Karl-Franzens-University Graz
Department of Geography and Regional Science

German Aerospace Center
Earth Observation Center
Remote Sensing Data Center
Department of Georisks and Civil Security
Team City and Society

Deep Learning Based Exposure Analysis of Landslide-Prone Areas in Medellín, Colombia

Master's Thesis

in

Applied Physical Geography and Mountain Research

Submitted by:

Raphael **Tubbesing**, B.Sc.

Supervision:

Ao.Univ.-Prof Mag. Dr.rer.nat. Wolfgang **Sulzer**

University of Graz

Dr. Michael **Wurm**

German Aerospace Center

Graz, September, 2022

Abstract

In the last century, Medellín grew into one of Colombia's largest cities. Today, the city continues to grow primarily due to the influx of internally displaced people (IDP's), who have been forced to leave their homes at the country side due to natural disasters or drug-related violence. Since the internally displaced are mostly low-income farmers and peasants, they are migrating to the larger cities in search of greater security and jobs. In Medellín, the new residents mostly settle informally on the steep slopes to the east and west of the city. Due to limited space and steep topography, such settlements are often built in areas with medium and high probability of landslides. However, not only free land area within the municipal boundaries are exploited by the build-up of new settlements, but also free land beyond the border of the municipality, which causes the city to grow into the rural area.

The study therefore seeks to find out how many residents are prone to potential landslide activity in the context of the pattern of migration. To analyze the exposure, the population is disaggregated down to the individual building block level. Such an approach requires precise building footprints to locate the population in relation to landslide-prone areas. Although the city has a cadaster including building footprints, it is more imprecise and incomplete towards the outskirts of the city, where landslide susceptibility is most pronounced. The incompleteness is due to the high population dynamics, which makes it quite difficult to maintain an up-to-date cadaster. But since Medellín's geospatial data service provides an orthophoto from 2019, a deep learning-based building extraction is applied to generate a more comprehensive building footprint dataset. This will be the main data source for the exposure analysis.

The respective deep learning architecture is a U-Net has been refined with the EfficientNetB2 as a backbone and eventually fine-tuned. It could show very accurate results, while still facing some challenges, like the field-of.view of the image tiles, that is sometimes too small for the vast rooftop landscapes, which leads to misclassifications.

The exposure analysis of population exposed to landslide hazard could prove the importance of having a more up-to-date data basis. While the number of residents living in formal settlements is not to different from the cadaster and the deep-learning derived building footprints, those numbers of residents of the informal settlements are much higher in the more actual deep learning derived dataset. A strong increase could also be found in the population exposed to medium and high landslide hazard. Further analyses facilitate the impression, that the poorer and the more vulnerable population has distinctively higher exposure to considerable landslide hazard, when using the deep-learning derived dataset. These findings show the strength of remote sensing techniques in order to retrieve actual building footprint data, that is clearly important for the estimation of potential consequences of landslide-prone areas.

Zusammenfassung

Im letzten Jahrhundert wuchs Medellín zu einer der größten Städte Kolumbiens heran. Heute wächst die Stadt vor allem durch den Zustrom von Binnenvertriebenen, die aufgrund von Naturkatastrophen oder Gewalt im Zusammenhang mit Drogen gezwungen waren, ihre Heimat auf dem Land zu verlassen. Da es sich bei den Binnenflüchtlingen meist um einkommensschwache Landwirte und Bauern handelt, wandern sie auf der Suche nach mehr Sicherheit und Arbeitsplätzen in die größeren Städte ab. In Medellín lassen sich die neuen Bewohner meist informell an den steilen Hängen im Osten und Westen der Stadt nieder. Aufgrund des begrenzten Platzes und der steilen Topografie werden solche Siedlungen oft in Gebieten mit mittlerer und hoher Wahrscheinlichkeit von Erdrutschen errichtet. Aber nicht nur freie Flächen innerhalb der Gemeindegrenzen werden durch den Bau neuer Siedlungen genutzt, sondern auch Flächen außerhalb der Stadtgrenze, wodurch die Stadt weiter in den ruralen Raum wächst.

Im Zuge der vorliegenden Masterarbeit wird daher versucht herauszufinden, inwieweit die Bewohner, im Zusammenhang mit dem Migrationsmuster, potenziellen Erdrutschaktivitäten ausgesetzt sind. Um die Gefährdung zu analysieren, wird die Bevölkerung auf die Ebene einzelner Gebäudeblöcke

disaggregiert. Ein solcher Ansatz erfordert präzise Gebäudegrundrisse, um die Bevölkerung in Bezug auf die erdrutschgefährdeten Zonen zu lokalisieren. Obwohl die Stadt über ein Kataster mit Gebäudegrundrissen verfügt, wird es zum Stadtrand hin, wo die Erdrutschgefahr am stärksten ausgeprägt ist, ungenauer und unvollständiger. Die Unvollständigkeit ist der hohen Bevölkerungsdynamik geschuldet, die die Instandhaltung eines aktuellen Katasters recht schwierig macht. Da der Geodatendienst von Medellín ein Orthofoto aus dem Jahr 2019 anbietet, wird eine Deep-Learning-Gebäudeextraktion angewandt, um einen umfassenderen Gebäudegrundrissdatensatz zu generieren. Dieser wird die Hauptdatenquelle für die Expositionsanalyse sein.

Die genannte Deep-Learning-Architektur ist ein U-Net, das mit dem EfficientNetB2 als Backbone ergänzt und schließlich feinabgestimmt wurde. Es wurden sehr genaue Ergebnisse erzielt, auch wenn es immer noch einige Herausforderungen in der Methodik gibt, wie z. B. dem Sichtfeld der Bildkacheln, das manchmal zu klein für die großen Dachlandschaften ist, was entsprechend zu Fehlklassifizierungen führt.

Die Expositionsanalyse der erdrutschgefährdeten Bevölkerung konnte zeigen, wie wichtig es ist, über eine aktuellere Datengrundlage zu verfügen. Während sich die abgeleiteten Einwohnerzahlen in den formellen Siedlungen nicht sehr stark zwischen den Kataster- und den Deep-Learning-basierten Gebäudefußabdrücken unterscheiden, sind die Einwohnerzahlen in den informellen Siedlungen im letzteren Datensatz bedeutend höher. Ein starker Anstieg konnte auch bei der Bevölkerung festgestellt werden, die einem mittleren und hohen Erdrutschrisiko ausgesetzt ist. Weitere Analysen unterstützen den Eindruck, dass die ärmere und stärker gefährdete Bevölkerung bei Verwendung des Deep-Learning-Datensatzes deutlich stärker von Erdrutschen bedroht ist. Diese Ergebnisse zeigen die Stärke von Fernerkundungstechniken bei der Gewinnung aktueller Gebäudegrundrissdaten, die für die Abschätzung der potenziellen Folgen von erdrutschgefährdeten Gebieten bedeutsam sind.

Danksagung

Durch die lange Reise vom Anfang meiner Masterarbeit bis hin zum Abschluss der selbigen habe ich viel gelernt. Insbesondere über das Leben und über mich selbst. Zwar auch fachlich, aber fachliches zu lernen ist häufig einfacher und greifbarer auch wenn es häufig theoretischer Natur ist. Für die Möglichkeit diesen Weg einzuschlagen, möchte ich Dr. Michael Wurm danken. Gerade die Zusammenarbeit mit ihm und die Zeit meiner praktischen Arbeit im Untergeschoss des DFD‘s, war sehr aufregend, spaßig und inspirierend. die Zeit danach ist qualitativ leider Covid-19 zum Opfer gefallen. Entsprechend der Zeit vor der Pandemie möchte ich auch der ganzen „Keller-Crew“ (u.a. Peter, Elli, Gabo, Patti, Henri, Moritz, Anna und Lorenz) danken, die diese Zeit besonders gemacht haben. Außerdem möchte ich natürlich noch Ariane danken, die mich quasi den ganzen Weg seit dem Studium bis jetzt begleitet hat, da wir auch irgendwie denselben gegangen sind. Danke also für die geistige Inspiration, mentale Unterstützung, die guten Gespräche, das gute Essen und wahrscheinlich Weiteres.

Auch wenn nicht vor Ort anwesend, möchte ich mich auch noch bei Robi und Josi für die Unterstützung bedanken. Durch das Telefon oder die Webcam bestand immer ein Kanal für Austausch, Unterstützung, gutem Rat und viel Spaß.

Vom Institut für Geographie und Raumforschung der Karl-Franzens-Universität möchte ich Ao.Univ.-Prof Mag. Dr.rer.nat. Wolfgang Sulzer für die Herausforderungen in der Fernerkundung, die schönen Aufenthalte im Feld rund um die Pasterze und für das Vertrauen danken.

Zum guten Schluss noch ein großer Dank an das gesamte Institut für Geographie und Raumforschung der Karl-Franzens-Universität im Bereich der physischen Geographie. Ich habe selten Dozenten und Professoren kennenlernen dürfen, die so viel Mühe aufbringen, um ihre Inhalte in Kursen und Vorlesungen so spannend und interaktiv wie möglich zu gestalten. Zudem auch noch ein besonderer Dank an Ao.Univ.-Prof. Mag.phil. Dr.rer.nat. Gerhard Karl Lieb, der immer ein offenes Ohr für die Studierenden hat und mit seinen Taten und Ratschlägen meines Wissens nach noch jedes Problem der Studis gelöst hat.

Vielleicht sind zwei abschließende Aphorismen auch ein Dank, die in der Lage sind etwas zurückzugeben:

UNABHÄNGIGKEIT UND SELBSTERKENNTNIS. Wir scheitern immer wieder daran gute Gewohnheiten und Techniken für unseren Alltag und unser eigenes Leben zu erlernen. Dabei versteifen wir uns fast immer darauf einzelne Bausteine zu lernen und sie uns zu eigen zu machen. Das ist wie das richtige Puzzleteil mit verbundenen Augen zu finden und anzulegen. Das heißt, wenn wir beispielsweise versuchen uns besser abzugrenzen, von Dingen oder Menschen, lernen wollen nein zu sagen, ist es meistens relativ schwer umzusetzen und langfristig zu erlernen. Das liegt daran, dass diese Fähigkeiten in erster Linie ein externer Baustein sind und wenig mit uns selbst oder unseren bestehenden Fähigkeiten/Kompetenzen zu tun hat. Wenn wir aber sagen, wir wollen zu jeder Zeit aufrichtig und redlich sein, dann ist es wahrscheinlich, dass wir das in einigen Situationen schon recht gut beherrschen und wir gleichzeitig nur dann ganzheitlich aufrichtig sind, wenn wir uns auch abgrenzen können. Somit komplettieren wir eine Fähigkeit, die wir schon besitzen und versuchen nicht eine ganz neue in uns einzubauen.

DAS GESCHRIEBENE WORT. Zwar ist die wissenschaftliche Schreibweise immer sehr klar, deutlich und sollte unmissverständlich geschrieben sein, aber das Schöne an der philosophischen Schreibweise ist, dass der Leser oder die Leserin nur über tiefgründiges Nachdenken und dem Einbinden der eigenen Erfahrung versteht, was der Autor meint. Damit wird aber nicht nur das Geschriebene klar, es bekommt auch einen individuellen Einfluss durch die Gedanken der Leser. Dies entspricht der Auffassung, dass es keine universelle Wahrheit gibt, aber es sehr wohl möglich ist sich einer solchen, auch im Konsens, möglichst nahe zu bringen.

Table of contents

List of figures

List of tables

Abbreviations

AOI	Area of Interest
C&I	Commerce & Industry
CNN	Convolutional Neural Network
DAGRD	Administrative Department of Disaster Management
DAP	Administrative Department of Planning
DL	Deep Learning
DNN	Deep Neural Network
FARC	Revolutionary Armed Forces of Columbia
FCN	Fully Convolutional Network
FOV	Field of View
GIS	Geographic Information System
GLCM	Grey Level Co-occurrence Matrix
GPU	Graphics Processing Unit
HOG	Histogram of Orientated Gradients
iDMC	Internal Displacement Monitoring Center
IDP	Internally Displaced Persons
LBP	Local Binary Patterns
LSSd	Low Socioeconomic Status, dense
LSSl	Low Socioeconomic Status, loose
MSS	Medium Socioeconomic Status
PUI	Proyecto Urbano Integral
ReLU	Rectified Linear Unit
RGB	Red-Green-Blue
RS	Remote Sensing
SVM	Support Vector Machines
UAS	Unmanned Aircraft System
VHR	Very High Resolution

1 Introduction: Medellín, a vulnerable city?

Medellín, Columbia, situated in the Aburrá Valley as part of the Central Cordilliera, is surrounded by steep mountain slopes to the east and the west, on which mainly poorer dwellers live (Ferrari et al. 2018). In the first decades of the 20th century, Medellín was a major industrial base; build upon the earnings of gold mining and coffee production. This resulted in a considerable concentration of jobs and education, and thus in a significant influx of migrants from the countryside. These people were mainly low-income farmers and peasants, settling informally on the margins of the city and accordingly on the slopes (Betancur 2007, Ferrari et al. 2018). The city grew from around 138,000 inhabitants in 1938 to around 1,150,000 in 1973 (Betancur 2007). Further migration intensification has been caused by conflicts between paramilitary forces, guerrilla groups and the military (Ferrari et al. 2018, Betancur 2007). Today, the city counts approximately four million inhabitants (UN DESA, 2018).

In recent years, the city has pulled much attention in public and among urban planners, due to its advances in its socio-economic situation after the escalation of the drug war between 1990 and 2002 has ended. It was the beginning of the so called ‘social urbanism’. Characteristic for the social urbanism were the ‘Integrated Urban Projects’ (span. ‘Proyecto Urbano Integral’ (PUI)), which aimed towards the integration of the more indigent residents and of the informal settlements, located especially on the steep slopes of Medellín. The spatial integration of the disadvantaged social groups living in these deprived areas has been realized by providing mobility with the expansion of cable car lines, the urban metro, and the installation of escalators (Ferrari et al. 2018, Heinrichs & Bernet 2014). Additionally, investments in cultural and educational infrastructure, like libraries in the poorest areas of the cities, have been made (Ferrari et al. 2018, Humphrey & Valverde 2017).

Despite the successes of integrating the poor and violence-plagued population of Medellín’s urban periphery, which can be measured in a distinct decrease in the homicide rate, the social situation in the concerned barrios (Eng. neighborhoods: smallest administrative unit) is still precarious. The pacification and securitization

have not only been undertaken by the governmental military and police, but also in an intertwined cooperation with paramilitary forces and criminal gangs or companies (Civico 2012). This led to fragmented governance in the city and to urban micro-territories, controlled by illegal armed gangs, who e.g. extort rents or recruit local youth as foot soldiers. Consequently, between five and fifteen thousand persons per year flee their homes from one neighborhood to another (Marston 2019, Humphrey & Valverde 2017). The internal displacement, as a result of criminal violence in general, has been rated as highly urban by the Internal Displacement Monitoring Centre (iDMC). But there is still a high amount of internally displaced persons (IDPs), which flee their homes at the countryside and seek for shelter in cities like Medellín, where they contribute to the expansion of informal settlements at the slopes of the city. The conflicts in rural areas are driven by several armed groups, like the Revolutionary Armed Forces of Colombia (FARC), who rival for control of land, land mining, cocoa plantations, and drug–trafficking routes. Displacement associated with conflicts affected about 145,000 people in 2018 (iDMC 2019). But there is also a high amount of displacement associated to natural disasters, since a big part of Colombia is located in the Andean Mountain range, where active volcanoes and seismic activity is present. Further risk is related to landslides and floods. Natural disasters coerced 67,000 people to leave their homes in 2018 (IDMC 2019).

All these affected people are potential new inhabitants for the informal settlements in Medellín. The fact that a) the slopes around Medellín are prone to landslide activity, b) the population density there is very high, and c) the relatively basic building fabric of the housings of those still growing settlements, let the population become significantly vulnerable and threatened in their physical and economic integrity (cf. Vega & Hidalgo 2016).

Landslides, as the predominant form of natural hazards in Medellín are mostly caused by rainfall events. The patterns of precipitation show a bimodal distribution throughout the year with maxima from April to May and from October to November. Also, short duration precipitation events with high intensities can occur during the rest of the year (Klimeš & Escobar 2010). In addition, the bedrock consists of structurally weak and highly erosive dunite rock (Claghorn et al. 2016). Due to the combination of the resulting deep regolith, steep relief, phases of intense rain and

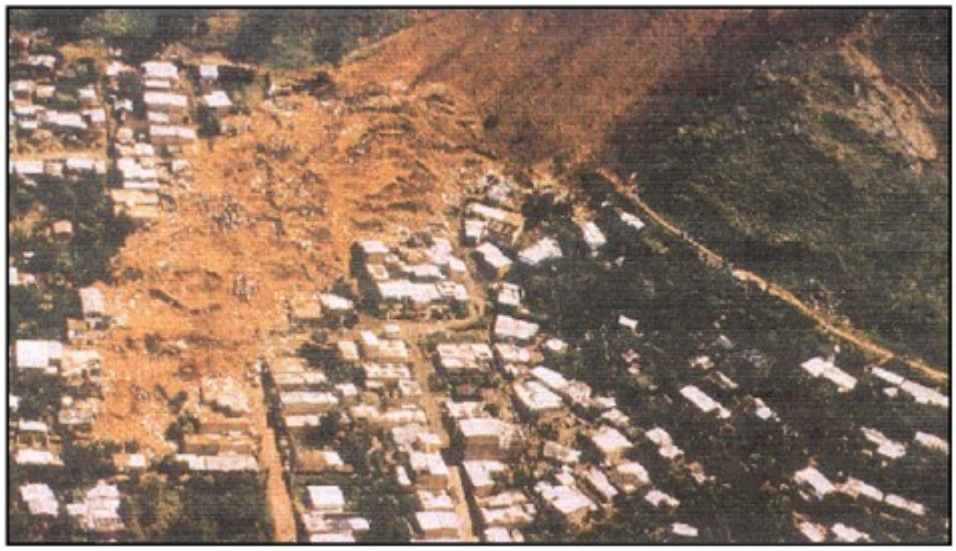


Figure 1: *In 1987 a landslide in Villa Vallantina, Medellín, caused appr. 500 casualties triggered by a leakage at a water pipeline. Ojeda 2006, p.10.*

the high population density at the slopes, the area around the city can be assessed as highly prone to landslides hazards. These are likely to cause high rates of fatalities (see Claghorn & Werthmann 2015, Ojeda 2006). The past proves the severe situation, since 854 people died in landslides in Medellín's Aburrá Valley in the last 80 years (Claghorn et al. 2016). According to estimates from O'Shea (2014), 284,000 people are at risk at the slopes of the Aburrá Valley and nearly 350,000 people are expected to be at risk by 2030.

Although Columbia is situated on the Nazca-Southamerica-Caribe plate convergence zone, earthquakes which had been registered in the east of Medellín between 1566 and 1999, haven't had a higher magnitude than M = 4 and are mostly below M = 3 (Klimeš & Excobar 2010, Ojeda & Donelly 2006). Therefore, seismic activity prevails not as a major trigger mechanism, but it still represents a probable hazard and is also a topic under investigation (see Acevedo 2017, Vega & Hidalgo 2017, Vega & Hidalgo 2016).

A more important aspect are human activities, which can exacerbate or trigger landslides, as happened in 1987 (fig. 1), when the most fatal landslide caused about 500 casualties and affected further 1,300 – 1,700 residents (Aristizábal & Gómez 2007, Ojeda 2006, García 2005). This event has been triggered by a leakage at a water pipeline, which maintains the water supply in the peripheral settlements (Aristizábal & Gómez 2007). But also overflowing drainage channels, a blocked

drain or even an informal carwash have also reinforced landslide events in the past (Claghorn & Werthmann 2015). Consequently, it is of high importance to counter the prevailing risk with mitigation strategies. This master thesis will serve as a basis for these strategies by measuring the exposure of the population living in landslide-prone areas. This approach uses dasymetric cartography to increase the resolution of the population distribution compared to the smallest census unit, the neighborhood level. With building footprints and a hazard map as ancillary data it is possible to estimate exposed population locally using the census data. Since the city of Medellín is determined by high population dynamics, especially due to IDPs, it is important to have an up-to-date source for the building footprints to neither under- nor overestimate exposed population. Therefore, a building extraction algorithm has been applied on an orthophoto. The algorithm is based on deep neural networks (DNN), since they have been proven as a valuable tool for extracting buildings from very high-resolution images (e.g. Wurm et al. 2021, Stiller et al. 2019, Lu et al. 2018, Xu et al. 2018, Huang et al. 2016). In addition to creating a suitable architecture for this task, another objective is to compare the results of the exposure analysis with those obtained by using the official building footprints from the cadastral data. As it is difficult to maintain cadastral data in a highly dynamic city, the comparison of these two results will highlight how much incomplete data can affect the results.

