DGK Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 882

Yu Feng

Extraction of Flood and Precipitation Observations from Opportunistic Volunteered Geographic Information

München 2021

Bayerische Akademie der Wissenschaften

ISSN 0065-5325 | ISBN 978-3-7696-5294-9

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 376, Hannover 2021

DGK Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 882

Extraction of Flood and Precipitation Observations from Opportunistic Volunteered Geographic Information

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
Doktor-Ingenieur (Dr.-Ing.)
genehmigte Dissertation

von

M.Sc. Yu Feng

geboren am 09.10.1989 in Baotou, Nei Mongol, China

München 2021

Bayerische Akademie der Wissenschaften

ISSN 0065-5325 ISBN 978-3-7696-5294-9

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 376, Hannover 2021

Adresse der DGK:



Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100

e-mail post@dgk.badw.de ● http://www.dgk.badw.de

Prüfungskommission:

Vorsitzender: Prof. Dr.-Ing. habil. Christian Heipke

Referentin: Prof. Dr.-Ing. habil. Monika Sester

Korreferenten: Prof. Dr. Alexander Zipf (Universität Heidelberg)
Prof. Dr.-Ing. Winrich Voß

Tag der mündlichen Prüfung: 13.09.2021

© 2021 Bayerische Akademie der Wissenschaften, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325 ISBN 978-3-7696-5294-9

Abstract

Floods are the most common natural disaster, causing many deaths, injuries, and property damages. They are primarily caused by heavy precipitation events. Nowadays, with increasing urbanization, floods occur more frequently in cities. A comprehensive understanding of the current flood and precipitation situation is essential not only for the city's emergency management but also for the local residents. Today's flood and precipitation monitoring techniques have deficiencies in terms of temporal resolution, spatial resolution, or global coverage. Thus, additional sources of information need to be considered to achieve a more comprehensive monitoring.

Floods and heavy precipitation often draw the attention of the public, and may also affect their behaviors. Citizens share their observations on social media with text and photos. Also their behaviors change accordingly, e.g., driving more slowly than usual. These observations can be used as a source of information for flood and precipitation monitoring, as long as the corresponding geographic locations are available. These data sources are often referred to as Volunteered Geographic Information (VGI). VGI can be collected using a participatory approach or an opportunistic approach. Participatory approaches require a conscious and active participation by the users, e.g., by using a web portal or a mobile app. The unconscious information acquisition or information acquisition for a different purpose (e.g., photos shared on Instagram) is considered as the opportunistic approaches. It is becoming increasingly difficult to fully rely on the frequent voluntary participation of users. Due to its necessary efforts (and the often required disclosure of identity), the number of voluntary users is typically low. Therefore, opportunistic VGI is the focus of the thesis.

The goal is to investigate to what extent flood and precipitation observations can be extracted from VGI with minimal intentional user involvement. In terms of the information sources that are used, the thesis focuses on two aspects. One aims to extract precipitation indications from passive behavioral changes of road users, the other aims to extract flood-relevant information from users' active information provision on social media.

Precipitation events can lead to significant decreases in traffic speeds in the affected areas. This is different from slowdowns caused by local events, such as concerts or traffic accidents, which have a limited area of impact around the event location. As a proof-of-concept, a study is conducted to learn a precipitation indicator from the road speed observations collected by road speed detectors. A binary classifier was trained on six-month road speed records from New York City and achieved an accuracy of 91.74% and F_1-score of 78.34% when tested on the remaining two-month test data. This promising performance demonstrates the potential of using this information source to complement precipitation observations, especially for the areas which lack basic meteorological facilities.

Social media as a real-time data source can provide flood observations from users. This is also an opportunistic data source, as the users typically want to share the information with friends rather than uploading data to an emergency response web page or using an app provided by the fire brigade. In this thesis, a framework is built to collect and analyze social media data from Twitter and Instagram. Previous studies mainly focused on user-generated text. This thesis presents a very early attempt to use deep learning models to extract high-quality flood "eyewitness reports" from user-generated text and images. Further analyses identify spatiotemporal clusters and hotspot areas for flood events in Paris, London, and Berlin in 2016 and 2017. These detected clusters and

hotspots are the areas that attract the attention of users. In addition to the location of such events, the city's emergency management is very much interested in the severity of the flooding. A novel method is proposed to extract and map flood severity information. The severity corresponds to the inundation level. Typically, this would require a gauge rod. The approach in this thesis uses human as scales. After retrieving flood-relevant images, images containing people are classified into four severity levels by observing the relationship between body parts and their partial inundation, i.e., images are classified according to the water level with respect to different body parts, namely ankle, knee, hip, and chest. Locations of the Tweets are then used for generating a map of estimated flood extent and severity. This process is applied to an image dataset collected during Hurricane Harvey in 2017 as a proof of concept.

In summary, this thesis presents several new potentials of opportunistic VGI. Speed variation of road users can be used as a precipitation indicator, and social media data can provide flood eyewitness reports as well as water level estimates. These citizens' observations can complement existing monitoring technologies and provide new information for the city's emergency management.

Keywords: Volunteered Geographic Information, citizen science, flood, precipitation, social media, deep convolutional neural networks, flood mapping, traffic speed variation, crowdsourcing

Kurzfassung

Hochwasser sind die häufigste Naturkatastrophe und verursachen viele Todesfälle, Verletzungen und Sachschäden. Sie werden hauptsächlich durch Starkniederschlagsereignisse verursacht. Mit zunehmender Urbanisierung treten Hochwasser heutzutage häufiger in Städten auf. Ein umfassendes Verständnis der aktuellen Hochwasser- und Niederschlagssituation ist nicht nur für das Krisenmanagement der Stadt, sondern auch für ihre Bewohner unerlässlich. Die derzeitigen Verfahren zur Hochwasser- und Niederschlagsüberwachung weisen Defizite in Bezug auf die zeitliche und räumliche Auflösung sowie die globale Abdeckung auf. Daher müssen zusätzliche Informationsquellen in Betracht gezogen werden, um eine umfassendere Überwachung zu erreichen.

Hochwasser und Starkniederschläge ziehen oft die Aufmerksamkeit der Öffentlichkeit auf sich. Bürger teilen daher ihre Beobachtungen in sozialen Medien mithilfe von Fotos und Texten. Darunter hinaus passen sie ihr Verhalten den Gegebenheiten an, z.B. fahren sie langsamer als gewöhnlich. Diese Daten von und über Einzelpersonen können als Informationsquelle für die Hochwasser- und Niederschlagsüberwachung genutzt werden, sofern die entsprechenden geografischen Standorte verfügbar sind. Diese Informationsquelle wird oft als "Volunteered Geographic Information" (VGI) bezeichnet. VGI kann in partizipative oder opportunistische Ansätze unterschieden werden. Partizipative Ansätze erfordern eine bewusste und aktive Beteiligung der Nutzer, z.B. durch die Nutzung eines Webportals oder einer mobilen App. Die oben beschriebene Art der unbewussten Erfassung bzw. Erfassung zu einen anderen Zweck (wie z.B. Instagram Fotos) zählt zu den opportunistischen Ansätzen. Es wird immer schwieriger, sich ausschließlich auf die regelmäßige, freiwillige Teilnahme der Nutzer zu verlassen. Aufgrund des hierfür erforderlichen Aufwands (und der oft erforderlichen Preisgabe der Identität) ist die Anzahl freiwilliger Nutzer typischerweise gering. Daher steht die opportunistische VGI im Fokus der Arbeit.

Ziel ist es, zu untersuchen, inwieweit Hochwasser- und Niederschlagsbeobachtungen aus VGI mit minimaler bewusster Nutzerbeteiligung extrahiert werden können. In Bezug auf die verwendeten Informationsquellen konzentriert sich die Arbeit auf zwei Aspekte. Der erste Teil zielt darauf ab, Niederschlagshinweise aus passiven Verhaltensänderungen von Verkehrsteilnehmern zu extrahieren; der zweite Teil darauf, hochwasserrelevante Informationen aus der aktiven Informationsbereitstellung von Social-Media-Nutzern zu gewinnen.

Niederschlagsereignisse können zu signifikanten Verringerungen der Verkehrsgeschwindigkeit in den betroffenen Gebieten führen. Dies unterscheidet sich von den Verlangsamungen, die durch lokale Ereignisse verursacht werden, wie z.B. Konzerte oder Verkehrsunfälle, die einen begrenzten Einflussbereich um den Veranstaltungsort haben. Es wird eine Proof-of-Concept-Studie durchgeführt, um einen Niederschlagsindikator – aus den von den Geschwindigkeitsdetektoren gesammelten Beobachtungen auf den Straßensegmenten – zu lernen. Ein binärer Klassifikator wurde auf sechsmonatigen Straßengeschwindigkeitsaufzeichnungen aus New York City trainiert und erreichte eine Genauigkeit von 91,74% und einen F_1-Score von 78,34%, als dieser auf den verbleibenden zweimonatigen Testdaten getestet wurde. Diese vielversprechende Leistungsfähigkeit der Methode zeigt das Potenzial der Nutzung dieser Informationsquelle zur Ergänzung von Niederschlagsbeobachtungen, insbesondere für Gebiete, in denen es an grundlegenden meteorologischen Einrichtungen fehlt.

Soziale Medien als Echtzeit-Datenquelle können Hochwasser-Beobachtungen von Nutzern enthalten. In dieser Arbeit wird ein Framework entwickelt, um Social-Media-Daten von Twitter und

Instagram zu sammeln und zu analysieren. Bisherige Studien konzentrierten sich hauptsächlich auf benutzergenerierte Texte. Diese Arbeit stellt einen der ersten Ansätze dar, Deep-Learning-Modelle zu verwenden, um qualitativ hochwertige "Hochwasser-Augenzeugenberichte" aus nutzergenerierten Texten und Bildern zu extrahieren. Weitere Analysen identifizieren raum-zeitliche Cluster und Hotspot-Bereiche für die Hochwasser-Ereignisse in Paris, London und Berlin in den Jahren 2016 und 2017. Neben dem Ort solcher Ereignisse interessiert das Krisenmanagement der Stadt vor allem die Schweregrad des Hochwassers. In der Arbeit wird eine neuartige Methode vorgeschlagen, um Informationen zum Schweregrad von Hochwasser zu extrahieren und zu kartieren. Der Schweregrad entspricht der Höhe der Überflutung. Normalerweise würde dies eine Referenz erfordern. Der Ansatz in dieser Arbeit verwendet den Menschen als Maßstab, d.h. Bilder in denen Menschen im Wasser stehen, wobei der Wassserstand in Bezug auf verschiedene Körperbereiche klassifiziert, nämlich Knöchel, Knie, Hüfte und Brust. Die Standorte der Tweets werden dann für die Erstellung einer Karte der geschätzten Ausdehnung und Schwere des Hochwassers verwendet. Als Proof-of-Concept wird dieser Prozess auf einen Bilddatensatz angewendet, der während des Hurrikans Harvey im Jahr 2017 gesammelt wurde.

Zusammenfassend stellt diese Arbeit mehrere neue Potenziale der opportunistischen VGI vor. Geschwindigkeitsvariationen von Verkehrsteilnehmern können als Niederschlagsindikator genutzt werden, und Daten aus sozialen Medien können Hochwasser-Augenzeugenberichte sowie Wasserstandsschätzungen liefern. Diese Beobachtungen der Bürger können bestehende Beobachtungstechnologien ergänzen und neue Informationen für das Krisenmanagement der Stadt liefern.

Schlagwörter: Volunteered Geographic Information, Citizen Science, Hochwasser, Niederschlag, Soziale Medien, Deep Convolutional Neural Networks, Hochwasserkartierung, Verkehrsgeschwindigkeitsvariation, Crowdsourcing

Contents

1 Introduction **1**

1.1 Motivation . . . 1

1.1.1 Existing precipitation and flood monitoring techniques . . . 2

1.1.2 Volunteered Geographic Information (VGI) - citizens as sensors . . . 5

1.2 Goal of this thesis . . . 6

1.3 Outline . . . 7

2 Background **9**

2.1 Machine learning . . . 9

2.1.1 Supervised classification . . . 9

2.1.2 Classification evaluation . . . 12

2.2 Image interpretation with Deep Convolutional Neural Networks . . . 14

2.2.1 Basics of Convolutional Neural Network . . . 14

2.2.2 Common tasks in computer vision . . . 17

2.3 Text analysis with Natural Language Processing . . . 22

2.3.1 Bag-of-Words . . . 23

2.3.2 Word embedding . . . 24

2.3.3 TextCNN . . . 25

2.4 Spatial and spatiotemporal analyses . . . 26

2.4.1 Heatmaps and hot spot analysis . . . 26

2.4.2 Density-based clustering . . . 28

2.5 Opportunistic VGI data . . . 29

2.5.1 Characteristics and data sources of opportunistic VGI . . . 29

2.5.2 Structures and characteristics of Twitter data . . . 32

3 Related work **35**

3.1 User-provided precipitation observations . . . 35

3.2 User-provided flood observations . . . 36

3.2.1 Participatory approaches . . . 36

3.2.2 Opportunistic approaches . . . 37

3.3 Interpretation of flood observations from social media texts and images . . . 42

3.3.1 Text analysis for flood events . . . 42

3.3.2 Image analysis for flood event characterization . . . 44

3.3.3 Water level observations from social media posts . . . 46

3.4 Precipitation and traffic speed variation . . . 47

3.5 Research gap . . . 48

4 Precipitation indicator from road users' speed variation 49
4.1 Motivation . . . 49
4.2 Methodology . . . 49
4.2.1 Data . . . 49
4.2.2 Method . . . 50
4.3 Experiment and results . . . 51
4.4 Summary . . . 53

5 Methodology for the extraction of flood observations from social media VGI 57
5.1 Interpretation of flood-related social media texts . . . 57
5.1.1 Pre-processing and training data preparation . . . 57
5.1.2 Training text classifiers . . . 59
5.2 Interpretation of flood-related social media images . . . 60
5.2.1 Training image classifiers using single pre-trained model . . . 61
5.2.2 Training image classifiers by assembling multiple pre-trained models . . . 61
5.2.3 Detection of duplicate images . . . 62
5.3 Estimation of water level from flood-relevant images . . . 62
5.3.1 Learning a water level classifier with handcrafted features . . . 62
5.3.2 Baseline 1: Multiclass image classification with global deep features of the whole image . . . 65
5.3.3 Baseline 2: Mask R-CNN with extra branch for water level classification . . . 66

6 Experiments to extract flood observations from social media VGI 67
6.1 Social media data acquisition . . . 67
6.1.1 Floods in Europe in 2016 . . . 68
6.1.2 Hurricane Harvey in Texas, United States in 2017 . . . 69
6.2 Datasets for training classification models . . . 69
6.2.1 Text dataset annotated via keyword filtering and cross-referencing weather data . . . 69
6.2.2 Manually annotated pluvial flood image dataset . . . 69
6.2.3 MediaEval'17 MMSat benchmark dataset and its extension . . . 70
6.2.4 Image dataset for water level estimation . . . 71
6.3 Extraction of pluvial flood-relevant VGI based on social media texts and photos . . . 72
6.3.1 Training of the text classifier . . . 72
6.3.2 Training of the image classifier . . . 75
6.3.3 Detection of heavy rainfall and flood events . . . 77
6.3.4 Visualization of the pluvial flood relevant information . . . 81
6.3.5 Analyses and comparison with external data sources . . . 81
6.3.6 Summary . . . 83
6.4 Flood severity mapping from VGI by interpreting water level from images containing people . . . 84
6.4.1 Retrieval of flood relevant social media images . . . 85

6.4.2 Experiment and evaluation of water level estimation 87
6.4.3 Flood severity mapping for Hurricane Harvey in 2017 91
6.4.4 Summary 100

7 Discussion **101**
7.1 The inherent challenge of social media as opportunistic VGI 103
7.2 Limitations of the current social media processing pipeline 104

8 Conclusions and outlook **107**
8.1 Research questions 107
8.2 Outlook 108

List of Figures **111**

List of Tables **115**

Bibliography **117**

Acknowledgements **135**

Curriculum Vitae **137**

1 Introduction

1.1 Motivation

Floods, the most frequent natural disaster, cause a large number of casualties and property damage. According to the UN Office for Disaster Risk Reduction (UNISDR) and the Centre for Research on the Epidemiology of Disasters (CRED), 3,148 floods occurred between 1998 and 2017, accounting for 43.4 percent of all the natural disasters. Floods affected two billion people, or 45 percent of the total number of people affected by natural disasters, far more than any other disaster type. In addition, 142,088 people have been killed as a result of floods, and economic losses amounted to $656 billion (CRED and UNISDR, 2018).

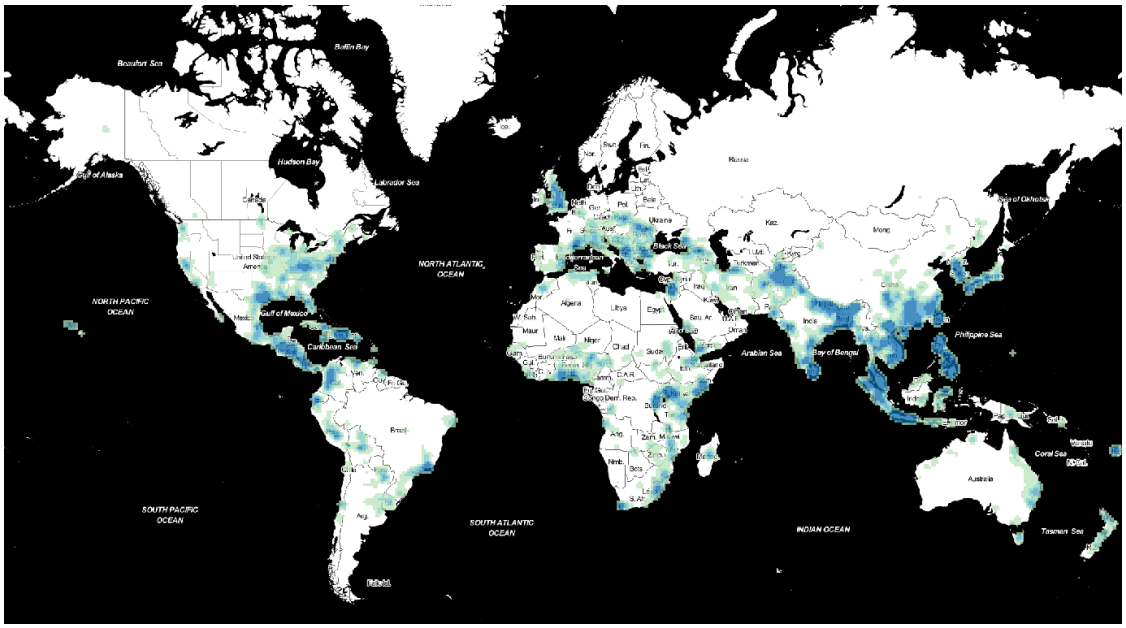


Figure 1.1: Global heat map for the large flood events 1985-2019.

Flood is a global challenge. Many places in the world are threatened by flooding, as visualized in the heat map of historical flood records[1] between 1985 and 2019 in Figure 1.1. More than one-third of the land area worldwide is flood-prone regions, in which about 82% of the world's population lives (Dilley et al., 2005). This includes many metropolitan cities, which have been plagued by flooding, such as Beijing in 2012 (Wang et al., 2016a), New York in 2014 (WSJ, 2014), Paris in 2016 (BBC, 2016), London in 2016 (BBC, 2016), Berlin in 2017 (B.Z., 2017), and Houston in 2017 (BBC, 2017).

Flood monitoring is one of the essential components of flood control and risk management. Observations of floods are necessary to government agencies of all countries for the protection of people's

[1]Data source: Dartmouth Flood Observatory, University of Colorado. `http://floodobservatory.colorado.edu` (Accessed on 31.01.2021)

lives and properties, emergency response, post-disaster damage analysis, and accumulation of historical flood data.

Floods are primarily caused by precipitation events. As shown in Figure 1.2, the top four leading causes based on historical flood records[2] are all related to precipitation events, namely heavy rain, torrential rain, tropical cyclone, and monsoonal rain. Other reasons, such as snowmelt, ice jam, or dam breaks, occur significantly less frequently. Thus, the monitoring of flood events is inseparable from the monitoring of precipitation events.

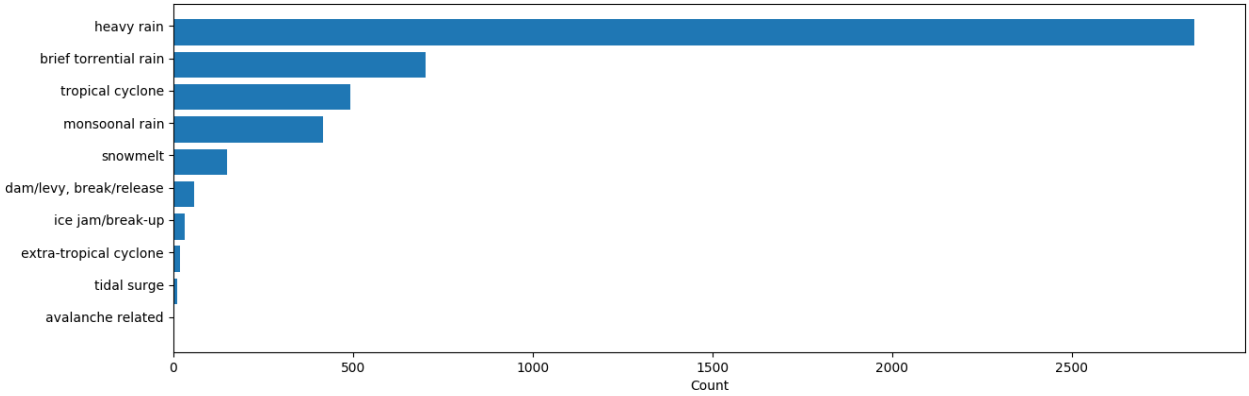


Figure 1.2: The main causes of the large flood events 1985-2019.

In general, there are three types of floods, namely coastal flood, fluvial (river) flood, and pluvial flood. Storm surges related to tropical cyclones and tsunamis are the leading causes of coastal floods. Consistent rainfall or snowmelt forces rivers to exceed capacity, which leads to a fluvial flood. Pluvial floods are caused by rapid and excessive rain and may occur both in urban and rural areas, not necessarily in the vicinity of water bodies. The floodwater over-saturates the ground and overflows the drainage systems (Maddox, 2014). In the last few decades, the density of urban development and the area of sealed land has increased dramatically, which leads to more severe flooding situations than ever before (Mård and Di Baldassarre, 2018).

1.1.1 Existing precipitation and flood monitoring techniques

In hydrology, there is a long history of monitoring precipitation and surface water. Different techniques and tools have been developed to collect observations on precipitation and flood events.

As for precipitation monitoring, the most often used measurement devices are rain gauges and weather radars. Both of the techniques can achieve precipitation monitoring in real-time, i.e., with a delay of only several minutes. Rain gauges are the most common devices. In contrast to weather radars, which provide areal information, rain gauges only measure the data at individual points. Rain gauges can be categorized as recording or non-recording rain gauges. Non-recording rain gauges aggregate observations on a daily basis and are available in sufficient density. Recording rain gauges provide a much higher temporal resolution (e.g., 10 min for German Weather Service - DWD data) but they are less densely distributed. For example, in Germany, the average density of recording rain gauges is one station per 1800 km^2, whereas 90 km^2 for non-recording rain gauges (Haberlandt and Sester, 2010). Contemporary rain gauges can provide automatic readings and transmit data in real-time, such as the UK Environment Agency's (EA) rain gauge measure-

[2]Data source: Dartmouth Flood Observatory, University of Colorado. `http://floodobservatory.colorado.edu` (Accessed on 31.01.2021)

ments[3]. Weather radar can be used to detect the type, distribution, movement, and evolution of precipitation in the atmosphere, such as the DWD weather radar data[4]. Weather radars often have a large coverage area, e.g., DWD radars have a coverage area of above 150 km and provide data with a spatial resolution of around 1 km (DWD, 2020b). However, the global distribution of such high-resolution measurements is extremely uneven. As shown in Figure 1.3, most weather radar stations are located in Europe, America, and the Asia Pacific regions. In Africa, central Asia, and the west coast of South America, much fewer weather radar stations exist. In addition, satellite-based precipitation measurement is an emerging product. It aims to improve the forecast of extreme events, which may cause natural hazards and disasters. The most popular data is GPM (Global Precipitation Measurement), which has a spatial resolution of 0.1° (approximately 10 km) and temporal resolution for 30 minutes. However, it can only achieve a near-real-time performance, where a minimum latency of four hours exists due to data acquisition (NASA, 2020).

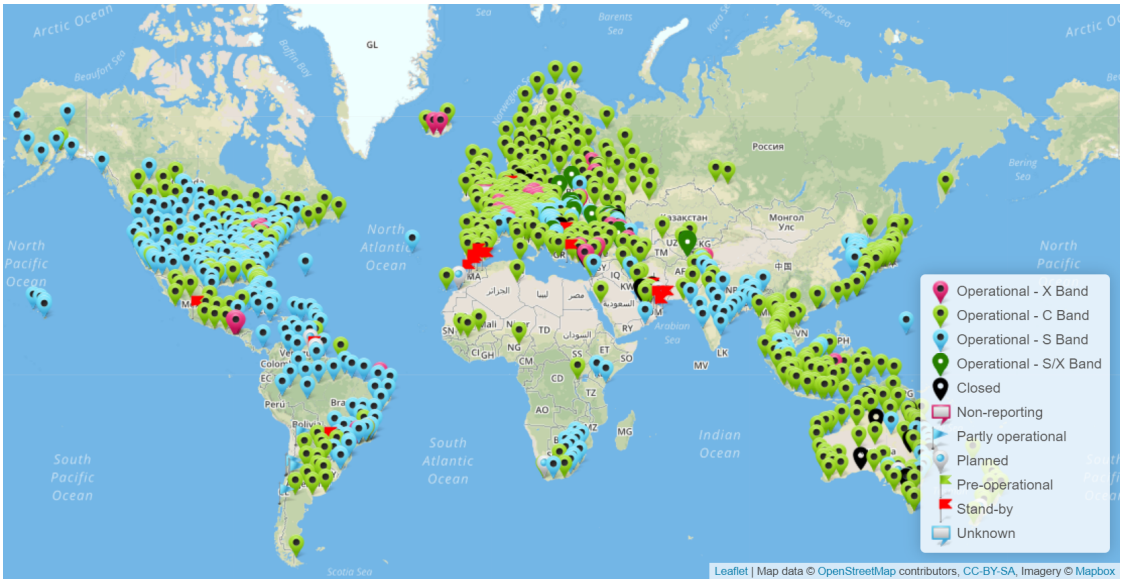


Figure 1.3: Global distribution of weather radars in the WMO radar database; status: 31.01.2021 (WMO, 2020).

As for flood monitoring, the most conventional way is using tide and river gauges. Staff gauge is the most basic tool to measure the water surface elevation. A long ruler is placed in the water body and is read by human operators. In addition, gauges using a float in the stilling well or pressure-actuated recording gauges are also commonly used (USGS, 2020a). Nowadays, gauges using ultrasonic and radar sensors are new techniques providing continuous water level recording (WMO, 2008). For example, in the United States, USGS collect data in 15-60 minutes intervals and transmit it to their offices every 1-4 hours. For critical events, the recording and transmission times may be more frequent. The data are available for visualization in a web map application within minutes of arrival (USGS, 2020b). Many countries have developed nation wide online river or tide gauge maps, e.g., river gauge from USGS[5], tide gauge from NOAA in the United States[6],

[3]Environment Agency - Rainfall API. `https://environment.data.gov.uk/flood-monitoring/doc/rainfall` (Accessed on 31.01.2021)
[4]Deutscher Wetterdienst (DWD) - Open Data Server. `https://www.dwd.de/EN/ourservices/opendata/opendata.html` (Accessed on 31.01.2021)
[5]USGS WaterWatch - Streamflow conditions. `https://waterwatch.usgs.gov/` (Accessed on 31.01.2021)
[6]NOAA Tides and Currents. `https://tidesandcurrents.noaa.gov/` (Accessed on 31.01.2021)

river gauge in UK and Ireland[7], and river gauge in Germany[8]. Such tide and river gauges are primarily focused on coastal and fluvial flood events. However, pluvial floods, which are normally caused by local, fast storm events with very high rainfall rates, are hard to be monitored and observed.

In contrast to the water level measurements at specific locations along rivers and coasts, remote sensing is a rapidly developing technique for acquiring water extent and depth information for large coverage areas. Intensive studies have been conducted on flood extent mapping from satellite remote sensing data. Methods for flood detection have been tested on different high-resolution remote sensing products, such as Landsat TM/ETM+ (Li et al., 2015), MODIS(Son et al., 2013), and TerraSAR-X (Martinis et al., 2015; Li et al., 2019). Researchers used the Normalized Difference Water Index - NDWI (Huang et al., 2018b), a modified NDWI (Rosser et al., 2017) or image analysis (Sarker et al., 2019; Feng et al., 2019) to obtain the water extent. With a given Digital Terrain Model (DTM), water depth can be further estimated (Singh et al., 2015).

However, airborne or satellite remote sensing products can hardly achieve real-time monitoring of flood events for the following three reasons. First, severe weather conditions limit the visibility of both products, especially because of the clouds that come along with heavy rain (Huang et al., 2018b). Second, the revisit time of the satellites limits the data availability (Feng et al., 2015). Commercial optical satellites sometimes need several days after an event to acquire high-resolution imagery (Ning et al., 2020). Third, airborne sensors such as Unmanned Aerial Systems (UAS) normally can only be deployed with a controllable risk after the events. All these limitations may result in the loss of first-hand information on a flood event. Therefore, for the flood events in urban areas, especially floods caused by short-time storms and heavy rainfall, observations from remote sensing are not able to achieve a satisfactory spatial and temporal resolution.

Table 1.1: Comparison between current precipitation and flood monitoring approaches.

Target	**Method and Examples**	**Efficiency**	**Spatial Resolution**	**Temporal Resolution**	**Limitation**
precipitation	rain gauge (EA, 2020)	real-time	isolated points	15min	spatial coverage
	weather radar (DWD, 2020a)	real-time ($<$ 10min)	1km	5min	global coverage
	GPM Satellite (NASA, 2020)	near-real-time (4h latency)	0.1°	30min	time delay
flood	river/tide gauge (Pegelonline, 2020)	real-time	isolated points	15min (best 1min)	spatial coverage
	remote sensing	near-real-time			weather/cloud
	- Sentinal-2 (ESA, 2020)		10m	5 days	revisit time
	- WorldView-3 (SIC, 2020)		0.31m	$<$ 1 day	high cost
	UAV	post-event only	0.05-0.3m	limited number of visits	weather

[7]Shoothill GaugeMap. `http://www.gaugemap.co.uk/` (Accessed on 31.01.2021)
[8]Pegelonline. `https://www.pegelonline.wsv.de/` (Accessed on 31.01.2021)

The above-mentioned precipitation and flood monitoring techniques are summarized in Table 1.1. It is worth noting that in precipitation monitoring, while the combination of rain gauges and weather radar provides both real-time and high temporal and spatial resolution, there are still many areas of the world that are not well covered by weather radar. Meanwhile, the GPM can achieve only a near-real-time performance. As for flood monitoring, river and tide gauges are located along rivers and coasts, while remote sensing data are limited by revisit time and weather conditions. Therefore, observations from the ground are needed as a supplement to the existing precipitation and flood monitoring framework. Observations should ideally come from real-time data sources with high temporal and spatial resolution.

1.1.2 Volunteered Geographic Information (VGI) - citizens as sensors

In the past few decades, the development of Web 2.0 and mobile Internet has enabled users to participate in the creation of Internet content anytime and anywhere. The popularity of mobile devices with positioning sensors makes location information easier to collect and share. With the fast development in the recent ten years, *crowdsourcing* is becoming an important information source for data acquisition. The word *crowdsourcing* itself is a blend word of *crowd* and *outsourcing* (Howe, 2006), which means outsourcing work to the crowd. Individuals can contribute to many tasks to achieve a cumulative result via the Internet. It has shown great success in various fields. For instance, *Wikipedia*[9], is a multilingual online encyclopedia, which is maintained by a community of volunteer editors. It is currently the largest and most popular online general reference work (Tancer, Bill, 2007).

Crowdsourcing is nowadays also a rapidly developing method of data acquisition in the field of Geo-spatial Science. *"Citizen as sensors"* is a well-known concept, where crowdsourcing is used to obtain geospatial information (Heipke, 2010). Volunteered Geographic Information (VGI), first coined by Goodchild (2007), denotes crowdsourced geospatial data. It is defined as "the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals". OpenStreetMap[10] (OSM) is one of the most successful VGI applications used for collaborative mapping. Contributors can insert, edit and delete map features based on their measurements with GPS devices or digitization according to aerial imagery. The collected data, in this case, is well-structured, which can be easily used for other web applications. In addition to collaborative mapping, users are also asked to provide their observations on topics from various disciplines with geographic location voluntarily. Birdwatchers share their geo-referenced birding records on citizen science websites, such as eBird.org and ornitho.de, to facilitate the understanding of their biological patterns (Sullivan et al., 2009). Mobile apps have been developed to collect citizens' reports on Land Use and Land Cover (LULC) types to complement the authoritative LULC survey (Laso Bayas et al., 2016, 2020).

The value of VGI for disaster management was first observed in the behavior of citizens during the Southern California wildfires of 2007–2009 and documented in (Goodchild and Glennon, 2010). Volunteers shared wildfire reports with locations on Flickr, interpreted remote sensing imagery by themselves, and established map sites presenting both VGI and official information. It was also noticed that volunteers could, in certain circumstances, provide more timely situation information than official sources.

VGI can be distinguished in two approaches, participatory and opportunistic. A participatory approach requires a conscious and active participation by the users. In contrast, the opportunistic ap-

[9] Wikipedia. `https://www.wikipedia.org/` (Accessed on 31.01.2021)
[10] OpenStreetMap. `https://www.openstreetmap.org/` (Accessed on 31.01.2021)

proach acquires the information in a quasi unconscious and passive manner. The above-mentioned studies were using the participatory approach, which relies heavily on user engagement. In order to do so, people have to install a specific App or use a dedicated web application to provide their input. In addition, users are often required to register for a new account. This is considered as inconvenient for users – especially if they are only infrequent users. As a result, the opportunistic VGI is more desirable and becomes the focus of this thesis.

One of the first sources of opportunistic VGI is social media, which has been used in a wide variety of studies. Texts from social media can be used for mining public opinions during elections (O'Connor et al., 2010), or influenza surveillance (Broniatowski et al., 2013). Since 2009, Twitter has been supporting users to share geographic locations (Stone, Biz, 2009). And since then, social media data were used to analyze different natural disasters, e.g., earthquakes (Sakaki et al., 2010), floods (Schnebele and Cervone, 2013), storms (Huang and Xiao, 2015), or fires (De Longueville et al., 2009). Social events were also studied, such as protest (Earl et al., 2013), social unrest (He et al., 2015), or stampede events (Zhu et al., 2019). Social media has become an essential source of information for many studies that want to utilize user-generated information. Therefore, in this thesis, social media is the primary source of information to extract flood-related VGI.

In addition to social media, user-provided trajectory data are also considered as opportunistic VGI, as they reflect the user behaviors in a geographic environment. Trajectory mining is an active research area, which has been used to generate road network (Lyu et al., 2017), detect traffic regulators (Cheng et al., 2020), or to produce terrain models (Massad and Dalyot, 2015). Severe precipitation and flood events can cause traffic slowdowns and even suspension of traffic flow. It therefore can be expected to also have effects on traffic and traffic-related VGI data. Today, VGI has become an important source of information for navigation service providers to provide real-time traffic information. Therefore, in this thesis it is intended to investigate whether an indication of precipitation can be obtained from vehicle speed data.

1.2 Goal of this thesis

The goal of this thesis is to explore to what extent precipitation and flood observations can be automatically extracted from the opportunistic VGI. This thesis mainly focuses on the following three research questions:

- **Can precipitation indications be extracted from the speed variation of road users?**

  Severe precipitation and flooding can cause traffic slowdowns or even suspensions. The slowdown on a few roads is probably due to local events, such as concerts, football matches, traffic accidents. Local events usually have a limited impact around the location of the event. If the slowdown affects most roads in an area, it is mainly caused by regional events, such as precipitation. To investigate this hypothesis, a proof-of-concept study is conducted to verify whether such a pattern can be identified from the speed variation of road users and consequently be used as a precipitation indicator.

- **What is the benefit of jointly exploiting text and images from social media to extract high-quality pluvial flood observations?**

  Social media users often provide texts and photos related to flood and rainfall events with geolocations. However, most of the previous applications only used keyword filtering or classical language processing methods to identify disaster-related social media posts. The

visual information in the photos is mostly unused. To answer this research question, a framework is developed to automatically collect and analyze textual and visual information from social media.

- **How can the flood severity information be interpreted from social media images and how far they are helpful for flood severity mapping?**

  Flood severity is related to the inundation level. In this thesis, a novel approach is developed which uses objects with known fixed size as levels to determine the water heights above ground. Automatic interpretation of this information from social media images can provide more detailed flood observations. People in floodwater can be used as targets to verify the feasibility of the idea. It is also interesting to know how this method performs compared to other data sources, such as remote sensing, when responding to a real-world flood event.

In addition to answering these research questions, the contributions of this thesis are

- This is the first proof-of-concept study to verify whether the road users' speed variations on multiple roads can be used as a precipitation indication.
- This work presents an early framework leveraging both text and image information to extract flood-related observations for mapping social media clusters and hotspots of real-world flood events.
- A novel approach is developed to analyze flood water levels for the scenarios when humans stand in floodwater.
- Images-based water level analysis is the first time applied to geotagged social media posts of a real-world flood event. The extracted information is used for flood severity mapping.

1.3 Outline

The remainder of this thesis is organized as follows: Chapter 2 introduces the basics of machine learning, the techniques for interpreting texts and images, and the tools for spatio-temporal analysis. In addition, data sources and characteristics of opportunistic VGI are summarized. Chapter 3 reviews the relevant studies on the acquisition and interpretation of flood and precipitation observations from VGI. Chapter 4 presents the method and experiment to extract binary precipitation observations from the speed variations of road users to answer the first research question. Chapter 5 introduces the proposed methods to extract flood observations from social media. Subsequently, Chapter 6 details the experiments with these presented methods, including the framework for social media data acquisition, and two case studies of the 2016 European floods and the 2017 flood event caused by Hurricane Harvey in Houston, USA. The pros and cons of opportunistic VGI for flood and precipitation monitoring are discussed in Chapter 7. In the last chapter, the research questions are answered and conclusions are drawn. Furthermore, future research directions are summarized based on the presented results.

2 Background

This chapter presents the basic technologies and background knowledge underlying the presented work. Section 2.1 introduces the fundamentals of machine learning for supervised classification as well as the evaluation metrics. In Section 2.2, computer vision techniques for image classification and interpretation using Deep Convolutional Neural Network (DCNN) are explained. In Section 2.3, Natural Language Processing (NLP) techniques for text classification are presented. Section 2.4 addresses the basic principles and concepts of spatial and spatiotemporal analyses, which are used in the analysis of opportunistic VGI. At last, the characteristics of opportunistic VGI is detailed in Section 2.5.

2.1 Machine learning

Machine learning, as an essential part of artificial intelligence, aims to build systems that can learn from sample data (Goodfellow et al., 2016b). There are many applications of machine learning in everyday life. From the most basic applications, such as email spam detection and credit card fraud detection, to more advanced autonomous driving cars that perceive their surroundings through sensor data (Yurtsever et al., 2020), machine learning acts as the theoretical basis. In general, machine learning algorithms are categorized into supervised and unsupervised learning. Both types require a dataset with features. However, only the dataset for supervised learning is associating examples with labels (Goodfellow et al., 2016b). As far as supervised learning is concerned, classification and regression are the two most basic tasks. The classification predicts which pre-defined category or categories an input belongs to, while the regression predicts numeric values as output. More attention is paid to the supervised classification approaches in this thesis because the flood- and rainfall-relevant targets are specific and can be categorized with proper labels. In Section 2.1.1, three representative machine learning algorithms are described. The metrics and methods used to evaluate the classification performance are presented in Section 2.1.2.

2.1.1 Supervised classification

Supervised classification aims to learn models based on data or features annotated with categorical labels. There are many models developed for this type of tasks. Only the basics of the methods used in this thesis are presented, namely Support Vector Machine (SVM) (Boser et al., 1992; Cortes and Vapnik, 1995), random forest (Ho, 1995), and Xgboost (Chen and Guestrin, 2016). Nowadays, these methods are well available as software packages in different programming languages, e.g., scikit-learn[1] (Pedregosa et al., 2011) in Python, WEKA[2] (Hall et al., 2009) in Java, and Xgboost[3] in multiple programming languages.

[1]Scikit-learn: machine learning in Python. `https://scikit-learn.org/` (Accessed on 31.01.2021)
[2]WEKA 3. `https://www.cs.waikato.ac.nz/ml/weka/` (Accessed on 31.01.2021)
[3]Xgboost. `https://github.com/dmlc/xgboost` (Accessed on 31.01.2021)

Support Vector Machine

Support Vector Machine (SVM) is a non-probabilistic learning model, which attempts to classify data points into binary categories with a learned decision boundary. This decision boundary can be linear and also non-linear via a kernel trick (Boser et al., 1992). As illustrated in Figure 2.1, a line (red) is estimated from binary labeled data points ($\boldsymbol{x_1}$,y_1),...,($\boldsymbol{x_n}$,y_n), where a separation with the largest margin is achieved.

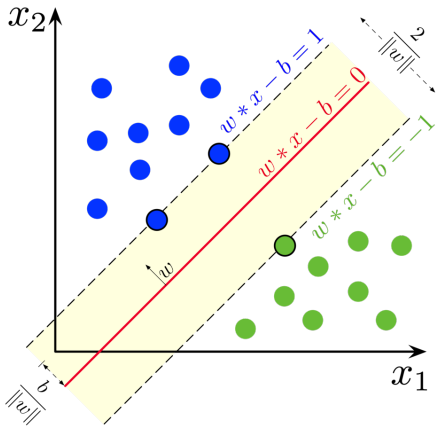


Figure 2.1: Example of margin of a linear SVM model (image under CC BY-SA 4.0).

The distance from the function to the nearest data point of each class is defined as the margin, and these data points represent the so-called support vectors. The hyper-plane (i.e., the red line in Figure 2.1) to be estimated is denoted as

$$\boldsymbol{w}^T\boldsymbol{x} - b = 0, \tag{2.1}$$

where $\boldsymbol{w}$ is the normal vector of this linear function and b is the intercept. The margin is denoted as $\frac{2}{||\boldsymbol{w}||}$. For this easily separable case shown in Figure 2.1, a hard margin solution $\boldsymbol{w}$ and b can be obtained via optimization, i.e., minimizing

$$\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \quad s.t. \quad y_i(\boldsymbol{w}^T\boldsymbol{x_i} - b) \geq 1 \quad for \quad i = 1,...,n. \tag{2.2}$$

Since hard margin SVM can only be used for easily separable data without outliers, hinge loss $\mathcal{L}_{hinge}(z) = max(0,\, 1 - z)$ is added to the cost function to allow a reasonable amount of outliers, which results in the cost function to minimize is

$$\min \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{n}\xi_i \quad s.t. \quad y_i(\boldsymbol{w}^T\boldsymbol{x_i} - b) \geq 1 - \xi_i \quad \xi_i \geq 0 \quad for \quad i = 1,...,n, \tag{2.3}$$

where ξ_i represents the deviations to the functional margin, and C is a penalized hyper-parameter that balances the over-fitting and under-fitting problems.

For the linearly inseparable cases, the kernel trick (Boser et al., 1992) is applied to transform the input features to a new feature space that is more easily separable by linear functions. Non-linear classification can be achieved without a significant increase in computation cost. RBF (Radial Basis Function) kernel is one of the most commonly used kernels,

$$K(\boldsymbol{x}, \boldsymbol{x'}) = \exp(-\frac{||\boldsymbol{x} - \boldsymbol{x'}||}{2\sigma^2}), \tag{2.4}$$

where $\boldsymbol{x}$ and $\boldsymbol{x'}$ are the input features, and σ determines the influence of a single data point. Further developments of SVM have supported multi-class classification by ensemble multiple binary SVM classifiers.

SVM is a method that is often used to train supervised classification models. It is applied in Chapter 4 to train the binary precipitation indicator and in Section 5.1 to train text classification models.

Random Forest

Random Forest is a representative supervised learning model, which is the ensemble of multiple binary decision trees as shown in Figure 2.2. This model intends to use a set of weak prediction models to create a single robust prediction model.

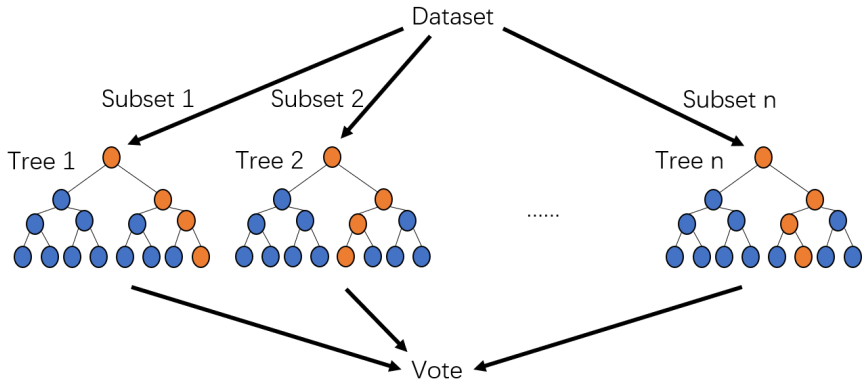


Figure 2.2: Example of a random forest model.

The binary decision tree is a standard tree model. The tree nodes correspond to binary conditions, and each tree leaf corresponds to a decision, i.e., the prediction for the classification task. Each node is a hard split of one of the feature dimensions, which subdivides the feature space recursively into regions. Each region corresponds to a tree leaf. A single tree can be learned with a set of supervised samples, using for example, ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993) or CART (Breiman et al., 1984). In the prediction process, a data sample is passed to the root node of a tree until it reaches one of the leaves, as shown in the highlighted nodes (in orange color) of every single tree in Figure 2.2.

Random forest (Ho, 1995) learns with bagging strategy, i.e., randomly sampling subsets from training samples with replacement. For each subset, a binary decision tree is learned. All these trees can be trained in parallel. In the splitting process of each node, features are chosen randomly. The final prediction of the model is obtained by a majority voting on the predictions from multiple independently trained decision trees with the same data input. Each tree is treated with equal importance.

Random forest is used in Chapter 4 to train the binary precipitation indicator and Chapter 5 to train text and image classification models.

Gradient boosting and Xgboost

Gradient boosting is one of the ensemble learning algorithms that learns with a boosting strategy. Boosting learns in an iterative way. In each iteration, a weak classifier (e.g., a decision tree) is learned. The misclassified samples in the current iteration are given with higher weights into the next iteration. Thus the model can focus more on those samples that are not easily classified. The final prediction is the sum of all the predictions of individual weak classifiers. In comparison to standard boosting methods, gradient boosting uses a differentiable loss function during the optimization. During the updates, the new predictor learns the residual error of the last predictor as illustrated in Figure 2.3.

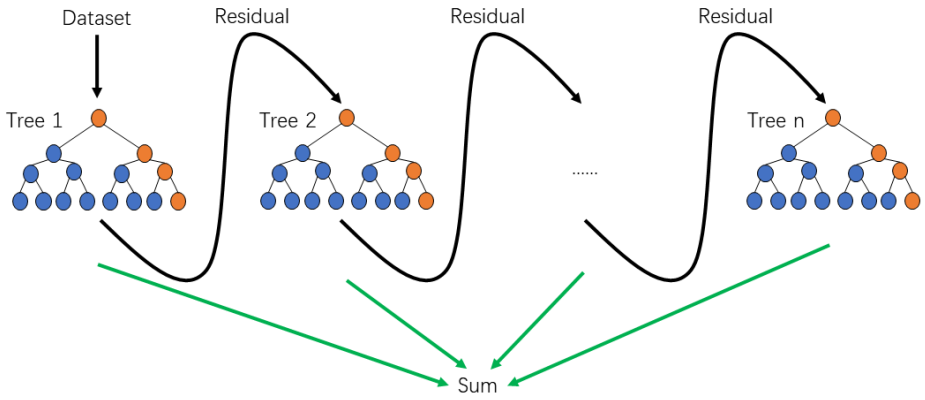


Figure 2.3: Illustration of Gradient Boosting Decision Trees.

Xgboost (Chen and Guestrin, 2016) is a scalable end-to-end implementation of a gradient boosting model (Friedman, 2002). It runs much faster and needs fewer computational resources, compared to the previous gradient boosting implementations. Xgboost additionally computes a second-order gradient during optimization and includes regularization for the cost function to prevent model overfitting. This method has received much attention for its excellent learning results and fast training speed.

Gradient boosting and Xgboost are used in Chapter 4 to train the binary precipitation indicator. Xgboost is used to train the image classification model (Section 5.2) as well as the water level estimation model (Section 5.3).

2.1.2 Classification evaluation

Evaluation is an essential step after training a supervised classifier. Based on a dataset with ground truth labels and predicted labels, multiple metrics are calculated to evaluate the performance of classification models. When comparing predicted labels with ground truth labels, there are four possibilities: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). As for binary classification, TP is the number of samples that the model predicts as “positive” and are actually also “positive”. TN is the number of samples that the model predicts as “negative” and are actually “negative” as well. Both TP and TN are the cases when the model predicts correctly. FP is the number of samples that the model predicts as “positive” but are actually “negative”. FN is the number of samples that the model predicts as “negative” but are actually “positive”. FP and FN are the cases when the model predicts wrongly. The numbers of samples for all these cases are often presented in a confusion matrix shown in Table 2.1.

Table 2.1: Example of a confusion matrix for binary classification.

		Actual class		
		Positive	Negative	Total
Predicted class	Positive	TP	FP	TP+FP
	Negative	FN	TN	FN+TN
	Total	TP+FN	FP+TN	N

With the number of samples for the above four cases, measures such as True Positive Rate (TPR, also known as recall), False Positive Rate (FPR), precision, F_1-score, and accuracy can be calculated with the equations as follows,

$$\text{TPR} = \text{Recall} = \frac{\text{TP}}{\text{TP+FN}}, \tag{2.5}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}}, \tag{2.6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}, \tag{2.7}$$

$$F_1 = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP+FP+FN}}, \tag{2.8}$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}. \tag{2.9}$$

Precision, recall, and accuracy are the most basic measures for classification models. F_1-score is the harmonic mean of precision and recall, which can better reflect the model's overall performance.

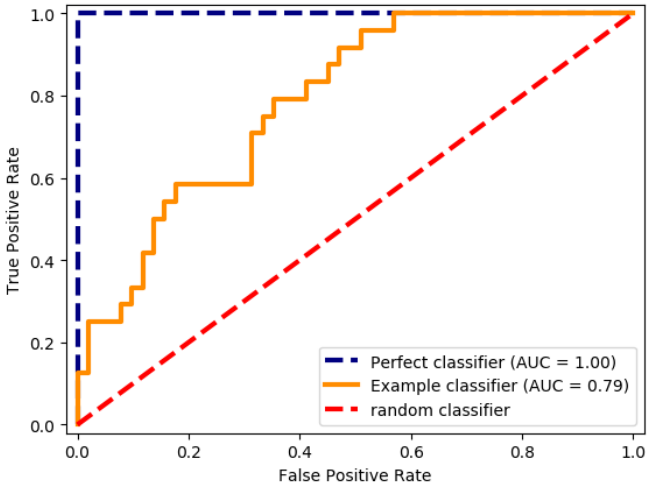


Figure 2.4: Example of Receiver Operating Characteristic (ROC) curves (adapted based on code example under BSD license).

