OULU 2022
C 818

ACTA
UNIVERSITATIS OULUENSIS

Hamid Darabi

MACHINE LEARNING TECHNIQUES FOR URBAN FLOOD RISK ASSESSMENT

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF TECHNOLOGY

ACTA UNIVERSITATIS OULUENSIS
C Technica 818

HAMID DARABI

MACHINE LEARNING TECHNIQUES FOR URBAN FLOOD RISK ASSESSMENT

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in Auditorium IT116, Linnanmaa, on 17 February 2022, at 10 a.m.

UNIVERSITY OF OULU, OULU 2022

Acta Univ. Oul. C 818, 2022

Supervised by
Associate Professor Ali Torabi Haghighi
Professor Bjørn Kløve

Reviewed by
Associate Professor Ali Mehran
Assistant Professor Olga Makarieva

Opponent
Associate Professor Saeed Golian

ISBN 978-952-62-3195-2 (Paperback)
ISBN 978-952-62-3196-9 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

PUNAMUSTA
TAMPERE 2022

Darabi, Hamid, Machine learning techniques for urban flood risk assessment.
University of Oulu Graduate School; University of Oulu, Faculty of Technology
Acta Univ. Oul. C 818, 2022
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Floods can cause severe damage in urban environments. In regions lacking hydrological and hydraulic data, spatial urban flood modeling and mapping can enable city authorities to predict the intensity and spatial distribution of floods. These predictions can then be used to develop effective flood prevention and management plans. In this doctoral thesis, flood inventory data for Mazandaran, Iran were prepared based on historical and field survey data from the Sari and Amol municipalities and the regional water company. Flood risk maps were produced using several machine learning (ML) algorithms: GARP, QUEST, RF, j48DT, CART, LMT, ANN-SGW, SVM, MAXENT, BRT, MARS, GLM, GAM, Ensemble, MLPNN, and MultiB-MLPNN models. The flood influencing factors used in modeling were precipitation, slope, curve number, distance to river, distance to channel, depth to groundwater, land use, and elevation. Two equal sets of points were identified randomly for both categories of flooded and non-flooded areas. Therefore, 113 (for Sari city) and 118 (for Amol city) locations for each category were identified. Each set is divided into training (70%) and testing (30%) groups. The flood locations were assigned a value of 1, and non-flood locations were assigned a value of 0. Different conditioning factors, including urban density, quality of buildings, age of buildings, population density, and socio-economic conditions were considered to analyze urban flood vulnerability. Several confusion matrix criteria were applied to evaluate the accuracy of the ML algorithms. The results demonstrated that the ANN-SGW (as the optimized model), GARP (as the standalone model), Ensemble (BRT, MARS, GLM, and GAM), and MultiB-MLPNN models (as the hybridized model) had the highest performance accuracy, with area under the curve (AUC) values of 0.963, 0.935, 0.925, and 0.847 respectively. The results also indicated that distance to channel played a major role in flood hazard determination, whereas population density was the most important factor in terms of urban flood vulnerability. These findings demonstrate that machine learning models can support flood risk mapping, especially in areas where detailed hydraulic and hydrological data are not available.

Keywords: confusion matrix, flood inventory data, flood risk mapping, NN-SGW model

Darabi, Hamid, Koneoppimistekniikat kaupunkitulvariskien arvioinnissa.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Teknillinen tiedekunta
Acta Univ. Oul. C 818, 2022
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Tulvat voivat aiheuttaa vakavia vahinkoja kaupunkiympäristössä. Alueilla, joista hydrologisia ja hydraulisia tietoja ei ole kattavasti saatavilla, kaupunkitulvien alueellinen mallinnus ja kartoitus avaavat mahdollisuuden viranomaisille arvioida tulvien alueellista jakautumista ja voimakkuutta. Mallinnus auttaa päätöksentekijöitä kehittämään toimivia tulvien ehkäisy- ja hallintasuunnitelmia. Tässä tutkimuksessa tulvainventointitiedot laadittiin Sarin ja Amolin kuntien sekä Iranin Mazandaranin vesiyhtiön historiallisten ja kenttätutkimusten tietojen perusteella. Tulvariskikarttoja tuotettiin useilla koneoppimisalgoritmeillä: GARP, QUEST, RF, j48DT, CART, LMT, ANN-SGW, SVM, MAXENT, BRT, MARS, GLM, GAM, Ensemble, MLPNN, ja MultiB-MLPNN mallit. Mallinnuksessa käytettyjä tulviin vaikuttavia tekijöitä olivat sadanta, maanpinnan kaltevuus, käyrän numero, etäisyys jokeen, etäisyys kanavaan, etäisyys pohjaveden pintaan, maankäyttö ja maanpinnan korkeus. Kaksi samanlaista pistejoukkoa tunnistettiin satunnaisesti sekä tulvivalla että tulvattomalla alueella ja siksi kullekin luokalle tunnistettiin 113 (Sarin kaupunki) ja 118 (Amolin kaupunki) sijaintia. Jokainen sarja on jaettu koulutusryhmiin (70 %) ja testausryhmiin (30 %). Tulvapaikoille määritettiin arvo 1 ja tulvattomille arvo 0. Kaupunkien tulvahaavoittuvuuden analysoinnissa arvioitiin erilaisia tekijöitä, kuten rakennustiheys, rakennusten laatu, rakennusten ikä, väestötiheys ja sosioekonomiset olosuhteet. ML-algoritmien tarkkuuden arvioimiseksi käytettiin useita sekaannusmatriisikriteerejä. Tulokset osoittivat, että ANN-SGW (optimoitu malli), GARP (erillisenä mallina), yhdistelmä-ensemble (BRT, MARS, GLM ja GAM) ja MultiB-MLPNN-mallit (hybridimallina) tuottivat muita paremman suorituksen tarkkuuden, AUC=0.963, AUC=0.935, AUC=0.925 ja AUC=0.847, edellä mainitussa järjestyksessä. Tulokset osoittivat myös, että etäisyys kanavaan oli tärkeässä asemassa tulvariskien määrittämisessä, kun taas väestötiheys oli haavoittuvuuden kannalta tärkein tekijä. Nämä havainnot osoittavat, että koneoppimismallit voivat auttaa tulvariskikartoituksessa erityisesti alueilla, joilla yksityiskohtaisia hydrauliikka- ja hydrologisia tietoja ei ole saatavilla.

Asiasanat: NN-SGW malli, sekaannusmatriisi, tulvariskien kartoitus, tulvien dataaineistot

**** Mother is the most precious gift ever from God ****
**** To the memory of my mother ****

****I am thankful to my brilliant wife for motivating me throughout my journey of thesis. Thereby, I dedicate my thesis to my lovely wife****

Acknowledgements

Throughout the writing of this dissertation, I have received a great deal of support and assistance. I would first like to thank my supervisors, Dr. Ali Torabi Haghighi and Professor Bjørn Kløve, and Dr. Pekka Rossi whose expertise was invaluable in formulating the research questions, conceptualization, methodology, supervising, and manuscript review and editing. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would like to take this opportunity to thank members of my doctoral training follow-up group Dr. Hannu Marttila and Dr. Katharina Kujala for giving their valuable time for the meetings and providing helpful guidance.

I would like also to acknowledge my colleagues from my other country at Iran, Norway and Australia for their wonderful collaboration. I want to thank you for your patient support and for all the opportunities I was given to further my research.

I consider myself fortunate for the opportunity to have this thesis pre-examined by Dr. Ali Mehran and Dr. Olga Makarieva for their constructive and valuable comments throughout my dissertation and help me to improve this thesis to its current format.

My appreciation also goes out to my family and friends for their encouragement and support all through my studies. Especially, I would like to thank my friends Ehsan Hassani Nejad Gashti and Abolfazl Jalali Shahrood for their great supports.

Finally, I could not have completed this dissertation without the great support of my lovely wife, who provided boundless love as well as happy distractions to rest my mind outside of my research.

I would like to express my deep gratitude for financial supports by the Maa- ja vesitekniikan tuki r.y. (MVTT), Olvi, and Sven Hallin Foundations and I also, would like to thank Sari and Amol municipality (In Iran) for providing relevant data.