1.1 State of the art

The current knowledge about the relevant research for this master's thesis will be presented in this section, which addresses different parts. First, the scientific context about the landslide hazard and risk in Medellín will be outlined for a better understanding of the relevance of this study (section 1.1.1). Since a DNN will be applied for creating building footprints, the method will be brought in the context of the state-of-the-art research of building extractions with DNNs in section 1.1.2.

1.1.1 Landslide-related research in Medellín

The morphology of the Aburrá Valley in Medellín and the ongoing high rates of urbanization and population dynamics force residents to settle on the west and east slopes of Medellín. Combined with the precipitation patterns and the underlying geological stratification and soil properties, many of the residents can be considered as being exposed to a high risk of landslides (Ferrari et al. 2018, Isaza-Restrepo et al. 2016, O'Shea 2014, Betancur 2007). Therefore, several studies have been conducted to gain more knowledge about the hazardous situation. The approaches of the studies presented below are on the one hand focusing on quantifying natural processes and on the other hand more qualitatively on mitigation strategies for a given risk.

As one of the quantitative oriented studies, Isaza-Restrepo et al. (2016) proposed a methodology for quantifying the hazard, vulnerability, and risk at 120 sites in Medellín with emphasis on geotechnical properties of the slopes. A further quite comprehensive investigation of landslide risk, taking geotechnical properties into account as well, has been conducted in three *communas* in the north east of Medellín by Vega et al. (2016). Additionally, their focus laid on the economic costs of the affected buildings. Klimeš & Escobar (2010) constructed a method for landslide susceptibility and hazard assessment by modifying pre-existing landslide inventories. Spatially, the study has been applied to the Iguaná Valley in the north-west of Medellín, covering 13% of the cities area. Also, not directly in Medellín, but at four different roadsides around the city, Muñoz et al. (2014) statistically analyzed the frequency and extent of landslides and the influence of rainfall and anthropogenic activities as triggering mechanisms. Further studies by Acevedo et al. (2017) and Vega & Hidalgo (2017) are dealing with seismic risk, whereas the first study evaluates the seismic risk for unreinforced houses. The latter one proposes a model to quantify arising costs to urban infrastructure under different scenarios of seismically induced landslides.

The studies that are focusing more on a qualitive analysis to develop suitable mitigation strategies for risk reduction have been carried out by Claghorn & Werthmann (2015), Claghorn et al. (2016), and Smith et al. (2020). All of them using certain slopes of the highly affected eastern part of Medellín, where residents mostly

settle informally. They developed pilot projects in low-income neighborhoods, to find strategies for alleviating risk. These are for example early warning systems, drainage improvements, but also slope stabilizing methods like forestation or micro-farming (Claghorn et al. 2016, Claghorn & Werthmann 2015). The focus for Smith et al. (2020) lies particularly on how risk mitigation strategies can be politically and socially accomplished. All of them incorporate and investigate the participation of the local community.

It shows that the research, which has been done in several different study areas in and around Medellín already provide valuable findings about geotechnical behavior of hazardous slopes (Muñoz et al. 2014, Klimeš & Escobar 2010), as well as risk assessment with statistical methods and estimations about life- (Isaza-Restrepo et al. 2016) and economic losses (Isaza-Restrepo et al. 2016, Vega & Hidalgo 2016).

In that context, this master's thesis is a quantitative orientated study and contributes to a present knowledge gap. As pointed out above the previous studies examined causes for landslide hazard, such as investigating the geotechnical properties or triggering mechanisms and dealt with landslide susceptibility. However, landslide susceptibility was estimated and measured in potential economic damage, but not in potential human losses. To complement current research this thesis will elaborate the exposure of inhabitants to landslides. It will further extent the data basis on which qualitative studies can rely to adapt risk mitigation strategies and on which political decisions or measures can be made.

1.1.2 Deep neural networks for remote sensing and the field of building extractions

In the current decade Deep Learning (DL) evolved to a very promising method in image analysis, as recurrent successes in Machine-Learning contests show. The fundament for the widespread applications of Neural Networks has been built when they were made more accessible to researchers by providing easier to use programming frameworks in common languages like Python. Before, one needed to have advanced low-level programming knowledge for building and applying a Neural Network (Chollet 2018).

While shallow Neural Networks in remote sensing (RS) have been used for roughly two decades, DNNs started to draw the attention of the RS community around the year 2014 (Ma et al. 2019). DL has been proven to be a very successful method in different RS applications like object detection, scene classification or semantic segmentation (Li et al. 2019, Ma et al. 2019, Zhu et al. 2017). Some more specific studies could also show accurate results at slum detection and building extraction in general, which is the corresponding topic under investigation in this master's thesis (see Wurm et al. 2021, Wurm et al. 2019, Stiller et al. 2019, Persello & Stein, 2017).

A key property, which differentiates DL from conventional classification algorithms is the ability of the algorithm to automatically create abstract and meaningful features for representation of image attributes. Although hand-crafted textural, geometric, and spectral attributes could also lead to accurate classification results, it is not possible to achieve an optimal balance between discriminability and robustness of image object details (Zhang et al. 2016). DNNs consist of multiple layers or levels of representation, based on the raw image data. They are obtained by simple non-linear functions applied to the pixel values by applying a moving kernel on an image tile. Every single created feature map or layer of representation of these tiles transforms the previous layer into a higher, slightly more abstract level using a general-purpose learning procedure (LeChun 2015, Zhang et al. 2016). The reason why tiles instead of whole scenes are used in the classification is that the graphics processing units (GPU) needs to be utilized to enable the computation of such an amount of calculations. However, bigger scenes won't fit into the memory of a GPU. A more detailed insight into the architecture of a DNN will be given in chapter 3.1 and 3.2.

DL as a tool for extracting buildings from very high resolution (VHR) optical imagery is often based on semantic segmentation. It is a pixel-wise classification procedure to infer objects from the image data. Since VHR data is available in spatial resolutions from 5 cm by now, many urban objects like even small rooftops can be potentially classified (Xu et al. 2018). The challenge which comes with such high spatial resolutions is the high number of pixels with high intra-class variability, that describe one image object. To achieve good classification accuracies with

conventional classification approaches, it is generally necessary to extract spatial-contextual features to capture all of the descriptive pixels with the help of e.g. a grey level co-occurrence matrix (GLCM), local binary patterns (LBPs), histogram of oriented gradients (HOGs) or filters that include the neighbourhood. However, it is not known beforehand which feature extractor yields the best improvements and each of its parameter needs to be tuned in a trial-and-error manner. Additionally, applying a lot of feature extractors can be computationally very intensive, and therefore time consuming, especially when applying large filters (Bergado et al. 2016, Mboga et al. 2017, Persello et al. 2017, Wurm et al. 2019). Cost intensive computation is a key limitation in this specific project, since the investigation area covers 55.4 km² with big contiguous areas of buildings (map of the investigation area can be found in section 2.2.2, fig. 6, and in section 4.3, fig. 22). Thus, using a DNN is expected to be handier, as it extracts spatial-contextual features automatically directly from the data (cf. Bergado et al. 2016, Mboga et al. 2017).

Although DNNs with several different architectures, e.g. Res-U-Net in Xu et al. (2018) (F1-score: 97,93% and 98,8% for respective study area), DeepResUnet in Yi et al. (2019) (F1-score: 91,9%), Mask-RCNN in Stiller et al. (2019) (overall accuracy: 88,8%), Support vector machine-based (SVM-based) fusion of Multiscale Convolutional Neural Networks (CNN) in Sun et al. (2019) (F1-score: 91%, 90% and 94% for respective study area), showed very good results in urban building extraction based on VHR imagery, there are still some demanding challenges to master, specific to extracting buildings.

Scale invariance for example considers the classification of buildings of different sizes, which is often a problem because of the small image patches, which are fed through the network due to limited GPU memory. That means each image patch or tile (in this study their extents are specified to 224 pixels edge length) is predicted separately, and therefore lacking a bigger field-of-view (FOV), and thus lacking spatial context. As bigger buildings may consist of several tiles, it is difficult for the classifier to assign each of the concerned tiles as one belonging to a bigger context (Ji et al. 2019).

Stiller et al. (2019) also addressed the challenge to accurately classify buildings with different morphological appearances and that it is recommendable to train the

network on all occurring building morphologies and arrangements in the study area. The study area in this project covers most parts of the slopes at the east and the west of the city, thus covering a big variety of building structures and morphologies. Especially, the very characteristic heterogeneous morphologies and organic arrangements of informal settlements (see Wurm et al. 2019), which make up for approximately 67% of the study area according to the cadastral data, are very different to the structures of the south east of Medellín. They consist more of bigger quadratic buildings blocks with different roof materials or also of single buildings with various morphologies in wealthy gated communities (spa. *barrios cerrados*) like e.g. in the *communa El Poblado* (cf. Sanín Naranjo 2010).

However, such high accuracies, which already have been achieved by scientists extracting buildings with DNNs while still facing essential problems and challenges, already proved the capability of those methods, and that there is a clear potential to improve classification results in the near future.

1.2 Scope of the study

The landslide related research in Medellín brought some important knowledge about slope stability, risk estimations, mitigation strategies and cost estimations in landslide-prone areas. While there are estimates of the economic consequences of damaged infrastructure, there are no estimates of the number of inhabitants.

The framework of this thesis builds upon three main data sources to enable an accurate estimation of the number of inhabitants, who are exposed to a certain landslide risk. These are:

- a publicly available geodata set with population data for each of the 249 *barrios* (smallest administrative unit in the urban area) (chapter 2.2.1).
- a mass movement threat map based on empirical studies and past events, that was jointly created by the university of Medellín and several municipal authorities (chapter 2.2.2),
- an also publicly available orthophoto covering the municipality of Medellín with a spatial resolution of 8 cm (chapter 2.2.3).

Methodically, the goal is to use building footprints to disaggregate the population data from the smallest administrative unit to a more local building block related level. The footprints will be derived from the orthophoto by a semantic segmentation approach using the deep learning architecture U-Net. In this major step the appropriate setting for the U-Net will be examined and defined. It is often time consuming to fit the hyperparameters, as well as fitting the general settings of the architecture to the data basis to enable an efficient processing as much as possible. To give a better understanding of the process and concept of defining the final model the method for this process will be presented in 3.5 and the results in chapter 4.1.

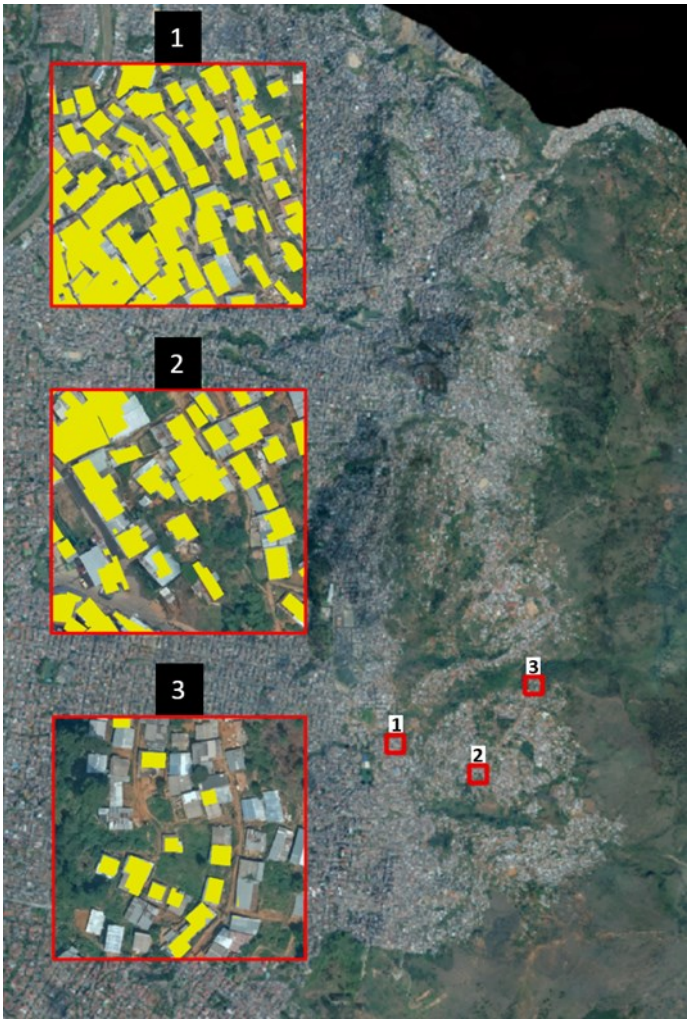


Figure 2: *Example of how the cadaster becomes increasingly inaccurate towards the margin of the city. Data basis: Secretaria de Tecnologia (municipio de Medellín).*

After generating the building footprints, they are intersected with the hazard map to identify built-up areas, which are located in zones with at least medium threat (namely in the Mass Movement Threat Map: 'Criticality 3 – medium threat'). After the identification of those sites the population will be estimated at a local, building block or sub-neighborhood level.

Choosing a building extraction from an orthophoto over the use of the available cadaster data for the disaggregation has two main reasons: firstly, the building footprints of the cadaster become more inaccurate in terms of spatial location and extent of the buildings towards the margins of the city. Secondly, the city experiences such high population dynamics, which has already been stressed in chapter 1, that the cadaster data is

not complete and cannot be expected as to be up-to-date (fig. 2). However, regardless of those spatial inaccuracies the cadaster data of Medellín is still quite comprehensive.

This cannot always be expected for cities dealing with natural hazards and high population dynamics, especially in economically weak countries and cities. Moreover, rural regions or smaller cities do not necessarily have a cadaster at all. For example, in the Aburrá valley sub-region only one out of ten municipalities actually provide cadastral information (Acevedo et al. 2017). Thus, acquiring an up-to-date Orthophoto, with an UAS (unmanned aircraft system) for example, might be significantly less expensive and time consuming than generating a new cadaster or keeping a cadaster up-to-date in a very dynamic urban environment in order to make information of built-up area available.

As the cadaster of Medellín is comparatively comprehensive and only becomes inaccurate towards the margins of the city it is a valuable source of comparison for the deep learning-based building footprints. By comparing both data sources, it is possible to exhibit how far the amount of affected population diverge as the cadaster becomes more incomplete, and consequently how important it might be to use modern remote sensing techniques to acquire up-to-date data.

This thesis represents an integrative approach at the interface between anthropogenic living spaces and natural processes, using remote sensing techniques in order to derive an estimation of population being exposed to landslide hazard. According to this approach, following research questions are raised:

- How accurate is the building extraction using DNN's?
- How good is the ability of a state-of-the-art neural network at learning a broad diversity of building morphologies?
    - What are the key limitations?
- Which uncertainties occur, when estimating affected population?
- How many residents are exposed to landslide risk?
    - How far do the statistics diverge from those derived from the cadaster?

- Is there evidence that informal settlements are more at risk of landslides, or conversely, are wealthier areas usually less at risk of landslides?

2 Investigation area and data

After presenting the historic and scientific context as well as introducing the framework of the thesis, this chapter will provide a detailed overview over the data, which will be utilized for the deep learning-based building extraction as well as for the following exposure analysis. Further on, the decision for the specific investigation area (also referred to as 'area of interest') will be made and clarified based on further details about the spatial socio-economical patterns in Medellín.

2.1 Settlement structure and related hazard susceptibility

As one of the biggest cities in Columbia, Medellín with its approximately 4 million inhabitants is located in the Central Cordillera. In this part of the Andean mountain range the city is embedded in the northern Aburrá Valley (span. *Valle de Aburrá*) in the northeast of Columbia (Hermelin, 2012). The Aburrá Valley is also a political and administrative region of which Medellín is the capital (Ferrari et al 2017). It is a densely populated region and highly urbanized. The whole area of the municipality extents over 376 km², whereas 101 km² belong to the urban zone and 275 km² belong to the townships (span. *corregimientos*) of the rural zone (Secretaria de Tecnologia (municipio de Medellín)).

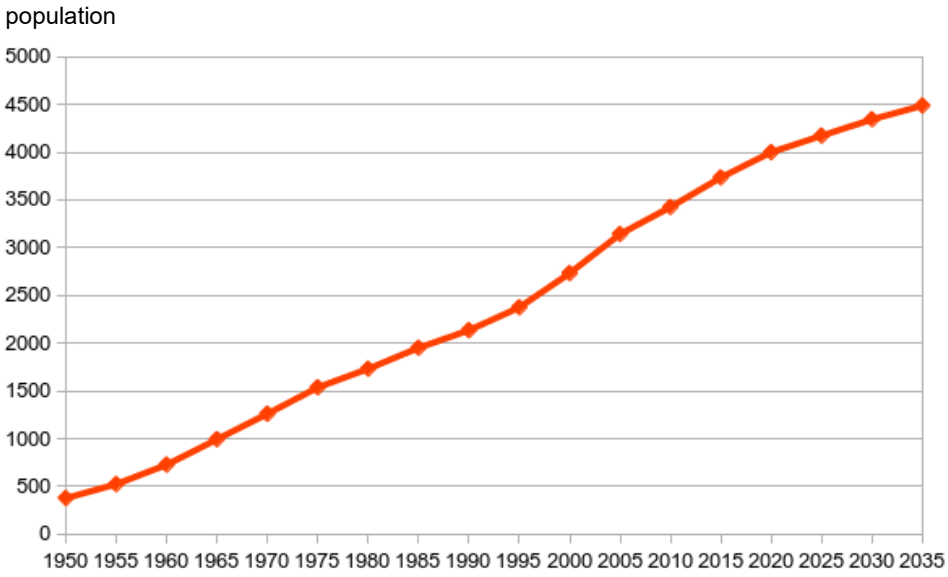


Figure 3: *Urban population in Medellín. The data is based on the census of 1951, 1964, 1973, 1985, 1993 and 2005 and was estimated in 2017 by the UN. Source: United Nations Department of Economic and Social Affairs, Population Division (2018).*

The administrative units of the municipality are divided into 16 communes (span. *Communas*) and 249 neighbourhoods (span. *Barrios*) in the urban zone and 5 *Corregimientos* divided into 54 veredas in the rural zones. In Medellín the strong urbanization dynamics are related to the political and drug-related violence in the past century, as well as due to natural disasters, as described in chapter 1. Due to a new migratory wave in the 1950s the city doubled its population within ten years. The new inhabitants settled informally at the margins of the city and the growth of the "informal city" started (Restrepo & Orsini 2012). As a consequence, the city

