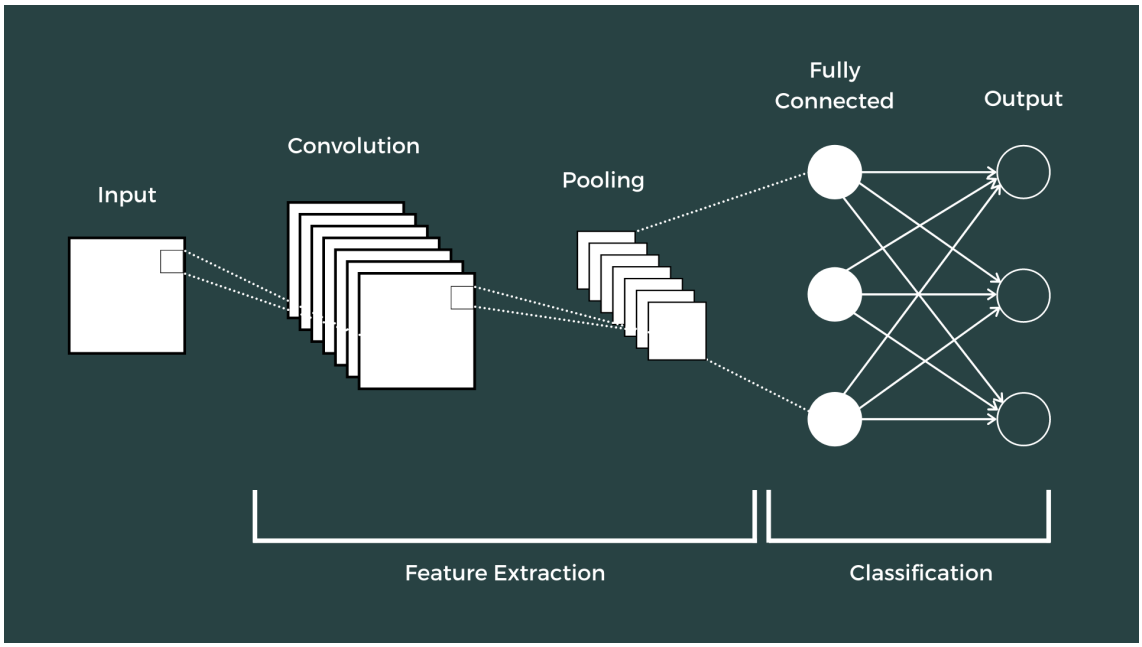


Figure 7. Convolutional neural network architecture[1]

The task of the convolutional layer is to take input and perform a series of convolution operations. In turn, the main goals of the pooling layer are to reduce computational tasks by reducing the number of parameters and reducing the likelihood of model retraining. The fully connected layer is responsible for taking a vector as input and finding a probability score for each label in the training data [30]. The first two layers (convolution and pooling) are responsible for feature extraction and can be repeated, while the last fully connected layer is responsible for classification.

This thesis uses two types of convolutional neural networks: one dimensional convolutional neural network (1D-CNN) architecture for text and meta modality and two dimensional convolutional neural network (2D-CNN) architecture for visual modality.

### 5.3.2. Transfer learning in deep learning

Transfer Learning is a machine learning method, that also used in deep learning where we reuse a pre-trained model as the starting point for a model on a new task [31]. Domains like natural language processing and image recognition are considered to be the hot areas of research for transfer learning [31]. In this thesis, pre-trained models as

[1] https://www.theclickreader.com/building-a-convolutional-neural-network/

VGG16 for visual modality and word2vec for text modality are used. The pre-trained models are trained on a large and general enough dataset and will effectively serve as a generic model and the key idea here is to leverage the pre-trained model's weighted layers to extract features [31].

In order to improve the accuracy of the obtained models, the fine-tuning method is used for text CNN models, since retraining or fine-tuning allows specialized features to better adapt to work with the new dataset. The use of fine-tuning in this case is extremely necessary due to the specific topic of COVID-19 and the fact that the word2vec pre-trained model dating from 2014 simply lacks some pre-trained vectors related to it.

### 5.3.3. Multimodal approach

The goal of multimodal deep learning is to create models that can process and link information using various modalities. Multimodal learning helps to understand and analyse better when various senses are engaged in the processing of information [32].

A more detailed multimodal approach and mono-modal and multimodal models architecture is presented in Section 6.

## 5.4. Model testing and data validation

In this thesis, dealing with the binary classification problem, since one input sample belongs to one of two classes - reliability = 1 or reliability = 0, the most common metrics, such as accuracy, confusion matrix, precision, recall, f1-score, as well as model accuracy and loss curves are used.

### 5.4.1. Confusion matrix

The confusion matrix is a 2x2 table that is often used to describe the performance of a model and explaining the probabilities of binary classification. The confusion matrix helps to visualise how accurate the model is at distinguishing two classes. It provides information not only about the errors made by the classifier, but also about the types of errors made.

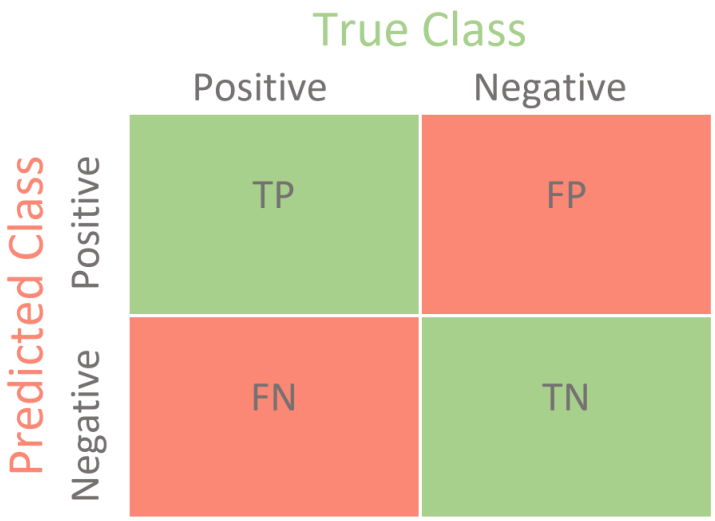


Figure 8. Confusion matrix example[1].

The top-left cell of the confusion matrix corresponds to True Positive or TP results. Indicates the number of times the model correctly classified a positive result correctly. Or in our case, how many reliability = 0 were classified correctly.

The top-right cell of the confusion matrix corresponds to False Positive or FP results. Indicates the number of times the model incorrectly classified a positive result as a negative result. Or how many reliability = 0 were misclassified as reliability = 1.

The bottom left cell corresponds to False Negative or FN results. Indicates the number of times the model incorrectly classified a negative result as positive. Or how many reliability = 1 were misclassified as reliability = 0.

The bottom right - True Negative or TN. Correctly classified negative results, or correctly classified reliability = 1 as reliability = 1.

### 5.4.2. Accuracy

Accuracy is a metric that describes how often a classifier correctly predicts the correct result, or percent of accurate prediction across all predictions. It's proportion of True Positive (TP) and True Negative (TN) in all evaluated classes.

[1] https://emilia-orellana44.medium.com/breakdown-confusion-matrix-2cf25842f1ae

Described by the formula:

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

where TP, TN, FP and TN are parameters detailed in the confusion matrix.

### 5.4.3. Precision

Another important metric is Precision. The precision describes how many detected items are truly relevant. Precision is the ratio between the True Positives (TP) and all the Positives and is calculated by the formula:

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

### 5.4.3. Recall

Recall is a metric that shows how many True Positives were detected. It is the ratio between the number of True Positives (TP) correctly classified as True Positives (TP) to the total number of Positives and is calculated by the formula:

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

### 5.4.4. F1 Score

The F1 Score is single metric denoting combination of Precision and Recall taking their harmonic mean. It is designed to be useful metric when classifying between unbalanced classes or other cases when simpler metrics could be misleading. F1 Score is calculated by the formula:

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

### 5.4.5. Learning curves

During the training of a model, the current state of the model at each step of the training algorithm can be evaluated. The shape and dynamics of a learning curve can be used to

diagnose the behaviour of a model. [33] The learning curves used in this thesis are divided into two types: optimisation learning curves used to estimate model loss and performance learning curves used to evaluate model accuracy.

The common dynamics that can be observed in learning curves are: underfit, overfit and good fit [37].

# 6.Multimodal architecture and tools

This section will describe the mono-modal and multimodal CNN architecture used for COVID-19 fake news detection. This section also describes the main tools and libraries, that are used in the experiment.

## 6.1. Modalities

This thesis focuses on multimodal CNN approach for COVID-19 fake news detection. Multimodal learning research has typically focused on developing models that combine multiple modes of data with varying structures such as sequential relationships between words in natural language and spatial pixel relationships in images [18]. The scope of multimodal approaches today is quite extensive and includes, for example, multimodal deep learning to predict movie genres[1] or house price estimation[2] from visual and textual features. Also, besides textual and visual data, multimodal approaches also use various numerical and categorical meta data, as well as audio and video modalities.

But mono-modal models are also used within the framework of the experiment. All models presented below were built using the 1D and 2D CNN from Section 5.3. The dataset used is described in more detail in Section 4 and consists of four modalities:

- COVID-19 News from reliable and unreliable sources or not fake and fake news - textual modality
- Images posted in these news - visual modality
- Tweets that link to this news - textual modality
- General meta-information on tweets - meta modality

Further, all the above models and techniques used to build them separately (in a mono-modal form). In addition, multimodal pairs and all four data modalities into one multimodal network will be presented.