17.02.2022 Hamid Darabi

List of abbreviations and definitions

ACC	Accuracy
AHP	Analytic Hierarchy Process
ANN	Artificial Neural network
ANN-SGW	Artificial Neural Network-Swarm intelligence and Grey Wolf
asl	above sea level
BA	Balanced accuracy
BM	Bookmaker informedness
BRT	Boosted Regression Tree
CART	Classification and regression tree
CN	Curve Number
CSI	Critical Success Index
DL	Deep Learning
FANP	Fuzzy Analytical Network Process
FDR	False Discovery Rate
FM	Fowlkes–Mallows index
FN	False negative
FNR	False Negative Rate
FOR	False Omission Rate
FP	False positive
FPR	False Positive Rate
GAM	Generalized Additive Model
GARP	Genetic Algorithm for Rule-set Prediction
GLM	Generalized Linear Model
GWO	Grey wolf optimizer
HSG	Hydrological Soil Group
J48DT	J48 Decision Tree
LULC	Land Use/Land Cover
MARS	Multivariate Adaptive Regression Spline
Maxent	Maximum entropy
MCC	Matthews Correlation Coefficient
MK	Markedness
ML	Machine learning
MLPNN	Standalone MultiLayer Perceptron Neural Network
MR	Misclassification Rate
MultiB-MLPNN	Multi-Boosting MultiLayer Perceptron Neural Network

NPV	Negative Predictive Value
PPV	Positive Predictive Value
PT	Prevalence Threshold
QUEST	Quick, Unbiased, and Efficient Statistical Tree
RF	Random Forest
RL	Reinforcement Learning
RMSE	Root Mean Square Error
ROC-AUC	Receiver Operating Characteristic-Area Under the Curve
SL	Supervised Learning
SVM	Support Vector Machine
TN	True negative
TNR	True Negative Rate
TP	True positive
TPR	True Positive Rate
TS	Threat Score
UL	Unsupervised Learning
USCS	United State Soil Conservation Service

List of original publications

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:

I Darabi, H., Choubin, B., Rahmati, O., Haghighi, A. T., Pradhan, B., & Kløve, B. (2019). Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *Journal of Hydrology*, *569*, 142–154. https://doi.org/10.1016/j.jhydrol.2018.12.002

II Darabi, H., Haghighi, A. T., Mohamadi, M. A., Rashidpour, M., Ziegler, A. D., Hekmatzadeh, A. A., & Kløve, B. (2020). Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. *Hydrology Research*, *51*(1), 127–142. https://doi.org/10.2166/nh.2019.090

III Darabi, H., Rahmati, O., Naghibi, S. A., Mohammadi, F., Ahmadisharaf, E., Kalantari, Z., Haghighi, A. T., Soleimanpour, A. M., Tiefenbacher, J. P., & Tien Bui, D. (2021). Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood. *Geocarto International*, 1–27. https://doi.org/10.1080/10106049.2021.1920629

IV Darabi, H., Haghighi, A. T., Rahmati, O., Shahrood, A. J., Rouzbeh, S., Pradhan, B., & Tien Bui, D. (2021). A hybridized model based on neural network and swarm intelligence-grey wolf algorithm for spatial prediction of urban flood-inundation. *Journal of Hydrology*, *603*, 126854. https://doi.org/10.1016/j.jhydrol.2021.126854

Contribution by Hamid Darabi to Papers I–IV:

I Designed the study with Prof. Bjørn and Dr. Ali. Conceptualization, methodology, data collection and acquisition, manuscript preparation and revising.

V Designed the study with Prof. Bjørn and Dr. Ali. Conceptualization, methodology, data collection and acquisition, manuscript preparation and revising.

VI Designed the study, conceptualization, data collection and acquisition, manuscript preparation, and editing with co-authors.

VII Designed the study, conceptualization, methodology, data collection and acquisition, manuscript preparation and revising with co-authors.

Contents

1 Introduction

1.1 Importance of the present study

Natural disasters, such as floods, droughts, earthquakes, landslides, hurricanes, and volcanic eruptions, are unexpected catastrophic events with atmospheric, geological, or hydrological origins. They can have an immense impact on humans and the environment, causing significant socio-environmental disruption, infrastructure damage, and economic losses (Xu et al., 2016; Yousuf et al., 2020). Over the past 30 years, an average of 330 natural disasters have been reported worldwide annually (Centre for Research on the Epidemiology of Disasters [CRED], 2020), typically causing more than $100 billion in damages every year (Pielke, 2018). Meteorological, climatological, and hydrological disasters such as floods and storms have become more frequent in recent decades (Kong et al., 2021). However, advances in technology have assisted in predicting some of them (Witting et al., 2020).

A flood is an overflow of water related to hydro-meteorological and geophysical conditions. Floods are the most common type of natural disaster, accounting for about 41% of all-natural disasters reported globally over the past 30 years (CRED, 2020). They are also among the most hazardous, damaging, and costly (Barredo, 2007), causing widespread human fatalities and property losses every year (Kuang & Liao, 2020). Urbanization and extreme weather conditions, such as heavy precipitation, can exacerbate flood losses. During extreme weather conditions, for example, floods can be part of a complex of hazards that includes mudslides and tidal surges (Few 2003; Wu et al., 2012).

In general, flood impacts are classified into two categories: direct and indirect (Albano et al., 2018; Carrera et al., 2015; Merz et al., 2010). Direct impacts arise through the physical contact of flood water with people, possessions, and other entities during the flood event; whereas indirect impacts arise after the flood event or outside of the flooded area (Albano et al., 2018; Koks, Bočkarjova et al., 2015). Impacts can also be categorized into tangible and intangible impacts, depending on whether they can be assessed in terms of economic value (Parker et al., 1987; Smith & Ward, 1998).

Floods in cities are often devastating because they involve high densities of people and assets concentrated in areas where flood potential is exacerbated by disturbances to nature (Cherqui et al. 2015; Kjeldsen 2010). Urbanization is usually

associated with a high proportion of impervious features, disturbed river/stream channels, and artificial storm drainage systems (Suriya & Mudgal, 2012). Therefore, urban flood mapping is an evolving challenge for city managers and policymakers working to reduce flood risk (Noh et al. 2016). Effective risk and vulnerability assessments require thorough knowledge of the conditions affecting flooding and its impacts (Ouma & Tateishi 2014). At the heart of such assessments are flood risk maps, which can be created in various ways, including through the use of hydrological and hydraulic models (Booij 2005; Masood & Takeuchi 2012), integration of analytic hierarchy process (AHP) and geographic information system (GIS) techniques (Ouma & Tateishi 2014), multi-criteria evaluation (Meyer et al. 2009), systems simulation (Amendola et al. 2000), and probability-based analysis (Jalayer et al. 2014). In addition, an emerging approach involves the application of machine learning methods (Termeh et al. 2018; Zhao et al. 2018). However, despite recent advances in using machine learning to create maps for flood risk assessments (Choubin et al., 2019), limitations still exist, and new hybridized and new optimized models are needed.

1.2 Different types of floods

Nied et al. (2014) outlined three different methods to describe flood events: i) using a flood event description (i.e. a detailed description provided by scientific case studies or grey literature documentation), ii) linking flood occurrence probability with large-scale atmospheric circulation patterns, or iii) classification of floods into flood types (i.e. grouping specific flood events into different clusters based on causal processes) (Blöschl et al., 2013; Delgado et al., 2014; Gaál et al., 2012). Following the third method, floods can be classified into different types based on weather-related circumstances, such as the spatial distribution and amount of rainfall, and the antecedent soil moisture conditions (Turkington et al., 2016). Floods can generally be categorized into three types depending on their characteristics: river floods, flash floods, and tidal floods (Turkington et al., 2016).

1.2.1 River floods (fluvial floods)

River flooding is a complex type of flood phenomenon that can be affected by changes in physical and terrestrial conditions and socio-economic (refers to the social setting and environmental conditions in which people live or in which something happens or develops) and climate conditions (Kundzewicz et al., 2010).

Precipitation for an extended period over a large area leads to an overflow and contributes to the river flood (Ikeuchi et al., 2017). However, river floods can happen along big or small rivers and are usually caused by an extended period of precipitation. River flood damage can be extensive, as the overflow can affect nearby structures, break dams and dikes, and impact surrounding swamp zones (Maddox, 2014). In some parts of the world, river floods can arise after a cold period due to the 'ice jam' phenomenon. This occurs when floating river ice breaks up and the pieces become clogged at a natural or man-made feature; this can block the water flow upstream such that the river breaks its banks (Wright, 2007). The intensity and duration of precipitation can determine the severity of a river flood. Other influencing factors include soil moisture conditions due to earlier rainfall events, and the topography near the river. In river flood modeling and prediction, models consider past precipitation as well as soil and terrain antecedent conditions, current river levels, and forecasted precipitation (Ikeuchi et al., 2017; Wang, Xia, et al., 2020).