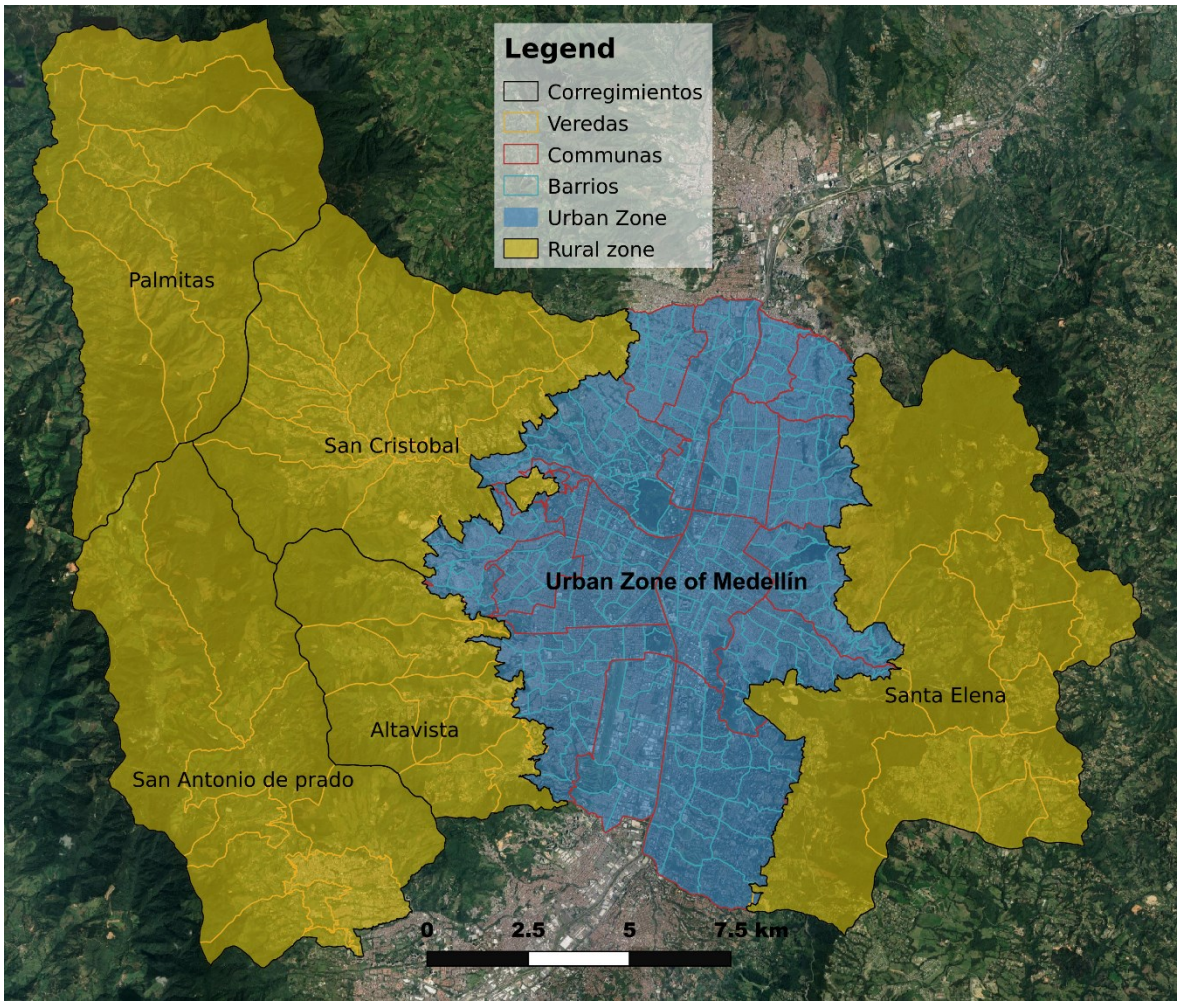


Figure 4: *Overview over the administrative boundaries and the extent of the city. Data basis: Secretaria de Tecnologia (municipio de Medellín); Google Maps.*

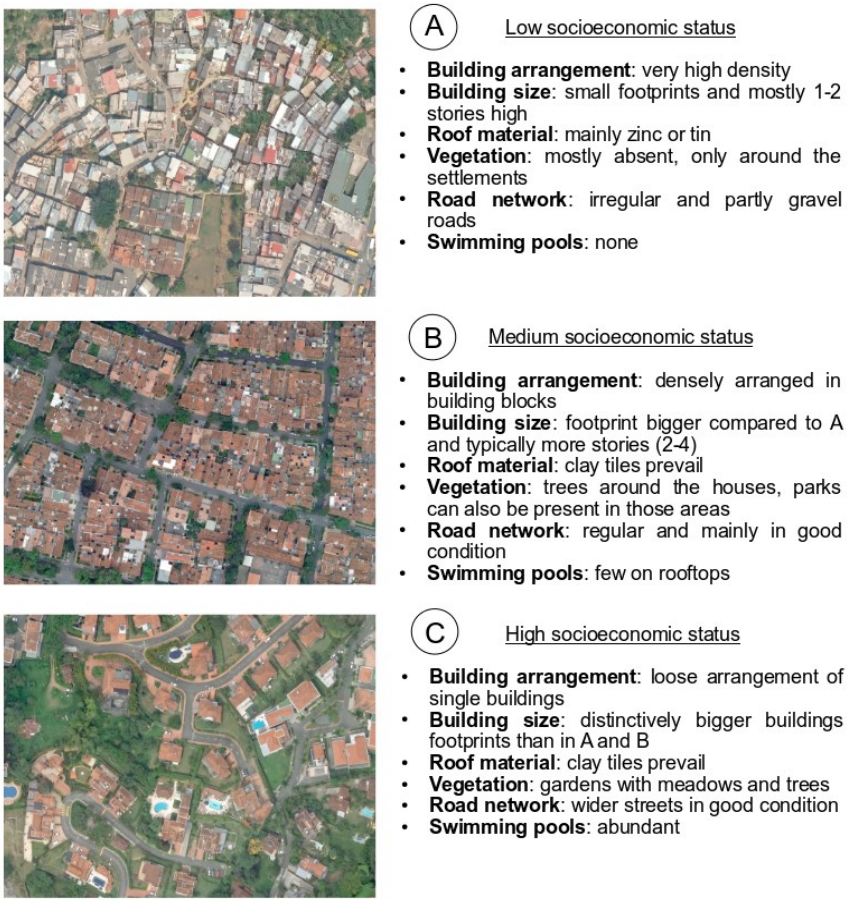


Figure 5: *The three mainly occurring residential building morphologies in the investigation area. Source: cf. Quinchía 2012.*

expanded to highly inaccessible areas. Additionally, infrastructure and public services were completely absent in the early stages of the informal settlements.

The uncontrolled expansion of the city led to a “profound segregation of the city’s physical, social and economic order” (Restrepo & Orsini 2012, p 137f). Due to the effort of the government, especially due to the implementation of the *social urbanism* (see chapter 1), the infrastructural and social situation is not as precarious as it was before, but still the buildings of the informal settlements consist of low-quality building fabric and the expansion of the city still progresses (Ferrari 2017, Vega 2016, Restrepo & Orsini 2012).

The observations from above using aerial imagery in this thesis, show that the aforementioned segregation leads to distinct differences in the city’s morphology. Quinchía (2012) names several features, which enable this visual distinction between areas with different socioeconomic status. These are mainly:

- building arrangements
- roof materials
- presence of landscape vegetation and urban amenities
- the shape and density of the road network
- and the presence of private swimming pools.

Figure 5 represents the three main appearances of residential buildings, which are present in the urban zone of the city. They are termed as low, medium and high concerning its socio-economic situation according to Quinchia (2012). The structure of fig. 5A shows a typical representation of informal settlements. The buildings are typically one to two stories high and have small footprints. They are densely and often irregularly arranged. The roofs mainly consist of zinc or tin tiles. The pathway infrastructure shows irregular patterns as well, compared to the medium and the high socio-economic areas of the city. Contrary to both of the other morphologies the roads are often not paved. Fig. 5B is a typical representation of residential areas with a medium socio-economic status. A very clear feature is the chessboard arranged building blocks with red roof tiles. Often, they are two to four stories high and have a small inner courtyard. Additionally, the streets are relatively wide and planted with trees at the sides. Similar characteristics apply for the areas with high socio-economic status, with the main difference that their arrangement is not chessboard-like. The building landscape is much more open (fig. 5C), which means that detached houses dominate the view from above. A further dominant feature is the abundance of swimming pools and the generous size of the gardens. Often the high socio-economic residential sites are embedded in gated-communities (cf. Quinchía 2012).

In chapter 2.3, as the investigation area will be introduced, a further discrimination of the building morphologies will be made in order to consider the predominant morphologies for the training of the deep learning model. This is more related to a technical background than to a socio-economic representation.

2.2 Utilized data for the building extraction and the analysis

2.2.1 Cadaster & geodata of Medellín

Numerous geographic datasets are provided by the city council through the geoinformation service ‘GeoMedellín’. The data available on their website (https://geomedellin-m-medellin.opendata.arcgis.com/) is free of charge. Its data can and has been downloaded as shapefiles. It is very comprehensive and shall strengthen the transparency of the municipality’s administration and facilitate participation of the inhabitants. A further motive for the free distribution of the cadastral data expressed by the major’s office is, that the open data policy shall allow third parties to work with the information and generate new knowledge (cf. Alcaldía Medellín 2020), which is also one intention of this thesis.

To get an overview of the administrative boundaries and the layout of the city, the following two layers were used: firstly, the *Límite Catastral de Comuna y Corregimiento* layer, which represents the boundaries of the communes in the urban zone, as well as in the townships in the rural zone. Secondly, the *Limite Barrio Vereda Catastral* layer, representing the boundaries of the *barrios* and the *veredas* in the urban and the rural zone, respectively. The neighborhoods are the smallest administrative unit in the urban zone, whereas the *veredas* are the smallest administrative unit of the rural zone (see fig. 4).

Despite administrative boundaries the cadaster data of the entire building plots of the municipality has been used. This layer is called *Construcciones y Mejoras del Predio* and contains not only spatial information but also further information about the typology such as: ‘transitory construction’, ‘swimming pool’ or ‘building’. However, it doesn’t take the division of property into account (Alcaldía Medellín 2020). These building plots or building footprints are used to later disaggregate the population data from the neighborhood to a local, building block related level. This will be conducted analogous to the population disaggregation with the deep learning-based building footprints.

A further used data source is the ground value layer of Medellín, which is represented in sub-neighborhood areas with a specific ground value. This data is

used in the exposure analysis to get a better understanding of how strong the poorer population is affected by medium or high landslide hazard.

2.2.2 Mass movement threat map

Areas in the city which are prone to landslide events are determined by the mass movement threat map, that was published in 2009 as part of the land-use plan of Medellín. It is essentially a review and adjustment of the previous land-use plan of 2006. Hazardous areas in terms of mass movements have been defined by the publishers (National University, AMVA, CORANTIOQUIA, Municipality of Medellín

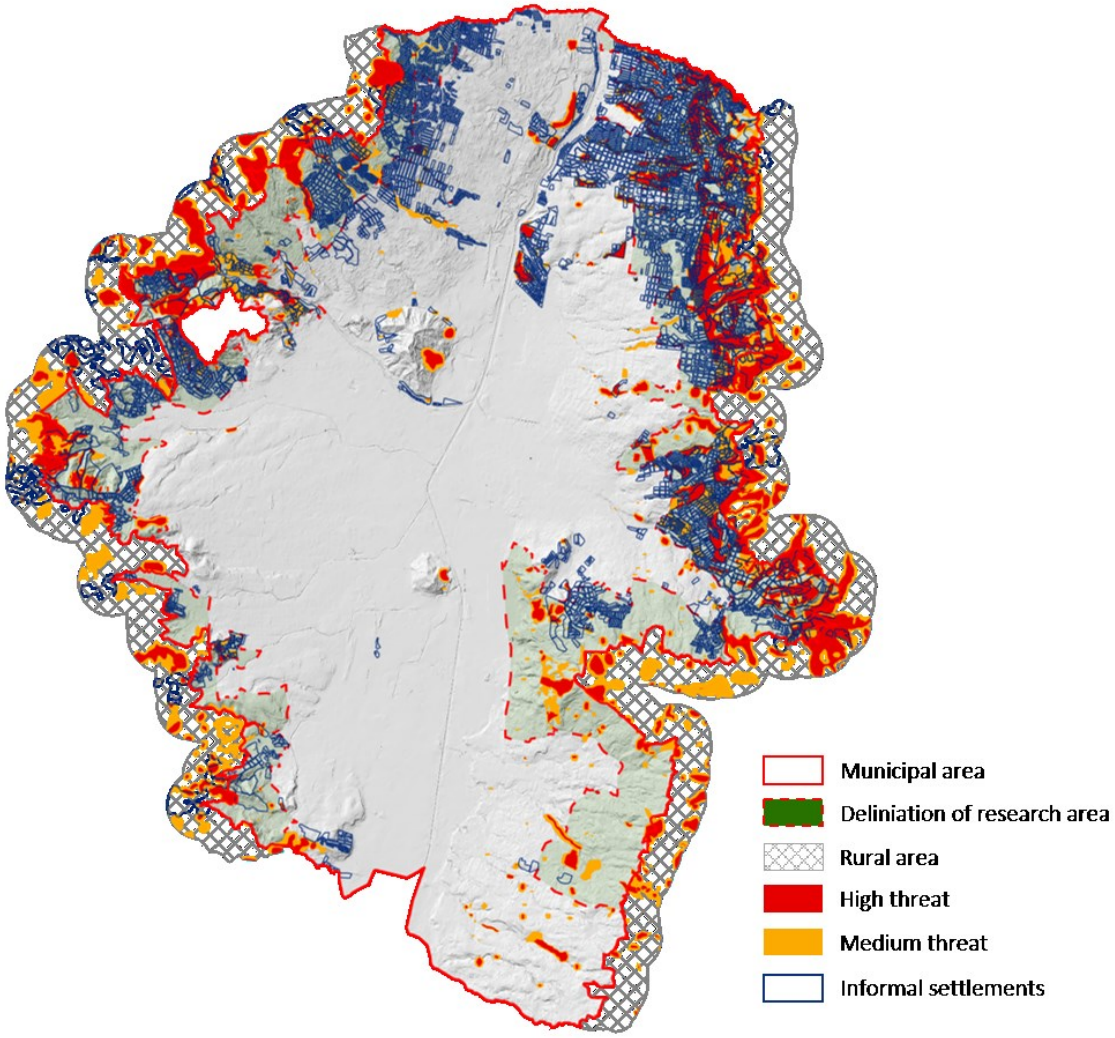


Figure 6: *The distribution of landslide hazard in the urban zone of Medellín. The shaded relief-based map provides an overview of the distribution of hazardous sites and of the distribution of informal settlements, as well. Data basis: Secretaria de Tecnologia (municipio de Medellín); National University, AMVA, CORANTIOQUIA, Municipality of Medellín and Municipality of Envigado, 2009.*

and Municipality of Envigado, 2009) as areas with latent danger of a physical event of natural origin or as one accidentally caused by human action with sufficient severity to cause loss of lives, injuries and health impacts in general. The same applies for infrastructure, livelihoods, or environmental resources affected.

These areas are represented in the threat map as five different classified categories of criticality, which are listed below.

- Criticality 1 (Probability between 0 – 0.2%): “very low threat”
- Criticality 2 (Probability between 0.2 – 3.2%): “low threat”
- Criticality 3 (Probability between 3.2 – 5.5%): “medium threat”
- Criticality 4 and 5 (Probability between 5.5 – 16.5%): “high threat”

The classes are based on an assessment of geomorphological parameters such as slope, underlying geology and geotechnical properties. Despite these information, which are the result of several years of geomorphological and geological studies conducted by the city of Medellín since 2006, also the inventory of mass movement events, prepared by the Administrative Department of Disaster Management (DAGRD) and the Administrative Department of Planning (DAP), have been considered in the classification. Thus, for identifying mass movement threat, the participants in that project made a qualitative analysis of the most recurrent rupture mechanisms and magnitudes, using geological and geomorphological information, morpho-dynamics and the characteristics of the most relevant historical events in recent decades (National University, AMVA, CORANTIOQUIA, Municipality of Medellín and Municipality of Envigado, 2009).

Besides the depiction of medium and high threat areas in figure 6, the investigation area and the distribution of the informal settlements has been visualized, as well. This will provide a better overview of the general landslide hazard, but also of the pronounced hazard for the particularly vulnerable population of Medellín living in informal settlements.

2.2.3 Orthophoto

The remote sensing data used for the building extraction is an orthophoto, and like the cadaster data, it can be accessed via the geoservice ‘GeoMedellin’ (Secretaria

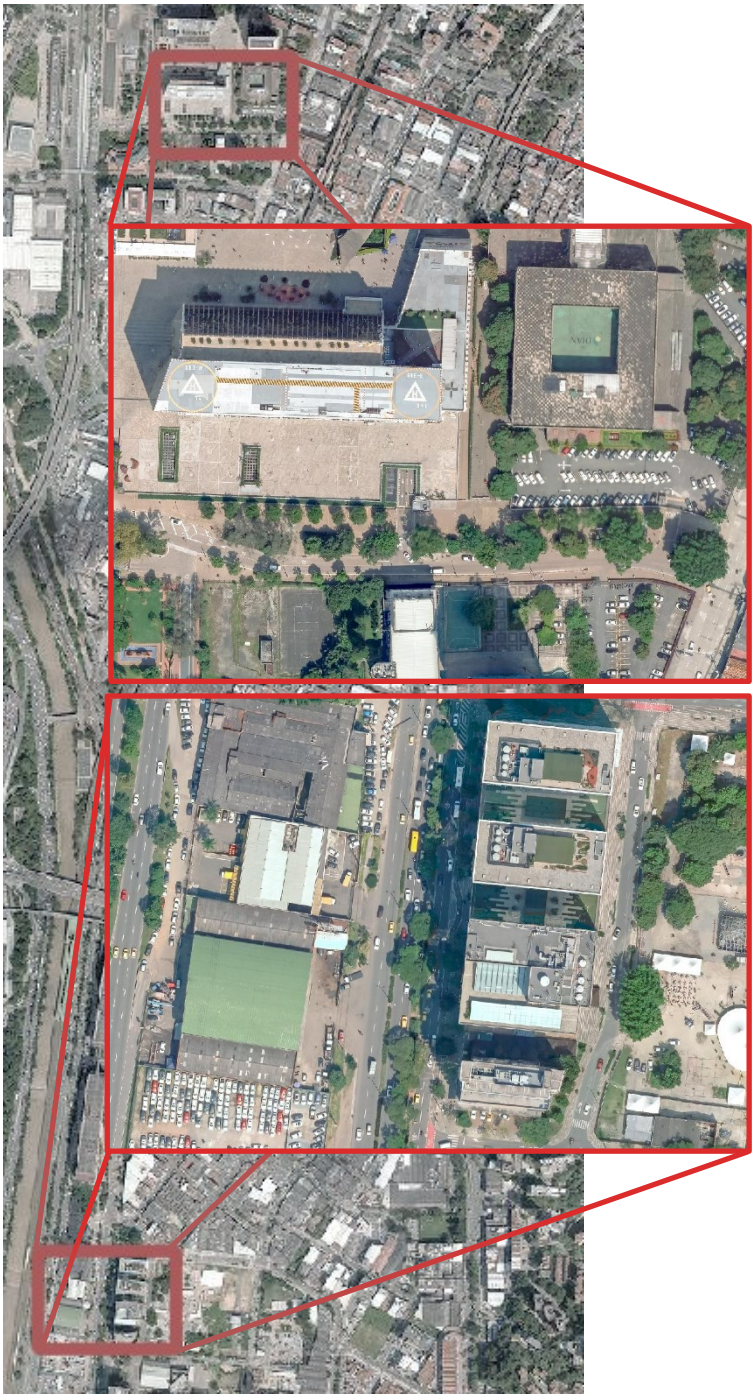


Figure 7: *Example to showcase the positional accuracy of the orthophoto.*

de Tecnologia (municipio de Medellín)). All the data from the portal can also be accessed through an ArcGIS Map Server to export and work with it, e.g. in a Geographic Information System (GIS). The orthophoto covers everything within the borders of the city. That are all 16 communes in the urban zone as well as the 5 townships in the rural zone.

According to the description of the dataset, it is a true orthophoto generated with photogrammetric methods in 2019. Typically, those kinds of datasets are generated by a large number of optical images, with a certain overlap (typically over 50%) to enable a multi-angle view for a more precise orthorectification. The images can be taken from an aircraft or an UAS. Traditional orthophotos are acquired through a perspective projection, which leads to scale variation and relief displacement. The rectification process aims to eliminate those errors. However, in large scale urban areas the relief has typically a high variation with sudden changes and therefore, major

artefacts can be created from double mapped areas. Apart from these artefacts, hidden areas ‘behind’ buildings often occur, due to the angle between the object and the nadir position of the sensor. With a high number of overlapping images, pixels in hidden areas can be filled with actual data, and distortion caused by a static perspective can be minimized (Chen et al. 2014, Habib et al. 2014).

The image section in figure 7 shows two sites with a relative distance of about two kilometers. By comparing the buildings in both images, one can see that the distortions are minimal and tend to occur only with high buildings. This is not unimportant when it comes to labelling houses for setting up the training data for the DL network. If a facade is visible it will most likely be classified as a building, because the texture and shape is similar. But since the orthophoto has such little positional errors, the calculation of the built-up area will be very accurate as long as the classification algorithm is accurate and the extracted building footprints correspond to the rooftops.

2.3 Introducing the investigation area

As indicated in the above sections, by pointing out which parts of the city are most exposed to landslide hazard, the areas of interest (AOI) in this thesis comprise the eastern and western slopes of the city. The delineation of the investigation area is derived by the anticipated goal to quantify population that is exposed to landslide hazard. Therefore, such neighborhoods have been chosen, where at least intermediate hazard is defined in the Amenaza hazard map. To choose whole administrative units for the analysis over a delineation by a distinct line, deciding if an area is of interest or not, would be too unhandy, because hazardous areas in the hazard map have all kinds of shapes and this would result in countless scattered little AOIs. Additionally, as the hazard map is just an estimation of a certain risk, it is more meaningful to comprise a bigger and contiguous AOI. The specific areas with different stages of risk can then be considered and locally assigned in the analysis afterwards.