Many machine learning classification methods provide probabilities as outputs. However, taking different thresholds for these probability values can lead to significant differences in the metrics

mentioned above. In order to achieve a more rational model comparison, Receiver Operating Characteristic (ROC) curves (see Figure 2.4) are calculated with TPR and FPR at all classification thresholds. It measures how good a classifier can differentiate positive and negative examples. A random classifier is represented by the red dashed line, where a perfect classifier is represented by the blue dashed line. The ROC curves are often placed between the random and perfect classifier, presented as the yellow line in this example. The Area Under the Curve (AUC) is a quantitative metric for comparing ROC curves, independent of the selection of thresholds. The AUC of a perfect classifier is 1.0, and the AUC of the random classifier is around 0.5.

2.2 Image interpretation with Deep Convolutional Neural Networks

With the popularity of smartphones and mobile Internet, social media users can easily use photos to share events of their lives. There are several social media platforms for photo-sharing, such as Instagram[4] and Flickr[5]. User-generated images are an essential form of opportunistic VGI, which can provide critical observations for flood events. However, the vast majority of images on social media are unrelated to flood events. Therefore, efficient extraction of flood-relevant images out of a large number of irrelevant images is a prerequisite for the further analysis of social media data.

Image classification is one of the basic computer vision tasks. It has developed rapidly since Krizhevsky et al. (2012) introduced Convolutional Neural Network (CNN) (LeCun et al., 1989) for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Deng et al., 2009). It exhibits superior performance compared to previous studies using hand-craft features, and has been quickly extended to solve various other tasks beyond image classification. Since then, neural networks have been built with more convolutional layers, i.e., deep convolutional neural networks (DCNN). DCNN has been successfully extended to a broader range of computer vision tasks, such as image segmentation, object detection, and human pose estimation.

This thesis applies these three tasks to the estimation of flood water levels. Therefore, in Section 2.2.1, the basics of Convolutional Neural Network are summarized. The network architectures used for image classification, image segmentation, object detection, and human pose estimation that contribute to the approaches developed in this thesis are outlined in Section 2.2.2.

2.2.1 Basics of Convolutional Neural Network

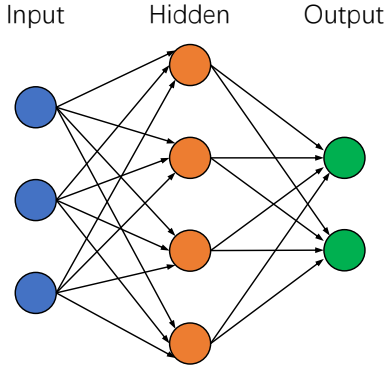


Figure 2.5: Illustration of an Artificial Neural Network.

[4]Instagram. `https://instagram.com/` (Accessed on 31.01.2021)
[5]Flickr. `https://www.flickr.com/` (Accessed on 31.01.2021)

Artificial Neural Network (ANN) is one of the machine learning algorithms, which is the basis of CNN. It consists of an input layer, an output layer, and a series of hidden layers in between. The basic structure of ANN is illustrated in Figure 2.5.

Fully connected layers are similar to a traditional multi-layer perceptron neural network (MLP). Each layer contains multiple nodes, i.e., the neurons. The nodes in one layer are connected to every nodes in another layer, where the values $\boldsymbol{x}$ of all input neurons are linearly combined with a weight $\boldsymbol{W}$ and a bias b. They are then passed to a non-linear *activation function* $g(z)$ that provides values for the nodes at the next layer. Such an operation is denoted as

$$\boldsymbol{a} = g(\boldsymbol{W}^T\boldsymbol{x} + b). \tag{2.10}$$

Many non-linear functions can be used as activation functions, such as sigmoid (Eq. 2.11), hyperbolic tangent - tanh (Eq. 2.12), and Rectified Linear Unit - ReLU (Eq. 2.13). The most commonly used activation function is ReLU. Compared with the other two, ReLU mitigates the vanishing gradient problem[6] and is also faster to compute.

$$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{2.11}$$

$$tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}, \tag{2.12}$$

$$ReLU(x) = \max(0,x). \tag{2.13}$$

At the end of an ANN, the output layer needs to provide outputs that are easy to interpret, especially for classification tasks. The sigmoid function is often used as the activation function at the output layer for binary classification tasks, which outputs values between 0 and 1. For multi-class classifications, softmax is mainly used to normalize the outputs for all categories to a vector whose elements add up to one. The softmax function is denoted as

$$softmax(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}, \tag{2.14}$$

where z is the output after applying the linear combination with K elements, corresponding to the pre-defined number of categories.

With the nesting of multiple structures like this, a feed-forward network $\hat{\boldsymbol{y}} = h(\boldsymbol{x};\theta)$ can be built to map the data inputs $\boldsymbol{x}$ to the outputs $\hat{\boldsymbol{y}}$. θ are the parameters to be learned, including the weights and biases between all adjacent two layers. The learning of this network is mainly achieved by backpropagation, which aims to minimize the errors (i.e., the *loss function* $\mathcal{L}$) between ground truth and feed-forward predictions. *The loss function* is highly task-dependent. For a multi-class classification task, cross-entropy (CE) loss is frequently used. For a dataset of N samples with K classes, CE loss $\mathcal{L}_{CE}$ is defined as

[6]The vanishing gradient problem is a typical problem encountered when training neural networks with gradient-based optimization methods. During backpropagation, the gradient decreases exponentially as the number of layers increases. This problem can lead to prolonged training in the early layers. Skip-connection, introduced by ResNet, was applied in DenseNet to solve this issue. However, the difference is that each convolutional layer in the dense block uses the feature maps of all previous layers as input. (Huang et al., 2017a)

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} y_{ij} log(\hat{y}_{ij}), \tag{2.15}$$

where $\hat{y}_{ij}$ corresponds to the softmax output for i^{th} sample at j^{th} class. y_{ij} is the one-hot coded ground truth label for i^{th} sample at j^{th} class. Binary cross-entropy (BCN) is a special case of CE when K is 2, which is used for a binary classification. BCE loss $\mathcal{L}_{BCE}$ is defined as

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{2} y_{ij} log(\hat{y}_{ij}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1-y_i) log(1-\hat{y}_i)]. \tag{2.16}$$

The Mean Squared Error (MSE) is a loss function often used in regression tasks, i.e., predicting real numbers instead of class labels. For a dataset of N samples, MSE loss $\mathcal{L}_{MSE}$ is defined as

$$\mathcal{L}_{MSE} = -\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{2.17}$$

where y_i is the reference ground-truth value, and $\hat{y}_i$ is the predicted value from the regression model. In this case, an L2-norm is used.

CNN is an algorithm for image classification that builds upon the ANN. The basic structure of a CNN is illustrated in Figure 2.6. Convolution is a basic operation frequently used in image processing, e.g., for blurring, sharpening, and edge detection. A convolution kernel slides over the entire image and calculates for each pixel the corresponding dot product between filter weights and the input raster cropped by the window. The filters in convolutional layers extend by the full depth of the input. The number of pixels that the filter moves at each step is called stride. The weights in the filter are learnable. Similar to ANN using backpropagation, these weights are updated each time incrementally with respect to the learning rate. Compared to applying a fully-connected layer on pixel-like data, a convolutional layer requires far fewer parameters, as the same parameters of the convolutional kernel are applied to the entire image. The convolutional layer's output is regarded as a feature map, which preserves the spatial structure of the input volume.

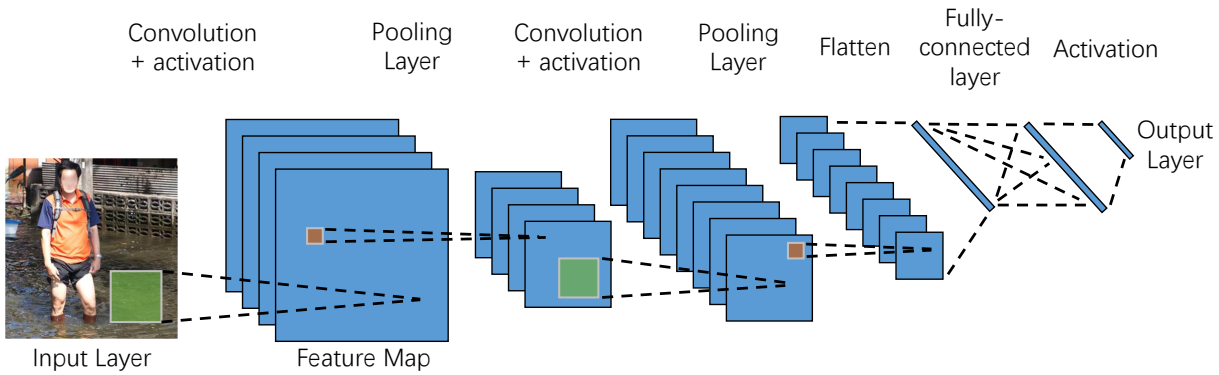


Figure 2.6: Illustration of a Convolutional Neural Network.

The convolutional layers are often followed by *activation function* and *pooling layers*. *Pooling layers* are often used to reduce the spatial size of the feature maps while increasing the receptive field of filters and preserving the spatial structure of features. The most commonly used pooling

layer is max-pooling, where the down-sampled cells are represented by the maximum values of the corresponding original cells. However, many network architectures in recent years have replaced the pooling layer with the convolution layer with an increased stride, according to the research from Springenberg et al. (2015). For example, applying a convolutional layer with stride two and padding can also reduce the feature maps to 1/4 of their spatial size.

After applying the convolutional layers, activation layers, and pooling layers, the extracted feature maps are flattened to generate a vector representation of the image for classification (see Figure 2.6). With the fully connected layers at the end of the neural network, classification predictions are obtained similar to an ANN.

The training of a neural network is the process of optimizing the weights of a network with respect to a loss function. Most *optimizers* used for deep learning apply minibatch-based learning, where the loss function is iteratively optimized based on only a small subset of the training samples at a time. Stochastic Gradient Descent (SGD) is a commonly used method, requiring the hyperparameter learning rate, which is challenging to set. It also suffers from problems such as small but consistent gradients, high curvature, or noisy gradients (Goodfellow et al., 2016c). In order to accelerate this process, methods with momentum and adaptive learning rate are more frequently used, such as RMSProp (Tieleman and Hinton, 2012), or Adam (Kingma and Ba, 2015) optimizers.

2.2.2 Common tasks in computer vision

With the development of computer vision, CNNs have achieved great success with image classification tasks. They have been extended to solve many other computer vision tasks, such as image segmentation, object detection, and human pose estimation. As these tasks contribute to the approaches developed in this thesis, representative solutions for these tasks are presented in the following.

Image classification

Image classification is an essential computer vision task, which assigns unique labels to images. The most classic tasks include handwritten digit recognition, or color image classification containing categories such as airplanes, cars, birds, cats, etc. Following the success of AlexNet (Krizhevsky et al., 2012), several networks have been proposed to improve the performance of image classification tasks by stacking more convolutional layers, e.g., the well-known VGG16 model (Simonyan and Zisserman, 2014) is a stack of 16 weight layers. In addition to building deeper models, GoogLeNet (a.k.a. InceptionV1) (Szegedy et al., 2015) applied dimension reduction using 1x1 convolution, which reduces the depth of feature maps while preserving the important features, and significantly improves computational efficiency. GoogLeNet applies filters with multiple kernel sizes and concatenates the resultant feature maps for image classification. Its successor, InceptionV3 (Szegedy et al., 2016) achieved a new state-of-the-art performance in 2015 by using consecutive small-kernel-size filters (3x3), instead of filters with 5x5 or 7x7 kernels.

However, building deeper models (i.e., stacking more convolutional layers) does not necessarily improve the model's performance. In this regards, He et al. (2016) demonstrated the drawbacks of simply stacking convolutional layers. Instead, they introduced the ResNet (Deep Residual Network). Rather than learning the desired mapping directly, the residual blocks in ResNet attempt to learn the residuals via shortcut connections. The shortcut connection sums the input and output of the current block and feeds it to the subsequent layers. With this structure, deeper CNNs can be built without degradation problems. The ResNet has further inspired the improvement of the

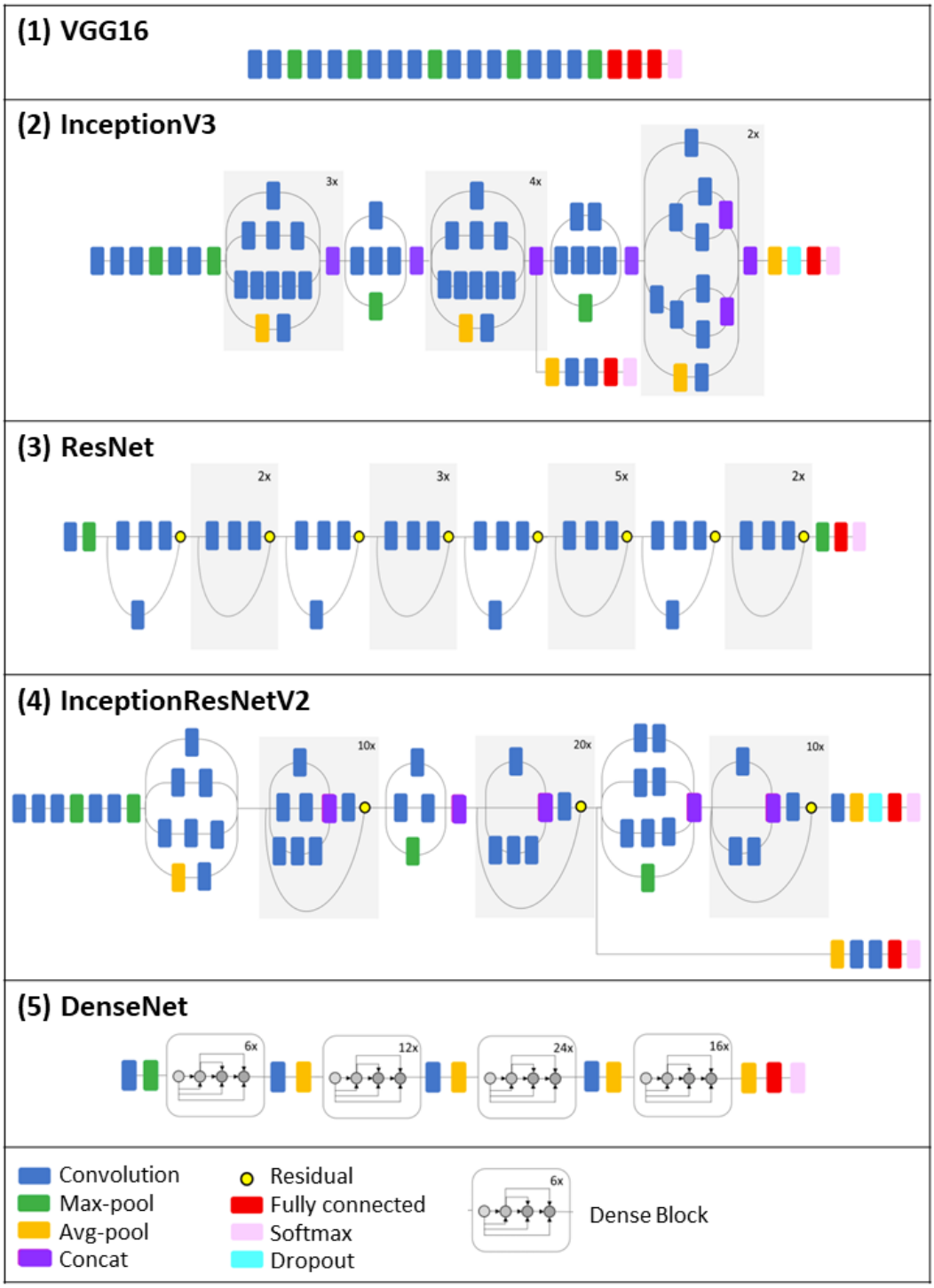


Figure 2.7: Overview of the network architectures: VGG16, InceptionV3, ResNet, InceptionResNetV2 and DenseNet for image classification (adapted based on Figures 1-6 from Mahdianpari et al. (2018) under CC BY 4.0).

Inception series. For example, InceptionResNetV2 (Szegedy et al., 2017) combines residual blocks with the inception module. This strategy further improves the image classification performance. Later on, DenseNet (Huang et al., 2017a) introduced a structure named dense block, where each convolutional layer in the block uses the feature map of all previous layers as inputs. Dense blocks encourage feature reuse, enhance feature propagation, mitigate the vanishing-gradient problem, and significantly reduce the number of parameters. In order to visualize the evolution of DCNN for image classification, Figure 2.7 presents an overview of some of the network architectures mentioned above, including VGG16, InceptionV3, ResNet, InceptionResNetV2, and DenseNet. These model architectures are used in this thesis to extract visual features to classify flood-relevant social media images.

Nowadays, DCNNs contain numerous convolutional layers and millions of parameters to train. However, for flood-related image classification, only a very limited number of training samples are available. When training a DCNN from scratch, this leads to a big concern of overfitting. A possible solution is to use a pre-trained model, where deep features are already learned. This model is then applied to other image recognition domains, as described in DECAF (Donahue et al., 2014). *Transfer learning* (Goodfellow et al., 2016a) is a common strategy for image classification with fewer training examples. Pre-trained DCNNs on a very large dataset (e.g., ImageNet) are often used to initialize a DCNN and adapt it by a custom output layer. A pre-trained DCNN can also be used as a feature extractor and these features can then be classified using machine learning classifiers, such as Support Vector Machine (SVM) or logistic regression. Transfer learning is utilized in many fields of research, such as the classification of satellite images (Nogueira et al., 2017b), or vehicles detection in RGB images or LiDAR data (Niessner et al., 2017; Ammour et al., 2017).

Semantic image segmentation

Image segmentation is another widely studied computer vision task that assigns labels to each pixel of an image. Various convolutional architectures have been developed for semantic image segmentation. Fully Convolutional Networks (FCN) (Long et al., 2015) are one of the earliest deep learning models for semantic image segmentation. It used the convolutional neural network for image classification as an encoder, where the images are encoded as features. With transpose convolutional layers, coarse features from the encoder are upsampled in a decoder. The decoder is used to decode the features to a prediction with the same spatial dimension as the input image. Thus, it uses the same number of upsampling operations as downsampling, which can provide at the end a full-resolution pixel-wise prediction. Since then, many approaches have been developed using such an encoder-decoder network architecture, e.g., UNet (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017).

Deeplabv3+ (Chen et al., 2018) is one of the state-of-the-art network architectures, which combines several DCNN architectures and components from previous research, including the atrous convolution, atrous depthwise separable convolution, Atrous Spatial Pyramid Pooling (ASPP), and encoder-decoder network architecture.

Atrous convolution (a.k.a. dilated convolution) was introduced in (Yu and Koltun, 2016), which is a technique to replace the pooling layers in FCN. It increases the receptive field of filters without losing detailed information when reducing the spatial resolution. Compared to a standard convolution described in Section 2.2.1, atrous convolution defines a spacing between the values in a kernel with respect to a hyperparameter – the dilation rate. Figure 2.8 illustrates an example, where atrous convolution is applied to an image using a 3x3 kernel with a dilation rate of 2.

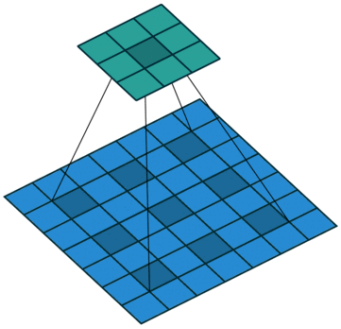


Figure 2.8: Example of atrous convolution (image under MIT License).

Depthwise separable convolution is a strategy used in Xception (Chollet, 2017), where a standard convolution is replaced by two steps, i.e., a depthwise convolution and a pointwise convolution (i.e., 1x1 convolution). It reduces the computation cost significantly while preserving the performance. Chen et al. (2018) adapted the depthwise convolution in Xception with atrous convolution and used it as the encoder of Deeplabv3+ (as the upper white blocks shown in Figure 2.9).

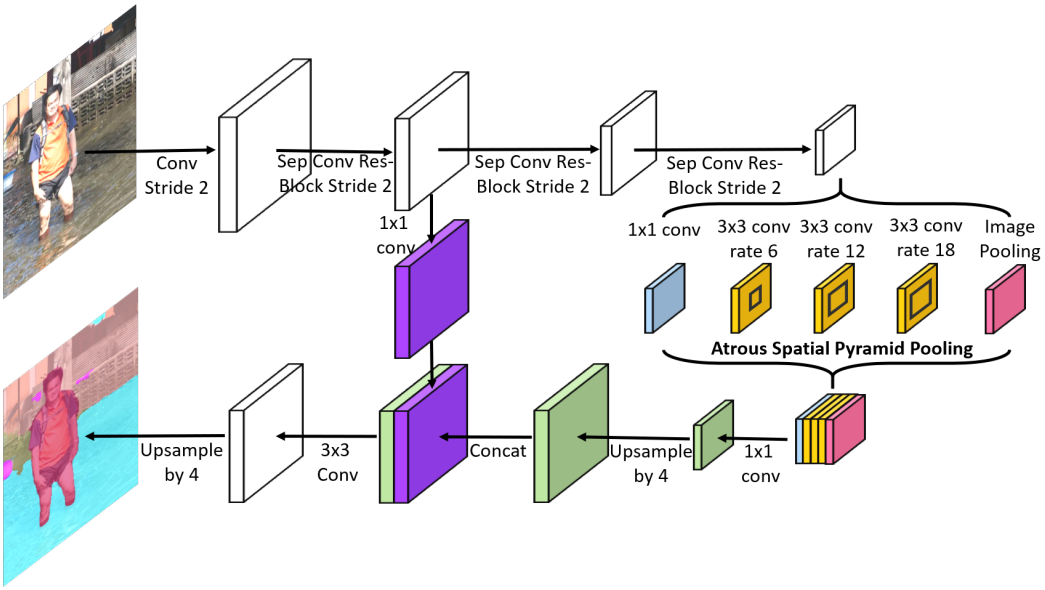


Figure 2.9: Network architecture overview of DeeplabV3+ (Chen et al., 2018).

Figure 2.9 presents the overview of Deeplabv3+ network architecture. The three-channel image is first fed to an encoder consisting of a set of convolutional blocks using atrous depthwise separable convolution and the Atrous Spatial Pyramid Pooling (ASPP). ASPP is oriented from Spatial Pyramid Pooling (SPP), which is introduced in DeeplabV2 (Chen et al., 2017). Deep features from atrous convolution with one 1x1 convolution and three 3x3 convolutions with dilation rates 6, 12, 18 are concatenated together with image pooling, which provides the image-level feature (denoted as the part in the parenthesis in Figure 2.9). With this, multi-scale contextual information is encoded by applying atrous convolution at multiple dilation rates. Encoder-decoder is the most basic network structure, which has been used in many approaches for semantic image segmentation. By deploying this structure, Deeplabv3+ can capture sharper object boundaries by adding the skip-connection between the encoder and decoder. The concatenated multi-scale deep features are then passed through a decoder network, where low-level features from the encoder branch are combined.

Deeplabv3+ has been used in Section 5.3.1 to provide the surrounding information for each detected person in the flood-relevant images.

Object detection

Object detection is a computer vision technique, which detects and delineates object instances of semantic categories on an image. Mask R-CNN (He et al., 2017) is one of the leading DCNN architectures that can detect objects with a bounding box and a semantic class. In addition, this architecture provides a segmentation mask for each detected object. It has a two-stage structure, as illustrated in Figure 2.10, which is explained in more detail in the following paragraphs.

The first stage is to generate detection proposals. A backbone network is used to generate deep features based on the entire input image. The backbone network can have a variety of options. A common choice is the ResNet (He et al., 2016) combined with Feature Pyramid Network (FPN) (Lin et al., 2017). FPN is an encoder-decoder-like network structure, which outputs image deep features at multiple scales. With these extracted multi-scale deep features, a Region Proposal Network (RPN) (Ren et al., 2016) predicts a set of Region of Interest (ROIs) as rectangular boxes. In order to adapt to different shapes of the ROIs, RoIAlign generates a fixed-size feature map for each ROI using a bilinear interpolation.

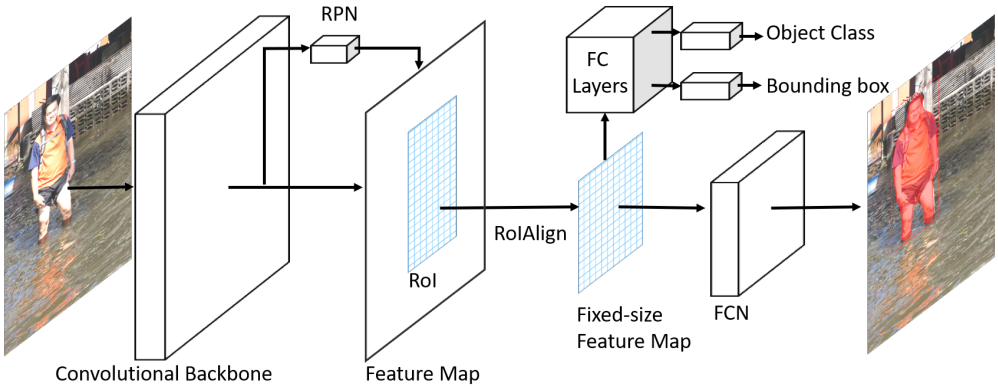


Figure 2.10: Network architecture overview of Mask R-CNN (He et al., 2017).

The second stage is detection, where the fixed size feature maps are fed to three branches of neural networks, which can predict the semantic class, regress the proposal to the object bounding box, and also a network branch using FCN to predict an object segmentation mask. Since there are multiple tasks to be achieved in one end-to-end model, the total loss function is a linear combination of the losses of all single branches.

Mask R-CNN is used in Section 5.3.1 to detect people in the social media images. The bounding box of each detected person provides a bottom-line indicating the visible part of a detected object.

Human pose estimation

Human pose estimation is another computer vision task, which has attracted much attention in recent years. Human joints, such as elbows, knees, etc., are identified from images or videos. The development of this technology has benefited many applications such as human-computer interaction (HCI), video surveillance, gaming, physiotherapy, movies, dancing, and sports (Sarafianos et al., 2016).

DeepPose (Toshev and Szegedy, 2014) is the very early attempt to apply a deep convolutional neural network for a single person's 2D pose estimation task. This model firstly learns a coarse keypoint location with an L2 regressor. The area around this predicted coordinate is then fed to a cascaded regressor to refine this coordinate. Instead of learning the Cartesian coordinate regressors, most recent approaches predict keypoints as heatmaps. Convolutional Pose Machines (CPM) (Wei et al., 2016) is a multi-stage solution, which inherits the cascaded structure of DeepPose. The first-stage predicts a coarse heatmap for each body part separately via a DCNN. The latter stages use both the deep features from the input image and heatmaps from the previous stage to predict the refined heatmaps for each body part. Cao et al. (2017) further combined CPM with Part Affinity Fields (PAF), which is able to achieve real-time pose estimation for multi-person. PAF learns to associate body parts based on the information in between. For each pixel around the keypoint connection, a direction vector is estimated and used as the ground truth value for PAF. A deep model learns to predict such a direction vector, which helps to reject incorrect connections between body parts. In (Cao et al., 2017), both the branches for body part locations and PAF are trained jointly, and the model predicts 2D keypoints for multiple people in an image.

OpenPose (Cao et al., 2019) is an open-source software for multi-person 2D pose detection. It is optimized based on a series of research including (Wei et al., 2016; Cao et al., 2017; Simon et al., 2017). Figure 2.11 illustrates the network architecture of OpenPose. Deep features are firstly extracted by a VGG-19 for the entire image and then fed to a two-branch network. The first branch predicts the PAFs with a cascaded structure of T_p stages. The second branch predicts confidence maps (i.e., heat maps) for each body part based on the input of PAFs and VGG deep features. It is a cascaded structure of another T_c stages. In the end, with a greedy parsing, the model iteratively detects people according to PAFs and a pre-defined tree structure.

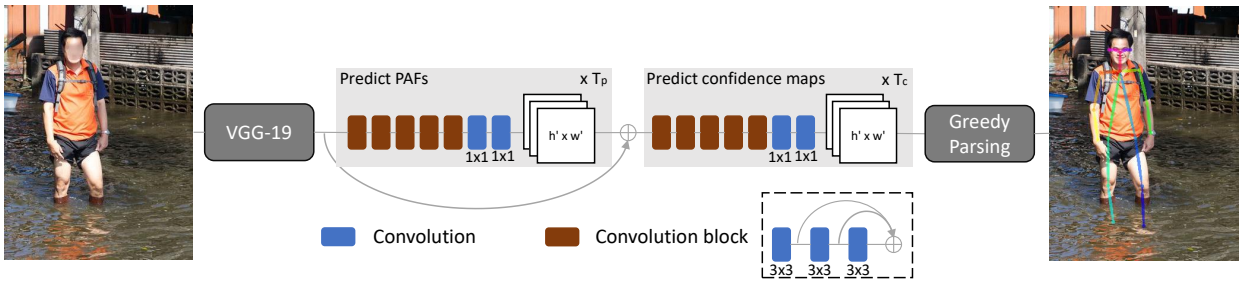


Figure 2.11: Network architecture overview of OpenPose (Cao et al., 2019).

OpenPose is used in Section 5.3.1, where the body keypoints are identified and used as a basis for the proposed water level estimation method.

2.3 Text analysis with Natural Language Processing

Text is the most basic form of user-generated content and accounts for the majority of users' input, especially on Twitter. It is another information source where users may provide their observations and subjective opinions on flood events. The understanding of user-generated texts is essential for analyzing opportunistic VGI. However, social media posts related to flood events account for only a very small proportion of the social media data stream. Therefore, the extraction of flood-relevant information is an essential initial step.

Natural Language Processing (NLP) is a technique to extract target documents from large amounts of textual data. Document classification is the task of assigning documents to pre-defined cate-

gories. Since the objectives in this thesis are specific and can be categorized with proper labels (i.e., flood-relevant and irrelevant), supervised document classification is needed. In the early days of NLP, text documents were represented as vector representations through statistics, treating each word as a feature dimension. The generated feature vectors of text documents can then be classified using classical machine learning models, such as SVM, Naive Bayes, random forest, etc. The basics of this approach are presented in Section 2.3.1. Nowadays, word embedding is a technique that has become more common. Words can be represented as word vectors, capturing the precise syntax and semantics through a neural network. Several representative word embeddings methods are summarized in Section 2.3.2. In the end, a CNN model for text classification using word embeddings is described in detail in Section 2.3.3.

2.3.1 Bag-of-Words

Bag-of-Words (BoW) model is a common method used in the early days of NLP for document classification. It is based on a simple assumption that a document is represented as a "bag" of independent terms, where the ordering of terms is ignored. A large collection of text documents that can be used for text analysis is called corpus. In practice, it is often in the form of large textual datasets with or without annotations, e.g., the IMDB dataset[7] containing 50,000 texts with positive and negative sentiment annotations, or the unannotated Google News dataset used for representation learning (Mikolov et al., 2013b).

Term frequency (tf) of each word is calculated according to all unique known words in the corpus. It is the raw count of each term in the sentence. It is the most basic way to represent the documents in the vector form, which can be easily adapted with machine learning algorithms for classification purposes. However, the importance of each word should be considered differently. Therefore, *document frequency*, the number of documents that contain each word, is used to mitigate the impact of terms that appear very frequently by scaling down their weights. The text documents are transformed into a sparse tf-idf (term frequency - inverse document frequency) matrix (Manning et al., 2008). It is a 1-V matrix, where V is the number of unique words in the whole corpus. Inverse document frequency indicates the rareness of the words. This value diminishes when the term occurs frequently. The tf-idf matrix can be calculated as follows:

$$\text{tf-idf}_{t,d} = tf_{t,d} \times idf_t, \tag{2.18}$$

$$idf_t = \log \frac{N_d}{df_t}, \tag{2.19}$$

where t stands for term index in the whole corpus, d for document index, N_d for total number of documents and df_t for document frequency of each word. This matrix can be calculated, e.g., using the methods offered by the scikit-learn library (Pedregosa et al., 2011). The classifiers can be trained based on this tf-idf matrix with the normal classification methods in machine learning.

This method has achieved satisfactory results for many early document classification tasks. However, due to the sparsity and high dimension of the matrix (e.g., the IMDB dataset has approximately 6.2 million dimensions when calculating the tf-idf matrix[8]), synonyms and phrases are not considered. Hence, the performance of such models is often limited.

[7] Large Movie Review Dataset. `http://ai.stanford.edu/~amaas/data/sentiment/`(Accessed on 31.01.2021)

[8] Sentiment Analysis of IMDB Movie Reviews - Kaggle. `https://www.kaggle.com/lakshmi25npathi/sentiment-analysis-of-imdb-movie-reviews` (Accessed on 31.01.2021)

2.3.2 Word embedding

Word embedding is the technique that represents words or phases as vectors of real numbers. Currently, the model to generate word embedding is mostly learned in an unsupervised manner. Word2vec (Mikolov et al., 2013a,b) provides the vector representations of words by a shallow two-layer neural network. There are two model architectures to generate the word embedding: CBOW (Continuous Bag-of-Words) and skip-gram (see Figure 2.12).

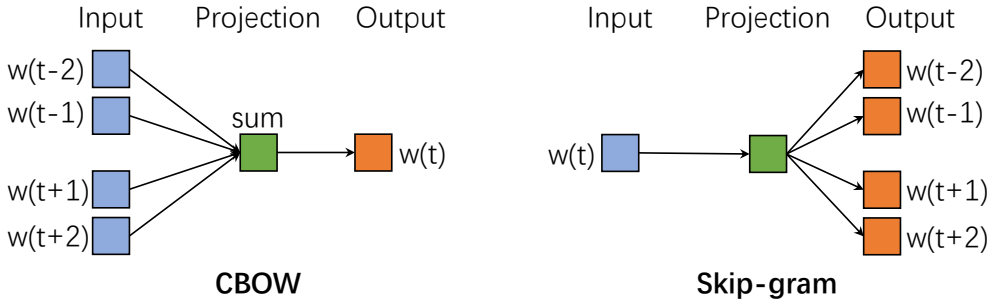


Figure 2.12: CBOW and skip-gram architectures for generating Word2vec word embedding (Mikolov et al., 2013a).

Based on the context, i.e., the words within a fixed size window in a sentence, CBOW predicts the target word while skip-gram predicts the context based on the target word. The objective is to find representations for words that are useful for these two tasks by maximizing the following probability for CBOW,

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}^{T}\log p(w_t \mid w_{t+j}), \tag{2.20}$$

and the following probability for skip-gram,

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}^{T}\log p(w_{t+j} \mid w_t), \tag{2.21}$$

where T is the number of words w_1, w_2, w_3, $\cdots$, w_T. c is the window size defining the training context. For example, for the sentence "It's raining and hailing like crazy" using a window size of 2, Table 2.2 shows the data input for CBOW while Table 2.3 shows it for skip-gram.

No.	Input	Output
1	raining, and	it's
2	it's, and, hailing	raining
3	it's, raining, hailing, like	and
4	raining, and, like, crazy	hailing
5	and, hailing, crazy	like
6	hailing, like	crazy

Table 2.2: Data input for learning Word2vec word embedding using CBOW

No.	Input	Output
1	it's	raining
2	it's	and
3	raining	it's
...	...	...
17	crazy	hailing
18	crazy	like

Table 2.3: Data input for learning Word2vec word embedding using skip-gram

The recommended window size for learning the word embedding is 10 for skip-gram and 5 for CBOW (Google, 2016). $p(w_t \mid w_{t+j})$ and $p(w_{t+j} \mid w_t)$ are defined by the softmax function

$$p(w_O \mid w_I) = \frac{exp\left(v'^{\mathsf{T}}_{w_O} v_{w_I}\right)}{\sum_{w=1}^{W} exp\left(v'^{\mathsf{T}}_{w} v_{w_I}\right)}, \tag{2.22}$$

where v_w and v'_w are the vector representation of the input and output word w_I and w_O. W is the number of unique words in the corpus. However, due to huge computation effort, this can be optimized using the approximations of the softmax function, such as Hierarchical Softmax or Negative Sampling, to calculate a probability value (Mikolov et al., 2013b).

Word2vec has achieved superior performance compared with previous state-of-the-art approaches (Mikolov et al., 2013a) by only learning the local statistics of words in a corpus. GloVe (Pennington et al., 2014) is another word embedding model that calculates the global word-word co-occurrence matrix. It is a count-based model using global matrix factorization methods to generate low-dimensional word representations. It further combines the local context window methods, which are focusing on a corpus's local co-occurrence statistics. Another important extension of Word2vec is fastText (Bojanowski et al., 2017; Joulin et al., 2017). It represents each word as a bag of character n-grams, e.g., the word ⟨flood⟩ is represented as ⟨fl, flo, loo, ood, od⟩ for character 3-grams. The skip-gram model is trained to learn the vector representation of words considering the subword information, enabling the model to capture the meaning of shorter words, suffixes, and prefixes. This model is also able to provide embedding for words not included in the corpus.

2.3.3 TextCNN

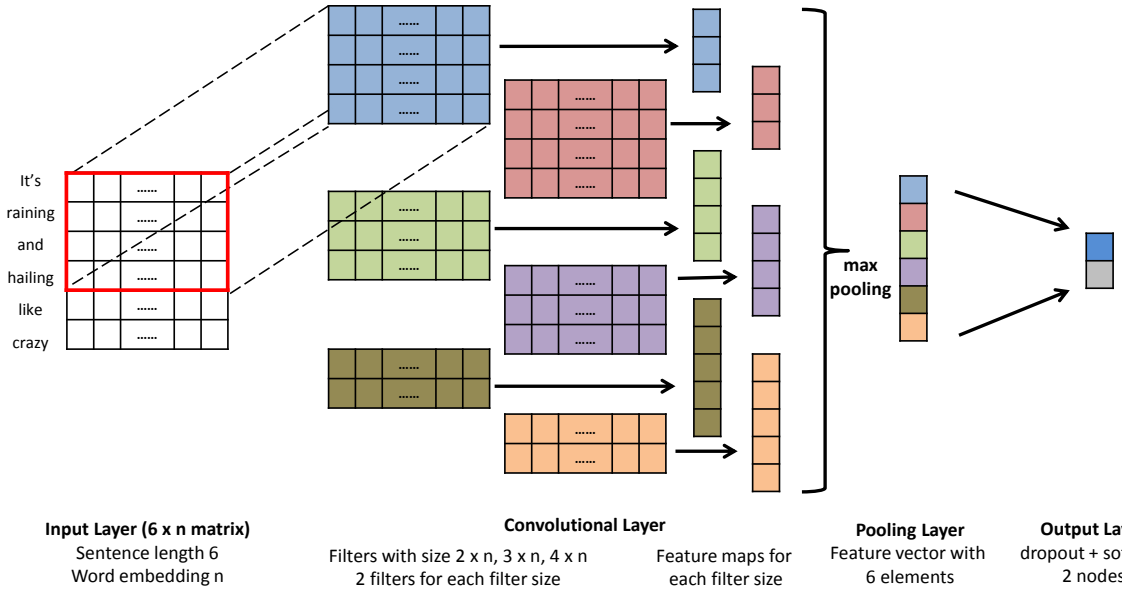


Figure 2.13: Illustration of TextCNN architecture used for text classification, adapted based on Figure 1 from (Zhang and Wallace, 2017).

Feature representations of sentences can be generated with word embedding pre-trained on large text corpora, such as Wikipedia. Early approaches, e.g., (Le and Mikolov, 2014), have tried to average the word vectors to provide sentence embedding. Kim (2014) proposed to use CNN to

encode word vectors into sentence embeddings. These embeddings are then used for classification using the softmax output layer. This method is also known as *TextCNN* in the subsequent studies.

The architecture of TextCNN is illustrated in Figure 2.13. For each word in a sentence, the corresponding word embeddings are searched to build a $k \times n$ matrix, where k is the sentence length, and n is the word embedding dimension. For example, a commonly used Word2vec word embedding is pre-trained on the Google News dataset with a dimension of 300 (Google, 2016). Convolutional filters of different sizes are applied to capture the information within the filter window. All these feature maps are max-pooled and concatenated to generate a sentence representation with the same size as the number of filters. Subsequently, the sentence representations are classified with the softmax output layer. As in the case presented in Figure 2.13, the model is a binary text classifier. In this thesis, TextCNN is used in Section 5.1.2 to train the text classifier to retrieve flood-relevant text documents.

2.4 Spatial and spatiotemporal analyses

Social media data are typically given as location points. In order to summarize them, spatial and spatiotemporal analyses are important tools to further leverage flood-related social media VGI. They are frequently applied on geotagged social media data to detect areas with a high density of VGI location points, and reveal the distribution of user activity over time and space.

There are mainly three branches of tools that can be used to analyze VGI points. The first branch is Kernel Density Estimation (KDE), which is often used to generate heatmaps from location points. The density of points can be presented as a raster image. However, such a visualization is mostly subjective. The hyperparameter required by KDE – bandwidth – severely affects the user's perception of the situation. In other words, even for the same VGI location points, different bandwidths can lead to different perceptions. Therefore, hot spot analysis, the second branch of tools, is more appropriate. This method can identify the statistically significant spatial clusters. Aggregation is often performed to gather the location points into spatial units. The hot spots are the regions that have a high number of location points in themselves and should be surrounded by regions that contain many location points. The statistical test shows whether they are clustered or dispersed, as well as their significance levels. However, this also brings new problems. A general challenge is the Modifiable Area Unit Problem (MAUP) (Ratcliffe, 2004), where the identified spatial patterns can vary with a changing spatial unit. The last branch of the tools is clustering, which is independent of the selection of spatial units. Since the number of clusters is normally unknown in advance, density-based clustering is often applied.

In this thesis, all three branches of methods have been used. The following subsections further explains these methods in detail.

2.4.1 Heatmaps and hot spot analysis

Both heatmap and hot spot analysis are spatial analysis tools in GIS to visualize a high density or cluster of spatial data. However, they are designed for different purposes. Heatmaps are often created using KDE. This is a non-parametric approach by applying kernels to estimate the probability density function

$$p(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{2.23}$$

where n is the number of data points, h is the smoothing parameter bandwidth, and K is a kernel function, e.g., uniform, triangular, Gaussian kernels, etc. By applying different kernels and bandwidths, the density of points can be visualized in different smoothness (as illustrated in Figure 2.14). This can be a bias that leads to different perceptions among end-users.

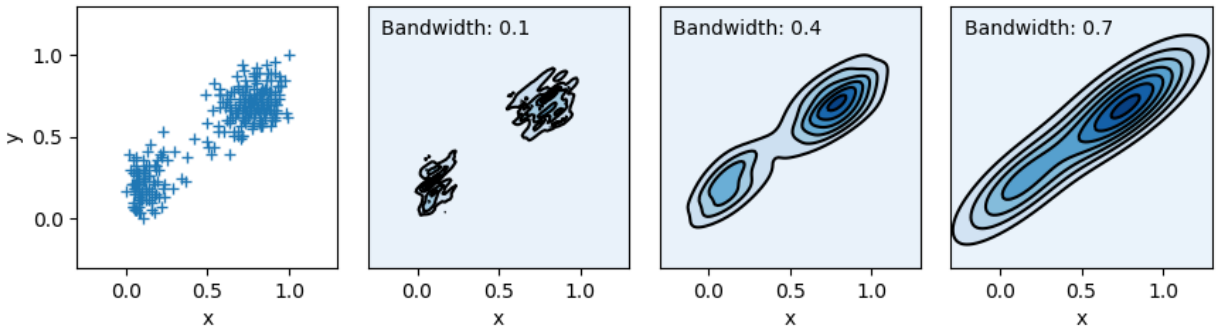


Figure 2.14: Example of KDE using different bandwidths.

Statistically significant spatial clusters of high values (hot spots) and low values (cold spots) can be identified by hot spot analysis. A single feature with a high value does not necessarily have to be a statistically significant hot spot. Statistically significant hot spots are features with high values, surrounded by other features also with high values (ESRI, 2019a). The social media VGI location points are often aggregated into spatial units by counting the points that fall into the unit, such as grids, hexagons, and polygons.

Getis-Ord Gi* (Ord and Getis, 1995) is one of the frequently used geostatistics methods for hot spot detection. This method also takes the local neighborhood into account. The principle of Getis-Ord Gi* is to compare local averages to global averages based on the z-score. The following equation is used to compute this z-score:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}, \tag{2.24}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j. n is the number of feature and

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}, \tag{2.25}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2}. \tag{2.26}$$

The statistical significance can be calculated using the resultant z-scores. A z-score of 1.65 represents a 90% confidence level, 1.96 for 95%, 2.58 for 99%, and 3.29 for 99.9% (Bohm and Zech, 2010).

In this thesis, the KDE method is used in Section 6.4.3.2 to generate a heatmap for property claims. Getis-Ord Gi* is used in Section 6.3.3 to detect hot spot regions where flood-related social media VGI data points accumulated.

2.4.2 Density-based clustering

Clustering is a task that aims at grouping similar objects. It is frequently applied in spatial analysis, where different distance measures can be used as the similarity measure, e.g., Euclidean distances. There are many ways to cluster spatial data points, such as partitioning clustering (e.g., k-means), hierarchical clustering, density-based clustering. As for social media data used in this thesis, the number of clusters is usually not known in advance. In addition, not every point needs to be clustered due to the fact that some of them are outliers. Therefore, density-based clustering methods are the natural choices, i.e., identifying regions of high point density while considering outliers.

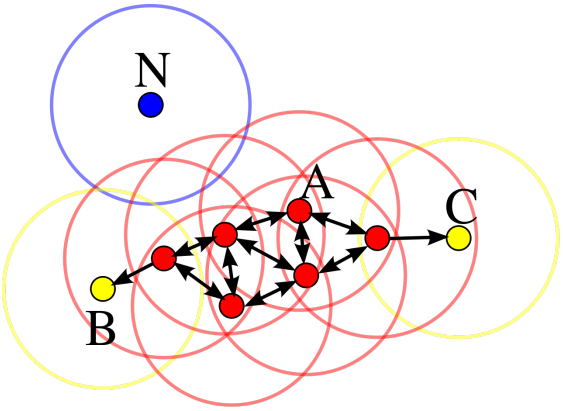


Figure 2.15: Example of DBSCAN (image under CC BY-SA 3.0).

Density-based spatial clustering of applications with noise (DBSCAN) (Ester et al., 1996) is the best known density-based clustering algorithm. It iterates all data points and marks the points with at least $MinPts$ points within a radius of ε as core points (as the red points shown in Figure 2.15). All the neighbors of the core points reachable within the radius of ε are grouped to form a cluster. These include non-core points that are regarded as border points (the yellow points B and C in Figure 2.15). This step iterates until no further clusters can be found. The points that are not reachable by any core point are considered as outliers (as the blue point N shown in Figure 2.15). Therefore, the minimum number of points $MinPts$ and the maximum spatial distance ε, are essential hyperparameters for this method. The choice of $MinPts$ is mostly task-oriented and needs to be determined based on domain knowledge. The radius ε can be selected by observing the k-distance graph, where the sharp change that occurs can be determined as the value of ε based on visual analysis.

In addition, such density-based clustering algorithms have been extended with a temporal dimension to perform spatiotemporal clustering. ST-DBSCAN (Birant and Kut, 2007) is such an extension of DBSCAN. In addition to the maximum spatial distance ε_{space} (denoted as ε in a standard DBSCAN), and the minimum number of points to form a cluster $MinPts$. The maximum time difference ε_{time} is added as an additional parameter for the temporal dimension.

In this thesis, DBSCAN is used in Section 6.4.3.1 to detect duplicated social media images that show same or very similar content in feature space. ST-DBSCAN is used in Section 6.3.3.1 to detect spatiotemporal clusters based on flood-related social media VGI data points.

2.5 Opportunistic VGI data

As mentioned in Section 1.1.2 there are two main forms of opportunistic VGI data considered in this thesis: one is social media and the other is mobility data. In this section, the characteristics and availability of both data types are introduced in Section 2.5.1. In addition, the data structure of Twitter data used in this thesis is detailed in Section 2.5.2.

2.5.1 Characteristics and data sources of opportunistic VGI

There are various sources of VGI data obtained in an opportunistic manner. Different platforms and service providers have different numbers of active users, and different access rights and developer policies. Different data sources of social media data and mobility data are compared in Table 2.4.

Social media data sources Currently, there are many social media platforms. Here, only five platforms with a large number of active users, which are often mentioned in VGI-related research, are compared, namely Facebook, Instagram, Weibo, Twitter, and Flickr. They all support users to post text, images, and videos. Most of them provide APIs (Application Programming Interface) for data acquisition and posting. However, not all of them can be used as data sources for flood monitoring, as described below.

The popularity of social media platforms is often measured by Monthly Active Users (MAU). Although Facebook has a large number of active users, most of the posts currently accessible are from the public pages of organizations. Posts from individual users are mostly invisible. Currently, most studies using social media data have been conducted based on Twitter or Weibo data. They provide access to postings from a large number of individual users. Real-time data can be collected using official APIs or web crawlers. Weibo is a social media platform mainly popular in China. Twitter is a data source available in most countries and regions around the world. Therefore, it is more suitable for the study of this thesis.

Instagram and Flickr are image-sharing services, where text cannot be shared without images and videos. Compared to Facebook, there are many more individual users whose postings are visible on Instagram. It does not provide real-time data streaming as Twitter does. Nevertheless, many Instagram users automatically share their posts on Twitter, and these posts appear as a shortened text with a URL link. With these links, user-generated texts can be complemented, and images and videos can be downloaded through the Instagram API. Although only some users do so, the large number of active users still provides a large amount of real-time posts from individual users. Flickr is mainly for image-sharing. In its early days, it was often used to host high-resolution photos for photographers. The images on Flickr are of relatively high quality, along with professional information such as camera parameters. However, its relatively small number of active users and relatively weak real-time nature make it difficult to be used for the purpose of flood monitoring in this thesis.

All these platforms allow users to post with geotags. Social media users can easily access their exact coordinates using smartphones. Since 2009, Twitter has been supporting users to share geographic locations (Stone, Biz, 2009). However, due to privacy concerns, fewer users are sharing precise locations, so that Twitter was shutting down this feature in 2019 (Porter, Jon, 2019). Most social media platforms now only allow users to select a location from a list of nearby locations. Flickr still allows users to geotag their photos, either by reading the EXIF of the images or by manually geotagging (Ding and Fan, 2019).

Table 2.4: Social media and mobility data sources for opportunistic VGI research.