---

[1] https://towardsdatascience.com/multimodal-deep-learning-to-predict-movie-genres-e6855f814a8a

[2] https://pyimagesearch.com/2019/02/04/keras-multiple-inputs-and-mixed-data/

### 6.1.1. Textual modalities architecture

Using the convolutional neural networks and word2vec pre-trained embeddings, a general architecture was developed for the two text modalities of news and tweets, shown in Figure 9 below.

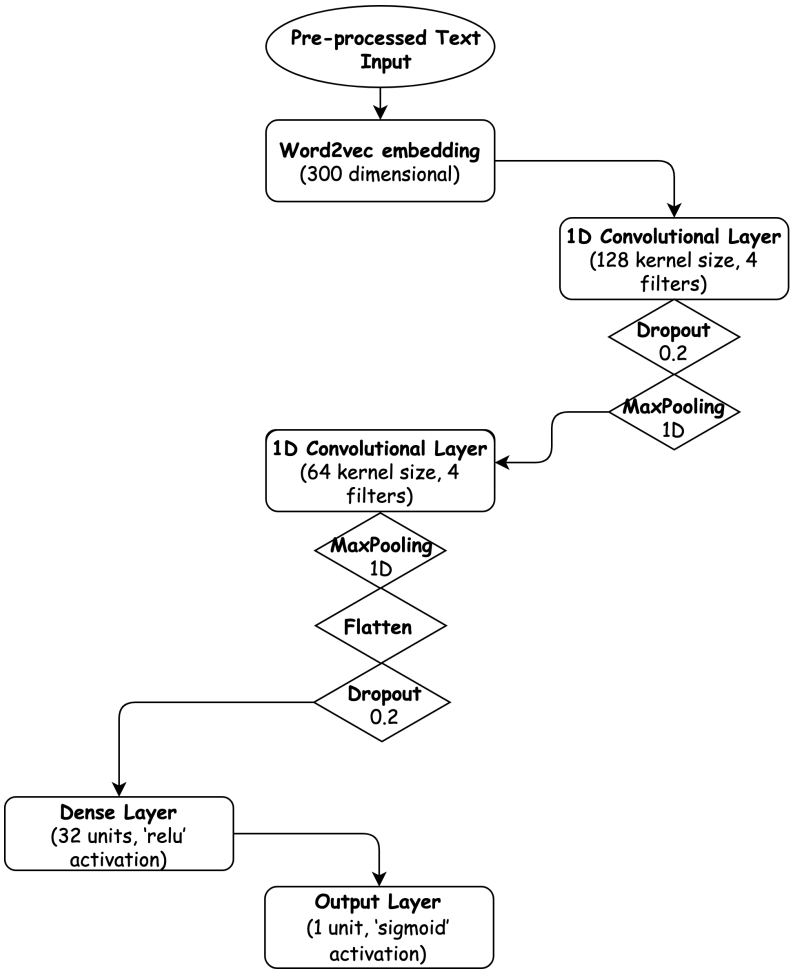


Figure 9. General textual architecture.

The text model consists of a pre-processed text input whose pre-processing methods are described in more detail in Section 4. This is followed by a 300 dimensional word2vec embedding layer that uses the pre-trained weights of the word2vec model. The number of common words in the word2vec embedding matrix is different for each of the models - for news it is 31169, for tweets 16050. After the word2vec embedding layer, two hidden 1D convolutional layers follow with a kernel size of 128 and 64, 4 filters and relu activation with a dropout rate of 0.2. After that, maxpooling with pool size 4 after

each of the models and flatten after the second convolutional layer. At the end, one fully connected layer with 32 nodes and output layer with 1 node.

### 6.1.2. Visual modality architecture

To build a visual modality model, a 2D convolutional neural network, pre-processed image input and VGG16 pre-trained model are used. Thus, the following architecture for the visual modality was developed, shown in Figure 10.

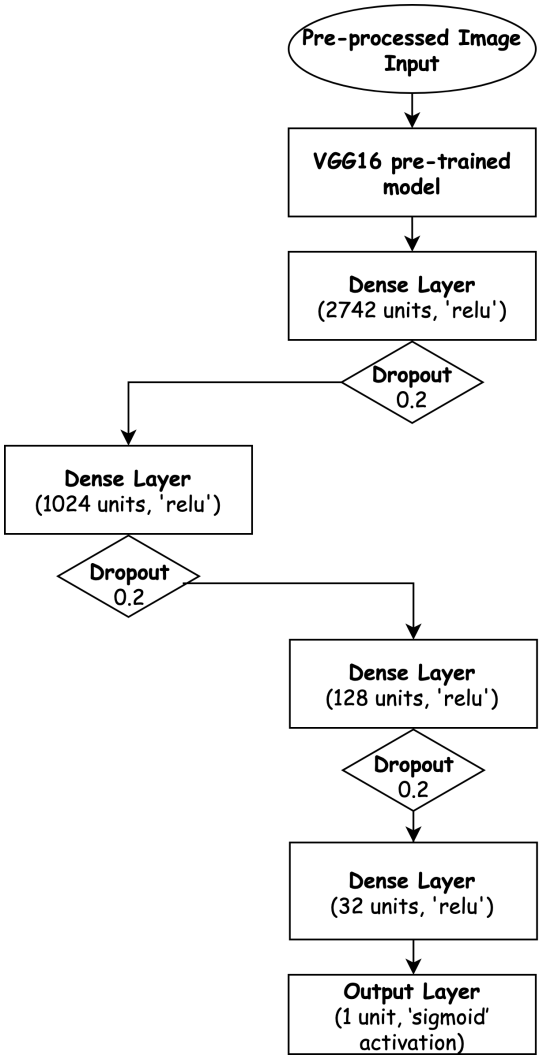


Figure 10. Visual architecture.

In addition to pre-processed image input and VGG16 pre-trained models, the presented visual model consists of three fully connected dense layers with dropout 0.2, as well as one more dense layer before fully connected output with 32 channels and relu activation

function. The final fully connected output layer has sigmoid activation with 1 node. This architecture provides the best accuracy and reduces overfitting.

### 6.1.3. Meta modality architecture

A one-dimensional convolutional neural network and a pre-processed meta-input are used to develop the meta modality architecture. Since convolutional neural networks are not limited to text and image input, this deep learning technique can also be used for a meta modality consisting of numeric data. In this case, no pre-trained model is used.

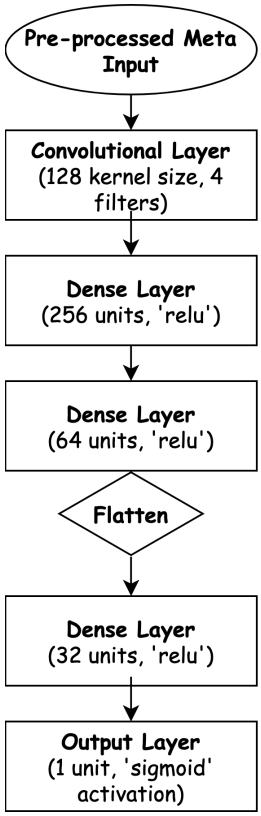


Figure 11. Metadata architecture.

The meta model consists of pre-processed numeric input. This is followed by a convolutional layer with 4 filters and a kernel size of 128. As well as 3 fully connected dense layers with 256, 128 and 32 channels and a relu activation function. Between the second and third fully connected layers is the flatten function. The final fully connected output layer has sigmoid activation with 1 node.

### 6.1.3. Multimodal pairs architecture

The mono-modal architecture presented earlier in this thesis were combined into logically related multimodal pairs: news articles and images, as well as tweets and meta -information.