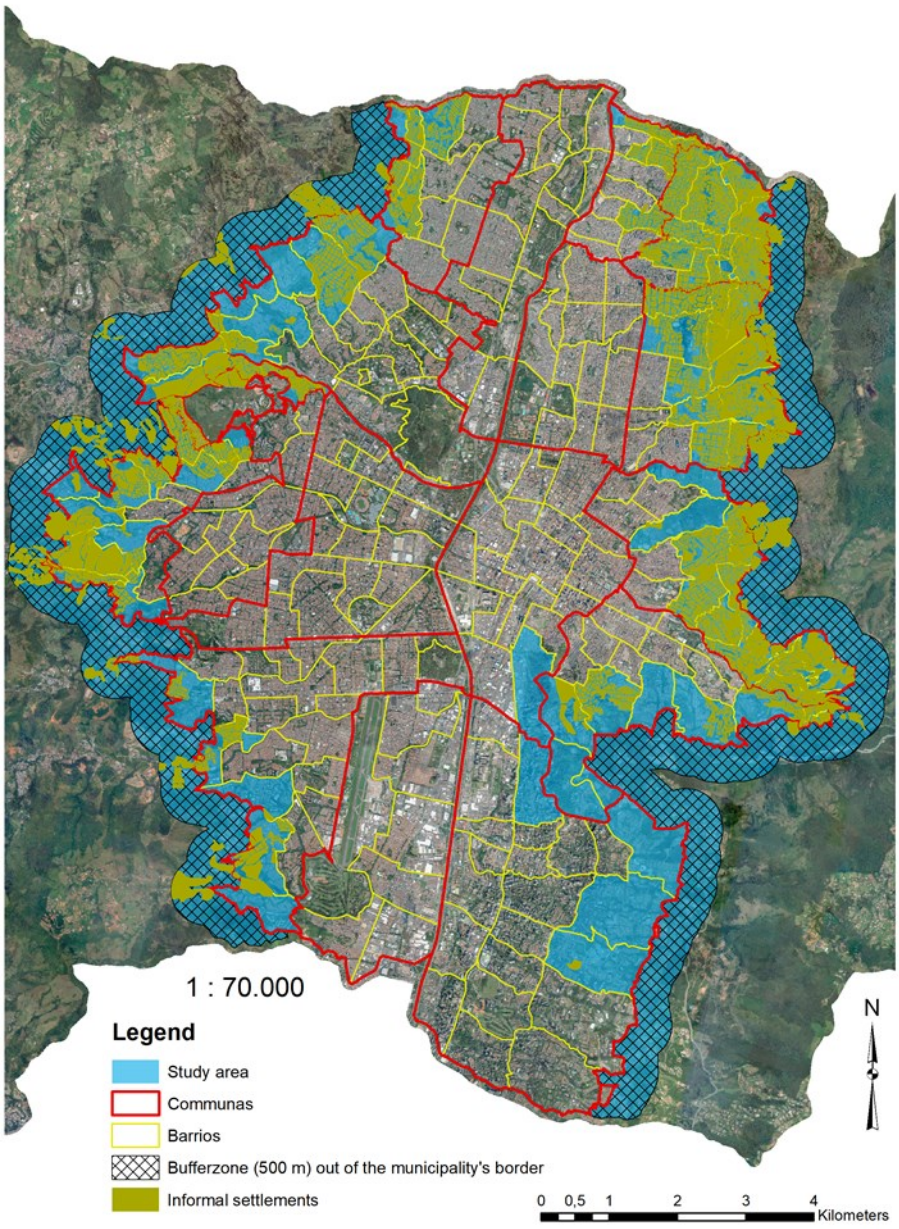


Figure 8: *The study area or neighborhoods of interest. The Map also includes the buffer zone beyond the municipality's border, where the informal settlements outgrow the official border of the urban zone. Data basis: Secretaria de Tecnologia (municipio de Medellín)*

Figure 8 highlights the neighborhoods of interest as well as a buffer zone, which extends the AOI by 500 meters beyond the municipality's border. This has been done, to also include the settlements which are now located outside of the city's urban administrative border, due to the steady expansion of residential areas in the last decades. The location of the informal settlements in the AOI are highlighted as building footprints (bold yellow) in the map. This overview underlines the large

relative share of such settlements and thus of the lower-income population in the whole AOI.

3 Methodology

3.1 Conceptualization of the deep learning approach

3.1.1 Basic concepts of deep learning in image analysis

In the vast landscape of DL architectures, there is a great variety of architectures available to choose from. These architectures can have very different structures and serve different purposes. In image classification tasks there are four basic classification approaches to choose from (see fig.10): image recognition, object detection, semantic segmentation and instance segmentation.

In image recognition the network analyses a scene and assigns one label to it. In object detection the algorithm is able to recognize and locate a specific object in the scene by drawing a bounding box. In image segmentation, which is also known as semantic segmentation, the desired classes are predicted for each pixel and thus creating the footprints of the found objects. Instance segmentation combines both, image segmentation and object detection. Having the objects footprint as well as the bounding box makes objects in scenes countable (Höser & Künzer, 2020).

In this thesis it is required to locate buildings in an aerial image and quantify their area. Therefore, sematic segmentation has been chosen as the underlying approach. Semantic segmentation in the context of remote sensing is mostly based on CNNs, which have one important advantage or capability, which enables meaningful classification results in imagery data. CNNs are able to learn representational features to an increasing abstract degree from the original input image. Since the calculation of these so-called feature maps are realized by kernel functions, the local arrangement of the pixel values can also be extracted. The feature maps are essentially texture features. But in contrast to conventional texture

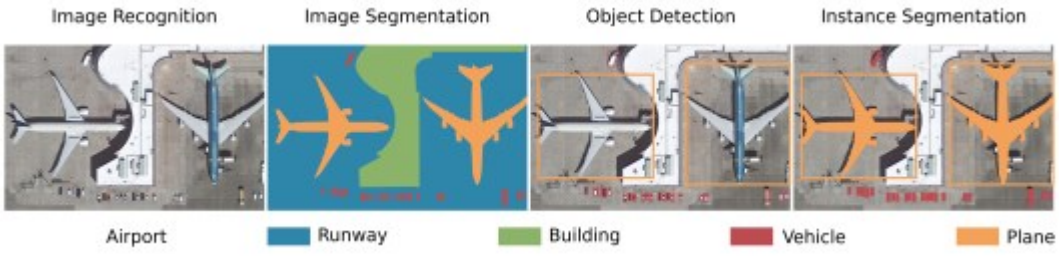


Figure 9: *The different types of deep learning-based image classifications. Hoeser & Kuenzer 2020, p.8.*

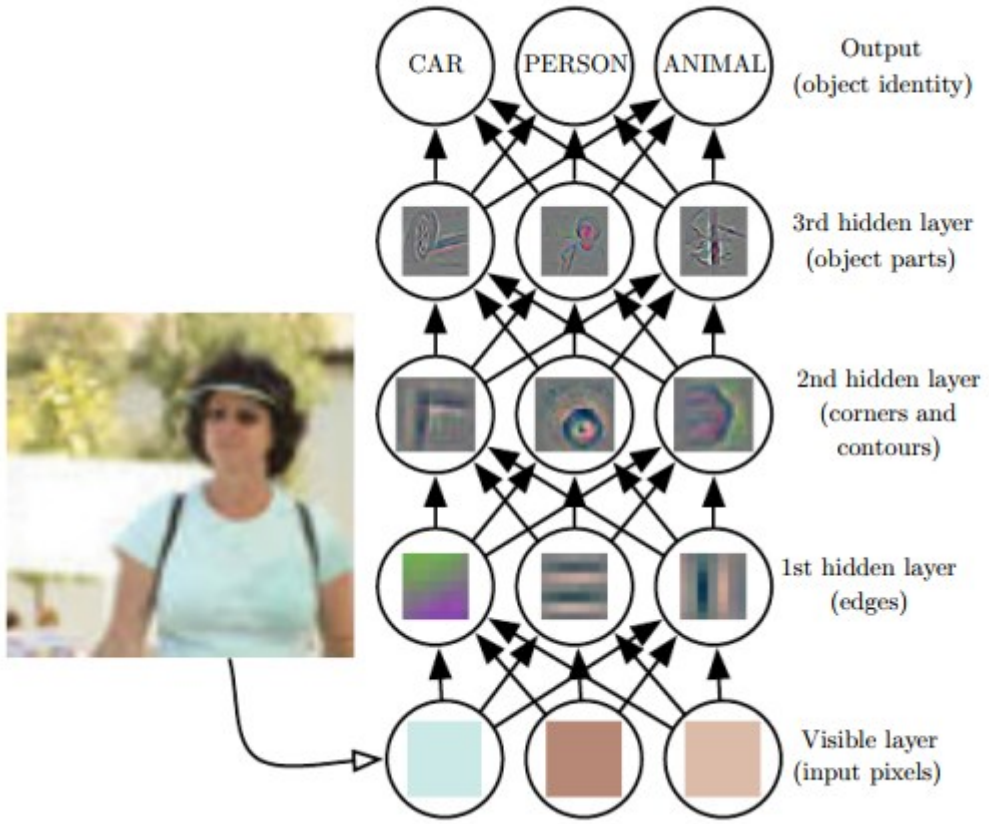


Figure 10: *Representation of the general concept of a deep learning network. The figure shows the typical structure, beginning with the visible layer (or input layer) and continuing with extracting image features in hidden layers to eventually yield the output layer or prediction. The principle how the features become more a more abstract with each consecutive hidden layer is also made visible. Goodfellow et al. 2016, p. 6.*

features like the GLCM, the feature maps can be calculated from a multilayer imagery. Figure 10 shows an example from Goodfellow et al., 2016, to get a visual and better understanding of the feature maps and how they get more abstract with increasing number of layers. Those layers typically referred to as convolutions. Besides the graphical depiction of abstract features, this illustration also shows the typical feed-forward manner in CNNs learning object representations. That means, that the information learned is always connected with the following layer of representation until the final prediction is made.

To stay with the basic concept of DL, it is meaningful to briefly explain some key terminology. The first layer of representation is the actual data, which is therefore called visible layer. The subsequent layers are called hidden layers, since they do not represent actual data, rather they have been created by non-linear functions to extract meaningful features, which are typically only visible for the neural network.

As mentioned before, the hidden layers create increasingly higher-level features. That can start for example with a local edge detection filter, where the next hidden layer seeks for arrangements of edges and so on, until actual objects or parts of objects are represented in those feature maps.

The smallest part of a neural network is called a ‘neuron’, which represents each pixel in an image fed through the network in semantic segmentation tasks. However, one neuron representing one pixel does not apply for object detection tasks like in fig 10. But, essentially the structure is the same. In most graphical representations of a multi-layered neural network the neurons are presented as circles. Each one holds the grey scale value of the input image or of each input band of a multiband image, like a red-green-blue band (RGB) Orthophoto. All neurons are so called fully-connected with trainable non-linear functions, which nowadays is mostly the Rectified Linear Unit (ReLU), depicted in fig. 11 (Höser & Künzer 2020, Goodfellow et al. 2016, LeCun et al. 2015).

During the training procedure of a neural network the weight vector and bias included in the ReLU function are adjusted iteratively, after each cycle when an image is fed through the network and a label is assigned per pixel. There are two major components, that enables learning and therefore adjusting the weight vectors and biases. This is on the one hand the gradient vector and on the other hand the objective function. The latter function summarizes all the differences between the classified output and the reference label for all samples and therefore for each

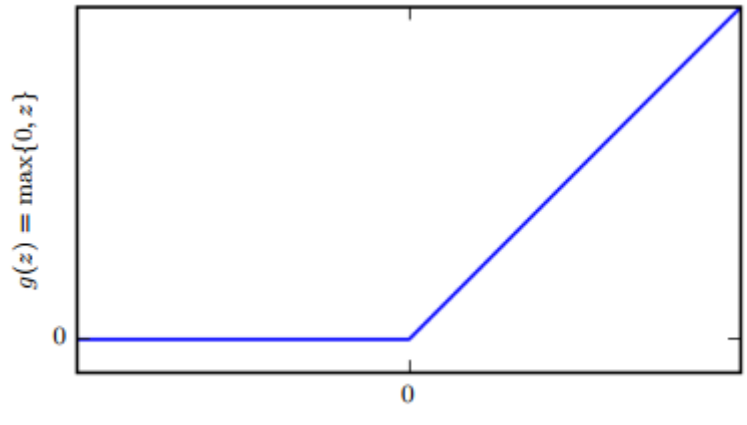


Figure 11: *The Rectified Linear Unit (ReLU) function. The weights and biases of each neuron are altered by that function, during each iteration of backpropagation. Goodfellow et al. 2016, p.175.*

neuron in the network. A common objective function for example is called cost function and calculates the mean squared error for the output. The gradient vector is calculated for each weight and it indicates by what amount the objective function increases or decreases after the weight has been adjusted. The weights will then be adjusted to the opposite of the gradient vectors. That means, that the opposite weight vector or gradient vector indicates the direction for the weights to change, in order to decrease the high dimensional cost function. This happens starting from the cost function at the output stage of the network, changing the neurons vectors layer by layer on the way back to the input layer. This is also called backpropagation. (Goodfellow et al. 2016, LeCun et al. 2015).

3.1.2 The applied U-Net architecture

In this thesis the semantic segmentation of buildings in an urban RGB Orthophoto utilizes the popular and still state-of-the-art U-Net architecture. It has already been applied for several remote sensing tasks. These are for example the detection of burned areas caused by wildfires (Knopp et al. 2020), water body extraction (Feng et al. 2018) and road extraction (Chen et al. 2021). But it has also been utilized for building extractions as for example in Wurm et al. (2021), Xu et al. (2018), and in Lu et al. (2018).

Originally the U-Net was built to track cell borders in medical image analysis. Its architecture is based on the Fully Convolutional Network (FCN) for semantic segmentation proposed by Long et al. (2015) and has been adapted to deal with little training data, which is a typical constraint in biomedical tasks. As shown in figure 12 the network consists of a contracting and of an expanding part. The contracting part is called encoder and the expanding part is called decoder. The main difference compared to the FCN is the decoder. After reducing the resolution and extracting deep features of the original input image the FCN has one up-sampling module to transfer the features to the original resolution, performing a pixelwise prediction. In contrast, the U-Net has got a decoder part, which is nearly equally structured like the encoder part. Combined with skip connections (indicated by the horizontal arrows in fig. 13), features of all four stages in the encoder and therefore, features with the whole spectrum of abstraction and resolution will be

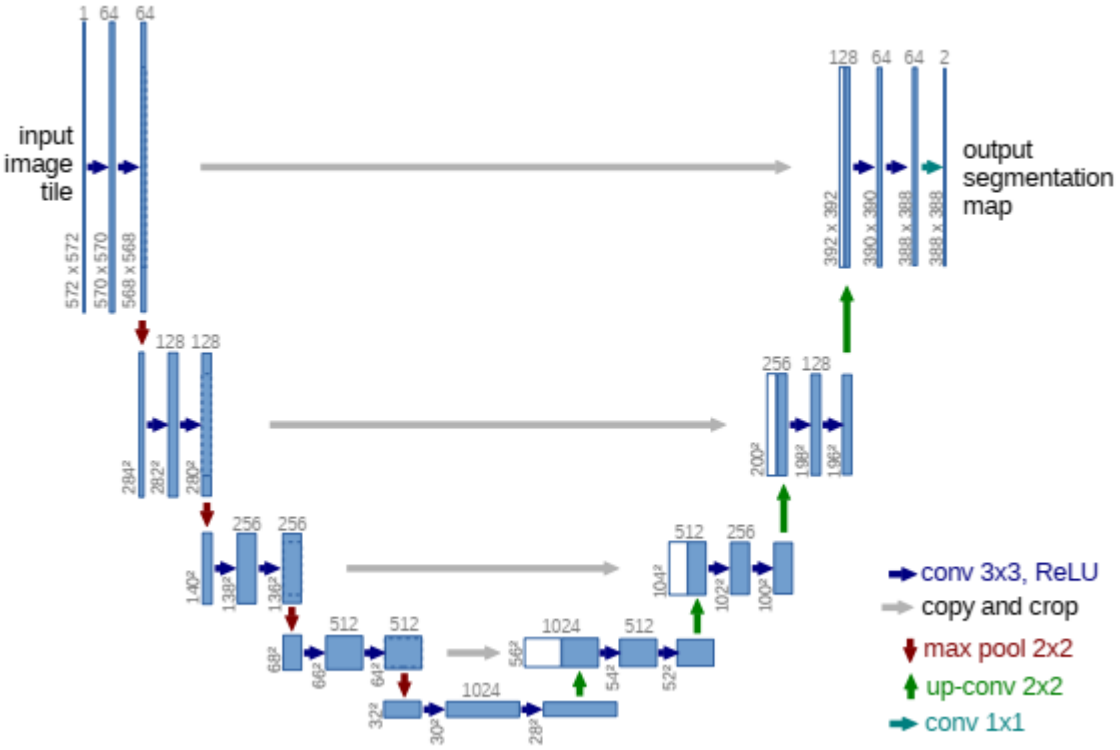


Figure 12: *The U-Net's architecture. Ronneberger et al. 2015, p.2.*

added to the decoder in the respective stage. This enables a more effective and detailed localization of the objects in the image. Additionally, Ronneberger et al. (2015) used extensive data augmentation to deal with less training samples. Data augmentation describes the process, where the input images will be transformed in many ways, like twisting, flipping, applying geometric distortion or transforming grey values. Those functions can be provided with a specific probability of occurrence. Besides getting a reliable model with better generalization ability despite small training sizes, data augmentation is often applied to yield spectral and geometric invariance, as well (Ronneberger et al. 2015, Goodfellow et al. 2016).

Dealing with a small sample size is also an asset for the building extraction in this thesis. Thus, the U-Net has been chosen as the main architecture. Having relatively little training data can be expected, since creating labels by hand is tedious and also the samples are quite limited due to the sampling design, where the four chosen building morphologies for training are trained separately and ought to be free from other building morphologies as far as possible. Details about the sampling design will be given in chapter 3.2.1.

In the depiction of the U-Net's architecture (fig. 13), the vertical numbers to the side of the blue layers indicate the resolution of the image. The horizontal numbers on top of these layers indicate the depth of the layers or the number of feature maps. Each of the arrows between the layers indicate the following kind of operation. One can see, that in each of the four stages in the encoder as well as at the bottom of the network two successive convolutions are followed by a ReLU function. The convolutions represent the typical layered structure explained in section 3.1.1, but without the output since the convolutions serve only the extraction of features. The dimensions of the kernel used in the convolution operations are three by three pixels of width. The transition from every stage is made up by a two by two max pooling operation with a stride (number of pixels used for each step of the moving kernel) of two. Max pooling means, that the feature map is down-sampled with a two by two moving kernel using the maximum pixel value within. At each down-sampling step the number of features is doubled.

As already mentioned above, the decoder part of the architecture has basically the same structure as the encoder part, where the max pooling operators are replaced by up-convolution operators in order to increase the spatial resolution and halve the number of the features again. However, due to the convolutions using a sliding window the extent of the image passed through the network reduces and the border pixels get lost.

In image recognition tasks it is common to change or modify the initial architecture to better fit ones needs. Especially, since the U-Net is around since 2015 it is meaningful to also modify the architecture to combine it with modern feature extraction methods. But nonetheless, it should be emphasized that the U-Net as the underlying architecture is still widely applied, which speaks for its success. In this thesis the U-Net will be modified with one of two tested backbones. This is a) the ResNet (core paper: He et al. 2016) and b) the EfficientNet (core paper: Tan & Le 2019). The ResNet could already achieve good results in combination with the U-Net in a building extraction task (Xu et al. 2018) and is very commonly used for feature extraction (cf. Höser & Künzer 2020.). The EfficientNet is more modern and has been designed to be more efficient in depth, width and resolution. The authors

focused on scaling a model in order to make it lighter, but also precise or even more precise in its predictive ability.

Both backbones replace the standard U-Net feature extraction modules and will be compared in an experimental approach (chapter 3.2.2) to use the best one of them for the final prediction.