Name	Data format	Location quality	Usage (MAU)	Access
Facebook	text/image/video	Places	2.74 billion[1]	Public pages via Facebook API
Instagram	text/image/video	Places	1.22 billion[1]	Public pages via Instagram API/ Shared posts on Twitter
Weibo	text/image/video	Places	511 million[1]	Weibo API using keyword/location search
Twitter	text/image/video	Places/ Coordinates	353 million[1]	Twitter Streaming API
Flickr	text/image/video	Places/ Coordinates	60 million[2]	Flickr API
Navigation service providers	Trajectories	Coordinates	-	Mostly commercial
Taxi data	Trips	Pickup & drop-off cell IDs	-	Monthly for New York City[3]
Taxi data	Trajectories	Coordinates	-	Datasets published by individual cities

MAU: Monthly Active Users
[1] Source: Most popular social networks worldwide as of January 2021. `https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/` (Accessed on 31.01.2021)
[2] Source: Work at Flickr. `https://www.flickr.com/jobs/` (Accessed on 31.01.2021)
[3] Source:TLC Trip Record Data. `https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page` (Accessed on 31.01.2021)

The use of social media data can also be restricted by platform policies. Flickr users can choose to share their work using different Creative Commons[9] licenses. Therefore, it is favored by VGI researchers because of the accessibility of the data. The other four social media platforms do not yet support the setting of copyright information. Twitter mentions in their privacy policy that "Twitter is public and Tweets are immediately viewable and searchable by anyone around the world."[10] According to the Developer Policy[11] of Twitter for non-commercial research, researchers are allowed to access Twitter data and redistribute data as Tweet IDs and User IDs. This also results in researchers not being able to publicly share text and images from Tweets as a dataset. Therefore, the example images used in the following parts of this thesis are mainly from Flickr, under the Creative Commons licenses. It is important to note that the quality of images on Flickr is relatively high. Images on Twitter and Instagram, on the other hand, are of varying quality,

[9] Creative Commons. `https://creativecommons.org/` (Accessed on 31.01.2021)
[10] Twitter Privacy Policy. `https://twitter.com/en/privacy` (Accessed on 31.01.2021)
[11] Twitter Developer Policy. `https://developer.twitter.com/en/developer-terms/policy#6._Be_a_Good_Partner_to_Twitter` (Accessed on 31.01.2021)

sometimes containing emojis, screenshots, GIF animations, etc. Twitter may also trans-code and compress user-uploaded photos[12].

Social media data quality Data quality is an inherent challenge when using social media, such as Twitter and Instagram. Several inherent limitations exist when using it for flood monitoring. Social media includes information mainly in three aspects, time, location, and content. Each aspect may introduce uncertainty to the applications using this data source.

Regarding the time information recorded in social media posts, such as for flood events, users can only post after seeing this event. There must be a time delay ranging from a few seconds to even a few days, and it varies from person to person. There is also little research quantifying this time delay.

With respect to location information, the locations reported by the user can be the location where the user observed an event, and not necessarily the location where the event occurred. In addition, social media users may send their posts with an inaccurate or even fake location. The investigation from (Cvetojevic et al., 2016) showed that the typical distances between the image content and photo upload location have a median value of 198.7m for Twitter in North America and the Caribbean (based on 154 posts) and 85m for Instagram posts (based on 251 posts from 16 countries worldwide).

With respect to the utility of the content, user reports may be personally biased and may even contain false or exaggerated content. Photos and videos are often edited by users, for example, by applying image filters to change color or brightness, overlaying texts on photos, or collaging several photos together. These features are challenging for the interpretation process because the content between text and images, or between images in a post, may be sometimes contradictory to each other.

Mobility data Navigation service providers, such as Google Maps, TomTom, HERE, INRIX, and Didi Chuxing, collect a large amount of trajectory data every day. Such data is also considered to be VGI collected in an opportunistic manner. Every time a user uses navigation and route planning services, the current locations are sent to the navigation service provider. Collecting this information also benefits their users in the way of, for example, providing real-time information on traffic conditions and road speeds. However, such real-time information is mostly commercially available and is one of the sources of profit for the navigation service providers. Other than navigation service providers, there are also taxi datasets released by the local transportation communities, e.g., the New York City (NYC) Taxis Trip Record Data, where the taxi trips are published on a monthly basis with the pick-up and drop-off timestamps and locations. The locations are in the form of cell IDs, that correspond to the taxi zones pre-defined by the city. In addition, there are also full trajectory records for a certain period of several days or even months available for individual cities, such as for Beijing[13], Shenzhen[14] and Chengdu[15] in China, Porto in Portugal[16].

[12]Twitter will now preserve JPEG quality for photo uploads on the web. `https://techcrunch.com/2019/12/11/twitter-will-now-preserve-jpeg-quality-for-photo-uploads-on-web/` (Accessed on 31.01.2021)

[13]T-drive. `https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/` (Accessed on 31.01.2021)

[14]Shenzhen Opendata (Chinese). `https://opendata.sz.gov.cn/` (Accessed on 31.01.2021)

[15]DataCastle Challenge (Chinese). `https://www.pkbigdata.com/common/zhzgbCmptDataDetails.html` (Accessed on 31.01.2021)

[16]ECML/PKDD 15. `https://www.kaggle.com/c/pkdd-15-predict-taxi-service-trajectory-i/overview` (Accessed on 31.01.2021)

2.5.2 Structures and characteristics of Twitter data

Real-time Twitter data posts can be obtained through the Twitter Streaming API in a JSON format. Each post contains several fields, including posting time, text content, text language, user information, whether it was a retweet or a reply to another Tweet, etc. Hashtags and URLs are also stored in a separate field *entities* when symbols # or *http* are used. Field *source* can be used to identify which application is used. They can be a web client, mobile apps, or shared posts from other social media platforms, such as Instagram and Facebook. Examples of Twitter posts are visualized in Figure 2.16.

Geotagged Tweets are the focus of this thesis. There are generally two types: one provides the exact coordinates recorded by the users' device, the other is a user-selected *place* which is represented as a bounding box. Figure 2.16 presents three examples. In the post on the left, the user provided the exact coordinates, and a city-level bounding box (i.e., the city Hannover in this example) was automatically assigned to the *place* field. For the rest two cases, the user selected a location from a list of nearby locations. The one presented in the middle selected a city district, where the corresponding bounding box is assigned to the *place* field. The one presented on the right selected a Point of Interest (POI), where the associated bounding box is a point. In both cases, the *coordinates* field is empty. The field *place_type* in *place* can be used to distinguish bounding boxes of different location level.

Currently, there are a large number of geotagged Tweets that are shared posts from Instagram, as many Instagram users choose to synchronize their posts on Twitter. These posts can be easily distinguished by the field *source*. Instagram only allows users to select one location from a list of nearby locations. This list is a mix of locations of different location types, which can be, e.g., a city name, a city district name, or a POI, as presented in Figure 2.17. However, they are all stored in the form of point coordinates when they are shared on Twitter.

To summarize, the characteristics of different kinds of geotags should be considered when aggregating these information.

```
{
 "created_at": "Mon Oct 02
               12:26:35 +
              0000 2017",
 "text": ...,
 "lang": "en",
 "retweeted": false,
 "user": {...}
 "source": "<a href=\"
    http://twitter.com
    /download/iphone\"
    rel=\"nofollow\">
    Twitter for iPhone
    </a>",

 "coordinates": {
  "coordinates":
     [9.71293439,
      52.38888322],
   "type": "Point"
 },

 "place": {
  "id":"48504653e183c91c",
  "url": ...,
  "place_type": "city",
  "name": "Hanover",
  "full_name": "Hanover,
               Germany",
  "country_code": "DE",
  "country": "Germany",
  "contained_within": [],
  "bounding_box": {
   "coordinates": [[
    [9.604388,52.305196],
    [9.918478,52.305196],
    [9.918478,52.454401],
    [9.604388,52.454401]
   ]],
   "type": "Polygon"
  },
  "attributes": {},
 },
 ...
}
```

```
{
 "created_at": "Wed Nov 25
               14:41:46 +
              0000 2015",
 "text": ...,
 "lang": "de",
 "retweeted": false,
 "user": {...}
 "source": "<a href=\"
    http://twitter.com
    /download/iphone\"
    rel=\"nofollow\">
    Twitter for iPhone
    </a>",

 "coordinates": null,



 "place": {
  "id":"2beeba19b04e7422",
  "url": ...,
  "place_type":
    "neighborhood",
  "name": "Nordstadt",
  "full_name": "Nordstadt,
                Hannover",
  "country_code": "DE",
  "country": "Germany",
  "contained_within": [],
  "bounding_box": {
   "coordinates": [[
    [9.69136,52.37706]
    [9.733297,52.37706],
    [9.733297,52.397283],
    [9.69136,52.397283]
    ]],
   "type": "Polygon"
  },
  "attributes": {}
 },
 ...
}
```

```
{
 "created_at": "Mon Oct 05
               13:02:29 +
              0000 2015",
 "text": ...,
 "lang": "en",
 "retweeted": false,
 "user": {...}
 "source": "<a href=\"
    http://twitter.com
    /download/iphone\"
    rel=\"nofollow\">
    Twitter for iPhone
    </a>",

 "coordinates": null,


 "place": {
  "id":"0952919bcd972002",
  "url": ...,
  "place_type": "poi",
  "name": "Universitaet
    Hannover Institut f.
    Stroemungsmech.",
  "full_name":
    "Universitaet
    Hannover Institut f.
    Stroemungsmech.",
  "country_code": "DE",
  "country": "Germany",
  "contained_within": [],
  "bounding_box": {
   "coordinates": [[
   [9.7129698,52.3886185],
   [9.7129698,52.3886185],
   [9.7129698,52.3886185],
   [9.7129698,52.3886185]
   ]],
   "type": "Polygon"
  },
  "attributes": {}
 },
 ...
}
```

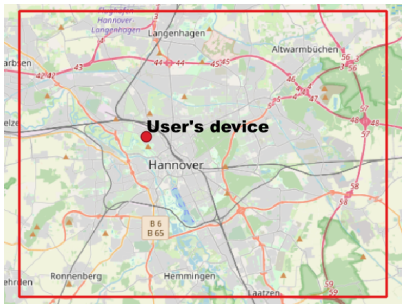







Figure 2.16: Examples of geotagged Tweets with exact coordinate (left), bounding box of a city district, and bounding box of a POI (right).

```
...
"full_text":"*** @
    Hanover, Germany",

"source": "<a href=\"
    http://instagram.com
    \"rel=\"nofollow\">
    Instagram</a>",

"coordinates": {
 "type": "Point",
 "coordinates":[9.7383,
        52.3722]
},

"place": {
 "id":"48504653e183c91c",
 "url": ...,
 "place_type": "city",
 "name": "Hanover",
 "full_name": "Hanover,
              Germany",
 "country_code": "DE",
 "country": "Germany",
 "contained_within": [],
 "bounding_box": {
  "type": "Polygon",
  "coordinates": [[
   [9.604388,52.305196],
   [9.918478,52.305196],
   [9.918478,52.454401],
   [9.604388,52.454401]
  ]]
 },
 "attributes": {}
},
...
```



```
...
"full_text": "*** @
    Hannover Nordstadt",

"source": "<a href=\"
    http://instagram.com
    \"rel=\"nofollow\">
    Instagram</a>",

"coordinates": {
 "type": "Point",
 "coordinates":[9.72078953
    , 52.39089835]
},

"place": {
 "id":"48504653e183c91c",
 "url": ...,
 "place_type": "city",
 "name": "Hanover",
 "full_name": "Hanover,
              Germany",
 "country_code": "DE",
 "country": "Germany",
 "contained_within": [],
 "bounding_box": {
  "type": "Polygon",
  "coordinates": [[
   [9.604388,52.305196],
   [9.918478,52.305196],
   [9.918478,52.454401],
   [9.604388,52.454401]
  ]]
 },
 "attributes": {}
},
...
```



```
...
"full_text": "*** @
   Geodaetisches Institut,
   Universitaet Hannover",

"source": "<a href=\"
    http://instagram.com
    \"rel=\"nofollow\">
    Instagram</a>",

"coordinates": {
 "type": "Point",
 "coordinates":[9.71230157
    ,52.38524343]
},

"place": {
 "id":"48504653e183c91c",
 "url": ...,
 "place_type": "city",
 "name": "Hanover",
 "full_name": "Hanover,
              Germany",
 "country_code": "DE",
 "country": "Germany",
 "contained_within": [],
 "bounding_box": {
  "type": "Polygon",
  "coordinates": [[
   [9.604388,52.305196],
   [9.918478,52.305196],
   [9.918478,52.454401],
   [9.604388,52.454401]
  ]]
 },
 "attributes": {}
},
...
```








Figure 2.17: Examples of geotagged Instagram posts shared on Twitter with different location levels: city-level location (left), city-district-level location (middle), and POI-level location (right).

3 Related work

With the popularity of the mobile Internet, data and information provided by individuals can be collected more easily and therefore receive much attention. Crowdsourcing is becoming a common way to acquire information. Many of the crowdsourcing applications focus on weather events and natural disasters. In this chapter, the latest research on obtaining precipitation and flood observations from volunteer citizens is outlined in Section 3.1 and Section 3.2 respectively. In addition, Section 3.3 reviews the techniques used for the automatic interpretation of natural disaster-related social media texts and images. Section 3.4 presents the relevant studies exploring the effects of precipitation on traffic speed and flow. Lastly, Section 3.5 summarizes the reviewed studies and identifies the research gaps.

3.1 User-provided precipitation observations

Rain gauges and weather radars are the most commonly used measurement devices for precipitation monitoring. Observations from both sources can be combined to achieve real-time precipitation monitoring with high spatial and temporal resolution. However, as described in Section 1.1.1, such observations are not well available for all regions in the world. There are still many parts of the world with limited meteorological monitoring networks, which implies vast areas in Africa, Asia and Latin America (Alfonso et al., 2015).

Therefore, crowdsourcing as an emerging information source is used to obtain such precipitation information from volunteer citizens with a participatory approach. Users were asked to measure the precipitation amount with a home-made gauge, e.g., using a simple rain gauge with funnel and ruler as in (Cifelli et al., 2005) or modified bottles as in (Illingworth et al., 2014). Then, the readings were sent back via email, social media, or web interfaces. The project mPING[1] provided a customized mobile application to support user reporting all kinds of precipitation events, including rainfall, hail, snow, and etc. Such participatory approaches of collecting VGI are often limited by the number of actively participating users.

Social media is one of the opportunistic information sources of VGI, which can be used to collect information from users. Since rainfall is a very common precipitation event, social media posts only appear in large numbers when it is particularly severe. Therefore, there are only a few studies aiming at extracting rainfall information from social media, e.g., (de Vasconcelos et al., 2016; Feng and Sester, 2017). More studies are focusing on snow, e.g., *UK Snow Map*[2] listens to the hashtag *#uksnow* in Twitter and visualizes the locations on an online map. Muller (2013) collected snow depth observations from social media users and generated a snow depth map through interpolation.

In addition to listening to social media to extract precipitation-relevant information, there are also studies and applications that rely on sensors to automatically obtain precipitation information. *WeatherUnderground*[3] is one of the participatory approaches to collect such observations. Participants need to buy a *Personal Weather Station (PWS)* and install it. The collected data are

[1] mPING. `https://mping.ou.edu/` (Accessed on 31.01.2021)
[2] UK Snow Map. `https://uksnowmap.com/` (Accessed on 31.01.2021)
[3] WeatherUnderground. `https://www.wunderground.com/` (Accessed on 31.01.2021)

transmitted to the server of *WeatherUnderground* to build a network of volunteered static sensors. There is also research on building dynamic sensor networks to obtain precipitation observations from information provided by road users. Manual selection of wiper speed is a strong indicator of precipitation intensity. The project *RainCars*[4] validated this idea first with computer experiments (Haberlandt and Sester, 2010) and then with laboratory experiments (Rabiei et al., 2013). In this way, taxis were used as moving rain gauges to provide high resolution precipitation data. However, this research is still in the proof-of-concept stage during project lifetime. Most cars were not yet able to easily share this wiper activity data in real-time. Additional data reading and transmission devices were required. The wiper information is available via the CAN bus of the cars, and future cars will be able to communicate this information. Both studies that automatically collected precipitation observations required the installation of additional sensors or devices, which limited the willingness of voluntary users to participate.

3.2 User-provided flood observations

The analysis and monitoring of natural disasters is one of the primary application fields for VGI (Yan et al., 2020). Among the various types of natural disasters, floods receive the most attention (Wang and Ye, 2018). In this section, research on the participatory and opportunistic acquisition of flood observations from voluntary users is outlined in Section 3.2.1 and in Section 3.2.2 respectively. In addition, possible scenarios for further use of this user-generated information are summarized.

3.2.1 Participatory approaches

The most straightforward way to collect disaster observations from voluntary users is through web interfaces and mobile applications. Mobile apps have been developed for crowdsourcing different disasters, such as *Did You Feel It?*[5] from USGS for earthquake (Atkinson and Wald, 2007), *iSeeFlood*[6] from University of Texas for flooding (Choe et al., 2017), and *FEMA Mobile App*[7] for disaster of all kinds. Additionally, *Ushahidi*[8] is a well-known platform on which different crowdsourcing projects can be deployed to collect reports from voluntary users. The data collected by *Ushahidi* has been used in the crisis mapping of the earthquake in Haiti in 2010 (Meier, 2012), the flood in Queensland in 2011 (McDougall and Temple-Watts, 2012), and the wild fire in Russia in 2010 (Asmolov, 2010).

In addition to setting up a system to collect user reports with location points, there are also collaborative mapping projects based on the well established VGI mapping platform, OpenStreetMap. U-flood project (Needham, 2017) is the realization of this idea, where users were asked to mark inundated roads on a map during Hurricane Harvey in 2017. The community updated 1,500 reports of voluntary observations and 991 roads were marked as inundated in Houston during the event (Chien, 2019). In addition, a similar approach has been used to analyze flood risk perception: Klonner et al. (2018) interviewed local residents in Chile and asked them to mark areas of flood risk using sketch maps based on their local knowledge. The study further confirmed the feasibility

[4]Smart wipers against floods. RainCars: The mobile measuring stations (German). `https://wissen.hannover.de/Forschen/Technik-Exzellenzcluster/Schlaue-Wischer-gegen-Hochwasser` (Accessed on 31.01.2021)
[5]Did You Feel It? `https://earthquake.usgs.gov/data/dyfi/` (Accessed on 31.01.2021)
[6]iSeeFlood. `https://www.iseeflood.org/` (Accessed on 31.01.2021)
[7]FEMA Mobile App. `https://www.fema.gov/about/news-multimedia/mobile-app-text-messages` (Accessed on 31.01.2021)
[8]Ushahidi. `https://www.ushahidi.com/` (Accessed on 31.01.2021)

of using sketch maps provided by volunteers to map flood events. The use of these applications primarily focused on emergency response purposes, or sharing information with local residents to improve situational awareness.

Moreover, the development of citizen science has also gained the attention of hydrologists. Various participatory crowdsourcing projects have been deployed to obtain more specific flood-relevant variables, such as water level, flow velocity, rather than mere flood reports with location. In order to obtain water level information from voluntary users, several platforms were developed. Users can provide water level readings of existing gauges via SMS messages, website, or mobile apps (e.g. Alfonso et al., 2010; Lowry and Fienen, 2013; Degrossi et al., 2014). A quality analysis shows that volunteers are able to provide accurate water level readings through training (de Brito Moreira et al., 2015).

More often, text descriptions and photos are collected from citizens via mobile apps or citizen science initiatives. Researchers manually analyze them to extract qualitative observations about the water level. For example, participatory approaches have been conducted to collect pictures about a flood event in Newcastle, UK, on the 28$^{\text{th}}$ of June 2012 from the citizens (Kutija et al., 2014). From the provided contributions, 12 images from 12 different places were manually annotated with water depth and used as validation for flood models. In the project *RiskScape* in Christchurch and Dunedin, New Zealand, people were asked to send photos of flood levels with time and place information after the flood peak. In Christchurch, 600 photos were received and assessed by professionals. However, the project in Dunedin was discontinued due to a lack of response (Le Coz et al., 2016).

Social media users rarely voluntarily share specific variables such as water level (Smith et al., 2017). To motivate users to contribute, in the project *PetaJakarta* (Ogie et al., 2019) invitation requests were sent to Twitter users who had mentioned flood-related keywords, to participate in a collaborative flood mapping initiative and provide water level information (See, 2019).

Nevertheless, motivating users to participate and provide information is difficult. As stated in the 90:9:1 rule observed by Nielsen (2006) for social media and online communities, only 1% of the users participate frequently and are responsible for most contributions while 90% only use or read. The remaining 9% contribute from time to time. Therefore, an opportunistic approach is desirable, where the information is acquired in a quasi-unconscious and passive manner, for instance, by exploiting information and data, which were provided for a different purpose.

3.2.2 Opportunistic approaches

Social media offers the possibility to collect thematic, spatio-temporal information in real-time. It is nowadays frequently used in emergency response. The emergency services such as 911 are often overloaded when a crisis happens, and people in the affected area often seek for help from social media (Cowan, 2017). In this case, the social media act as a platform, where critical information can be shared (e.g., Facebook Crisis Response, Iyengar, 2015).

Lots of applications were already built to detect or analyze various natural disaster events based on social media, such as earthquakes (Sakaki et al., 2010; Earle et al., 2011; Crooks et al., 2013), floods (Wang et al., 2016a; Schnebele and Cervone, 2013; Herfort et al., 2014), storms (Huang and Xiao, 2015; Yu et al., 2019), fires (De Longueville et al., 2009; Wang et al., 2016b). Among all the disastrous events, flood has attracted the most attention (Wang and Ye, 2018). Flood-related social media data can be analyzed standalone to detect spatio-temporal patterns or combined with other data sources to generate more detailed flood mapping results. In the following, these two aspects are presented in detail.

3.2.2.1 Spatial and temporal analysis

Since 2009, Twitter has started to support for geotagging (Stone, Biz, 2009). It also supports free access to the real-time data stream. Since then many studies have been conducted to understand natural disaster events by analyzing the spatiotemporal distribution of Twitter posts. Floods are one of the natural hazards, but the study of their spatiotemporal distribution is not much different from other disasters like earthquakes and wildfires. Therefore, the following review is not only limited to the analysis of flood-related social media. Similar studies for other disasters are also included. In the following, the research of social media VGI analysis of disasters is introduced from three aspects: temporal analysis, spatial analysis, and spatiotemporal analysis.

Temporal analysis is one of the most basic components and has been widely used in research on disaster-related social media posts. Time series of the number of social media posts over time are compared with significant events reported by local newspapers, for example, during a wildfire event (De Longueville et al., 2009), aiming to investigate the full course of the disaster event and the behavior of users on social media platforms. In addition, such temporal analysis can also be carried out over a longer time frame. Daly and Thom (2016) analyzed social media photos associated with fires over a seven-year period and linked peaks in the temporal analysis to 13 major fire events. With the time series, anomaly detection can be performed to detect events, e.g., using Seasonal Trend Decomposition (STL) as in (Chae et al., 2012).

Spatial analysis is mainly used to visualize the spatial distribution of social media posts containing location information. Earlier studies presented locations of disaster-relevant posts simply as point symbols on static maps or markers with pop-up windows on online maps.

Kernel Density Estimation (KDE) is often used to generate heatmaps from location points, where the concentration of location points can be represented as a raster image, e.g., for flood in (Cervone et al., 2016), for wildfire in (Wang et al., 2016b). In a KDE, the bandwidth or the radius of the kernel is a hyperparameter which is mostly chosen empirically. Since social media locations have a strong bias due to the uneven distribution of population and social media users, population data has been used to normalize the KDE results (Wang et al., 2016b). In addition, the location points can be aggregated to a variety of spatial levels, such as grids (MacEachren et al., 2011; Stefanidis et al., 2013), administrative polygons (Crooks et al., 2013), or Voronoi polygons (Cerutti et al., 2016; Wang et al., 2016a) by counting the number of posts. The distribution of these aggregated points can be visualized as choropleth maps that present the concentration of location points with respect to these spatial units. In general cases, population data from agencies (Cresci et al., 2015) or the averaged historical social media posts amounts (Feng and Sester, 2018) are used to normalize these aggregation results.

With the information aggregated to spatial units, hot spot analysis can be further performed, e.g., using Getis-Ord-Gi* (Ord and Getis, 1995). The details of this method are presented in Section 2.4.1. The output z scores and p scores represent the statistical significance of spatial clustering, based on the values in the spatial units. High or low values clusters (i.e., hot spots or cold spots) can be identified spatially. In terms of disaster-related VGI, Getis-Ord-Gi* has been applied to hotspot analysis of floods (Panteras and Cervone, 2018) and earthquakes (Resch et al., 2018). In addition, KIB (Kernel Interpolation with barriers) has been applied on the result of Getis-Ord-Gi* to provide a smoother visualization (Panteras and Cervone, 2018).

Even though such a mechanism of aggregating location points into areal units has been widely applied, this strategy may suffer from Modifiable Area Unit Problem (MAUP) as described in Section 2.4 (i.e., the identified spatial patterns can vary with a changing spatial unit). Clustering

Table 3.1: Standalone analysis of social media VGI for disasters.

Type	Information	Method and Visualization	Example Paper
temporal	**count of posts**	Number of posts as time series	De Longueville et al. (2009) Sakaki et al. (2010)
	anomaly/ event	STL on time series of LDA topics	Chae et al. (2012) Chae et al. (2014)
	sentiment score	Sentiment score as time series	Alam et al. (2020a)
spatial	**distribution**	Location points or markers on static or online maps	Alam et al. (2020a) Feng et al. (2020a)
	heatmap	KDE generated raster heatmap	Wang et al. (2016b) Cervone et al. (2016)
	heat regions	Aggregation counts to grids	Stefanidis et al. (2013)
		Aggregation counts to administrative polygons	Crooks et al. (2013) Cresci et al. (2015)
		Aggregation counts to Voronoi polygons	Cerutti et al. (2016) Wang et al. (2016a)
	hotspot	Hotspot detection on grids with Getis-Ord Gi*	Resch et al. (2018)
		Hotspot detection with Getis-Ord Gi* and interpolated with KIB	Panteras and Cervone (2018)
		Hotspot detection on points with text-interpreted urgency grades using Getis-Ord Gi* and Local Moran's I	Xing et al. (2019)
	clusters	DBSCAN	Daly and Thom (2016)
		OPTICS	Wang et al. (2016a)
spatio-temporal	**heatmaps over time**	Heatmaps generated separately by KDE for multiple time slots	Chae et al. (2014) Zhu et al. (2019)
		ST-KDE and presented in space-time cube	Kersten and Klan (2020)
	hot spots over time	Get-Ord Gi* applied on different dates	Kersten and Klan (2020)
	ST-clusters	OPTICS and visualized in time-space cube	Fuchs et al. (2013) Cerutti et al. (2016)
		ST-DBSCAN	Huang et al. (2018c) Kersten and Klan (2020)

is an approach that does not need predefined spatial units. Since the number of clusters is normally unknown in advance, density-based clustering is often applied, e.g., DBSCAN (Ester et al., 1996)

for wildfire events detection in (Daly and Thom, 2016), OPTICS (Ankerst et al., 1999) for urban flood event in (Wang et al., 2016a). Wang et al. (2016a) clustered the Weibo posts (similar to Twitter) in China during the flood event in Beijing in June, 2012.

Spatiotemporal (ST) analysis is often used to present the changes of disaster-reported locations over time. KDE has been applied to generate heatmaps at multiple temporal periods separately to discover the changes of spatial distribution patterns over time (Chae et al., 2014; Zhu et al., 2019). Instead of partitioning the time axis into intervals, ST-KDE takes time as an additional dimension for three-dimensional density estimation and has been visualized in the space-time cube (Kersten and Klan, 2020). In addition, ST-clusters are detected based on the VGI points with timestamps. OPTICS has been applied considering the temporal dimension, and the ST-clusters were visualized in the space-time cubes (Fuchs et al., 2013). The detected spatiotemporal clusters were manually validated with the evidence on the Internet, which confirmed the potential of social media data to be used as a distributed sensor for flooding. ST-DBSCAN has also been utilized in (Huang et al., 2018c; Kersten and Klan, 2020) to detect flood-related ST-clusters as events.

3.2.2.2 Integration with other information sources

Due to the sparseness and uncertainty of social media locations, standalone analyses of social media VGI can hardly provide information with full coverage and high-level details. Therefore, another branch of research focuses on the integration of social media VGI with other information sources for disaster mapping. The following three information sources are often combined with social media VGI for flood monitoring: Digital Terrain Models (DTMs), simulation results from hydraulic model, and remote sensing flood detection.

Digital Terrain Models (DTM) provide the basic relief information of an area. The terrain itself has bulges and depressions, which indicate where there is a high chance of flooding. A straightforward way is to estimate a flood surface with the water level information. There are a series of early studies utilizing social media for the analyses of Queensland floods in 2011 (McDougall, 2011a,b; McDougall and Temple-Watts, 2012). Texts, photos and videos from Flickr and Facebook were manually interpreted. In order to obtain precise water levels and exact locations for these user observations, field surveys were conducted with Real-time kinematic (RTK) GPS and conventional survey methods. In this way, 23 selected sites with photographs of the 2011 flooding in Brisbane, Australia, were verified. A flood surface was estimated with these measures and the flood extent was generated by subtracting the DTM by this flood surface (McDougall and Temple-Watts, 2012).

For the fluvial flood in 2013 in Dresden, Germany, Tweets were filtered by flood-related keywords. Experts or voluntary annotators were asked to estimate the relevance regarding inundation mapping and the water level from social media photos on a web-based platform. Five inundation depth estimates were used to estimate a flood surface with DTM via bilinear spline interpolation (Fohringer et al., 2015). Instead of estimating one global flood surface, Li et al. (2018) estimated a simple flood plane based on each water level estimate from either social media or river gauges. However, each estimate has an impact only on the area around it and decreases with distance (i.e., Inverse Distance Weighting - IDW). The flood probability of all estimates is summed and then normalized to the 0-100 range. The results show a high agreement to the USGS flood mapping results. However, methods that use only terrain information ignore the hydrological and hydraulic aspects of flood events.

Hydrodynamic models provide the estimation of flood-prone and inundation areas based on a DTM. Flood-related social media posts can be used as evidence to evaluate the flood modelling

Table 3.2: Analysis of social media VGI in combination with other sources of information.

Source	Method and Purpose	Example Paper
DTM	Estimation of a flood surface using interpolation of water levels from post-event survey on VGI locations	McDougall (2011a) McDougall (2011b) McDougall and Temple-Watts (2012)
	Estimation of a flood surface using interpolation of water levels from social media images	Fohringer et al. (2015)
	Estimation and integration of local flood surfaces based on water level estimates from gauges and VGI	Li et al. (2018)
simulation results	Validate flood hydraulics simulation	Aulov et al. (2014) Eilander et al. (2016) Smith et al. (2017)
remote sensing	Kernel smoothing generate VGI layer and merged with RS and others using a weighted sum overlay	Schnebele and Cervone (2013) Schnebele et al. (2014) Cervone et al. (2016)
	Integrate Tweets with NDWI flood detection by applying Gaussian kernel	Huang et al. (2018a)
	Integrate gauge data and Tweets with NDWI flood detection by applying kernel-smoothing and local morphological dilation	Huang et al. (2018b)
	Integrate Tweets with RS flood detection using maximum entropy and the least effort principle	Wang et al. (2018)

results (e.g. Aulov et al., 2014; Kutija et al., 2014; Smith et al., 2017). Aulov et al. (2014) validated the surge model forecasts from NOAA with social media data. Smith et al. (2017) applied hydrodynamic modelling for the 2012 flood events in Newcastle upon Tyne, UK with a 2D hydraulic model. Modelling results were compared with the locations of flood relevant social media posts. Eilander et al. (2016) applied flood mapping based on 888 water level mentions from social media texts during three days in 2015 in Jakarta, Indonesia, which is the most user active city on Twitter. Combined with DTM and hydraulic models, flood extent and water depth maps were generated.

Remote sensing is another important information source for flood monitoring, allowing information on the extent of flooding to be obtained for disaster management and emergency response. Flood extent can be extracted from remote sensing imagery to generate a flood probability map, e.g., using NDWI - Normalized Difference Water Index (e.g., Huang et al., 2018b), Modified NDWI (e.g., Rosser et al., 2017) or machine learning models (e.g., Sarker et al., 2019). However, for densely built-up urban areas, the performance of flood detection from remote sensing products is often compromised. In addition, occlusion due to observation angles and shadows of buildings and trees may also lead to misses of flood detections. In contrast, VGI data appear more frequently

in cities, as more users live there. Thus, flood-related social media posts with geolocation can be used as an ideal local complement to remote sensing flood detection.

VGI locations have been used to generate flood probability maps by applying kernel smoothing, e.g., with a quadratic kernel in (Schnebele et al., 2014), Gaussian kernel in (Cervone et al., 2016). They are merged with remote sensing detection or other data sources (e.g., flood hazard map based on DTM and river gauge data) with a weighted sum overlay. Due to their uncertainty, social media data were only given low weight (Schnebele et al., 2014; Cervone et al., 2016). Still it has a considerable effect, even if only using a small amount of VGI data, as demonstrated in (Schnebele and Cervone, 2013).

Based on a Digital Terrain Model, Huang et al. (2018a) queried the height of each flood-related VGI location and marked areas below that height as having a higher probability of flooding. In order to limit the impact range of individual VGI location, this probability decreases with increasing distance, which is similar to IDW (Inverse distance weighting). VGI points have been assigned with weights based on the NDWI wetness values around each point. By applying a weighted sum, the flood probability map was generated. In another research for the same event, Huang et al. (2018b) created a basic flood probability map on DTM and gauge observations, which was integrated with a flood probability map generated using quadratic kernel smoothing on NDWI. A local morphological dilation was applied to increase the flood probability for the area with VGI data points. This study showed that even though flood relevant information takes up only a very small proportion of the social media data streams, the geotagged flood relevant posts can still contribute to flood monitoring and extent mapping. In further, Wang et al. (2018) introduced a theoretical and algorithmic framework for heterogeneous data fusion of remote sensing data and social media data based on the maximum entropy and the least effort principle.

3.3 Interpretation of flood observations from social media texts and images

Flow velocity, flood extent, and water level are three flood-related pieces of information that can be extracted from social media and used for flood monitoring, mapping, and modeling purposes (Assumpção et al., 2018). Le Coz et al. (2016) and Le Boursicaud et al. (2016) estimated the water surface velocity from YouTube videos with Large Scale Particle Image Velocimetry (LSPIV, Fujita et al., 1998). Ground control points are needed as input for the LSPIV software, and they provide the scale of the video frames to estimate the flow velocity. The site needs to be surveyed after the events. Therefore these videos are difficult to be considered as opportunistic VGI for real-time flood monitoring. This thesis focuses mainly on the extraction of the other two flood-related information: flood extent and water level.

Flood-related social media posts are of great interest for estimating the extent of floods as presented in Section 3.2.2. In this section, methods of interpreting texts and images concerning floods are outlined in Section 3.3.1 and Section 3.3.2 separately. Not every flood-related post contains information about flood levels. Therefore, only some of the posts with clear indication of water level can be used for flood level mapping. Studies on water level estimation from social media texts and images are summarized in Section 3.3.3.

3.3.1 Text analysis for flood events

Text is the primary information source for the extraction of flood-relevant information. Keyword filtering is the most straightforward solution that has been used in many early studies to extract disaster-relevant social media posts. Predefined keywords are filtered to retrieve social media

posts related to flood (Fuchs et al., 2013; Murthy and Longwell, 2013; Fohringer et al., 2015; Li et al., 2018). Flood-relevant keywords are filtered based on the social media texts. For instance, "hoch" and "wasser" were filtered for flood mapping in Germany, 2013 (Fuchs et al., 2013). For multilingual regions keywords in multiple languages are often defined, e.g., German and English keywords are used to filter Tweets in Germany (Fohringer et al., 2015). For applications with a global focus, keywords related to nine types of disaster (including *earthquake*, *blizzard*, *tornado*, *drought/heatwave*, *cyclonic storm*, *hail/thunder*, *flood*, *tsunami*, *volcanic eruption*) have been filtered in 43 languages in (Dittrich and Lucas, 2014). However, keywords used for filtering are often ambiguous in meaning. For example, the keyword *flood* also has other possible meanings under different contexts, such as in *flood light*, the term *flooded by people*, thus leading to a limited performance in information retrieval.

In comparison, text classification in Natural Language Processing (NLP) provides better solutions for extracting disaster-related information. Social media documents are classified into binary or multiple categories with supervised classification based on the manually annotated corpus datasets, often prepared in advance.

Different types of features are summarized from social media posts. Statistical features, such as word n-grams (e.g., unigrams, bigrams), text length, the number of hashtags, user mentions, URLs, whether it is a retweet, whether it is a reply to another Tweet, POS (Part-of-Speech) tags[9], etc. are commonly used (Sakaki et al., 2010; Yin et al., 2012; Imran et al., 2013; Karimi et al., 2013; Cresci et al., 2015). *tf-idf* (term frequency–inverse document frequency) (Salton and Buckley, 1988) is a special case of word n-grams, where the word frequency is normalized by word document frequency (detailed in Section 2.3.1). It represents the importance of the word to a document based on the whole corpus, which has been frequently used for text classification tasks, e.g., in (Xiao et al., 2018; Khare et al., 2018). The statistical features can be used to train a binary or multi-class classifier using supervised machine learning (ML) methods, such as Support Vector Machine (SVM), random forest, logistic regression, naive Bayes.

Word embedding is the technique that aims to represent words or phrases as vectors of real numbers. This strategy has gained more attention after the rise of artificial neural networks. With a shallow neural network (presented in Section 2.3.2), vector representations of words can be learned based on a large corpus in an unsupervised manner. Word2vec (Mikolov et al., 2013a,b), the most well-known model, was widely applied for the text classification of flood-related social media texts. Since the vector representations are learned for individual words, sentences can be classified in multiple ways. Sentences can be represented by the averaged word vectors (Tkachenko et al., 2017; Bischke et al., 2017a) and classified using classic machine learning methods. Stowe et al. (2016) demonstrated that such features outperformed many combinations of statistical features and showed the highest importance in the ablation study for both the binary and multi-class text classification task. There are also further developed deep learning solutions to summarize sentence representations for a supervised text classification, such as TextCNN (Kim, 2014) in (Huang et al., 2019), and LSTM/Bi-LSTM (Liu et al., 2016; Zhou et al., 2016) in (Lopez-Fuentes et al., 2017; Sit et al., 2019).

Meanwhile, word embedding with better performance has been developed. GloVe (Pennington et al., 2014), an improvement from word2vec, captured both global statistics and local statistics of a corpus. In addition, fastText (Bojanowski et al., 2017; Joulin et al., 2017) represents each word as a bag of character n-grams (for more details see Section 2.3.2). The word representation is the sum of the character n-grams (Bojanowski et al., 2017). A model similar to word2vec is

[9]POS (Part-of-Speech) tagging is a classic task in Natural Language Processing, which marks tokens in a sentence with their corresponding part-of-speech categories, such as noun, verb, etc.

Table 3.3: Extraction of disaster-related social media posts based on texts.

Input	Category	Method*	Example Paper
text	binary	keyword filtering	Fuchs et al. (2013) Murthy and Longwell (2013) Fohringer et al. (2015) Li et al. (2018)
		tf-idf features + classic ML	Hanif et al. (2017) Xiao et al. (2018)
		word2vec + classic ML	Tkachenko et al. (2017) Bischke et al. (2017a)
		word2vec/fastText + TextCNN	Feng et al. (2018) Huang et al. (2019)
		word2vec/GloVe + LSTM/Bi-LSTM	Lopez-Fuentes et al. (2017) Sit et al. (2019)
		tf/tf-idf/word2vec + classic ML	Moumtzidou et al. (2018)

* Methods are represented in the form of *features + classification methods*, e.g., statistical features + classic ML. *classic ML* includes common machine learning methods such as Naive Bayes, logistic regression, SVM, random forest, etc.

trained to learn the vector representation of a word considering the subword information. These two improved word embedding models have also been used in disaster-related text classification, e.g., (Feng et al., 2018; Huang et al., 2019; Lopez-Fuentes et al., 2017; Sit et al., 2019).

Table 3.3 summarizes the aforementioned studies based on the applied methods. It can be observed that studies from 2017 mainly adopted word embedding techniques instead of statistical features.

3.3.2 Image analysis for flood event characterization

Images are also a key component of social media data, that are frequently used for the extraction of flood-relevant post. Many early studies rely on human interpretation to extract visual observations of flood events (e.g., in Kutija et al., 2014; Fohringer et al., 2015; Le Coz et al., 2016).

Research on automatic extraction of flood-related posts has emerged in recent years, all thanks to the rapid development of computer vision, and in particular the success of DCNN. Before using DCNN, engineered visual features, such as SIFT, SURF, and their derivatives, have been used for detecting flood (Jing et al., 2016a,b) from social media images in the years around 2016. *Multimedia Satellite (MMSat) Task* in the *MediaEval'17* benchmarking initiative (Bischke et al., 2017b) is a well-known task in the community to retrieve flood-relevant Flickr posts based on visual and textual features. The organizers prepared engineered visual features, such as Auto color correlogram (Huang et al., 1997), Color and Edge Directivity Descriptors (Chatzichristofis and Boutalis, 2008), etc., which have been used to retrieve flood relevant images in (Tkachenko et al., 2017; Zhao and Larson, 2017; Hanif et al., 2017).

The use of DCNN for disaster-related image classification started around 2016. The image classification algorithms using deep learning are rarely trained from scratch. Instead, transfer learning techniques are commonly adopted to fine-tune models trained on much larger datasets, such as

ImageNet (Deng et al., 2009), Places365 (Zhou et al., 2017a). Lagerstrom et al. (2016) fine-tuned an early DCNN model named OverFeat (Sermanet et al., 2013), which is pre-trained on the ImageNet dataset. More deep learning models for image classification have appeared since then, such as VGG (Simonyan and Zisserman, 2014), InceptionV3 (Szegedy et al., 2016), ResNet (He et al., 2016), etc. Many of them have also been used for image classification tasks related to natural disasters.

Table 3.4: Extraction of disaster-related social media posts based on images.

Input	Category	Method*	Example Paper
image	binary	SIFT-like features + classic ML	Jing et al. (2016a) Jing et al. (2016b)
		multiple engineered visual features + classic ML	Hanif et al. (2017) Tkachenko et al. (2017) Zhao and Larson (2017)
		pre-trained DCNN + classic ML	Bischke et al. (2017a) Avgerinakis et al. (2017)
		fine-tune pre-trained DCNN	Lopez-Fuentes et al. (2017) Nogueira et al. (2017a) Huang et al. (2019)
		ensemble of SVMs trained on features from multiple pre-trained DCNNs	Ahmad et al. (2017b) Ahmad et al. (2017a) Ahmad et al. (2018)
		concatenation of features from multiple pre-trained DCNNs + softmax/SVM	Said et al. (2018) Feng et al. (2018)

* Methods are represented in the form of *features + classification methods*, e.g., statistical features + classic ML. *classic ML* includes common machine learning methods such as Naive Bayes, logistic regression, SVM, random forest, etc.

As for transfer learning, two strategies are commonly used. One replaces the output layer of a pre-trained network (e.g., 1000 categories for ImageNet) with a softmax layer corresponding to the desired categories, such as flood-relevant and irrelevant. This process is also considered as fine-tuning of pre-trained networks. This strategy has been used for the *MediaEval'17 MMSat* task in (Lopez-Fuentes et al., 2017; Nogueira et al., 2017a) and other later-on studies (Huang et al., 2019). Another is to view DCNN as a feature generator and further apply classic machine learning methods, such as SVM, to the extracted features. This strategy has been explored for the *MediaEval'17 MMSat* task in (Ahmad et al., 2017b; Avgerinakis et al., 2017). In addition, machine learning classifiers can be trained on the deep features from different pre-trained DCNNs (pre-trained on both ImageNet and Places dataset). The final output is the fusion of prediction scores of multiple classifiers, e.g., the ensemble of SVM classifiers for *MediaEval'17 MMSat* task in (Ahmad et al., 2017a,b). This fusion strategy was further investigated in (Ahmad et al., 2018), where better results were achieved by introducing scene-level information, i.e., a pre-trained model on the Places dataset. There are also experiments with an ensemble of multiple DCNNs at feature level as in (Said et al., 2018) and (Feng et al., 2018).

More compact models have been developed in recent years, where social media images can be categorized by the type of disaster (Alam et al., 2020b), covering the most frequent natural disasters such as earthquakes, fires, floods, hurricanes, landslides, etc. EfficientNet (Tan and Le, 2019) outperforms the current commonly used models (e.g., ResNet, DenseNet) by around 1-2% on weighted average F_1-scores. Besides classification, there are also studies focusing on image retrieval. Barz et al. (2018) proposed an approach, which can retrieve not only flood relevant images, but also images containing evidence for an inundation depth estimation.

Table 3.4 summarizes the aforementioned studies that target images as input. It can be observed that the application of DCNN disaster-related image classification started to emerge around 2017, while fine-tuning and ensemble of pre-trained DCNN models are dominant approaches till now.

3.3.3 Water level observations from social media posts

Flood level estimation is an emerging task that has received much attention in recent years. The in-time estimation of flood extent and depth improves situational awareness and is beneficial for hydrological studies. A few studies tried to derive water level from social media text via template matching. Combinations of numbers and length units (e.g., *m*, *cm*, *in*) are searched in the user-generated texts (Eilander et al., 2016; Li et al., 2018). Pre-defined keywords like "knee-deep" have also been used as water level indicators (Smith et al., 2017). Despite the success of the above efforts, social media users who mentioned flood depth in texts during flood events are rare (Smith et al., 2017).

Visual information from social media contributes more to the water level estimation. In many early studies, water levels were manually extracted from social media images that contained objects of known size submerged in water (Assumpção et al., 2018). The most commonly used indicators for such a manual analysis are standing people and wheels of cars in water (Kutija et al., 2014). This interpretation is relatively easy for humans, however, it is a nontrivial problem for computers. Even though modern deep learning technologies can successfully interpret the relevance of photos or texts to flood events, the extraction of more detailed severity information from images has been explored in only a few studies.

With the development of DCNNs, the efficiency and accuracy of image classification and object recognition has been greatly improved. Pereira et al. (2019) classified social media images into three water level categories (i.e., no flood, below 1 m, and above 1 m). Deep features, extracted from the entire image by DenseNet (Huang et al., 2017a) and EfficientNet (Tan and Le, 2019), were used. Other studies explored the application of object detection to assess water levels. Partially submerged objects in the water received more attention. Chaudhary et al. (2019, 2020) detected person, car, bus, bicycle, and house by a Mask R-CNN model (He et al., 2017) as water level indicators. With the local deep features around these detections, objects are classified into 11 water levels, which correspond to the water height intervals in real numbers, e.g., 0 cm, 1 cm, 10 cm, 21 cm, until 170cm.

With the improved performance of human keypoint detection, e.g., OpenPose[10], scholars started to investigate the possibility of using human keypoints to estimate water levels. Quan et al. (2020) made use of detected human pose and well-designed rules to compare the relation between body keypoints and person segments. Multiple empirical thresholds were applied on ratios between different body parts to represent such a water level situation. Two categories (i.e., above the knee and below the knee) were assigned for the images, containing people in the flood scenarios.

[10]OpenPose. `https://github.com/CMU-Perceptual-Computing-Lab/openpose` (Accessed on 31.01.2021)

Vehicles have also been used as water level indicators. Recently, Park et al. (2021) estimated the orientation of a car in flood scenarios by calculating the similarity between the car detected by Mask R-CNN and a 3D rendered car on a horizontal plane. The submerged area is estimated to predict water level in real numbers.

Table 3.5: Flood water level estimation from social media posts.

Input	Prediction	Method	Example Paper
text	real numbers	text template matching	Eilander et al. (2016) Smith et al. (2017) Li et al. (2018)
image	3 (above 1m/ below 1m/no)	fine-tune pre-trained DCNN	Pereira et al. (2019)
	11 water levels	adapted Mask R-CNN using local deep features	Chaudhary et al. (2019) Chaudhary et al. (2020)
	2 (above/below the knee)	apply pre-designed rules on Mask R-CNN and human keypoints outputs	Quan et al. (2020)
	real numbers	estimate car orientation by comparing detected car with 3D rendered model	Park et al. (2021)

3.4 Precipitation and traffic speed variation

Precipitation events can have an impact on traffic speeds and volumes. Many previous studies focus on finding a general model, e.g., using a non-linear regression model (Lam et al., 2013), on representing the influence of precipitation events on traffic flow and density. Statistical analysis of the effect of weather conditions on vehicle speed has also been conducted (Jägerbrand and Sjöbergh, 2016). The impact of weather conditions on macroscopic urban travel times was investigated regarding different intensities of rain, snow, and temperature levels (Tsapakis et al., 2013). The precipitation information was also used to improve the prediction of macro traffic flow with LSTM (long short-term memory) networks (Jia et al., 2017). A recent approach applied correlation analysis, principal component analysis, and LASSO (Yang and Qian, 2019) to predict travel time with additional weather information. In most of the cases above, the goal of the research is to extract the correlation between traffic speed and the amount of precipitation. Some others aim to achieve better predictions of traffic flow or travel times with additional precipitation data.

Conversely, there is only a little research about extracting precipitation information from moving vehicle data. A previous research has employed motorcars as moving rain gauges. Using windscreen wipers' activities as sensors, it is possible to estimate the precipitation amount and improve the spatial resolution of precipitation data (Haberlandt and Sester, 2010). In recent years, some data-driven approaches have also been proposed to learn the rainfall or weather conditions from observed data. Prasad et al. (2013) learned a tree-based rainfall indicator from weather records, such as humidity, pressure, temperature, etc. Sathiaraj et al. (2018) learned a random forest classifier regarding normal and abnormal weather condition from the hour of the day, temperature, precipitation, visibility, and wind speed. Here, the situation is considered abnormal if the normal traffic volume is exceeded by one standard deviation.

Vehicles' trajectories data is an opportunistic VGI that has also been used to investigate the impact of precipitation on the traffic. One previous research analyzed road speed data of nine stormy days in Shenzhen, China (Li, 2017), where the average speed of roads was estimated based on map-matched taxi trajectories. She et al. (2019) could estimate the flood inundation areas on a grid basis by comparing the difference in trajectory point density between a normal day and a rainy day. It was based on the assumption that a flood would completely block the roads.

3.5 Research gap

Based on the review of related work, the following research gaps were identified.

- **New precipitation indicator from opportunistic VGI**

  There are few opportunistic methods available to collect precipitation observations from citizens. Social media are used however there is mostly a focus on events such as snow and other extreme weather events like hail and storms. Applications such as monitoring wiper activities of vehicles requires users to install additional devices in their vehicles for data reading and transmission. Therefore, a data source is more desirable which can be easily and abundantly provided by users.

  When precipitation occurs, the behavior of car drivers is naturally affected, mainly by slowing down the speed of their vehicles. Therefore, using vehicle speed provided by road users as an indicator of precipitation is a novel data source worth exploring, which has not yet been considered in any previous studies.

- **Extraction of flood-relevant social media VGI**

  Since pluvial flood is one of the disasters that severely affects people and is normally directly caused by heavy rainfall events, a system is desirable to efficiently extract the voluntarily posted Tweets relevant to rainfall and flooding and detect such events. Prior to this research, such studies mainly rely on manual annotation, keyword filtering, or classic NLP models to extract flood-relevant posts from social media for analysis regarding floods.

  In this thesis, a framework for the extraction of flood-related social media VGI is proposed. Deep learning models for text and images are trained for an automatic extraction. VGI from social media can be effectively collected and applied to real-world flood event analysis.

- **Flood severity mapping from social media VGI**

  Water levels are information that can be extracted from social media images when objects of known size are identified submerged by floodwater. Previous studies either use global deep feature of the whole image or local deep features around detected targets, such as people, cars. However, such implicit features do not accurately capture the relationship between the target and the floodwater. Although many of these social media posts are associated with locations, the extracted information has not yet been further used for mapping purposes.

  In this thesis, a novel method for flood level estimation is proposed by combining the outputs from multiple computer vision techniques, including object detection, human keypoint detection, and semantic segmentation.

4 Precipitation indicator from road users' speed variation

Intense precipitation and flooding lead to traffic slowdowns or even suspensions. Vehicles' speed can be measured and captured by trajectory data or by traffic speed detectors. The research objective of this chapter is to investigate the feasibility of using road speed variation of multiple road segments as a precipitation indicator for a city. This research is also presented in (Feng et al., 2020b).