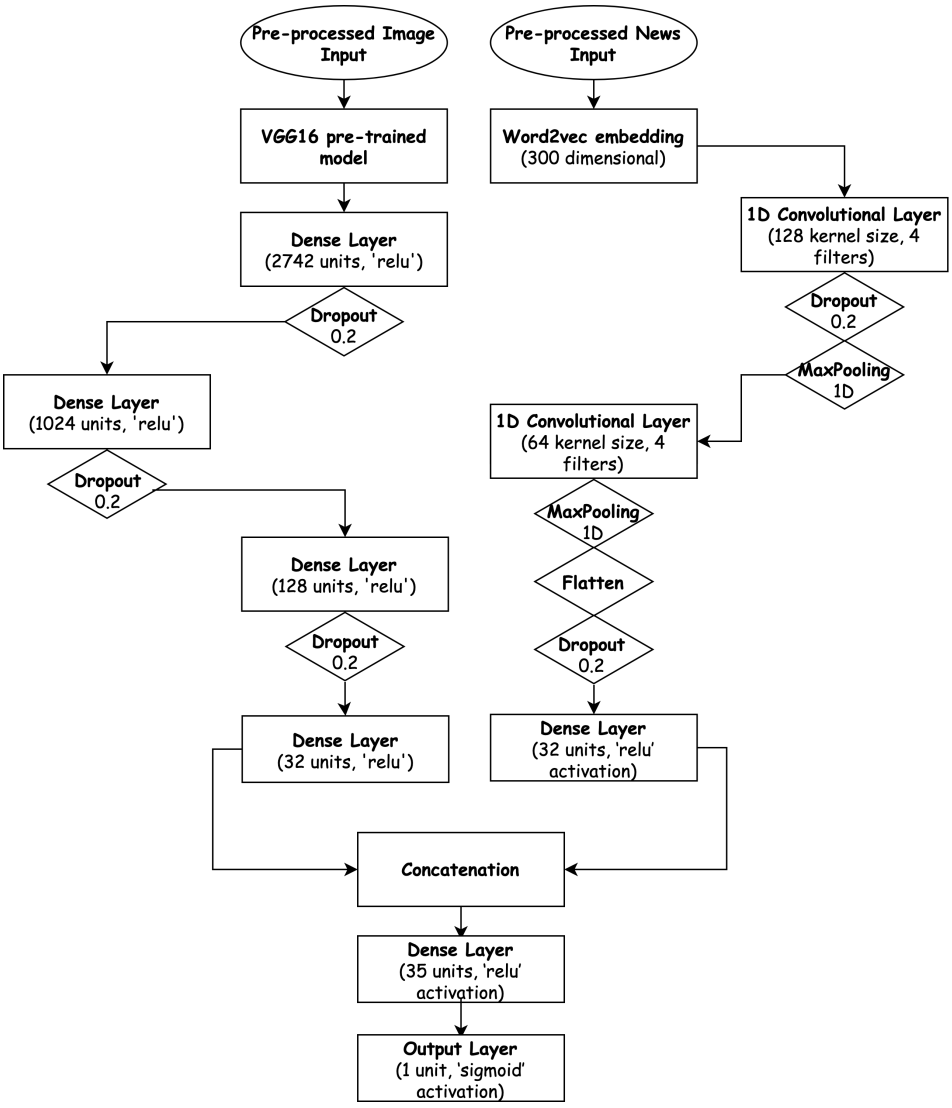


Figure 12. News articles and images multimodal architecture.

Combining two mono-modal news and images models into one multimodal one is due to the operation of concatenation of their weights outputs from the last fully connected dense layers of each of the models. Next, the multimodal model is followed by a fully connected layer with 35 channels and a relu activation function. The final layer is output layer with 1 unit and sigmoid activation function.

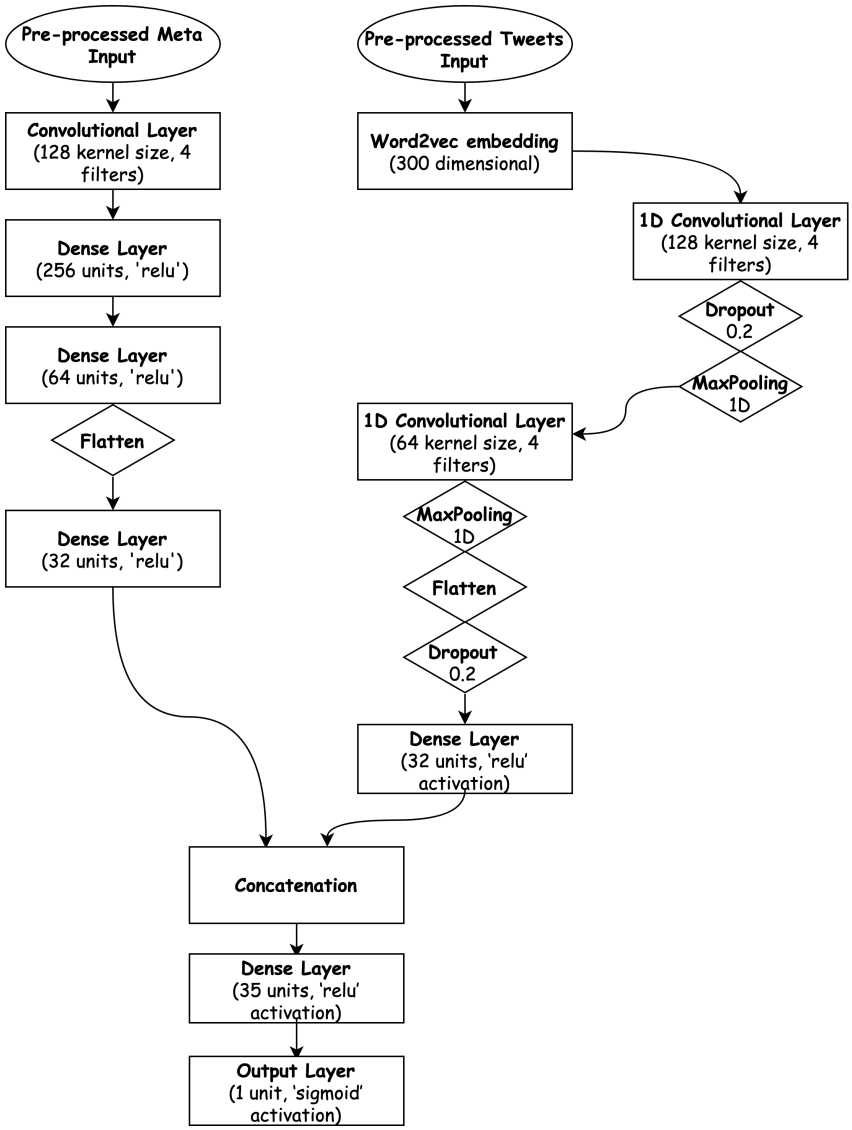


Figure 13. Tweets and metadata multimodal architecture.

The architecture and concatenation of the two tweet and metadata models into one multimodal pair is exactly the same as for the news and image models detailed above.

### 6.1.4. Multimodal network architecture

To combine all four mono-modal models into one multimodal network, the same principle was used as when combining into multimodal pairs.

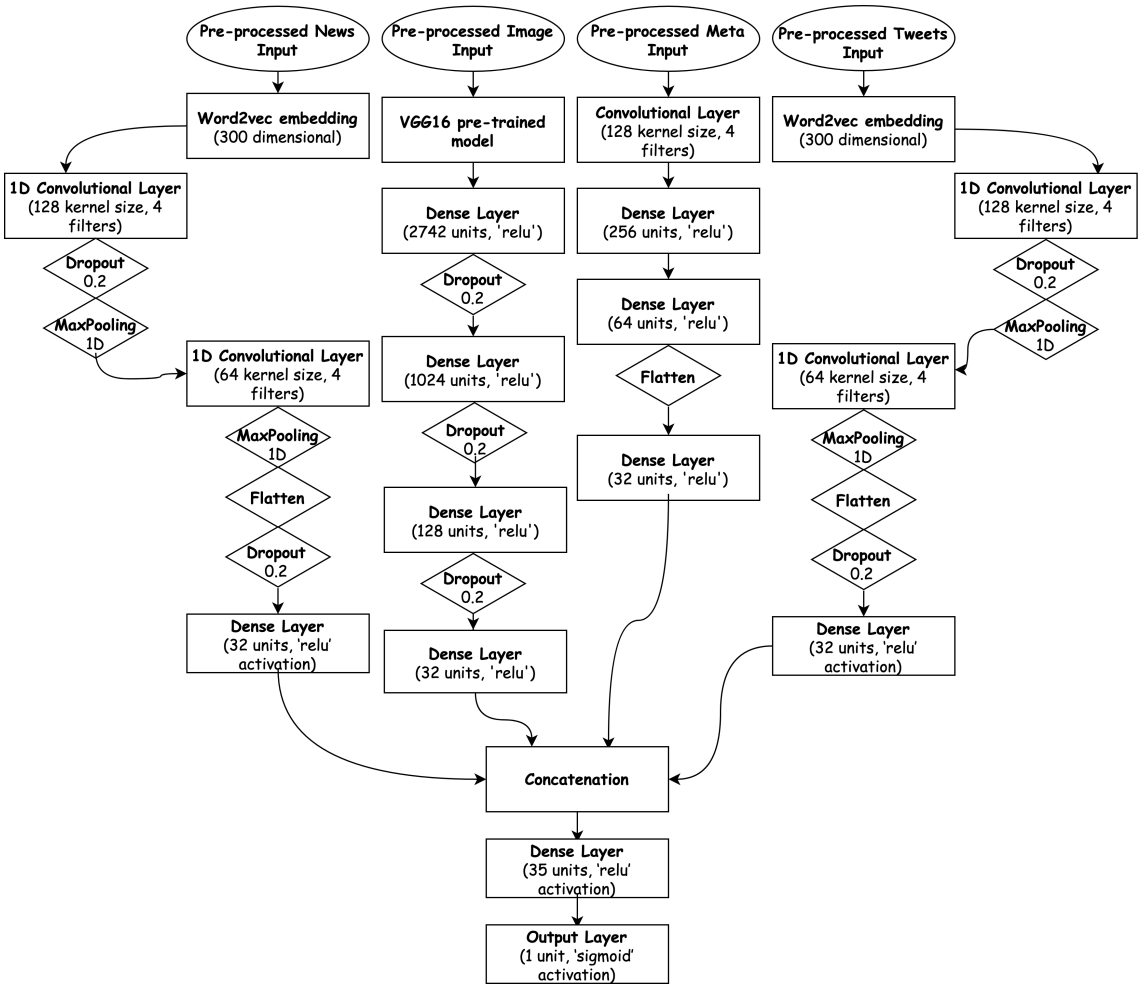


Figure 14. Multimodal network architecture.

By concatenating the weights of the last fully connected layers of four models: news, images, tweets, and meta-information, a multimodal network can be created. After concatenation, this multimodal network has one fully connected layer with 35 channels and relu activation, and the final layer with 1 unit and sigmoid activation.