1.2.2 Flash floods (pluvial floods)

Flash floods are frequently caused by heavy precipitation events and typically happen in mountainous areas with steep topography. They are among the most dangerous and costly natural disasters worldwide (Zhai et al., 2020). An intensive rainfall event occurring in a small area causes water to quickly flow downstream, resulting in a flash flood (Saharia et al., 2017; Trigo et al., 2016; Zhai et al., 2020). Flash floods are characterized by an intensive and high velocity of water flow and caused by heavy precipitation in a short period in a small area or on nearby elevated topography. Flash floods can also occur through the rapid release of water from upstream to downstream. They are very destructive and hazardous not only because of the force of the water but also because of the careering debris flow (Khosronejad et al., 2020).

Extreme precipitation is a major cause of flood-related damage to humans and assets in two ways: i) when precipitation intensity is higher than infiltration capacity, pluvial or flash floods can arise, ii) and when high-water levels in river canals exceed bank heights or cause dam breaks, fluvial or river floods can arise (Moftakhari et al., 2017; Yin et al., 2016). Flash floods are dominated by short-term high-intensity precipitation and usually occur more frequently than river floods. While river floods usually have more severe floods and therefore cause serious human and economic damage (Moftakhari et al., 2017; Yin et al., 2016).

1.2.3 Tidal surge (coastal flood)

Tidal or coastal floods are the inundation of areas near coastal zones by seawater, which can severely impact residential areas and industrial infrastructure, as well as tourism (Irawan et al., 2021). Coastal floods have destructive impacts on coastal communities such as population displacement, loss of life, and damage to coastal infrastructure due to flooding conditions and high velocity storms and winds. A major source of coastal floods is extreme storm events happening at the same time as high tide (Leijnse et al., 2020; Marfai & King, 2008). In coastal floods, water flows over low land, and in many regions, damages caused by tidal surges are expected to increase significantly, as coastal infrastructure (refers to structures, systems, and facilities built along coastlines) and value of possessions increase (Hinkel et al., 2014). The severity of a coastal flood is determined by several factors, such as the size, strength, direction, and speed of the storm. Also, damages due to tidal surges have been affected by onshore and offshore topography. To recognize the magnitude and probability of a tidal surge, coastal flood models can use data and information from historical storms (Leijnse et al., 2021; Tanim & Goharian, 2020).

1.3 Urban flood

Many cities are located in flood plains because the land there is flat and suitable for urban development (Das et al., 2020). Urban flood is a severe and inescapable natural hazard which arises due to heavy rainfall events and limited drainage in the urban environment. Cities have a large percentage of impervious areas that prevent effective infiltration of rainfall into soil, leading to high surface runoff and flooding potential (Tingsanchali, 2012). Urban environments with complex geophysical and hydrometeorological conditions may face multiple threats from various types of floods, such as river floods, flash floods, and tidal floods (Wu et al., 2018). Therefore, floods can cause devastation of urban communities and loss of life, with numerous harmful socio-economic effects (Paper II; Yang et al., 2020). As urbanization spreads, flood management is getting increasingly challenging and the concept of flood resilience receives growing attention (Kuang & Liao, 2020). During the last decades, different types of urban floods are slowly being given more attention. Furthermore, flood risk and vulnerability for different types of floods have been addressed in several studies (Fernández & Lutz, 2010; Paper I; Paper II). Furthermore, several researchers have considered the hazards associated with

different flood types and built models involving geographic information systems (GIS) and socio-economic variables (Paper II). In the previous studies, some researchers have considered flood risk and vulnerability variables, and, in the context of meteorological and hydrological conditions they have considered some related factors (Ahmad and Simonovic, 2013; Eini et al., 2020; Karmakar et al., 2010; Koks, Jongman et al., 2015; Masood & Takeuchi, 2012; Nasiri et al., 2016). Therefore, there is a crucial need to consider different flood types in the analysis of urban flood hazards, vulnerability, and risk (Yang et al., 2020).

1.4 Machine learning

Machine learning (ML) appeared in the late 1950s, from the field of artificial intelligence (AI) and it has been conceptualized as a subsection of AI, which delivers machines the capability to learn and improve robotically from data and previous experiences without explicit programming (Haenlein & Kaplan, 2019; Poole et al., 1998; Wang, Ryoo, et al., 2020). AI is well-defined as a capability system to understand and interpret external data appropriately to learn from data and to employ those learnings to reach explicit aims. AI branches are strictly related to language processing, computer vision, and ML (Haenlein & Kaplan 2019; Wang, Ryoo, et al., 2020).

ML was first used (Samuel, 1959) to refer to a process by which heuristics pruned possibilities in a program, and it is now the scientific field that focuses on training algorithms (Kohavi & Provost, 1998). ML contains methods and designed algorithms to learn the original patterns in the data and make predictions based on the patterns (Dzyabura & Yoganarasimhan, 2018; Ma & Sun, 2020). ML is one of the most prominent and famous research fields in knowledge identification and the discovery of hidden patterns from datasets. It emphasizes the training of intelligent models so that model prediction can be more accurate and faster (Arunakranthi et al., in press; Raju & Bonagiri, in press). ML combines statistical and outcome optimization in various applications including machine vision and anatomy, biometry, bioinformatics, face identification, natural language processing, information retrieval, pattern recognition, consumer analysis, and content storage and retrieval (Raju & Bonagiri, in press). The ever-increasing amounts of available data have produced inspiring achievements in machine learning and artificial intelligence. The recent developments, successes, and popularity of machine learning have focused on using past and historical data to better predict unobserved consequences (Haenlein & Kaplan 2019).

1.4.1 Different type of ML

Machine learning uses three types of learning techniques including supervised learning, unsupervised learning, and reinforcement learning.

1) Supervised learning (SL) is the construction of algorithms that produce general patterns and assumptions using externally supplied instances. SL is divided into two steps, training and testing, and it contains various algorithms with different functions that use inputs to create preferred outputs. SL classification algorithms aim to categorize new data based on prior knowledge and information, by training a model on input data so that it can forecast future outputs. Also, the aim of SL is to build a model that produces predictions based on evidence as independent variables. The main purpose of SL models is to learn how to predict a random variable based on a set of random variables. SL uses classification and regression algorithms to develop predictive machine learning models: classification algorithms predict discrete responses, whereas regression algorithms predict continuous responses (Crisci et al., 2012; Nasteski, 2017; Singh et al., 2016).

2) Unsupervised learning (UL) discovers hidden patterns or basic structures in data, and it is a type of algorithm that learns patterns (use to draw inferences) with unlabeled responses and data. Clustering is the most common UL technique, and applications for clustering include gene sequence analysis, market research, and object recognition. Since UL does not require labeled data and is complementary to current processing methods, it can be applied for data exploration to identify preferred feature information (Watson, 2020).

3) Reinforcement learning (RL) is a complicated processing method that identifies the relationship between response and predictive variables, and it maximizes the reward of an agent and decision-maker for candidate actions (Lohi & Tiwari, in press). The reinforcement learning models are used to obtain effective methods and resources, and they work by rule-based methods, and the functions in the reinforcement learning are calculated for occurrence. Reinforcement learning algorithms interact with an environment, so a feedback loop exists between the learning system and its experiences (Karthiban, 2020). Deep learning (DL) as a strong autonomous learning ability, is more reliable and with highly performance-conscious learning problem in which the error rate is much lower. DL is often referred to as deep formal learning and hierarchy (Dimiduk et al., 2018). The common DL algorithms include Convolution Neural Network (CNN), Stacked Auto Encoder (SAE), Recurrent Neural Network (RNN), Deep Belief Network (DBN), and Feedforward Neural Network (FNN) (Xu et al., 2021).

1.5 Data mining vs machine learning

Data mining (DM), known as Knowledge Discovery in Databases (KDD), is a means of discovering and extracting structures, patterns, and useful information from large complex datasets. There are two steps to DM including model building and pattern detection (Baker, 2010; Hand & Adams, 2015). DM is a tool that is applied by a user to determine new, accurate, precise, and valuable patterns in datasets (Chen et al., 1996; Hand & Adams, 2015). But ML includes algorithms which allow to the model to learn without user intervention. ML is a collection of the models and tools to make a smarter algorithm and eliminate the human component (more data and a better model leads to higher accuracy) (Arunakranthi et al., in press; Raju & Bonagiri, in press).

1.6 Performance analysis

In performance analysis of the ML algorithms, a known error matrix is the confusion matrix, which allows the performance of supervised models to be assessed. The four elements of the confusion matrix are True Positive and True Negative (TP and TN), False Positive and False Negative (FP and FN), see Table 1 (Frattini et al., 2010). Different types of cutoff-dependent metrics derived from the confusion matrix include Positive and Negative Predictive Values (PPV and NPV), Sensitivity (S), Specificity (S), and Efficiency (E).