4.1 Motivation

Many events can influence the speed of traffic. Local events, such as concerts, football matches, or traffic accidents, normally have a limited influence range around the event's location. However, regional events, especially inclement weather conditions, such as rain, snow, mist, and haze, can lead to a significant reduction of traffic speed for much larger areas. Such traffic variation patterns may reoccur for similar events with similar severity, which makes the presence of such events predictable.

In order to learn a precipitation indicator from the variation of vehicle speed, sufficient observations are needed to cover a reasonable number of positive examples during precipitation events. As an opportunistic VGI, trajectories of vehicles are one of the major sources of information to provide traffic speed and flow observations. The abundantly collected trajectory data by navigation service providers or local taxi service communities meet this requirement. Publicly available datasets rarely cover a very long period. Real-time traffic speed observations are also available at the transportation departments of many cities. They use traffic speed detectors to collect this data, primarily covering the city's main roads.

In this work, a precipitation indicator is to be learned based on the traffic speed variation patterns of multiple road segments. To the best of our knowledge, this is the first attempt to learn the precipitation information from the road speed observations directly. A binary precipitation indicator is trained, which can detect precipitation events directly from traffic speed observations. This chapter is organized as follows. In Section 4.2, the data and method used for this research are introduced. Section 4.3 presents the results and evaluations of the proposed method. In the last section, there is a short summary of this work.

4.2 Methodology

In this section, the data used for this research is introduced, and the proposed method for training a precipitation indicator is explained.

4.2.1 Data

The traffic speed observations used in this chapter are available at New York OpenData[1], which is provided by the Traffic Management Center (TMC) of New York City Department of Transporta-

[1]Real-Time Traffic Speed Data - New York OpenData. `https://data.cityofnewyork.us/Transportation/Real-Time-Traffic-Speed-Data/qkm5-nuaq` (Accessed on 31.01.2021)

tion (NYCDOT). It covers mostly the major arterials and highways within the city limits, where NYCDOT has installed traffic speed detectors. A subset of this traffic speed data is used, from 8th of August 2017 to 25th of April 2018. In this dataset, 135 road segments were observed (shown in Figure 4.1 left). Each road is associated with one road id. Two of them were not used in this analysis because they did not contain enough observations for the time series analysis. Therefore, the data from 133 road segments are used in this study. The observations are generally in 15-minute intervals, and significant gaps can be observed for most road segments. For the same time range, precipitation intensity data and textual weather descriptions from Central Park Station in New York (shown in Figure 4.1 right) were retrieved via the Weather Underground API[2]. The textual weather descriptions included weather conditions, such as fair, cloudy, rain, shower, snow, etc. In this study, only the weather information from this single station was considered. This dataset has a lower and unevenly distributed sampling rate, ranging from 10 minutes (minority) to one hour (majority).

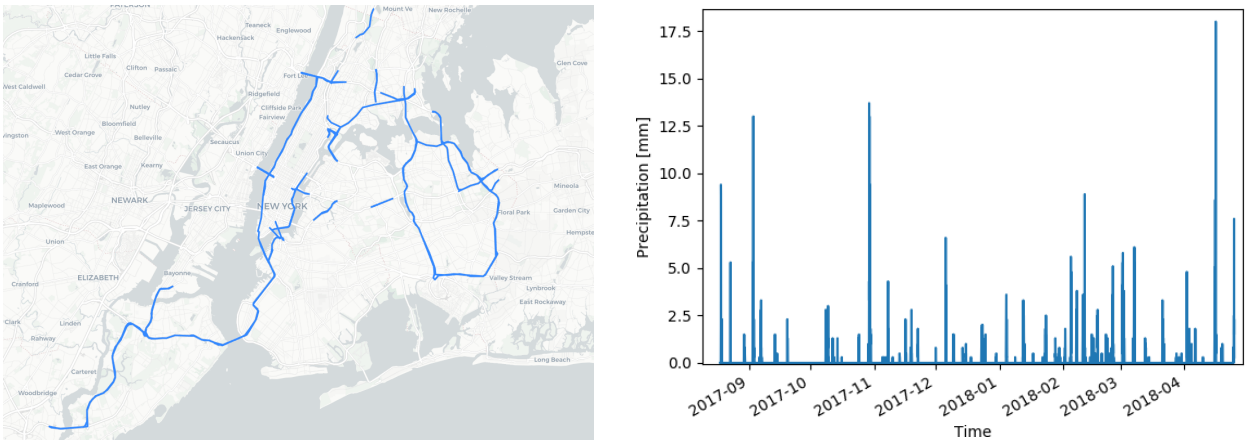


Figure 4.1: Spatial distribution of road segments in New York City (left, Basemap: OpenStreetMap) and precipitation data retrieved from Weather Underground (right).

4.2.2 Method

Since the traffic speed observations contain strong seasonal effects, the traffic speed data were modeled with *Prophet* – an open-source software from Facebook (Taylor and Letham, 2018). *Prophet* is a time series analysis tool based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality. It is regarded as a robust tool against missing values and outliers, which should make it well suited for this dataset containing gaps and noise. Observations of each road segment were modeled separately. An example is shown in Figure 4.2. The time series were decomposed into the trend, seasonal and residual signals. The extracted residuals represent the difference between the actual observations and the periodic model, which can indicate the anomaly level as compared to the normal traffic state. Since the given data is within one year, both the weekly and daily seasonalities were considered. Similar to the example in Figure 4.2, it was observed that the model fits nicely to the data in most cases.

The precipitation records have non-uniform sampling intervals (mostly 1 hour). However, the traffic speed data has a higher sampling rate in general. Therefore, pre-processing for both datasets was necessary. All of the speed residual records 15 minutes before and after each weather record timestamp were searched. The residual values of all road segments are averaged to represent the

[2]Weather Underground API. `https://www.wunderground.com/weather/api/` (Accessed on 31.01.2021)

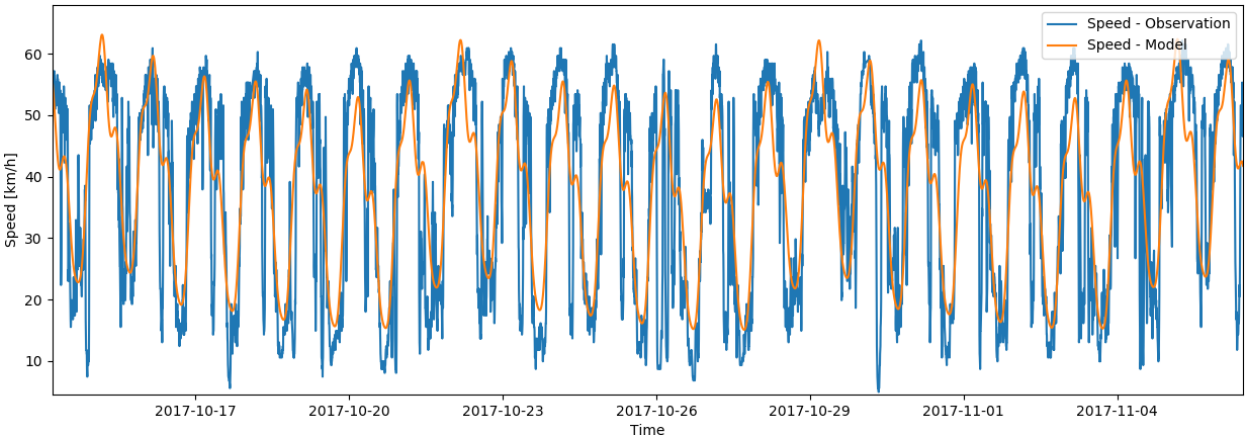


Figure 4.2: Example results achieved by Prophet, speed observations versus predictions, for one road segment within 24 days.

anomaly state compared to the normal state. If no record can be found within this range, NaN will be filled in. In this way, for each timestamp from the weather records, there are 133 values corresponding to the 133 road segments, representing the anomaly state of the whole area. This is exactly the input feature vector for training the models. Using several classical machine learning methods, binary classifiers were trained to generate binary predictions of precipitation events based on these features.

4.3 Experiment and results

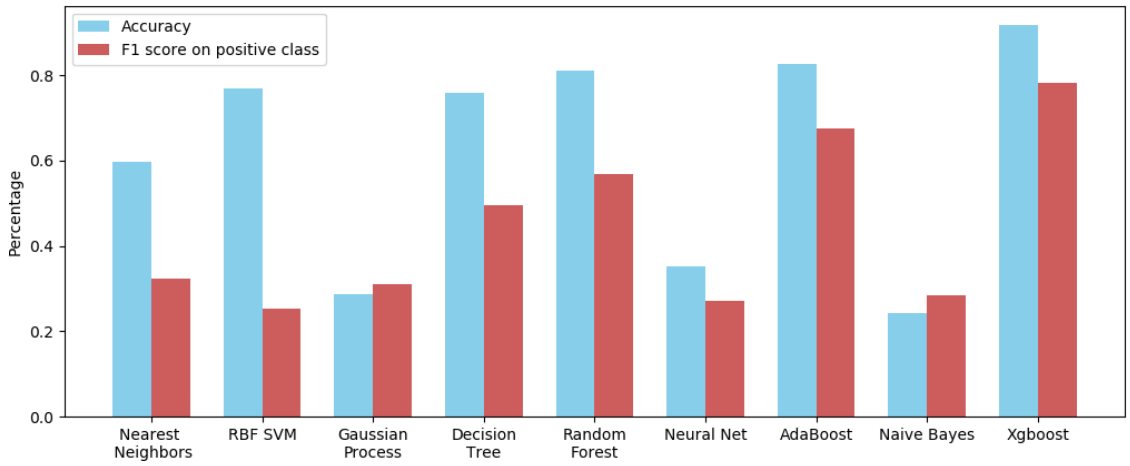


Figure 4.3: Comparison of some of the classical machine learning methods.

The training and test sets were separated at the date 28th of February 2018. The observations before that date were used for training, the remaining part for testing. The training set covers almost 6 months. Very light precipitation events normally do not have a significant impact on traffic speed. Therefore, only the weather records above 0.5 mm precipitation were selected as positive examples. From the training set, 408 weather records were extracted, and the same number of negative examples were randomly selected from the records with sunny or cloudy weather conditions based on the textual weather descriptions. In this way, a balanced dataset was built for training the models. Features were extracted based on the timestamp of each weather record. Several classical machine-learning methods were compared, such as Nearest Neighbors, SVM (Support Vector

Machine) with RBF (Radial Basis Function) kernel, Random Forest, AdaBoost, Xgboost, for this binary classification task. The implementations with default settings from scikit-learn (Pedregosa et al., 2011) were used to train the binary classifiers. The achieved overall accuracy and F_1-score on the positive class are shown in Figure 4.3.

From the results above, Xgboost outperformed the other classifiers and achieved an overall accuracy of 91.74% and F_1-score of 78.34% on the positive class on the two-month test set. More details regarding the evaluation of this model are shown in Table 4.1 and 4.2. From the results below, the model has achieved high precision and recall on the negative class, with slightly lower precision and recall on the positive class. The number of false positives and false negatives is small compared with the true positives and true negatives.

Table 4.1: Precision, recall and F_1-score on test set for Xgboost model.

		Precision	Recall	F_1-score
No Precipitation	0	0.93	0.97	0.95
Precipitation	1	0.85	0.73	0.78

Table 4.2: Confusion matrix on test set for Xgboost model.

		Prediction - 0	Prediction - 1
True Label	0	1312	45
	1	96	255

Besides precipitation, there may be other reasons which lead to a slowdown in traffic. Regional events, such as mist and haze, may also lead to similar traffic speed variation patterns. Therefore, the weather conditions of the false positive predictions were further analyzed. Out of the 45 false-positive predictions, 29 are associated with a description of light rain, light snow, haze, or mist, which all indicate adverse weather conditions, but still result in 0 mm precipitation records. For the true negative predictions, 81 out of 96 are less or equal to 1 mm precipitation, and all of them are less than 3.3 mm.

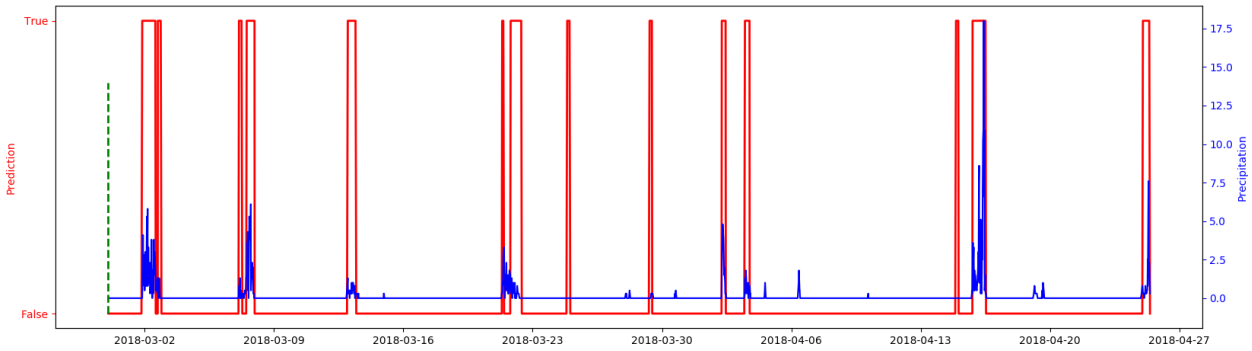


Figure 4.4: Predictions on 2-month traffic speed test dataset with Xgboost. The blue line indicates the precipitation amount in millimeter and red line indicates the prediction from the Xgboost model.

Figure 4.4 shows the binary predictions of the Xgboost model on the 2-month test data compared with the given precipitation amount. This set is totally independent of the training data. As can be seen, almost all of the precipitation events are covered by the positive predictions. The model

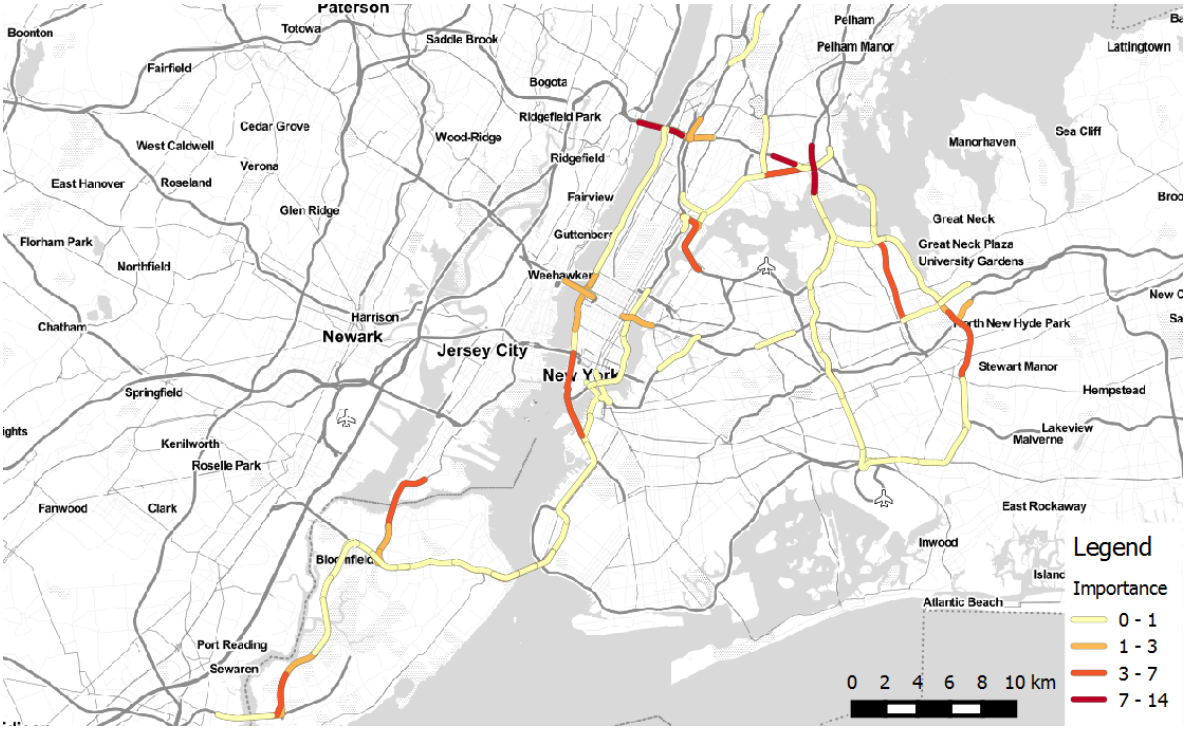


Figure 4.5: Importance of each road, as determined by Xgboost (Basemap: OpenStreetMap).

also ignored the events with very little precipitation. Only two false alarms were given, both of them being very short.

A second output of the Xgboost classifier is the feature importance. Since each feature corresponds exactly to one road in the dataset, the importance of each road for the overall classification result can be determined. For the Xgboost classifier, this is shown in Figure 4.5. The highlighted roads have higher importance than the others. Therefore, these roads can be considered to be more closely related to the precipitation events than the others because they play a more important role for the model to make a reasonable prediction. The reasons for their importance could be analyzed with respect to other information sources in further work, e.g., historical inundation records, or number of accidents during precipitation events.

Figure 4.6 and 4.7 show examples of two precipitation events on 15^{th} and 25^{th} of April 2018. The color indicates the speed observation, slower (blue) or faster (red) than the *Prophet* estimated model, for each individual road (identified in the figure by their individual road id). The black and green lines represent the start and end times of the precipitation event based on the textual weather description. On the right side, the corresponding precipitation amounts and binary prediction from the Xgboost model are compared. In both examples, a significant slowdown of the traffic can be observed visually when the precipitation event happens, which is in line with the expectation. Comparing the start and end time extracted from textual descriptions, the model makes only positive predictions when the precipitation amount increases. The time range of significant precipitation events was successfully identified in both cases.

4.4 Summary

In this chapter, a proof-of-concept study is presented, which can indicate precipitation events based on traffic speed variation patterns. Seasonal trend decomposition is used to eliminate the

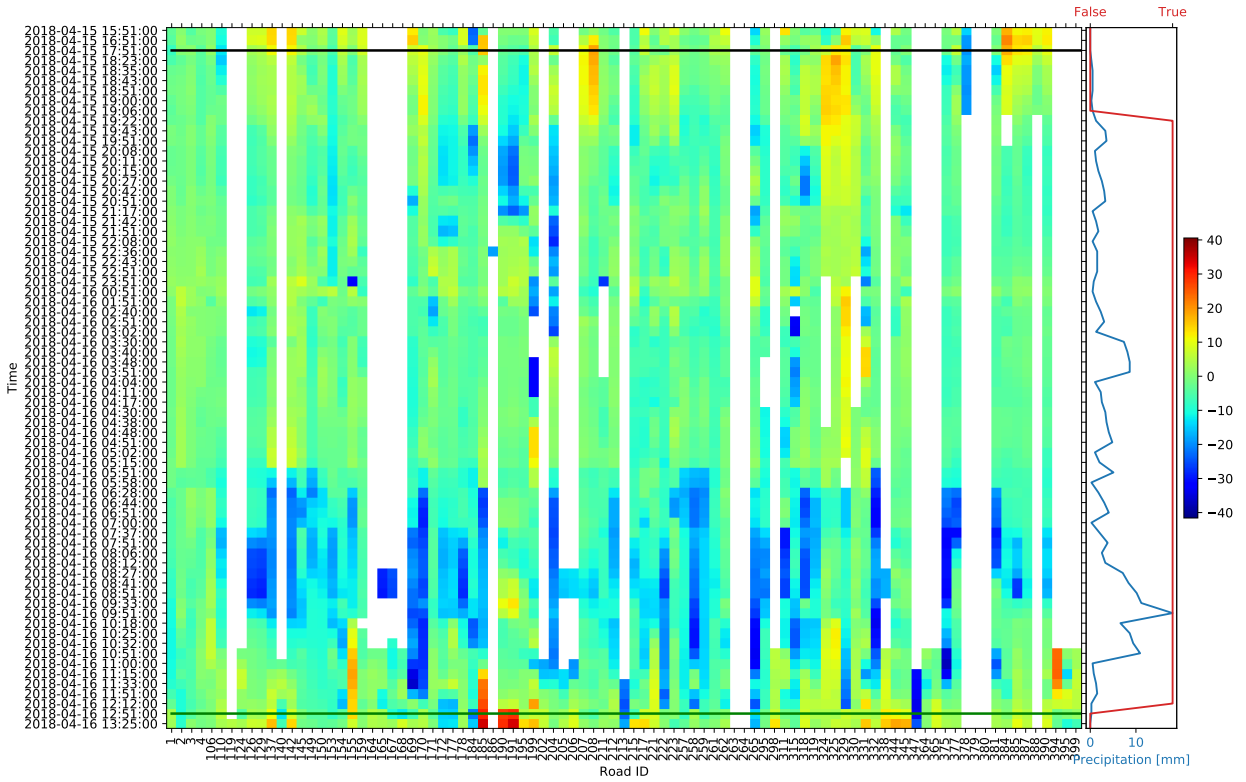


Figure 4.6: Speed variation pattern for a precipitation event on 15th of April 2018. The color indicates the speed observation, slower (blue) or faster (red) than the Prophet estimated model. The black and green lines represent the start and end times of the event based on the text description. On the right side, the blue line indicates the corresponding precipitation amount, and the red line indicates the prediction from the Xgboost model.

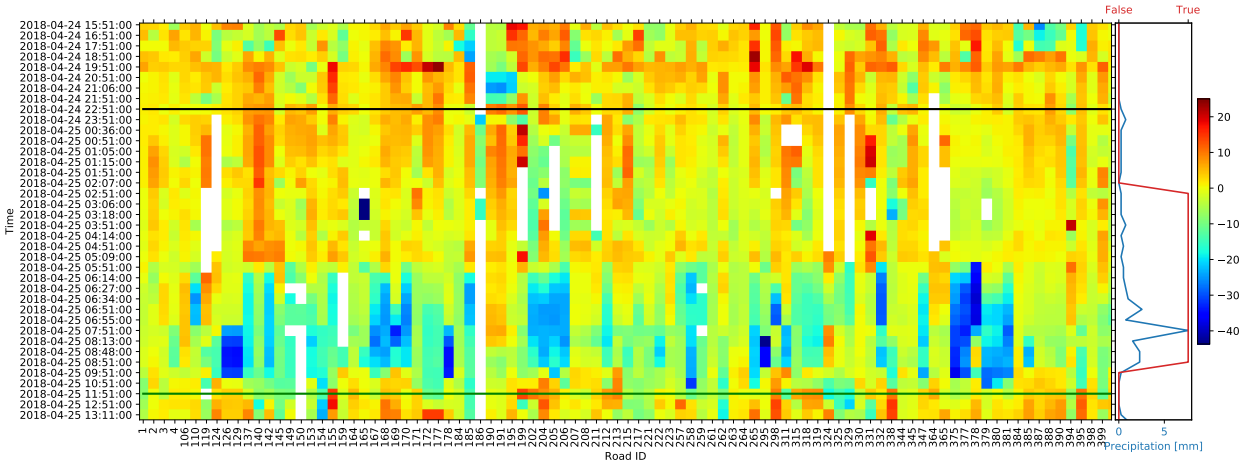


Figure 4.7: Speed variation pattern for a precipitation event on 25th of April 2018. The color indicates the speed observation, slower (blue) or faster (red) than the Prophet estimated model. The black and green lines represent the start and end times of the event based on the text description. On the right side, the blue line indicates the corresponding precipitation amount, and the red line indicates the prediction from the Xgboost model.

daily and weekly periodic effects of the traffic speed observations. Residuals of this model are used as features that indicate the anomaly level of the traffic as compared to the normal traffic state. Several machine learning methods are compared, and finally, Xgboost is chosen to train a classifier,

which makes predictions with respect to precipitation events based on these features. Since only a limited amount of positive examples are available within the 8-month observation range, this work was confined to a binary classifier. It has achieved a promising performance, an accuracy of 91.74% and an F_1-score of 78.34% for the positive class. This indicator is able to successfully identify most of the precipitation events during the two months of testing.

However, the purpose of this work is not to replace the weather stations but rather to prove that such an indicator could be learned from the road speed data. It is also not expected to achieve a performance similar to a set of distributed weather stations. Rather, this is expected to be a potential by-product, which could be extracted from huge amounts of currently available traffic data.

5 Methodology for the extraction of flood observations from social media VGI

Social media, one of the essential forms of opportunistic VGI, can provide valuable real-time information for flood monitoring. Prior to this research (Feng and Sester, 2018), most of the applications for flood monitoring have merely used keyword filtering or classical language processing methods to identify disaster relevant documents based on user generated texts. As the quality of social media information is often under criticism, the precision of information retrieval plays a significant role for further analyses. Thus, in this research, high quality eyewitness reports of rainfall and flood events are retrieved from social media by applying deep learning approaches on user-generated texts and photos. Apart from merely classifying posts as flood relevant or not, more detailed information, e.g., the flood severity, can be further extracted based on image interpretation. A novel method is proposed in this thesis to estimate water level for social media images containing people in floodwater.

The methods and processes proposed in this thesis for social media data interpretation and analysis mainly focus on three aspects. Section 5.1 describes the methods used to retrieve flood-related social media texts. Section 5.2 presents the approaches used to retrieve flood-relevant social media images. Section 5.3 introduces the method proposed to estimate water levels from social media images containing people in floodwater. The methods described in this chapter are also presented in (Feng and Sester, 2018) and (Feng et al., 2020a).

5.1 Interpretation of flood-related social media texts

Text is one of the most dominant forms of social media data. Flood-related posts can be extracted based on an understanding of the text content. In addition to keyword filtering, which was widely used in early studies, NLP-based text classification has been used in recent years. However, it requires a very time-consuming annotation process. In order to reduce human efforts, an automatic annotation process applying keyword filtering and querying weather data is proposed.

In this way, for the interpretation of social media texts, raw texts collected by crowdsourcing are firstly pre-processed, filtered with pre-defined keywords, and then automatically labeled using historical rainfall records. Five classical NLP methods and one deep learning method using word embedding are applied to train the text classifiers.

5.1.1 Pre-processing and training data preparation

Since social media posts contain lots of noise, a pre-processing step is needed. Besides the raw text as the most relevant information, also the fields creation time, coordinates, source, media, user's screen name, language and text (as detailed in Section 2.5.2) were used for the analyses. During the text pre-processing, the punctuation marks, numbers and URL were removed from the raw text. Emojis were not removed as some of them are also associated with the flood events.

In NLP, reducing stop-words and stemming are standard techniques as pre-processing steps. Stop-words are the most common words in a language, such as articles, pronouns or prepositions.

Stemming is the process, which reduces each word to the root form, such as from "flooding" to "flood". For different languages, different lists of stop-words and stemming algorithms have to be applied. However, not all languages are supported by both stop-word lists and stemming algorithms. The most frequently used languages within the study areas of this thesis (i.e., Europe West in Figure 6.1) are *English*, *French*, *German*, *Italian*, *Spanish*, *Portuguese* and *Dutch*. Therefore, a stop-word list[1], which supported all of the seven languages above, was used. Subsequently, different stemming algorithms from Natural Language Toolkit (NLTK) library[2] were applied on the sentences in different languages.

Many Twitter bots automatically send messages, such as weather reports, advertisements or weather forecast (examples are shown in Table 5.1). These messages are regarded as noise information. Most of them normally have similar contents or similar text structure after stemming and removing stop-words. These Tweets are often sent repeatedly, which is also a way to automatically detect them: if text messages of one user had similar contents or structures for more than three times, this user was added to a black list, the Tweets sent by these users were then filtered out from the input data stream.

Table 5.1: Examples of Tweets with similar structure of texts.

No.	Text
1	Wind 13.4 mph NW. Barometer 1023.6 hPa, Rising slowly. Temperature 10.2 °C. Rain today 0.0 mm. Humidity 99%
2	Wind 3 kts NW. Barometer 1025.5 hPa, Rising slowly. Temperature 8.8 °C. Rain today 0.0 mm. Humidity 81%
3	Wind 14.4mph NW. Barometer 1034.1hPa, Rising slowly. Temperature 9.3°C. Rain today 0.0mm. Forecast Settled fine
4	Wind 2.2 mph NW. Barometer 1032.5 mb, Rising slowly. Temperature 10.9 °C. Rain today 7.2 mm. Humidity 99%

With this approach, a 30-day collection of geotagged Tweets from Western Europe in June 2016 (as later detailed in Section 6.1.1), 3.6 million geotagged Tweets (from 473,004 users) could be reduced to 2.9 million (from 468,051 users). This means that these 4953 blocked users sent on average 149 Tweets during 30 days, and thus behaved obviously different from ordinary social media users. In the pre-processing steps, there was no normalization of texts and no grouping of synonyms.

Labeling training data is a typical problem for most of the supervised learning approaches, as large amounts of training data are required. In previous research using machine learning, Tweets were manually annotated (Sakaki et al., 2010; Karimi et al., 2013). For instance, the crowdsourcing service Amazon Mechanical Turk was used to employ annotators for labeling texts in (Karimi et al., 2013). At the end, they collected 5747 annotated Tweets for training their classifiers. Thus, the number of training datasets is limited by the annotation time and budget.

In order to automate the labeling process, the novel idea of this thesis was to link Tweets with known precipitation information. Automatic annotation based on a priori knowledge can generate a large amount of annotated data, but at the same time it may also introduce label noise. Some recent studies are focusing on training of deep neural network models with noisy labels. The result shows that, the performance of these models is not much affected when small parts of the

[1]Google Code Archive - stop-words. `https://code.google.com/archive/p/stop-words/` (Accessed on 31.01.2021)
[2]Natural Language Toolkit. `http://www.nltk.org/` (Accessed on 31.01.2021)

dataset were not precisely labeled (Patrini et al., 2017). As the goal of this thesis was to achieve an automatic labelling procedure, historical weather data are considered as a suitable indicator for identifying whether a Tweet is relevant to rainfall events, as pluvial floods are directly caused by heavy rainfalls and fast storms. An important data source is Weather Underground[3], a platform that offers a Weather API[4] to query historical weather records based on the date and location on city level. In order to automatically label Tweets into positive and negative examples, keyword filtering was used to search for potential candidates. Weather data were then inquired only for the Tweets containing pluvial flood-relevant keywords.

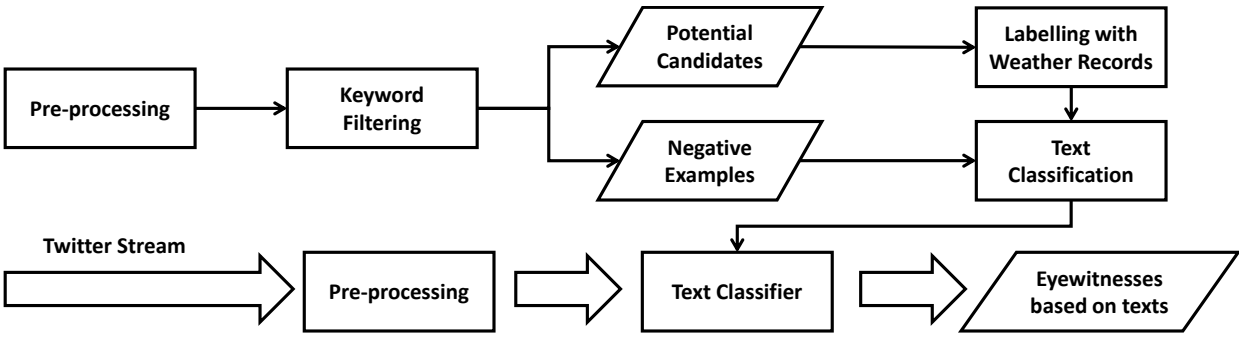


Figure 5.1: Workflow for training the text classifiers.

The whole procedure is shown in Figure 5.1. First, the collection of Tweets was pre-processed. For the following keyword filtering, a keyword list (as shown in Table 5.2) which contains the concepts such as "flood", "inundation", "rain" and "storm" in all the seven languages was used. All posts which contained the keywords were then looked up in the historical weather records via the Weather API. When the Weather API reported a rainfall, this Tweet was assigned with a positive label. If not, it was labeled as negative. By that, all the potential candidates for text classifier training were automatically labelled based on the weather records. Since there were many more positive examples than the negative after this step, random selected Tweets without any keywords were used as a supplement of the negative training examples to balance the training data.

Table 5.2: Keywords used for generating training dataset.

Language	**Keywords**
English	flood, inundation, deluge, rain, storm
French	inondation, inonder, crue, pluie, orage
German	hochwasser, flut, überschwem, überflut, regen, starkregen, regnen, sturm, unwetter, gewitter
Italian	inondazione, inondare, allagamento, pioggia, diluvio, borrasca, tempestad
Spanish	inundar, inundación, diluvio, aguacero, lluvia, tormenta
Portuguese	inundar, inundação, dilúvio, chuva, chover, tempestade
Dutch	overstroming, zondvloed, stortvloed, regen, storm

5.1.2 Training text classifiers

Following the preparation of a balanced training dataset, text classifiers were trained with five frequently used classical NLP methods, namely naive Bayes (McCallum et al., 1998), random

[3]Weather Underground. `https://www.wunderground.com/` (Accessed on 31.01.2021)
[4]Weather API: Introduction. `https://www.wunderground.com/weather/api/d/docs` (Accessed 07.11.2017)

forest (Breiman, 2001), SVM with linear kernel (Dumais et al., 1998), SVM with RBF kernel (Joachims, 1998) and logistic regression (Genkin et al., 2007). All these methods are trained based on tf-idf features. As additional method, deep learning using DCNN for sentence classification was used.

Classical NLP methods For the classical NLP methods, the text documents were first transformed into a sparse tf-idf (term frequency - inverse document frequency) (Manning et al., 2008) matrix, also called the 1-V matrix, where V is the number of unique words in the whole corpus. Term frequency is the raw count of each term in the sentence. Inverse document frequency indicates the rareness of the words. This value diminishes when the term occurs frequently. More details about tf-idf matrix is presented in Section 2.3.1. This matrix was calculated using the methods offered by scikit-learn library (Pedregosa et al., 2011). With the normal classification methods in machine learning, the classifiers could be trained based on this tf-idf matrix. Naive Bayes was firstly applied, which is the most basic method for text classification in NLP. It was used as a baseline to demonstrate the performance of the other methods. Random forest, logistic regression, SVM with linear kernel and SVM with RBF kernel are also methods frequently used NLP methods for text classification and the corresponding classifiers were trained separately.

DCNN for text classification As introduced in Section 2.3.2, word embedding is a technique that represents each word in the sentences by a word vector. The vector representation using word embedding need a much lower dimension than the one-hot vector representation in tf-idf. The skip-gram strategy of the Word2vec (Mikolov et al., 2013b) was used in this research. The word embeddings were generated based on 20 million unfiltered Tweets collected within the study area of western Europe from 1 July 2016 to 15 December 2016. The total size of the vocabulary is 934,063 and the average number of words of each Tweet is 6.01. In this case, the python implementation of word2vec in the Gensim[5] library was used to train this model, which has a default vector dimension set as 300 (i.e., the words are represented by vectors of 300 real values).

DCNNs were then applied on the word embedded sentences with the TextCNN structure adopted from Kim (2014) containing one convolutional layer, one max-pooling layer and one output layer (as detailed in Section 2.3.3 and illustrated in Figure 2.13). The output layer has two nodes, which are "pluvial flood relevant" and "irrelevant", respectively. After the convolution on the input matrix, feature maps are generated. Then max-pooling is applied on each feature map and a feature vector with the same size as the number of filters is generated. Subsequently, predictions are generated by the soft-max function. The implementation of this DCNN was based on the Tensorflow[6] framework.

5.2 Interpretation of flood-related social media images

Images are another essential component of social media data, which have been used only in recent years to extract flood-related observations. Since the number of flood-related images collected for training is limited, image classifiers are more suitable to be trained by fine-tuning a pre-trained model (i.e., transfer learning) rather than training a new model from scratch. In this section, two solutions using transfer learning are presented. The first model is trained to retrieve social media images related to pluvial floods in Section 5.2.1, where a single pre-trained DCNN model Inception-V3 is used for feature extraction. Fusing multiple models is a common strategy to optimize model

[5]Gensim (Version 0.13.4.1). `https://radimrehurek.com/gensim/models/word2vec.html` (Accessed on 31.01.2021)
[6]Tensorflow (Version 1.0.1). `https://www.tensorflow.org/` (Accessed on 31.01.2021)

performance. To explore better-performing models for flood image retrieval, Section 5.2.2 presents a feature fusion approach for image classification where four pre-trained models were considered. In addition, as duplicated images often appear in social media, a method to detect duplication is introduced in Section 5.2.3.

5.2.1 Training image classifiers using single pre-trained model

To interpret whether a user generated photo is relevant to rain and floods or not, a binary image classifier can be built. For training such a model, large amounts of training examples are required, which should contain images annotated as positive and negative. A common approach to cope with the problem of labelling large amounts of training examples is to use transfer learning (Goodfellow et al., 2016a). The pre-trained DCNN can serve as a feature generator by removing the output layer. The rest of the weights in the pre-trained model stay unchanged, and the output for each image is then a fixed-size feature vector. As described in the DECAF (Donahue et al., 2014) framework, features can be classified with the classical machine learning such as SVM or logistic regression. Logistic regression was applied since it is a method frequently used for binary classification. The ensemble methods, such as random forest (Breiman, 2001) and gradient boosted trees (Friedman, 2001), were also tested. For the three methods above, the implementation in scikit-learn library were used. Furthermore, the Xgboost (Chen and Guestrin, 2016) implementation of the gradient boosted tree was used. Multilayer perceptron with one hidden layer using back propagation was also tested and the implementation was based on the Tensorflow framework.

The pre-trained DCNN utilized for the extraction of pluvial flood-related social media images is the GoogLeNet Inception-V3 (Szegedy et al., 2015). It was trained based on the ImageNet 2012 Challenge dataset (Deng et al., 2009). This dataset contains 1.2 million images categorized into 1000 classes. This pre-trained model is available at the Tensorflow repository[7]. From the description of this model, it can achieve a top-5 error with 4.2% on the test dataset (Szegedy et al., 2015). After removing the output layer, the output for each image is a feature vector with 2048 values. The feature classification is subsequently conducted using a classical machine learning approach.

5.2.2 Training image classifiers by assembling multiple pre-trained models

Deep learning models with different architectures perform differently in image classification tasks. Therefore combining features from different models and models pre-trained based on different datasets has the potential to achieve a better feature representation. Therefore, an alternative solution for the extraction of flood related images is to use the features from multiple pre-trained models. This strategy is also used in Feng et al. (2018). The images were classified based on these features using either Xgboost (Chen and Guestrin, 2016) or fully-connected (FC) layers both with two softmax outputs (shown in Figure 5.2). The FC layers consist of two dense layers followed by batch normalization. Dropout of 50% was applied at the output layer. The softmax outputs on the positive class provide the confidence score of flood relevance. The final class prediction is based on a 0.5 threshold of this score.

Different pre-trained models are available, which were considered as the basic feature extractors: *InceptionV3* (Szegedy et al., 2016), *DenseNet201* (Huang et al., 2017b), *InceptionResNetV2* (Szegedy et al., 2017). They were all trained based on ImageNet and could achieve a

[7]Pre-Trained Inception-V3 model available on Tensorflow repository. `http://download.tensorflow.org/models/image/imagenet/inception-2015-12-05.tgz` (Accessed on 31.01.2021)

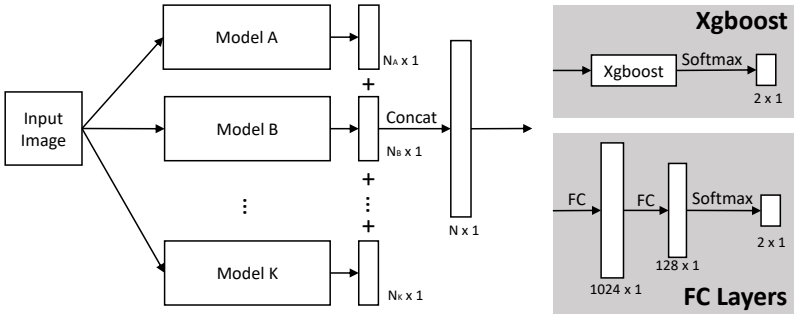


Figure 5.2: Feature fusion model for image classification into two classes: flood relevant and irrelevant.

top-5 classification accuracy of 0.936, 0.937 and 0.953, respectively. Since CNN models pre-trained on Places365 (Zhou et al., 2017a) were reported to have a better performance due to their scene-level features (Ahmad et al., 2019), a VGG16 network pre-trained on Places365 (Kalliatakis, 2017) was considered in addition. The experiments comparing different combinations are presented in Section 6.4.1.

5.2.3 Detection of duplicate images

In many cases, social media users may apply photo editing or add extra texts to others' images, therefore it is not sufficient to apply pixel-level comparisons to detect such duplicates. Because of this, a deep feature based duplication detection was developed. Images were firstly processed to feature vectors with a pre-trained deep model. In this case, a light-weight model, ResNet18 (He et al., 2016) was used, which generates 512 dimensional feature vectors from resized input images of $224 \times 224 \times 3$. The assumption is then, that similar images should also be close to each other in feature space, which can be revealed using clustering algorithms. In this work, the features were clustered using DBSCAN, a density-based clustering method. The application of this approach is shown in Section 6.4.3.1.

5.3 Estimation of water level from flood-relevant images

The estimation of water levels from social media images is a task that has not been much studied. Previous studies have used deep features from the whole image or local deep features based on the detected objects to classify water depth. However, these methods implicitly represent the proportion of the object that is not occluded by floodwater. Therefore, in this thesis, an explicit representation of the target-flood relationship is proposed and used for water level estimation in Section 5.3.1. For a comparison, the methods using global deep features and local deep features are used as baselines and are detailed in Section 5.3.2 and 5.3.3.

5.3.1 Learning a water level classifier with handcrafted features

A straightforward way to determine the water level from social media images is to analyze an object of known size, which is partially covered by water. The relative proportions of human bodies are well known and thus a rough estimation of the parts covered by water can straightforwardly be determined – as opposed, e.g., to buildings or vegetation. Thus, the task is to identify the body parts which are not covered by water. The aim is to determine qualitative measures in relation to human body, namely the classes ankle, knee, hip, and chest. In order to do so, three separate neural

networks were used to provide the fundamental information. First, an object detection network detected people as bounding boxes. Using the second network for body keypoint detection, each person's body parts were identified. Finally, the third network is a segmentation network, which was used to provide the surrounding information around the persons. This information is mainly used to reject the detected people who are occluded by targets other than water. The workflow of the proposed water level estimation model is presented in Figure 5.3.

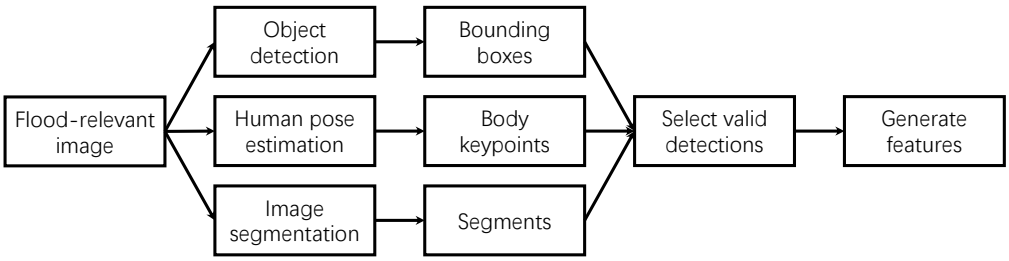


Figure 5.3: Workflow of water level estimation model

The first neural network is Mask R-CNN (He et al., 2017), which is one of the state-of-the-art frameworks for object detection. For each detected single object instance, it outputs a class label and a bounding box. The *Keras* implementation (Abdulla, 2017) of this network was used which applied the weights pre-trained on the MS COCO dataset (Lin et al., 2014). The detection determines whether the image contains people which can be subsequently used for water level estimation.

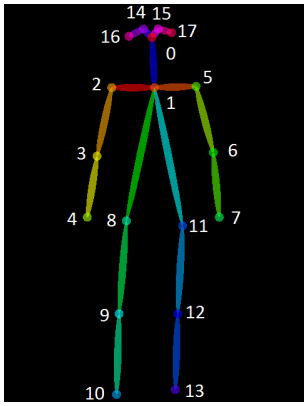


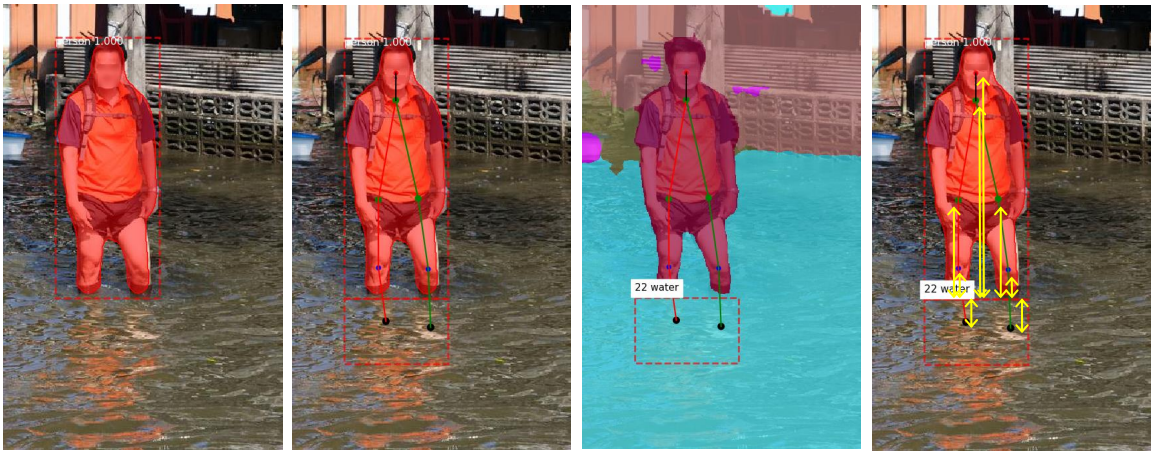
Figure 5.4: Output of OpenPose with 18 body keypoints (OpenPose, 2018).

Using the second neural network, people in the scene are detected and body keypoints are identified. In this work, OpenPose (Cao et al., 2019) was used to detect multi-person keypoints. It is a multi-stage CNN, which provides not only the detected body keypoints but also their corresponding confidence scores. The model detects 18 landmark points of the human body (OpenPose, 2018) as shown in Figure 5.4. Not all of the detected keypoints are relevant for water level estimation. Therefore, only the keypoints 0, 1, and 8-13 were selected to represent the human body, the others were neglected.

The third neural network aims at the identification of surrounding pixels of a person by semantic image segmentation. Especially, two categories are focused on, namely ground and water. In this work, Deeplabv3+ (Chen et al., 2018) was used, which is one of the state-of-the-art architectures for semantic segmentation. Specifically, a Deeplabv3+ network pre-trained on the ADE20K dataset was used (Tensorflow, 2019) for semantic image segmentation, which achieves a 82.52% pixel-wise

accuracy on the ADE20K validation set. The ADE20K dataset (Zhou et al., 2017b) was annotated with more than 250 classes, including the two classes of interest.

With outputs from the above-mentioned models, it is possible to determine the water height relative to the human body. The difference between the bounding box of the human shape and the keypoints indicates the proportions of the body parts hidden by the water. In order to determine the water level, a classifier was built, which is based on a feature vector, created by the sequence of the steps shown in Figure 5.5.



(1) People detection with pre-trained Mask R-CNN.

(2) Overlay with body keypoint detection to check validity and mark the area beneath (1/4 box height).

(3) Query the most frequent label within the area beneath from image segmentation.

(4) Extraction of distance features for people connected with water or ground area.

Figure 5.5: Steps for extracting handcrafted distance features (example image under CC BY-NC-SA 2.0).

First, the pre-trained Mask R-CNN was used to detect people, resulting in bounding boxes. Then, body keypoints were overlaid, and only the bounding boxes with corresponding body keypoint detections were preserved. In this way, only the people, which could be detected by both, the object detection model and the body keypoint model, were retained for further analysis.

In the third step, the waterline was hypothesized to be at the bottom line of the bounding box of a person, which was detected by the Mask R-CNN. An area beneath the bounding box with a box height of 1/4 of the given bounding box is marked. In this box, the most frequent class label from the segmentation results is queried (e.g., water, ground, but also classes such as cars, boats). Only the people connected to an area of ground or water were kept. Preservation of ground is necessary, as the segmentation algorithm detects water segments only in case of a severe flood. For most of the other cases (e.g., ankle level flood), flooded areas are mostly predicted as being ground. Thus, both classes – ground and water, were considered as the focus classes.

Lastly, a distance feature vector $\mathbf{D_{box_bottom}}$ was calculated from the water line (box bottom) to all used keypoints (8 values, see Figure 5.5(4)). These distances were further normalized by the box height to eliminate the influence of the unknown scale,

$$\mathbf{D_{box_bottom}} = \frac{y_{bottom} - \mathbf{Y_{keypoints}}}{y_{bottom} - y_{top}} \tag{5.1}$$

where $\mathbf{Y_{keypoints}}$ is a vector of y-coordinates of all the used keypoints in the image coordinate system, y_{top} and y_{bottom} are y-coordinates of the top and bottom line of the bounding box. These built the Feature Group 1 (FG 1).

Two additional groups of features were further considered: Feature Group 2 (FG 2) contains the confidence scores of the OpenPose keypoint detection (8 values), which indicate how well each keypoint can be detected. Lastly, Feature Group 3 (FG 3) is a binary value, which indicates whether the person is connected to a water area or a ground area. The features of each group are summarized in Table 5.3.

No.	FG1
1	Dist_Nose
2	Dist_Neck
3	Dist_RHip
4	Dist_RKnee
5	Dist_RAnkle
6	Dist_LHip
7	Dist_LKnee
8	Dist_LAnkle

No.	FG2
1	Conf_Nose
2	Conf_Neck
3	Conf_RHip
4	Conf_RKnee
5	Conf_RAnkle
6	Conf_LHip
7	Conf_LKnee
8	Conf_LAnkle

No.	FG3
1	Connect_to_water_ or_ground_segment

Table 5.3: Feature names in each feature group.

Thus, in total, a feature vector of 17 values was used to represent one person, consisting of two feature groups of 8 values each, and one additional binary feature. Then, a classic machine learning method, such as SVM (Support Vector Machine) or random forest, could be applied to determine the water height relative to the body frame, in terms of the water level classes *ankle*, *knee*, *hip*, *chest* and in addition, *no evidence*. In this work, a more state-of-the-art classifier Xgboost (Chen and Guestrin, 2016) was used for training the water level estimation model.

The annotations of the flood levels are per image, while the water depth estimation is per instance (i.e., each person in the image). This creates a potential problem, since simply assigning the image level annotations to each instance may mislead the training process. As an example, an image may show several people standing in different water levels, while some others are sitting in boats. Thus, it can be regarded as a Multiclass Multiple Instance Learning (MIL) problem. All images are regarded as bags of instances. Only the annotations of the bags are given. The model, however, needs to predict each instance in the image. One of the possible strategies is pseudo labelling. The bag annotation was assigned to each instance in the image and the model was trained. The instance level annotations were updated based on the confidence score of the softmax outputs. If the confidence score is above a relatively high value (0.85 in this work), it means the model is very sure about its prediction. Thus, this instance label was replaced with its predicted label and trained this model again. This step was repeated until no further updates happened for the instance level annotations.

