Chapter 14
Forecasting Internally Displaced People's Movements with Artificial Intelligence

Ayaka Oishi, Takeshi Teshima, Kunikazu Akao, Tsuyoshi Kano, Megumi Kiriha, Naoki Kojima, Takuya Sasaki, Kentaro Takahira, Jumpei Takami, Tomonari Takeuchi, Kenji Tajima, Chihiro Noda, Hikaru Hirose, and Shigeki Yamanaka

Abstract The rise of big data and artificial intelligence (AI) has paved the way for data-driven interventions in the field of international development. In this paper, a group of researchers (i) summarizes policies and implications of the use of advanced technology in the field and (ii) presents the result of the study they conducted which applies machine learning to forecast internally displaced people's (IDPs) movements in the Democratic Republic of the Congo. Despite methodological limitations, the results confer an exposition on how machine learning models and open-source data could enhance the predictive insights of forced displacement. Our approach could be used to predict not only IDP flows but also refugee flows, expanding the use of machine learning for social good. To counter future crises triggered by climate change and the COVID-19 pandemic, we believe our approach has a great possibility to support the effective distribution of limited funds and supplies. This study underscores the benefit of AI and highlights issues in implementations. Future research will need to widen target regions and periods as well as to include the pragmatic aspects of the implementations.

A. Oishi (✉)
Peloria Insights, Washington, D.C., USA
e-mail: ayaka.oishi@peloriainsights.com

T. Teshima
The University of Tokyo, Tokyo, Japan
e-mail: teshima@ms.k.u-tokyo.ac.jp

K. Akao
International Organization for Migration, Freetown, Sierra Leone

T. Kano
University of Michigan, Ann Arbor, MI, USA

M. Kiriha
NTT Data, Tokyo, Japan

R. Boateng et al. (eds.), *Digital Innovations, Business and Society in Africa*, Advances in Theory and Practice of Emerging Markets,
https://doi.org/10.1007/978-3-030-77987-0_14

Keywords Conflict · IDPs · Machine learning · Artificial intelligence · Democratic Republic of the Congo

14.1 Introduction

Forced displacement is a significant challenge to humanity affecting the lives of 79.5 million people worldwide (UNHCR, 2020c). The displaced are vulnerable, especially under the COVID-19 pandemic: they have limited medical access, limited income source under lockdowns, and exposure to the illicit economy. Meanwhile, the recent progress of artificial intelligence (AI) gave rise to new forms of solutions to this humanitarian challenge.

Given the recent development of technology as well as a related policy discussion, the primary objective of this study is twofold: (i) to sketch out the landscape of this novel field and to identify pitfalls based on the desktop research, especially highlighting how international organizations and governmental entities have addressed the issue of forced displacement leveraging the latest technology, and (ii) to experiment the utility of data-driven analysis on the forced displacement to better respond the pitfall we identify through the desktop research.

First, we find that international organizations such as the United Nations High Commissioner for Refugees (UNHCR), International Organization for Migration (IOM), and the World Bank have strongly promoted leveraging the novel technology for their operations. We confer the recent implications of these organizations in both strategic level and programmatic level. At the same time, it becomes clear that such implementations tend to be limited in locations, mainly in the Middle East and

N. Kojima
Link and Motivation, Tokyo, Japan

T. Sasaki · K. Tajima
Koei Research & Consulting, Tokyo, Japan

K. Takahira
Kyoto University, Kyoto, Japan

J. Takami
PASCO, Tokyo, Japan

T. Takeuchi · C. Noda
ICT for Development, Tokyo, Japan

H. Hirose
Peace Tech Foundation, Tokyo, Japan

S. Yamanaka
Actit, Tokyo, Japan

North Africa (MENA), and to involve private data due to the sensitive nature of displacement information, which we assume makes it difficult for researchers to collectively and freely produce insights for other regions than MENA.

Given this imperative, second, we examine the utility of machine learning methods with public data to predict the movement of internally displaced people (IDPs) in the less studied but critical country, the Democratic Republic of the Congo (DRC), where conflicts have continued since the 1990s. This study targets the most war-tone provinces in the eastern region, namely, Ituri, North Kivu, and South Kivu, and predicts the destination of the IDPs associated with conflict events. The dependent variable, the destination goal of IDPs, is obtained from a dataset provided by the Internal Displacement Monitoring Center (IDMC), while the main independent variable, namely, conflict events, is obtained from the Armed Conflict Location and Event Data Project (ACLED).[1]

Methodologically, we employ supervised machine learning to make predictions of IDP flows. We observe that, by incorporating more training data, the model's performance shows an acute improvement in two measures, namely, precision and recall. Although our result itself does not confer an actionable intelligence to decision-makers, it underscores the possibility to predict IDP flows associated with political violence and the importance of making more data available.

Based on these findings, we argue that the implications of data-driven approach in the international development which we find to currently be biased can be better adjusted by combining machine learning and open-source data. This study also sheds lights on the potential use case where IDP predictions rationalize the resource allocation under the current environment of pressing demand for humanitarian aid caused by the disaster, pandemic, or climate change.

Following the introduction, this paper provides the landscape of the use of advanced technology in the field of international development in Sect. 14.2 and highlights the key issues that are overlooked by previous studies in Sect. 14.3. After providing an overview of the conflict in our target country, the DRC, in Sect. 14.4, we provide detailed descriptions of methodologies and variables in Sect. 14.5. Following the empirical results presented in Sect. 14.6, this paper provides discussions and possible implications in Sects. 14.7 and 14.8 and concludes in Sect. 14.9.

[1]We also explored the potential of satellite imagery in tracking human movements due to our assumption that it may lead to additional dataset for human displacement. The current findings are still limited so that we report our attempt in the appendix.

14.2 The Landscape of the Use of Advanced Technology in International Development

The recent technological advancement of AI and big data has transformed many aspects of our lives. The field of international development is not an exception: as of 2019, 36 UN agencies have investigated and experimented how to improve their response to global challenges leveraging the new technologies (International Telecommunication Union, 2019). The trend is observed both in the high-level strategies of international actors and in the number of implications produced by academics and practitioners.

This section outlines the recent development of AI-related policies in the field of international development and confers some examples which specifically address forced displacement issues by using technology. We find that international organizations tend to have high-level strategies where they envision new intervention and evaluation mechanisms, while bilateral aid organizations tend to have limited or no such master plans.

14.2.1 Strategies

International organizations and bilateral aid organizations have made strategies on (1) supporting refugees/IDPs and (2) using data for better intervention. As a result of persecution, conflict, violence, human rights violations, or other events disturbing public order, the number of forcibly displaced people worldwide is estimated to be up to 79.5 million (UNHCR, 2020c).

In September 2016, the United Nations General Assembly decided to develop a global compact for safe, orderly, and regular migration. One year after GCM, the United Nations General Assembly affirmed the Global Compact on Refugees (IOM, 2016). These are the framework for more predictable and equitable responsibility-sharing, recognizing for governments, and international organizations. Even though these Global Compacts do not have compelling force, they still have certain impact on the international agencies as they incorporate the migration/refugee issue in their main strategies. For instance, the IDF committed 2.2 billion US dollars to refugees and host communities (World Bank, 2020).

Each multilateral organization is keen to use advanced technology as a solution. The World Bank has a new strategy for systematizing digital solutions in fragile and conflict-affected situations (World Bank, 2020). The UNHCR also plans to actively pursue innovative ways to amplify refugees' voices and take advantage of new technologies to enhance our ongoing dialogue with them and their connectivity with the global community. The IOM recently launched the Migration Data Portal as a unique access point to timely comprehensive migration statistics and reliable information about migration data globally (Message from Iraqi friends, Kodansha (2003), Tokyo Japan).