In our study, sensitivity (i.e. the True Positive Rate (TPR)), specificity (i.e. the False Positive Rate (FPR)), and efficiency determined the proportion of actual flooded locations that are accurately identified. The other cutoff-dependent metrics used, PPV and NPV, are the ratios of positive and negative results, respectively. Thus, model performance can be assessed by applying primary statistical criteria (TP and TN) in the confusion matrix (Fletcher et al., 2018). All cutoff-dependent metrics used for model performance assessment were applied. Generally, the accuracy assessment of ML algorithms is represented by i) cutoff-dependent metrics and ii) cutoff-independent metrics.

Table 1. The four basic elements of the confusion matrix (Modified from Frattini et al. 2010).

Observed	Predicted	
	Absence	Presence
Absence	True negative (TN)	False positive (FP), Type I error
Presence	False negative (FN), Type II error	True positive (TP)

1.6.1 Cut-off-dependent evaluation metrics

Cutoff-dependent metrics were extracted based on the confusion matrix. The components of the confusion matrix are true negative (TN), true positive (TP), false negative (FN), and false positive (FP) (defined in Table 2), where FP and FN are the numbers of pixels incorrectly categorized (type I and type II errors) and TN and TP are the numbers of pixels correctly categorized. True positive rate (TPR) (also known as 'recall', 'sensitivity' or 'hit rate') indicates the proportion of positive observations that have been properly classified as true. True negative rate (TNR) (also known as 'specificity' or 'selectivity') indicates the proportion of negative observations that have been properly classified as negative. False positive rate (FPR) (also known as 'fall-out' or 'type I error') indicates the proportion of negative observations that have been falsely classified as true. False negative rate (FNR) (also known as 'miss rate' or 'type II error') indicates the proportion of positive observations that have been falsely classified as negative. The *F*-score (also known as the *F*-measure) integrates precision evaluation metrics into one metric: the higher the score, the better the model. The *F*-score is calculated by Eq. 13 (Table 2). Accuracy (ACC) (also known as 'efficiency') is another evaluation metric for accuracy assessment of models, which determines how many positive and negative observations were appropriately classified. The Matthews correlation coefficient (MCC) assesses the performance of model based on the correlation degree between observations and predictions (Matthews, 1975). MCC ranges from −1 to 1, (1 shows perfect agreement between observed and predicted data and −1 shows significant disagreement). The Misclassification Rate (MR) uses both the false negative and false positive components and thus indicates an overall error rate. Other measurements, including Positive Predictive Value (PPV), Negative Predictive Value (NPV), False Discovery Rate (FDR), False Omission Rate (FOR), Prevalence Threshold (PT), Threat Score (TS) or Critical Success Index (CSI), Balanced accuracy (BA), Fowlkes-Mallows index (FM), Informedness or

bookmaker informedness (BM), and Markedness (MK) can be calculated by equations presented in Table 2.

1.6.2 Cut-off-independent evaluation metrics

The Area Under the Curve (AUC)-Receiver Operating Characteristic (ROC) Curve (AUC-ROC curve), is a cutoff-independent metric, and the most important metric for evaluating ML models in accuracy assessment (as it is effective in organizing models and assessing their performance visually). It can be calculated by Eq. 9. It is known as a decisive metric where values closer to 100% indicate better performance of a model (Fawcett, 2006; Pontius & Schneider, 2001). The AUC-ROC simply plots the TPR (sensitivity) on the *Y*-axis against the FPR (1−sensitivity) on the *X*-axis. ROC-AUC is considered a real evaluation metric of the model because it simultaneously and fairly calculates the overall quality of the ML models and includes all components of the confusion matrix (Frattini et al., 2010). AUC ranges from 0 to 1, and it can be calculated by Eq. 19. Higher values indicate better model performance.

Table 2. Confusion matrix metrics.

Metric name	Equation	
True positive rate	$TPR\ (\text{sensitivity or hit rate}) = \frac{TP}{TP+FN}$	(1)
True negative rate	$TNR\ (\text{specificity or selectivity}) = \frac{TN}{TN+FP}$	(2)
False positive rate	$FPR\ (\text{fall-out}) = \frac{FP}{FP+TN}$	(3)
False negative rate	$FNR\ (\text{miss rate}) = \frac{FN}{TP+FN}$	(4)
Positive predictive value	$PPV\ (\text{precision}) = \frac{TP}{TP+FP}$	(5)
Negative predictive value	$NPV = \frac{TN}{TN+FN}$	(6)
False discovery rate	$FDR = \frac{FP}{FP+TP}$	(7)
False omission rate	$FOR = \frac{FN}{FN+TN}$	(8)
Prevalence threshold	$PT = \frac{\sqrt{TPR(TNR-1)}+TNR-1}{(TPR+TNR-1)}$	(9)

Metric name	Equation	
Threat score	$TS = \frac{TP}{TP + FN + TP}$	(10)
Accuracy	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	(11)
Balanced accuracy	$BA = \frac{TPR + NPR}{2}$	(12)
F-score	$F\text{-score} = \frac{2TP}{2TP + FP + FN} = \frac{2 \times (PPV \times TPR)}{PPV + TPR}$	(13)
Matthews correlation coefficient	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (FN + TN) \times (FP + TN) \times (TP + FN)}}$	(14)
Fowlkes-Mallows index	$FM = \sqrt{\frac{TP}{TP + FP} \times \frac{TP}{TP + FN}} = \sqrt{PPV \times TPR}$	(15)
Informedness	$BM = TPR + TNR - 1$	(16)
Markedness	$MK = PPV + NPV - 1$	(17)
Misclassification rate	$MR = \frac{FP + FN}{FP + FN + TP + TN}$	(18)
AUC-ROC curve	$AUC = \int_0^1 f(FPR)\, d\, FPR = 1 - \int_0^1 f^{-1}(TPR)\, d\, TPR$	(19)

1.7 Objective of the study and thesis structure

The principal objective of this study was to employ novel machine learning methods in urban flood modelling. The work assessed different modelling approaches and several types of factors that affect urban floods. Different hazards, vulnerabilities, and risks were assessed using the different models. Key sub-tasks included: i) contribution of conditioning factors, ML techniques and GIS to obtain urban flood maps, ii) applying the new ML algorithms to urban flood mapping, iii) developing ensemble, hybridized and optimized ML models and using them to determine the importance of different influential factors in urban flood maps.

Therefore, the study focused on urban flood assessment and provided methods to predict the urban flood maps and provide key information to city planners for better safety by addressing the following research questions:

- Are new machine learning techniques suitable for urban flood risk mapping?
- How can we select the best machine learning model to urban flood risk mapping?

– How can we create effective machine learning models for urban flood risk mapping?

To answer these questions, three main tasks (i.e., work packages, WPs) were defined as follows:

i. Urban flood risk mapping using new data and machine learning techniques.
ii. A comparative study of machine learning techniques using evaluation metrics.
iii. Developing novel and robust ensemble, optimized and hybridized machine learning-based approaches for urban flood assessment.

This doctoral thesis is based on four peer-reviewed original scientific publications (Papers I–IV). Their aims and contribution to the thesis goals are presented in Table 3.

Table 3. Summarizing of the scope and structure of the work for papers included in the thesis.

Stage of research	Time					
	2019		2020		2021	
	1st half	2nd half	1st half	2nd half	1st half	2nd half
WP1	Paper I	Paper I				
WP2			Paper II	Paper II	Paper III	Paper III
WP3					Paper IV	Paper IV
Writing thesis					Thesis	Thesis

Paper I addressed the first work package. Paper II and III addressed the second and third work packages. Paper III and IV addressed all three work packages.

The novelty of the study lies in: i) developing a spatial framework for urban flood mapping by training ML algorithms on flood inventory data and data associated with influential urban flood risk factors, ii) developing ensemble and hybridized models for urban flood mapping, where newly developed models will employ the advantages of previous standalone models. Key sub-tasks include the following: i) Combination of ML algorithms, GIS, by employing socio-environmental conditioning factors in urban flood vulnerability, hazard, and risk mapping. ii) Developing an optimized model. iii) Accuracy assessment of the urban flood mapping using new cutoff-dependent and cutoff-independent metrics. iv)

Rank the importance of different socio-environmental conditioning factors in urban flood.

This study focused on applying ML algorithms to available hydrological data in the R software environment and aimed to develop a framework to produce flood risk maps for urban environments in data-scarce regions where there is no detailed information about hydrological and hydraulic characteristics. In such areas, long-term and continuous rainfall data are not available, which presents a challenge to researchers. In addition, numerical and hydraulic-based models cannot be used in such areas, because they require several kinds of data (comprehensive and time-series). In these conditions, ML techniques helped us to predict flood risk situation by considering several previous heavy rainfall events that resulted in devastating floods.