3.2 Data Processing

3.2.1 Data setup: pre-processing

The first step in the classification process is the preprocessing of the orthophoto in order to set up the training data. Therefore, training areas have been chosen and manually digitized in the scene as the first step. With reference to Stiller et al. (2019) it is meaningful to train all urban building structures, which occur in the area of prediction. Thus, the decision for training areas has been done under consideration of the different building morphologies presented in chapter 2.1 and depicted in figure 5. In contrast to the above presented morphologies, the residential sites with low economic status are split into two different sets of training data, where it has been differentiated between dense and loose structures. This has been done, because there are wide areas with loose settlement structures at the very margin of the urban zone of the city. This means the classifier has to recognize small single buildings with bare ground and vegetation in the background instead of building blocks or wide heterogenous roof landscapes with little or no background. Furthermore, the

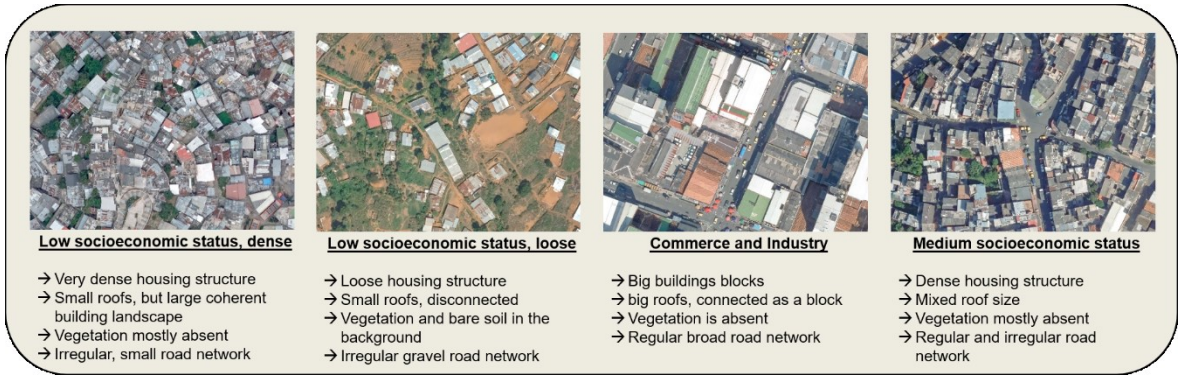


Figure 13: *Overview over the four training sites. These have been defined as the four mainly occurring building morphologies, characterized by specific properties.*

inspection of the investigation area shows that non-residential sites with commerce and industry make up for a reasonable area. As the appearance of those areas differ in terms of roof sizes and materials it has been added as an own training site.

The constellation of data on which the deep learning network is trained on is typically found to be divided into three sections of each set of training data. These three sections are the actual training part, the validation part and the test part (see fig. 14). In this thesis the training, validation and testing parts account for 60%, 20% and 20% of each training area, respectively.

The training areas, which are presented in figure 13, were selected so that each represents only one particular building morphology. This enables the investigation of the achieved accuracy for each settlement structure individually using the respective test part.

After the training areas have been defined, each of them is labeled manually. In order to fit the requirements of the applied deep learning framework, each training set is tiled into image tiles with an edge length of 224 by 224 pixels and an overlap of 1/3, which corresponds to 74 pixels. The overlap is necessary because of the border effects induced by the convolutions in the U-Net architecture. Therefore, the

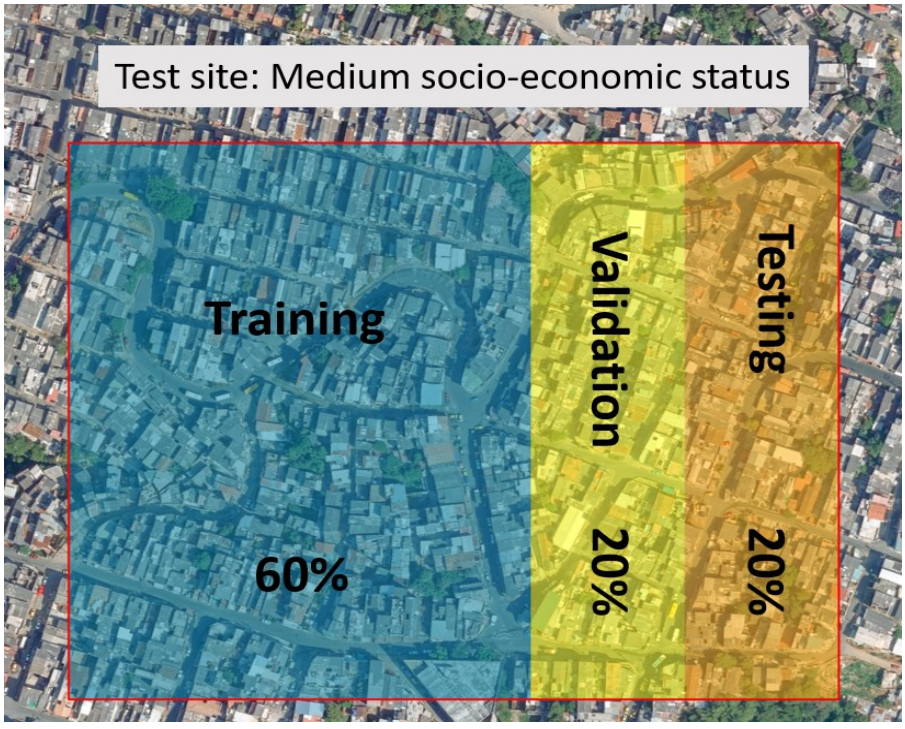


Figure 14: *Schematic illustration of the splitting between training, validation and testing data. Data basis: Secretaria de Tecnologia (municipio de Medellín).*

prediction of pixels becomes less accurate towards the border of each image tile. The tiling also applies for the labelled images, since they need to be fed analogous with the actual orthophoto tiles into the network during the training process.

Because the tiling produces a massive amount of data the investigation area has been divided into 15 sections before the final prediction. Each section has been classified separately and combined again with the other results afterwards. This is a recommendable procedure when classifying such a big area or dataset in order to reduce the risk of losing big parts of the result, e.g. when an error occurs during the classification or even before the actual processing, e.g. during the copying of files.

3.2.2 Experimental: finding a suitable deep learning framework and parameters

When the basic network architecture for the building extraction has been found, the exact setup needs to be clarified. However, there is no manual or textbook, which can recommend an optimal architecture and parameters for the network. The same applies for most classifications, especially for the whole field of machine learning. The reason is that the success or achievement of high accuracies depends on many factors, that each study and dataset have its individual requirements and approaches. Some of the many factors, which require adaption of the neural network architecture and the parameters, are the radiometric-, spatial and spectral resolution of the used imagery, the desired type and number of classes to be classified, as well as the image content. The image itself can represent all kinds of landcover with varying complexity. In order to find the best possible settings in this thesis, multiple classifications have been conducted at the four training sites iteratively in an experimental manner (results in chapter 4.1), where the 20% test data of each site is used for the accuracy assessment.

As mentioned in chapter 1.1.2 the FOV lacks spatial context with 8cm spatial resolution combined with the limitation of the image tiles to 224 pixels. This results in a resolution of 17.92 m on the ground. To better evaluate the field-of-view issue

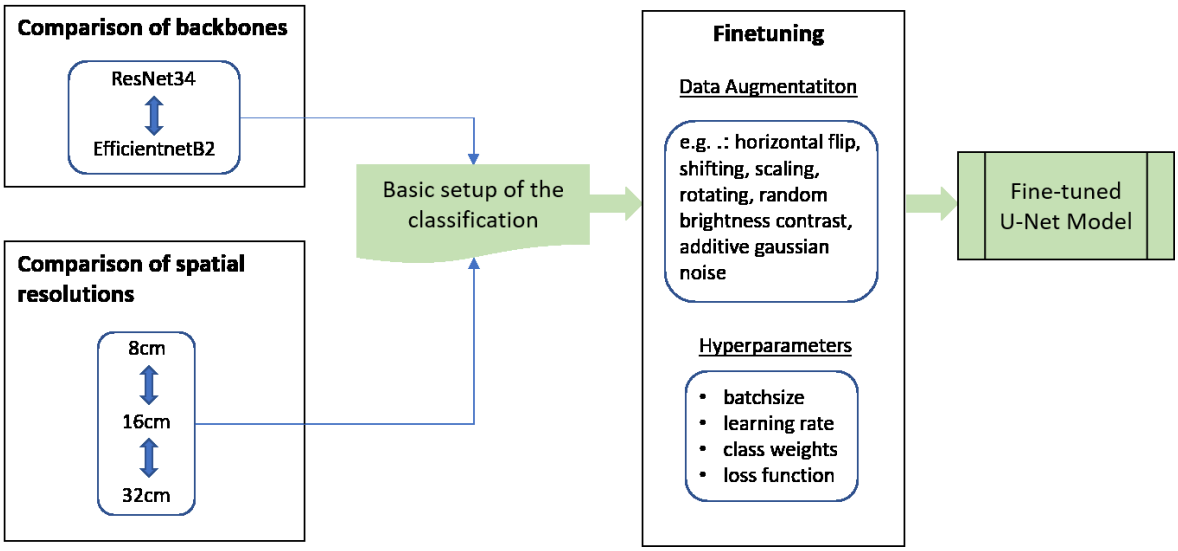


Figure 15: *Conceptual workflow of the experimental approach, where the best possible setup for the final classification is sought.*

and finding a good compromise between detailed image information and spatial context, the experiments have been conducted with three different spatial resolutions. These three different resolutions are the results of halving the resolution by averaging, once for 16 cm and again for 32 cm. This increases the FOV per edge for each image tile to 32.84 m and 71.68 m at the ground, respectively.

All three spatial resolutions have been tested with the two backbones ResNet34 and EfficientNetB2 on all four training sites. After evaluating the performance of the backbones in combination with the different resolutions, the best one will be picked for the final fine tuning of the neural network. The two main components to be adjusted are the constellation of the data augmentation and the underlying hyperparameters. The best performing final setting will be sought iteratively and applied on the whole investigation area afterwards.

Besides assessing the performance at each training area separately, an average classification performance is calculated as well.

The metrics chosen are the precision, recall and the F1-Score. The metrics are described by following equations:

$$\boldsymbol{Precision} = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad \text{(eq. 2)}$$

$$\boldsymbol{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad \text{(eq. 3)}$$

$$\boldsymbol{F1-Score} = 2x\ \frac{Precision*Recall}{Precision+Recall} \qquad \text{(eq. 4)}$$

Precision and Recall are metrics, that can be used in binary classifications and are particularly used for unequally distributed classes. Thus, it is well suited for the comparison of the model's ability in extracting building in areas with different building morphologies. Especially, the training sites 'low socioeconomic status, dense' (LSSd) and 'low socioeconomic status, loose' (LSSl), which have very different distributions of the positive and the negative class. To enable the same conditions for the backbones, the same hyperparameters as the baseline have been used. These are mainly the same data augmentation, leaning rate (0.0001), batch size of 4, class weights (buildings: 2, background: 0.5) and a training time of 40 epochs.

3.2.3 Postprocessing and further processing steps for statistical analysis

After the final prediction has been made, the image tiles need to be stitched together. In order to receive a coherent scene, a majority vote has been applied for each pixel in the overlapping areas. These are up to 9 predictions from all adjacent tiles. The accuracy assessment in chapter 4 has been made after applying the majority vote. To enable a more precise estimation of affected population, false positive misclassifications have been deleted manually prior to the exposure analysis.

Besides disaggregating and analyzing the population data based on the prediction from the DNN on the one hand and based on the cadaster data on the other hand to make a comparison of both results afterwards, newly built houses will be estimated as well. This is also a very valuable information, since an evaluation of newly built houses can reveal, in which zone of criticality construction projects have been realized and if they can be found in economically wealthier or poorer sites of the investigation area, according to the ground value dataset provided by the municipality.

Filtering out newly built-up areas is possible, since the cadaster data is older and accordingly less complete than the building footprints from the orthophoto. However,

since the footprints have no clear timestamp and no indication about the completeness, the following processing steps essentially show the methodological potential to identify newly built-up areas with remote sensing techniques. Several processing steps have been made to gain such information: First, the building footprints from the prediction have been polygonised. The basic idea is, that every building footprint which cannot be selected by the cadaster is supposed to be newly built-up area. However, if the prediction is simply polygonised, whole building blocks would consist of only a single polygon. In that manner a small building in the cadaster can select a complete building block, which might not be recorded in the cadaster. To avoid that, each polygon has been limited to a maximum of 15 vertices during the polygonization. After the selection and the following removal of building shapes, which already occurred in the cadaster, some small polygons belonging to preexisting houses remained. That is because the building footprints of the prediction are frequently bigger than the ones of the cadaster and therefore outer parts of those buildings cannot be selected and remain in the layer. To remove these residual small detached polygons, the multi-polygons have been merged to single features and all polygons with a smaller area than 20 square meters have been deleted.

The threshold for the maximum vertices used during the polygonization and the removal of polygons under 20 square meters have been found to be best fitting in order to remove the noisy data patches, but still keep all residential buildings. The threshold is based on thorough visual inspection of the prediction, cadaster and the orthophoto.

3.3 Data Analysis: The exposure analysis

The last and final part of the methodological workflow is the exposure analysis. In this approach, a basic population disaggregation method has been applied. After the generation of building footprints of the DL-based building extraction, the population data for each neighborhood can be disaggregated to the built-up area. To be technically more specific, the built-up area resulting from the DL model can

be termed as the disaggregation weight (Šimbera 2020). In that case Šimbera (2020) calls it ‘areal weighting’, which is the baseline of disaggregation weights. Of course, this also implies that there are more sophisticated methods or more sophisticated disaggregation weights, but these methods typically require more auxiliary data like a land-use map or a building height map to make more precise assumptions about the spatial distribution of the population. Acquiring such data, converting it to the right format and implementing it to the workflow would be out of the scope and the focus of the thesis. Consequently, the method applied in this master thesis assumes an equally distributed population over the built-up area.

As the building footprints are generated by semantic segmentation in a densely built area, resulting in a partly coherent binary mask, the estimation of landslide-prone inhabitants refers rather to the building block level than to the single building itself.

The following steps for generating the final statistics are carried out on the mentioned DL-based building footprints and on those from the cadaster accordingly. Therefore, the analysis will not only estimate the inhabitant’s exposure to landslide, but also the relative statistical difference between both datasets. This will clarify the importance of an up-to-date data basis in a very dynamic urban environment.

The main step for calculating the exposure of the residents at hazardous sites for each neighborhood in the investigation area is described by following equation:

$$P_r = \frac{Ba_r}{Ba_t} * P_t \qquad \text{(eq. 5)}$$

Where P_r and Ba_r are the population and the built-up area at risk, respectively, and P_t and Ba_t are the population and the built-up area in total for each barrio, respectively. The landslide-prone areas are represented by the medium and high-risk areas of the Mass Movement Threat Map (see 2.2.2).

Based on that calculation to yield a total number of affected populations, more specific statistics are made to get a more detailed impression of the local situation. For example, special attention is paid to informal settlements to find out to what extent they belong to the buildings in medium and high threat areas, as the population living there can be considered the most vulnerable population due to their socio-economic situation and the substandard building fabric. (cf. Echeverri & Cadena-Gaitán 2019, Vega & Hidalgo 2017, Vega et al. 2017).

Further, it will be investigated how many newly built areas can be identified and to what extent they have been built in designated hazardous sites. With the help of the ground value dataset it will be analyzed if those new residential sites are mainly located in poorer or also in wealthier locations, under the assumption that locations with low ground value are an indication of population with higher vulnerability. To support this assumption, the range of ground values of informal settlements, among other categories of settlements, are investigated.

4 Results & discussion

4.1 Experimental approach of deep learning frameworks: evaluation

Table 1 gives an overview of the metrics for both backbones, the Resnet34 and EfficientnetB2. To firstly address the spatial resolution of the orthophoto and accordingly the FOV issue, one can clearly see the high differences of the precision and recall for 8 cm concerning both backbones. Additionally, the recall is mostly relatively low. By looking at the 32 cm resolution, the opposite becomes apparent, while the 16 cm results are relatively balanced. Figure 16 shows a representative example of the FOV in all three resolutions. It shows that the 8 cm tile completely covers parts of roofs and has no real textural context. As a result, the classifier shows no big issues for correctly classifying building structures. But in turn, the prediction there is not very consistent, which leads to a low recall with many false negatives.

The prediction on the 32 cm resolution image often lacks precision, which becomes visible by relatively high false positive predictions (tab. 1). It can be assumed that the low precision results from too little training data, which reduces by factor 16 when reducing the spatial resolution from 8 cm to 32 cm. However, increasing the size of training areas would not only drastically increase manual annotation, but it would hinder the idea of separately training different building morphologies, as well.

Table 1: *Performance comparison of both tested backbones. LSSd = Low Socioeconomic Status, dense; LSSl = Low Socioeconomic Status, loose; C&I = Commerce and Industry; MSS = Medium Socioeconomic Status.*

		Resnet34		
		8 cm	16 cm	32 cm
LSSd	Precision:	0,89	0,82	0,8
	Recall:	0,58	0,98	0,94
	F1:	0,7	0,89	0,86
LSSl	Precision:	0,99	0,51	0,51
	Recall:	0,71	0,98	0,97
	F1:	0,83	0,67	0,67
C&I	Precision:	0,88	0,87	0,79
	Recall:	0,77	0,96	0,96
	F1:	0,82	0,92	0,87
MSS	Precision:	0,97	0,81	0,94
	Recall:	0,56	0,87	0,93
	F1:	0,71	0,89	0,94
Average metrics	Precision:	0,93	0,75	0,76
	Recall:	0,66	0,95	0,95
	F1:	0,77	0,84	0,84

		EfficientnetB2		
		8 cm	16 cm	32 cm
LSSd	Precision:	0,87	0,85	0,81
	Recall:	0,79	0,96	0,98
	F1:	0,82	0,9	0,89
LSSl	Precision:	0,97	0,66	0,49
	Recall:	0,39	0,95	0,99
	F1:	0,56	0,79	0,66
C&I	Precision:	0,92	0,92	0,68
	Recall:	0,84	0,97	0,96
	F1:	0,88	0,94	0,91
MSS	Precision:	0,98	0,95	0,94
	Recall:	0,67	0,94	0,98
	F1:	0,79	0,95	0,96
Average metrics	Precision:	0,94	0,85	0,73
	Recall:	0,67	0,96	0,98
	F1:	0,76	0,90	0,86

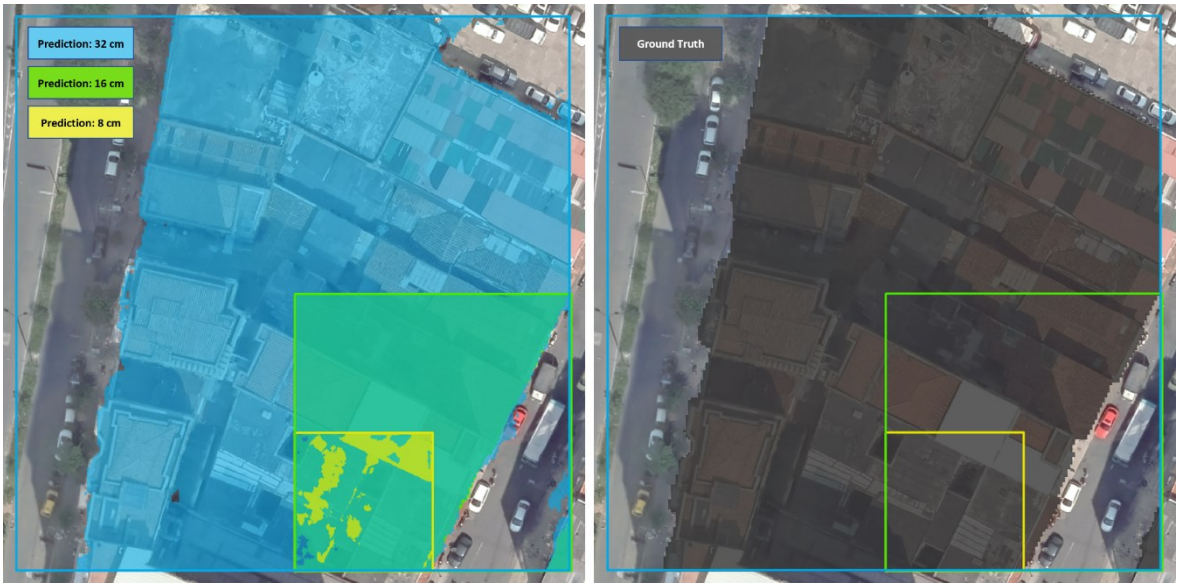


Figure 16: *The scene shows a building block of the test site 'Commerce & Industry'. It demonstrates the FOV issue in capturing global context of the urban built-up landscape.*

It can be assumed, that it won't be possible to increase the training data by e.g. factor 10 or higher and ensure that the samples in the training area consistently belong to one certain morphology at the same time. Additionally, if the operator chooses several training sites for one morphology to train, there won't be much data left to classify.