Concerning the refugee policies in bilateral aid organizations, whether they employ new technology depends on the situation that each country faces regarding asylum seekers and refugees (Czaika & Mayer, 2011). For example, Western countries tend to allocate their funds for prevention measures to the sending countries of asylum seekers rather than for IDPs to control the refugee migration to their countries, while Japan focuses on emergency funds due to less influence by asylum seekers (Czaika & Mayer, 2011).

As for the approach for data-driven interventions in bilateral aid organizations, some organizations have published digital strategies that focus not only on the benefits but also on the risks of data-driven development. While digital technology such as AI and machine learning is expected to strengthen accountability, transparency, and public engagement in humanitarian assistance (USAID, 2020; DFID, 2018), the USAID (2020) claims that international stakeholders need to discuss (1) how to protect the privacy of vulnerable populations, including IDPs and refugees, and (2) how to prevent amplifying biases that might be present in historical data when they are used for detecting patterns and making predictions.

14.2.2 Implications

Corresponding to such high-level strategies, the number of implementations of the advanced technology for international development has skyrocketed. In the refugee/IDP area, precedents of varying purposes and methodologies underscore the significance of the potential contribution enabled by technology. Table 14.1 categorizes such previous projects into five groups based on their objectives and methodologies.

Category 1: Object Detection Using Satellite Imagery
This category's objective is to visualize physical objects associated with, for instance, refugee/IDP tents and shelters or damaged sites due to natural disasters. The majority of projects in this category leverage remote sensing data and AI. Machine learning such as autoencoders is used for object detection in images, while satellite imagery is used to count various kinds of structures, for instance, tents and shelters of different materials, in refugee and IDP sites. Without automation, these tasks can be labor-intensive and time-consuming, and they may require experienced experts for adequate decision-making to respond to people's needs in crisis. Conversely, by leveraging the advanced technology, the United Nations Global Pulse and United Nations Institute for Training and Research built a deep learning model to automatically count the numbers of structures in multiple refugee camps in Africa and the Middle East (Quinn et al., 2018). A similar research was conducted by using deep learning for effective refugee tent extraction near Syria-Jordan border (Lu et al., 2020). In Bangladesh, satellite imagery and random forest algorithms are used to quantify deforestation caused by the establishment of Rohingya refugee camps (Hassan et al., 2018). There is also an activity among the

Table 14.1 Precedents using AI technologies for the refugee/IDP

#	Objective	Overview	Project example
1	Object detection using satellite imagery	Analysis of objects (tents, facilities, forests, geographic structures, etc.) in the satellite images. The main focus is on the accuracy of classification	Geographic Analysis of Refugee Camps in Africa and the Middle East by the United Nations Global Pulse and United Nations Institute for Training and Research (Quinn et al., 2018)
			Landslide Risk Analysis of Refugee Camps in Bangladesh by North South University (Ahmed et al., 2020)
			Deforestation through the establishment of refugee camps in Bangladesh by University of Florida and Virginia Tech (Hassan et al., 2018)
2	Visualization of migration routes chosen by refugees/IDPs	Visualization of the routes people use for traveling from one place to another, etc. Use Twitter, mobile usage data, and a variety of other data	Stop Corona Virus in the DRC by the government, GSMA, and other partners (GSMA, 2020)
			Distribution of city residents and travelers of Twitter data in the world by Mapbox (Migration Data Portal, 2020)
			Immigration in Facebook Data in the world by Facebook (Facebook, 2013)
			Trace Together in Singapore by the Government (Government of Singapore, 2020)
3	Prediction of the destinations and forecasts of trends in displacement of refugees/IDPs	Find patterns in the movement of people and their lives at the destination, and based on this, predict what kind of life can be lived under what conditions and where to move to for an ideal life, etc.	GeoMatch by Immigration Policy Lab (Immigration Policy Lab, 2020), Annie MOORE (Matching Outcome Optimization for Refugee Empowerment) used by HIAS in the USA and other agencies in the UK and Sweden (Teytelboym, 2020)
			Jetson in Somalia by the UNHCR Innovation Service (UNHCR Innovation Service, 2019; Parater, 2020)

(continued)

Table 14.1 (continued)

#	Objective	Overview	Project example
4	Sentiment analysis of both refugees/ IDPs and host communities	Analysis of the psychological conditions of host communities, refugees and migrants, refugee children, etc. and the correlation between behavior and psychological factors	Analysis of Psychological Factors in the Decision to Return Refugees in Syria by the World Bank (The World Bank, 2019)
			Correlation analysis of children's pictures and psychological trauma of Syrian refugees in Iraq, Jordan, and Lebanon (Baird et al., 2020)
			Sentiment analysis of host communities (in Europe) toward refugees and refugee reception in Greece by the UN Global Pulse and UNHCR (UN Global Pulse, 2017)
5	Improvement of work efficiency	Use of AI to improve business efficiency, e.g., to assist prioritization of tasks	Improving operational efficiency with AI for asylum procedures in refugee support groups in the USA by Microsoft, ASAP, and KIND (Spelhaug, 2019)

global data science community named "The Crowd AI mapping challenge" that aimed at detecting and mapping buildings using satellite imagery for humanitarian response (Crowd AI, 2018).

Category 2: Visualization of Migration Routes Chosen by Refugees/IDPs
The objective is to visualize human movement. Projects in this category often employ mobile data and SNS usage as well as AI. In fact, there are growing applications of mobile usage data for international development because tracking human movements is critical in the recent COVID-19 crisis. In the DRC where the national census has not been conducted since 1984, the government, GSMA, several mobile companies, and other partners implemented a project to analyze mobile usage data and visualize the population mobility trends, which also can be used for projects to prevent an epidemic (GSMA, 2020). Similarly, SNS data is also seen to be highly promising in the field of international development. Facebook conducted a project to map internal and international migrations using aggregated and anonymized Facebook profiles on departure and destination locations (Facebook, 2013). Mapbox, a company providing mapping platform services, supported a project to visualize human mobility within several major cities by using Twitter data and by updating the maps automatically using AI (Migration Data Portal, 2020).

Category 3: Prediction of the Destinations and Forecasts of Trends in Displacement of Refugees/IDPs
The main objective of this category is to predict forced displacement and to forecast future trends in human mobility. Some projects in this category also aim to identify host communities that best fit the refugees. Data used in this type of analysis varies from population data, local market data, climate indicators, to conflict dataset. For

example, in an attempt to predict displacements in Somalia, the UNHCR Innovation Service built a tool called Jetson, which is described as “a machine-learning based application [that] measures multiple variables to see how changes over time that affect movement of UNHCR’s persons of concern, particularly refugees and internally displaced people” (UNHCR Innovation Service, 2019; Parater, 2020). Also, HIAS, a refugee resettlement agency of the USA, created a software called Annie™ MOORE. It uses “advanced machine learning and state-of-the-art integer optimization methods” to support the agencies that work on the appropriate placement of refugees to raise the chances for their employment and to ensure their access to services to meet their needs (Teytelboym, 2020).

Category 4: Sentiment Analysis of Both Refugees/IDPs and Host Communities

This category includes projects that analyze the human sentiment of both refugees/IDPs and people in the host communities and those that serve to improve the quality of lives of the displaced population after settlement. They obtain data from multiple sources, such as online media and children’s paintings. For example, one research reviewed refugee children’s drawings and metadata to understand the correlation between their psychological well-being and the exposure to violence experienced during displacement. Also, another study analyzed Twitter data to see the interactions among refugees as well as to understand the xenophobia sentiment in host communities toward displaced populations. There is also a study that uses survey data collected by humanitarian agencies such as the UNHCR and analyzes them by developing models with machine learning techniques.