## 6.2. Tools

This thesis uses Google Colaboratory, the Python programming language and its libraries: Numpy, Pandas, OpenCV, Keras, Pickle and Tensorflow

### 6.2.1. Google Colaboratory

Google Colaboratory[1] or "Google Colab» for short, is a developed by Google free Jupyter notebook environment and research tool that that allows to write and execute

[1] https://colab.research.google.com

Python code in the browser. It requires a serverless environment for interactive development of machine learning and deep learning models with no setup and runs entirely in the cloud. In addition, Google Colaboratory supports many popular machine learning Python libraries as well as powerful GPUs.

### 6.2.2. Python

Python is one of the leading programming languages in machine learning area for its extensive collection of libraries and packages, simple syntax, code flexibility and readability. As of Q3 2021 according to SlashData[1], Python is used by over 70% of machine learning developers and data scientists. Python is the preferred programming language of choice for machine learning for some of the giants in the IT world including Google, Instagram, Facebook, Dropbox, Netflix, Walt Disney, YouTube, Uber, Amazon, and Reddit [34].

In fact, as of April 2022, Google Colab is using Python 3.7, so this particular version of Python and some of the most widespread Python libraries below will be used in this thesis:

- **Numpy and Pandas** for data pre-processing and analysis. These libraries allow merging and filtering of data, gathering from different external sources.
- **OpenCV** for image data pre-processing.
- **Keras** for CNN deep learning architecture. It allows fast calculations and prototyping, as it uses the GPU in addition to the CPU of the computer.
- **Tensorflow** for working with deep learning by setting up and training.
- **Pickle** to save and load complex objects such as pre-processed image data and pre-trained word2vec embedding matrices.

[1] https://aster.cloud/2021/12/09/top-programming-languages-most-popular-and-fastest-growing-choices-for-developers/

# 7.Experiments & results

This section provides section training model parameters and the main results of the metrics obtained during the experiment, using the tools described in Section 6.2. The metrics include the confusion matrix, accuracy, precision, recall, f1-score and learning curves. In addition, an appropriate analysis of each model was performed. All presented results are obtained on a test data.

## 7.1. Training model parameters

In this experiment, the same training model parameters are used for training all models, both mono-modal and multimodal.

Table 2. Training model parameters

| Parameter name | Value |
|---|---|
| Number of epochs | 20 |
| Batch size | 16 |
| Optimizer | Adam |
| learning rate | 0.0005 |
| beta 1 | 0.9 |
| beta 2 | 0.999 |
| epsilon | 1E-08 |
| decay | 0.0 |

Number of epochs means the number of complete passes through the training dataset [38]. Batch size is the number of training examples utilised in one iteration [39]. This is followed by the Adam optimisation algorithm for gradient descent with appropriate learning rate, beta1, beta2, epsilon and decay parameters.

Binary crossentropy is used as a loss function. In the context of this experiment, binary cross entropy compares each of the predicted probabilities to actual class output which can be either unreliable (0) or reliable (1). It then calculates the score that penalises the probabilities based on the distance from the expected value [36].

Methods such as earlystopping and L2 regularization are used to prevent overfitting of models. Earlystopping allows model training to be completed if the model stops learning. The main indicator for this is test data loss, and if more than two epochs in a row loss does not decrease, then earlystopping occurs. L2 regularisation is used in datasets with complex features, combats overfitting by making the weights small and thus solving multicollinearity problems, which is certainly important for data represented by several different modalities.

Random baseline plays an important role in assessing the success and accuracy of the model in this experiment. Having a dataset with unbalanced classes with 809 fake news and 2194 reliable news, a simple random guessing method can be inefficient, and therefore ZeroR or Zero Rule is used as a random baseline. This method predicts all values for the largest class. In this case, this means that the method will predict all records in the dataset as reliable or 1. In the end, 2194 records out of 3014 will be predicted correctly, which is 73%. The accuracy of training models on the test data is above 73% mean that model learns from extracted features.

The reproducibility of the results is also important. In order to get similar results every time the models are trained, this thesis uses the approach described in the Keras FAQ entitled “How can I obtain reproducible results using Keras during development?”[1].

## 7.2. Mono-modal CNN models

The mono-modal models are trained using the CNN architecture presented in Section 6.1. For each modality, its own specific model architecture is used.

### 7.2.1. News article model

For the text modality of news articles, pre-trained word2vec embeddings are used. Based on the learning curves of the model, the following conclusions can be drawn: the increase in accuracy on the test data is quite linear, ranging between 84 and 86% and gradually increasing with the number of epochs with overfitting at the end. Model loss

[1] https://keras.io/getting_started/faq/#how-can-i-obtain-reproducible-results-using-keras-during-development

is also linear with small peaks ranging from 0.7 to 0.5. A visual representation using learning curves is shown in Figure 15.

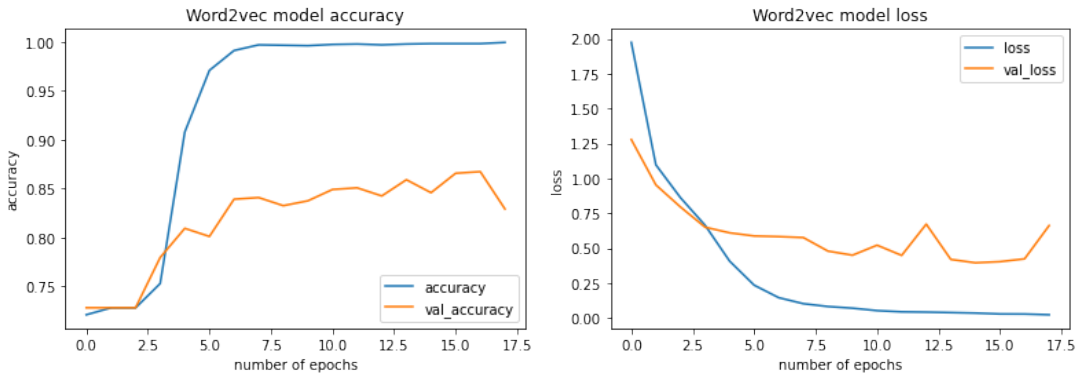


Figure 15. News article mono-modal model learning curves.

In this experiment, the best CNN news article mono-modal model is the one with the lowest test loss. This approach is also applicable to all other models presented in this Section. The resulting model managed to achieve the following results for the test dataset in the main metrics presented in Table 3.

Table 3. News article model main metrics

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fake news | 84.58% | 74.39% | 70.52% | 72.40% |
| Not fake news | 84.58% | 88.38% | 90.23% | 89.30% |

First of all, it should be noted that the random baseline of accuracy was successfully exceeded. This fact specifies that the training model is successful and the model learns from extracted features. In contrast to accuracy, there is some misbalance in the rest of the metrics, relative to the two classes presented (fake and not fake news). Concerning not fake news, the results of precision, recall and f1-score metrics show results close to 90%. While fake news has significantly lower results, ranging from 70% to 74%. A more detailed classification probability distribution is presented in the confusion matrix in Figure 16.

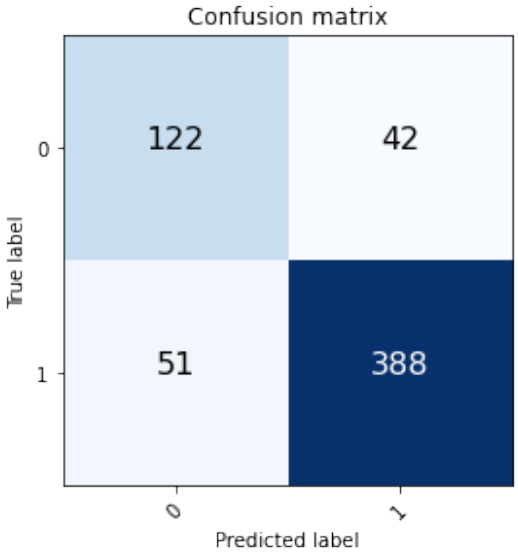


Figure 16. News article model confusion matrix

In this case, the model correctly predicted 122 COVID-19 fake news and 388 non-fake news. But incorrectly predicted 42 news articles that were fake but were predicted as not fake. Also, this model incorrectly predicted 51 not fake news that were predicted as fake.

### 7.2.2. Tweet-based model

For the text modality of tweets, pre-trained word2vec embeddings are used. Compared to news article model, this CNN tweet-based mono-modal model has a smoother linear increase in accuracy and a decrease in loss. This is shown in Figure 17 below.

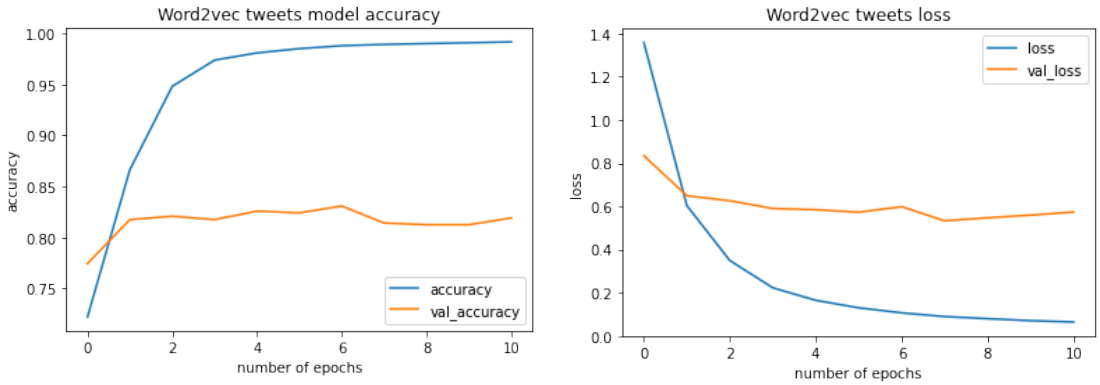


Figure 17. Tweet-based mono-modal model learning curves.