Another issue is the reasoning of the final prediction for the whole image. The persons classified as *N* (no evidence) by the classifier were firstly excluded. In the case when all of the persons were excluded, the final prediction of the image is *N*. For the remaining persons, a majority vote was used to make the final prediction for the image. If the votes are equal, the prediction with a higher confidence score based on the softmax output was taken.

5.3.2 Baseline 1: Multiclass image classification with global deep features of the whole image

For a comparison with the proposed method, a simple multiclass classification was applied using global deep features as baseline. The same feature fusion architecture as described in Section 5.2.2 was applied, where features generated by pre-trained *DenseNet201*, *InceptionV3* and *Inception-ResNetV2* on ImageNet were concatenated and then classified with Xgboost. Instead of a binary

classification, softmax outputs were generated for all five water level classes. From this, the performance of this model indicates, whether the global deep features are beneficial for water level estimation.

5.3.3 Baseline 2: Mask R-CNN with extra branch for water level classification

In order to classify the water level based on the local deep features around each person, the implementation of Mask R-CNN with *Keras* provided by (Abdulla, 2017) was extended with an extra classification branch for water level classification as the second baseline. The default parameter settings were used for Mask R-CNN. A backbone network, ResNet101, was used for extracting deep features at different spatial scale, which is also known as FPN (Feature Pyramid Network). The RPN (Region Proposal Network), mask branch, classification branch, box branch were trained based on the feature maps generated from FPN separately. An extra branch was added which is the same as the classification branch for water level estimation. It classified with a cross-entropy loss based on the output of FPN. For the original parts, such as FPN, RPN, box branch and classification branch, the weights were initialized with a model pre-trained on the MS COCO dataset. The object detection parts of the network were frozen and the custom water level classification branch was trained on the dataset. Furthermore, as it is assumed that considering the area below the detected persons might contribute to the water level classification, an adapted version of this network architecture fed both the FPN outputs from the object area and the area of 1/4 of the box height beneath the object to the water level classification branch. The network architecture is shown in Figure 5.6.

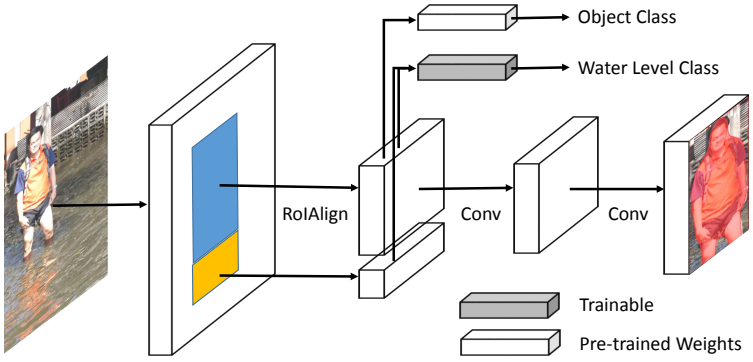


Figure 5.6: Network architecture of baseline 2: Mask R-CNN with water level classification branch using local deep features (example image under CC BY-NC-SA 2.0).

The idea of this baseline is similar to (Chaudhary et al., 2019). The main difference is that they trained the model from scratch based on pixel-level annotations of their flood level dataset and part of the MS COCO dataset, whereas this baseline inherits the object detection function directly from pre-trained weights. The flood level dataset was only used for tuning the FPN and training the water level classification branch. Therefore, this method only needs to provide water level labels for each person instance and the model can be trained with the pseudo labelling strategy with one single label for the whole image as described in Section 5.3.1. Thus, the annotation effort for this baseline is much less. Additionally, Chaudhary et al. (2019) used not only people, but also many other object classes, such as houses and cars, whereas the dataset used in this research is annotated only based on the water level measure of persons. Even though this baseline is not the same as the work from Chaudhary et al. (2019), it generally represents the ability of water level classification, which makes use of the local deep features around detected persons.

6 Experiments to extract flood observations from social media VGI

In this chapter, the experiments to extract flood observations from social media VGI are presented. In Section 6.1, the infrastructure for social media data acquisition is introduced. In addition, the datasets prepared for supervised classification of flood observations are described in Section 6.2. In Section 6.3, the flood-related social media posts during flood events in Europe in 2016 are extracted and analyzed. Apart from merely classifying posts as flood relevant or not, Section 6.4 extracted flood severity observations based on image interpretation and prepared a VGI-based flood severity map for Hurricane Harvey in 2017. The experiments in this chapter are also presented in (Feng and Sester, 2018) and (Feng et al., 2020a).

6.1 Social media data acquisition

The social media data used in this thesis is from Twitter. Twitter reported having 187 million daily active usage worldwide (Twitter, 2020), which leads to a large amount of user-generated data. Because of its public Streaming API (Twitter, 2021a), the real-time data streams of Twitter can be accessed. However, the number of Tweets that can be crawled is restricted by the request limit (Twitter, 2021b). The API allows a pre-filtering according to a geographical bounding box, keywords or languages. Therefore, instead of collecting Tweets globally, Twitter data are collected for five study areas individually to mitigate the request limit. These study areas are defined as shown in Figure 6.1, namely *US West*, *US Middle*, *US East*, *Europe West*, and *Europe East*.

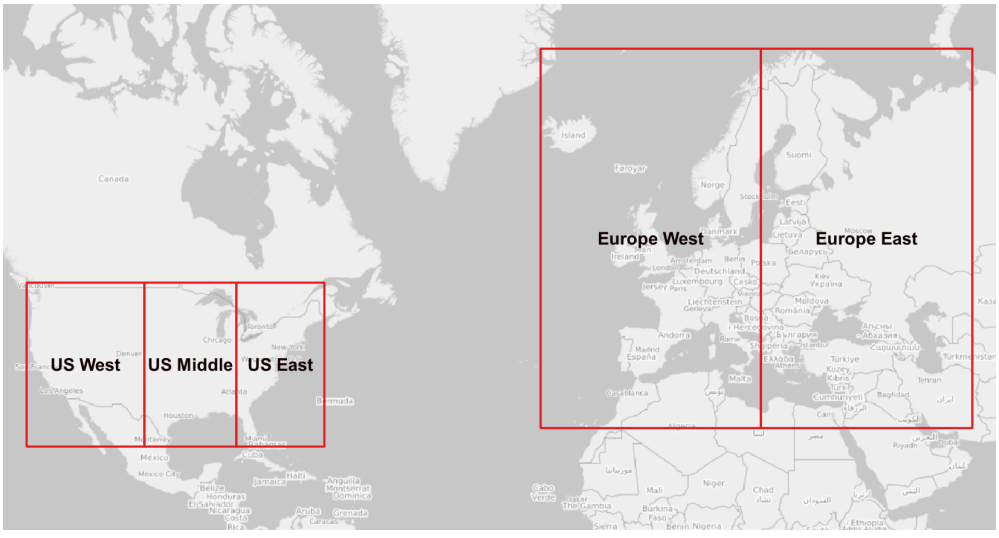


Figure 6.1: Study areas for collecting Twitter data (Basemap: OpenStreetMap).

In addition, the data stream was also filtered according to language and preserved only the Tweets in seven frequently used languages within the study areas of focus, namely *English*, *French*, *German*, *Italian*, *Spanish*, *Portuguese* and *Dutch*. These are also the languages currently well supported by the NLP tools. At this step, no keyword filtering was applied to the Streaming API. By restricting the area and filtering with respect to languages, the limitation by the Streaming API is greatly overcome. Higher completeness of the crowdsourcing data has been achieved.

In order to collect data from all study areas simultaneously, five processes were deployed to collect Tweets at the same time. The Tweets were downloaded in JSON format and subsequently stored in a MongoDB[1] database. Many Tweets may have a different number of fields. Therefore, the MongoDB database, as a NoSQL database (Moniruzzaman and Hossain, 2013) is ideal for this kind of data, as it does not require all the documents to have exactly the same fields. The data collection began on 15th of May 2016. During the collection process, there have been intentional interruptions in some of the study areas to achieve a proper allocation of storage resources. From the extensive data collection, two datasets were identified for further investigation and case studies. One focused on floods in Europe[2] in 2016. The other is the Tweets collected during Hurricane Harvey[3] in 2017.

6.1.1 Floods in Europe in 2016

In late May and early June, intense rainfall has caused severe flooding in several European cities. Flash floods affected Braunsbach (Sim, 2016a) and Simbach am Inn (Sim, 2016b), Germany, on May 29, 2016, and caused several deaths. The estimated overall losses amounted to EUR 2.6 billion (Munich RE, 2017). Heavy rainfall also caused the Seine River in France to burst its banks, and the river was 6 meters higher than the normal water level. The Louvre museum was closed to evacuate the masterpieces (The Guardian, 2016). In the UK, London, Manchester and many of the major cities experienced frequent flooding in June due to heavy rainfall (BBC, 2016; Independent, 2016). Therefore, the Twitter posts collected within the study area *Europe West* are retrieved to analyze the social media behaviours during these pluvial flood and fluvial (river) flood events. From 1st of June 2016 to 28th of October 2016, about 18 million geotagged Tweets were collected within the study area *Europe West*.

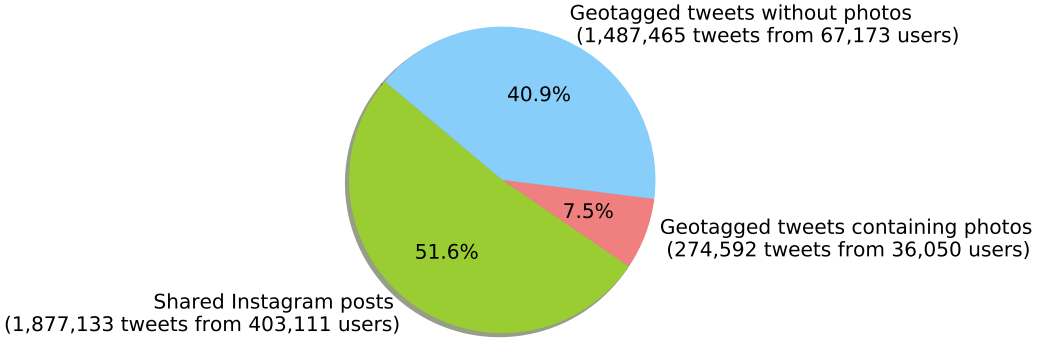


Figure 6.2: Proportion of geotagged Tweets containing photos.

A sample of one month's data from 1st of June 2016 to 30th of June 2016 is used to demonstrate the percentage of geotagged Tweets with photos (Figure 6.2). A total of 3.6 million geotagged Tweets were collected from 473,004 users. 59.1% of the geotagged Tweets contain photos or references to photos on Instagram. The majority of these Tweet are shared Instagram posts with shortened text and URL link. Therefore, an extension to download these Instagram posts with full texts and images was also developed to improve the completeness of the collected VGI data.

[1] MongoDB. `https://www.mongodb.com/` (Accessed on 31.01.2021)
[2] Wikipedia - 2016 European floods. `https://en.wikipedia.org/wiki/2016_European_floods` (Accessed on 31.01.2021)
[3] Wikipedia - Hurricane Harvey. `https://en.wikipedia.org/wiki/Hurricane_Harvey` (Accessed on 31.01.2021)

6.1.2 Hurricane Harvey in Texas, United States in 2017

Together with Hurricane Katrina in 2005, Hurricane Harvey was regarded as the costliest hurricane by National Hurricane Center (NOAA, 2018). It inflicted a damage of $125 billion in Southeast Texas, especially the Houston metropolitan area. The strong precipitation led to a severe flood in the Houston area from 25th of August to the 1st of September 2017. Therefore, the Twitter posts collected within the study area *US Middle* are retrieved to analyze the social media images during this hurricane event. Spatially, this data collection covered the whole disastrous area, and temporally, it covered all eight days with significant flood events. From 25th of August to the 1st of September 2017, a total of 150,227 Tweets with either geo-coordinates or location information were retrieved in the Houston area.

6.2 Datasets for training classification models

In order to interpret social media data in an automatic manner, supervised classification has been performed in the experiments of this thesis. To this end, data have to be annotated with pre-defined categories. Four datasets were prepared, one for text classification (Section 6.2.1), two for image classification (Sections 6.2.2 and 6.2.3), and one for water level estimation (Section 6.2.4).

6.2.1 Text dataset annotated via keyword filtering and cross-referencing weather data

With the process described in Section 5.1.1, a text dataset is prepared by combining pre-defined keywords and weather data. From 1st of July 2016 to 28th of October 2016, about 14.4 million of geotagged Tweets were collected within the study area *Europe West*. After keywords filtering, 51,732 Tweets (from 36,002 users) were identified as potential training examples. Using the Weather API[4], 36,469 (70.5%) of them were labeled as positive and 15,263 as negative. In order to coarsely verify the automatic labeling, 100 randomly selected Tweets were manually checked which were labeled as positive by the weather API: 94 of them are correctly labeled.

Training on an imbalanced dataset may lead to over-prediction of the presence of the majority class (Wei and Dunbrack Jr, 2013). For a binary text classification, a balanced training dataset is beneficial. Therefore, 21,206 randomly selected Tweets without any keywords were used as a supplement of the negative training examples to balance the training data. In this way, a balanced dataset was prepared for training the text classifiers. The final training dataset contained totally 72,938 Tweets with 65,772 unique words. They were sent by 50,701 users. The average number of words for each document after pre-processing is 6.5.

6.2.2 Manually annotated pluvial flood image dataset

The dataset for training the image classifiers has three subsets, which have been collected and labeled by one annotator. Each of them contains 7600 images. The first subset contains social images irrelevant to flooding or rainfall events. The second subset contains images relevant to flood and rainfall events. The third subset contains images of water surfaces, such as rivers, lakes, coast or swimming pools.

Social media images not related to flooding or rainfall events were collected in the first subset. Photos in social media have their own distribution on each topic, such as artworks, selfies or photos

[4]Weather API: Introduction. `https://www.wunderground.com/weather/api/d/docs` (Accessed on 07.11.2017)

of the surroundings (as shown in Figure 6.3a). They all have their own proportion in the overall data stream. In order to preserve this proportions in the dataset, random selections of Tweets with photos from 1st of July 2016 to 28th of October 2016 were given to this annotator. At the end, 7600 images unrelated to flooding or rainfall events were collected.

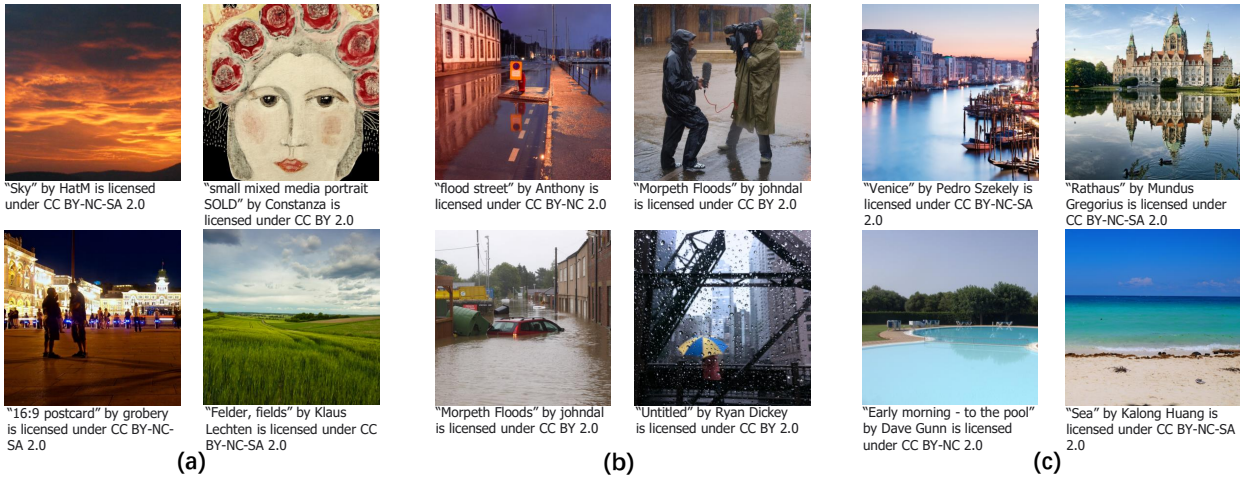


Figure 6.3: Examples of training dataset: (a) rainfall and flooding irrelevant images, (b) relevant images and (c) images of water surfaces.

Flood- and rainfall-relevant images were collected in the second subset. Since the proportion of such photos is very small with respect to the whole data stream, it is very time consuming to collect many positive examples by filtering the Twitter data collected in this research. Therefore, images relevant to flood and rainfall events were manually collected from the Internet using a search engine and the search tools provided by Twitter and Instagram. As it is concentrated on extracting evidences for flood and rainfall events, this subset included scenarios such as people or vehicles standing beside or in the water, raindrops on the windows or on objects as well as wet or flooded streets (as shown in Figure 6.3b).

Images of water surfaces such as lakes and rivers are collected in a third subset. Images of flood scenes and water surfaces often have similar visual features. However, they can be distinguished by human, for example, based on the image brightness and the targets of interest in the image. Therefore, this subset was collected in the same way as the second subset, which contains images of water surfaces, such as rivers, lakes, coast or swimming pools (as shown in Figure 6.3c). In this dataset, flooding and rainfall relevant photos were excluded. It is worth mentioning that the photos in the first subset also contain some photos of the water surfaces, however, the amount of such photos is very small and only with respect to the distribution of normal social media images in the data stream.

6.2.3 MediaEval'17 MMSat benchmark dataset and its extension

In addition to the image dataset collected for extracting social media posts for pluvial floods, existing benchmark datasets are also considered to evaluate the proposed methods. DIRSM (Disaster Image Retrieval from Social Media) is a benchmark dataset offered by *MediaEval'17 MMSat Task* (Bischke et al., 2017b), where Flickr images were manually annotated as being flood relevant or irrelevant.

Since images from social media such as Twitter or Instagram may vary largely in quality (as explained in Section 2.5.1), three additional data sources were introduced, namely the annotated

image collection from Section 6.2.2, 4000 randomly selected images from the two-class Weather Classification Dataset (Lu et al., 2014), and images annotated as containing scenarios where roads are not passable during the flood from *MediaEval'18 MMSat Task* (Bischke et al., 2018). This dataset was named *Extended DIRSM*. In this way, a relatively balanced dataset was built, which is beneficial for both training and evaluation. The distribution between training vs. test and positive vs. negative examples is summarized in Table 6.1.

Table 6.1: Number of positive and negative examples for dataset.

Dataset	Number of Negative Examples	Number of Positive Examples
DIRSM	3360 (train) + 840 (test)	1920 (train) + 480 (test)
Extended DIRSM	9945 - 3360 (DIRSM, train) - 2000 (Two-class weather, cloudy) - 2000 (Two-class weather, sunny) - 2585 (Own collection partly from Section 6.2.2)	9625 - 1920 (DIRSM, train) - 1206 (MediaEval'18, road not passable) - 6499 (Own collection partly from Section 6.2.2)

6.2.4 Image dataset for water level estimation

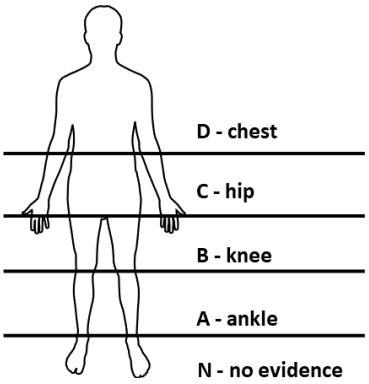


Figure 6.4: Annotation rules for water level estimation of single person.

To the best of our knowledge, there is no public dataset or benchmark available for the task of water level estimation. The only comparable dataset which appeared in previous research is used for the work of Chaudhary et al. (2019), where 7000 images were annotated pixel-wise into 11 water level classes. The images were collected from various Internet sources, such as news articles, search engines, and social media. However, it is not yet publicly available. In this work, images from flooding or heavy rainfall scenarios which contain at least one person were collected with similar data sources. As the final goal of this work is to provide one water level estimation for each geotagged social media image, only *one* water level estimate for each image is of interest. Thus, there is no need to provide pixel-wise labels for every image. For this reason, instead of an annotation of all image pixels, the whole image was annotated with one single label, which is much less time-consuming. The images were annotated into five classes with the rules shown in Figure 6.4. N stands for all persons who have no evidence for water level estimation, e.g., standing on wet ground, standing on the river bank, or sitting in a boat. From A to D, the label is associated with the water level at ankle, knee, hip and chest. Regarding the case when multiple people stand in the water (in different heights), these images were annotated with the label of the majority. The images were annotated according to this rule by one annotator.

1375 images containing persons in flood or heavy rainfall situations were collected and annotated into the above mentioned 5 classes. Additional 325 images from the MS COCO dataset were introduced as class N (no evidence), which contains mostly the situation of people standing on the ground with no significant water level. From each class, 50 images were randomly selected as the test set and kept unseen during training. In total, 1700 images were used; the composition of the dataset is summarized in Table 6.2.

Table 6.2: Composition of train set and test set for water level estimation

Class Name	Train Set	Test Set
N - No evidence	450	50
A - Ankle	250	50
B - Knee	250	50
C - Hip	250	50
D - Chest	200	50

6.3 Extraction of pluvial flood-relevant VGI based on social media texts and photos

The workflow of the proposed approach is shown in Figure 6.5. In this case, only the social media posts including both texts and photos are analyzed. Classifiers for texts and images are trained and applied separately, and the individual evidences are combined. In the end, events are detected by spatiotemporal analysis.

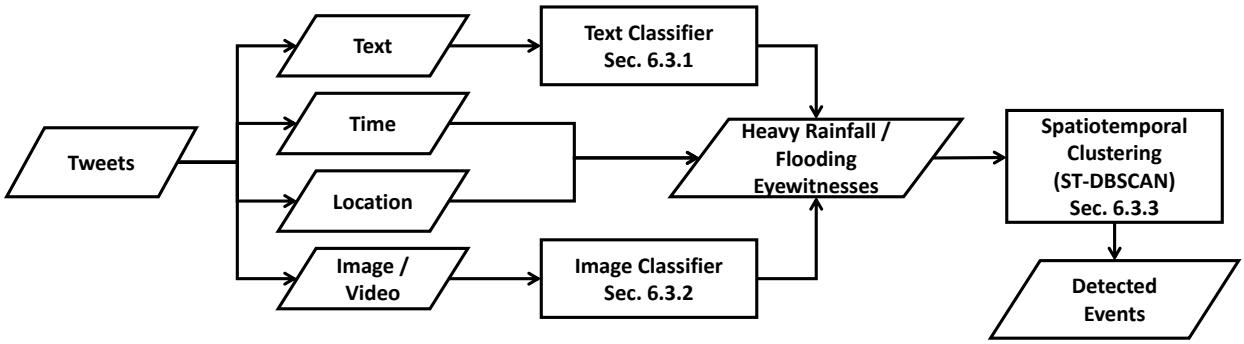


Figure 6.5: Workflow for the extraction of pluvial flood relevant VGI.

6.3.1 Training of the text classifier

The methods introduced in Section 5.1.2 were used to train the text classifiers. 10% of the dataset was randomly selected as test set (as described in Section 6.2.1) and used only for methods comparison. Since most of the classification methods need hyperparameter tuning, grid search with 5-fold cross validation on the remaining 90% of dataset was used to find the optimal hyperparameters for each method (as summarized in Table 6.3).

After training the models with the optimal hyperparameters, the performance of all methods are compared and evaluations were given based on the test set with the metrics such as the accuracy, precision, recall and F_1-score. F_1-score, precision and recall are the metrics calculated based on one single class, the flood and rainfall relevant class. The results are shown in Figure 6.6 and Table 6.4. The ROC (Receiver Operating Characteristic) curves (as shown in Figure 6.7) for each method and area under the curve (AUC) were also calculated and used as criteria for comparing the text

Table 6.3: Parameters used for training the text classifiers.

Method	Parameters
Random Forest	max_depth = 60, n_estimators = 300
Logistic Regression	C = 1.0, penalty = 'l2'
SVM (Linear Kernel)	C = 1.0, gamma = 'auto'
SVM (RBF Kernel)	C = 100.0, gamma = 0.01
TextCNN	learning_rate=0.001

classifiers. All experiments in this research were performed on a PC with Intel Core i7-4790 CPU, 16 GB RAM and one NVIDIA GeForce Titan X GPU. The runtime for training the models is also summarized in Table 6.4.

As shown in Figures 6.6 and 6.7, six text classification methods were compared. The deep learning method using word2vec word embedding and TextCNN outperformed the other methods and achieved an accuracy of 78.68%. The AUC for ROC of this method is also larger than the others. Except for the naive Bayes, the rest of classical NLP methods using tf-idf matrix as input perform relatively similar. Due to its performance, the trained model using TextCNN was embedded into this application. Concerning runtime, TextCNN needs significantly more time for training. Naive Bayes and logistic regression are the methods which could be trained with less time. However, for an operational use, only the prediction time is relevant, which is similar for all classifiers.

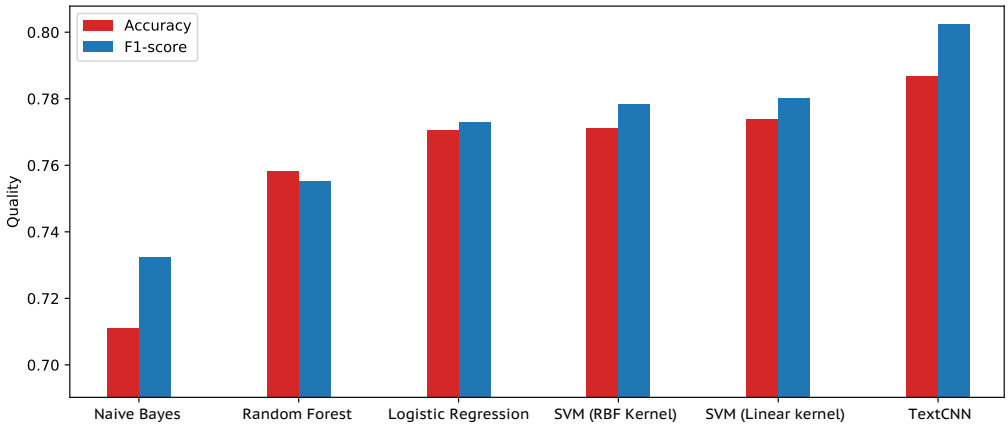


Figure 6.6: Comparison of text classification methods on test set.

Table 6.4: Evaluation of text classification methods.

Method	Accuracy	Precision	Recall	F_1-Score	Runtime (s)
Naive Bayes	71.09	69.29	77.69	73.25	**0.02**
Random Forest	75.82	**77.97**	73.24	75.53	182.1
Logistic Regression	77.05	77.93	76.66	77.29	0.53
SVM (RBF Kernel)	77.12	76.87	78.81	77.83	286.0
SVM (Linear Kernel)	77.39	77.32	78.71	78.01	207.2
TextCNN	**78.68**	75.98	**85.03**	**80.25**	1124.8

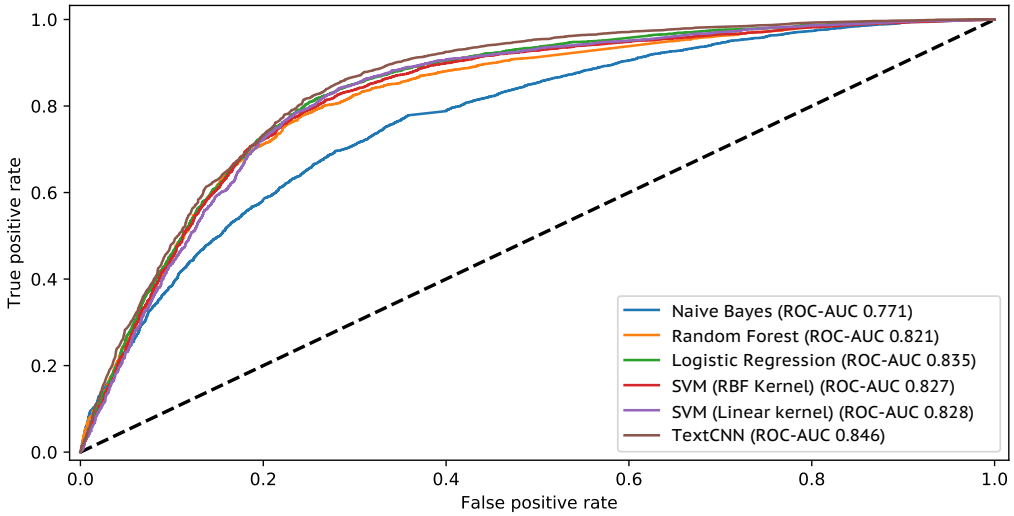


Figure 6.7: ROC curves of text classification methods.

In order to demonstrate the performance of the model trained on such an automatically labeled dataset, further validation can be performed based on manually labeled data. However, a dataset specifically prepared for pluvial flood (i.e., including observations on both rainfall and flood events) does not exist. Therefore, a manually annotated Twitter dataset[5] for a general flood event in Queensland in 2013 was used for a comparison. It contains 6,019 Tweets for training, and 3,011 for testing. Of these, 1,625 were positive labels and 1,386 were negative labels. Alam et al. (2018) compared in their experiments three training strategies. The first approach is a supervised model, which learns with Word2vec word embeddings similar to TextCNN. The second approach is a semi-supervised model named self-training, which first trained a supervised model with labeled data and then generated annotations for approximately 21,000 unlabeled Twitter texts. Only the generated annotations with a confidence level greater than 0.75 are used to retrain a new model. The third approach is a graph-based semi-supervised model (Yang et al., 2016), which learns the feature representations of text documents in a graph and encodes the labeled and unlabeled text for semi-supervised classification.

Table 6.5: Evaluation of text classification on 2013 Queensland floods dataset.

Method	**Num. of Input**	**AUC**	**Prec.**	**Rec.**	**F_1**
Alam et al. (2018)					
- Supervised	6,019 (manual labeled)	80.14	80.08	80.16	80.16
- Semi-supervised(Self-training)	6,019+21k (unlabeled)	81.04	80.78	80.84	81.08
- Semi-supervised(Graph-based)	6,019+21k (unlabeled)	92.20	**92.60**	**94.49**	**93.54**
This work					
- TextCNN	72,938 (auto-labeled)	**95.95**	85.88	82.50	82.32

The performance comparison is presented in Figure 6.5 based on the weighted average AUC, Precision, Recall, and F_1-score. On this test set, the model trained in this section achieved a performance better than the supervised model and semi-supervised model using self-training. It has also achieved an AUC higher than the graph-based semi-supervised model. The models presented in Alam et al. (2018) all need a certain amount of labeled data to initialize the semi-

[5]CrisisNLP. `https://crisisnlp.qcri.org/` (Accessed on 31.01.2021)

supervised training process. It is worth noting that the model proposed in this section does not require any manual annotations. The annotation process is performed automatically using keyword filtering and querying the corresponding weather records based on date and geotags. Still, it has demonstrated an overall performance better than a model similar to TextCNN trained on labeled dataset.

6.3.2 Training of the image classifier

Similar to training the text classifiers, 90% of the dataset was used for training. Hyperparameters for each method were tuned by 5-fold cross-validated grid-search. The evaluation was given based on the rest 10%, namely the test set. The hyperparameters used for training the final model for each method are summarized in Table 6.6.

The classification methods used for transfer learning were tested firstly on Subset 1 and Subset 2 (as introduced in Section 6.2.2) and the evaluations are given with accuracy, precision, recall and F_1-score on the test set. The ROC curves for each method and AUC were also used as criteria for evaluation. With the same computer as described in Section 6.3.1, the training time for each method was also recorded.

Table 6.6: Parameters used for training the image classifiers.

Method	**Subset 1 and Subset 2**	**Subset 2 and Subset 3**
Logistic Regression	C = 1000.0 penalty = 'l1'	C = 10000.0 penalty = 'l2'
Random Forest	max_depth = 60 n_estimators = 300	max_depth = 30 n_estimators = 300
Multilayer Perceptron	num_hidden_units = 8 learning_rate = 0.005	num_hidden_units = 8 learning_rate = 0.01
Gradient Boosted Trees	n_estimators = 300 learning_rate = 0.05	n_estimators = 150 learning_rate = 0.1
Xgboost	eta = 0.32, gamma = 0.01 max_depth = 15	eta = 0.32, gamma = 0.05 max_depth = 15

As shown in Table 6.7 and Figure 6.8, the classifier which was trained based on transfer learning achieved the best performance using the Xgboost implementation of gradient boosted trees. Both the accuracy and F_1-score reached 92.8% and the AUC of ROC achieved the maximum compared with other methods (as shown in Figure 6.9). It was followed by the random forest and gradient boosted trees; even the worst case, a simple logistic regression, could achieve an accuracy of about 88%, which shows the transfer learning approach can really distinguish raining or flooding scenarios from social media images. When comparing the runtime of each classification method, the gradient boosted trees from scikit-learn is much more time consuming than Xgboost. The multilayer perception is the method with the least training time.

Even though high accuracy and high F_1-score were achieved on the test dataset, the classifier was still found to be not optimal classifying the images containing water surfaces, i.e., water surfaces such as lakes or rivers were sometimes classified as positive. Therefore, a second classifier was trained only to distinguish the pluvial flood-relevant images from the scenarios containing water surfaces. The same transfer learning approach was utilized but only with different input data, which contained 7600 images relevant to raining and flooding and another 7600 images containing only images of lakes, rivers, as in the second and third subsets of dataset shown in Figure 6.3b,c.

Table 6.7: Evaluation of image classification methods.

Method	Accuracy	Precision	Recall	F_1-score	Runtime (s)
Logistic Regression	88.86	90.04	87.52	88.76	138.8
Multilayer Perceptron	89.07	**97.45**	80.36	88.09	**22.9**
Random Forest	91.33	94.97	87.38	91.02	117.9
Gradient Boosted Trees	92.52	93.42	**91.58**	92.49	669.8
Xgboost	**92.95**	94.36	91.44	**92.88**	121.2

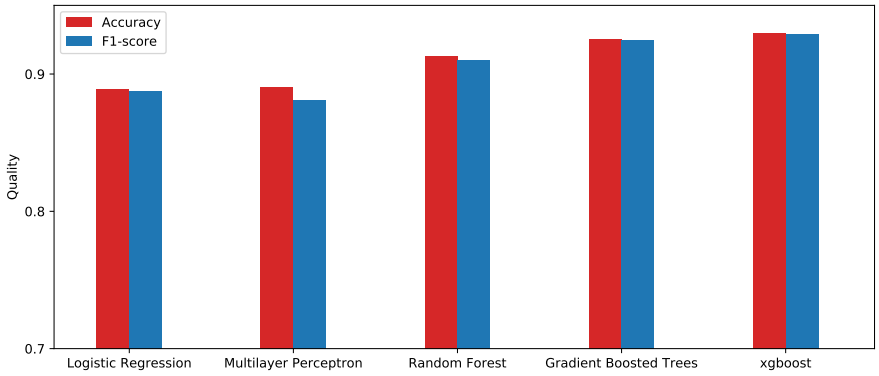


Figure 6.8: Comparison of image classification methods on test set.

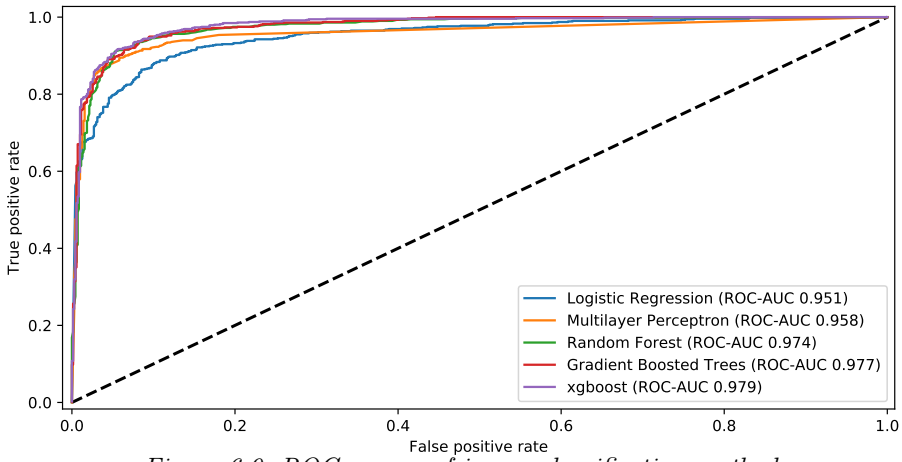


Figure 6.9: ROC curves of image classification methods.

As shown in Table 6.8 and Figure 6.10, similar performance as for the first classifier was observed, however, with lower values. Xgboost outperforms the other methods and has a highest accuracy and F_1-score. It has also achieved the largest AUC for ROC (as shown in Figure 6.11). Comparing to the performance achieved using pluvial flood-irrelevant images of any content as negative examples, the classifier accuracy of 87.38% indicated that it is more complicated to distinguish rain and flood images from images with lakes or rivers.

Therefore, together with a pre-trained model used as a feature generator, the two trained Xgboost models were embedded in this application. Only the images predicted by both classifiers as positive were regarded as rain and flood relevant images. A visual inspection of the wrongly classified photos (false positives) was conducted. They can be generally grouped into three categories as the examples presented in Figure 6.12. Firstly, many photos with water surfaces in relative dark color were wrongly classified. Secondly, the images containing reflecting area (e.g., windows, floor),

which was similar to water reflection were sometimes not well classified. Lastly, photos containing fountains or springs, which have contents like water drops, were also hard to be correctly classified.

Table 6.8: Evaluation of image classification methods.

Method	Accuracy	Precision	Recall	F_1-score	Runtime (s)
Logistic Regression	84.07	84.53	84.95	84.74	221.3
Random Forest	85.55	87.63	84.11	85.84	158.1
Multilayer Perceptron	86.25	**89.15**	83.78	86.38	**16.1**
Gradient Boosted Trees	86.95	88.36	86.29	87.31	425.3
Xgboost	**87.38**	88.72	**86.79**	**87.74**	134.2

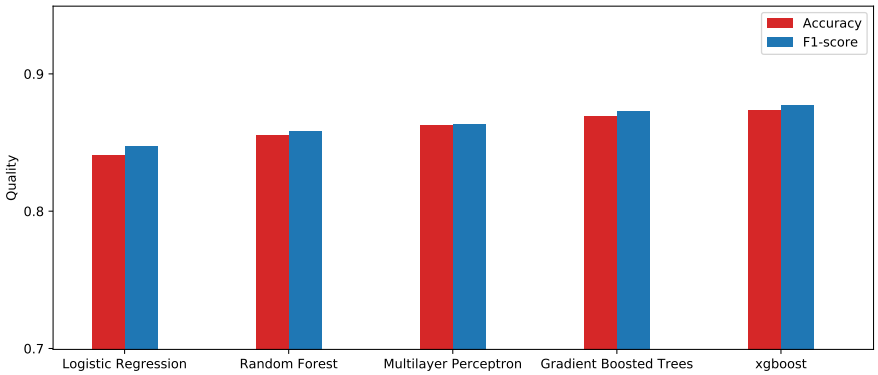


Figure 6.10: Comparison of image classification methods on test set.

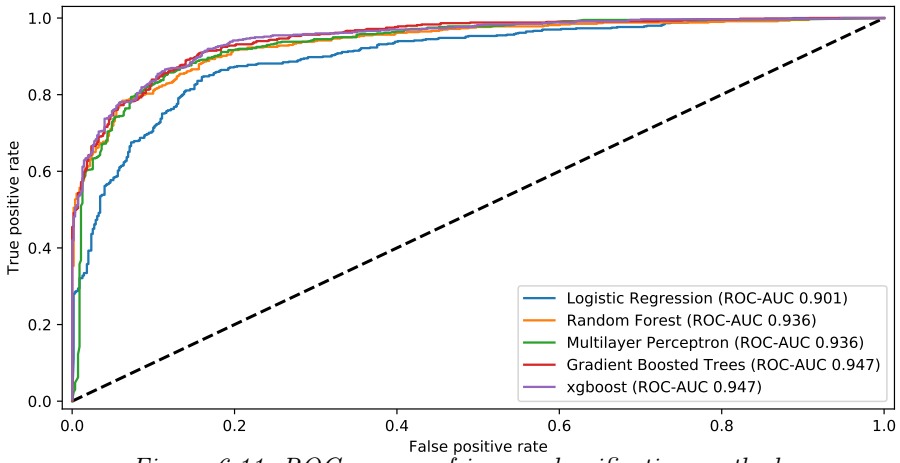
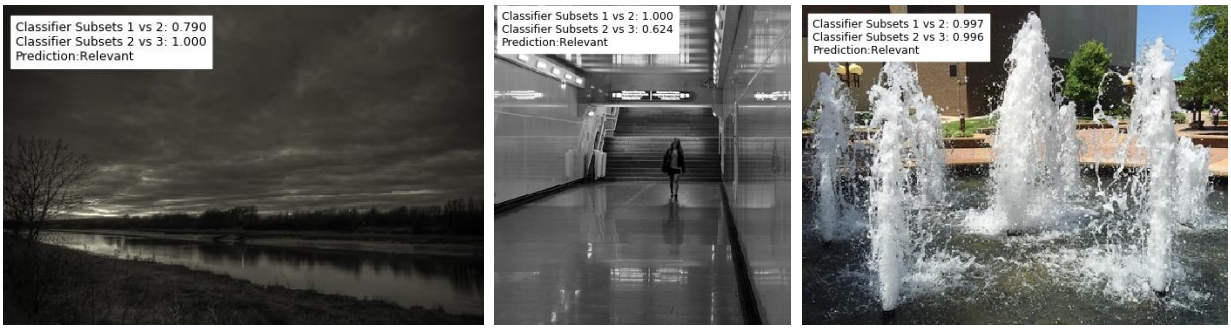


Figure 6.11: ROC curves of image classification methods.

6.3.3 Detection of heavy rainfall and flood events

In order to detect heavy rainfall and flood events, the geotagged Tweets containing both texts and images were processed. Only the Tweets with positive predictions from both filters were regarded as high quality observations for such events, and the corresponding geo-locations are treated as pluvial flood related areas. These points were aggregated subsequently to detect events with spatiotemporal clustering and a hot spot map was generated using Getis-Ord Gi* (Ord and Getis, 1995) with respect to the city administrative regions. As presented in Section 2.5.1, the locations of the social media posts are of heterogeneous quality. The contents of some Tweets may



"dark skies over the river lune" by "These * Are * My * Photons" is licensed under CC BY-NC 2.0

"Reflections" by Sergey Kochkarev is licensed under Public Domain

"Fountain" by 10001dp is licensed under CC BY-NC-SA 2.0

Figure 6.12: Three typical failure cases of the classifiers: water surfaces in relative dark color (left), images containing reflecting area (middle), and photos containing fountains or springs (right).

not be always associated with the posted coordinates. Therefore, spatiotemporal clustering is used for event detection, which requires a minimum number of posts at a certain location. Moreover, the main focus of this research lies on pluvial flood events, thus during the training of classifiers, texts and images containing general rainfall-relevant information were taken into consideration.

6.3.3.1 Event detection with spatiotemporal clustering



Figure 6.13: Spatiotemporal cluster detected by ST-DBSCAN (pluvial flood in London on 26th of June 2016, green markers are the aggregated Tweets only for visualization).

Spatiotemporal clustering can be used to detect events based on spatiotemporal patterns among data points. ST-DBSCAN (Birant and Kut, 2007) is an extension of the density based clustering method DBSCAN (Ester et al., 1996) into spatiotemporal space. Three parameters are needed for this method: the maximum spatial distance ε_1, the maximum time difference ε_2 and the minimum number of points to form a cluster $MinPts$. Since users may send several posts at the same place and same time with very similar contents, posts from one single user may already be enough to create the spatiotemporal clusters without confirming from others. Therefore, instead of using the

minimum number of Tweets, $MinPts$ is redefined in this case as the minimum number of different Twitter users.

For cities of different size, the spatial distribution of Twitter users varies significantly. According to the literature, the cell size of intense rain is generally less than 10 km in the UK (Begum and Otung, 2009) and rain with duration of 3 h contributes the most to total summer precipitation (Thorp and Scott, 1982). With this guidance, different combinations were checked and visually compared. At the end, an optimal setting used for the results in London (as shown in Figure 6.13) is $\varepsilon_1 = 8$ km, $\varepsilon_2 = 1.5$ h and $MinPts = 3$ users.

6.3.3.2 Polygon based hot spot detection with Getis-Ord Gi*

Getis-Ord Gi* (Ord and Getis, 1995) is one of the frequently used geostatistics methods for hot spot detection. This method also takes the local neighbourhood into account. In this case, administrative polygon data for the cities were used to represent the local neighbouring relations. This method was applied to find the statistical hot spots for the extracted rainfall and flood-relevant Tweets. The principle of Getis-Ord Gi* is to compare local averages to global averages. The results after applying this method are the z-scores, which represent the statistical significance. They indicate the particular value for each polygon relative to the global average. The z-scores are frequently used to determine the confidence threshold. The statistical significance can be calculated using the resultant z-scores. A z-score of 1.65 represents a 90% confidence level, 1.96 for 95%, 2.58 for 99%, and 3.29 for 99.9% (ESRI, 2019b).

The number of Tweets in each part of the city is different because of the difference in social media users' density. A simple hot spot detection directly based on the number of topic-relevant Tweets may frequently lead to the appearance of hot spots at the city center or somewhere more people are living. To avoid this, the total number of Tweets in the same city over a 90-day period was aggregated and the average number of Tweets collected per day in each polygon was calculated. An example in Paris, France was generated and shown in Figure 6.14. The polygons represent the 80 administrative districts provided by Open Data Paris[6]. It is obvious to find that the areas including places of interest or shopping zones in Paris are highlighted. This statistic was used as a basis for inspecting the places where normally few Tweets are sent, but suddenly a large number of Tweets appears at that area. This may indicate a more reasonable hot spot region for pluvial flood event.

The coordinates of many Tweets are only in city level, for instance, user may provide the single point coordinate representing 'Paris, France' when they sent a social media post. Such imprecise coordinates were also recorded. Thus, these Tweets representing the cities were filtered out before the hot spot detection. After that, the ratios of the number of filtered Tweets and the daily average number of Tweet were calculated for each polygon. Based on the Tweets collected in Paris on 3rd of June 2016, a map of ratios (as shown in Figure 6.15) was generated. This ratio map is then used as the input for Getis-Ord Gi* hot spot detection. From the result, a map of the z-scores (as shown in Figure 6.16), a situation in Paris could be identified, showing that the regions along the riverbank of the Seine were highlighted during this fluvial (river) flood event. Comparing with Figure 6.15, it could achieve a better neighbouring consistency.

[6]Quartiers Administratifs - Open Data Paris. `https://opendata.paris.fr/explore/dataset/quartier_paris/information/` (Accessed on 31.01.2021)

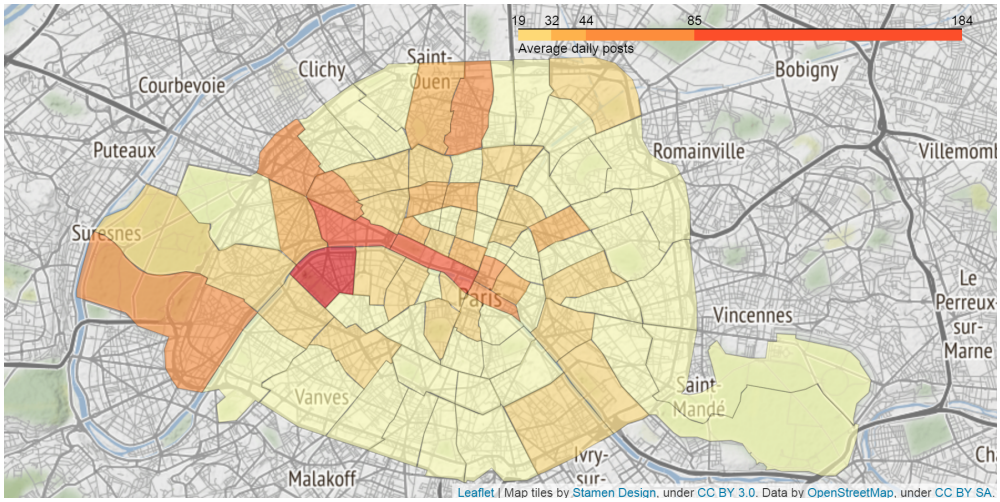


Figure 6.14: Map of daily average number of Tweets based on aggregation of 90 days' Tweets.

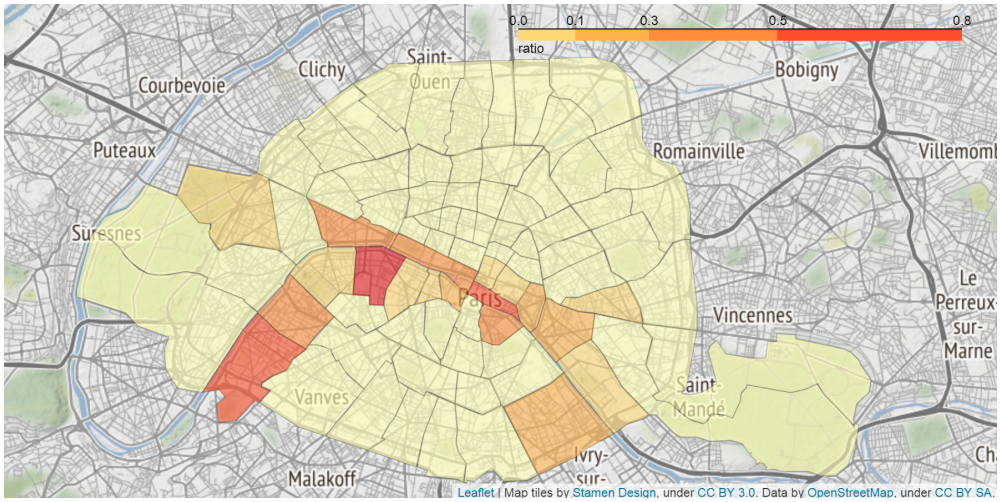


Figure 6.15: Ratio map on 3rd of June 2016 in Paris.

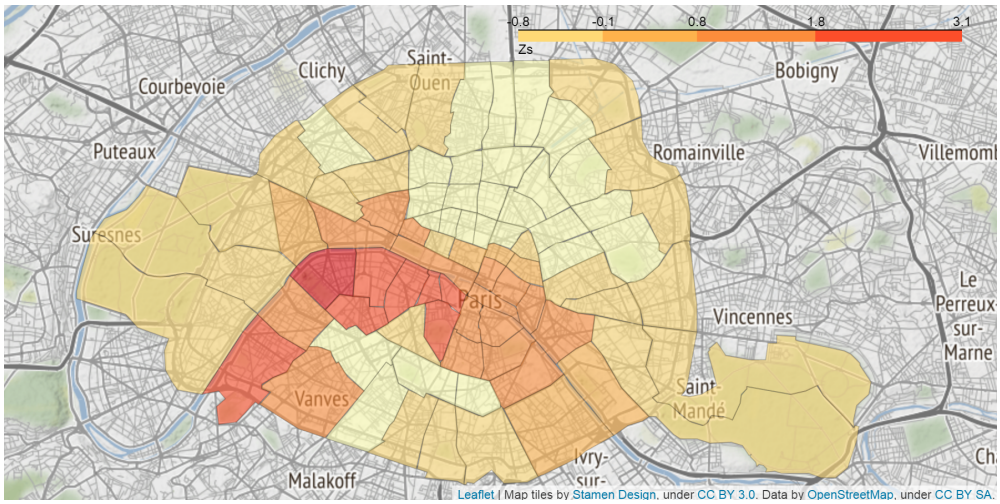


Figure 6.16: Hot spots detected by Getis-Ord Gi on 3rd of June 2016 in Paris.*

6.3.4 Visualization of the pluvial flood relevant information

A further test of this framework was applied during the pluvial flood in Berlin, Germany on 29$^{\text{th}}$ of June 2017. A heavy rainfall stroke Berlin and led to severe inundation in the city and failure of the drainage systems (B.Z., 2017). This application could generate for each day a report with social media posts regarding rainfall or flood events. The posts are then visualized as clustered point markers. After clicking the marker clusters, the detailed information of each Tweet can be accessed by opening the links in pop-up window at the user given locations. By this approach, overlaps of data points are avoided. Spatiotemporal clusters are visualized as a light blue circle and the radius is set as the bigger eigenvalue calculated based on the data points belong to the same spatiotemporal cluster. Hot spots are also detected based on the prediction from both text and image classifiers and visualized as a choropleth map (as shown in Figure 6.17). The polygons represent the 138 regions defined by Life-World Oriented Spaces (LOR)[7], which is a partition of the city of Berlin frequently used for statistic and demography. It is also available under Berlin Open Data[8].