Category 5: Improvement of Work Efficiency

This category is quite different from the others in terms of the objective and the use of AI. While others track movements and lives of refugees/IDPs to help them, this category focuses on helping an operation process in the organizations that support asylum seekers. In a collaboration between Microsoft and two NPOs called Asylum Seeker Advocacy Program (ASAP) and KIND, they took advantage of Microsoft’s AI tools, such as speech-to-text artificial intelligence and an Azure-based database, to improve their work efficiency. It helped the staff members of ASAP to “efficiently track changing court dates and prioritize cases most in need of emergency legal services” in delivering legal assistance service to asylum seekers (Spelhaug, 2019). This is a unique and great example of multi-sector collaboration, where various parties bring their knowledge and technologies together to achieve a common goal.

14.3 Issues of Past Studies

Despite the enormous preceding contributions to the realm of international development, the past studies remain to have key issues. First, most of the implementations are limited to specific areas such as MENA or some host countries such as the

USA. The projects mentioned in Sect. 14.2 were largely conducted in Yemen, Somalia, Iraq, or Syria, whereas only a few examples target other countries such as Bangladesh or Nigeria. This bias makes sense as countries in MENA suffer refugee and IDP issues due to their unstable political dynamics. However, it does not explain why these cases overlook other critical regions such as sub-Saharan Africa or Latin America that also produce high numbers of forced displacements. For example, the number of forcibly displaced people in the DRC is the second highest in the world (UNHCR, 2020a). This selection bias may be partly because their ongoing conflicts make it hard to get quality data of displacement flow. Moreover, the forested geographical condition of these regions might limit the usability of satellite imagery. However, we argue that those less-focused countries are the ones that need to be better served by advanced technologies. Their long-lasting humanitarian crises not only cost millions of lives, but it also destabilizes the whole region.

Second, these precedents often rely on closed data that is not publicly available. For instance, Project Jetson led by the UNHCR made a strategic partnership with other institutes such as the World Meteorological Organization, the Met Office in the UK, academia, and other UN institutions to access proprietary data (UNHCR Innovation Service, 2019). In another case, a group of data scientists at Omdena leveraged satellite imagery and granular displacement data provided by the UNHCR to measure and predict displacement mobility in Somalia. Again, this data was not publicly available. The dependence on closed data serves its purpose to secure privacy and to ensure ethical aspects of the project deliverables. Yet, it ironically inhibits researchers to conduct similar studies and build collective knowledge in the field.

14.4 Part II

Acknowledging the issues raised by the precedents in part I, this part of the study conducts an empirical analysis by targeting a less-studied country and only using open-source data and investigates the utility of machine learning methods on the topic of forced displacement. This study choses the DRC as the main target country given the high number of forced displacement and humanitarian crisis of IDPs.

14.4.1 IDP Crisis in the DRC

The number of forcibly displaced persons in the DRC is the second highest in the world (UNHCR, 2020a). Most of these displacements are due to the long-lasting conflicts in the eastern region, notably North Kivu, South Kivu, and Ituri provinces. This section confers the overview of conflicts in the country and the key drivers of forced displacement.

14.4.4 Overview of Displaced Persons: Refugees and IDPs

The displacement situation in the DRC, with more than 630,500 Congolese refugees and 4.49 million IDPs, is one of the most complex, challenging, prolonged, and forgotten crises in the world (UNHCR, 2018a). The number of IDPs has doubled since 2015, and approximately 428,000 persons have been displaced in the 3 months between October and December 2017 alone. In 2017, 120,000 Congolese fled to neighboring countries as refugees, namely, Uganda, Angola, Zambia, the United Republic of Tanzania, Burundi, the Republic of Congo, and Rwanda, joining the 510,000 already in exile. In addition, several thousand have also fled to southern Africa and to other countries such as the Central African Republic, Chad, Kenya, and South Sudan and even to the outside of Africa (UNHCR, 2018a).

14.4.5 IDP Situation in Three Provinces

Numerous armed groups continue to fight the FARDC and the United Nations Stabilization Mission (MONUSCO), causing significant harm to civilians. Intense fights in the Beni, Masisi, Rutshuru, and Lubero areas of North Kivu led to a large-scale displacement of people who escaped to neighboring provinces to become IDPs. In South Kivu, fighting in the Fizi and the Uvira areas led to the displacement of 50,000 people in 3 months in 2019. In addition, conflicts between the Lendu and Hema communities in Ituri province caused a mass displacement in early 2018 and June 2019, and Lendu militias attacked the national army and the civilians (UNHCR, 2019). As a result of these fightings, overall, 4.6 million IDPs have evacuated within the three provinces: 1.95 million IDPs in North Kivu, 0.98 million IDPs in South Kivu, and 1.67 million IDPs in Ituri (UNHCR, 2020b). In spite of a large population of IDPs, 93% of them are accommodated by host families and communities. Only 5% of the IDPs in North Kivu live in IDP sites and 7% in Ituri. There are no official IDP sites in South Kivu but spontaneous sites where 4% of IDPs live exist (UNHCR, 2019). This divergence of IDP locations renders most IDPs invisible to aid providers and makes it hard to deliver the support. The living condition of IDPs is severe: 34% of Ituri's population depends on some kind of humanitarian aid (Norwegian Refugee Council, 2020). Schools and hospitals are often attacked by armed groups, causing instability in the host communities.

14.5 Methodology

To explore further possibilities of data analysis for conflict resolution, we consider predicting the destinations of IDP movements associated with conflict events and human rights violations.

14.4.2 Human Rights Violation in Three Provinces

The conflicts in the eastern region, North Kivu, South Kivu, and Ituri, date back to 1993, when the DRC was the Republic of Zaire. Various human rights violations have occurred in the eastern region since 1993, caused by three major factors, the failure of the democratization process in the Republic of Zaire, the influence of Rwanda genocide, and the Second Congo War (OCHA, 2010), and the conflict is still continued. In fact, the OCHA/UNJHRO (2020) reported that the cases of human rights violations from January to June 2020 in these three provinces were 1864 cases in North Kivu, 680 in Ituri, and 475 in South Kivu, marking the highest in the country. In addition, more than 1300 people (655 victims in Ituri, 617 victims in North Kivu, and 79 victims in South Kivu) were killed by armed perpetrators during the same period. The UN Security Council (2019) also condemns other types of human rights violations other than killing such as the recruitment and the use of child soldiers in armed groups, conflict-related sexual violence, and attacks against civilians.

14.4.3 Armed Groups and Mines in Three Provinces

A recent study estimated that approximately 120 armed groups operate in North and South Kivu provinces and that most of them are small-sized and primarily ethnocentric (Stearns & Voge, 2017). It has been reported that 57% of the human rights violations are committed by such non-state armed groups, e.g., Djugu-based armed assailants, FDLR, Nyatura, or Mayi-Mayi, while the other 43% are by state authorities including the Armed Forces of the Congo (FARDC) and Police Nationale Congolaise (PNC) (OCHA & UNJHRO, 2020). This dual nature of perpetrators, namely, the government and the non-government agents, is one of the unique features of conflicts in the DRC. This paper analyzes the political violences by both types. In addition, since the armed groups are mostly formed by ethnic or tribal identity, we take ethnic distribution into consideration. Another feature of this conflict is the deep relations to conflict minerals: tin, tantalum, tungsten, and gold (3TG). Surprisingly, 35.7% of total gold mines in the eastern region, 15.3% of tin, 11.3% of tantalum, and 9.5% of tungsten are estimated to be associated with either the DRC army or non-state armed groups (Hanai, 2019). Their illegal mining activities are a large part of their financial resources, while such activities themselves accelerate conflicts. Therefore, our analysis includes the locations of the mining sites, assuming that they influence refugee and IDP movements.

14.5.1 Data Sources

Here, we describe the data sources we employed for generating the predicted/predictor variables of our experiment. See Sect. 14.6.2 for the list of generated variables and how they were obtained.