The tweet-based model after the first epoch has a noticeable increase in accuracy and a drop in loss, and then a „plateau" follows, with almost constant values relative to accuracy of approximately 82% and a loss of 0.6.

Table 4. Tweet-based model main metrics

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fake news | 81.43% | 62.20% | 67.11% | 64.56% |
| Not fake news | 81.43% | 88.61% | 86.25% | 87.41% |

Like the news article model, the accuracy of the tweet-based model outperforms the random baseline, and also has similar trends regarding the classification of fake and not fake news. In terms of accuracy, this model is more than 3% inferior to the news article model. As for the rest of the metrics, the results of predictions of not fake news in this case are also much higher than fake ones.

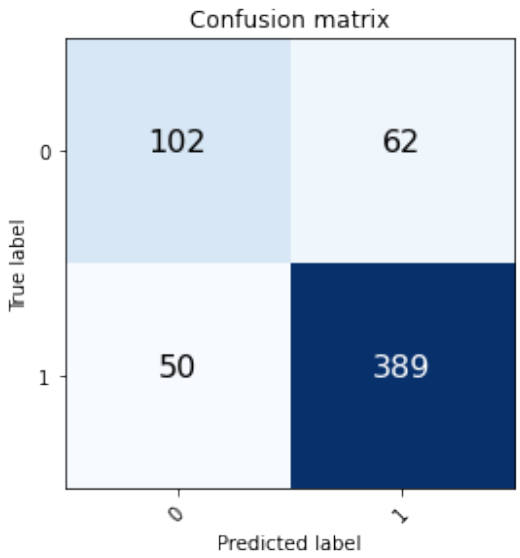


Figure 18. Tweet-based model confusion matrix

Using only the tweet-based model, it is possible to predict an almost equal amount of not fake COVID-19 news compared to the news article model (389 vs 388). In addition, both models make an equal number of errors in predicting not fake news as fake.

Although the number of correctly predicted COVID-19 fake news is much lower at only 102 cases.

### 7.2.3. Image-based model

For the visual modality of images, VGG16 pre-trained weights are used. As can be seen in Figure 19 below, the image-based CNN mono-modal model has a clear tendency to overfitting, which may be due to an insufficient set of test images for classification, as well as the lack of specific features that distinguish between images used in reliable and unreliable COVID-19 news articles.

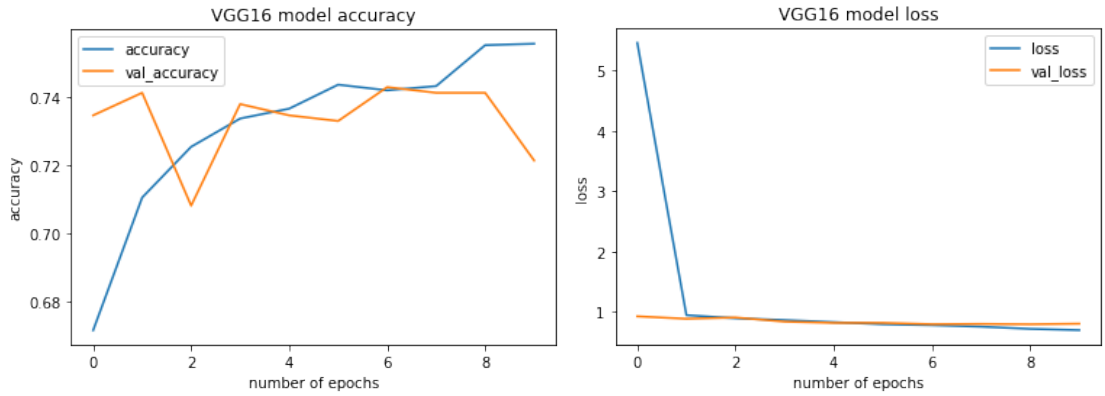


Figure 19. Image-based mono-modal model learning curves

Almost regardless of epoch, the image-based model accuracy on the test data is between 72 and 74%, which corresponds to the minimum value of the random baseline. And also the model loss is almost always close to 1. More accurate results of the main image-based visual model metrics are presented in Table 5.

Table 5. Image-based model main metrics

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fake news | 74.30% | 12.20% | 65.52% | 20.51% |
| Not fake news | 74.30% | 97.49% | 74.83% | 84.67% |

Relative to other previously presented mono-modal models, the image-based model has an accuracy only slightly higher than the random baseline with a result of 74.30%. In

addition, anomalies are observed in other metrics. Precision for not fake news is 97.49%, while for fake news it is only 12.20%. The same imbalance in the results applies to the f1-score metric (84.67% vs 20.51%). This may indicate that the model predicts almost all instances for the largest class corresponding to a reliability of 1. Image-based model classification probability distribution is presented in the confusion matrix in Figure 20.

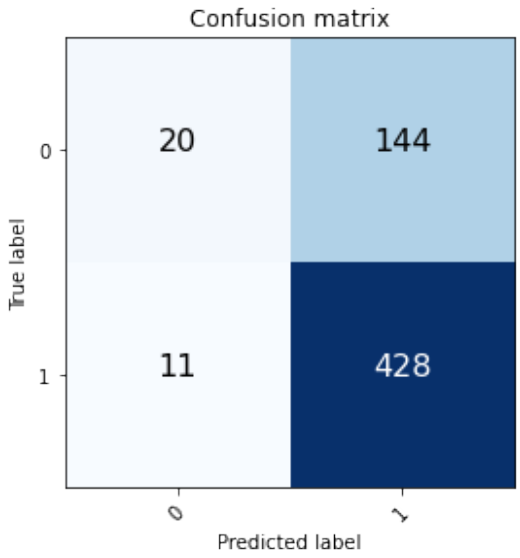


Figure 20. Image-based model confusion matrix

In this case, the confusion matrix gives the clearest idea of how correct and incorrect predictions are distributed. This model defines almost all news articles as reliable or not fake. While 31 COVID-19 news articles are classified as fake by the image-based model, only 20 of them are actually fake. At this moment, image-based mono-modal model in the most inaccurate of the presented models. And one can make an unambiguous conclusion that only due to the images in the COVID-19 news articles it is impossible to accurately determine whether this news is fake or not.

### 7.2.4. Tweet meta-based model

Like the image-based visual model, the accuracy of the tweet meta-based CNN mono-modal model are only slightly better than the random baseline value of 73%, as shown in Figure 21.

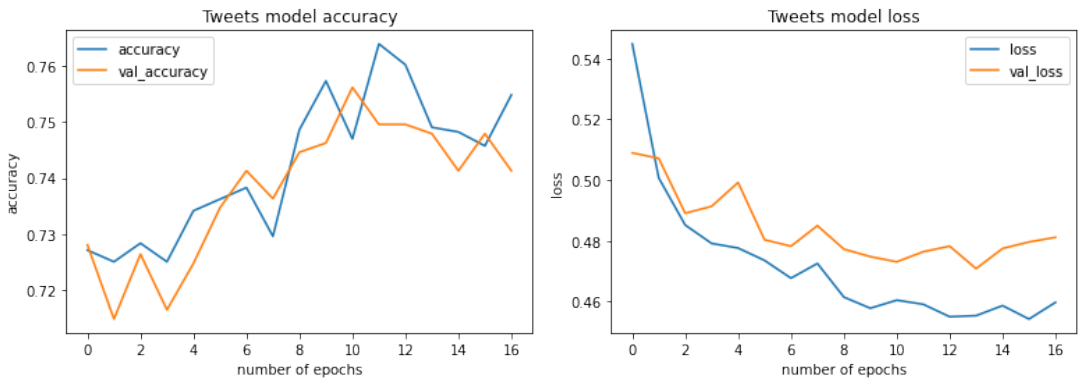


Figure 21. Tweet meta-based mono-modal model learning curves

Accuracy of the model has a fluctuating shape and has values from 72 to 75%, increasing slightly with each epoch for both the training data and the test data. Although the model loss graph is more linear, the decrease in model loss is generally insignificant and amounts to only 0.02-0.03 for the test data set and 0.08 for the main data set over 17 epochs.

Table 6. Tweet meta-based model main metrics

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fake news | 74.79% | 54.88% | 53.57% | 54.22% |
| Not fake news | 74.79% | 82.23% | 82.99% | 82.61% |

As in all mono-modal CNN models presented earlier, the precision, recall, and f-1 score metrics have much higher results in relatively not fake news. And although for not fake news the values of these metrics are 82-83%, for fake news it is only 54-55%. In general, such a small spread in the metric values for each individual class may indicate that the number of False Negative and False Positive instances is approximately equal. The visual representation of this statement can be seen in Figure 22.

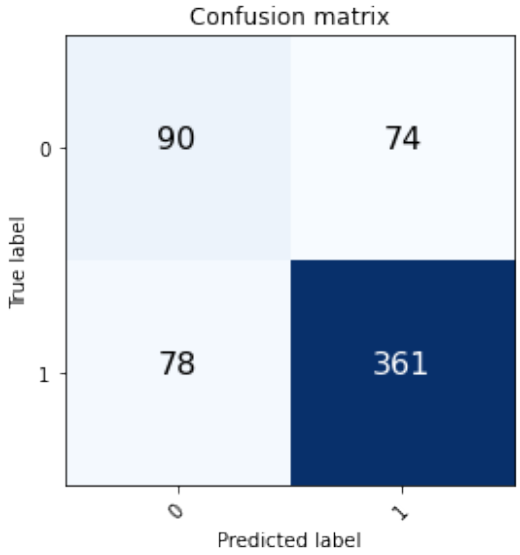


Figure 22. Tweet meta-based model confusion matrix

In the given confusion matrix, it can be determined that the model has a much more balanced confusion matrix than the visual-based model, but in general it is inferior in terms of the number of correctly predicted both fake and not fake news to text mono-modal models. Only 361 not fake and 90 fake news articles out of a total of 603 COVID-19 news articles can be correctly predicted using tweet meta information alone.