Table 2: *Performance of the fine-tuned model compared to the baseline setup. Relative differences are given in brackets.*

EfficientnetB2 (16 cm)		baseline	fine-tuned
LSSd	Precision:	0,85	(+0,04) 0,89
	Recall:	0,96	(-0,01) 0,95
	F1:	0,9	(+0,02) 0,92
LSSl	Precision:	0,66	(+0,12) 0,78
	Recall:	0,95	(+0,03) 0,98
	F1:	0,79	(+0,08) 0,87
C&I	Precision:	0,92	(+0,03) 0,95
	Recall:	0,97	(-0,03) 0,94
	F1:	0,94	(+0,01) 0,95
MSS	Precision:	0,95	(-0,01) 0,94
	Recall:	0,94	(-0,04) 0,9
	F1:	0,95	(-0,03) 0,92
Average metrics	Precision:	0,85	(+0,04) 0,89
	Recall:	0,96	(-0,02) 0,94
	F1:	0,90	(+0,02) 0,92

By examining the statistics for the 16 cm resolution image, the prediction seems to be relatively balanced between precision and recall. Only the test site LSSI has a weak precision. It seems plausible, that for this test site and this resolution the training data is too little. This can originate from a much lower amount of buildings within the training site compared to the other sites.

The comparison of the Resnet34 and the EfficientnetB2 for the 16 cm resolution dataset shows a nearly consistent advantage of the EfficientnetB2 over all metrics. This results in an 0.06 higher F1-score on average, which mainly originates from the distinct higher precision of 0.1 on average.

Because the U-Net combined with the EfficientnetB2 as the backbone applied on the 16 cm resolution dataset could be clearly identified as the best setup for the final classification, the next step includes the fine-tuning of the network. This step mainly has been conducted in a trial and error manner to find the best combination of batch size, learning rate, class weights and also the parameters of the data augmentation. Compared to the baseline setup described above, the change of the class weights to 1.5 (buildings) and 0.5 (background), a slightly enhanced data augmentation and the combination of the cross-entropy and Jaccard loss function could increase the model performance. The change of the loss function setup is inspired by Lu et al.

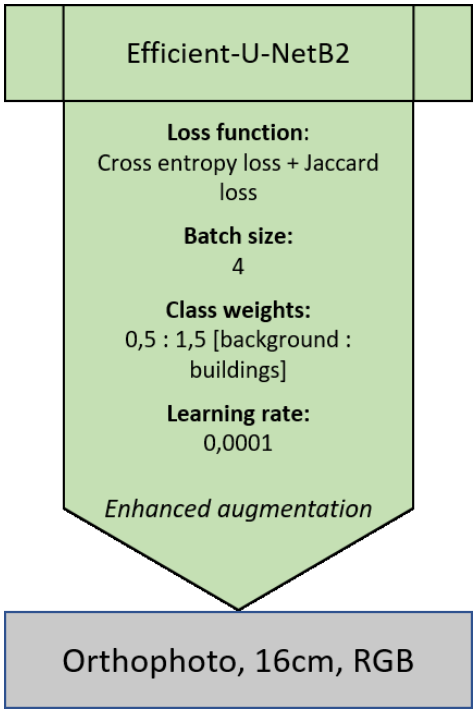


Figure 17: *Final classification setup.*

(2018), who are addressing the issue of effectively training a DL building extraction model with high intra-class variability, which is also true for the training data in this thesis. According to the authors the combination with Jaccard loss helps the network to better balance foreground and background during the training.

In the average metrics the fine-tuned model demonstrates an overall improvement. While showing a slight decrease in the recall by 0.02, the precision increases by 0.04 resulting in an F1-score improvement by 0.02 to 0.92. Except for the MSS test site, the predominant change in the performance is characterized by a distinct increase in precision and a better balance between recall and precision.

The final DL setup, as it is applied on the whole investigation area, is summarized in figure 17.

4.2 Deep learning-based building extraction: quality analysis

Following the examination of different DL approaches in section 4.1, the building extraction of the whole investigation area is conducted using the best performing DL setup presented in figure 17. After the prediction of the ~700,000 image tiles, they have been combined to a coherent scene with a majority vote for each pixel in the overlapping areas. Further, the prediction is manually cleared from misclassifications. This manual correction, however, only corrects for false positive standalone polygons, because they are relatively easy to find and fast to correct. Considering the size of the investigation area it would not be feasible to do more manual correction, which considers either inconsistent and too small building footprints or those, which are too large incorporating surrounding background. Nonetheless, regarding a total area of 893,427 m² (6.3% of the raw classification) in cleared false positive misclassifications, a significant bias in the exposure estimation could be eliminated in that manner.

As indicated above, there is no ground truth data available for the prediction of the defined investigation area. Therefore, isolated false positive objects will be categorized semantically as the most frequent misclassifications, allowing the inference of objects in the VHR image, which are not sufficiently trained as

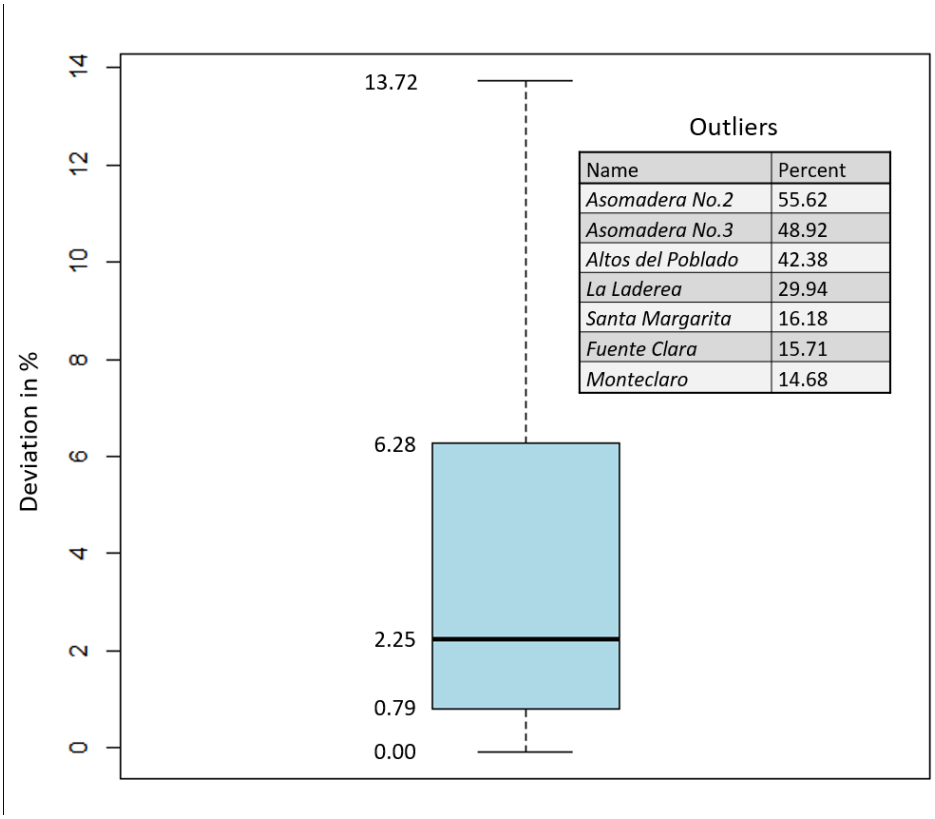


Figure 18: *The percentage change of building area distribution between the cleared and the raw prediction. The values incorporated are barrio-wise.*

background to allow for a more accurate differentiation of buildings and background (see fig. 20).

To get a quantitative overview of the quality of the classification, the boxplot in figure 18 shows the deviation from the manually cleared prediction to the raw prediction in percent for all 87 neighborhoods of interest.

The distribution shows, that the neighborhoods predominately have low deviations from the raw classification with a median at 2.25%. At the same time, it is clearly visible, that there are some neighborhoods achieving very high deviations. The visual inspection of the concerned neighborhoods reveals, that their area is mostly comparatively small and they have a low number of buildings. That means, that the share of potential true positives is relatively little compared to the background inhibiting a potential for misclassifications. With figure 19 as a representative example of these neighborhoods, it becomes apparent that forest and paved roads are a typical source for false positives. This can be explained by the fact, that this kind of urban morphology has not been trained and they are distinctively different from the trained morphologies. The loose arrangements, organically shaped and

Figure 19: *Typical misclassifications (red) in wealthier areas of the city. The blue signature indicates building shapes of the DL network which are not corrected.*

partly wide paved roads and comparatively much vegetation, especially little forests, are all unique features, which would have been needed to be trained to avoid misclassifications. However, it must be considered that learning a new and distinctively different morphology, can lead to less accurate predictions in the other morphologies, which are predominantly informal settlements and building landscapes with very little background.

Figure 20 summarizes the most prominent objects of misclassifications. One

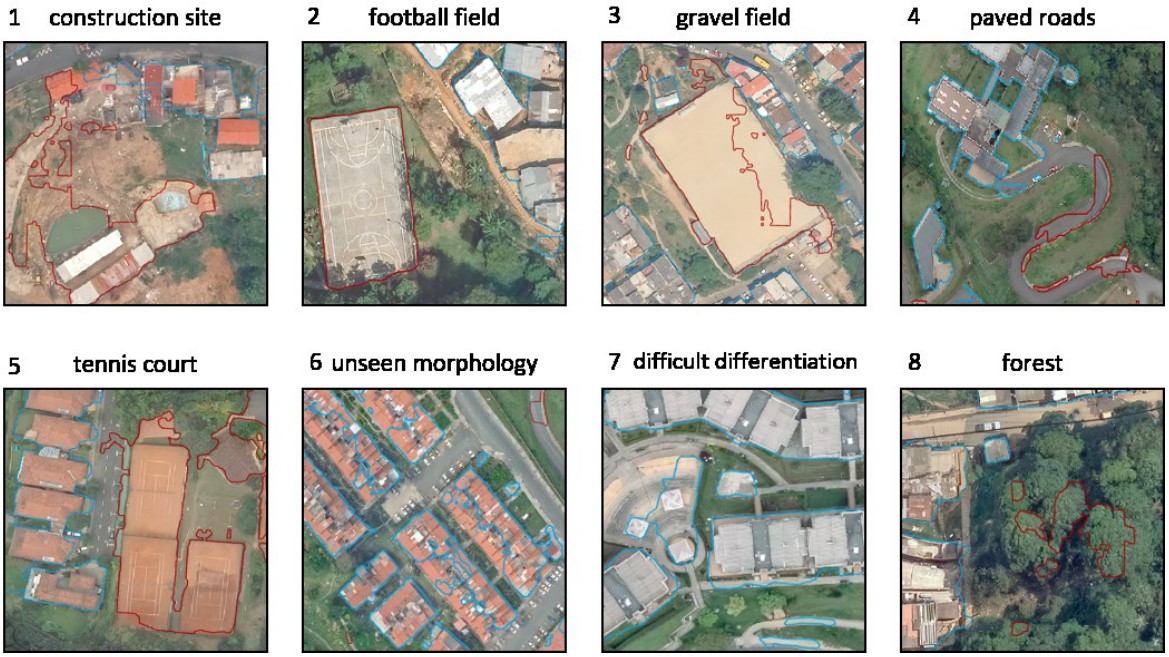


Figure 20: *Prominent misclassified objects divided into eight categories.*

general important observation, which has been made throughout the whole

prediction and which is represented by 2,3 and 5, is the seemingly high confidence of some false positives. High confidence in that regard is indicated by the sharply derived footprints. The clearest outlines of those misclassifications are achieved at football fields. It is likely, that these kinds of errors occur, because their texture and shape is very similar to that of certain rooftop landscapes or building morphologies. As one visual example, this similarity can be observed with the tennis court, where the buildings to the left are quite similar in shape and texture. Image 7 of figure 20 also shows an example, where it is difficult for the classifier to differentiate between building and background. In that case, there is an inner courtyard between densely arranged buildings, which has a very similar color compared to the surrounding buildings. Also, the distinct edges towards the grass is a typical feature of buildings.

Misclassifications with less confidence and shapes untypical for buildings are those of forests, construction sites and paved roads. It is likely that these kinds of false positives are better to eliminate by extended training of these objects as background, compared to the false positives examined before. The last category of inaccuracies is termed as unseen morphology, which is depicted in figure 20/6, as well as in figure 19. As mentioned above, these inaccuracies are to laborious to manually correct and are therefore not captured statistically. However, they mostly appear in wealthier neighborhoods (according to Quinchía (2012) and fig. 5), accounting for a minor part of hazardous areas. Based on the observation that these kinds of misclassifications often analogously occur with those captured statistically in figure 18, the built-up area prone to landslide activity with medium and high hazard has been evaluated for the extreme values and the upper whisker in the boxplot (equates to over 6.28% deviation from the raw prediction). Neighborhoods, where the building footprints fit accurately with the orthophoto have been dropped from this selection. Neighborhoods with visibly unprecise building footprints (like those in fig.

Table 3: Relative share of inaccurately classified neighborhoods.

less accurate neighborhoods	share [%]	absolute [m²]	total research area [m²]
total area	22.8	3,023,133	13,256,378
hazardous areas	8.7	316,549	3,636,565
threat level: very low	74.8	60,202	80,521
threat level: low	27.7	2,646,381	9,539,292
threat level: medium	9.7	234,948	2,419,887
threat level: high	6.7	81,601	1,216,677

19 or fig. 20 image 6 and 7), which haven't been selected due to their low false positive single features, have been added to the selection.

The share of the neighborhoods with less accurate predictions add up to 22.8% for the total area, whereas 8.7% lie in hazardous areas (medium and high threat areas). As 8.7% do not seem inconsiderable, it must be emphasized that this statistic only shows the amount of area with more errors and therefore, have a higher expected bias in exposed population, but these concerned barrios inhibit not completely erroneous information. Furthermore, it can be seen that most of the areas with a very low threat level lie in these less accurate areas, which are therefore not 'harmful' for the exposure analysis.

When comparing the absolute numbers, it becomes clear, that areas with very low threat in general are relatively rare in the investigation area. Areas with the highest threat level are only represented with 6.7%. In general, the bias from the neighborhoods with less accurate predictions can be assessed as relatively low for the relevant medium and high threat areas.

Since table 3 shows the percentage share of hazardous areas in the less accurate neighborhoods, the rest of the share is represented by the neighborhoods, which can be identified as accurately predicted. Thus, 91.3% of the medium and high threat levels lie in neighborhoods, that have accurate building footprints and that feature very little errors. The following figure 21 visualizes some representative examples of these sites. At the same time those examples also represent the morphologies and arrangements of buildings, which predominantly occur in the investigation area. The roof types do not differ much throughout the examples, whereas the arrangements show some particular differences. Figure 21D shows the kind of arrangement, which has been defined above as LSSI. The irregular housing structures with some vegetation and open soil in its background is mostly found at the margin of the city's municipal boundaries. Figure 21C can also be assigned to one homogenous morphology, that is MSS. The houses are densely arranged, but organized in blocks, surrounded by chess-board like paved roads. This is a clear

Figure 21: *Examples of representative building morphologies which are accurately classified.*

indication of formal settlements. Figure 21A and 21B are examples of sites with mixed morphology. Figure 21B is a mixture of the MSS and the LSSd morphology, because the building arrangements to the right side of the image is more irregular

and denser, whereas the buildings towards the left side are more regularly arranged and a little bit less dense. The same applies for figure 21A, except for the difference that there is also a building block in the middle of the image, which corresponds to the C&I morphology.

All of these examples have little to no misclassifications, which represents the predictive ability of the U-Net for the majority of the investigation area.

4.3 Exposure analysis

The statistics created in the following subchapter, 4.3.1 and 4.3.2, are split into two parts and each of them consists of two categories. One part compares the statistics of the municipal with the rural area and the other part compares those of the formal with the informal settlements. Each of these statistics will be calculated and visualized for the classified and for the cadaster-based building footprints. The last two subchapters, 4.3.2 and 4.3.4, will only utilize the DL-based statistics.

Figure 22 gives an overview over the location of both settlement types and the delineation of the municipal and rural area.

The location of the settlement types shows a distinct pattern. Especially the eastern and north-eastern slopes consist nearly completely of informal settlements. Towards the inner border of the investigation area one can see some formal settlements. Those typically continue from there towards the center of the city. Contrary to that, the south east consists mainly of formal settlements. This part of the city is mainly wealthy and its appearance corresponds mainly to that shown exemplary in figure 19. The wealthy neighborhood El Poblado, mentioned above in the context of gated communities, is also part of this area.

The pattern of the informal and formal settlements on the western slopes can be described as patchy and rather mixed.

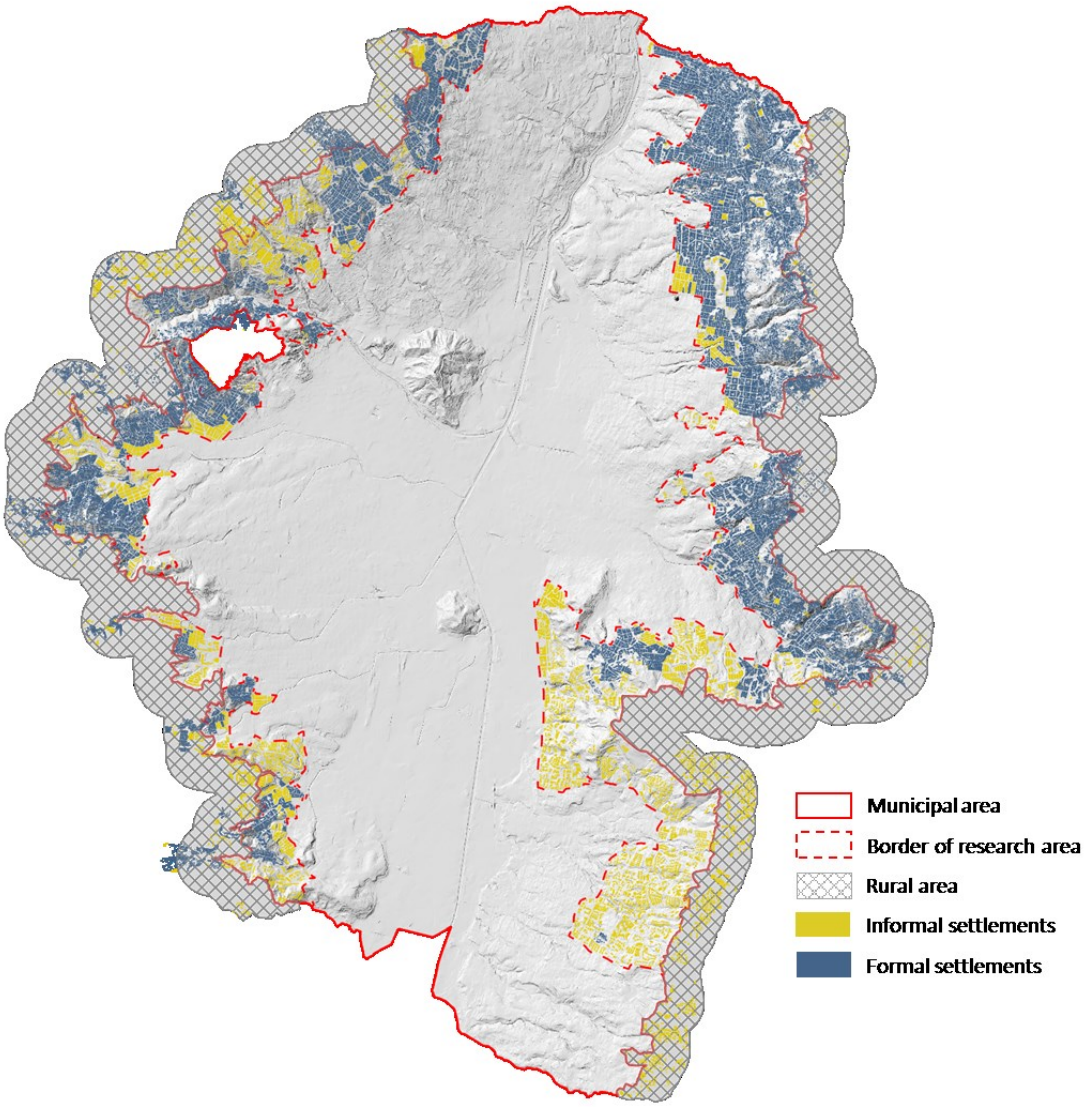
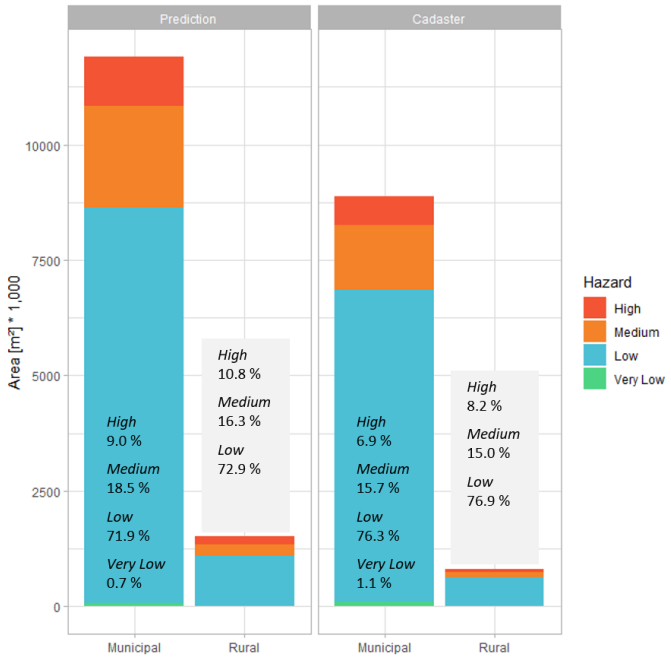