Internal Displacement Updates (IDU): this dataset was provided by the IDMC via private communication. The IDU data is a collection of IDP movement information curated by the IDMC as a secondary data source collected from multiple primary sources, namely, the Office for the Coordination of Humanitarian Affairs (OCHA), United Nations High Commissioner for Refugees (UNHCR), IOM Displacement Tracking Matrix (IOM DTM), Intersos, European Civil Protection and Humanitarian Aid Operations (ECHO), and Radio Okapi. The IDU data contains the start date, the end date, and the destination of each reported IDP movement, along with other information. We first filtered out the records whose destination information was missing, and then we aligned the granularity of the destination to the same administrative level, namely, territory, for all the records. We treated the destination as the predicted variable. Since some records had few but multiple destinations, the predicted variable is essentially a list of destinations instead of a single destination. The number of destination territories that appeared in IDU data was 16. The preprocessed data contained records from Nov. 15, 2018, to Sept. 07, 2020.

Armed Conflict Location & Event Data (ACLED): this dataset was collected from the ACLED project website (Raleigh et al., 2010). It is a collection of the dates, actors, locations, fatalities, and types of political violence reported since 1997. ACLED is one of the most commonly used datasets in the peace and conflict study. While it has been pointed out to have a methodological limitation causing possible reporting bias and uneven quality in the dataset (Eck, 2012), this study relies on ACLED as it provides political events recorded at an ideal granularity. Each event has a pair of longitude and latitude, which was later used for calculating auxiliary input features for the predictors such as the routing distance from the conflict location to the destination candidates.

OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world (OpenStreetMap Contributors, 2017). We use the data from OSM for reverse geocoding (i.e., looking up addresses by their global coordinates) via Nominatim (2012/2020), an open-source search engine for OSM data.

Open Source Routing Machine (OSRM) is an open-source router designed for use with the OSM data (Open Source Routing Machine, 2011/2020). It allows us to compute the walking distance and the walking duration from one location to another, specified by their longitudes and latitudes.

Territory Ethnic Composition (TEC) data: this dataset shows the different ethnic groups of the DRC by territory, published by the American Red Cross based on the work of Abbe Leon de Saint Moulin (American Red Cross, 2019/2020). The ethnic composition of each territory is referred to as the destination characteristics.

Subnational Population Statistics (SPS): this dataset shows subnational population in the DRC by subnational level, country, province, and territory in 2019 and 2020, contributed by the UNOCHA (OCHA DR Congo, 2019).

Mining Site Information (MSI): this dataset shows the mining site locations and the number of workers at the time recorded from Jan. 02, 2009, to July 19, 2019 (International Peace Information Service, 2019). We only use the latest information when multiple records exist for the same mining site.

14.5.2 List of Variables and Feature Engineering

We considered each IDU record as a data point, and we attached some additional data to each record as the predictor variables.

Predicted Variable We regarded the destination column of the IDU data to be the predicted variable. The variable is essentially a list of the territory names which are the reported destinations of the IDPs. Technically, the list of the territory names was converted to a fixed-length array of 0 or 1, indicating whether each candidate territory was a destination (1 if selected as a destination, 0 otherwise). The list of all candidate territories was determined by taking the unique list of territories that appeared in the IDU dataset as the destination.

Destination-Conflict Cross Features To create the input variables reflecting the information of the conflicts that occurred prior to each IDU record, we first associated the recent conflict records of ACLED, namely, those which reportedly occurred within 2 weeks before the reported starting date of the IDP movement. The number of conflict records associated with each IDU record ranged from 33 to 113, excluding those IDU records which were not assigned any conflict records. From the list of recent conflicts associated with each IDU record, we created the following predictor variables.

Routing distance: obtained using IDU, ACLED, and OSRM. For each recent conflict record associated with the IDU record, we obtained the representative coordinate of the conflict from ACLED, and we calculated the walking distance and duration from the coordinate to the representative coordinates of the destination candidate territories. We use the minimum walking distance and the minimum walking duration as predictive features out of those for the associated conflicts.

Ethnic composition dissimilarity: obtained by combining TEC and OSM. We first mapped the conflict coordinates to the corresponding territories by reverse geocoding using OSM. Then, the territories were converted to a list of its major ethnic groups by looking up in the TEC. In the end, to each pair of conflict record and destination candidate, we assigned 0 or 1 indicating the dissimilarity in the ethnic composition of the two territories (0 if the two territories share at least one ethnic group is common in their ethnic compositions and 1 if none is common. If

TEC has a missing record for the territory, we define this dissimilarity to be 1). We take the sum of the binary variable over all the associated conflict events to form the predictive feature.

Destination-Specific Features

Territorial population: obtained from SPS. We use the population at territorial level in 2019 as the destination feature, assuming that the population partly explains the cultural or economical importance of that area.

Wikipedia page importance: obtained from OSM via Nominatim. To each destination candidate territory, we assigned an importance score (in the range [0, 1]) calculated by Nominatim. The score is based on the page-link information on Wikipedia (Wikipedia-Wikidata, 2020/2020).

Mining site information: obtained by MSI via simple aggregation. As described in Sect. 14.4, the mining industry has been a key conflict driver in the eastern DRC. It is a general understanding in the field of civil war that mining concessions foster political violence in the region, hence resulting in the increase of IDPs (Sekeris et al., 2013). Based on this assumption, we count the mining sites and sum up the workers by each territory and assign these aggregation values to the corresponding territory as their features.

14.5.3 Integrated Dataset

We constructed our dataset by collecting the variables explained in Sect. 14.4.2. Among all IDU records, we disregarded the ones whose associated set of recent conflicts was empty. The resulting dataset consisted of 52 records of 8 variables (7 predictive features and 1 predicted variable):

1. *Walking distance* from recent conflict locations (aggregated over the recent conflict events by taking the minimum)
2. *Walking duration* from recent conflict locations (aggregated in the same way as the walking distance)
3. *Ethnic composition dissimilarity* (aggregated over the recent conflict events by taking the summation)
4. *Territorial population*
5. *Wikipedia page importance*
6. *Total number of mining sites*
7. *Total number of workers in the mining sites*
8. (Predicted variable) binary vector indicating whether each of the 16 candidates, namely, Aru, Beni, Bunia, Djugu, Irumu, Kabalo, Kalehe, Kisangani, Lubero, Mahagi, Masisi, Moba, Mwenga, Nyunzu, Rutshuru, and Walikale, was a destination. Each variable except the predicted variable and the ethnic composition dissimilarity was normalized by subtracting the average and dividing the standard deviation. The ethnic composition dissimilarity was normalized by dividing the value of one of the randomly selected destinations, namely, Nyunzu.

14.5.4 Problem Setup and Evaluation Metric

Here, we describe how we formulated the destination prediction as a machine learning problem setup.

The Multi-label Classification Problem Technically, we formulated the problem as multi-label classification (Tsoumakas et al., 2010). It is a variant of the classification problem where multiple labels are assigned to each instance. More concretely, we formulated the problem as predicting a binary vector indicating whether each of the destination candidates was selected as destination ("0" indicating "not selected" and "1" indicating "selected") for each record of IDU.

Evaluation Metric To evaluate the performance of the algorithm, we combine two evaluation metrics called precision and recall, devised for evaluating a rare phenomenon in the fields such as weather forecasting and information retrieval. We use these metrics since our multi-label classification is imbalanced, i.e., for each record of IDU, only a few of the destination candidates have the truth value "1." In such a case, it is easy to achieve more than 80% accuracy by predicting "0" for all destination candidates, which is unreasonable because this trivial predictor is not informative. On the other hand, the precision and the recall of this trivial predictor would be both 0, appropriately evaluating the uninformative predictions.