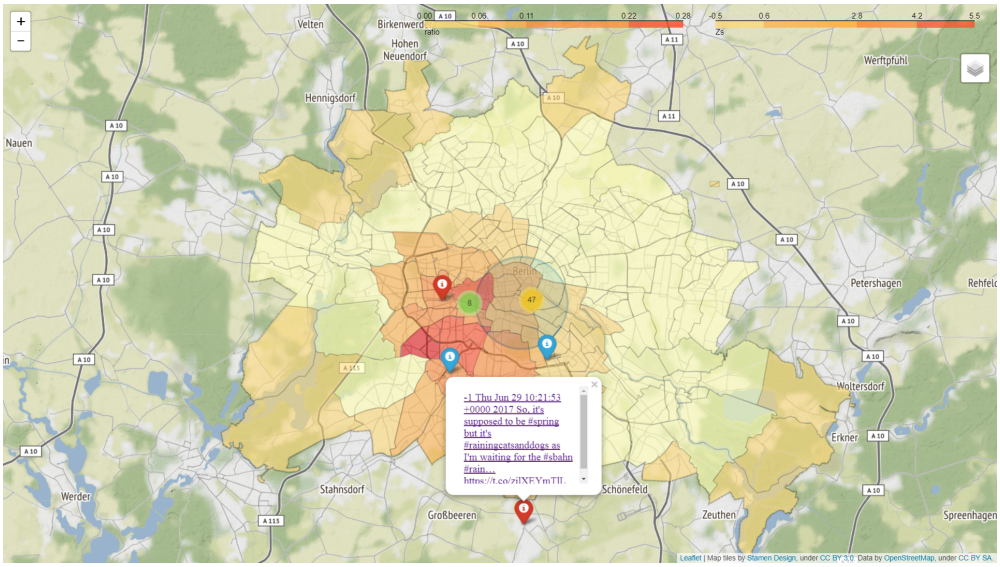


Figure 6.17: Screen-shots of the web map application (pluvial flood in Berlin on 29th of June 2017).

6.3.5 Analyses and comparison with external data sources

In the previous sections, user-generated texts and photos were used to identify flood and rainfall relevant social media posts. As the aim is to evaluate whether the extracted Tweets are relevant for the real world events, additional data can be used for a correlation analysis. Since pluvial floods are normally associated with heavy rainfall events, rainfall intensity can be an additional information which is latently related to the occurrence number of flooding relevant Tweets. In this case, the precipitation data recorded by Weather Underground[9] was accessed. As classifiers were trained separately for images and texts, three strategies can be compared, namely image based

[7]Lebensweltlich orientierte Räume (LOR) - Land Berlin. `https://www.stadtentwicklung.berlin.de/planen/basisdaten_stadtentwicklung/lor/` (Accessed on 31.01.2021)

[8]Geometrien der LOR-Bezirksregionen Berlins, Stand: 07/2012 - Berlin Open Data. `https://daten.berlin.de/datensaetze/geometrien-der-lor-bezirksregionen-berlins-stand-072012` (Accessed on 31.01.2021)

[9]Weather Underground. `https://www.wunderground.com/` (Accessed on 31.01.2021)

filtering, text based filtering and filtering based on both texts and images. Two case studies in Paris and London are given.

For the first case study, correlation analysis is conducted based on the Tweets filtered during 45 days from 17$^{\text{th}}$ of May 2016 to 30$^{\text{th}}$ of June 2016 in Paris. In this time range, a fluvial flood event has happened. 111,500 geotagged Tweets containing both texts and images were collected. After filtering by the text classifier, 2093 Tweets are classified as flood relevant. 6431 Tweets are classified as flood relevant based on user generated photos. With the confirmation from both text and image classifiers, 690 flooding relevant Tweets were extracted. Subsequently, these extracted Tweets were manually checked, 616 of them are correctly classified, thus a precision about 89.3% was achieved.

Since each day may have a different numbers of Tweets in total, ratios between the topic relevant Tweets and total number of Tweets on the same day are calculated for the three strategies. As shown in Figure 6.18, proportions of Tweets filtered by the three strategies are presented and the red solid line indicates the precipitation data in millimeter. Correlations between the results from the three strategies and precipitation were calculated and summarized in Table 6.9. From the results, only a relative small correlation exists between the text based filtering and the precipitation records, and the other two strategies are almost uncorrelated with the precipitation data. A peak can be identified from the VGI data on 3$^{\text{rd}}$ of June 2016, which is exactly the fluvial flood event on 3$^{\text{rd}}$ of June 2016 (BBC, 2016). It should be noted, that there was no rain on that day, as indicated by the very low precipitation value. The peak identified by the VGI filter therefore identifies the peak in the fluvial flood and not in the rainfall.

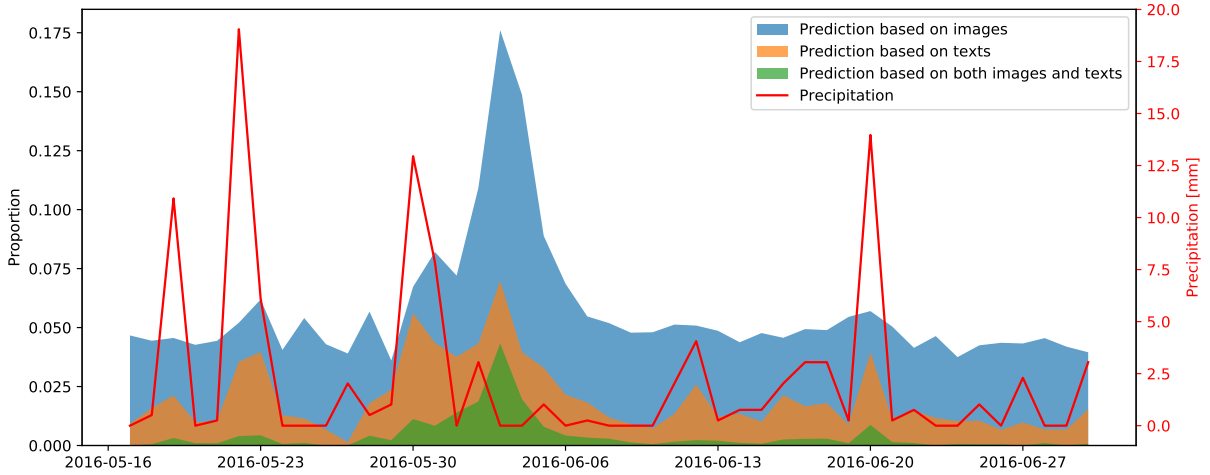


Figure 6.18: Comparison of the retrieval strategies (Paris, 17th of May – 30th of June 2016).

Table 6.9: Correlations between the proportion of topic related Tweets and rainfall intensity (Paris, 17th of May – 30th of June 2016).

Prediction	**Correlation**	***p*-Value**
Prediction based on images	0.0108	0.9439
Prediction based on texts	0.4927	0.0006
Prediction based on both images and texts	0.1063	0.4870

For the second case study, the correlation analysis is conducted based on the Tweets filtered from 17$^{\text{th}}$ of June 2016 to 30$^{\text{th}}$ of June 2016 in London. As shown in Figure 6.19 and Table 6.10, a

much stronger correlation can be identified compared to the previous case. On 23$^{\text{th}}$ of June 2016, a pluvial flood happened in London (BBC, 2016) and the peak on that day can also be identified. In this case, image based filtering has higher correlation than the others, which shows that the filtering by the image classifier is more sensitive to the real rainfall events.

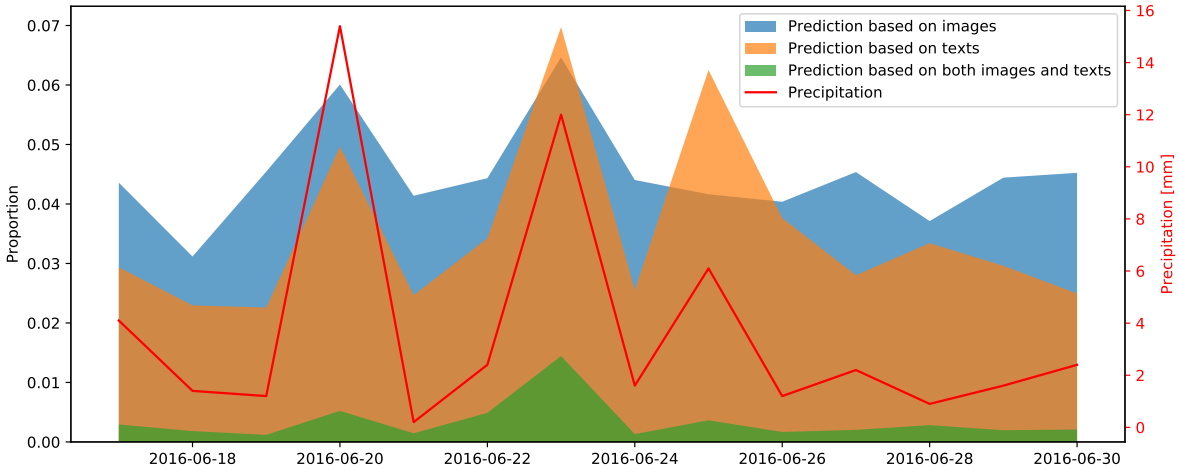


Figure 6.19: Comparison of the retrieval strategies (London, 17th – 30th of June 2016).

Table 6.10: Correlations between the proportion of topic related Tweets and rainfall intensity (London, 17th – 30th of June 2016).

Prediction	**Correlation**	***p*-Value**
Prediction based on images	0.8360	0.0002
Prediction based on texts	0.7685	0.0013
Prediction based on both images and texts	0.7208	0.0036

In summary, from the two case studies above, it can be found that there is a strong correlation with pluvial flooding within a time range, however, when fluvial flooding occurs, the correlation becomes weaker. Instead of using precipitation, river gauges can be considered as a potential data source for calculating such correlations. Therefore, the approach presented in this research is able to detect pluvial flood events, but is not able to distinguish pluvial flood from fluvial and coastal floods.

Furthermore, it was also noticed that pluvial flood events are different from fluvial flood in the sense of spatial distribution of the relevant Tweets. As a matter of fact to be seen clearly in Figure 6.20 right, a fluvial flood event occurs close to a river, therefore most of the relevant information are accumulated near the river. However, for a pluvial flood event (as shown in Figure 6.20, left), the extracted Tweets distribute much evenly in space. In this way, there is also great potential to distinguish different types of flood events from the spatial patterns of the extracted social media posts.

6.3.6 Summary

In summary, this section has described a framework to collect, process and analyze pluvial flood relevant information from the social media platform Twitter. The extraction of relevant information takes not only the textual information into consideration, but also user-generated photos as

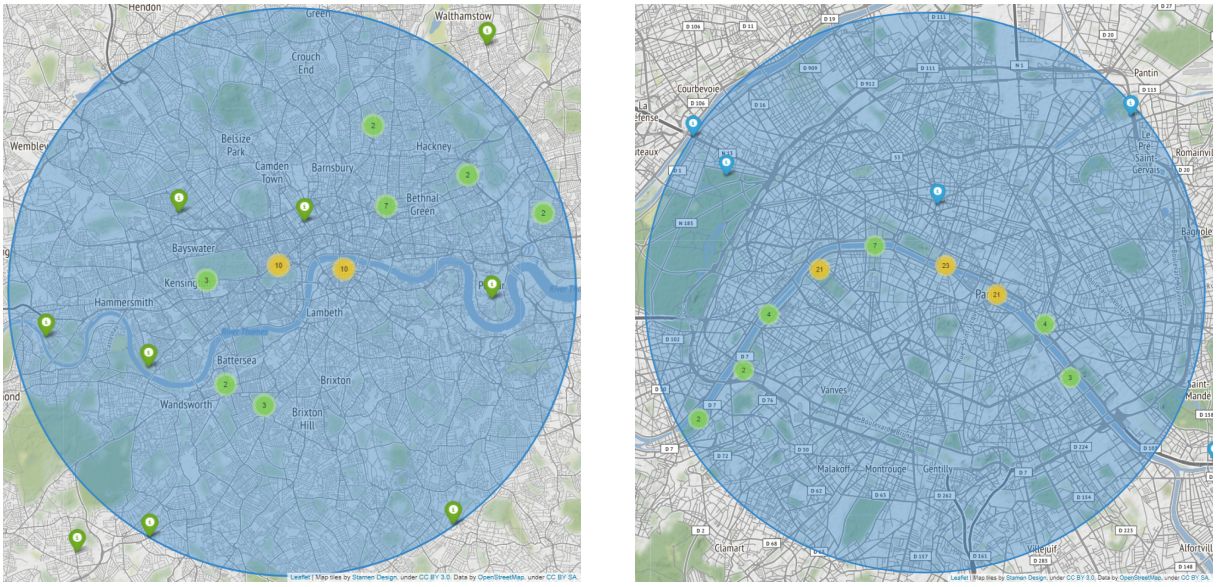


Figure 6.20: Comparison of pluvial flood event (left, London, 23th of June 2016) and fluvial flood event (right, Paris, 3rd of June 2016).

supplements to find high quality eyewitness reports for such events. These individual cues for events are subsequently aggregated using spatiotemporal clustering to extract significant clusters in space and time and ignore the outliers. Finally, a document in the form of a map was generated. It visualizes the high quality topic relevant Tweets, the spatiotemporal clusters and hot spots of the city for each day. In this research, fixed text and image training datasets were evaluated, and real Twitter stream data was filtered. Different filtering strategies are compared with respect to the precipitation data. The case study in London provided evidence that the extracted number of flood and rainfall relevant Tweets are correlated with the precipitation records. The work demonstrated in this research is part of a real-time pluvial flood forecasting system presented in (Rözer et al., 2021).

6.4 Flood severity mapping from VGI by interpreting water level from images containing people

In order to further extract more detailed information from flood-related VGI, this experiment mainly focuses on social media images. In this study, further experiments were conducted to obtain an image retrieval model with better performance. With these retrieved flood-relevant social media images, a water level estimation model is trained and compared with two baseline methods. Lastly, locations of the Tweets are used for generating a map of estimated flood extent and severity. As a proof of concept, this process was applied to an image dataset collected during Hurricane Harvey in 2017.

The overview of the whole proposed workflow is visualized in Figure 6.21. It has three main components, namely (1) retrieval of flood relevant social media posts, (2) duplication detection, and (3) water level estimation from images containing persons. The experiments in this section are also presented in (Feng et al., 2020a).

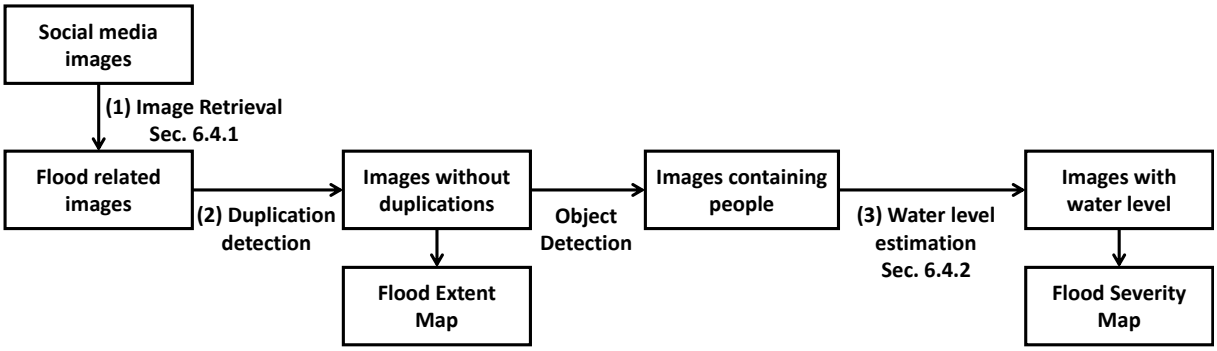


Figure 6.21: Workflow of the process to extract flood extent and flood severity from social media data.

6.4.1 Retrieval of flood relevant social media images

In order to extract flood-relevant VGI from massive social media data, a retrieval step is always essential for all kinds of further applications. The ultimate objective of this study is concerned with the extraction of information on inundations, which can be used later for flood severity mapping. The study in the previous subsection focused on pluvial floods and aimed to obtain social media posts regarding rainfall or flooding. In contrast, the research described in the following could rely on new computer vision strategies for image classification and also on new, publicly available datasets. Therefore, there was also the chance to train a better performing image classifier.

With the method using the ensemble of deep features presented in Section 5.2.2, an image classifier is trained to identify flood-relevant images. Firstly, image classifiers were trained with different combinations of pre-trained models, namely *InceptionV3* only, three models (containing *InceptionV3*, *DenseNet201*, and *InceptionResNetV2*), and four models (three models plus VGG16, pre-trained on Place365). Each combination was trained either with FC layers or Xgboost. The combination using *InceptionV3* for feature extraction and Xgboost for classification is the same strategy as used in the previous experiments of Section 6.3.2. Since it was desired to limit the complexity of the whole framework, only one model was trained in this section to determine if an image is relevant to the flood. Images of lakes and rivers were not specifically considered here, as it has already been demonstrated in the previous approach.

The models were firstly evaluated based on the DIRSM dataset. In order to compare the result with existing work, the same metrics as in those tasks were used. Since a ranking retrieval system is to be built, precision of the top-related documents is more relevant. For this reason, cut-offs were applied on the ranked retrieval results and the precision was calculated. As the number of positive examples is 480, the metrics precision at cut-off 480 (P@480) and average precision at cut-offs 50, 100, 150, 240 and 480 (AP@{50, 100, 150, 240, 480}) were used for evaluation. Two hundred images were randomly selected from each of the positive and negative training examples, and they were used as the validation set. Early stopping with a patience of 6 epochs was applied when the validation loss did not constantly improve. The comparison with previous research using the same dataset is summarized in Table 6.11. From the results, it was concluded that the Xgboost classifier has generally outperformed the models using FC layers. The combination of three models achieves the best results and it also indicates that combining a VGG16 pre-trained on Place365 is not beneficial in this work.

Secondly, in order to adapt to the larger variety of images from Twitter and Instagram, models were trained on the extended DIRSM dataset. From each of the two categories, 800 images were randomly selected as the validation dataset and 1000 images as the test set. The models were

Table 6.11: Evaluation of different approaches on MMSat Task in MediaEval'17 and comparison with this approach.

Methods	P@480	AP@{50,100,150,240,480}
Tkachenko et al. (2017)	50.95	62.75
Zhao and Larson (2017)	51.46	64.70
Lopez-Fuentes et al. (2017)	61.58	66.38
Hanif et al. (2017)	64.88	80.98
Nogueira et al. (2017a)	74.60	87.88
Dao et al. (2018)	77.62	87.87
Avgerinakis et al. (2017)	78.82	92.27
Ahmad et al. (2017a)	84.94	95.11
Bischke et al. (2017a)	86.64	95.71
Ahmad et al. (2017b)	86.81	95.73
this approach		
FC - InceptionV3	82.92	93.57
FC - 3 models	87.08	97.25
FC - 4 models	85.00	93.17
Xgboost - InceptionV3	86.46	96.96
Xgboost - 3 models	**89.17**	**97.53**
Xgboost - 4 models	88.75	97.37

trained on the rest of the images. Since the combination of three models demonstrated the best performance, this strategy was used on both the DIRSM training set and extended DIRSM training set and then evaluated on the DIRSM test set and extended DIRSM test set.

Table 6.12: Evaluation of model performance based on precision, recall and F_1-scores on positive class, Overall Accuracy (OA) and Area Under Curve (AUC).

Trainset	Method	DIRSM test set					Ext. DIRSM test set				
		Prec.	Rec.	F_1	OA	AUC	Prec.	Rec.	F_1	OA	AUC
DIRSM	FC	91.44	82.29	86.62	90.76	0.967	95.53	70.50	81.13	83.60	0.950
DIRSM	Xgboost	89.31	88.75	**89.03**	**92.05**	**0.972**	98.55	74.80	**85.05**	**86.85**	**0.976**
ext. DIRSM	FC	90.68	81.04	85.59	90.08	0.964	94.22	86.40	90.14	90.55	0.972
ext. DIRSM	Xgboost	85.35	91.04	**88.10**	**91.06**	**0.971**	92.51	92.60	**92.55**	**92.55**	**0.982**

The purpose of this work differs slightly from a ranked retrieval system, as in the *MMSat* task, since it aims at rejecting off-topic posts efficiently. In this case, the false negative error plays a role. Thus, for the evaluation on the extended DIRSM dataset, the metrics such as precision, recall, F_1-score, on the positive class, Overall Accuracy (OA) and Area Under Curve (AUC), were used. The performance of the models is summarized in Table 6.12. The ROC curves of the trained models are compared in Figure 6.22.

From the evaluation on both test datasets, the Overall Accuracy and AUC of the Xgboost models are significantly higher. For the DIRSM test set, the benefits of introducing more annotated images are not obvious, however, both metrics are significantly improved on the extended DIRSM test set. This means that introducing more annotated images makes the classifier more adaptive to the images coming from Twitter or Instagram.

In summary, a flood image classifier was trained for social media image classification with state-of-the-art performance, which can filter out most of the off-topic images with an accuracy of 92.55%.

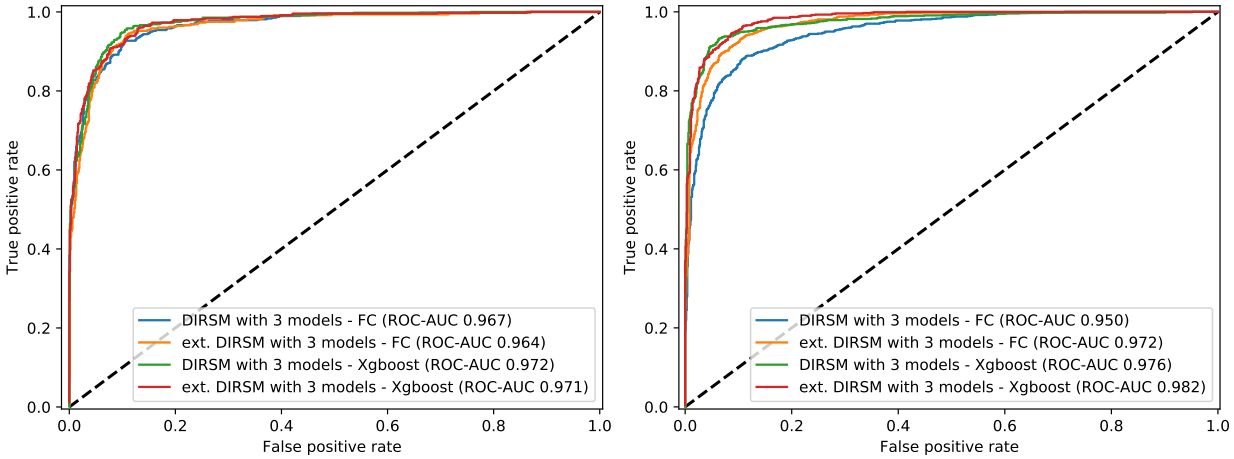


Figure 6.22: Evaluation of models on DIRSM test set (left) and extended DIRSM test set (right).

6.4.2 Experiment and evaluation of water level estimation

After filtering out off-topic images and removing duplicates, the third component is to estimate the water level, based on objects with known dimensions standing partly in the water. In this work, people were selected as the targets, because – according to the observation – they are the most common objects in social media image datasets. In the following, the proposed method presented in Section 5.3.1 is evaluated by comparing it with two baseline methods.

The proposed model and the two baselines, were trained on the same dataset (as presented in Section 6.2.4), where 20% of the data were used for validation and the rest for training. During the experiments, it was observed that many of the wrong predictions were due to the very small size of people at greater distances. Therefore, it is required that the number of pixels of the detected people segments must be larger than 0.1% of the total pixel number of the whole image. The important parameters used for training the models are listed in Table 6.13.

Table 6.13: Parameters for all methods.

Method	Parameters
Ours	**Xgboost** {max-depth:2, eta:0.3, objective:multi-softmax, silent:1, num-class:5, num-round:300, early-stopping-rounds:20}
Baseline 1	**Xgboost** {max-depth:2, eta:0.3, objective:multi-softmax, silent:1, num-class:5, num-round:300, early-stopping-rounds:20}
Baseline 2	**Mask R-CNN** {batch-size:1, max num epochs:80, steps per epoch:300, learning rate:0.00005, early-stopping patience: 10}

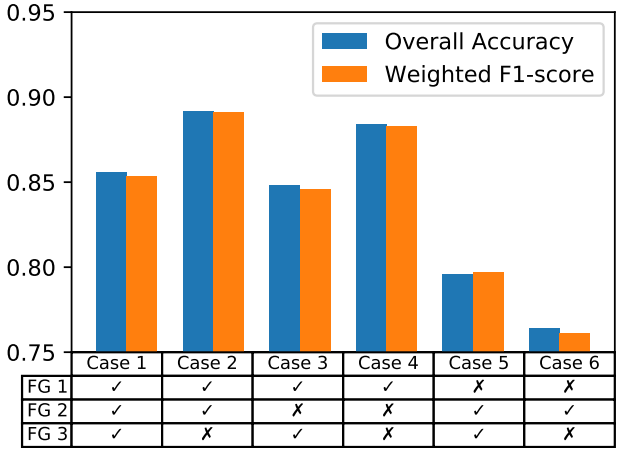


Figure 6.23: Evaluation of different combinations of feature groups performed on test set.

Selection of features All the combinations of the three feature groups (FG) (see Table 5.3 in Section 5.3.1) were firstly analyzed for the proposed method as shown in Figure 6.23. The model was trained with different feature groups separately. The overall accuracy and weighted average F_1-score on the test set were used as the performance measure. It is identified that FG 1 (distances of keypoints to water line) plays an important role in the classification, and a significant performance drop can be observed when FG 1 is excluded (see cases 5 and 6). For all cases using FG 1, a performance of over 85% has been achieved. For most of the cases, including FG 3 (binary label indicating whether the connecting area is water or ground) is less beneficial. Combining FG 2 (OpenPose confidence scores) can slightly improve the performance. Lastly, it can be observed that the combination of FG 1 and FG 2 achieves the best results. Therefore, this strategy (i.e., case 2) was used to train the model and compared with these two baseline methods described in Section 5.3.2 and 5.3.3.

Qualitative evaluation Some qualitative evaluations are shown in Figure 6.24, where five example images are presented with different water levels. The example images were collected from the Flickr album "Flood - Thailand" (ebvImages, 2011), published under CC BY-NC-SA 2.0 license. These images were kept unseen during the training of the models. The ground truth (GT) and predictions for each image from this model together with the baselines are given. From the results, it can be observed that this model can ignore the majority of the persons showing no evidence to water level. Based on the bounding box, the features can present the proportion of visible and non-visible body parts. In baseline 2, the water level estimations contain many wrong predictions, especially for the people showing no evidence to water level. Baseline 1 predicts a knee level flooding more frequently, and also cannot distinguish images showing no evidence of water level properly.

Additionally, some failed cases of this approach are presented in Figure 6.25. In general, they are three common situations. On the left image, the segmentation network cannot provide a reliable prediction as the boat pixels are mostly predicted as water in this image. Therefore, these people have been classified as standing in the water to the hip (C) or chest (D) level. Sitting people in the water can also hardly provide reliable evidence for water level estimation. As the example shown in the middle, the three sitting people on the left hand side cannot be rejected properly. It leads to a wrong prediction of this image. The third failure case on the right is caused by water reflection, where both the object detection and body keypoints estimation failed.



Figure 6.24: Qualitative evaluation of the proposed approach compared with the baselines (example images under CC BY-NC-SA 2.0).



Figure 6.25: Example failure cases of this approach, caused by segmentation failure - left, sitting people - middle, and water reflection - right (example images under CC BY-NC-SA 2.0).

Quantitative evaluation The results of quantitative analysis is presented in Table 6.14 and Figure 6.26, where the confusion matrix, overall accuracy, and weighted average F_1-score are given for the best model from different experimental settings. Since this is a classification task with five categories, to present the overall performance, the F_1-scores were averaged according to the proportion of the number of examples per category, i.e., 50 images per category in this test set. Analyzing the results reveals that the proposed method achieves the best performance, compared to the two baselines. According to the confusion matrices, more examples are located at the diagonal of the matrix. It achieves over 89% accuracy and weighted average F_1-score on the test set of 250 images. Baseline 1 has in general difficulties distinguishing neighbouring water levels. Baseline 2 can be improved by introducing the features from the area beneath the detected box, however, it is still not as good as the proposed method. There are many images which were assigned with water level labels, even though there is no evidence for flooding. In summary, the proposed method is a suitable solution for water level estimation and can be used for flood severity mapping.

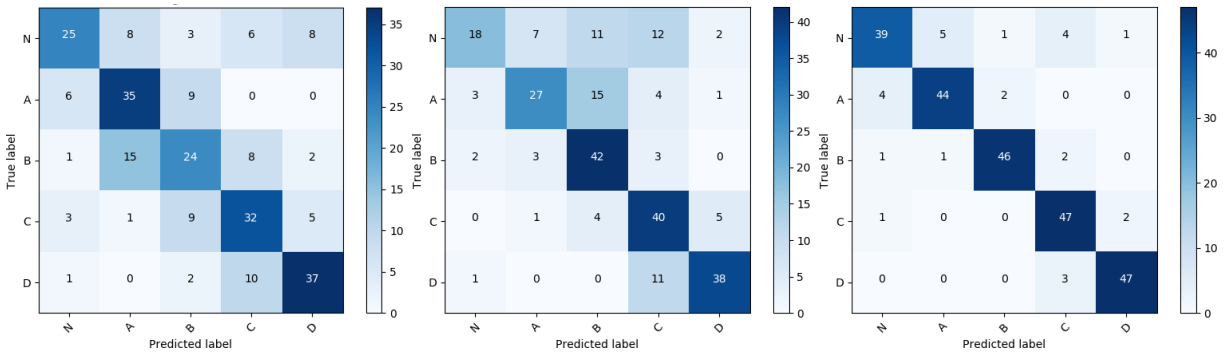


Figure 6.26: Comparison of confusion matrices on the test set using baseline 1 (left), baseline 2 using 1/4 area beneath (middle) and the proposed method (right).

Table 6.14: Quantitative comparison of models for water level estimation.

Method	Overall Accuracy	Weighted Avg. F_1
Baseline 1 - global deep features	61.20%	60.95%
Baseline 2 - adapted Mask R-CNN with no area beneath	57.60%	52.94%
Baseline 2 - adapted Mask R-CNN with 1/4 area beneath	66.00%	64.94%
Ours - handcrafted distance features	89.20%	89.14%
Our model fused with baseline 1	**90.00%**	**90.01%**

Furthermore, it is noticed that combining global and local information is a common strategy to optimize model performance. Therefore, a decision fusion is introduced which fuses the softmax outputs from our model using hand-crafted features and baseline 1 using global deep features. The proposed model makes the final decision based on voting, thus the person predicted as the voted result with the highest confidence score according to softmax outputs is selected for fusion. Both the softmax outputs from the two models are linearly combined with weights. By empirically setting the same weights for our model and baseline 1, this combination has achieved a slightly higher model accuracy and weighted average F_1-score of 90% on the test set.

6.4.3 Flood severity mapping for Hurricane Harvey in 2017

In order to show the benefits of the proposed processing pipeline for flood severity mapping, it was applied to a severe flood event caused by Hurricane Harvey in 2017. Many studies have been conducted by researchers and national agencies in the last few years, which can provide additional information for comparison and discussion.

As introduced in Section 6.1.2, from 25$^{\text{th}}$ of August to the 1$^{\text{st}}$ of September 2017, a total of 150,227 Tweets with either geo-coordinates or location information were collected in the Houston area. 28,833 of them contained URLs for photos; the photos were, however, not downloaded at that time. After deleting duplicate messages based on identical texts, 20,399 unique Tweets were retrieved. Two years later, on 13$^{\text{th}}$ of June 2019, 20,824 valid images were downloaded for further image analysis. In the following, the application of the proposed process is presented, followed by the visualization of three mapping possibilities presenting the extracted information.

6.4.3.1 Processing of social media images

Social media users may share images copied or duplicated from others. As such images often demonstrate severe flood situations and seemingly appear at multiple locations in a city, they can significantly mislead the mapping results. Therefore, the detection of duplicate images is an essential step before flood mapping. Thus, the processing of social media images has the following three steps in this application, (1) retrieve the flood relevant images, (2) remove duplicates of the images predicted as relevant, and (3) estimate the flood severity from the image collections.

Social media filtering The binary classifier as trained in Section 6.4.1 was applied on all downloaded images to retrieve the ones relevant to flood events. Since the model can provide an output with confidence score, the images were categorized into eight predefined groups with the thresholds 99%, 95%, 80%, 50%, 20%, 5% and 1%, as visualized in Figure 6.27. As shown in the bar diagram, 13,658 (65.6%) of the collected images are surely irrelevant to the flood event, while 3,142 (15.1%) are relevant; uncertainty exists for the remaining 19.3% of the images.

Duplication detection with deep features As described in Section 5.2.3 duplicated images are identified by checking high similarities in their features derived from a pre-trained DCNN model. The duplication detection was applied to the flood relevant images, where a 50% threshold was applied to the confidence score of the outputs. A total of 4,601 images were used for duplication detection. Feature vectors were first generated with a pre-trained ResNet18 and then clustered using DBSCAN. DBSCAN requires two parameters, *eps* and *minPts*, which represent the distance between the features in a cluster, and the required minimum number of elements in a cluster. *minPts* in this case is 2 because the aim is to include also duplicated image pairs. A suitable *eps* can be determined by a k-Nearest Neighbor Graph. As described in (Sander et al., 1998), by analyzing the sorted k-distances, good values are in the "valley". Different from most other applications of

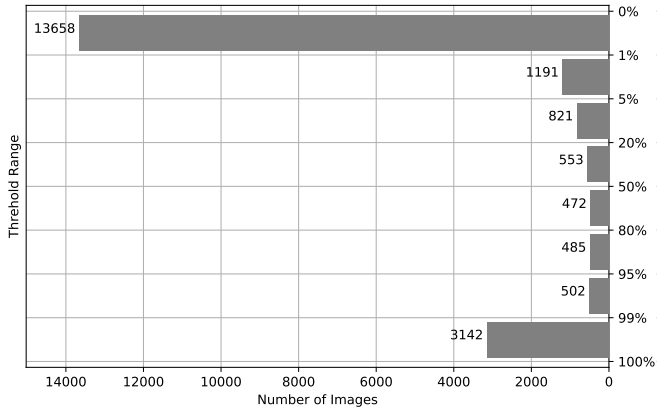


Figure 6.27: Distribution of the model predicted flood relevance scores for the images collected during Hurricane Harvey.

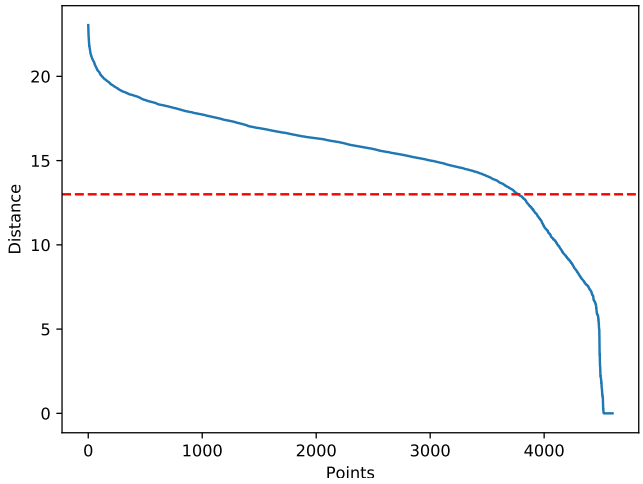


Figure 6.28: Sorted 2-distance plot for image deep features.

DBSCAN, in this case, the majority of images is considered as "noise" for DBSCAN, where the clusters of duplicate images are the minority. Then, the second significant turning point at 13 was selected from the graphical representation shown in Figure 6.28. Among the retrieved 4,601 images, 207 clusters were identified. The clusters were checked manually and only three clusters were found to contain non-duplicate images, whereas all the remaining clusters did represent duplicate and near-duplicate images. To select the most relevant image for a cluster, the earliest posted image was preserved and all later ones were deleted. In total, 653 duplicate images were eliminated in this step, and finally 3,948 images were available for further processing. Images from the largest cluster are shown in Figure 6.29, which cover different duplication cases, such as clipping, changing colour, and adding text.

Water level estimation The resulting flood relevant images were further processed with the water level estimation model as described in Section 5.3. In this case, only images highly relevant to the flooding are considered, i.e., with a confidence score over 99%. After applying all the above described filtering processes, 676 flood-related images remained for the water level estimation. In order to evaluate the performance of the proposed model for this real event, the images were annotated based on the annotation rules described in Section 6.2.4 which lead to the confusion matrix shown in Figure 6.30. The overall accuracy of the proposed model is 76.18% with a weighted average F_1-score 77.18%. The number of false positives and false negatives between the four water



Figure 6.29: Examples of the duplicate images from the largest cluster of DBSCAN result.

level classes are relatively small. However, there are many images, which are supposed to show no evidence for water level estimation (i.e., class N), classified with a water level class. Comparing these results with the ones in Section 6.4.2 (90%), the reduction of performance may be due to two aspects. One is the image quality. Compared to the images collected for training this model, social media images from Twitter and Instagram are often of poorer quality. Users may overlay texts on photos, make collage from several photos, and apply image filters changing brightness and color. Many photos are also resized, compressed, or cropped by the users. The other aspect is that the training examples of class N can cover only a small fraction of the cases encountered in reality. Especially, people in the scenarios with other postures than standing have a higher chance to be wrongly predicted by the classifier (e.g., sitting as shown in Figure 6.25).

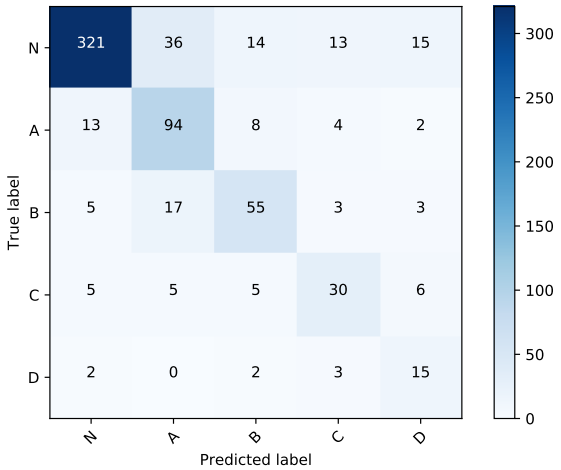


Figure 6.30: Confusion matrix of the water level estimation on social media images with flood relevance over 99%.

6.4.3.2 Flood mapping from VGI

After estimating water level from social media images containing people, the next step is to link these estimates to the locations on the map, with the goal of providing a map of the flood extent and a map of the flood severity. In order to evaluate the results, it is essential to have ground

truth to compare. This is difficult, as an exact ground truth comparable to VGI is not available. The following datasets have been selected as reference: There is a dataset with property claims from the U.S. Federal Emergency Management Administration (FEMA, 2018a). An additional dataset - Harvey flood depths grid dataset - contains modeled inundation from FEMA (2018b). Furthermore, there is a map with flood extent marked by remote sensing detection from the Dartmouth Flood Observatory (DFO, 2017).

In the following paragraphs, the mapping possibilities are presented with the extracted information in three aspects. Firstly, the individual severity estimations together with the associated text and image are visualized as markers with pop-ups. Secondly, flood extent was determined from VGI by aggregating the locations of flood related posts. Lastly, flood severity was determined from VGI by aggregating water level estimates.

Map of individual severity estimation The locations of Tweets are generally given in three types, see also Section 2.5.2. Type 1 – Tweets provide exact geo-coordinates, which is a rare case, covering only 3.29% of the total amount of the data collected for this research. Type 2, 33.72% of the retrieved Tweets, provide the location information corresponding to an area, where a bounding box is normally given. Type 3 (62.99%) are retrieved Tweets that are shared Instagram posts, for which both geo-coordinates and bounding boxes are available as the examples demonstrated in Figure 2.17. However, the saved geo-coordinates may represent either a point location (POI) or an administrative area such as city and district. The recorded bounding box normally represents the corresponding city-level bounding-box.



Figure 6.31: Map of social media posts with severity predictions as markers (Basemap: OpenStreetMap).

The most straightforward way to present the extracted information on a web map is using markers with symbols representing the flood severity situation. In dense areas, markers are clustered. As

shown in Figure 6.31, users can click into the cluster to inspect individual Tweets on the web map. For Types 1 and 3, the Tweets were located to the given coordinates. However, city-level Tweets do not provide much information about where the observations were taken. Therefore, the Tweets with city level geo-coordinates were excluded, such as for the City of Houston, or Harris County. For Type 2, where the Tweets have only bounding boxes, the Tweets were positioned at the locations of the box centres.

This visualization can provide a straightforward overview on the spatial distribution of individual flood level related Tweets. This map can provide detailed information about the flood severity at individual locations, together with an exact image. However, it does not provide an integrated overview of the flood extent and corresponding severity. Thus, in the following, maps for flood extent and severity are developed. These maps are then compared with existing maps provided by authorities after the disaster.

Map of flood extent In order to get an overview of the flood situation, the point information given with the Tweets have to be extended to areal information, typically using spatial interpolation methods. However, in this case, the social media posts are very sparsely and unevenly distributed in space. The main factor for inundation - terrain - varies from regions to regions significantly. Thus, interpolation can hardly reflect the real situation between observations. Additionally, the locations of the Tweets may refer to either a point location or a bounding box. Therefore, instead of interpolation, aggregation of the information to spatial units is a more reasonable representation for the flood situation.

In the United States, the most commonly used spatial units in geography are census tracts. They are relatively permanent statistical subdivisions of a county, which have on average about 4,000 inhabitants (U.S. Census Bureau, 2015). Boundary files for Texas were downloaded from the U.S. Census Bureau (2018) and tracts around Harris County were extracted. They covered most of the Houston metropolitan area. This area contains 966 census tracts in total.

Due to the uncertainty of the location of the posts, not just census tracts in which posts coordinates lies were considered, but the following strategy was applied. As investigated by Cvetojevic et al. (2016), the typical distances between the image content and photo upload location have a median value of 198.7m for Twitter in North America and the Caribbean, and 85m for Instagram posts. Thus, all tracts lying with a buffer of 200m around post coordinates were marked. In Figure 6.32, the census tracts where Tweets were sent are marked with light grey colour, and the census tracts where flood relevant Tweets were sent with dark grey colour. For the Tweets with only bounding boxes (type 2), all the intersected tracts were marked. The area (with holes) marked with a red boundary in Figure 6.32 is the flood extent estimated by VGI.

Remote sensing has been widely applied for flood extent mapping and is used in this research as a baseline. The remote sensing detection was created by the Dartmouth Flood Observatory. They extracted the maximum observed flooding for Hurricane Harvey from NASA MODIS, ESA Sentinel 1, ASI Cosmo SkyMed, and Radarsat 2 data (DFO, 2017), shown as blue pixels (of size 85 m $\times$ 85 m) in Figure 6.33. Pixels were aggregated to census tracts by overlay and marked in gray. It can be observed that there is less flooding observed in the city center, whereas more flood pixels are detected outside the city or along the river in the city.

As an additional data source, the property claims for hurricane Harvey from FEMA (2018a) was used as a reference. It contains property claims with dates, loss types (e.g., electric current, wind, flood, water damage), and locations (both in text and coordinate). From 27$^{\text{th}}$ of August to 2$^{\text{nd}}$ of September 2017, in total, 226,167 property claims were collected in Texas and 38,422 of them are caused by flood or water damage. Even though this data were collected for insurance purpose,

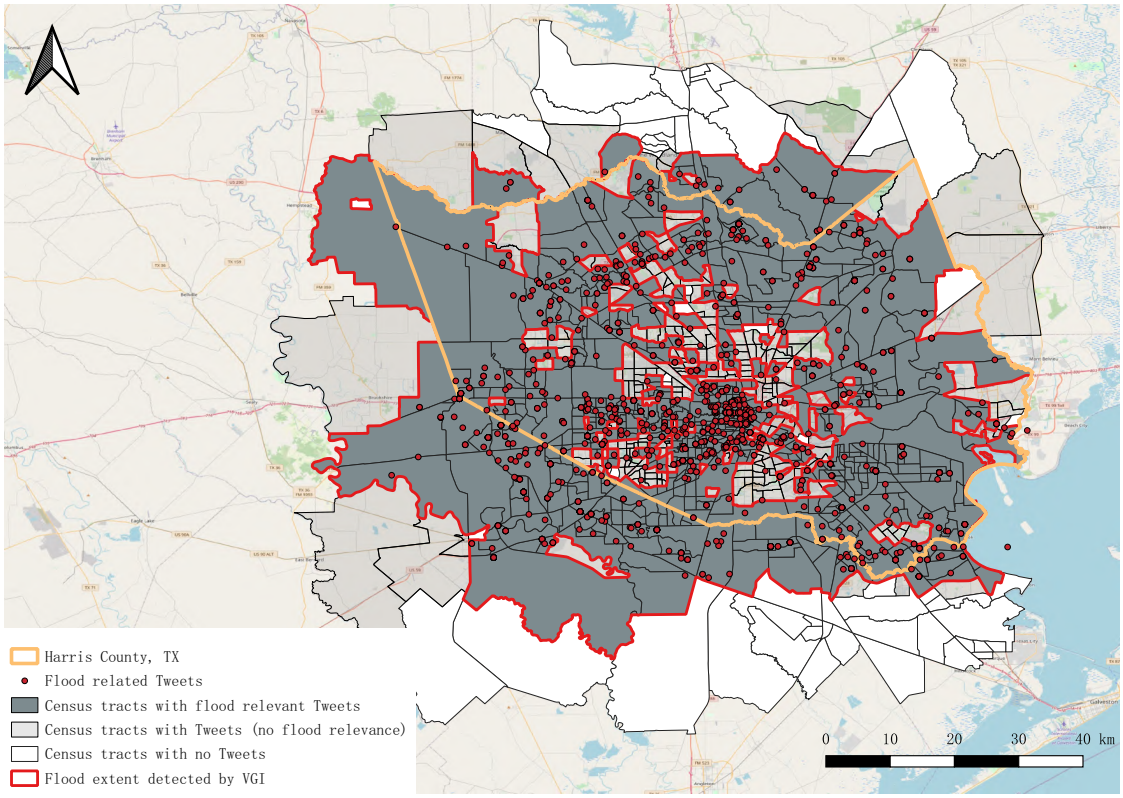


Figure 6.32: Locations of flood relevant Tweets with overlaid census tracts as the flood extent detected by VGI.

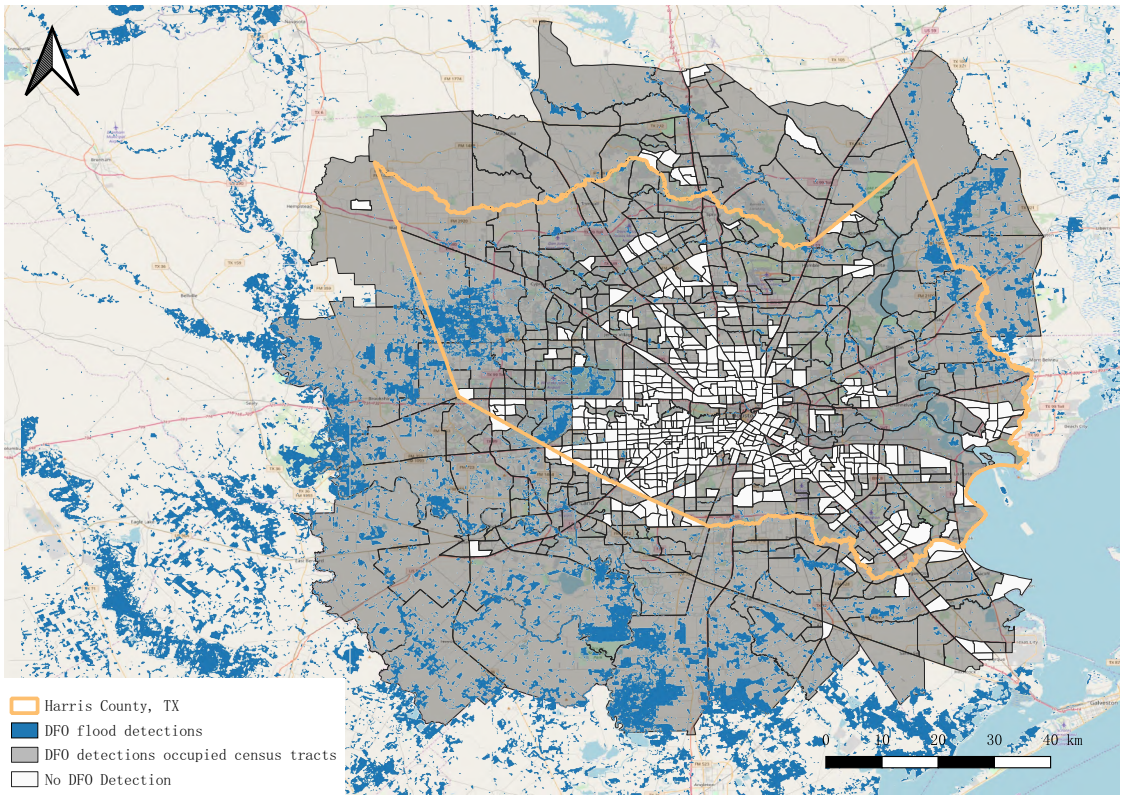


Figure 6.33: Maximum observed flooding mapped from NASA MODIS, ESA Sentinel 1, ASI COSMO-SkyMed, and RADARSAT 2 data from Dartmouth Flood Observatory (DFO, 2017) and the overlaid census tracts.