2 Study area

This study was carried out in the two cities (Sari and Amol cities) in the North of Iran. The cities selected for the present study faced many disastrous floods that occurred yearly, damaging homes and infrastructure, disrupting traffic, trade, and public services. These two cities are located on the southern coast of the Caspian Sea in the Mazandaran province, Iran (one of the most important geographical areas by the semi-humid climate and high annual precipitation). The Sari and Amol cities lie near the outlet of the Tajan and Haraz watersheds with 4352 km^2 and 1100 km^2 in area, respectively. The Tajan and Haraz rivers passing through the heart of Sari and Amol cities respectively, then reaching the Caspian Sea and sometimes the riverine floods are main reason to be flood the mentioned cities (Mohseni-Bandpei & Yousefi, 2013).

2.1 Sari city

Sari city is extended between 35°58′39″ to 36°50′12″ North and 52°56′42″ to 53°59′32″ East, and it is one of the biggest cities in Mazandaran province, Iran. The altitude of the Sari city is between 9 and 82 m above sea level (asl) (Fig. 1). The population of Sari city is 296,417, which makes it the second populated city in the North of Iran (Zali et al., 2016). The area of the Sari city is about 42 km^2, and it is located at the outlet of the Tajan watershed (4352 km^2). The Tajan river is passing through the Sari and then discharging into the Caspian Sea. The residential areas of the Sari city are surrounded mainly by orchards, agricultural land, and high forested mountains. The climate type in the region is semi-humid with an aridity index of 0.73 which was calculated by precipitation (735 mm/year) and potential evapotranspiration (1004 mm/year) for the period 1986–2016, which was recorded at Sari weather station and obtained from the Iranian Meteorological Organization (IRIMO). In Table 4, meteorological variables (precipitation [mm], evapotranspiration [mm], temperature [°C]) are presented for the study Sari city.

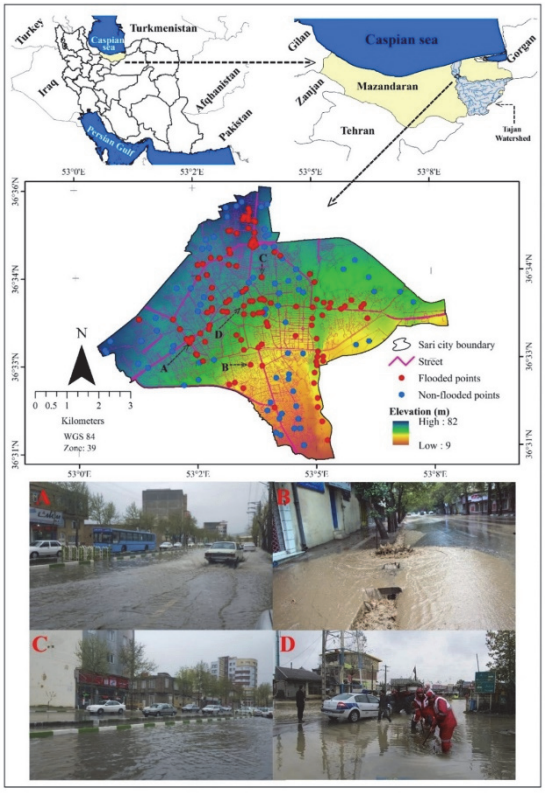

Fig. 1. Location of Sari city in Iran (top panel), and flood event in Sari city (bottom panel A–D) (Adapted, with permission, from Paper I © 2018 Elsevier B.V.).

Table 4. Meteorological variables of the Sari city in 1986–2016 (Data from IRIMO).

Climate variable	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Precip. (mm)	60.44	70.11	80.94	41.04	32.13	32.36	24.41	32.07	70.46	102.22	108.52	80.29
ET[1] (mm)	27.60	30.37	46.92	72.14	102.46	143.06	156.89	155.62	106.75	82.00	50.66	30.14
T_{min} (°C)	3.85	4.70	7.53	11.40	16.53	20.87	23.15	23.43	20.83	15.65	9.33	5.28
T_{avg} (°C)	8.60	9.12	12.08	16.16	21.20	25.34	27.45	28.21	25.21	20.26	14.08	9.95
T_{max} (°C)	13.34	13.55	16.62	20.91	25.88	29.81	31.75	32.99	29.59	24.86	18.82	14.61

[1]ET = Evapotranspiration

2.2 Amol city

Amol city is extended between 36°26′02″ to 36°29′45″ North and 52°19′14″ to 53°23′50″ East. The population of Amol city is 237,528, which makes it as the third most populated city in Mazandaran province in northern Iran (Lotfi et al., 2016). The altitude of Amol city is between 59–137 above sea level (asl), and the area of Amol city is 27 km^2 (Fig. 2). Amol city is located on the outlet of the Haraz watershed, and the Haraz river passes through the Amol city, then discharging into the Caspian Sea. The residential areas of Amol city are surrounded mainly by orchards, agricultural land, and high mountains covered by forest. The mean annual precipitation in the region is 680 mm with the semi-humid climate (Sahin, 2012). Over recent years, urban development has led to considerable changes in hydrological processes and drainage systems in the Amol city, which led to increase the number of urban flood events. Notable flood events occurred on November 12, 2012 (with damage to more than 40 residential areas), September 16, 2015 (31 areas), November 24, 2015 (40 areas), and October 15, 2016 (10 areas) (Sedaghat et al., 2016). In the Table 5, meteorological variables (precipitation [mm], evapotranspiration [mm], temperature [°C]) are presented for the study Amol city.

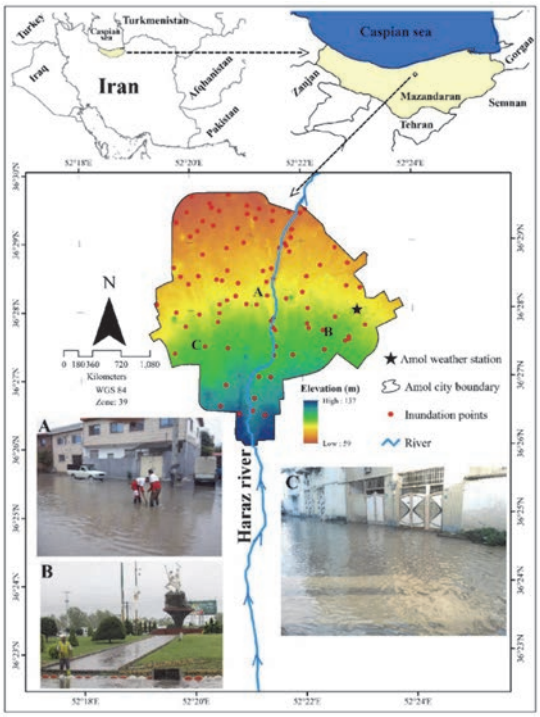

Fig. 2. Location of Amol city in Iran (top panel) and flood event in Amol city (bottom panel A–C) (Adapted under CC BY 4.0 license from Paper II © 2020 Authors).

Table 5. Meteorological variables of the Amol city in 1986–2016 (Data from IRIMO).

Climate	Month											
variable	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Precip. (mm)	59.58	58.42	63.49	35.26	20.86	20.26	22.10	28.59	61.85	111.63	115.66	82.73
ET[1] (mm)	30.94	29.29	38.96	68.81	96.01	127.19	140.35	147.74	121.85	83.33	58.42	39.94
T_{min} (°C)	3.33	4.35	7.38	11.43	17.02	20.82	22.24	22.28	20.12	15.11	8.86	4.69
T_{avg} (°C)	8.10	8.62	11.61	15.54	20.84	24.66	26.46	27.14	24.45	19.77	13.68	9.43
T_{max} (°C)	12.88	12.89	15.84	19.65	24.67	28.51	30.67	32.01	28.78	24.46	18.52	14.17

[1]ET = Evapotranspiration

3 Materials and methods

The current study has been arranged as following sections; i) Collecting flood inventory data, ii) Preparing flood influencing factor, iii) Applying new ML algorithms in urban flood mapping and developing an ensemble, optimized and hybridized models, iv) Accuracy assessment of the in urban flood mapping using new cutoff-dependent/independent metrics. Methodological flowchart of the research of the current study is presented in Fig. 3. In the following sections data used and methodology are described in detail.