Precision and recall are both evaluation metrics of prediction results in binary classification. Precision is defined as the ratio of the correctly predicted instances among those for which the prediction was "1" (or positive), and recall is defined as the ratio of the positively predicted instances (i.e., those for which the prediction was "1") among those whose true label is "1." Together, these two metrics evaluate the characteristics of our algorithm: how it is accurate while maintaining sensitivity. We provide additional figures showing the results in terms of a few other performance metrics in Appendix E.

14.5.5 Model and Training

Model selection and corresponding parameter estimations are crucial for predicting the movement flow. Descent parameters can be obtained by optimizing the parameters to fit the model to explain the obtained data. We refer to this optimizing process as training.

Model We adopted a machine learning-based approach because of its appropriate characteristics to the refugee movement forecast. Compared to agent-based models like the ones proposed by Suleimenova et al. (2017), ML-based approaches are resilient with the sudden changes in a data structure (e.g., some features which were available in the modeling process cannot be observed when the model is in operation). This characteristic is essential in that the organizations generally need rapid predictions of refugee movements when conflicts occur. Besides, these characteristics make the prediction model more flexible to incorporate an alternative dataset which can be a good indicator for the refugee movement but whose relationship with the people movement is not always clear (e.g., Twitter logs at the time of conflicts).

To produce a prediction as precise as possible while maintaining the above characteristics, we employ eXtreme Gradient Boosting, or XGB (Chen & Guestrin, 2016), which is a commonly used machine learning algorithm based on gradient boosted trees. XGB is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. We chose this supervised learning model based on the following two reasons. First, it is empirically shown that XGB tends to produce high performance with structured tabular data like our dataset (Chen & Guestrin, 2016). Second, it is relatively easier to interpret the result compared to other nonlinear techniques like deep neural networks. The easiness of interpolation is often essential in the context where the model results lead to political decisions. To predict the whole destination choice of IDPs (expressed by a binary vector), we construct an XGB model for each territory to learn the probability of which each territory is chosen as a destination and combine their predictions into a binary vector expressing the whole destination selection.

Training and Experiment Procedure To train or find decent parameters, we split our dataset into test dataset and training dataset and feed only the training dataset into our model. We first randomly selected 30% of the 52 records and held it out as the test data. Out of the remaining 70%, in order to see the performance for different training data sizes, we randomly selected a varied number of data to use for training on the way how the training data size increased incrementally. The loss function of the model is logistic loss which is composed from the summation from all training data samples. Our model learns the parameters which minimize this objective function. We use the default hyperparameters specified in the XGBoost GitHub repository (XGBoost, 2014/2020), namely, the max depth of the tree is 6, the learning rate is 0.3, and the minimum child weight is 1. We repeated the initial procedure of test split 10 times to incorporate different train-test splits. After each test split, we varied the training data size in {5, …, 36} and randomly selected the specified number of data to train on.

14.6 Empirical Results

14.6.1 Performance Improvement for a Varied Number of Training Data

Figure 14.1 shows the performance improvement of the trained predictor trained on varied train data sizes. In both precision and recall, we can see that the performance improves in both metrics as we use more training data. The highest scores achieved at the training data size 30 are 0.175 for the precision and 0.186 for the recall.

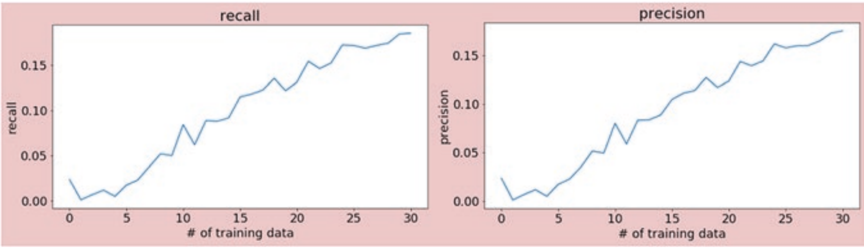

Fig. 14.1 Improvement of performance metrics (precision and recall; the higher the better) for varied sizes of training data

14.6.2 Implications

The values of the highest scores indicate that, with the current availability of data, we can build an alert system that can be trusted approximately only once in six alerts and misses 80% of the cases that should be alerted. Even though this result is superior to a random classifier's theoretical results (F1 scores of our model and a random classifier are 0.18 and 0.12 each), these performances are certainly unsatisfactory for actual deployment of the system to the real world. Nevertheless, the parallel upward trend in the evaluation metrics is clear, meaning that we can anticipate building a predictor that achieves high sensitivity (recall) and precision together by collecting more data.

14.7 Limitation and Caveat

While the results confer a robust exposition on how machine learning models and open-source data could enhance the predictive insights of forced displacement, it should be underscored that there are methodological challenges in our approach. First, it is worth noting that the data quality of displacement data should be improved in many ways. This study relies on the IDMC dataset to measure the flow of IDP in the DRC, as it was the most holistic and granular data source the research needed. Yet, the data the IDMC provides is secondary information, which is the collection of primary information provided by other organizations. This implies that the quality and methods to datafy IDP flow depend on the various primary providers and the numbers could be inconsistent based on who and how to count them. Although the IDMC does provide robust quality assurance, the possible misrepresentation of IDP data cannot be deniable. Furthermore, even with first-hand sources, measuring the number of moving persons in the midst of conflicts is fundamentally challenging, and there is no single way to be a hundred percent accurate on counting. This reminds researchers that the findings of this type of study, which heavily rely on data, are insights reflecting a counted data, and not necessarily the representation of the real phenomena on the ground. Together with the methodological advancement of modeling techniques, the sophistication of the datafying process is highly demanded.

Secondly, although the study indicates the significant improvement in predictive performance, it does not confer a high predictive accuracy to the extent where our models can be applied to real decision-making situations. With the greater size of data, future studies shall validate our methodological approach and could inform practitioners and decision-makers in actual circumstances.

14.8 Policy Implications

This paper presents great extensibility for implications that can serve practitioners at international organizations or governments to better respond and manage humanitarian challenges that IDPs face. Moreover, our approach could be used to predict not only IDP flows but also refugee flows, expanding the use of machine learning for social good. In fact, the UNHCR points out that its ineffective protection monitoring mechanism currently causes delay or a failure of service delivery to forcibly displaced persons and that it could be mitigated by building and leveraging an early warning system (UNHCR, 2018b). This shows the high demand of predictive insights. The long-term predictions may also rationalize the resource allocation that host countries require. Although it was not in the scope of our research, predicting the numbers of IDPs or refugees may provide a rationale for donors and host communities to determine how much resources, e.g., financial, material, and human resources, are required. To counter future crises triggered by climate change and the COVID-19 pandemic, we believe our approach has a great possibility to support the effective distribution of limited funds and supplies.

14.9 Conclusion

In the first part, we identified the previous applications of advanced technologies to the topic of forced displacement and pointed out the biased case selection issue as well as the potential of open-source data. Given this limitation of current implications, in the second part, we chose one of the less studied area, the DRC, and demonstrated that predictive performances improve as the number of training data increases using only open-source data and XGB model. The result shows the large potential of advanced technologies and open data in this field. Similar to most of the precedents, this research was conducted as an experimental project without an intention to integrate the results into real operations. Building off of our findings, the future research will need to widen target regions and periods as well as to include the pragmatic aspects of the implementations.

Acknowledgement The authors wish to extend their special thanks to Arata Otake, who provided enormous support for this study. Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.

Appendix

Introduction

The correlation between refugee migration and the environment change has gotten attention. This is because deforestation and land development are frequently seen in countries where refugees or IDPs arrive from neighboring countries due to conflicts (Van der Geest et al., 2010). The environmental causes of migration have been studied separately from its environmental impacts. This appendix shows the dynamics of migration and vegetation in the DRC and evaluates the relationship between three population-classified migrations and the environment change. We examine the utility of satellite imagery in measuring human movements from provinces into the destination of IDPs associated with political violence. As for the environmental impacts of migration, we look at three destination regions: Kisangani, Kitchanga, and Bukombo. Datasets of movements are utilized to calculate in-migration flows in this appendix and correlate with vegetation dynamics obtained from a remote sensing dataset. Please note that the first paragraph of a section or subsection is not indented. The first paragraphs that follows a table, figure, equation, etc. does not have an indent, either.