### 7.2.5. Mono-modal models general table

In the following general Table 7 presents all mono-modal models and their results. A common visual presentation of all results in a single table allows to determine the most effective CNN mono-modal model for COVID-19 fake news detection.

Table 7. General table of all mono-modal models

| Model | Correctly classified instances | Incorrectly classified instances | Accuracy | Precision (%fake/ not fake) | Recall (%fake/ not fake) | F1-Score (%fake/ not fake) |
|---|---|---|---|---|---|---|
| News | **510** | **93** | **84.58%** | **74.39**/88.33 | **70.52**/**90.23** | **72.40**/**89.30** |
| Tweets | 491 | 112 | 81.43% | 62.20/88.61 | 67.11/86.25 | 65.56/87.41 |
| Images | 448 | 155 | 74.30% | 12.20/**97.49** | 65.52/74.83 | 20.51/84.67 |
| Meta | 451 | 152 | 74.79% | 54.88/82.23 | 53.57/82.99 | 54.22/82.61 |

Not taking into account the anomaly that occurred in the precision of the image-based model for all other results obtained, the best unimodal CNN model for COVID-19 fake news detection is the news article model. According to the results, the news article model is able to correctly classify 510 instances and incorrectly classify 93 instances.

## 7.3. Multimodal CNN pairs

The multimodal CNN pairs are trained using the CNN architecture presented in Section 6.1. For each pair, its own specific model architecture is used. This section will also compare multimodal pairs with the best mono-modal model.

### 7.3.1. News article and image-based multimodal pair

Presented multimodal CNN pair combines two models: news article model and image-based model. It is worth noting that this model has a similar model accuracy curve relative to the number of epochs and a smoother model loss curve compared to the mono-modal news article model. Visually, these results are presented in Figure 23.

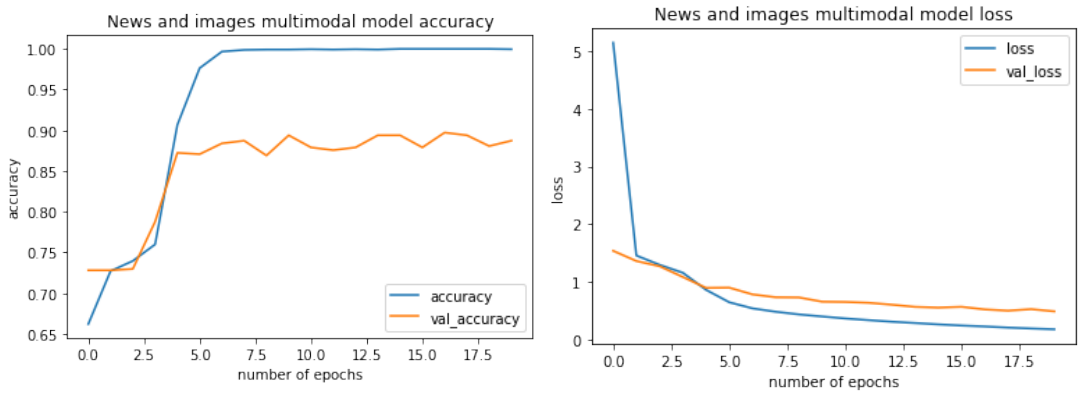


Figure 23. News and image-based multimodal pair learning curves

After the first epoch, the loss of the test model smoothly decreases from 1 to 0.5, and the accuracy increases to 86-89%. In terms of accuracy, as shown in Table 8, the multimodal pair model outperforms news article and image-based mono-modal models, but the model loss remains about the same.

Table 8. News and image-based model main metrics

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fake news | 88.72% | 68.29% | 87.50% | 76.71% |
| Not fake news | 88.72% | 96.36% | 89.05% | 92.56% |

The resulting multimodal pair outperforms the most efficient CNN mono-modal news article model and shows a significant gain in all metrics except precision. Accuracy of this model is higher than that of the best mono-modal model by 4.14%. As for the rest of the metrics, fake news recall is higher by 17%, and f1-score by 4.5% with a result of 76.71%. The improvement of the model concerns not only the prediction of fake news, but also not fake news - this is the only model presented so far that has an f1-score above 90%.

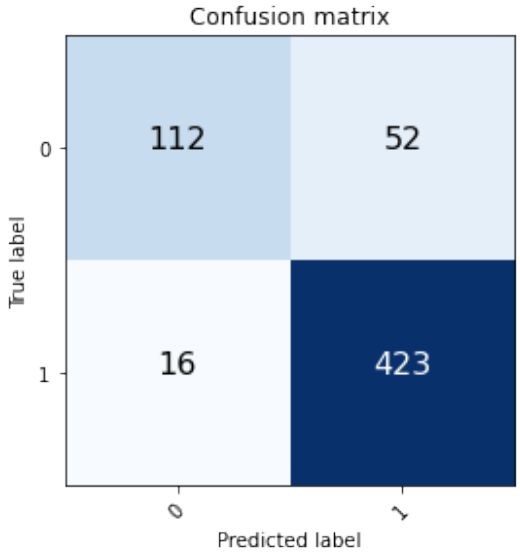


Figure 24. News and image-based model confusion matrix

The model shows much better efficiency in detecting not fake news than the best mono-modal model - their number increased from 388 to 423, and the number of non-fake news that the model predicted as fake was significantly reduced from 51 to 16.

### 7.3.2. Tweet and meta-based multimodal pair

The second multimodal CNN pair, consisting of tweets and metadata, has accuracy and loss curves similar to the tweet mono-modal model. This may be due to the small number of trainable parameters in the mono-modal meta data model. The multimodal pair has exactly the same „plateau" and susceptible to rapid overfitting and early stopping.

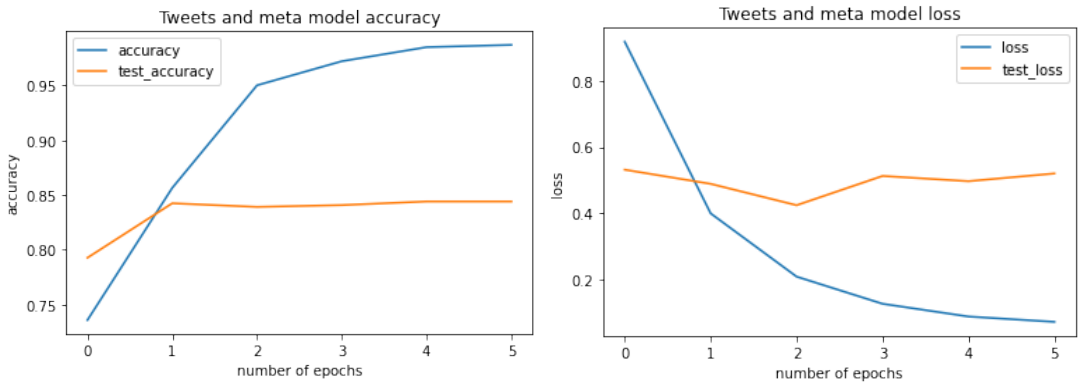


Figure 25. Tweet and meta-based multimodal pair learning curves

This one model has greater accuracy and comparable loss compared to the tweet mono-modal model. It is also comparable in accuracy to the best mono-modal news article CNN model, but generally inferior to the first multimodal pair. Accuracy of the model is on average in the range of 83 to 85%.

Table 9. Tweet and meta-based model main metrics

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fake news | 83.91% | 70.73% | 70.30% | 70.52% |
| Not fake news | 83.91% | 88.84% | 89.04% | 88.94% |

The resulting multimodal pair also shows an increase in all metrics relative to the best used mono-modal tweet model from this pair. The increase compared to the tweet model in terms of accuracy is 2.48%, f1-score increase for COVID-19 fake news is 4.96%, and f1-score increase for not fake news is 1.53%.

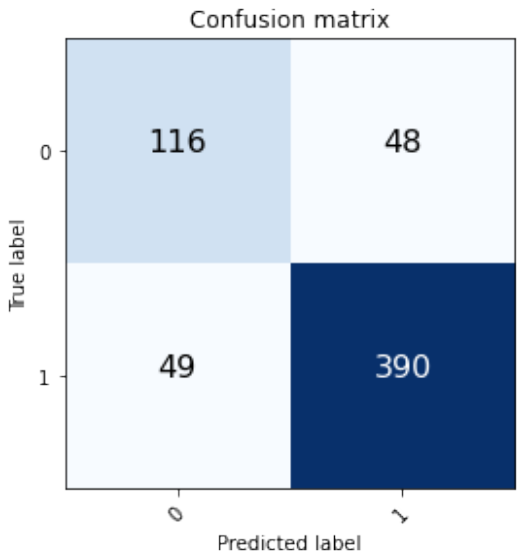


Figure 26. Tweet and meta-based model confusion matrix

The visual representation in the form of a confusion matrix shows that the model is very similar to the mono-modal news article model, having almost the same number of correctly and incorrectly classified instances. Thereby, the number of correctly

classified COVID-19 news is 506, compared to 510 for the mono-modal news article model.