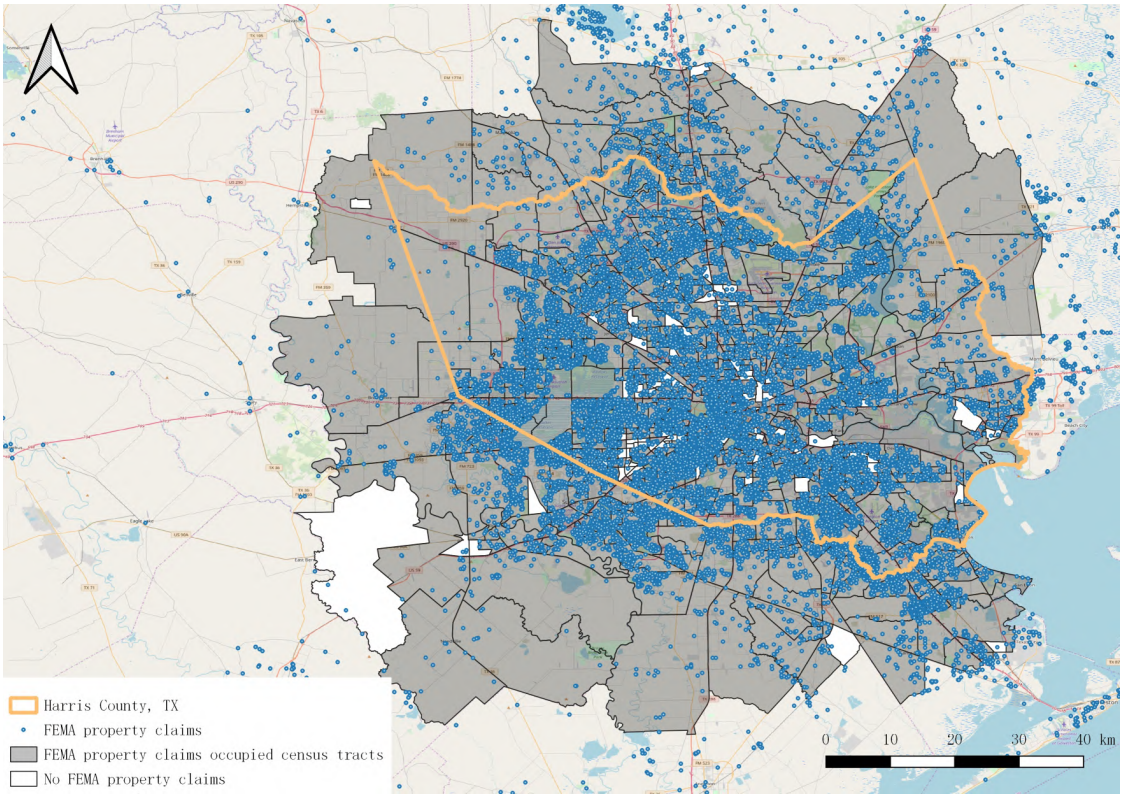


Figure 6.34: FEMA property claims and the overlaid census tracts. Data source: FEMA (2018a).

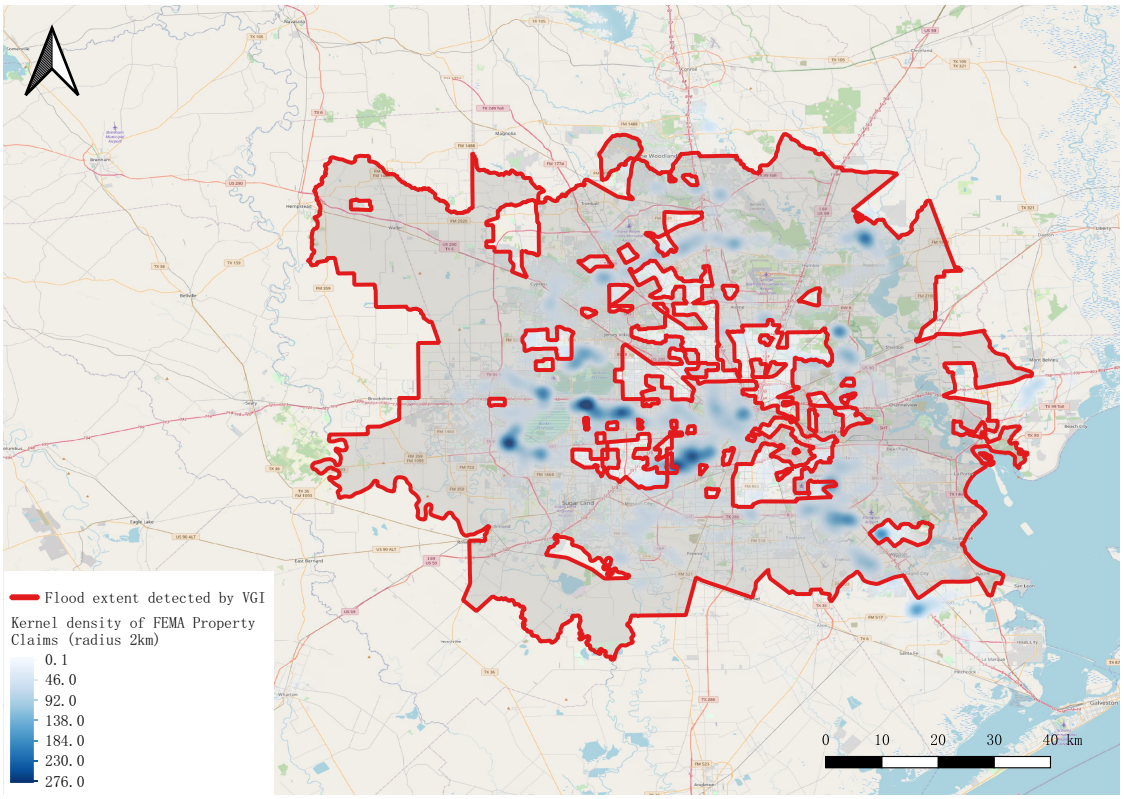


Figure 6.35: FEMA property claims density map and the flood extent detected by VGI. Data source: FEMA (2018a).

Table 6.15: Comparison of water extent mapping from different information sources.

Method	Precision	Recall	F_1-score	Accuracy
VGI	96.54%	64.68%	77.46%	64.70%
RS	**97.10%**	51.77%	67.53%	53.31%
VGI+RS	96.35%	**81.68%**	**88.41%**	**79.92%**

Table 6.16: Confusion matrices of water extent mapping from different information sources.

Method		VGI		RS		VGI+RS	
Predicted		0	1	0	1	0	1
True	0	39	21	46	14	32	28
Labels	1	320	586	437	469	166	740

the data may still contain errors as mentioned in (FEMA, 2019). In addition, the latitudes and longitudes in this data were truncated to one decimal point (i.e., around 10 km) to protect privacy. In order to aggregate the points to census tracts considering this probable uncertainty, only the tracts with 3 or more claims were considered as flooded regions. This threshold was chosen based on the distribution of the number of claims in each cell. Since this distribution is skewed, log transformation was applied to the data to achieve a distribution similar to a normal distribution. Afterwards, a confidence interval of 2-sigma was applied to this transformed distribution. The cells with a transformed score smaller than this interval were regarded as outliers, which correspond to the cells with 1 or 2 claims in the original distribution. Hurricane Harvey was a great disaster which led to huge losses, thus most of the tracts contain property claims caused by flood or water, which is shown in Figure 6.34.

The FEMA property claims were then used as ground truth and compared to the flooded tracts detected from remote sensing and VGI. The two results were combined by a logical OR operation. The precision, recall, F_1-score of the positive class and overall accuracy at census tract level are summarized in Table 6.15 and the confusion matrices are summarized in Table 6.16.

According to Table 6.15, remote sensing detection achieved the best precision but also a low recall. Based on a visual comparison between the remote sensing detection (Figure 6.33) and the reference (Figure 6.34), many false negatives are located in the city centre. Even though VGI provided only very sparse spatially distributed data points, it was able to mark the flooded census tracts with only a slightly lower precision but a higher recall compared to the remote sensing detection. Based on a visual comparison between the VGI based detection and the reference, more census tracts in the city centre are correctly detected. However, due to the lack of observations in census tracts where no Tweets are available (the white tracts shown in Figure 6.32), there are still many false negatives which lead to a low recall of 64.7%. Simply combining the VGI and remote sensing detection achieves a much better overall accuracy and F_1-score, which shows the complementary properties of VGI. With this, it is demonstrated that VGI can be used as a supplement data source for flood extent mapping, especially beneficial for urban areas. The absence of remote sensing detection in urban areas is often due to occlusions caused by different viewing angles, and shadows from buildings and trees.

To regionalize the information, kernel density estimation with a radius of 2 km was applied on the FEMA property claims and overlaid with the area VGI marked as flood extent in Figure 6.35. It can be identified that almost all the “heat regions” are located within the red border of the flood

extent marked by VGI, especially the two heat regions in the west and southwest of the city center. The flood extent from VGI excludes the census tracts where there are no significant heat regions in the city's northeast, northwest and southeast. Although users did send Tweets in most of these areas, but no images related to the flood appeared.

Map of flood severity The Harvey flood depth grid dataset was used as the reference to evaluate the performance of flood severity mapping. It has a 3 m resolution and was published by FEMA on 15th of November 2017 (FEMA, 2018b). It was generated based on High Water Marks from on-site follow-up field surveys and Digital Terrain Models in the form of a Triangulated Irregular Network (TIN). Four quality assurance measures (namely identifying dips, spikes, duplication, and inaccurate/unrealistic measurements) were applied. In addition, water areas (e.g., lakes and rivers) were removed based on authoritative data (U.S. Census Bureau, 2019). The flood depth data in the study area are visualized in Figure 6.36. Since the severity estimation from VGI is at census tract level, the water depths were aggregated to census tracts by calculating the maximum flood depth to represent the most severe situation of each census tract (shown in Figure 6.37).

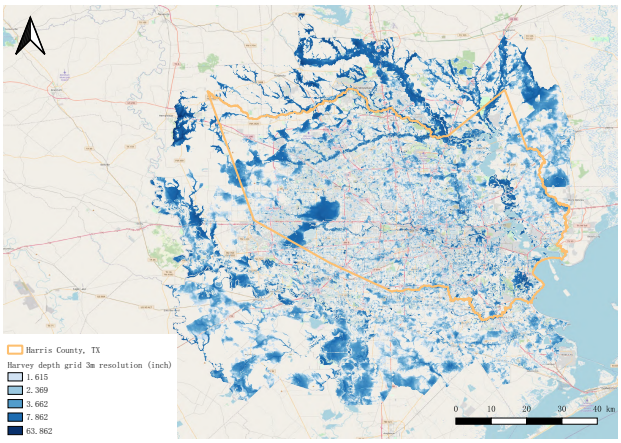


Figure 6.36: FEMA Harvey flood depth grid (FEMA, 2018b).

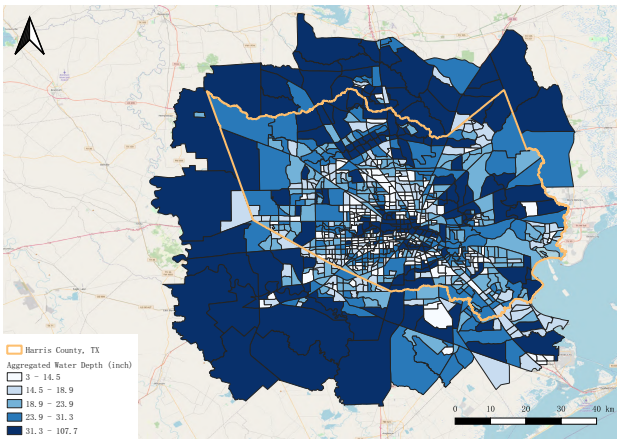


Figure 6.37: Aggregated flood depth tracts with max depth values.

Among the 966 census tracts observed in this study, 323 could provide flood severity estimations based on the interpretation of social media images. Flood severity estimations were aggregated to tracts (shown in Figure 6.38) according to the most frequent flood severity class. Subsequently, the correlation between VGI estimated flood severity and water depth from FEMA was calculated to evaluate the performance of the VGI based flood severity mapping. Since the VGI based flood severity estimations are ordinal and skewed while the modeled water depths are continuous and skewed, Spearman's rank correlation is an appropriate correlation coefficient to use, according to Mukaka (2012). The result ($r = 0.1836$, $n = 323$, $p < 0.001$) indicates a weak positive monotonic correlation between these two variables. This is based on the interpretation for positive correlation (weak: $r > 0.1$, moderate: $r > 0.4$, strong: $r > 0.7$ and perfect: $r = 1$) in (Akoglu, 2018). It is also statistically significant because of the p-value < 0.05.

Due to the sparse and uneven distribution of VGI, the number of the VGI data points available in each census tract is sometimes very limited. 140 out 323 tracts have only 1 or 2 valid images for severity mapping. Nevertheless, this real-time flood severity map can already provide an integrated overview of flood severity. It can be overlaid with the cluster maps presented in Figure 6.31, which allows to inspect the individual observations in detail. It is also worth noting that, even though this information is few and sparse, it is normally available well in advance of other observations, such as remote sensing or field surveys, which is valuable during the emergency response phase.

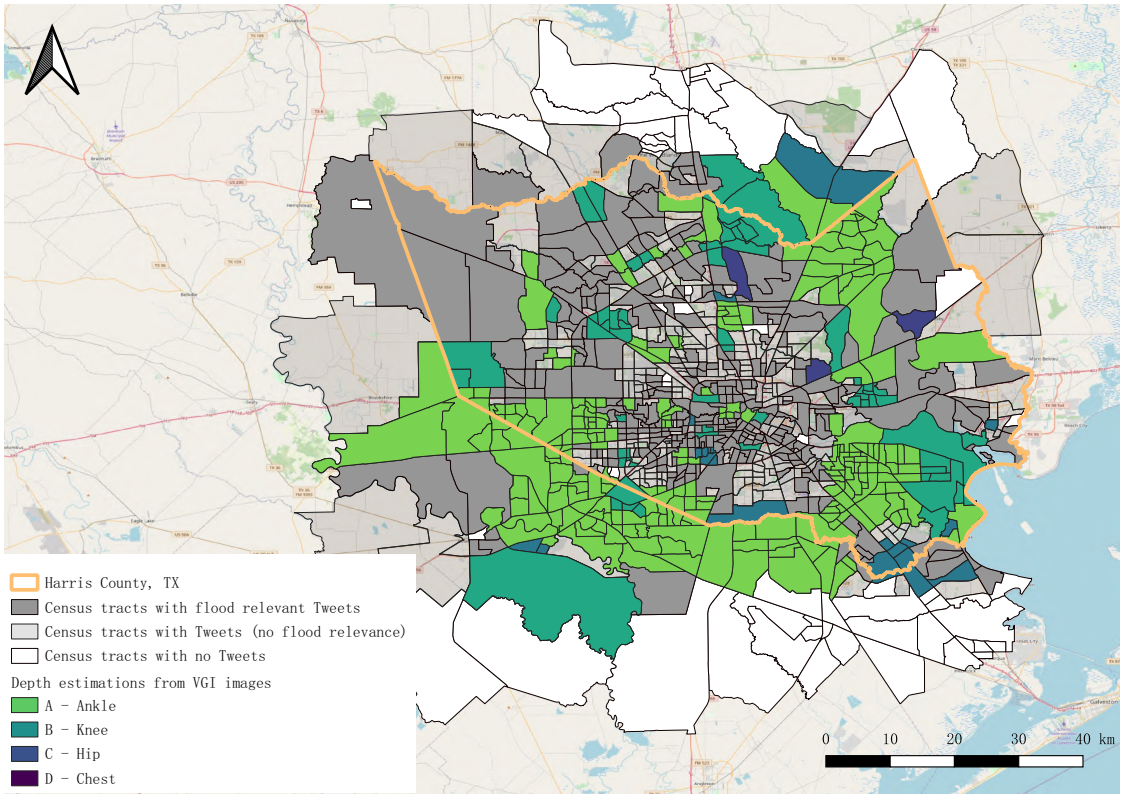


Figure 6.38: Flood severity map derived from the water level estimations of VGI images.

6.4.4 Summary

In this research, a novel process is proposed to map the flood severity from social media images and apply it to a real flood event as a proof of concept. The process includes the collection and filtering of social media images with respect to flood relevant pictures, as well as elimination of similar (and thus potentially duplicated) images. Furthermore, the flood relevant images containing people were classified into four flood severity levels according to the water level with respect to different body parts of people present in the scene. The water level estimation on a representative dataset achieved an accuracy of 90%. Compared with previous studies, the proposed model achieved fine-grained water level classification with less annotation effort.

The trained model was then applied to a social media image dataset collected during Hurricane Harvey in 2017. Flood extent was estimated based on this information, which correctly marked over 62% of the regions where people have claimed flood or water damage. Flood severity was mapped and compared with the modeled flood depth grid. The result indicates a weak positive monotonic correlation to the reference data. In addition, the severity map can serve as a flood severity information which is available well ahead of remote sensing detection.

7 Discussion

This thesis presented two aspects of studies using opportunistic VGI to observe flood and precipitation events.

Speed variation of road users as a precipitation indicator A new precipitation indicator was explored by learning vehicles' speed variation and using it as indicator for precipitation. In this proof-of-concept, vehicle speed detector data collected by a traffic department were used. The experiments showed promising results using the road speed data to learn a precipitation indicator. As summarized in Section 2.5.1, there are several opportunistic VGI sources that can also be used to provide road speed observations, e.g., vehicle trajectories from Floating Car Data (FCD). Compared to road speed data from the traffic department covering only a limited number of roads, vehicle trajectories provide road speed and volume estimations for many more roads and larger areas. With the same principle as presented in this thesis, the precipitation indicator can be trained with many more features. The current study was only able to distinguish precipitation events in the temporal dimension. Exploiting Floating Car Data, on the other hand, may provide information for a spatial temporal differentiation.

Social media as an opportunistic VGI data source for flood observations Two studies were conducted. Each study proposed a framework for processing social media VGI, but each study had its own focus. The first study set up a framework to collect, process, and analyze pluvial flood relevant information from the social media platforms: Twitter and Instagram. Deep learning models were trained separately for texts and images to extract high-quality observations for flood events. A late fusion was used to combine these two information sources, which allows for the extraction of flood-related posts even when one type of information is missing.

The framework was applied for several real-world flood events in western Europe in 2016 and 2017. With spatiotemporal clustering, significant clusters in space and time were extracted. Hotspots were detected. The social media posts were aggregated for each administrative region, and normalized by the number of posts that users typically make within the same region. This avoids constant hotspots in populated areas. An interactive map visualization is provided with high-quality pluvial flood-relevant Tweets, spatiotemporal clusters, and city hot spots. This framework allows to observe multiple flood events easily. Further comparisons were made between pluvial flood and fluvial (river) flood in relation to the precipitation data, and the differences in spatial distribution were revealed. The proposed framework identified areas of high interest and concern to social media users, but these areas often do not necessarily correspond to areas where flood events are more severe. Real-time information about the severity of a flood hazard is also a very important information for the city's emergency management. This prompted the second study.

The second study focused only on social media images and introduced a method to estimate the flood level. It designed and bundled different methods to collect, retrieve, and analyze social media images of flood events. The different elements in the process successfully extracted flood-relevant information, removed duplicates, and classified the water level. Flood severity information was extracted by analyzing these user-uploaded images. People in the scene are the targets, where component level information (i.e., human pose) was used to support the water level classification, which achieved a very high accuracy and weighted average F_1-scores of over 89%. Compared with the baselines using deep features of the whole image and deep features around the detected people,

the proposed method achieved better performance. Thus, the component-level information has proved to be beneficial. Furthermore, the annotation effort is significantly reduced, where each photo is annotated with only a single water level label instead of a time-consuming pixel-level annotation.

This technical process was applied to a real event, Hurricane Harvey in 2017. The high weighted average F_1-scores of the benchmark experiments could not be achieved (77.18% as opposed to 90.01%) – this is attributed to the fact that in the Harvey scenario, many Twitter and Instagram photos were used, which are of lower quality than the data used in the training. Social media users may often overlay extra texts on photos, make collages from several photos, and also upload blurred photos in bad light conditions. These unexpected situations are challenging not only for the proposed water level estimation method but also for most of the computer vision algorithms to detect objects, segment images, or extract human poses. Pre-trained models specifically learned on social media images would be beneficial in providing better inputs to the proposed water level estimation model.

Due to differences in people's size and unknown camera perspectives, automatic interpretation of water levels from social media images can hardly reach centimeter-level accuracy, as e.g., water engineers would need. However, it is very useful for applications with lower accuracy requirements, e.g., emergency response, or to improve residents' situation awareness. The water level extracted from social media images is an intuitive indicator of flood severity. Rescuers and citizens would not take actions according to the extracted information alone but also combine it with their own interpretation of the information. These identified areas deserve special attention of emergency management agencies. In addition, the automatic flood severity interpretation from images extends the usefulness of VGI and provides users with the most evident information efficiently. Also this information can in principle be filtered and searched for images showing similar water level.

Model-based interpretation is hard to be perfect. Thus it can be beneficial to use a visual analytic approach with the human in the loop. Since the proposed method has eliminated most off-topic Tweets and Tweets that show no evidence of flooding, users need much less effort to verify model predictions and improve location quality. As for the 20,824 images during the 8 days of Hurricane Harvey, only 330 images with water level prediction needed to be validated for correctness.

Since remote sensing data used for disaster monitoring usually has a time delay, the information extracted from VGI can provide city managers with timely information on flood extent and severity. As presented in Section 6.4.3.2, VGI can provide more observations for populated areas, whereas remote sensing is good at detecting floodwater in less constructed areas. Therefore, VGI can be used as a good complement to remote sensing flood detection and delineation.

The extracted flood severity map demonstrates only a weak correlation to the modeled results gained from FEMA. However, by inspecting in combination with the individual water level estimation markers as presented in Section 6.4.3.2, decision-makers can get an intuitive situation awareness of, where severe inundation happens and how severe the situation is at the very moment. The water depth data provided by FEMA was acquired with different data sources and compiled after the event. However, the observations extracted from social media are individual, local observations, sparsely distributed with limited location accuracy at a particular time. Thus a large discrepancy between these two datasets can be expected.

In the following, the limitations of the current work are discussed from two aspects, one is the inherent challenge of social media VGI in general, and the other are elements where the current workflow can be improved and further investigated.

7.1 The inherent challenge of social media as opportunistic VGI

As presented in Section 2.5.1, data quality is an inherent challenge for using social media as an opportunistic VGI. The quality can be discussed with respect to time, location, and content for the application of flood monitoring.

Time quality In terms of time, there is typically a delay from the time the user observed the event to the time the photo was uploaded. This delay ranges from seconds to days and varies from individual to individual, which can hardly be detected or quantified. It has been observed in this research that the data points identified as noise by spatiotemporal clustering are often flood observations shared by users at a later time or in a very remote location. The only way to refine an unclear or uncertain time is if the user mentions it in the text explicitly or it is implied in their pictures (e.g., daylight or illumination), or it is confirmed by other users. However, there is no doubt that this is a task that is difficult to automate.

Location quality As discussed in Section 2.5.1, the locations reported by social media users can be the locations where users observed an event, and not necessarily the location where the event occurred. In addition, the user selected locations (i.e., the *place* field in Twitter data) are a mix of different location types, which can be, e.g., a city name, a city district name, or a POI. Therefore, assessing the suitability of the location quality of social media data for the task at hand is a necessary step. As for the extraction of flood observations, social media VGI can be used as a source of information to raise situational awareness. However, for the task of the verification of detailed flood simulation results of hydrologists, very precise geolocations of flood level reports are required, at least at the meter or decimeter level. Current location quality (i.e., median offset of around 200m for Twitter) does not meet this need. If social media data are to be used for this purpose, it is still necessary to perform manual corrections or eliminate posts with poor location quality.

In addition, social media users may send their posts with an inaccurate or even fake location. Two common situations were observed. One is that people retweet or share information from other users' observations or news media images at their current location. Many of these cases can be detected and eliminated by the proposed image duplication detector based on the assumption that their shared images are the same or similar. The other situation is that people assign a wrong location intentionally or unintentionally. This case cannot be easily solved by the interpretation of the social media text and image alone.

Content quality In terms of content information, especially for extracting flood-relevant information from images, photo editing and low-quality images are great challenges in many cases. In general, these problems can be mitigated when several posts at a certain location and time are available. A majority filter concerning the semantics can be applied. In addition, fake news and content is often spread very fast on social media. However, it is not yet considered in the frameworks proposed in this research. Even if there are studies that detect such fake information based on text interpretation (Zhou and Zafarani, 2020), validation based on other sources of information and other users is still necessary. It also requires more sophisticated algorithms to achieve a robust performance.

Even if social media VGI are sparse and are provided with varying intensity in space and time (and quality), also interesting inferences can be drawn: for example, if there are many Tweets in a region, but no flood-relevant Tweets, then there is a high probability that there is no flood event.

7.2 Limitations of the current social media processing pipeline

There are also limitations of the frameworks proposed in this thesis. Some are inherent limitations of the method, and some others can be considered for further improvement.

Water level estimation Concerning the water level estimation using people as targets, three limitations of the current method can be identified. First, the result is a qualitative estimate of the water level. However, it is still not sufficient to derive the water depth observations in metric units that e.g. hydrologists expect. Although it can be approximated using the average height of a person and the common ratio of the body parts, this may lead to a large uncertainty due to the large individual differences in human height. In addition, the current model cannot consider the difference between adults and children. Thus, children in the scene may cause an overestimation of the flood severity. Second, there are posts containing multiple images with different water levels at the same location. Currently, a voting strategy is used to aggregate multiple flood severity estimations to the location on the map. In order to solve this problem, additional information from the scene has to be taken into account. Third, the number of images that can be used for water level estimation is limited. In this work, 676 out of 3,142 flood-relevant images could be used for water level mapping during Hurricane Harvey. The number of useful images might be much fewer for a less significant event or events in less populated areas. In order to increase the number of usable images, additional fixed-size objects should be introduced, such as cars or bikes.

VGI based flood mapping Due to the sparseness of social media posts in time and space, standalone social media data analysis presents only very limited information. Some studies tried to combine social media information with other sources of information such as remote sensing flood detection, terrain information, and river gauge measurements. Most of these studies applied two-dimensional kernels at the VGI locations, e.g., Huang et al. (2018b). However, the kernel bandwidth is an empirical hyperparameter chosen differently for each study. Some other recent research based on terrain information tends to overly trust social media location information to conduct mapping in a very detailed level. For example, Li et al. (2018) rely on the height of VGI locations and user-mentioned water levels to estimate a local water surface - a horizontal plane. Huang et al. (2018a) assume that the areas of higher flood probability are those below the flood-related VGI locations, based on the digital terrain model. Such approaches are risky in the absence of manual verification because many of the Tweets containing precise coordinates are from shared Instagram posts. As described above, these coordinates may very often correspond to the location of cities, urban districts, or other polygonal areas such as university campuses, hospitals, etc., and not to an exact point location.

Thus, in this thesis, the results of flood mapping are presented in an aggregated way, i.e., combining the information of individual posts in pre-defined spatial units (i.e., administrative polygons and census tracts). The use of the proposed framework is recommended to be limited to raising situational awareness and early warning. Nevertheless, aggregating to spatial units may lead to another concern: the Modifiable Area Unit Problem - MAUP (Ratcliffe, 2004). The identified spatial patterns can be different when using another definition of spatial units.

Framework and time efficiency Social media is a real-time data source. With the models trained in advance, this real-time property can be preserved by setting up a proper infrastructure to analyze the data. OpenPose and Mask R-CNN have been proved to achieve a real-time performance in (Cao et al., 2019; He et al., 2017). Nowadays, there are also emerging solutions to achieve a real-time performance on semantic segmentation (e.g., Yu et al., 2018). Xgboost used in the methods for classification can also be deployed as a real-time online service (Negrey and Yang, 2018).

Nevertheless, a systematic time budget calculation is still needed, which was, however, beyond the scope of the current work.

Online deep learning For a framework that needs to run in real-time, its ability to learn and self-correct continuously is intuitively important. However, this is not yet included in the proposed process in this thesis. As with many existing software, users can report if there were any misclassifications during use. Based on these additional negative samples, the model can be further fine-tuned to better fit real-world scenarios. Online Deep Learning (ODL), an active research area in machine learning, offers many solutions (e.g., Sahoo et al., 2018) to this problem. It has also been performed for the retrieval of crisis-related social media data in (Nguyen et al., 2016). Therefore, this could be an add-on component to build a more intelligent real-time system.

8 Conclusions and outlook

In this thesis, approaches to extract precipitation and flood observations from opportunistic VGI were investigated. The answers to the research questions are summarised in Section 8.1. Section 8.2 outlines future research directions based on the results of this thesis.

8.1 Research questions

With respect to the research questions defined in Section 1.2, the experiments demonstrate the following results.

- **Can precipitation indications be extracted from the speed variation of road users?**

  Chapter 4 presented a proof-of-concept experiment to train a binary precipitation indicator based on the road users' speed of multiple roads. The proposed method first modeled the traffic speed data with a seasonal trend decomposition to eliminate the daily and weekly periodic effects of the road speed observations. Residuals of this model were used as features that indicate the anomaly level compared to the normal traffic state. Since only a limited amount of positive examples were available within the observation range, the experiment was confined to a binary classifier. Several machine learning models were trained on a six-month road speed dataset and compared based on the subsequent two months of data. The best model has achieved a promising performance with an accuracy of 91.74% and an F_1-score of 78.34%. The good performance of the model demonstrates the feasibility of exploiting this information.

  It is certainly not intended to use it as a replacement for the current weather stations. However, cities' transportation departments and navigation service providers are regularly collecting this information from traffic speed detectors or voluntary users' trajectory data (FCD) in large volumes anyway. This experiment demonstrates the potential using these speed data to provide a by-product for precipitation monitoring, which might be beneficial for the areas lacking basic meteorological facilities.

- **What is the benefit of jointly exploiting text and images from social media to extract high-quality pluvial flood observations?**

  Section 6.3 presented a framework to collect, process, and analyze pluvial flood-relevant information from Twitter and Instagram. Previous studies mainly focused on user-generated text only. This thesis presented a very early attempt to apply deep learning models on both text and images to extract pluvial flood-related social media posts. An additional contribution of this thesis is to automatically annotate training examples for text classification by filtering pluvial flood-related keywords and querying the corresponding weather information based on date and geotags. The model could achieve a performance comparable to that of a model trained on a manually annotated training dataset. As river and lake images are easily confused with flood scenarios, they were specially considered to enhance the robustness of the image classification model.

  Further analyses were able to identify spatiotemporal clusters and hotspot areas for the flood events in Western Europe in 2016 and 2017. Spatial and temporal characteristics of social

media posts during a pluvial flood in London and a fluvial (river) flood in Paris were analyzed using this framework.

- **How can the flood severity information be interpreted from social media images and how far they are helpful for flood mapping?**

  Section 6.4 presented a novel approach to extract flood severity information from opportunistic VGI. Flood-relevant images were filtered out of social media. Similar and potentially duplicated images were eliminated. People in flood-relevant images were used as scales, where qualitative water level estimates with respect to different body parts (namely ankle, knee, hip, and chest) were obtained. The evaluation on a representative dataset demonstrated an accuracy of 90%. It should also be highlighted that the proposed model achieved fine-grained water level classification with less annotation effort compared with previous studies.

  Furthermore, the proposed pipeline method was applied to a real flood event, Hurricane Harvey in 2017, as a proof-of-concept. Flood extent was estimated based on VGI, which correctly marked over 62% of the regions where people have claimed flood or water damage. The results further showed that VGI can be used as a supplement to remote sensing observations for flood extent mapping and is beneficial, especially for urban areas, where the infrastructure often occludes water in remote sensing images. Flood severity was mapped based on the interpretation of social media images and was compared with the modeled flood depth map. The result indicates a weak positive monotonic correlation to the reference data. An integrated overview of flood severity can be provided for the early stages of emergency response based on the extracted water level information.

8.2 Outlook

Regarding the use of vehicle behavior as a precipitation indicator, this thesis used road speed data from traffic speed detectors as a substitute of opportunistic VGI. This is because there are few publicly available trajectory datasets that can be used for this experiment, covering long periods of time and many precipitation events. However, navigation providers such as TomTom, HERE, Didi Chuxing have such data and use it for a long time to monitor road conditions and vehicle speeds. Compared to traffic speed detectors, GPS trajectories provide speed data that can cover more roads and collect spatial road speed information at a finer granularity. This can provide more features to learn a precipitation indicator. With a longer observation period containing more examples, this approach can be further investigated to see if an indicator can be learned that can estimate also the severity of precipitation events. With denser road speed observations, also local analyses are possible. Separate models can be trained for individual areas to determine which areas are experiencing such precipitation events and when traffic is starting to be affected.

Regarding the interpretation of flood-relevant information from social media VGI, there are three aspects, which have been mentioned in Section 7.2. In addition to people, other objects with approximately known dimensions can be analyzed to extract water levels, such as vehicles and bicycles. Considering these objects with additional component-level information, more flood-related images can be used to increase the number of effective observations. In this thesis, the data collected during flood events were analyzed offline. However, the efficiency of the entire system needs to be further investigated to test how much time the VGI can be ahead of the information from remote sensing flood detection in practice. Furthermore, online learning can be considered, where models can be tuned with new training data and higher robustness can be achieved.

Regarding general future research directions, there are the following four aspects.

- **VGI quality on time, location, and content**

  The quality issues for VGI were thoroughly discussed in Section 7.1 with respect to time, location, and content. Similar issues have also been mentioned in several previous reviews or studies (Klonner et al., 2016; Yan et al., 2020). Even though deep learning-based algorithms extract textual and visual information far more accurately and efficiently than before, automatically verifying or correcting the time and location of user postings remains a great challenge. This is important for applications that require more precise geolocation of VGI, e.g., the validation of flood simulation results (Rözer et al., 2021).

  Social media posts with precise GPS locations are becoming rare, as users are more likely to provide abstract locations than their actual coordinates. Hu and Wang (2020) found that social media users tend to provide precise address or road intersection information in their texts under emergency situations, especially when asking for help. When detailed location information was mentioned by the user in the text, the Named-Entity Recognition (NER) based Geoparser (e.g., Wang et al., 2019) has demonstrated a great potential to provide the posts with more precise geolocation. In this way, some of the user-generated locations can be verified, and more posts without location information can be used for VGI-based flood analysis.

  Efforts have also been made to take advantage of visual information, aiming to infer geographic coordinates using image retrieval techniques, e.g., in (Muller-Budack et al., 2018). As for the well-known Im2GPS dataset (Hays and Efros, 2008), reasonable location inferences can only be achieved at the country or even continental level (Vo et al., 2017). However, for urban scenarios with street-level geo-tagged images, e.g., the San Francisco Dataset (Chen et al., 2011), a 2D image retrieval-based method has achieved 10m-level localization error standalone or 5m-level after combining with a local Structure from Motion (SfM) reconstructions (Sattler et al., 2017). With the availability of detailed 3D LiDAR data and high-definition maps in many cities and many regions, the positioning of social media photos is expected to be easier, and the accuracy of visual localization is expected to be improved, as demonstrated in (Brenner, 2009; Cattaneo et al., 2019, 2020). As more geo-tagged images and LiDAR data accumulate, and as computing power increases, vision-based localization has great potential to provide precise location estimation also for social media images.

- **Fake user-generated contents**

  Fake news and contents on Twitter are often retweeted by many more users and spread far more rapidly (Vosoughi et al., 2018). Fake news regarding natural disasters may lead to misallocation of resources, and in extreme cases, endanger people's lives (Johnson, 2020). Detection of fake news has been studied mainly based on text interpretation (Zhou and Zafarani, 2020). However, this requires not only the understanding of current social media content provided by users but also the comparison and verification with information from other sources and other users.

  Current developments in computer vision, such as Generative Adversarial Networks (GANs), can generate photo-realistic human faces (Karras et al., 2019) and videos (e.g., text-driven video synthesis in Thies et al., 2020). It is believed that even novices can also become proficient in using these techniques to generate fake pictures or videos in the near future. Thus it is essential to develop reliable methods to detect fake posts.

- **Surveillance cameras**

  Many cities already have well-developed traffic monitoring infrastructure, thus flood-related observations can also be obtained through an automatic image interpretation. The surveillance cameras are mostly static and with well-known geo-location. If observations from real-time camera data from multiple locations are integrated with data from existing flood-related sensor networks, there is the potential to provide more comprehensive flood risk observations for cities at a low cost. These observations could potentially also improve the risk perception of city's emergency management.

- **Video and LiDAR, potential social media content for disaster-related VGI analysis**

  In recent years, online short videos are receiving increasing attention. Conventional social media platforms, such as Facebook, Twitter, Instagram, Weibo, all support users to upload videos. Tiktok[1], as an emerging social media platform for short videos, is very popular among young people. Disaster-related social media videos have been considered in early-stage studies via manual interpretation. However, only few studies explored the possibility of adopting automated procedures for analyzing social media videos. Meanwhile, user-generated videos are often shaking or blurry, posing challenges for traditional video analytics. The dynamic nature of videos facilitates the retrieval of other essential information. For floods, scholars started to use social media videos to estimate water flow velocity, e.g., Le Boursicaud et al. (2016) extracted flow velocity of rivers with LSPIV from YouTube videos. However, such an approach needs to survey ground reference points for calibration.

  The growing popularity of LiDAR on mobile devices (e.g., Apple iPad Pro, iPhone 12 pro) has made it much easier for users to obtain 3D measurements. Many applications have been developed to measure the size of objects and the height of people[2]. With this, social media users can potentially provide more precise water level estimations, more precise geo-location, or coordinates of reference points for flow velocity estimation with their mobile devices in the future.

Even though the information provided by voluntary users is sparsely distributed and contains inherent uncertainties in time, location, and content, experiments have shown that it can complement current remote sensing flood detection in areas where they are inadequate. Therefore, city governors and emergency management agencies could consider to include such an automatic social media VGI interpretation and analysis component into their existing emergency management systems.

[1] Tiktok. `https://www.tiktok.com/` (Accessed on 31.01.2021)

[2] How to measure someone's height with the iPhone 12 - Macworld UK. `https://www.macworld.co.uk/how-to/how-measure-someones-height-with-iphone-12-3797186/` (Accessed on 31.01.2021)

List of Figures

1.1 Global heat map for the large flood events 1985-2019. 1
1.2 The main causes of the large flood events 1985-2019. 2
1.3 Global distribution of weather radars in the WMO radar database; status: 31.01.2021 (WMO, 2020). 3

2.1 Example of margin of a linear SVM model (image under CC BY-SA 4.0). 10
2.2 Example of a random forest model. 11
2.3 Illustration of Gradient Boosting Decision Trees. 12
2.4 Example of Receiver Operating Characteristic (ROC) curves (adapted based on code example under BSD license). 13
2.5 Illustration of an Artificial Neural Network. 14
2.6 Illustration of a Convolutional Neural Network. 16
2.7 Overview of the network architectures: VGG16, InceptionV3, ResNet, Inception-ResNetV2 and DenseNet for image classification (adapted based on Figures 1-6 from Mahdianpari et al. (2018) under CC BY 4.0). 18
2.8 Example of atrous convolution (image under MIT License). 20
2.9 Network architecture overview of DeeplabV3+ (Chen et al., 2018). 20
2.10 Network architecture overview of Mask R-CNN (He et al., 2017). 21
2.11 Network architecture overview of OpenPose (Cao et al., 2019). 22
2.12 CBOW and skip-gram architectures for generating Word2vec word embedding (Mikolov et al., 2013a). 24
2.13 Illustration of TextCNN architecture used for text classification, adapted based on Figure 1 from (Zhang and Wallace, 2017). 25
2.14 Example of KDE using different bandwidths. 27
2.15 Example of DBSCAN (image under CC BY-SA 3.0). 28
2.16 Examples of geotagged Tweets with exact coordinate (left), bounding box of a city district, and bounding box of a POI (right). 33
2.17 Examples of geotagged Instagram posts shared on Twitter with different location levels: city-level location (left), city-district-level location (middle), and POI-level location (right). 34

4.1 Spatial distribution of road segments in New York City (left, Basemap: OpenStreetMap) and precipitation data retrieved from Weather Underground (right). . 50
4.2 Example results achieved by *Prophet*, speed observations versus predictions, for one road segment within 24 days. 51
4.3 Comparison of some of the classical machine learning methods. 51
4.4 Predictions on 2-month traffic speed test dataset with Xgboost. The blue line indicates the precipitation amount in millimeter and red line indicates the prediction from the Xgboost model. 52
4.5 Importance of each road, as determined by Xgboost (Basemap: OpenStreetMap). . 53

4.6 Speed variation pattern for a precipitation event on 15th of April 2018. The color indicates the speed observation, slower (blue) or faster (red) than the *Prophet* estimated model. The black and green lines represent the start and end times of the event based on the text description. On the right side, the blue line indicates the corresponding precipitation amount, and the red line indicates the prediction from the Xgboost model. 54
4.7 Speed variation pattern for a precipitation event on 25th of April 2018. The color indicates the speed observation, slower (blue) or faster (red) than the *Prophet* estimated model. The black and green lines represent the start and end times of the event based on the text description. On the right side, the blue line indicates the corresponding precipitation amount, and the red line indicates the prediction from the Xgboost model. 54

5.1 Workflow for training the text classifiers. 59
5.2 Feature fusion model for image classification into two classes: flood relevant and irrelevant. 62
5.3 Workflow of water level estimation model 63
5.4 Output of OpenPose with 18 body keypoints (OpenPose, 2018). 63
5.5 Steps for extracting handcrafted distance features (example image under CC BY-NC-SA 2.0). 64
5.6 Network architecture of baseline 2: Mask R-CNN with water level classification branch using local deep features (example image under CC BY-NC-SA 2.0). 66

6.1 Study areas for collecting Twitter data (Basemap: OpenStreetMap). 67
6.2 Proportion of geotagged Tweets containing photos. 68
6.3 Examples of training dataset: (a) rainfall and flooding irrelevant images, (b) relevant images and (c) images of water surfaces. 70
6.4 Annotation rules for water level estimation of single person. 71
6.5 Workflow for the extraction of pluvial flood relevant VGI. 72
6.6 Comparison of text classification methods on test set. 73
6.7 ROC curves of text classification methods. 74
6.8 Comparison of image classification methods on test set. 76
6.9 ROC curves of image classification methods. 76
6.10 Comparison of image classification methods on test set. 77
6.11 ROC curves of image classification methods. 77
6.12 Three typical failure cases of the classifiers: water surfaces in relative dark color (left), images containing reflecting area (middle), and photos containing fountains or springs (right). 78
6.13 Spatiotemporal cluster detected by ST-DBSCAN (pluvial flood in London on 26th of June 2016, green markers are the aggregated Tweets only for visualization). . . 78
6.14 Map of daily average number of Tweets based on aggregation of 90 days' Tweets. . 80
6.15 Ratio map on 3rd of June 2016 in Paris. 80
6.16 Hot spots detected by Getis-Ord Gi* on 3rd of June 2016 in Paris. 80
6.17 Screen-shots of the web map application (pluvial flood in Berlin on 29th of June 2017). 81
6.18 Comparison of the retrieval strategies (Paris, 17th of May – 30th of June 2016). . . 82
6.19 Comparison of the retrieval strategies (London, 17th – 30th of June 2016). 83
6.20 Comparison of pluvial flood event (left, London, 23th of June 2016) and fluvial flood event (right, Paris, 3rd of June 2016). 84

6.21 Workflow of the process to extract flood extent and flood severity from social media data. 85
6.22 Evaluation of models on DIRSM test set (left) and extended DIRSM test set (right). 87
6.23 Evaluation of different combinations of feature groups performed on test set. . . . 88
6.24 Qualitative evaluation of the proposed approach compared with the baselines (example images under CC BY-NC-SA 2.0). 89
6.25 Example failure cases of this approach, caused by segmentation failure - left, sitting people - middle, and water reflection - right (example images under CC BY-NC-SA 2.0). 90
6.26 Comparison of confusion matrices on the test set using baseline 1 (left), baseline 2 using 1/4 area beneath (middle) and the proposed method (right). 90
6.27 Distribution of the model predicted flood relevance scores for the images collected during Hurricane Harvey. 92
6.28 Sorted 2-distance plot for image deep features. 92
6.29 Examples of the duplicate images from the largest cluster of DBSCAN result. . . . 93
6.30 Confusion matrix of the water level estimation on social media images with flood relevance over 99%. 93
6.31 Map of social media posts with severity predictions as markers (Basemap: OpenStreetMap). 94
6.32 Locations of flood relevant Tweets with overlaid census tracts as the flood extent detected by VGI. 96
6.33 Maximum observed flooding mapped from NASA MODIS, ESA Sentinel 1, ASI COSMO-SkyMed, and RADARSAT 2 data from Dartmouth Flood Observatory (DFO, 2017) and the overlaid census tracts. 96
6.34 FEMA property claims and the overlaid census tracts. Data source: FEMA (2018a). 97
6.35 FEMA property claims density map and the flood extent detected by VGI. Data source: FEMA (2018a). 97
6.36 FEMA Harvey flood depth grid (FEMA, 2018b). 99
6.37 Aggregated flood depth tracts with max depth values. 99
6.38 Flood severity map derived from the water level estimations of VGI images. 100

List of Tables

1.1 Comparison between current precipitation and flood monitoring approaches. 4

2.1 Example of a confusion matrix for binary classification. 13
2.2 Data input for learning Word2vec word embedding using CBOW 24
2.3 Data input for learning Word2vec word embedding using skip-gram 24
2.4 Social media and mobility data sources for opportunistic VGI research. 30

3.1 Standalone analysis of social media VGI for disasters. 39
3.2 Analysis of social media VGI in combination with other sources of information. . . 41
3.3 Extraction of disaster-related social media posts based on texts. 44
3.4 Extraction of disaster-related social media posts based on images. 45
3.5 Flood water level estimation from social media posts. 47

4.1 Precision, recall and F_1-score on test set for Xgboost model. 52
4.2 Confusion matrix on test set for Xgboost model. 52

5.1 Examples of Tweets with similar structure of texts. 58
5.2 Keywords used for generating training dataset. 59
5.3 Feature names in each feature group. 65

6.1 Number of positive and negative examples for dataset. 71
6.2 Composition of train set and test set for water level estimation 72
6.3 Parameters used for training the text classifiers. 73
6.4 Evaluation of text classification methods. 73
6.5 Evaluation of text classification on 2013 Queensland floods dataset. 74
6.6 Parameters used for training the image classifiers. 75
6.7 Evaluation of image classification methods. 76
6.8 Evaluation of image classification methods. 77
6.9 Correlations between the proportion of topic related Tweets and rainfall intensity (Paris, 17^{th} of May – 30^{th} of June 2016). 82
6.10 Correlations between the proportion of topic related Tweets and rainfall intensity (London, 17^{th} – 30^{th} of June 2016). 83
6.11 Evaluation of different approaches on *MMSat Task* in *MediaEval'17* and comparison with this approach. 86
6.12 Evaluation of model performance based on precision, recall and F_1-scores on positive class, Overall Accuracy (OA) and Area Under Curve (AUC). 86
6.13 Parameters for all methods. 87
6.14 Quantitative comparison of models for water level estimation. 90
6.15 Comparison of water extent mapping from different information sources. 98
6.16 Confusion matrices of water extent mapping from different information sources. . . 98

Bibliography

Abdulla, W., 2017. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. `https://github.com/matterport/Mask_RCNN`, (Accessed on 31.01.2021).

Ahmad, K., Pogorelov, K., Riegler, M., Conci, N., Halvorsen, P., 2017a. CNN and GAN Based Satellite and Social Media Data Fusion for Disaster Detection. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017. pp. 1–3.

Ahmad, K., Pogorelov, K., Riegler, M., Ostroukhova, O., Halvorsen, P., Conci, N., Dahyot, R., 2019. Automatic detection of passable roads after floods in remote sensed and social media data. Signal Processing: Image Communication 74, 110–118.

Ahmad, K., Sohail, A., Conci, N., De Natale, F., 2018. A comparative study of global and deep features for the analysis of user-generated natural disaster related images. In: 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP). IEEE, pp. 1–5.

Ahmad, S., Ahmad, K., Ahmad, N., Conci, N., 2017b. Convolutional neural networks for disaster images retrieval. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017. pp. 1–3.

Akoglu, H., 2018. User's guide to correlation coefficients. Turkish journal of emergency medicine 18 (3), 91–93.

Alam, F., Joty, S., Imran, M., 2018. Domain adaptation with adversarial training and graph embeddings. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1077–1087.

Alam, F., Ofli, F., Imran, M., 2020a. Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes harvey, irma, and maria. Behaviour & Information Technology 39 (3), 288–318.

Alam, F., Ofli, F., Imran, M., Alam, T., Qazi, U., 2020b. Deep learning benchmarks and datasets for social media image classification for disaster response. arXiv preprint arXiv:2011.08916 .

Alfonso, L., Chacón, J. C., Peña-Castellanos, G., 2015. Allowing citizens to effortlessly become rainfall sensors. In: 36th IAHR World Congress edited, The Hague, the Netherlands. pp. 1520–0477.

Alfonso, L., Lobbrecht, A., Price, R., 2010. Using mobile phones to validate models of extreme events. In: 9th international conference on hydroinformatics, Tianjin, China. pp. 1447–1454.

Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., Zuair, M., 2017. Deep learning approach for car detection in uav imagery. Remote Sensing 9 (4), 312.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., Sander, J., 1999. Optics: Ordering points to identify the clustering structure. ACM Sigmod record 28 (2), 49–60.

Asmolov, G., 2010. Russia: Crowdsourcing assistance for victims of wildfires. Global Voices .

Assumpção, T. H., Popescu, I., Jonoski, A., Solomatine, D. P., 2018. Citizen observations contributing to flood modelling: opportunities and challenges. Hydrology and Earth System Sciences 22 (2), 1473–1489.

Atkinson, G. M., Wald, D. J., 2007. "Did You Feel It?" intensity data: a surprisingly good measure of earthquake ground motion. Seismological Research Letters 78 (3), 362–368.

Aulov, O., Price, A., Halem, M., 2014. Asonmaps: A platform for aggregation visualization and analysis of disaster related human sensor network observations. In: Proceedings of the ISCRAM. pp. 802–806.

Avgerinakis, K., Moumtzidou, A., Andreadis, S., Michail, E., Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I., 2017. Visual and textual analysis of social media and satellite images for flood detection@ multimedia satellite task mediaeval 2017. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39 (12), 2481–2495.

Barz, B., Schröter, K., Münch, M., Yang, B., Unger, A., Dransch, D., Denzler, J., 2018. Enhancing flood impact analysis using interactive retrieval of social media images. Archives of Data Science, Series A (Online First) 5 (1), 06.

BBC, 2016. Flash flooding causes chaos in parts of england. `https://www.bbc.com/news/uk-england-london-36471889`, (Accessed on 31.01.2021).

BBC, 2016. In pictures: Flash floods and flashes of lightning. `https://www.bbc.com/news/uk-36472468`, (Accessed on 31.01.2021).

BBC, 2016. Paris floods: Seine at 30-year high as galleries close. `https://www.bbc.com/news/world-europe-36446635`, (Accessed on 31.01.2021).

BBC, 2017. Houston floods: Harvey rains to worsen texas city's plight. `https://www.bbc.com/news/world-us-canada-41079746`, (Accessed on 31.01.2021).

Begum, S., Otung, I. E., 2009. Rain cell size distribution inferred from rain gauge and radar data in the UK. Radio Science 44 (02), 1–7.

Birant, D., Kut, A., 2007. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & knowledge engineering 60 (1), 208–221.

Bischke, B., Bhardwaj, P., Gautam, A., Helber, P., Borth, D., Dengel, A., 2017a. Detection of flooding events in social multimedia and satellite imagery using deep neural networks. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017.

Bischke, B., Helber, P., Schulze, C., Srinivasan, V., Dengel, A., Borth, D., 2017b. The multimedia satellite task at mediaeval 2017: Emergency response for flooding events. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017.

Bischke, B., Helber, P., Zhao, Z., de Bruijn, J., Borth, D., 2018. The multimedia satellite task at mediaeval 2018 emergency response for flooding events. In: Working Notes Proceedings of the MediaEval 2018 Workshop, Sophia Antipolis, France, 29-31 October 2018.

Bohm, G., Zech, G., 2010. 4.6 confidence intervals. In: Introduction to statistics and data analysis for physicists. Vol. 1. Desy Hamburg, pp. 116–119.

Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5, 135–146.