Materials and Methods

To assess the relationship between human migration and vegetation dynamics, we extract figures from datasets by the Internal Displacement Monitoring Center (IDMC) data and remote sensing data from Google Earth Engine. Migration data are utilized to determine in-migration at Kisangani, Bukombo, and Kitchanga. Remote sensing data are applied to calculate the trends in green and water content of vegetation change in time-series. The datasets and methods of extraction are described in more detail below.

Migration Data

The dependent variable, the destination of IDPs, is identified by leveraging the dataset provided by the IDMC. In Tshopo Province, including Kisangani, intercommunity tensions produced the displacement of more than 34,000 people in April 2020. In addition, more than 6000 people have fled toward Bukombo town, in

Table 14.2 Summary of target migration data

Destination	Start date	End date	Qualifier	Figure
Kisangani	2020-04-01	2020-04-30	More than	34,000
Bukombo	2020-08-24	2020-09-07	More than	6000
Kitchanga	2019-01-07	2019-01-24	Approximately	1000

Adapted from the IDMC (2020)

Rutshuru territory, North Kivu, fleeing clashes between armed groups. Furthermore, about 1000 people fled clashes between government forces and armed groups occurring near villages west of Kitchanga, in Masisi territory, North Kivu, moving temporarily to stay in Kitchanga town. Table 14.2 shows migration data from the IDMC for each destination.

Remote Sensing Data and Indices

In this analysis we use two indices: normalized difference vegetation index (NDVI) and normalized difference water index (NDWI). NDVI is designed to understand the status of vegetation using remotely sensed data applied to an easy formula. It represents the quantity and activity level of vegetation. Vegetation is usually seen in green. This is because of the characteristics of light, as shown on the right figure, that vegetation especially reflects green light well and does not reflect red light so much among the lights from the sun. Therefore, a lot of green light can be reached to the eyes of humankind, which consequently makes us see the vegetation green. Besides the green light, vegetation has a characteristic to show a high reflectance on part of the near infrared of the band (NIR). Vegetation index is calculated making the best use of these characteristics (Geospatial Information Authority of Japan, 2020). NDWI is an indicator of the water area on the ground surface and the amount of water contained in vegetation. It is known that the reflection of light by water or snow is the largest in the visible light band and the smallest in short-wavelength infrared radiation (SWIR). This is caused by the absorption of SWIR by water, and NDWI takes advantage of these properties.

Calculation of vegetation and water index of each data observed is as follows.

1. **Calculation of NDVI**

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

2. **Calculation of NDWI**

$$NDWI = \frac{Red - SWIR}{Red + SWIR}$$

Red: Visible red channel
NIR: Near infrared channel
SWIR: Shortwave infrared channel

To extract vegetation and water content indices, we use Google Earth Engine to compute a time-series of both values over three regions between Jan. 2019 and Sept.

2020. Google Earth Engine includes various kinds of satellite dataset and we use Landsat-8, Sentinel-2, and MODIS Aqua data for this analysis to create time-series graphs. Detailed information about satellite's products and quality assurance bands can also be obtained here:

Landsat8: USGS Landsat 8 Surface Reflectance Tier 1: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1_SR
Sentinel-2: Sentinel-2 MSI: MultiSpectral Instrument, Level-2A: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR
MODIS Aqua Daily NDVI: https://developers.google.com/earth-engine/datasets/catalog/MODIS_MYD09GA_006_NDVI?hl=en
MODIS Aqua Daily NDWI: https://developers.google.com/earth-engine/datasets/catalog/MODIS_MYD09GA_006_NDWI?hl=en

To analyze the relation between migration and vegetation or water content in vegetation, we create polygon data of destinations, Kisangani, Bukombo, and Kitchanga, and calculate the NDVI and NDWI, respectively. We calculate the average NDVI and NDWI per day, 15 days, and month to identify densely and sparsely vegetated areas. We argue that averaged NDVI and NDWI are reasonably accurate proxy for the availability of natural resources depending on their livelihoods, because the greenness of the environment is largely determined by rainfall and soil conditions (Fig. 14.2).

Results

The results show a significant but weak correlation between migration and vegetation cover at each level. Due to use of three satellites and a narrow range of target destinations, daily NDVI and NDWI time-series graphs are fluctuated. We select images having cloud less than 10%, but there are still effects on both indices, especially on daily NDVI and NDWI time-series graphs. For this, 15-daily and monthly

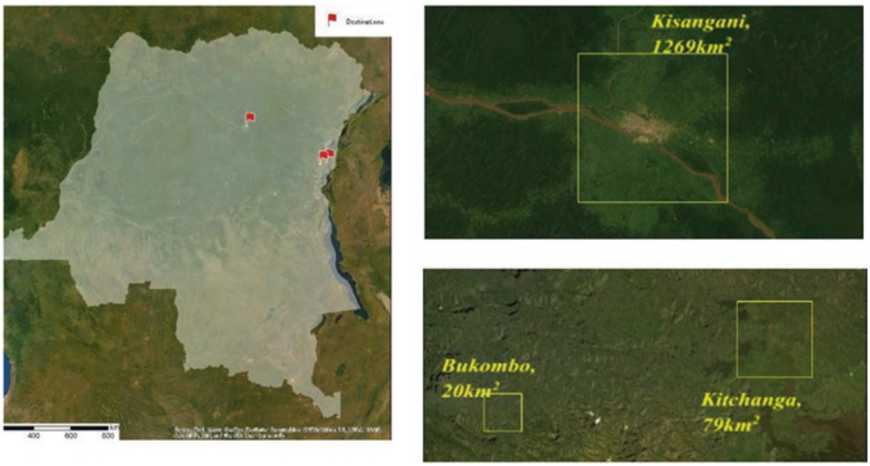

Fig. 14.2 Destinations on migration. (Google Earth Engine, Google)

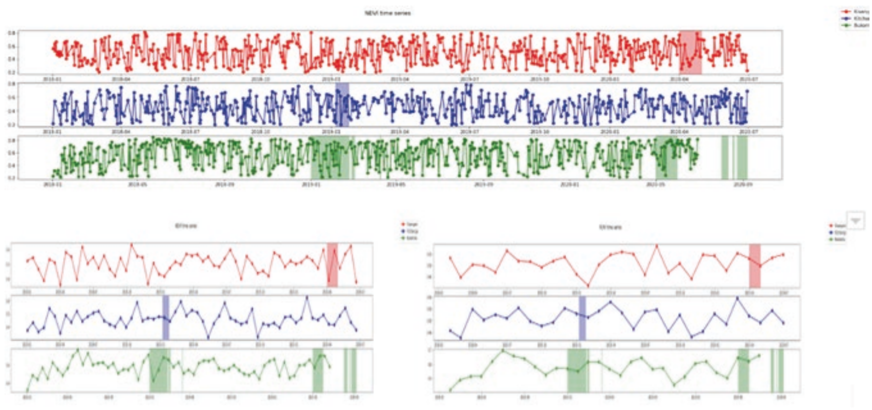

Fig. 14.3 Daily, 15-daily, and monthly NDVI time-series graphs (red Kisangani, blue Kitchanga, green Bukombo)