Based on the above learning curves, common metrics and confusion matrices of two multimodal pairs, the following conclusions can be drawn:

- The multimodal news article and image pair is more efficient at predicting reliable news articles, and has significantly fewer not fake news mispredicted as fakes (16 vs 49).
- On the other hand, the number of correctly predicted fake news and incorrectly predicted fake news as not fake news is almost equal.
- For a multimodal pair of news article and images, all metrics for both COVID-19 fake news and not fake news have significantly higher values.

Accordingly, the best multimodal pair and the best model at the moment is the multimodal news and images model with 535 correctly classified instances and 68 incorrectly classified instances.

## 7.4. Multimodal network

The multimodal CNN network consists of all the modalities presented in this thesis concatenated into a single model. The architecture of this multimodal network is presented in Section 6.1.

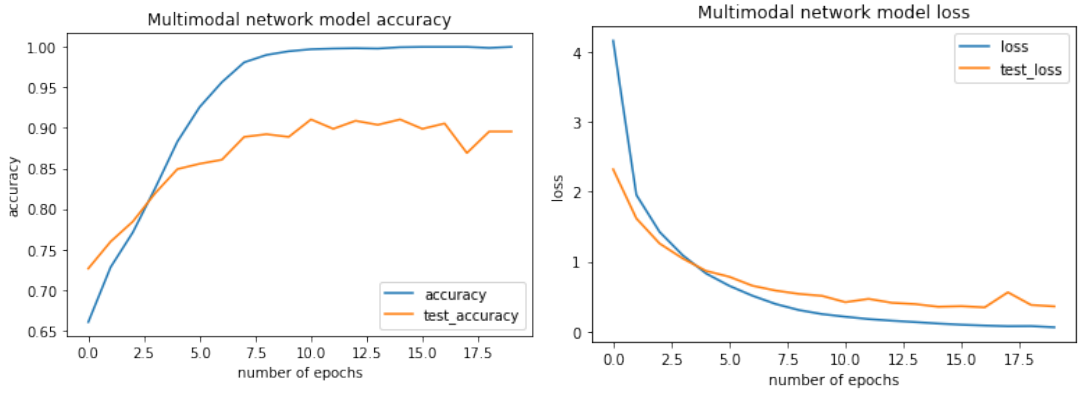


Figure 27. Multimodal network learning curves

In general, the model has a linear increase in accuracy with overfitting on the last epochs, and the loss is slightly less than that of other models and ranges from 0.5 to 0.35. This means that this model is able to more accurately predict the reliability of new data. It is also the only one of all presented mono-modal and multimodal models that can overcome the 90 percent accuracy threshold. All the main metrics of the multimodal network are presented in Table 10.

Table 10. Multimodal network main metrics

| Class | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Fake news | 90.55% | 77.44% | 86.39% | 81.67% |
| Not fake news | 90.55% | 95.44% | 91.89% | 93.63% |

The resulting multimodal network provides excellent performance and accuracy. Thereby, it outperforms all models presented so far in all metrics except the recall measure. The accuracy of this model is 1.83% higher, f1-score for COVID-19 fake news is 4.96% higher; the increase in f1-score for not fake news is 1.07% compared to the best model of a multimodal pair of news article and images.

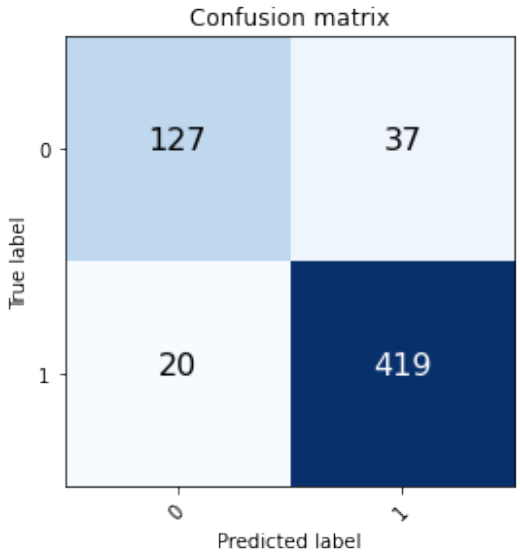


Figure 28. Multimodal network confusion matrix

According to the confusion matrix presented above, this model predicts fake news better than all other models, with a result of 127 instances. In addition, the multimodal network makes the least errors when predicting COVID-19 fake news articles as not fake. For other predictions, it can be compared to the multimodal pair of news and images.

Finally, the multimodal network presented in this section is by far the best and most accurate CNN model capable of predicting COVID-19 fake news articles with 546 correctly classified instances and 57 incorrectly classified instances.

# Conclusion and future work

To date, the spread of fake news, as well as disinformation from unreliable sources, has reached truly widespread proportions. They distort people's perceptions and opinions on a variety of issues, and this kind of information is becoming increasingly difficult to control. Thus, having an ideological and financial motivation, fake news negatively affects the dissemination of unbiased and reliable information regarding the COVID-19 pandemic. Therefore, it is extremely important to develop an approach that could most effectively distinguish reliable sources of information about COVID-19 from unreliable and fake ones.

As part of the thesis, a multimodal dataset of 3000 records is used, obtained from various repositories dedicated to COVID-19 fake news detection, which provide multimodal information of news articles on coronavirus, including textual, visual and network information. Using CNN deep learning architecture, word2vec and VGG16 pre-trained models, various mono-modal models, multimodal pairs and a common multimodal network were built based on the available modalities. After that, their comprehensive analysis was carried out using the appropriate metrics, as a result of which it can be clearly shown which of the models performs better result.

In general, all the obtained CNN models showed an accuracy result exceeding the random baseline, so the choice of convolutional neural network architecture can be considered successful for such tasks. The results of the experiments showed that the multimodal CNN architecture is able to provide an increase in accuracy and efficiency for COVID-19 fake news detection. As the modality increases, each of the models outperforms its mono-modal results. During the experiment, the best of the mono-modal CNN models of news articles achieved an accuracy of 84.58%, an f1-score of 72.40% for COVID-19 fake news class and an f1-score of 89.30% for not fake news class. Whereas the multimodal pair of news articles and images exceeds this result by more than 4% with accuracy of 88.72% and more than 3% relative to the f1-score for COVID-19 fake news class and not fake news class with results of 76.71% and 92.56%, respectively. The most effective and accurate model presented in this thesis for

COVID-19 fake news detection is a model that combines all four modalities in a multimodal network, with an accuracy of 90.55% and an f1-score for COVID-19 fake news class of 81.67%. This means that the multimodal CNN network is able to provide a 6% increase in accuracy and a 9% increase in f1-score over the best mono-modal model.

To improve the presented results, the following possibilities for further work should be considered:

- Data related:
    - Use a larger dataset. Since in this thesis we are limited to a dataset of 3000 records, using a larger dataset can lead to even better results.
- Modality related:
    - Use more modalities. For example, consider the time between the publication of a news and the appearance of the first tweets related to that news item as an additional modality.
    - Also, additional modalities can be hashtags, audio or video materials given in news articles.
- Pre-trained models related:
    - Use other pre-trained word embedding models like BERT, GloVe and FastText.
    - Use other image recognition models like Xception, ResNet50 and Inception.
- Implementing possibilities:
    - Implementation of the resulting models into a website for checking the facts about COVID-19.

To sum it up, due to the use of multimodal approaches and the convolutional neural network deep learning architecture, significant increase in the detection of COVID-19 fake news can be achieved. Thus, the methods used in this thesis are able to effectively

solve this problem, potentially reducing the negative impact of misinformation on society, the social sphere and the economy.

# References


[1] K. Lambert, „Computer Applications to Library", *ED-Tech Press,* pp. 36-37, 2020.

[2] L. Wu, F. Morstatter, K. M. Carley and H.Liu, „Misinformation in Social Media: Definition, Manipulation, and Detection", *ACM SIGKDD Explorations Newsletter*, vol. 2, pp. 80-90, 2019. [Online]. Available: https://kdd.org/exploration_files/8._CR.10.Misinformation_in_social_media_-_Final.pdf. [Accessed 7 May 2022].

[3] Z. Bastick, „Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation", *Computers in Human Behaviour*, vol. 116, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563220303800. [Accessed 7 May 2022].

[4] D. A. Scheufele and N. M. Krause, „Science audiences, misinformation, and fake news", *Proceedings of the National Academy of Sciences*, vol. 116, no.16, 2019. [Online]. Available: https://www.pnas.org/doi/pdf/10.1073/pnas.1805871115. [Accessed 7 May 2022].

[5] M. Westerlund, „The Emergence of Deepfake Technology: A Review", *Technology Innovation Management Review*, vol. 9, pp. 40-53, 2019. [Online]. Available: https://timreview.ca/article/1282. [Accessed 7 May 2022].

[6] F. Olan, U. Jayawickrama, E. O. Arakpogun, J. Suklan and S. Liu, „Fake news on Social Media: the Impact on Society", *Information Systems Frontiers*, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s10796-022-10242-z#article-info. [Accessed 7 May 2022].

[7] E. Tandoc and Z.W. Lim, „Defining "Fake News": A typology of scholarly definitions", *Digital Journalism*, vol 6(3), pp. 1-17, 2017. [Online]. Available: https://www.researchgate.net/publication/319383049_Defining_Fake_News_A_typology_of_scholarly_definitions. [Accessed 7 May 2022].