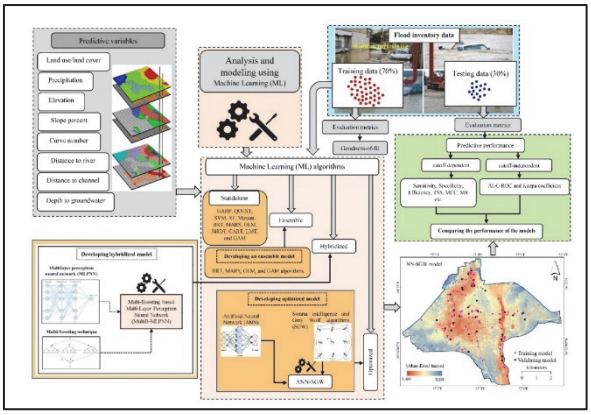

Fig. 3. Methodological flowchart of the study.

3.1 Materials

3.1.1 Influencing factors in the urban flood hazard

There are no instructions universally for choosing influencing factors on the urban flood. In the current study, eight hazard influencing factors and five vulnerability influencing factors were employed based on the literature review to produce the urban flood hazard and vulnerability maps (Chen et al., 2009; Fernández et al., 2010; Thieken et al., 2005). The hazard influencing factors were selected as precipitation, land use/land cover (LULC), curve number (CN), elevation, slope percent, distance to channel, distance to river, and depth to groundwater. All the hazard influencing factors were prepared in the raster format in ArcMap 10.5

environment. Precipitation: daily precipitation data were obtained from the IRIMO over the Mazandaran province, Iran to prepare the precipitation raster map using interpolation techniques. Land use/land cover: since runoff generation significantly depends on the various LULC types, LULC map for 2015 was obtained from Sari and Amol cities authority and based on this map, the study area was subdivided into three main group patterns: open spaces (including agricultural area, orchard, and parks,), urban districts (residential buildings, industrial areas, commercial and business buildings), and water body (river). Elevation: A Digital Elevation Model (DEM) was obtained from Sari and Amol cities authority with 5 m resolution. The elevation of the study area ranges 9–82 and 59–137 m asl for Sari and Amol cities, respectively. Slope percent: the slope percent affects the water velocity, and it plays an important role in flooding conditions (Wang et al., 2015). The slope maps were extracted from the DEM of the studies areas in ArcGIS 10.5 to quantify topographic controls on hydrological processes. Curve number (CN): is a function of land use patterns and hydrological condition, antecedent soil moisture, and soil type which is developed for the first time by the United State Soil Conservation Service (USCS). The CN maps of the studied areas were extracted based on the LULC maps, the Hydrological Soil Group (HSG) maps, and a lumped CN value, using the ArcCN-runoff tool in the ArcMap environment (Zhan & Huang, 2004). Distance to river: the banks of the Tajan and Haraz Rivers are the main flood-prone zones in Sari and Amol cities respectively. Hence, distance to those rivers plays an important role in urban flood conditions in the Sari and Amol cities. Distance to channel: or drainage systems in the urban environments collect surface runoff, and it has an important role in evacuation of the surface water and prevention of the urban inundation. The maps of ‘distance to river’ and ‘distance to channel’ were prepared using the Euclidian Distance module (Paper I) in ArcMap 10.5 environment for Sari and Amol cities. Depth to groundwater level: since infiltration capacity significantly depends on the groundwater level and soil moisture, depth to groundwater directly affects the surface runoff during high-intensity precipitation (Fernández & Lutz, 2010; Yin & Li, 2001). The groundwater level data used in this study were obtained from Iranian Water Resources Management Company (IWRMC). Among the influencing factors on urban flood hazard, elevation, slope, precipitation, depth to groundwater, distance to river and distance to channel were continuous raster layer, while LULC and CN were categorical raster layers. Fig. 4 shows hazard influencing factors on the urban flood over the Sari city.

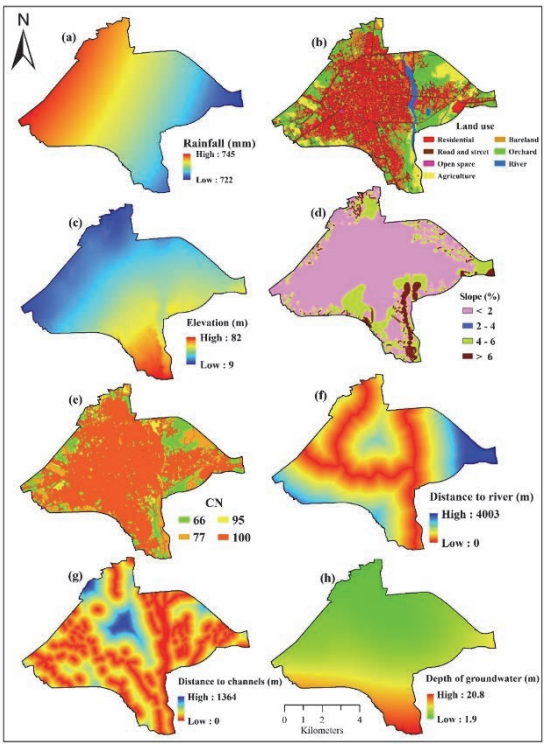

Fig. 4. Hazard influencing factors on the urban flood over the Sari city (Reprinted, with permission, from Paper I © 2018 Elsevier B.V.).

3.1.2 Influencing factors in the urban flood vulnerability

Flood vulnerability is produced by the conditions of the social, economic, and physical factors (United Nations Development Programme [UNDP], 2004). Various factors influence vulnerability of the urban flood and their inclusion depends on availability of the data (Dayal et al., 2018). The current study considered the quality of buildings, urban density, population density, age of buildings, and socio-economic conditions in urban flood vulnerability mapping over the Sari city (Fig. 5). The urban density which classified into four categories of high, medium, low, very low density. Since quality and age of buildings reflect building condition it has been considered as significant factors on the urban floods damage. Quality of buildings was classified into very high, high, medium, low, very low-quality classes. Low quality buildings are dangerous for inhabitants and are

more vulnerable to natural hazards, especially floods (Gerl et al., 2014; Schubert & Sanders, 2012). The architectural structure of a city reflects its historical development or age of buildings. Buildings in the Sari city were divided into five age classes: very old, old, medium, new, and newest. Data on the age and quality of buildings were obtained from Sari and Amol cities authority in 2015. Population density states the number of people occupying a given area, and density-populated areas are usually associated with high vulnerability to natural hazards such as earthquakes and floods (Güneralp et al., 2017). In this study, population density was classified into high, medium, low, very low-density categories. Socio-economic data states depth information on the inherent properties and behavior of society and humans within a specific region. These types of information are valuable in considering intangible impacts of floods (Kaspersen & Halsnæs, 2017). In this study, socio-economic map was considered as A, B, C, D, and E conditions reflecting very good, good, moderate, weak, and very weak conditions respectively.

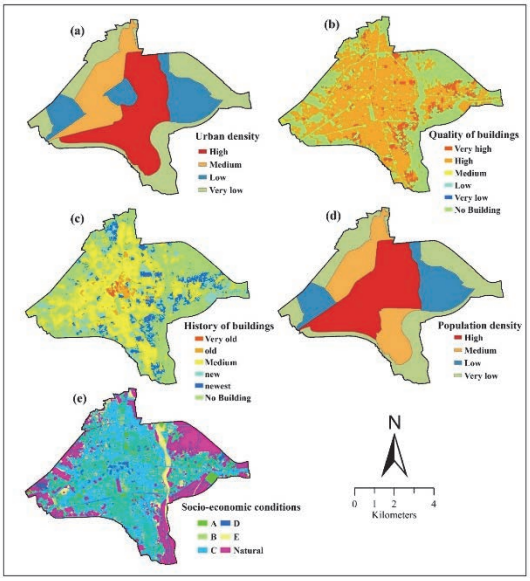

Fig. 5. Vulnerability influencing factors on urban flood over the Sari city (Reprinted, with permission, from Paper I © 2018 Elsevier B.V.).

3.1.3 Flood inventory data

All available historical database on floods were used as auxiliary data (flood and non-flood locations during the several past flood events over the Sari and Amol cities). Also, field surveys were conducted in Sari and Amol cities to identify flooded and non-flooded areas during intensive precipitation. Two equal sets of points were identified randomly for both categories of flood and non-flood areas. Therefore, 113 and 118 locations were selected respectively for each category over the Sari and Amol cities. Each set is divided into training (70% of data: 79 and 82 points over the Sari and Amol cities, respectively) and testing (30% of data: 34 and 35 points over the Sari and Amol cities, respectively) groups. The flood locations were allocated by a value of 1 and non-flood locations were assigned by a value of 0.