Boser, B. E., Guyon, I. M., Vapnik, V. N., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory. pp. 144–152.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A., 1984. Classification and regression trees. CRC press.

Brenner, C., 2009. Extraction of features from mobile laser scanning data for future driver assistance systems. In: Advances in GIScience. Springer, pp. 25–42.

Broniatowski, D. A., Paul, M. J., Dredze, M., 2013. National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. PloS one 8 (12).

B.Z., 2017. Zu viel Wasser für Berlin: Stadt Versinkt im Verkehrs-Chaos - B.Z. Berlin. http://www.bz-berlin.de/berlin/unwetterwarnung-berlin-wetter, (Accessed on 31.01.2021).

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y., 2019. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. IEEE transactions on pattern analysis and machine intelligence 43 (1), 172–186.

Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y., 2017. Realtime multi-person 2d pose estimation using part affinity fields. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1302–1310.

Cattaneo, D., Sorrenti, D. G., Valada, A., 2020. Cmrnet++: Map and camera agnostic monocular visual localization in lidar maps. arXiv preprint arXiv:2004.13795 .

Cattaneo, D., Vaghi, M., Ballardini, A. L., Fontana, S., Sorrenti, D. G., Burgard, W., 2019. Cmrnet: Camera to lidar-map registration. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, pp. 1283–1289.

Cerutti, V., Fuchs, G., Andrienko, G., Andrienko, N., Ostermann, F., 2016. Identification of disaster-affected areas using exploratory visual analysis of georeferenced tweets: application to a flood event. Association of Geographic Information Laboratories in Europe: Helsinki, Finland 5.

Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., Waters, N., 2016. Using twitter for tasking remote-sensing data collection and damage assessment: 2013 boulder flood case study. International Journal of Remote Sensing 37 (1), 100–124.

Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., Ertl, T., 2012. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In: 2012 IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE, pp. 143–152.

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., Ebert, D. S., 2014. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. Computers & Graphics 38, 51–60.

Chatzichristofis, S. A., Boutalis, Y. S., 2008. Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In: International Conference on Computer Vision Systems. Springer, pp. 312–322.

Chaudhary, P., D'Aronco, S., Moy de Vitry, M., Leitão, J. P., Wegner, J. D., 2019. Flood-water level estimation from social media images. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W5, 5–12.

Chaudhary, P., D'Aronco, S., Leitão, J. P., Schindler, K., Wegner, J. D., 2020. Water level prediction from social media images with a multi-task ranking approach. ISPRS Journal of Photogrammetry and Remote Sensing 167, 252–262.

Chen, D. M., Baatz, G., Köser, K., Tsai, S. S., Vedantham, R., Pylvänäinen, T., Roimela, K., Chen, X., Bach, J., Pollefeys, M., et al., 2011. City-scale landmark identification on mobile devices. In: CVPR 2011. IEEE, pp. 737–744.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40 (4), 834–848.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Computer Vision – ECCV 2018. Springer International Publishing, pp. 833–851.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794.

Cheng, H., Zourlidou, S., Sester, M., 2020. Traffic control recognition with speed-profiles: A deep learning approach. ISPRS International Journal of Geo-Information 9 (11), 652.

Chien, J., 2019. Validating the quality of crowdsourced data for flood modeling of hurricane harvey in houston, texas .

Choe, C., Seo, S., Sreetharan, S., 2017. Real-time, crowd-sourced flood mapping & analytics via iseeflood. In: meeting of New York State Floodplain and Stormwater Managers Association Conference, Binghamton, New York.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1251–1258.

Cifelli, R., Doesken, N., Kennedy, P., Carey, L. D., Rutledge, S. A., Gimmestad, C., Depue, T., 2005. The community collaborative rain, hail, and snow network: Informal education for scientists and citizens. Bulletin of the American Meteorological Society 86 (8), 1069–1078.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20 (3), 273–297.

Cowan, J., 2017. When 911 was overloaded, desperate harvey victims turned to social media for help. `https://www.govtech.com/em/disaster/When-911-Failed-Them-Desperate-Harvey-Victims-Turned-to-Social-Media-for-Help.html`, (Accessed on 31.01.2021).

CRED, UNISDR, 2018. Economic losses, poverty and disasters 1998-2017. The United nations Office for Disaster Risk Reduction , 33.

Cresci, S., Cimino, A., Dell'Orletta, F., Tesconi, M., 2015. Crisis mapping during natural disasters via text analysis of social media messages. In: International Conference on Web Information Systems Engineering. Springer, pp. 250–258.

Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J., 2013. # earthquake: Twitter as a distributed sensor system. Transactions in GIS 17 (1), 124–147.

Cvetojevic, S., Juhasz, L., Hochmair, H., 2016. Positional accuracy of twitter and instagram images in urban environments. GI_Forum 2016 1, 191–203.

Daly, S., Thom, J. A., 2016. Mining and classifying image posts on social media to analyse fires. In: Proceedings of the ISCRAM 2016 Conference – Rio de Janeiro, Brazil, May 2016. pp. 1–14.

Dao, M.-S., Quang Nhat Minh, P., Kasem, A., Haja Nazmudeen, M. S., 2018. A context-aware late-fusion approach for disaster image retrieval from social media. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval. ACM, pp. 266–273.

de Brito Moreira, R., Degrossi, L. C., de Albuquerque, J. P., 2015. An experimental evaluation of a crowdsourcing-based approach for flood risk management. In: Proceedings of the 12th Workshop on Experimental Software Engineering (ESELAW), Lima, Peru. pp. 22–24.

De Longueville, B., Smith, R. S., Luraschi, G., 2009. " omg, from here, i can see the flames!" a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In: Proceedings of the 2009 international workshop on location based social networks. pp. 73–80.

de Vasconcelos, L. E. G., dos Santos, E. C., Neto, M. L., Ferreira, N. J., de Vasconcelos, L. G., 2016. Using tweets for rainfall monitoring. In: Information technology: New generations. Springer, pp. 1157–1167.

Degrossi, L. C., Albuquerque, J. P. d., Fava, M. C., Mendiondo, E. M., 2014. Flood citizen observatory: a crowdsourcing-based approach for flood risk management in brazil. In: Proceedings of the International Conference on Software Engineering and Knowledge Engineering. pp. 570–575.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp. 248–255.

DFO, 2017. DFO Flood Event 4510 - Hurricane Harvey, Texas and Lousiana. https://floodobservatory.colorado.edu/Events/2017USA4510/2017USA4510.html, (Accessed on 31.01.2021).

Dilley, M., Chen, R. S., Deichmann, U., Lerner-Lam, A. L., Arnold, M., 2005. Natural disaster hotspots: a global risk analysis. The World Bank.

Ding, X., Fan, H., 2019. Exploring the distribution patterns of flickr photos. ISPRS International Journal of Geo-Information 8 (9), 418.

Dittrich, A., Lucas, C., 2014. Is this twitter event a disaster? In: 17th AGILE Conference on Geographic Information Science. Connecting a Digital Europe through Location and Place, Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, E., June, 3-6, 2014. Ed.: J. Huerta.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In: International conference on machine learning. pp. 647–655.

Dumais, S., Platt, J., Heckerman, D., Sahami, M., 1998. Inductive learning algorithms and representations for text categorization. In: Proceedings of the seventh international conference on Information and knowledge management. pp. 148–155.

DWD, 2020a. Niederschlagsradar (bild und film). https://www.dwd.de/DE/leistungen/radarbild_film/radarbild_film.html, (Accessed on 31.01.2021).

DWD, 2020b. Radarverbund. https://www.dwd.de/DE/derdwd/messnetz/atmosphaerenbeobachtung/_functions/Teasergroup/radarverbund_teaser5.html?nn=452870, (Accessed on 31.01.2021).

EA, 2020. Environment agency rainfall api. https://environment.data.gov.uk/flood-monitoring/doc/rainfall, (Accessed on 31.01.2021).

Earl, J., McKee Hurwitz, H., Mejia Mesinas, A., Tolan, M., Arlotti, A., 2013. This protest will be tweeted: Twitter and protest policing during the pittsburgh g20. Information, Communication & Society 16 (4), 459–478.

Earle, P. S., Bowden, D. C., Guy, M., 2011. Twitter earthquake detection: earthquake monitoring in a social world. Annals of Geophysics 54 (6).

ebvImages, 2011. Flood - Thailand. https://www.flickr.com/photos/ebvimages/albums/72157628033411293, (Accessed on 31.01.2021).

Eilander, D., Trambauer, P., Wagemaker, J., Van Loenen, A., 2016. Harvesting social media for generation of near real-time flood maps. Procedia Engineering 154, 176–183.

ESA, 2020. Sentinel-2. https://sentinel.esa.int/web/sentinel/missions/sentinel-2, (Accessed on 31.01.2021).

ESRI, 2019a. How Hot Spot Analysis (Getis-Ord Gi*) works. https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-statistics-toolbox/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm, (Accessed on 31.01.2021).

ESRI, 2019b. What is a z-score? What is a p-value? https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/what-is-a-z-score-what-is-a-p-value.htm, (Accessed on 31.01.2021).

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. KDD'96. AAAI Press, p. 226–231.

FEMA, 2018a. U.S. Federal Emergency Management Administration (FEMA) - Harvey Damage Assessments and Claims, HydroShare. https://doi.org/10.4211/hs.73c4f3dcff884a6da2c0982df769987c, (Accessed on 31.01.2021).

FEMA, 2018b. U.S. Federal Emergency Management Administration (FEMA) - Harvey Flood Depths Grid, HydroShare. https://doi.org/10.4211/hs.165e2c3e335d40949dbf501c97827837, (Accessed on 31.01.2021).

FEMA, 2019. National Flood Insurance Program (NFIP) Data Frequently Asked Questions (FAQs). https://www.fema.gov/media-library-data/1562164218054-5da0fdaa74b5ab246c16ceb96f456af4/NFIP_Data_Frequently_Asked_Questions_FAQs.pdf, (Accessed on 31.01.2021).

Feng, Q., Liu, J., Gong, J., 2015. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier–a case of yuyao, china. Water 7 (4), 1437–1455.

Feng, Y., Brenner, C., Sester, M., 2020a. Flood severity mapping from volunteered geographic information by interpreting water level from images containing people: A case study of hurricane harvey. ISPRS Journal of Photogrammetry and Remote Sensing 169, 301–319.

Feng, Y., Brenner, C., Sester, M., 2020b. Learning a precipitation indicator from traffic speed variation patterns. Transportation research procedia 47, 203–210.

Feng, Y., Sester, M., 2017. Social media as a rainfall indicator. In: Societal Geo-Innovation: short papers, posters and poster abstracts of the 20th AGILE Conference on Geographic Information Science. Wageningen University & Research 9-12 May 2017, Wageningen, the Netherlands.

Feng, Y., Sester, M., 2018. Extraction of pluvial flood relevant volunteered geographic information (VGI) by deep learning from user generated texts and photos. ISPRS International Journal of Geo-Information 7 (2), 39.

Feng, Y., Shebotnov, S., Brenner, C., Sester, M., 2018. Ensembled convolutional neural network models for retrieving flood relevant tweets. In: Working Notes Proceedings of the MediaEval 2018 Workshop, Sophia Antipolis, France, 29-31 October 2018. pp. 1–3.

Feng, Y., Tang, S., Cheng, H., Sester, M., 2019. Flood level estimation from news articles and flood detection from satellite image sequences. In: Working Notes Proceedings of the MediaEval 2019 Workshop, Sophia Antipolis, France, 27-30 October 2019. pp. 1–3.

Fohringer, J., Dransch, D., Kreibich, H., Schröter, K., 2015. Social media as an information source for rapid flood inundation mapping. Natural Hazards and Earth System Sciences 15 (12), 2725–2738.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics , 1189–1232.

Friedman, J. H., 2002. Stochastic gradient boosting. Computational statistics & data analysis 38 (4), 367–378.

Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S., Stange, H., 2013. Tracing the german centennial flood in the stream of tweets: first lessons learned. In: Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information. ACM, pp. 31–38.

Fujita, I., Muste, M., Kruger, A., 1998. Large-scale particle image velocimetry for flow analysis in hydraulic engineering applications. Journal of hydraulic Research 36 (3), 397–414.

Genkin, A., Lewis, D. D., Madigan, D., 2007. Large-scale bayesian logistic regression for text categorization. Technometrics 49 (3), 291–304.

Goodchild, M. F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69 (4), 211–221.

Goodchild, M. F., Glennon, J. A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth 3 (3), 231–241.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016a. Section 15.2 – transfer learning and domain adaptation. In: Deep learning. Vol. 1. MIT press Cambridge, pp. 328–343.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016b. Section 5.1 – learning algorithms. In: Deep learning. Vol. 1. MIT press Cambridge, pp. 99–110.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016c. Section 8.3.1 – momentum. In: Deep learning. Vol. 1. MIT press Cambridge, pp. 296–299.

Google, 2016. word2vec. `https://code.google.com/archive/p/word2vec/`, (Accessed on 31.01.2021).

Haberlandt, U., Sester, M., 2010. Areal rainfall estimation using moving cars as rain gauges-a modelling study. Hydrology and Earth System Sciences 14 (2010), Nr. 7 14 (7), 1139–1151.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., 2009. The weka data mining software: an update. ACM SIGKDD explorations newsletter 11 (1), 10–18.

Hanif, M., Tahir, M. A., Khan, M., Rafi, M., 2017. Flood detection using social media data and spectral regression based kernel discriminant analysis. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017.

Hays, J., Efros, A. A., 2008. IM2GPS: estimating geographic information from a single image. In: 2008 ieee conference on computer vision and pattern recognition. IEEE, pp. 1–8.

He, J., Hong, L., Frias-Martinez, V., Torrens, P., 2015. Uncovering social media reaction pattern to protest events: a spatiotemporal dynamics perspective of ferguson unrest. In: International conference on social informatics. Springer, pp. 67–81.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778.

Heipke, C., 2010. Crowdsourcing geospatial data. ISPRS Journal of Photogrammetry and Remote Sensing 65 (6), 550–557.

Herfort, B., de Albuquerque, J. P., Schelhorn, S.-J., Zipf, A., 2014. Exploring the geographical relations between social media and flood phenomena to improve situational awareness. In: Connecting a digital Europe through location and place. Springer, pp. 55–71.

Ho, T. K., 1995. Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. Vol. 1. IEEE, pp. 278–282.

Howe, J., 2006. The rise of crowdsourcing. Wired magazine 14 (6), 1–4.

Hu, Y., Wang, J., 2020. How do people describe locations during a natural disaster: An analysis of tweets from hurricane harvey. In: 11th International Conference on Geographic Information Science (GIScience 2021)-Part I. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017a. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K. Q., July 2017b. Densely connected convolutional networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4700–4708.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W.-J., Zabih, R., 1997. Image indexing using color correlograms. In: Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition. IEEE, pp. 762–768.

Huang, Q., Xiao, Y., 2015. Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. ISPRS International Journal of Geo-Information 4 (3), 1549–1568.

Huang, X., Li, Z., Wang, C., Ning, H., 2019. Identifying disaster related social media for rapid response: a visual-textual fused cnn architecture. International Journal of Digital Earth , 1–23.

Huang, X., Wang, C., Li, Z., 2018a. A near real-time flood-mapping approach by integrating social media and post-event satellite imagery. Annals of GIS 24 (2), 113–123.

Huang, X., Wang, C., Li, Z., 2018b. Reconstructing flood inundation probability by enhancing near real-time imagery with real-time gauges and tweets. IEEE Transactions on Geoscience and Remote Sensing 56 (8), 4691–4701.

Huang, Y., Li, Y., Shan, J., 2018c. Spatial-temporal event detection from geo-tagged tweets. ISPRS International Journal of Geo-Information 7 (4), 150.

Illingworth, S. M., Muller, C. L., Graves, R., Chapman, L., 2014. Uk citizen rainfall network: a pilot study. Weather 69.

Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F., Meier, P., 2013. Extracting information nuggets from disaster-related messages in social media. Proceedings of ISCRAM, Baden-Baden, Germany .

Independent, 2016. Uk weather: London and south east hit by flash flooding – in pictures — the independent. https://www.independent.co.uk/news/uk/home-news/uk-weather-london-flooding-floods-south-east-pictures-forecast-a7097316.html, (Accessed on 31.01.2021).

Iyengar, R., 2015. Facebook has activated safety check in india for the chennai floods. https://time.com/4134203/facebook-safety-check-chennai-flooding-rains/, (Accessed on 31.01.2021).

Jägerbrand, A. K., Sjöbergh, J., 2016. Effects of weather conditions, light conditions, and road lighting on vehicle speed. SpringerPlus 5 (1), 505.

Jia, Y., Wu, J., Xu, M., 2017. Traffic flow prediction with rainfall impact using a deep learning method. Journal of advanced transportation 2017.

Jing, M., Scotney, B. W., Coleman, S. A., McGinnity, M. T., 2016a. The application of social media image analysis to an emergency management system. In: 2016 11th International Conference on Availability, Reliability and Security (ARES). IEEE, pp. 805–810.

Jing, M., Scotney, B. W., Coleman, S. A., McGinnity, M. T., Zhang, X., Kelly, S., Ahmad, K., Schlaf, A., Gründer-Fahrer, S., Heyer, G., 2016b. Integration of text and image analysis for flood event image recognition. In: 2016 27th Irish Signals and Systems Conference (ISSC). IEEE, pp. 1–6.

Joachims, T., 1998. Text categorization with support vector machines: Learning with many relevant features. In: European conference on machine learning. Springer, pp. 137–142.

Johnson, B., 2020. Fake news during disasters putting people's lives at risk, warns intel bulletin. https://www.hstoday.us/subject-matter-areas/emergency-preparedness/fake-news-during-disasters-putting-peoples-lives-at-risk-warns-intel-bulletin/, (Accessed on 31.01.2021).

Joulin, A., Grave, E., Bojanowski, P., Mikolov, T., April 2017. Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Association for Computational Linguistics, pp. 427–431.

Kalliatakis, G., 2017. Keras VGG16 Places365 - VGG16 CNN models pre-trained on Places365-Standard for scene classification. https://github.com/GKalliatakis/Keras-VGG16-places365, (Accessed on 31.01.2021).

Karimi, S., Yin, J., Paris, C., 2013. Classifying microblogs for disasters. In: Proceedings of the 18th Australasian Document Computing Symposium. pp. 26–33.

Karras, T., Laine, S., Aila, T., 2019. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4401–4410.

Kersten, J., Klan, F., 2020. What happens where during disasters? a workflow for the multifaceted characterization of crisis events based on twitter data. Journal of Contingencies and Crisis Management 28 (3), 262–280.

Khare, P., Burel, G., Alani, H., 2018. Classifying crises-information relevancy with semantics. In: European Semantic Web Conference. Springer, pp. 367–383.

Kim, Y., 2014. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1746–1751.

Kingma, D. P., Ba, J., 2015. Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. pp. 1–15.

Klonner, C., Marx, S., Usón, T., Porto de Albuquerque, J., Höfle, B., 2016. Volunteered geographic information in natural hazard analysis: a systematic literature review of current approaches with a focus on preparedness and mitigation. ISPRS International Journal of Geo-Information 5 (7), 103.

Klonner, C., Usón, T., Marx, S., Mocnik, F.-B., Höfle, B., 2018. Capturing flood risk perception via sketch maps. ISPRS International Journal of Geo-Information 7 (9), 359.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105.

Kutija, V., Bertsch, R., Glenis, V., Alderson, D., Parkin, G., Walsh, C., Robinson, J., Kilsby, C., 2014. Model validation using crowd-sourced data from a large pluvial flood. In: 11th International Conference on Hydroinformatics, New York, USA, 17-21 August 2014. CUNY Academic Works.

Lagerstrom, R., Arzhaeva, Y., Szul, P., Obst, O., Power, R., Robinson, B., Bednarz, T., 2016. Image classification to support emergency situation awareness. Frontiers in Robotics and AI 3, 54.

Lam, W. H., Tam, M. L., Cao, X., Li, X., 2013. Modeling the effects of rainfall intensity on traffic speed, flow, and density relationships for urban roads. Journal of Transportation Engineering 139 (7), 758–770.

Laso Bayas, J. C., See, L., Bartl, H., Sturn, T., Karner, M., Fraisl, D., Moorthy, I., Busch, M., van der Velde, M., Fritz, S., 2020. Crowdsourcing lucas: Citizens generating reference land cover and land use data with a mobile app. Land 9 (11), 446.

Laso Bayas, J. C., See, L., Fritz, S., Sturn, T., Perger, C., Dürauer, M., Karner, M., Moorthy, I., Schepaschenko, D., Domian, D., et al., 2016. Crowdsourcing in-situ data on land cover and land use using gamification and mobile technology. Remote Sensing 8 (11), 905.

Le, Q., Mikolov, T., 2014. Distributed representations of sentences and documents. In: International conference on machine learning. PMLR, pp. 1188–1196.

Le Boursicaud, R., Pénard, L., Hauet, A., Thollet, F., Le Coz, J., 2016. Gauging extreme floods on youtube: application of lspiv to home movies for the post-event determination of stream discharges. Hydrological Processes 30 (1), 90–105.

Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R., Dramais, G., et al., 2016. Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in argentina, france and new zealand. Journal of Hydrology 541, 766–777.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L., 1989. Backpropagation applied to handwritten zip code recognition. Neural Computation 1 (4), 541–551.

Li, L., Chen, Y., Yu, X., Liu, R., Huang, C., 2015. Sub-pixel flood inundation mapping from multispectral remotely sensed images based on discrete particle swarm optimization. ISPRS Journal of Photogrammetry and Remote Sensing 101, 10–21.

Li, Q., 2017. Effects of the rainstorm on urban road traffic speed–a case study of shenzhen, china. Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci 42, 2.

Li, Y., Martinis, S., Wieland, M., 2019. Urban flood mapping with an active self-learning convolutional neural network based on terrasar-x intensity and interferometric coherence. ISPRS Journal of Photogrammetry and Remote Sensing 152, 178–191.

Li, Z., Wang, C., Emrich, C. T., Guo, D., 2018. A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 south carolina floods. Cartography and Geographic Information Science 45 (2), 97–110.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, pp. 740–755.

Liu, P., Qiu, X., Huang, X., 2016. Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101 .

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440.

Lopez-Fuentes, L., van de Weijer, J., Bolanos, M., Skinnemoen, H., 2017. Multi-modal deep learning approach for flood detection. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017. pp. 1–3.

Lowry, C. S., Fienen, M. N., 2013. Crowdhydrology: crowdsourcing hydrologic data and engaging citizen scientists. GroundWater 51 (1), 151–156.

Lu, C., Lin, D., Jia, J., Tang, C.-K., 2014. Two-class weather classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3718–3725.

Lyu, H., Sheng, Y., Guo, N., Huang, B., Zhang, S., 2017. Geometric quality assessment of trajectory-generated vgi road networks based on the symmetric arc similarity. Transactions in GIS 21 (5), 984–1009.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J., 2011. Senseplace2: Geotwitter analytics support for situational awareness. In: 2011 IEEE conference on visual analytics science and technology (VAST). IEEE, pp. 181–190.

Maddox, I., 2014. Three common types of flood explained. `http://www.intermap.com/risks-of-hazard-blog/three-common-types-of-flood-explained`, (Accessed on 31.01.2021).

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y., 2018. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sensing 10 (7), 1119.

Manning, C. D., Schütze, H., Raghavan, P., 2008. Section 6.2—term frequency and weighting. In: Introduction to information retrieval. Cambridge university press, pp. 107–109.

Mård, J., Di Baldassarre, G., 2018. Urbanization effects on floods: A global assessment. EGUGA , 13167.

Martinis, S., Kersten, J., Twele, A., 2015. A fully automated terrasar-x based flood service. ISPRS Journal of Photogrammetry and Remote Sensing 104, 203–212.

Massad, I., Dalyot, S., 2015. Towards the production of digital terrain models from volunteered gps trajectories. Survey Review 47 (344), 325–332.

McCallum, A., Nigam, K., et al., 1998. A comparison of event models for naive bayes text classification. In: AAAI-98 workshop on learning for text categorization. pp. 41–48.

McDougall, K., 2011a. Understanding the impact of volunteered geographic information during the queensland floods. In: Proceedings of the 7th International Symposium on Digital Earth (ISDE 7). Western Australian Land Information System, pp. 66–74.

McDougall, K., 2011b. Using volunteered information to map the queensland floods. In: Proceedings of the 2011 Surveying and Spatial Sciences Conference: Innovation in Action: Working Smarter (SSSC 2011). Surveying and Spatial Sciences Institute, pp. 13–23.

McDougall, K., Temple-Watts, P., 2012. The use of lidar and volunteered geographic information to map flood extents and inundation. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 1, 251–256.

Meier, P., 2012. How crisis mapping saved lives in haiti. National Geographic 2.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013a. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 .

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J., 2013b. Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119.

Moniruzzaman, A., Hossain, S. A., 2013. NoSQL Database: New Era of Databases for Big data Analytics-Classification, Characteristics and Comparison. International Journal of Database Theory and Application 6 (4).

Moumtzidou, A., Andreadis, S., Gialampoukidis, I., Karakostas, A., Vrochidis, S., Kompatsiaris, I., 2018. Flood relevance estimation from visual and textual content in social media streams. In: Companion Proceedings of the The Web Conference 2018. pp. 1621–1627.

Mukaka, M. M., 2012. A guide to appropriate use of correlation coefficient in medical research. Malawi medical journal 24 (3), 69–71.

Muller, C. L., 2013. Mapping snow depth across the west midlands using social media-generated data. Weather 68 (3), 82–82.

Muller-Budack, E., Pustu-Iren, K., Ewerth, R., 2018. Geolocation estimation of photos using a hierarchical model and scene classification. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 563–579.

Munich RE, 2017. Topics geo 2016 – natural catastrophes 2016 – analyses, assessments, positions. `https://www.munichre.com/content/dam/munichre/contentlounge/website-pieces/documents/TOPICS_GEO_2016-de3.pdf`, (Accessed on 31.01.2021).

Murthy, D., Longwell, S. A., 2013. Twitter and disasters: The uses of twitter during the 2010 pakistan floods. Information, Communication & Society 16 (6), 837–855.

NASA, 2020. Gpm data downloads. `https://gpm.nasa.gov/data-access/downloads/gpm`, (Accessed on 31.01.2021).

Needham, H., 2017. We're mapping flooded streets in real time. here's how to help. `https://wxshift.com/news/blog/were-mapping-flooded-streets-in-real-time-heres-how-to-help`, (Accessed on 31.01.2021).

Negrey, N., Yang, T., 2018. Serving real-time scikit-learn and XGBoost predictions. `https://cloud.google.com/blog/products/gcp/serving-real-time-scikit-learn-and-xgboost-predictions`, (Accessed on on 31.01.2021).

Nguyen, D. T., Joty, S., Imran, M., Sajjad, H., Mitra, P., 2016. Applications of online deep learning for crisis response using social media information. arXiv preprint arXiv:1610.01030 .

Nielsen, J., 2006. The 90-9-1 rule for participation inequality in social media and online communities. `https://www.nngroup.com/articles/participation-inequality/`, (Accessed on 31.01.2021).

Niessner, R., Schilling, H., Jutzi, B., 2017. Investigations on the potential of convolutional neural networks for vehicle classification based on rgb and lidar data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4, 115.

Ning, H., Li, Z., Hodgson, M. E., et al., 2020. Prototyping a social media flooding photo screening system based on deep learning. ISPRS International Journal of Geo-Information 9 (2), 104.

NOAA, 2018. Costliest u.s. tropical cyclones tables updated. https://www.nhc.noaa.gov/news/UpdatedCostliest.pdf, (Accessed on 31.01.2021).

Nogueira, K., Fadel, S. G., Dourado, Í. C., de Oliveira Werneck, R., Muñoz, J. A., Penatti, O. A., Calumby, R. T., Li, L., dos Santos, J. A., da Silva Torres, R., 2017a. Data-driven flood detection using neural networks. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017.

Nogueira, K., Penatti, O. A., Dos Santos, J. A., 2017b. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition 61, 539–556.

O'Connor, B., Balasubramanyan, R., Routledge, B., Smith, N., 2010. From tweets to polls: Linking text sentiment to public opinion time series. In: ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. pp. 122–129.

Ogie, R. I., Clarke, R. J., Forehead, H., Perez, P., 2019. Crowdsourced social media data for disaster management: Lessons from the petajakarta. org project. Computers, Environment and Urban Systems 73, 108–117.

OpenPose, 2018. OpenPose Demo - Output. https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/02_output.md, (Accessed on 31.01.2021).

Ord, J. K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geographical analysis 27 (4), 286–306.

Panteras, G., Cervone, G., 2018. Enhancing the temporal resolution of satellite-based flood extent generation using crowdsourced data for disaster monitoring. International Journal of Remote Sensing 39 (5), 1459–1474.

Park, S., Baek, F., Sohn, J., Kim, H., 2021. Computer vision–based estimation of flood depth in flooded-vehicle images. Journal of Computing in Civil Engineering 35 (2), 04020072.

Patrini, G., Rozza, A., Krishna Menon, A., Nock, R., Qu, L., 2017. Making deep neural networks robust to label noise: A loss correction approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1944–1952.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. the Journal of machine Learning research 12, 2825–2830.

Pegelonline, 2020. PEGELONLINE. https://www.pegelonline.wsv.de/gast/hilfe#hilfe_pegelparameter_MNW, (Accessed on 31.01.2021).

Pennington, J., Socher, R., Manning, C. D., 2014. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543.

Pereira, J., Monteiro, J., Estima, J., Martins, B., 2019. Assessing flood severity from georeferenced photos. In: Proceedings of the 13th Workshop on Geographic Information Retrieval. pp. 1–10.

Porter, Jon, 2019. Twitter removes support for precise geotagging because no one uses it. https://www.theverge.com/2019/6/19/18691174/twitter-location-tagging-geotagging-discontinued-removal, (Accessed on 31.01.2021).

Prasad, N., Reddy, P. K., Naidu, M. M., 2013. A novel decision tree approach for the prediction of precipitation using entropy in sliq. In: 2013 UKSim 15th International Conference on Computer Modelling and Simulation. IEEE, pp. 209–217.

Quan, K.-A. C., Nguyen, V.-T., Nguyen, T.-C., Nguyen, T. V., Tran, M.-T., 2020. Flood level prediction via human pose estimation from social media images. In: Proceedings of the 2020 International Conference on Multimedia Retrieval. pp. 479–485.

Quinlan, J. R., 1986. Induction of decision trees. Machine learning 1 (1), 81–106.

Quinlan, J. R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Rabiei, E., Haberlandt, U., Sester, M., Fitzner, D., 2013. Rainfall estimation using moving cars as rain gauges-laboratory experiments. Hydrology and Earth System Sciences 17 (2013), Nr. 11 17 (11), 4701–4712.

Ratcliffe, J. H., 2004. The hotspot matrix: A framework for the spatio-temporal targeting of crime reduction. Police practice and research 5 (1), 5–23.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence 39 (6), 1137–1149.

Resch, B., Usländer, F., Havas, C., 2018. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. Cartography and Geographic Information Science 45 (4), 362–376.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.

Rosser, J. F., Leibovici, D., Jackson, M., 2017. Rapid flood inundation mapping using social media, remote sensing and topographic data. Natural Hazards 87 (1), 103–120.

Rözer, V., Peche, A., Berkhahn, S., Feng, Y., Fuchs, L., Graf, T., Haberlandt, U., Kreibich, H., Sämann, R., Sester, M., et al., 2021. Impact-based forecasting for pluvial floods. Earth's Future , e2020EF001851.

Sahoo, D., Pham, Q., Lu, J., Hoi, S. C., 2018. Online deep learning: learning deep neural networks on the fly. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. pp. 2660–2666.

Said, N., Pogorelov, K., Ahmad, K., Riegler, M., Ahmad, N., Ostroukhova, O., Halvorsen, P., Conci, N., 2018. Deep learning approaches for flood classification and flood aftermath detection. In: MediaEval 2018.

Sakaki, T., Okazaki, M., Matsuo, Y., 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. pp. 851–860.

Salton, G., Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. Information processing & management 24 (5), 513–523.

Sander, J., Ester, M., Kriegel, H.-P., Xu, X., 1998. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. Data mining and knowledge discovery 2 (2), 169–194.

Sarafianos, N., Boteanu, B., Ionescu, B., Kakadiaris, I. A., 2016. 3d human pose estimation: A review of the literature and analysis of covariates. Computer Vision and Image Understanding 152, 1–20.

Sarker, C., Mejias, L., Maire, F., Woodley, A., 2019. Flood mapping with convolutional neural networks using spatio-contextual pixel information. Remote Sensing 11 (19), 2331.

Sathiaraj, D., Punkasem, T.-o., Wang, F., Seedah, D. P., 2018. Data-driven analysis on the effects of extreme weather elements on traffic volume in atlanta, ga, usa. Computers, Environment and Urban Systems 72, 212–220.

Sattler, T., Torii, A., Sivic, J., Pollefeys, M., Taira, H., Okutomi, M., Pajdla, T., 2017. Are large-scale 3d models really necessary for accurate visual localization? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1637–1646.

Schnebele, E., Cervone, G., 2013. Improving remote sensing flood assessment using volunteered geographical data. Nat. Hazards Earth Syst. Sci 13, 669–677.

Schnebele, E., Cervone, G., Kumar, S., Waters, N., 2014. Real time estimation of the calgary floods using limited remote sensing data. Water 6 (2), 381–398.

See, L., 2019. A review of citizen science and crowdsourcing in applications of pluvial flooding. Frontiers in Earth Science 7, 44.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. Overfeat: Integrated recognition, localization and detection using convolutional networks. 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings .

She, S., Zhong, H., Fang, Z., Zheng, M., Zhou, Y., 2019. Extracting flooded roads by fusing gps trajectories and road network. ISPRS International Journal of Geo-Information 8 (9), 407.

SIC, 2020. Worldview-3 satellite sensor (0.31m). `https://www.satimagingcorp.com/satellite-sensors/worldview-3/`, (Accessed on 31.01.2021).

Sim, D., 2016a. Germany: Deadly flash floods leave braunsbach buried under rocks, trees and car wrecks. `https://www.ibtimes.co.uk/germany-deadly-flash-floods-leave-braunsbach-buried-under-rocks-trees-car-wrecks-1562851`, (Accessed on 31.01.2021).

Sim, D., 2016b. Germany declares second disaster area as floods kill at least four and devastate towns in bavaria. `https://www.ibtimes.co.uk/germany-deadly-flash-floods-leave-braunsbach-buried-under-rocks-trees-car-wrecks-1562851`, (Accessed on 31.01.2021).

Simon, T., Joo, H., Matthews, I., Sheikh, Y., 2017. Hand keypoint detection in single images using multiview bootstrapping. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 1145–1153.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Singh, K. V., Setia, R., Sahoo, S., Prasad, A., Pateriya, B., 2015. Evaluation of ndwi and mndwi for assessment of waterlogging by integrating digital elevation model and groundwater level. Geocarto International 30 (6), 650–661.

Sit, M. A., Koylu, C., Demir, I., 2019. Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of hurricane irma. International Journal of Digital Earth 12 (11), 1205–1229.

Smith, L., Liang, Q., James, P., Lin, W., 2017. Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. Journal of Flood Risk Management 10 (3), 370–380.

Son, N., Chen, C., Chen, C., Chang, L., 2013. Satellite-based investigation of flood-affected rice cultivation areas in chao phraya river delta, thailand. ISPRS journal of photogrammetry and remote sensing 86, 77–88.

Springenberg, J. T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2015. Striving for simplicity: The all convolutional net. In: 3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings. pp. 1–14.

Stefanidis, A., Crooks, A., Radzikowski, J., 2013. Harvesting ambient geospatial information from social media feeds. GeoJournal 78 (2), 319–338.

Stone, Biz, 2009. Location, Location, Location - Twitter Blog. https://blog.twitter.com/en_us/a/2009/location-location-location.html, (Accessed on 31.01.2021).

Stowe, K., Paul, M., Palmer, M., Palen, L., Anderson, K. M., 2016. Identifying and categorizing disaster-related tweets. In: Proceedings of The fourth international workshop on natural language processing for social media. pp. 1–6.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., Kelling, S., 2009. ebird: A citizen-based bird observation network in the biological sciences. Biological conservation 142 (10), 2282–2292.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI'17. AAAI Press, pp. 4278–4284.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114.

Tancer, Bill, 2007. Look Who's Using Wikipedia. http://content.time.com/time/business/article/0,8599,1595184,00.html, (Accessed on 31.01.2021).

Taylor, S. J., Letham, B., 2018. Forecasting at scale. The American Statistician 72 (1), 37–45.

Tensorflow, 2019. Tensorflow deeplab model zoo. https://github.com/tensorflow/models/blob/master/research/deeplab/g3doc/model_zoo.md, (Accessed on 31.01.2021).

The Guardian, 2016. Europe floods: Seine could peak at 6.5 metres as louvre closes doors. https://www.theguardian.com/world/2016/jun/02/deaths-as-flash-floods-hit-france-germany-and-austria, (Accessed on 31.01.2021).

Thies, J., Elgharib, M., Tewari, A., Theobalt, C., Nießner, M., 2020. Neural voice puppetry: Audio-driven facial reenactment. In: European Conference on Computer Vision. Springer, pp. 716–731.

Thorp, J. M., Scott, B. C., 1982. Preliminary calculations of average storm duration and seasonal precipitation rates for the northeast sector of the united states. Atmospheric Environment (1967) 16 (7), 1763–1774.

Tieleman, T., Hinton, G., 2012. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning.

Tkachenko, N., Zubiaga, A., Procter, R., 2017. Wisc at mediaeval 2017: Multimedia satellite task. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017.

Toshev, A., Szegedy, C., 2014. Deeppose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1653–1660.

Tsapakis, I., Cheng, T., Bolbol, A., 2013. Impact of weather conditions on macroscopic urban travel times. Journal of Transport Geography 28, 204–211.

Twitter, 2020. Q3 2020 Selected Financials and Metrics. https://s22.q4cdn.com/826641620/files/doc_financials/2020/q3/Q3-2020-Selected-Financials-and-Metrics.pdf, (Accessed on 31.01.2021).

Twitter, 2021a. Twitter API. https://developer.twitter.com/en/docs/twitter-api, (Accessed on 31.01.2021).

Twitter, 2021b. Twitter API - Rate limits. https://developer.twitter.com/en/docs/rate-limits, (Accessed on 31.01.2021).

U.S. Census Bureau, 2015. Census Tracts. https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf, (Accessed on 31.01.2021).

U.S. Census Bureau, 2018. Cartographic Boundary Files - Shapefile. https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2018.html, (Accessed on 31.01.2021).

U.S. Census Bureau, 2019. TIGER2019 - AREAWATER. https://www2.census.gov/geo/tiger/TIGER2019/AREAWATER/, (Accessed on 31.01.2021).

USGS, 2020a. How Does the USGS Collect Streamflow Data? https://www.usgs.gov/special-topic/water-science-school/science/how-does-usgs-collect-streamflow-data?qt-science_center_objects=0#qt-science_center_objects, (Accessed on 31.01.2021).

USGS, 2020b. USGS Current Water Data for the Nation. https://waterdata.usgs.gov/nwis/rt, (Accessed on 31.01.2021).

Vo, N., Jacobs, N., Hays, J., 2017. Revisiting im2gps in the deep learning era. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2621–2630.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359 (6380), 1146–1151.

Wang, H., Skau, E., Krim, H., Cervone, G., 2018. Fusing heterogeneous data: A case for remote sensing and social media. IEEE Transactions on Geoscience and Remote Sensing 56 (12), 6956–6968.

Wang, X., Ma, C., Zheng, H., Liu, C., Xie, P., Li, L., Si, L., 2019. Dm_nlp at semeval-2018 task 12: A pipeline system for toponym resolution. In: Proceedings of the 13th International Workshop on Semantic Evaluation. pp. 917–923.

Wang, Y., Wang, T., Ye, X., Zhu, J., Lee, J., 2016a. Using social media for emergency response and urban sustainability: A case study of the 2012 beijing rainstorm. Sustainability 8 (1), 25.

Wang, Z., Ye, X., 2018. Social media analytics for natural disaster management. International Journal of Geographical Information Science 32 (1), 49–72.

Wang, Z., Ye, X., Tsou, M.-H., 2016b. Spatial, temporal, and content analysis of twitter for wildfire hazards. Natural Hazards 83 (1), 523–540.

Wei, Q., Dunbrack Jr, R. L., 2013. The role of balanced training and testing data sets for binary classifiers in bioinformatics. PloS one 8 (7), e67863.

Wei, S.-E., Ramakrishna, V., Kanade, T., Sheikh, Y., 2016. Convolutional pose machines. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 4724–4732.

WMO, 2008. Guide to Hydrological Practices: Vol. I: Hydrology—From Measurement to Hydrological Information, 6th Edition. Vol. 168. World Meteorological Organization, Geneva, Switzerland.

WMO, 2020. WMO Radar Database. https://wrd.mgm.gov.tr/Home/Wrd, (Accessed on 31.01.2021).

WSJ, 2014. Record rains batter long island. https://www.wsj.com/articles/record-breaking-rain-floods-long-island-new-york-region-1407947088, (Accessed on 31.01.2021).

Xiao, Y., Li, B., Gong, Z., 2018. Real-time identification of urban rainstorm waterlogging disasters based on weibo big data. Natural Hazards 94 (2), 833–842.

Xing, Z., Su, X., Liu, J., Su, W., Zhang, X., 2019. Spatiotemporal change analysis of earthquake emergency information based on microblog data: A case study of the "8.8" jiuzhaigou earthquake. ISPRS International Journal of Geo-Information 8 (8), 359.

Yan, Y., Feng, C.-C., Huang, W., Fan, H., Wang, Y.-C., Zipf, A., 2020. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in giscience. International Journal of Geographical Information Science 34 (9), 1765–1791.

Yang, S., Qian, S., 2019. Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents. IEEE Intelligent Transportation Systems Magazine 11 (3), 12–28.

Yang, Z., Cohen, W., Salakhudinov, R., 2016. Revisiting semi-supervised learning with graph embeddings. In: International conference on machine learning. PMLR, pp. 40–48.

Yin, J., Lampert, A., Cameron, M., Robinson, B., Power, R., 2012. Using social media to enhance emergency situation awareness. IEEE intelligent systems 6, 52–59.

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N., 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 325–341.

Yu, F., Koltun, V., 2016. Multi-scale context aggregation by dilated convolutions. 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings .

Yu, M., Huang, Q., Qin, H., Scheele, C., Yang, C., 2019. Deep learning for real-time social media text classification for situation awareness–using hurricanes sandy, harvey, and irma as case studies. International Journal of Digital Earth 12 (11), 1230–1247.

Yurtsever, E., Lambert, J., Carballo, A., Takeda, K., 2020. A survey of autonomous driving: Common practices and emerging technologies. IEEE Access 8, 58443–58469.

Zhang, Y., Wallace, B. C., 2017. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 253–263.

Zhao, Z., Larson, M., 2017. Retrieving social flooding images based on multimodal information. In: Working Notes Proceedings of the MediaEval 2017 Workshop, Dublin, Ireland, September 13-15, 2017.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2017a. Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence 40 (6), 1452–1464.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A., 2017b. Scene parsing through ade20k dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 633–641.

Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., Xu, B., 2016. Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 3485–3495.

Zhou, X., Zafarani, R., 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR) .

Zhu, R., Lin, D., Jendryke, M., Zuo, C., Ding, L., Meng, L., 2019. Geo-tagged social media data-based analytical approach for perceiving impacts of social events. ISPRS International Journal of Geo-Information 8 (1), 15.

Acknowledgements

First and foremost, my deepest gratitude goes to my supervisor Prof. Dr.-Ing. habil. Monika Sester. She has been an excellent role model for me, always inspiring me to be curious about the unknown and full of courage to explore new territories. I'm proud of and very grateful for my time working together with her. Her insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I would also like to thank Prof. Dr.-Ing. Claus Brenner. His comments and fruitful discussions vastly improved the quality of my work.

I would like to extend my sincere thanks to the chair of the examination committee, Prof. Dr.-Ing. habil. Christian Heipke, who guided me a lot during my study in Hannover and supported me for the research stay at the Ordnance Survey in the UK. Besides, I would like to thank Prof. Dr. Alexander Zipf and Prof. Dr.-Ing. Winrich Voß for acting as referees of this dissertation and providing valuable discussions.

I would like to thank all my colleagues at the Institute of Cartography and Geoinformatics and all the friends I met in Hannover. I have a nice time with you together. I enjoy working and discussing with you, as well as sharing the joys and sorrows of life. Special thanks go to my office-mate, Dr. Hao Cheng, who was always available for discussion and provided a lot of help. I would like to thank Dr. Fabian Bock, Dr. Lin Chen, Dr. Paul Czioska, Dr. Udo Feuerhake, Stefan Fuest, PD Dr. Hai Huang, Dr. Sabine Hofmann, Dr. Junhua Kang, Wentong Liao, Dr. Miao Lin, Prof. Dr. Philipp Otto, Dr. Aaron Peche, Florian Politz, Dr. Hu Wu, Dr. Alexander Schlichting, Dr. Xin Wang, Qing Xiao, Wenjun Xie, Chun Yang, Dr. Juntao Yang, and Dr. Xin Zhao for their generous help over the years. I would also like to thank Hans-Georg Mosler, Bing Liu, Jiachun Feng, Mengchen Li, Zihao Fu, and Kuiyan Chen for the many wonderful memories we shared together in Germany.

In addition, I would like to acknowledge the support from the research projects *EVUS* (Real-Time Prediction of Pluvial Floods and Induced Water Contamination in Urban Areas) and *Trans-MiT* (Resource-optimized transformation of combined and separate drainage systems in existing quarters with high population pressure), both funded by the German Ministry of Education and Research (BMBF). Furthermore, the support of GPU donated by NVIDIA Corporation is also gratefully acknowledged.

I would like to thank my parents for their understanding, support, and encouragement so that I can have the opportunity to study in Germany and pursue this goal.

Lastly, I want to thank my wife, Zeng Zhe. With her love and companionship, my life is full of light and joy.

Curriculum Vitae

Personal Information

Name	Yu Feng
Date of Birth	9th October 1989
Place of Birth	Baotou, Nei Mongol, China
Nationality	Chinese

Education

10.2012–12.2015	**M.Sc. Geodesy and Geoinformatics** Leibniz University Hannover, Germany
09.2008–06.2012	**B.Sc. Geographic Information System** South China Agricultural University, Guangzhou, China

Experience

Since 12.2015	**Research Assistant** Institute of Cartography and Geoinformatics, Leibniz University Hannover, Germany
11.2019	**Visiting Researcher** Ordnance Survey, Southampton, United Kingdom
08.2014–11.2015	**Student Research Assistant** Institute of Cartography and Geoinformatics, Leibniz University Hannover, Germany
10.2014–12.2014	**Internship** Institute of Transportation Systems, German Aerospace Center (DLR), Braunschweig, Germany
10.2011–03.2012	**Internship** Guangdong Institute of Eco-environmental Science & Technology (GDEST), Guangzhou, China

Hannover, 22nd June, 2021

Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover

(Eine vollständige Liste der Wiss. Arb. ist beim Geodätischen Institut, Nienburger Str. 1, 30167 Hannover erhältlich.)

Nr. 349	HOBERG, Thorsten:	Conditional Random Fields zur Klassifikation multitemporaler Fernerkundungs daten unterschiedlicher Auflösung (Diss. 2018)
Nr. 350	SCHILLING, Manuel:	Kombination von klassischen Gravimetern mit Quantensensoren (Diss. 2019)
Nr. 351	MILLER, Dominik:	Seismic noise analysis and isolation concepts for the ALPS II experiment at DESY (Diss. 2019)
Nr. 352	ALI, Bashar:	Optimierte Verteilung von Standorten der Schulen unter dem Einfluss des demografischen Wandels am Beispiel Grundschulen (Diss. 2019)
Nr. 353	ZHAO, Xin:	Terrestrial Laser Scanning Data Analysis for Deformation Monitoring (Diss. 2019)
Nr. 354	HAGHIGHI, Mahmud Haghshenas:	Local and Large Scale InSAR Measurement of Ground Surface Deformation (Diss. 2019)
Nr. 355	BUREICK, Johannes:	Robuste Approximation von Laserscan-Profilen mit B-Spline-Kurven (Diss. 2020)
Nr. 356	BLOTT, Gregor:	Multi-View Person Re-Identification (Diss. 2020)
Nr. 357	MAAS, Alina Elisabeth:	Klassifikation multitemporaler Fernerkundungsdaten unter Verwendung fehlerbehafteter topographischer Daten (Diss. 2020)
Nr. 358	NGUYEN, Uyen:	3D Pedestrian Tracking Using Neighbourhood Constraints (Diss. 2020)
Nr. 359	KIELER, Birgit:	Schema-Matching in räumlichen Datensätzen durch Zuordnung von Objektinstanzen (Diss. 2020)
Nr. 360	PAUL, Andreas:	Domänenadaption zur Klassifikation von Luftbildern (Diss. 2020)
Nr. 361	UNGER, Jakob:	Integrated Estimation of UAV Image Orientation with a Generalised Building Model (Diss. 2020)
Nr. 362	COENEN, Max:	Probabilistic Pose Estimation and 3D Reconstruction of Vehicles from Stereo Images (Diss. 2020)
Nr. 363	GARCIA FERNANDEZ, Nicolas:	Simulation Framework for Collaborative Navigation: Development - Analysis – Optimization (Diss. 2020)
Nr. 364	VOGEL, Sören:	Kalman Filtering with State Constraints Applied to Multi-sensor Systems and Georeferencing (Diss. 2020)
Nr. 365	BOSTELMANN, Jonas:	Systematische Bündelausgleichung großer photogrammetrischer Blöcke einer Zeilenkamera am Beispiel der HRSC-Daten (Diss. 2020)
Nr. 366	OMIDALIZARANDI, Mohammad:	Robust Deformation Monitoring of Bridge Structures Using MEMS Accelero-meters and Image-Assisted Total Stations (Diss. 2020)
Nr. 367	ALKHATIB, Hamza:	Fortgeschrittene Methoden und Algorithmen für die computergestützte geodätische Datenanalyse (Habil. 2020)
Nr. 368	DARUGNA, Francesco:	Improving Smartphone-Based GNSS Positioning Using State Space Augmentation Techniques (Diss. 2021)
Nr. 369	CHEN, Lin:	Deep learning for feature based image matching (Diss. 2021)
Nr. 370	DBOUK, Hani:	Alternative Integrity Measures Based on Interval Analysis and Set Theory (Diss. 2021)
Nr. 371	CHENG, Hao:	Deep Learning of User Behavior in Shared Spaces (Diss. 2021)
Nr. 372	MUNDT Reinhard Walter:	Schätzung von Boden- und Gebäudewertanteilen aus Kaufpreisen bebauter Grundstücke (Diss. 2021)
Nr. 373	WANG, Xin:	Robust and Fast Global Image Orientation (Diss. 2021)
Nr. 374	REN, Le:	GPS-based Precise Absolute and Relative Kinematic Orbit Determination of Swarm Satellites under Challenging Ionospheric Conditions (Diss. 2021)
Nr. 375	XU, Wei:	Automatic Calibration of Finite Element Analysis Based on Geometric Boundary Models from Terrestrial Laser Scanning (Diss. 2021)
Nr. 376	FENG, Yu:	Extraction of Flood and Precipitation Observations from opportunistic Volunteered Geographic Information (Diss. 2021)

Die Arbeiten werden im Rahmen des wissenschaftlichen Schriftenaustausches verteilt und sind nicht im Buchhandel erhältlich. Der Erwerb ist zu einem Stückpreis von € 25,00 bei den herausgebenden Instituten möglich.