Figure 22: *Map showing the separation into informal and formal settlements as well as the separation into the location of the municipal area (limited by the investigation area) and the rural area. The building footprints correspond to those of the classification.*

4.3.1 Share of hazard classes

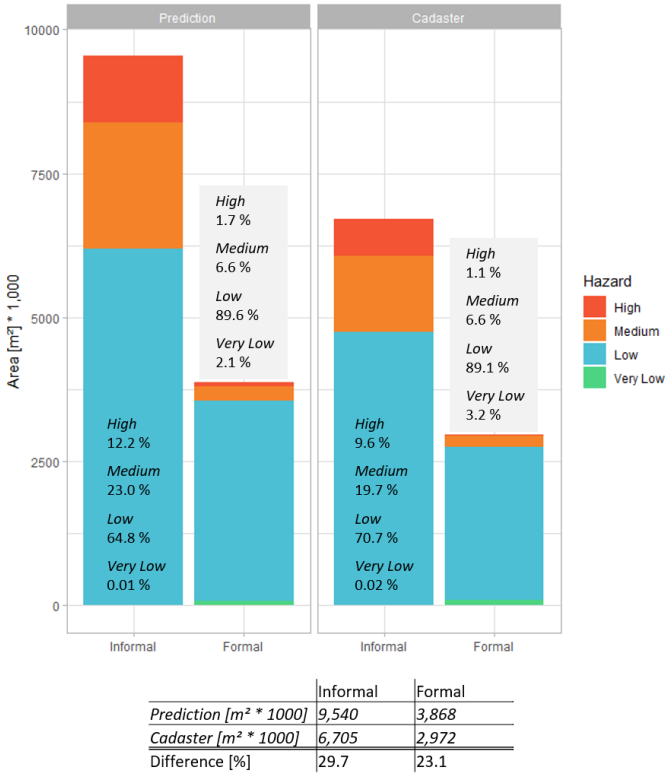


	Municipal	Rural
*Prediction [m² * 1000]*	*11,898*	*1,510*
*Cadaster [m² * 1000]*	*8,871*	*806*
Difference [%]	25.4	46.6

Figure 23: *Municipal vs. rural: The share of hazard class compared between the built-up area of the prediction and the cadaster.*

Figure 23 comprises different statistics to identify the relation between the built-up area within the borders of the municipality and beyond them, in the hereby called rural area. The first clear observation in this bar plot is, that the 'very low' hazard class is nearly not represented. Only a share of 0.7% (prediction) of all buildings within the municipality are located in those areas. This fact supports the intention of delineating the investigation area by the criterion, that the neighborhoods include at least a certain landslide hazard for the majority of their area.

By comparing the respective share of hazard classes of the prediction and the cadaster, it becomes apparent that the prediction covers more built-up area associated with medium and high landslide hazard. Thus, the relative shares of the very low and low hazard class of the cadaster-based building footprints are lower. The same applies for the rural area, besides the absence of the class 'very low'. In figure 24, where the share of hazard classes is calculated for informal and formal settlements, the same pattern is true for the informal settlements, but not for the formal settlements. Moreover, the observation, that the medium and high hazard areas increase towards the border of the city (fig. 6) and the accuracy of the cadastral building footprints decrease in many areas (example in fig. 2) supports the picture which is drawn by these statistics.



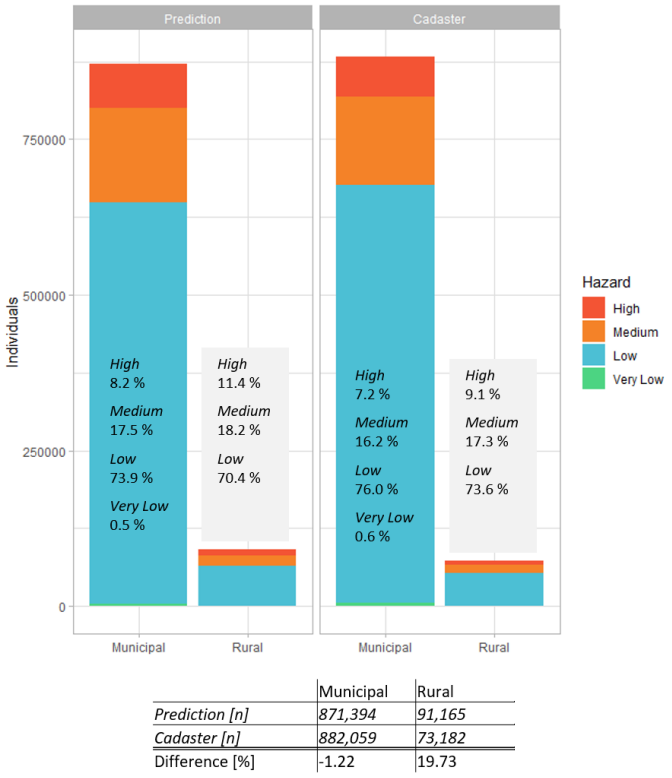
	Informal	Formal
*Prediction [m² * 1000]*	*9,540*	*3,868*
*Cadaster [m² * 1000]*	*6,705*	*2,972*
Difference [%]	29.7	23.1

Figure 24: *Informal vs. formal: The share of hazard class compared between the built-up area of the prediction and the cadaster.*

As a consequence, it can be stated that the lack of information about landslide prone built-up area in the cadaster concerns both, the rural and the municipal area, but not formally raised buildings. However, this finding must be confirmed in the modelled population numbers in the next subchapter (4.3.3), because the bias of the systematically bigger building footprints in the prediction is then considered. The consideration of the bias is realized by using the cadaster-based built-up area on the one hand and that of the prediction on the other hand, in order to calculate and make use of the respective population density per neighborhood in the population disaggregation.

The overall area of the building footprints of both used data sources can be found beneath each of the bar plots. The initial observation, that in general the cadaster-based building footprints are smaller than the actual buildings in the orthophoto, can be confirmed, since the total built-up area of the prediction is always significantly higher. The overall difference between both datasets also show, that the difference of the rural built-up area is nearly twice as high (46.6%) as the difference for the municipal area (25.4%). This supports the previous described pattern, that the cadaster gets more and more inaccurate and incomplete towards the outer border of the investigation area. As the comparison of the informal and formal settlements in figure 24 shows, this described inaccuracy has a bigger impact on the informal than on the formal settlements.

4.3.2 Affected population



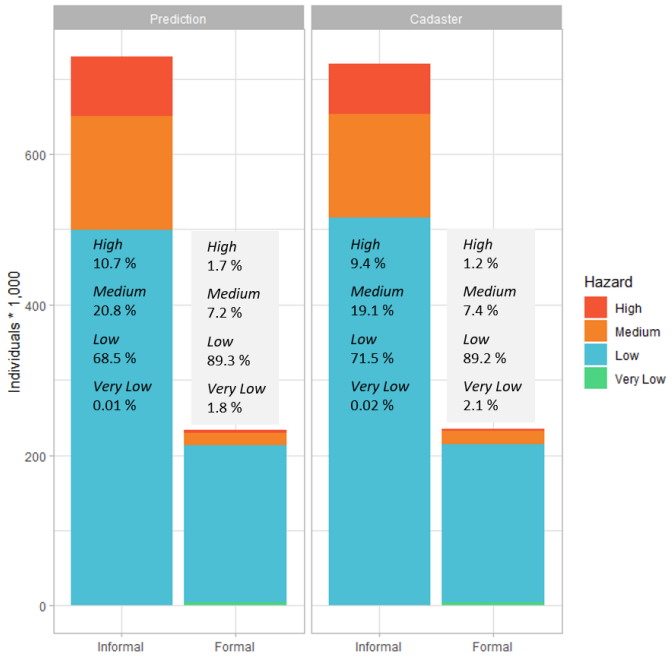
	Municipal	Rural
Prediction [n]	*871,394*	*91,165*
Cadaster [n]	*882,059*	*73,182*
Difference [%]	-1.22	19.73

Figure 25: *Municipal vs. rural: The share of hazard class compared between the modelled population of the prediction and the cadaster.*

While the analysis of the extracted building footprints and those of the cadaster in the subchapter before was focusing on the built-up area, it will now be analyzed how these found patterns will affect the modelled number of inhabitants.

The previous observation that the overall difference of the built-up area for the formal areas are lowest and that the cadaster visually seems to be more or less complete can now be confirmed, as the overall deviation of the modelled population numbers is lowest for the formal settlements as well (fig. 26). Furthermore, the difference in relative numbers of each hazard class are quite similar, as well. Nonetheless, the highest difference here are the 0.5% for the residents living in highly landslide prone areas, which translates to a difference of 1,107 inhabitants in absolute numbers (see table 4).

A stronger pronunciation of the high and medium threat levels in the prediction-based statistic, is also found in the modelled population data. That concerns, again, the municipal and rural area as well as the informal settlements. Whereas the formal settlements are nearly equal at the sites with medium threat. The higher relative numbers of affected population in areas with high threat range from 0.5% (formal)



	Informal	Formal
Prediction [n]	**729,315**	**233,245**
Cadaster [n]	**720,764**	**234,477**
Difference [%]	1.17	-0.53

Figure 26: *Municipal vs. rural: The share of hazard class compared between the modelled population of the prediction and the cadaster.*

to 2.3% (rural). But in absolute numbers the biggest difference is found for the informal settlements with 10,503 more affected inhabitants than modelled with the cadaster data. In turn, the lowest positive difference can also be found for the formal areas with 1,107 more affected inhabitants.

For the medium threat class, the relative differences range from -0.2% (formal) to 2.0% (rural).

In general, these numbers conclude where the biggest differences in quality or completeness of both datasets are, not only based on the visual impression. It seems that the prediction is more comprehensive, but also the statistics of affected inhabitants suggest that the positive differences are related to an incomplete cadaster. Especially, because the statistics of the formal settlements are quite similar, where the cadaster is particularly very comprehensive. The only bigger uncertainty in this comparison is the lack of classification accuracy in some areas as described in 4.2. That suggests, that it would have been useful to determine different reference areas in the investigation area to identify a certain systematic bias between the cadaster- and the DL-based population disaggregation. However, even without such reference the formal settlements provide a good hint to assure that bigger positive deviations, such as the rural residents who are exposed to a high landslide hazard, indicate incompleteness of the building footprints in the cadaster.

To provide concluding statistics of the overall population exposed to landslide hazard (medium and high hazard), table 5 shows the statistics again for both

Table 4: *Statistical overview over the modelled population data with consideration of the different hazard classes and the settlement categories.*

	Municipal			Rural		
Hazard class	Prediction	Cadaster	Difference	Prediction	Cadaster	Difference
High	*71,744*	*63,852*	*7,892*	*10,350*	*6,634*	*3,716*
Medium	*152,080*	*142,472*	*9,608*	*16,591*	*12,666*	*3,925*
Low	*643,316*	*670,573*	*-27,257*	*64,222*	*53,883*	*10,336*
Very low	*4,252*	*5,164*	*-912*	*0*	*0*	*0*

	Informal			Formal		
Hazard class	Prediction	Cadaster	Difference	Prediction	Cadaster	Difference
High	*78,066*	*67,563*	*10,503*	*4,029*	*2,922*	*1,107*
Medium	*151,830*	*137,874*	*13,956*	*16,841*	*17,262*	*-421*
Low	*499,321*	*515,203*	*-15,882*	*208,218*	*209,253*	*-1,035*
Very low	*97*	*124*	*-24*	*4,155*	*5,040*	*-885*

datasets. The uneven total results of the prediction and the cadaster in the left and right table are owed to rounding errors.

The modelling of population data with DL-footprints is not only more accurate in the overall numbers of affected population, but it is also more accurate in the relative relation between the hazard classes for each of the four categories. This is because an incomplete cadaster has higher population densities for each house, as the population density is specific to the total build up area (see equation 5). If one would assume, that at a certain site in the city the cadaster is complete at a highly hazardous site and incomplete at the adjacent low hazardous site, the modelled numbers would overestimate the population modelled for the highly hazardous part and underestimate the population for the other part. Additionally, the relation between both would be strongly biased as well. Although the relations between the different categories and hazard classes in figure 25 and figure 26 are not very different between the cadaster and the prediction, this bias could dramatically affect the results, when calculating similar statistics for local sites within neighborhoods.

Table 5: *Overall numbers of landslide-prone population (medium and high hazard) for the prediction- and the cadaster-based building footprints.*

	Datasets	
	Prediction	Cadaster
Municipal	*223,824*	*206,324*
Rural	*26,941*	*19,300*
Total	**250,765**	**225,624**

	Datasets	
	Prediction	Cadaster
Informal	*229,896*	*205,437*
Formal	*20,870*	*20,184*
Total	**250,766**	**225,621**

Therefore, a comprehensive building footprint dataset is very important for modelling population numbers for areas which are of interest.

4.3.3 Distribution of ground value

As the Geoservice of Medellín offers the ground value layer of the city as open source data, the dataset will be used to find patterns and relations between the distribution of ground value and landslide hazard in different areas. Therefore, the first comparison follows the same principle as the previous analyses: informal vs. formal settlements and municipal vs. rural areas.

This comparison is provided in figure 27 as boxplots, which show the distribution of the ground value for each category. This plot shows, that specifically the ground value in the rural areas, just outside of the border of the municipality, has very low prices and therefore, can be expected to be populated by low-income population with higher vulnerability. In contrast, the formal settlements have a wide distribution of ground value and incorporate the wealthiest parts of the investigation area. But since the ground value in the municipality includes both, informal as well as formal areas, and the boxplot show just a slightly higher distribution of ground values, it becomes evident that areas with relative high ground value concern only smaller

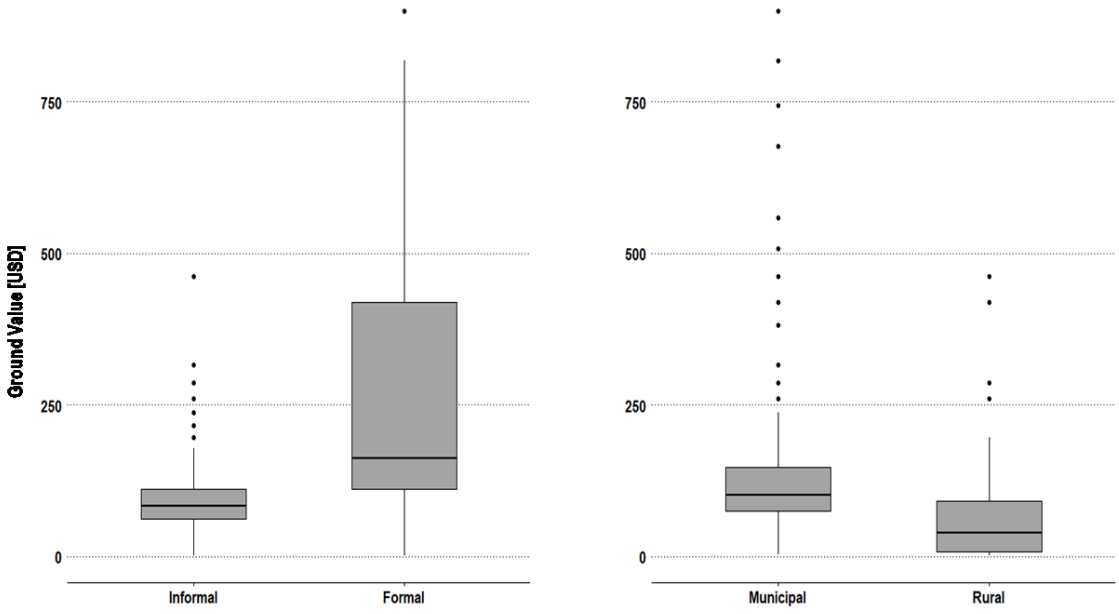


Figure 27: *The ground value distribution for the informal and formal settlements as well as for the municipal and rural areas.*

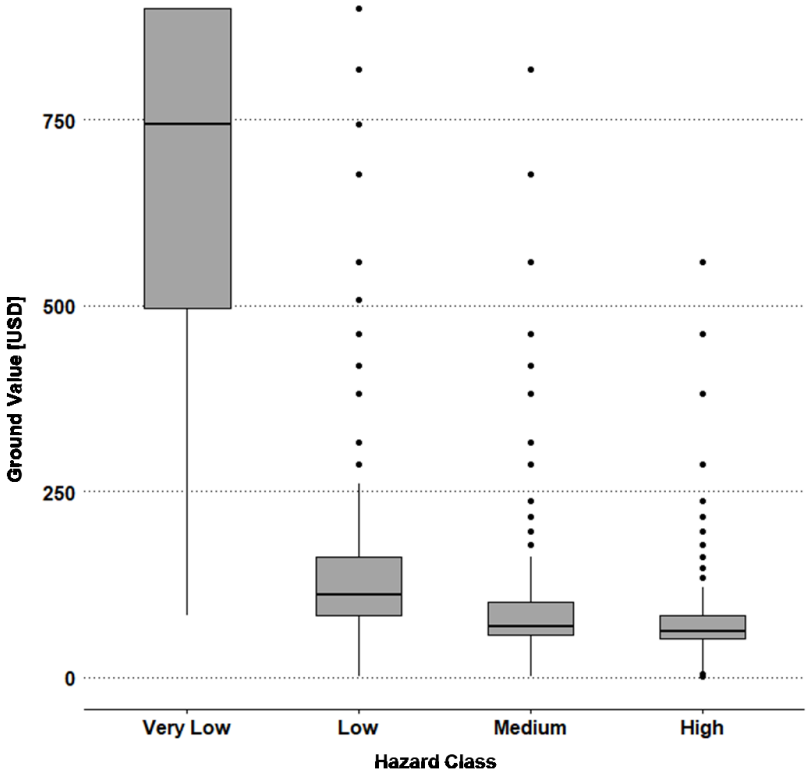
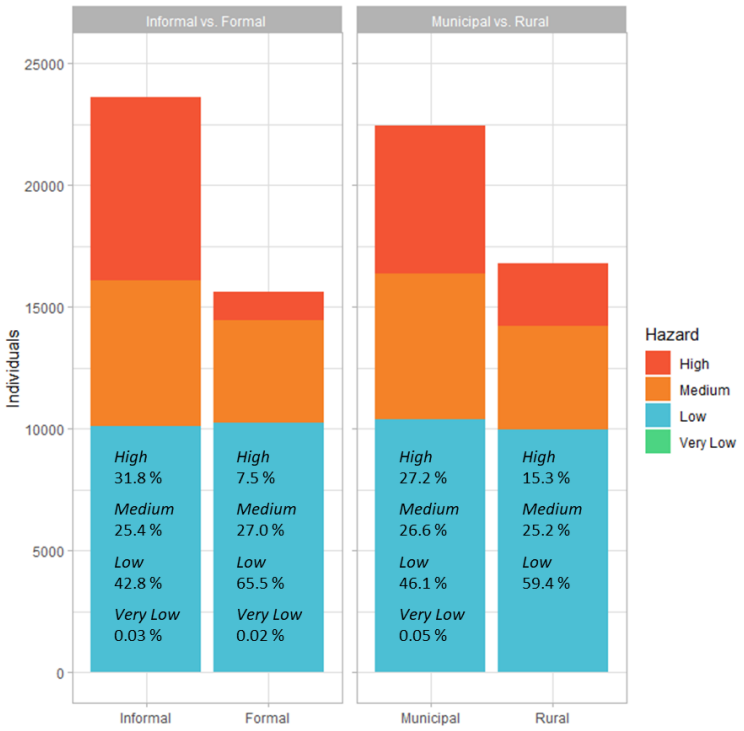


Figure 28: *The ground value distribution for each hazard class.*

areas. Thus, the statistical impact on the distribution of the formal settlements on the municipality is mainly represented in the upper whisker and numerous extreme values.