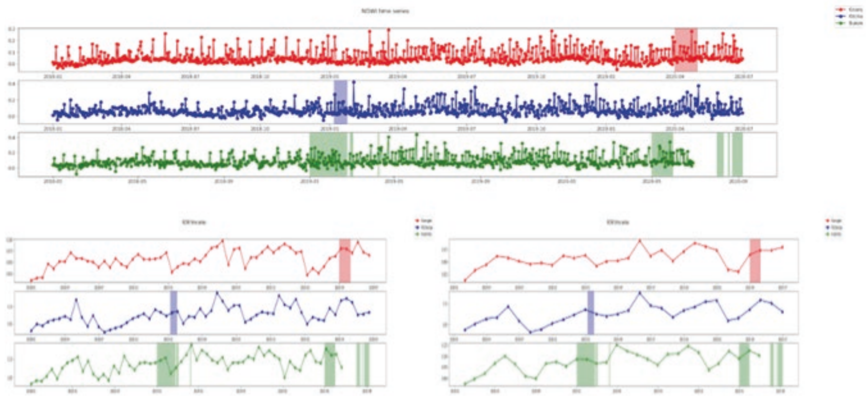

Fig. 14.4 Daily, 15-daily, and monthly NDWI time-series graphs

time-series graphs are suited for our analysis. Area with less populated migration, i.e., Kitchanga has low fluctuation in both indices compared to other overpopulated migration flows, i.e., Kisangani and Bukombo, and those trends toward NDVI and NDWI are more positive. The color bands in Fig. 14.3 determine the periods of migrations from the IDMC dataset (Fig. 14.4).

Conclusion

This substudy was conducted with the aim of expanding the emerging body of knowledge on migration-environment relationships by exploring the multiple ties between human mobility and vegetation and water content in vegetation dynamics in the DRC. We evaluate the role of environment changes in the DRC's three migration flows and explore the impact of migration on vegetation cover and water content in vegetation. Despite limitations in our analysis, this research provides an insightful exploration of migration and vegetation dynamics in migrations. Our findings suggest that although valid fluctuations are not indicated, the environmental factors play an important role in causing migration within the DRC depending on the area of target. A possible explanation could be that our target areas are too narrow to extract both indices and also the DRC does not have annual seasonal variation with constant temperature which may cause visible fluctuation of NDVI and NDWI. Indeed, other studies (Müller et al., 2016; Hassan et al., 2018) show the significance of utilizing NDVI and NDWI for the migrant's flow relation to vegetation change, but they set larger areas than our analysis. To assess the effective environmental impact of migration, larger areas are required to extract for both indices.

Detailed Experiment Results

Here, we show additional figures and tables to supplement the results shown in Sect. 14.5. Figure 14.5 shows the elements of the confusion matrix, the F1 score, and the average precision (AP), in relation to the training data size. Among these, the

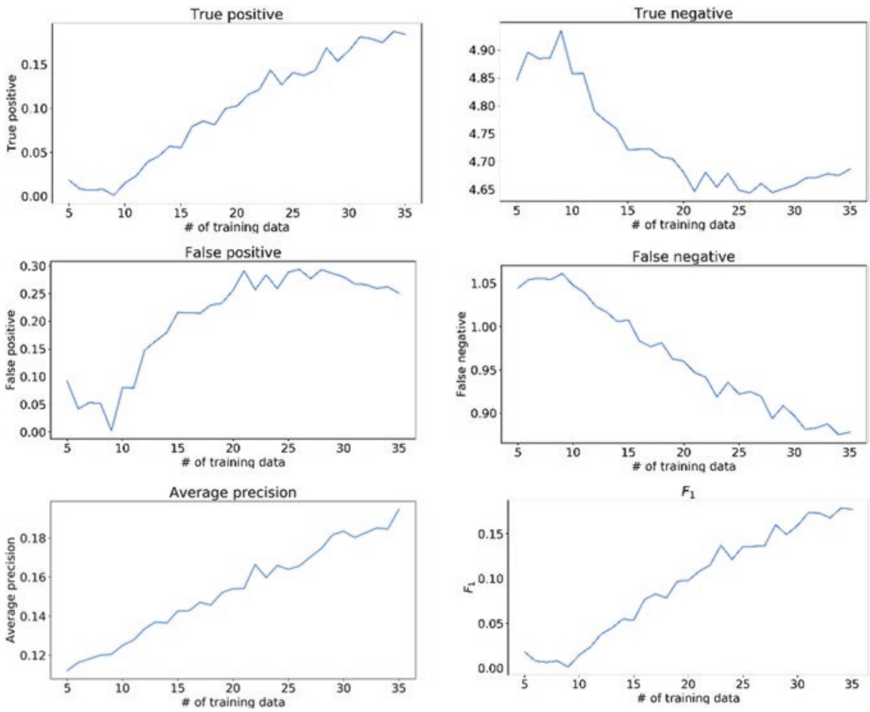

Fig. 14.5 Additional evaluation metrics: elements of the confusion matrix (# of true positives, true negatives, false positives, and false negatives), F1 score, and average precision

Table 14.3 F1 scores for each destination candidate

[a]Sample	Aru	Beni	Bunia	Djugu	Irumu	Kabalo	Kalehe	Kisangani	Lubero
5	0.02	0.02	0.01	0.02	0.02	0	0.03	0.03	0.01
10	0.02	0.04	0	0.01	0.01	0.03	0	0.01	0.01
15	0.01	0.09	0.03	0.06	0.07	0.07	0.08	0.04	0.04
20	0.03	0.12	0	0.12	0.1	0.09	0.15	0.04	0.08
25	0.04	0.17	0.03	0.19	0.15	0.1	0.24	0.07	0.11
30	0.1	0.16	0.01	0.2	0.17	0.17	0.25	0.04	0.12
35	0.05	0.21	0.01	0.29	0.2	0.18	0.3	0.02	0.1

[a]Sample	Mahagi	Masisi	Moba	Mwenga	Nyunzu	Rutshuru	Walikale	Average
5	0.02	0	0.01	0	0.03	0.04	0.03	0.02
10	0.01	0.01	0.01	0.02	0	0.03	0.03	0.01
15	0.01	0	0.05	0.06	0.06	0.17	0.03	0.05
20	0.03	0.04	0.08	0.17	0.11	0.3	0.1	0.1
25	0.03	0.05	0.12	0.17	0.19	0.39	0.13	0.14
30	0.07	0.05	0.1	0.21	0.19	0.5	0.22	0.16
35	0.05	0.09	0.1	0.2	0.2	0.55	0.3	0.18

[a]*Sample* shows the sample size, and *Average* shows the average of the figures over all destination candidates

average precision indicates the area under the precision-recall curve. In both the F1 and the AP, we observe similar trends as those of precision and recall, i.e., the scores improve as the training data size increases (Table 14.3).

References

Ahmed, N., Firoze, A., & Rahman, R. M. (2020). Machine learning for predicting landslide risk of Rohingya refugee camp infrastructure. *Journal of Information and Telecommunication, 4*(2), 175–198.

American Red Cross. (2020). *DRC tribes and ethnic cultural dataset* [Data set]. https://data.humdata.org/dataset/drc_cultural_social-data. (Original work published 2019).

Baird, et al. (2020). *Identifying psychological trauma among Syrian refugee children for early intervention: Analyzing digitized drawings using machine learning*. https://escholarship.org/uc/item/0fc7m5zm

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).