3.2 Methodology

3.2.1 GIS layer preparation

At the first step, all necessary data for urban flood hazard and vulnerability layers have been successfully acquired, the user would need to prepare the datasets for their input into the machine learning algorithms. This requires the creation of raster files. There are following steps that needed to be undertaken to process the GIS data so that it can be used by machine learning algorithms. i) Appropriate coordinate system, the user would need to secure all the datasets. Before starting the data mining in machine learning algorithms, the user should be sure that all datasets have the same coordinate system for raster and vector layers. ii) Appropriate layer projection, the user would need to determine the projection system. iii) Produce a raster file with the same raster size. iv) Determining the same spatial extent of the raster and vector maps to the area of interest. v) Deleting extra attributes and creating new attributes value based on the different types of the machine learning algorithms requirements and language programing in the next steps.

3.2.2 Urban flood vulnerability map

In this study, the weight and rank values were assigned to conditioning factors on urban flood vulnerability map using expert knowledge and Fuzzy Analytical

Network Process (FANP) and Analytic Hierarchy Process (AHP) for the Sari and Amol city respectively. Influencing factors incorporated on the vulnerability are presented in Fig. 4 for the Sari city. The relative weights of influencing factors in flood vulnerability mapping were determined using FANP as a decision-making technique incorporating the Fuzzy set theory with Analytical Network Process (ANP). FANP is conducted in five steps: i) Transformation of the problem to a network structure. This phase was carried out using Fuzzy DEMATEL to design a network structure from influencing floods vulnerability factors. ii) Pairwise comparisons of variables based on their importance, using Fuzzy extent analysis. iii) Estimation of the preliminary super-matrix based on the weights of the influencing floods vulnerability factors. iv) Calculation of the weighted super-matrix through multiplying the preliminary super-matrix by cluster weights. v) Conversion of the weighted super-matrix into a limit super-matrix and determination of priorities and the importance of the influencing floods vulnerability factors (Buyukozkan & Cifci, 2012; Wu & Lee, 2007). The AHP, FANP, and Fuzzy DEMATEL methods were used in the expert choice and MATLAB software, respectively, and then final layers with their weights were overlaid in the ArcMap environment.

3.2.3 Machine learning algorithms for urban flood hazard

Machine learning aims to develop methods and algorithms to learn and forecast data. ML algorithms are important for several reasons, e.g., they can i) handle complex systems with huge data, ii) solve specialized problems by specialized machine learning methods, iii) be used for forecasting and classification purposes, and iv) be applied to describe the performance of the data input concerning historical data records (Demolli et al., 2019). We applied the following ML algorithms to predict urban flood hazard using influencing factors: Genetic Algorithm for Rule-set Prediction (GARP), Quick, Unbiased, and Efficient Statistical Tree (QUEST), Support Vector Machine (SVM), Random Forest (RF), Maximum entropy (Maxent), Boosted Regression Tree (BRT), Multivariate Adaptive Regression Spline (MARS), Generalized Linear Model (GLM), Generalized Additive Model (GAM), Support Vector Regression (SVR), Classification And Regression Tree (CART), J48 Decision Tree (J48DT), and Logistic Model Tree (LMT). In the Table 6, different ML algorithms used in this study are presented.

Table 6. Machine learning algorithms used in this study.

ML algorithm	Description
Genetic Algorithm for Rule-set Prediction (GARP)	GARP has exposed excellent predictive ability in different fields. It is inspired by different genetic evolution models that analyze the relationship between flood inventory dataset and topo-hydrological influencing factors through iterative processes and conditional rules (Peterson et al., 2002; Qin et al., 2015). In this study the GARP model produces a few predictions for urban flood inundation through iterative processes to improve the stability of output. Executing numerous runs to obtain different results and outputs using the best-subset method is important in selecting the best output with optimum parameters. The user defines the set of models that reaches harmony between omission and commission error thresholds (Anderson, 2003; Boeckmann & Joyner, 2014). The GARP output is a raster map over the study area, which can be used in ArcMap environment to illustrate flood-prone areas (Boeckmann & Joyner, 2014). In this study, Desktop GARP software was run to obtain the urban flood hazard map over the Sari city.
Quick, unbiased, and efficient statistical tree (QUEST)	QUEST, a popular data-mining model, is a tree-structured classification algorithm that produces a growing binary-split decision tree. Also, QUEST yields subsets of the data that are as homogeneous as possible with respect to the response or dependent variable (Lee & Park, 2013; Ture et al., 2009). It utilizes a linear discriminate analysis method in splitting tree nodes which uses a sequential tree growing method (Ierodiaconou et al., 2011). Also, the unbiased procedure does not use an exhaustive variable search routine in choosing splitting rules (Sut & Simsek, 2011). Furthermore, the QUEST model employs imputation instead of surrogate splitting to deal with missing values. QUEST can simply handle categorical and continuous factors (Lee & Lee, 2015).
Support Vector Machine (SVM)	SVM as a popular supervised learning algorithm enables the computer to learn how to analyze data and it has many advantages such as good empirical performance (compared with other algorithms such as artificial neural networks), easy training process, avoidance of local minima, relatively suitable mathematics for multi-dimensional data, and a tradeoff between complexity and error (Chen et al., 2017).

ML algorithm	Description
Random Forest (RF)	RF is a classification and regression technique which suggested by Breiman (2001) for the first time. Specifically, it is an ensemble of trees (it is based on assembling decision trees) which is constructed from a training dataset and internally validated to obtain a response variable by given predictor variables (Boulesteix et al., 2012). Two powerful advantages of ML techniques are used by RF, including random feature selection and bagging. For in bagging procedure, each tree is trained on a bootstrap sample of the training data and predictions are made by majority vote of trees and in the random selection procedures, a subspace of feature predictions split at each node during growing a tree (Jiang et al., 2007).
Maximum entropy (Maxent)	For the first time Maxent was suggested by Phillips et al. (2006) and designed for spatial distribution modeling and specifically ecological modeling. It uses maximum entropy principle to relate presence-only data to predictor/independent variables and response/dependent variables to estimate a spatial distribution. Concerning pattern in the ML algorithms, a presence-only feature forces Maxent to follow a solution in which it indirectly solves a discriminative problem through the Bays' rule (Chen et al., 2017; Urbani et al., 2015).
Boosted Regression Tree (BRT)	BRT is a tree-based model that belongs to the Gradient Boosting Modelling (GBM) family. It combines several regression trees models to learn and weigh them (by assigning individual weights to every sample point of the training dataset), to describe the relationship between the predictor and response variables. It uses several models to improve the performance of a single model (Hu et al., 2018; Littke et al., 2017; Wang et al., 2018).
Multivariate Adaptive Regression Spline (MARS)	MARS proposed by Friedman (1991) to establish relations between predictor and response variables. It can estimate general functions of multi-dimension arguments, also MARS as an adaptive modeling process is based on non-parametric and nonlinear statistics (Samui, 2013). There is no assumptions MARS modeling about the underlying functional relationships between predictor and response variables (Zhang & Goh 2016). MARS with the simple rule-based functions enable predictions and it makes complex relationship modeling between predictor and response variables (Leathwick et al., 2006).
Generalized Linear Model (GLM)	Non-linear relationship has been assumed by GLM between the predictor and response variables and it is defined by three components: a systematic component that relates a parameter to the predictor variables, a random component which specifies a distribution for predictor and response variables, and a link component that connects the systematic and random components (Guisan et al., 2002; Koubbi et al., 2011).