The boxplots in figure 28 serve the purpose to visualize a relation between the ground value and the exposure to landslide hazard. Initially, it is important to note that the boxplot of the areas with very low hazard is not as representative as the other three hazard classes. This is because of the little representation of areas with very low hazard in this analysis. Nonetheless, this boxplot shows at least an indication, that areas with a very low landslide hazard tend to have the highest ground values in that comparison. With respect to the other classes it can be stated, that the ground value decreases with higher hazard classes.

4.3.4 Newly built houses



	Informal vs. Formal		Municipal vs Rural	
Hazard class	Informal	Formal	Municipal	Rural
High	*7,507*	*1,172*	*6,108*	*2,571*
Medium	*5,987*	*4,221*	*5,978*	*4,229*
Low	*10,095*	*10,237*	*10,354*	*9,978*
Very low	*8*	*3*	*11*	*0*
Total	23,597	15,633	22,451	16,778
Total (High and Medium)	13,494	5,393	12,086	6,800

Figure 29: *Newly built houses: the stacked bar plots show the modelled population for each hazard class and each settlement category. The absolute numbers are given in the table below.*

What the previous chapter showed, is that the affected population modelled is significantly higher when using the DL-derived building footprints instead of the cadaster-based building footprints. This is related to the ability to derive much more comprehensive up-to-date building footprints (with respect to the timestamp of the orthophoto) by remote sensing techniques. To add on to this finding, GIS-based techniques have been applied to yield a layer which contains new building footprints. But before presenting the results it is important to mention, that not all derived shapes are necessarily and strictly speaking -new- built-up areas, compared to the cadasters timestamp. It is also possible, that some buildings or rooftop landscapes are only partly covered by the single features due to spatial inaccuracies. Thus, the other part of the building structure would then be identified as being new, which might not correspond to ground truth.

By examining the modelled population numbers in figure 29, the first apparent finding is the very high overall number for the rural area. With 16,778 new residents it deviates just by 25,3% from the new residents in the municipal area. This gives a

strong indication that the city outgrows its boundaries. But it does not seem, that the present landslide hazard in the city is taken into consideration by the city for new residents. Especially, the informal settlements grew by a considerable 57.2% in areas with medium and high landslide hazard. Even if the growing formal settlements with 35.5% in such areas on the other hand are much lower, it is still a high number and concerns 5,393 residents.

Moreover, when comparing the overall hazard class distribution of the affected population in informal settlements (fig. 26) and the distribution of the population in new built-up areas (fig. 29), it shows these new residents are in medium and high hazard areas settling to a much higher share (57.2% compared to 31.5% overall). Additionally, the overall landslide prone population in the high hazard class are approximately half as high as those in the medium hazard class, whereas the newly exposed population in the high hazard class is by approximately 6% higher than in the medium one. These statistics clearly show, that the population dynamics in recent years still lead to an increase of exposed inhabitants in the city.

The ground value distribution for the municipal and the rural area, as well as for the informal settlements (fig. 30), show a similar distribution to those of the boxplot showing the results for the whole investigation area (fig. 27). Only the ground values of the formal settlements are lower. But at the same time, the pattern for the four categories are quite similar.

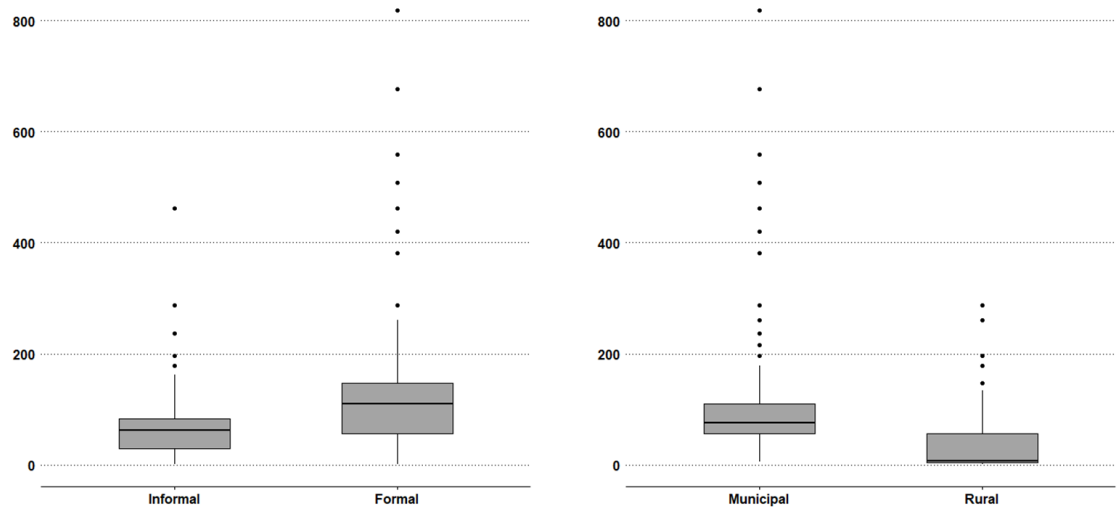


Figure 30: *Newly built houses: the ground value distribution for the informal and formal settlements as well as for the municipal and rural areas.*

The hazard distribution, shown in figure 31, is very similar for each class. Only the low class has a wider distribution of ground values. The Median, however, is nearly on the same level like the other. There is just a little tendency of lower ground values with higher hazard classes. Hence, there is no clear pattern in these distributions, which could lead to a distinctive statistical inference.

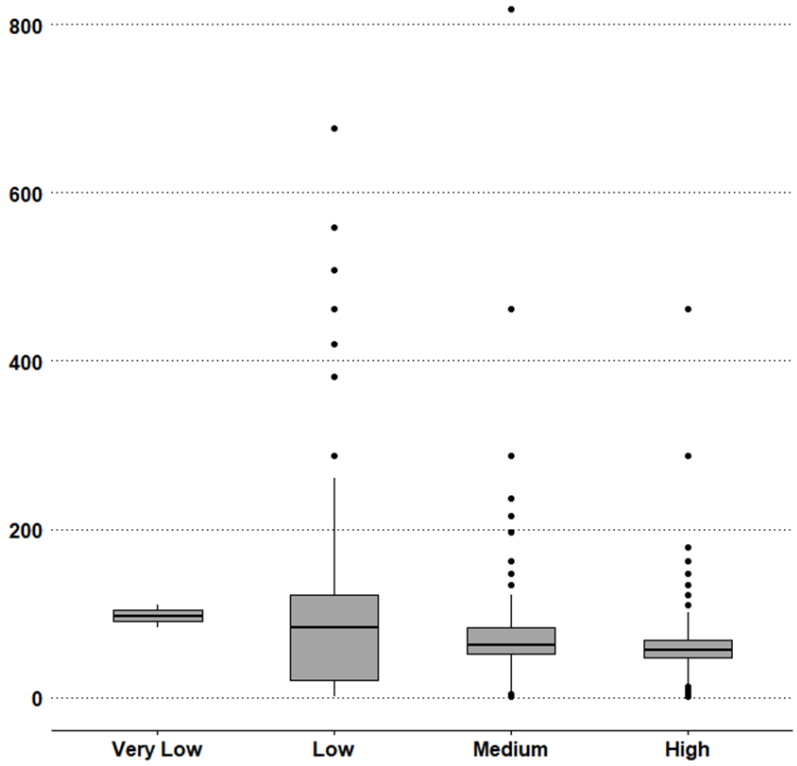


Figure 31: *Newly build houses: the ground value distribution for each hazard class.*

5 Conclusions

In the context of the history of the 20th century, Medellín has been developed to a fast-growing city, which today, is populated by around four million inhabitants. Because the city lies in the Aburrá valley and extends towards steep mountain slopes to the west and the east, the predominantly poor immigrants needed to settle on these slopes to find space for their new homes. These homes are raised informally with basic material. However, due to the geomorphology and the climate, the slopes are heavily influenced by high landslide susceptibility, which leads to a high exposure of the population. However, there are no studies trying to model this exposure until today.

In this thesis, a framework has been developed to most accurately analyze the exposure of the inhabitants of Medellín. With a refined deep learning U-Net model, building footprints have been extracted from a VHR orthophoto, enabling a precise population disaggregation to a local building block related level, in order to intersect this data with the Amenaza hazard map. Since the final results in the exposure analysis rely on the accuracy of the building footprints, the methodological approach and accuracy has been evaluated thoroughly. The applied DL model showed very accurate results for the majority of the investigation area. Clear inaccuracies where found in the south east, where the building morphology with its background is distinctively different from the rest of the investigation area. Although, these inaccuracies had not a critical implication on the exposure analysis, it made clear that the DL model strongly relies on the trained morphologies. This might pose new challenges to the deep learning community and raises the question how much different building morphologies and textures can explicitly be trained in one model, while assuring high classification accuracies for each morphology. Because as far as the literature in this thesis has been reviewed, building extractions are predominantly applied on local parts of cities, meaning that they mostly comprise not many different building morphologies having a similar background. Hence, it is not stressed as a main issue how many different spectral information belonging to only two classes can be learned with a deep learning network and when it might be meaningful to train certain parts of a city in separate models. Only Lu et al. (2018) could be found as a source, that addressed that issue of a high intra-class variability

and a lack of global context when extracting buildings from a VHR scene. They changed the architecture of the U-Net and introduced a procedure, where the U-Net learns two different resolutions separately. The feature maps are then merged and reweighted after the up-sampling process. However, it is difficult to rebuild and implement such an architecture without in-depth expert knowledge or detailed description about the steps to be conducted.

The statistics in the exposure analysis could reveal that the cadaster is mostly inaccurate for informal settlements and for the rural area of the investigation area. Furthermore, the percentage of population living in areas with medium and high landslide probability, is considerably higher than in the cadaster-based data. At the same time, the total modeled population and its relative proportion in each hazard class are comparable for both datasets for the formal area. That supports the localization of where the cadaster lacks information. Thus, it becomes clear how important it is to have an up-to-date building footprint dataset for estimating population exposed to landslide hazard. Especially, when the city expands towards areas with medium and high landslide hazard, concerning mainly relatively poor inhabitants.

The approach presented in this thesis is all in all very promising. However, there are still improvements necessary to make the classification more robust in terms of accuracy and evaluation. Therefore, future studies need to examine the inherent challenges of applied studies, and not just compete for the most accurate DL architecture applied on well-known publicly available datasets, which provide sufficient data for training and evaluation. This would help generate more reliable and relevant data to support the need for political measures in areas, where natural hazards threaten the economical and physical integrity of people.

6 References

Acevedo, A. B., Jaramillo, J. D., Yepes, C., Silva, V., Osorio, F. A., & Villar, M. (2017): Evaluation of the seismic risk of the unreinforced masonry building stock in Antioquia, Colombia. Natural Hazards, 86(1), 31-54.

Alcaldía Medellín (2020): Open Data. Alcaldía Medellín. https://geomedellin-m-medellin.opendata.arcgis.com/ (last access: 29.04.2022)

Aristizábal, E., & Gómez, J. (2007): Inventario de emergencias y desastres en el Valle de Aburrá. Originados por fenómenos naturales y antrópicos en el período 1880-2007. Gestión y ambiente, 10(2), 17-30.

Bergado, J. R., Persello, C., & Gevaert, C. (2016): A deep learning approach to the classification of sub-decimetre resolution aerial images. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1516-1519.

Betancur, J. J. (2007): Approaches to the regularization of informal settlements: the case of PRIMED in Medellin, Colombia. Global Urban Development Magazine, 3(1), 1-15.

Brand, P., & Dávila, J. D. (2011): Mobility innovation at the urban margins: Medellín's Metrocables. City, 15(6), 647-661.

Chen, Z., Wang, C., Li, J., Xie, N., Han, Y., & Du, J. (2021): Reconstruction bias U-Net for road extraction from optical remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 2284-2294.

Chen, Y., Briese, C., Karel, W., & Pfeifer, N. (2014): True orthophoto generation using multi-view aerial images. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(3), 67.

Chollet, F. (2018): Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG.

Civico, A. (2012): “We are illegal, but not illegitimate.” Modes of policing in Medellin, Colombia. PoLAR: political and legal anthropology review, 35(1), 77-93.

Claghorn, J., & Werthmann, C. (2015): Shifting ground: Landslide risk mitigation through community-based landscape interventions. Journal of Landscape Architecture, 10(1), 6-15.

Echeverri, A., & Cadena-Gaitán, C. (2019): Strategies for Sustainable Communities: Landscape-Based Landslide Risk Alleviation in Medellín. Technology | Architecture + Design, 3(1), 42-44.

Feng, W., Sui, H., Huang, W., Xu, C., & An, K. (2018): Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model. IEEE Geoscience and Remote Sensing Letters, 16(4), 618-622.

Ferrari, S. G., Smith, H., Coupe, F., & Rivera, H. (2018): City profile: Medellin. Cities, 74, 354-364.

García, M. (2005): El deslizamiento de Villatina, in: Hermelín, M.: Desastres de origen natural en Colombia, 1979-2004. Universidad EAFIT.

Goodfellow, I., Bengio, Y., & Courville, A. (2016): Deep learning. MIT press.

Habib, A. F., Kim, E. M., & Kim, C. J. (2007): New methodologies for true orthophoto generation. Photogrammetric Engineering & Remote Sensing, 73(1), 25-36.

Heinrichs, D., & Bernet, J. S. (2014): Public transport and accessibility in informal settlements: Aerial cable cars in Medellín, Colombia. Transportation research procedia, 4(1), 55-67.

Hermelin, M.: The natural environment, in: Arbaux, M. H., Restrepo, A. E., & Giraldo, J. (2012): Medellín: environment urbanism society. Universidad EAFIT, 33-50.

Höser, T., & Künzer, C. (2020): Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. Remote Sensing, 12(10), 1667.

Humphrey, M., & Valverde, E. (2017): A tale of dual cities: The urban miracle and humanitarian crisis maps of Medellín (2002–2015). Journal of Iberian and Latin American Research, 23(2), 159-177.

Internal Displacement Monitoring Centre (iDMC) (2019): Colombia. <http://www.internal-displacement.org/countries/colombia?page=1>. Accessed: 15.11.2019.

Isaza-Restrepo, P. A., Carvajal, H. E. M., & Montoya, C. A. H. (2016): Methodology for quantitative landslide risk analysis in residential projects. Habitat International, 53, 403-412.

Ji, S., Wei, S., & Lu, M. (2019): A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. International journal of remote sensing, 40(9), 3308-3322.

Klimeš, J., & Rios Escobar, V. (2010): A landslide susceptibility assessment in urban areas based on existing data: an example from the Iguaná Valley, Medellín City, Colombia. Natural Hazards and Earth System Sciences, 10(10), 2067-2079.

LeCun, Y., Bengio, Y., & Hinton, G. (2015): Deep learning. Nature, 521(7553), 436-444.

Li, J., Huang, X., & Gong, J. (2019): Deep neural network for remote-sensing image interpretation: Status and perspectives. National Science Review, 6(6), 1082-1086.

Long, J., Shelhamer, E., & Darrell, T. (2015): Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 3431-3440.

Lu, K., Sun, Y., & Ong, S. H. (2018): Dual-resolution U-Net: Building extraction from aerial images. 2018 24th International Conference on Pattern Recognition (ICPR), IEEE. 489-494.

Knopp, L., Wieland, M., Rättich, M., & Martinis, S. (2020): A deep learning approach for burned area segmentation with Sentinel-2 data. Remote Sensing, 12(15), 2422.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019): Deep learning in remote sensing applications: A meta-analysis and review. ISPRS journal of photogrammetry and remote sensing, 152, 166-177.

Marston, J. (2019): The urban displaced: fleeing criminal violence in Latin American cities. Global Report on Internal Displacement. Background paper to the main report.

Mboga, N., Persello, C., Bergado, J., & Stein, A. (2017): Detection of informal settlements from VHR images using convolutional neural networks. Remote sensing, 9(11), 1106.

Muñoz, E., Martínez-Carvajal, H., Arévalo, J., & Alvira, D. (2014): Quantification of the effect of precipitation as a triggering factor for landslides on the surroundings of Medellín-Colombia. Dyna, 81(187), 115-121.

Ojeda, J., & Donnelly, L. (2006): Landslides in Colombia and their Impact on Towns and Cities". IAEG, 112, 1-13.

O'Shea, C. (2014). Shifting ground: precarious settlements and geological hazard in Medellin, Colombia. Landscape Architecture Frontiers, 2(4), 148-160.

Persello, C., & Stein, A. (2017): Deep fully convolutional networks for the detection of informal settlements in VHR images. IEEE geoscience and remote sensing letters, 14(12), 2325-2329.

Quinchía, J. E. P.: The potential of satellite remote sensing as a tool for urban and environmental planning in the Aburrá valley, in: Arbaux, M. H., Restrepo, A. E., & Giraldo, J. (2012): Medellín: environment urbanism society. Universidad EAFIT, 114-129.

Restrepo, A. E., & Orsini, F. M.: Informality and social urbanism in Medellín, in: Arbaux, M. H., Restrepo, A. E., & Giraldo, J. (2012): Medellín: environment urbanism society. Universidad EAFIT, 132-156.

Ronneberger, O., Fischer, P., & Brox, T. (2015): U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. 234-241.

Sanín Naranjo, P. (2010): ¿ De ciudad abierta a ciudad cerrada? Configuraciones socio-espaciales en el barrio El Poblado, Medellín. Territorios, (23), 123-142.

Secretaria de Tecnologia (municipio de Medellín) (2020): GeoMedellín. Open data service. Alcaldía Medellín. https://geomedellin-m-medellin.opendata.arcgis.com/ (last access: 29.04.2022)

Šimbera, J. (2020): Neighborhood features in geospatial machine learning: the case of population disaggregation. Cartography and Geographic Information Science, 47(1), 79-94.

Smith, H., Coupé, F., Garcia-Ferrari, S., Rivera, H., & Castro Mera, W. E. (2020): Toward negotiated mitigation of landslide risks in informal settlements: reflections from a pilot experience in Medellín, Colombia. Ecology and Society, 25(1).

Stiller, D., Stark, T., Wurm, M., Dech, S., & Taubenböck, H. (2019): Large-scale building extraction in very high-resolution aerial imagery using Mask R-CNN. 2019 Joint Urban Remote Sensing vent (JURSE), IEEE, 1-4.

Sun, G., Huang, H., Zhang, A., Li, F., Zhao, H., & Fu, H. (2019): Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images. Remote Sensing, 11(3), 227.

United Nations, Department of Economic and Social Affairs, Population Division (UN DESA) (2018): World Urbanization Prospects 2018: The 2018 Revision, Online Edition.

Vega, J. A., Hidalgo, C. A., & Marín, N. J. (2017): Landslide Risk: Economic Valuation in The North-Eastern Zone of Medellin City. IOP Conference Series: Materials Science and Engineering, 245(6), 062010.

Vega, J. A., & Hidalgo, C. A. (2017): Risk Assessment of Earthquake-Induced Landslides in Urban Zones. Workshop on World Landslide Forum, 953-963.

Vega, J. A., & Hidalgo, C. A. (2016): Quantitative risk assessment of landslides triggered by earthquakes and rainfall based on direct costs of urban buildings. Geomorphology, 273, 217-235.

Wurm, M., Droin, A., Stark, T., Geiß, C., Sulzer, W., & Taubenböck, H. (2021): Deep learning-based generation of building stock data from remote sensing for urban heat demand modeling. ISPRS International Journal of Geo-Information, 10(1), 23.

Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019): Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS journal of photogrammetry and remote sensing, 150, 59-69.

Xu, Y., Wu, L., Xie, Z., & Chen, Z. (2018): Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. Remote Sensing, 10(1), 144.

Yi, Y., Zhang, Z., Zhang, W., Zhang, C., Li, W., & Zhao, T. (2019): Semantic Segmentation of urban buildings from vhr remote sensing imagery using a deep convolutional neural network. Remote Sensing, 11(15), 1774.

Zhang, L., Zhang, L., & Du, B. (2016): Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine, 4(2), 22-40.

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017): Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8-36.