[8] Statista, „Programmatic advertising revenue of misinformation publishing sites in the United States and worldwide in 2021.", *Statista Research Department*, 2021. [Online]. Available: https://www.statista.com/statistics/1253292/misinformation-publishers-advertising-revenue/. [Accessed 7 May 2022].

[9] T. J. Ackermann, „Study Finds 'Fake News' Has Real Cost: $78 Billion", 2019. [Online]. Available: https://www.bgp4.com/2019/11/22/study-finds-fake-news-has-real-cost-78-billion/. [Accessed 7 May 2022].

[10] European Comission, „A multi-dimensional approach to disinformation", *Report of the independent High level Group on fake news and online disinformation,* 2018. [Online]. Available: https://www.ecsite.eu/sites/default/files/amulti-dimensionalapproachtodisinformation-reportoftheindependenthighlevelgrouponfakenewsandonlinedisinformation.pdf. [Accessed 7 May 2022].

[11] D. Funke and D. Flamini, „A guide to anti-misinformation actions around the world", 2020. [Online]. Available: https://www.poynter.org/ifcn/anti-misinformation-actions/#germany. [Accessed 7 May 2022].

[11] D. Funke and D. Flamini, „A guide to anti-misinformation actions around the world", 2020. [Online]. Available: https://www.poynter.org/ifcn/anti-misinformation-actions/#germany. [Accessed 7 May 2022].

[12] L. Miner, „EU lawmakers, member states endorse act aimed at tackling harmful online content", 2022. [Online]. Available: https://www.euronews.com/2022/04/23/eu-lawmakers-member-states-endorse-act-aimed-at-tackling-harmful-online-content. [Accessed 7 May 2022].

[13] A. Thota, P. Tilak, S. Ahluwalia and N. Lohia, „Fake News Detection: A Deep Learning Approach", *SMU Data Science Review*, vol. 1, 2018. [Online]. Available: https://scholar.smu.edu/cgi/viewcontent.cgi?article=1036&context=datasciencereview. [Accessed 7 May 2022].

[14] N. O'Brien, S. Latessa, G. Evangelopoulos and X. Boix, „The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors", 2018. [Online]. Available: https://cbmm.mit.edu/sites/default/files/publications/fake-news-paper-NIPS.pdf. [Accessed 7 May 2022].

[15] O. Ajao, D. Bhowmik and S. Zargari, „Fake News Identification on Twitter with Hybrid CNN and RNN Models", *SMSociety '18: Proceedings of the 9th International Conference on Social Media and Society*, pp. 226-230, 2018. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1806/1806.11316.pdf. [Accessed 7 May 2022].

[16] A. Wani, I. Joshi, S. Khandve, V. Wagh and R. Joshi, „Evaluating Deep Learning Approaches for Covid19 Fake News Detection", 2021. [Online]. Available: https://arxiv.org/pdf/2101.04012.pdf. [Accessed 7 May 2022].

[17] A. Glazkova and T. Trifonov, „Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection", *Constraint@AAAI2021*, 2021. [Online]. Available: https://arxiv.org/pdf/2012.11967.pdf. [Accessed 7 May 2022].

[18] D. Marasco, „Multimodal Deep Learning Approaches And Applications", 2021. [Online]. Available: https://www.clarifai.com/blog/multimodal-deep-learning-approaches. [Accessed 7 May 2022].

[19] Z. Wang, Z. Yin and Y. Argyris, „Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning", *IEEE Journal of Biomedical and Health Informatics*, 2020. [Online]. Available: https://arxiv.org/pdf/2012.13968.pdf. [Accessed 7 May 2022].

[20] C. Raj and P. Meel, „ConvNet frameworks for multi-modal fake news detection", *Applied Intelligence*, 2021. [Online]. Available: https://www.researchgate.net/profile/Priyanka-Meel/publication/350422061_ConvNet_frameworks_for_multi-modal_fake_news_detection/links/61e189565779d35951aa294f/ConvNet-frameworks-for-multi-modal-fake-news-detection.pdf. [Accessed 7 May 2022].

[21] S. Singhal, R.R. Shah, T. Chakraborty and P. Kumaraguru, „SpotFake: A Multi-modal Framework for Fake News Detection", 2019. [Online]. Available: https://www.researchgate.net/profile/Shivangi-Singhal/publication/337791172_SpotFake_A_Multi-modal_Framework_for_Fake_News_Detection/links/5e01f131299bf10bc374598a/SpotFake-A-Multi-modal-Framework-for-Fake-News-Detection.pdf?origin=publication_detail%5D. [Accessed 7 May 2022].

[22] A. Kirchknopf, „Automated Identification of Information Disorder in Social Media from Multimodal Data", Master Thesis, St. Pölten University of Applied Sciences, 2020.

[23] J. Browniee, „Discover Feature Engineering, How to Engineer Features and How to Get Good at It”, 2020. [Online]. Available: https://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/. [Accessed 7 May 2022].

[24] Google Code Archive, „word2vec”, 2013. [Online]. Available: https://code.google.com/archive/p/word2vec/. [Accessed 7 May 2022].

[25] C. Nicholson, „A Beginner's Guide to word2vec and Neural Word Embeddings”, 2020. [Online]. Available: https://wiki.pathmind.com/word2vec. [Accessed 7 May 2022].

[26] G. Boesch, „VGG Very Deep Convolutional Networks (VGGNet) – What you need to know”, 2022. [Online]. Available: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/. [Accessed 7 May 2022].

[27] Neurohive, „VGG16 – Convolutional Network for Classification and Detection”, 2018. [Online]. Available: https://neurohive.io/en/popular-networks/vgg16/. [Accessed 7 May 2022].

[28] A. Jacovi and O. S. Shalom, Y. Goldberg, „Understanding convolutional neural networks for Text Classification”, 2020. [Online]. Available: https://arxiv.org/pdf/1809.08037.pdf. [Accessed 7 May 2022].

[29] R. Kumar, R.K. Dohare, H. Dubey and V. P. Singh, „Applications of Advanced Computing in Systems”, *Proceeding of International Conference on Advances in Systems Control and Computing*, 2021.

[30] The Click Reader, „Building a convolutional neural network”, 2020. [Online]. Available: https://www.theclickreader.com/building-a-convolutional-neural-network/. [Accessed 7 May 2022].

[31] P. Baheti, „A Newbie-Friendly Guide to Transfer Learning”, 2022. [Online]. Available: https://www.v7labs.com/blog/transfer-learning-guide. [Accessed 7 May 2022].

[32] J. Summaira, X. Li, A. M. Shoib, S. Li and J. Abdul, „Recent Advances and Trends in Multimodal Deep Learning: A Review”, 2021.

[33] J. Browniee, „How to use Learning Curves to Diagnose Machine Learning Model Performance”, 2019. [Online]. Available: https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/. [Accessed 7 May 2022].

[34] S. Gupta, „What Is the Best Language for Machine Learning?”, 2021. [Online]. Available: https://www.springboard.com/blog/data-science/best-language-for-machine-learning/. [Accessed 7 May 2022].

[35] Webwise, „Explained: What is False Information (Fake News)?”, 2020. [Online]. Available: https://www.webwise.ie/teachers/what-is-fake-news/. [Accessed 9 May 2022].

[36] S. Saxena, „Binary Cross Entropy/Log Loss for Binary Classification”, 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/03/binary-cross-entropy-log-loss-for-binary-classification/. [Accessed 9 May 2022].

[37] J. Browniee, „How to use Learning Curves to Diagnose Machine Learning Model Performance”, 2019. [Online]. Available: https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/. [Accessed 9 May 2022].

[38] J. Browniee, „Difference Between a Batch and an Epoch in a Neural Network”, 2018. [Online]. Available: https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/. [Accessed 11 May 2022].

[39] A. Murphy, „Batch size (machine learning)”, 2019. [Online]. Available: https://radiopaedia.org/articles/batch-size-machine-learning. [Accessed 11 May 2022].

[40] X. Zhou and A. Mulay, „ReCOVery: A Multimodal Repository for COVID-19 News Credibility Research”, 2020.

[41] M. Chen and K. P. Subbalakshmi, „MMCoVaR: Multimodal COVID-19 Vaccine Focused Data Repository for Fake News Detection and a Baseline Architecture for Classification”, 2021.

[42] L. M. Molera, „All About Model Validation”, 2022. [Online]. Available: https://explore.mathworks.com/all-about-model-validation. [Accessed 11 May 2022].

# Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Ivan Švaiger

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis , supervised by Nadežda Furs

   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

# Appendix 2 – Links to source code

ReCOVery repository pre-processing: https://colab.research.google.com/drive/1ysecdVUUYnGB0WZko-4NfyTAQixRXuRx?usp=sharing

MMCoVaR repository pre-processing: https://colab.research.google.com/drive/1ahk3k6jbUgh7x9AamYMzRFJuSUlzjHuk?usp=sharing

Combining repositories into a single dataset: https://colab.research.google.com/drive/1_pTnFudyj7mQ-mIYYVwBjlLSErSduEpl?usp=sharing

Final multimodal COVID-19 dataset: https://drive.google.com/file/d/1YeXEe2ep32yUHk3JsXQIqIYpI56iZ1kU/view?usp=sharing

Images pre-processing: https://colab.research.google.com/drive/1NiRxECz_odW94YOSJwscfcHkLy2JaVD1?usp=sharing

Text pre-processing, mono-modal and multimodal models: https://colab.research.google.com/drive/1j72MAN8XBwZFSthSEzKdTUJ9TO3J1zjM?usp=sharing