Crowd AI. (2018). *Mapping challenge building missing maps with machine learning*. https://www.crowdai.org/challenges/mapping-challenge

Czaika, M., & Mayer, A. (2011). Refugee movements and aid responsiveness of bilateral donors. *Journal of Development Studies, 47*(3), 455–474. https://doi.org/10.1080/00220388.2010.492855

DFID. (2018). *Digital strategy 2018–2020*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/701443/DFID-Digital-Strategy-23-01-18a.pdf

Eck, K. (2012). In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Cooperation and Conflict, 47*(1), 124–141. Retrieved November 7, 2020, from http://www.jstor.org/stable/45084688

Facebook. (2013). *Coordinated migration*. Posted on 18 December, 2013. https://www.facebook.com/notes/facebook-data-science/coordinated-migration/10151930946453859/

Geospatial Information Authority of Japan. (2020). *NDVI data*. Retrieved from https://www.gsi.go.jp/kankyochiri/ndvi_e.html

Government of Singapore. (2020). *Trace together*. https://www.tracetogether.gov.sg/

GSMA. (2020). *Using mobile big data to help inform the fight against COVID-19 in the Democratic Republic of Congo*. Posted on 29 July, 2020. https://www.gsma.com/mobilefordevelopment/blog/using-mobile-big-data-to-help-inform-the-fight-against-covid-19-in-the-democratic-republic-of-congo/

Hanai, K. (2019). Current situation and issues of conflict related minerals in the Democratic Republic of the Congo (in Japanese, コンゴ民主共和国における紛争資源問題の現状と課題). *International Affairs, 682*. http://www2.jiia.or.jp/kokusaimondai_archive/2010/2019-06_003.pdf?noprint. Accessed 5 Sept 2020.

Hassan, M., Smith, A., Walker, K., Rahman, M., & Southworth, J. (2018). Rohingya refugee crisis and forest cover change in Teknaf, Bangladesh. *Remote Sensing, 10*(5), 689. https://doi.org/10.3390/rs10050689

Immigration Policy Lab. (2020). *GeoMatch connecting people to places*. https://immigrationlab.org/geomatch/

International Organization for Migration (IOM). (2016). *Global compact for migration*. https://www.iom.int/global-compact-migration

International Peace Information Service. (2019). *Artisanal mining sites in Eastern DRC 2009–2019*. https://data.humdata.org/dataset/cod_mines_curated_all_opendata_p_ipis

International Telecommunication Union. (2019). *United Nations Activities on Artificial Intelligence (AI)*. https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2019-1-PDF-E.pdf

Lu, Y., Koperski, K., Kwan, C., & Li, J. (2020). Deep learning for effective refugee tent extraction near Syria-Jordan border. *IEEE Geoscience and Remote Sensing Letters, 2020*, 1.

Migration Data Portal. (2020). *Mapping population distributions on city level, using Twitter data*. Posted on 28 September 2020. https://migrationdataportal.org/data-innovation/mapping-population-distributions-city-level-using-twitter-data

Müller, M. F., Yoon, J., Gorelick, S. M., Avisse, N., & Tilmant, A. (2016). Impact of the Syrian refugee crisis on land use and transboundary freshwater resources. *Proceedings of the National Academy of Sciences of the United States of America, 113*(52), 14932–14937. https://doi.org/10.1073/pnas.1614342113

Nominatim (3.5.2). (2020). [Computer software]. osm-search. https://github.com/osm-search/Nominatim. (Original work published 2012).

Norwegian Refugee Council. (2020). *Fear for safety for people in Ituri affected by armed groups attacks and intercommunal violence*. Norwegian Refugee Council. https://www.nrc.no/news/2020/february/dr-congo-new-displacement-in-ituri/. Accessed 8 Sept 2020.

OCHA. (2010). *Democratic Republic of the Congo, 1993–2003*. United Nations High Commissioner for Human Rights. https://www.ohchr.org/Documents/Countries/CD/DRC_MAPPING_REPORT_FINAL_EN.pdf. Accessed 8 Sept 2020.

OCHA DR Congo. (2019). *Democratic Republic of the Congo—Subnational Population Statistics* [Data set]. https://data.humdata.org/dataset/democratic-republic-of-the-congo-administrative-level-0-2-population-statistics

OCHA/UNJHRO. (2020). *Overview of the human rights situation during the first semester 2020*. United Nations High Commissioner for Human Rights (OCHA) and United Nations Joint Human Rights Office in the DRC (UNJHRO) MONUSCO – OHCHR. https://monusco.unmissions.org/sites/default/files/unjhro_-_analysis_of_the_human_rights_situation_in_drc_2020_6_month_eng_copy.pdf. Accessed 8 Sept 2020.

Omdena. (2019). *Predicting climate change and forced displacement*. https://omdena.com/projects/ai-climate-change/. Accessed 5 Sept 2020.

Open Source Routing Machine. (2020). *Project OSRM*. https://github.com/Project-OSRM/osrm-backend. (Original work published 2011).

OpenStreetMap Contributors. (2017). *Planet dump*. Retrieved from https://planet.osm.org. https://www.openstreetmap.org

Parater, L. (2020). *Jetson: Insights into building a predictive analytics platform for displacement*. Retrieved October 10, 2020, from https://www.unhcr.org/innovation/jetson-insights-into-building-a-predictive-analytics-platform-for-displacement/

Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., & Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2128), 20170363.

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED-armed conflict location and event data. *Journal of Peace Research, 47*(5), 651–660.

Sekeris, P., De Luca, G., Maystadt, J.-F., & Ulimwengu, J. (2013). Mineral resources and conflicts in DRC: A case of ecological fallacy. *Oxford Economic Papers, 66*. https://doi.org/10.1093/oep/gpt037

Spelhaug, J. (2019). Helping refugees and displaced persons by shifting the approach to how we help nonprofits. *Microsoft on the Issues*. https://blogs.microsoft.com/on-the-issues/2019/06/19/helping-refugees-and-displaced-persons-by-shifting-the-approach-to-how-we-help-nonprofits/

Stearns, J., & Voge, C. (2017). The landscape of Armed Groups in Eastern Congo Fragmented, politicized networks, KIVU SECURITY TRACKER. https://kivusecurity.nyc3.digitaloceanspaces.com/reports/5/Landscape%20of%20Armed%20Groups%20Essay%20KST.pdf. Accessed 5 Sept 2020.

Suleimenova, D., Bell, D., & Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific Reports, 7*(1), 1–13.

Teytelboym, A (2020). *Annie™ Moore: Increasing employment of resettled refugees using matching, Machine learning and integer optimization*. University of Oxford. https://www.economics.ox.ac.uk/impact/annie-moore-increasing-employment-of-resettled-refugees-using-matching-machine-learning-and-integer-optimisation

The World Bank. (2019). *The mobility of displaced Syrians an economic and social analysis*. https://www.worldbank.org/en/country/syria/publication/the-mobility-of-displaced-syrians-an-economic-and-social-analysis

Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). Mining multi-label data. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 667–685). Springer. https://doi.org/10.1007/978-0-387-09823-4_34

UN Global Pulse. (2017). *Understanding movement and perceptions of migrants and refugees with social media* (Project Series, No. 28, 2017. https://reliefweb.int/sites/reliefweb.int/files/resources/Social%20Media%20Forced%20Displacement%20Europe%20Refugee%20Crises%20_0.pdf

UN Security Council. (2019). *Final report of the Group of Experts on the Democratic Republic of the Congo*. United Nations Security Council. https://reliefweb.int/sites/reliefweb.int/files/resources/S_2019_469_E.pdf. Accessed 5 Sept 2020.

UNHCR. (2018a). *Congolese situation – Responding to the needs of displaced Congolese and refugees – Supplementary Appeal, January–December 2018*. UN High Commissioner for Refugees, https://reporting.unhcr.org/node/20239. Accessed 10 Sept 2020.

UNHCR. (2018b). *Independent evaluation of UNHCR's response to the L3 emergency in the Democratic Republic of Congo, Evaluation Report, December 2018*. UN High Commissioner for Refugees, https://www.unhcr.org/en-us/5c5419fe4.pdf. Accessed 3 Nov 2020.

UNHCR. (2019). *Position on returns to North Kivu, South Kivu, Ituri and adjacent areas in the Democratic Republic of the Congo Affected by Ongoing Conflict and Violence – Update II*. UN High Commissioner for Refugees, https://www.refworld.org/docid/5d6d794a4.html. Accessed 10 Sept 2020.