ML algorithm	Description
Generalized Additive Model (GAM)	A GAM is specified as a sum of smooth functions and it also provides estimation using a combination of the local scoring and the back-fitting algorithms (Dominici et al., 2002). It is best used for more than purely exploratory analysis, for which its smoothing parameter is a key component (Hastie & Tibshirani, 1990, p. 1–11). It uses a link function to establish a relationship between the mean of the response and a smoothed function of the predictor variables. The merrier of GAM lies in its ability to deal with highly non-linear and non-monotonic relationships between the response and predictor variables (Guisan et al., 2002).
J48 Decision Tree (J48DT)	J48 decision tree (J48DT) is a new classification tree-based technique that has been employed in only a few studies. Predictive ability of the algorithm has been assessed and validated in a previous study using statistical and receiver operating characteristic (ROC) curve methods (Pham et al., 2017). For binary classification variables (here flood or non-flood locations), J48DT creates a tree, which includes a root node and internal branch and leaf nodes. The root node consists of the input dataset, the internal branch nodes relate to the decision function, and the leaf nodes indicate the production of a specified contribution of the dataset. Compared with other decision tree models, J48DT is considered best in terms of classification accuracy, which is trained in two steps: i) building, and ii) pruning the classification tree (Zhao & Zhang, 2008).
Classification and regression tree (CART)	Classification and regression tree (CART) is a recursive partitioning algorithm that used to predict continuous dependent and categorical predictor conditioning factors (Zhao et al., 2016). CART, which was proposed for the first time by Breiman et al. (1984), has gained in popularity during recent years. However, because of overfitting and high sensitivity of CART to minor changes in all training input data, it is viewed as an unstable ML algorithm (Erdal & Karakurt, 2013).
Logistic model tree (LMT)	Logistic model tree (LMT), a ML algorithm suggested by Landwehr et al. (2005), is a hybrid classification model combining logistic regression (LR) with decision tree (DT) functions, which can boost the precision of prediction (Chen et al., 2017). In LMT, variable splitting is performed with the maximum information ratio obtained. For splitting categorical and numerical variables with c specific values, the node has c child nodes and two child nodes, respectively (Lee & Jun, 2018). For pruning the nodes of the tree, the LMT algorithm employs a Logit Boost function to set up the logistic regression method, and it uses cross-validation to discover multiple Logit Boost iterations to stop overfitting of all training input data (Arabameri et al., 2019).

ML algorithm	Description
Artificial Neural network (ANN)	ANN models are a set of meta-heuristic population-based optimization algorithms. Computational ANN models have been broadly employed in different scientific fields, for prediction purposes in most cases (Madsen et al., 2017; Zhang et al., 2020). The ANN model has strong self-learning, self-compatibility, fitness, error tolerance, and extension capabilities, inspired by biological nervous systems. It shows high performance in fitting multivariate analysis by better learning efficiency (Zhao et al., 2019). The purpose of ANN model optimization is to get an optimum solution (Shirwaikar et al., 2019). Within the data mining field, the ANN algorithm has been used to solve many practical problems. In this study, it was used for urban flood susceptibility mapping.
Grey wolf optimizer (GWO)	GWO is a novel population-based and heuristic optimization algorithm that forms a pyramid with the most influential wolf at the remaining wolves in descendant importance in lower situations (Bui et al., 2019). GWO was developed by Mirjalili et al. (2017) according to the social hunting nature of grey wolves. It mathematically mimics the social leadership hierarchy and hunting behavior based on the grey wolves working together to detention prey with a clear cooperation. The GWO is divided into three stages for predation: encircling, hunting, and attacking (Niu et al., 2019). GWO, with low number of parameters and ease of implementation, has faster evolutionary programming and faster convergence than the swarm intelligence algorithm (Daniel et al., 2017). Search agents (wolves) in GWO, as swarm-based algorithms, are arbitrarily situated in the d-dimensional space. After each iteration, the positioned and fitness values of wolves are updated and the optimal solution for the problem is the final location of the top wolf (Bui et al., 2019).

3.2.4 Developing ensemble model

Based on the accuracy assessment and performance of the standalone ML algorithms, the ensemble model was built from the best ML models. The process for building the ensemble model elaborated the following stages: i) selecting the best models based on the confusion matrix metrics results, ii) combining the selected ML algorithms using an R programming language to exploit all the advantages of the selected ML algorithms, and iii) assessing the ensemble model using the confusion matrix evaluation metrics and selecting the most important influencing factors.

3.2.5 Developing hybridized and optimized models

3.2.6 Standalone MultiLayer Perceptron Neural Network (MLPNN)

Standalone Multilayer Perceptron Neural Network (MLPNN), the best known of the artificial neural networks (as ML algorithms that mimic neural network functions in the human body (Aleotti & Chowdhury, 1999)), contains neurons at three levels: input layer, hidden layer, and output layer. The independent input factors were considered in the input layer that is driving factors (here influencing factors in urban flood hazard) affecting the dependent factor (urban flood occurrence) in the output layer. Hidden layer provides relationship between the input and output layers. The number of hidden layer neurons is essential to find the most accurate predicted target value by try and error procedures (Gong et al., 1996).

3.2.7 Multi-Boosting MultiLayer Perceptron Neural Network (MultiB-MLPNN)

Multi-Boosting (MultiB) as a technique for combining two AdaBoost and Wagging algorithms, is a robust algorithm proposed by Webb (2000). MultiB thereby inherits the advantages of the two original algorithms and can reduce the bias and variance of AdaBoost and wagging, respectively. Given an urban-flood concept in this study, five steps were considered as the working procedures of the MultiB algorithm including 1) Weighted-bootstrap sampling is performed on the urban-flood inventory data to create subsets. 2) Based on the learning procedures of the algorithms, each subset is adopted to build a classifier (here MLPNN). 3) In each subset, the weights of all samples are reset using the continuous Poisson distribution. 4) The sub-classifiers are re-trained using the updated subsets. 5) The final MultiB model is formed based on the trained sub-classifiers (Delen et al., 2005). In this study, the input layer was constructed from influencing urban flood hazard factors, the output layer was based on the flooded and non-flooded zones, and the hidden layer transformed the input layer to the output layer. The hybridized (MultiB-MLPNN) configuration was carried out using MLPNN as the base-classifier, with training iteration 500, the purlin transfer function and the logistic sigmoid activate function as suggested in Bui et al. (2016). The number of classifiers and neurons in the hidden layer was determined through trial-and-error, the results showed that 15 classifiers and one neuron in the hidden layer were best for the flood dataset. The training mechanism of MultiB for urban hazard flood

modeling were as following: 1) Fifteen subsets of data were generated from the training dataset using the weighted bootstrap sampling technique, and the weight of all samples was set to 1. 2) Each data subset was used to build a base classifier; thus 15 classifiers were created. 3) Based on the training error of the 15 classifiers, all samples in each subset were re-weighted using the continuous Poisson distribution. 4) The 15 classifiers were trained again using the updated subset weighting. 5) The MultiB-MLPNN model was derived by aggregating the final 15 classifiers.

3.2.8 Designing and proposing an artificial neural network model optimized by swarm intelligence and grey wolf algorithms (ANN-SGW)

The swarm intelligence algorithm is one of the major popular optimization models applied by researchers in different subfields (Yang, 2014). In recent years, the swarm intelligence algorithm has been used to imitate behavior in nature (Şenel et al., 2019), e.g., to understand crowd behavior in biological systems (Qasim & Bhatti, 2019). It is a family of nature-inspired modeling approaches based on the collective behavior of social swarms in nature (e.g., honeybees, ant colonies, birds, fish). However, the swarm intelligence algorithm seeks to identify an optimal result from the social actions of many individuals (Mosa et al., 2019), according to the premise that individuals' interactions lead to intelligent behaviors at the group level (Zedadra et al., 2018). It comprises three steps: i) Initialize the swarm of predictor variables as dimensional space, ii) set parameters of the swarm algorithm (e.g., maximum iteration and population); and iii) predict a robustness value and randomly determine the primary best situation (Bui et al., 2019).

Optimization of models is important to achieve an optimum solution in a complicated d-dimensional space. When problem-solving needs too much time, optimized approaches can help; however, an overall optimized algorithm is not guaranteed (Şenel et al., 2019). Optimization is also a crucial activity in model design, allowing programmers and planners to produce better designs, saving time and costs. Many optimization problems in engineering are complex and challenging to solve with conventional optimization models, such as mathematical programming. By hybridizing a model with different algorithms, it is possible to combine the advantages of the algorithms and produce a better-optimized model (Cui & Bai, 2019; Jia et al., 2019). In this study, an ANN model was optimized using novel optimizer algorithms (swarm intelligence and GWO) in a parallel

computing approach, to improve spatial prediction of urban inundation susceptibility. Compared with actual data, the optimized model (ANN-SGW) was used to recognize the most flood-prone areas in the study city. The proposed ANN-SGW model has a simple construction, with a low number of factors and simulations, avoiding the complexity of large-scale networks, and combines swarm intelligence and GWO with ANN's advantages.

3.3 Model training

In this study, our modeling approach used different algorithms to narrate response or dependent variables (here, urban flood maps) to predictor or independent variables (influencing urban flood factors). Each ML algorithms was run based on learning procedures using flooded points (flood inventory data) and independent variables. Each ML algorithm also trained a portion of the flood inventory data and was then verified on another portion flood inventory.

3.3.1 Accuracy assessment

The accuracy assessment of the ML algorithms, ensemble, optimized, and hybridized models were measured by evaluating the agreement between flood inventory data and models outputs (here urban flood hazard maps) in terms of both presence or flooded points and absence or non-flooded points (Paper I). Three different evaluation approaches were used for measuring the performance of the standalone ML algorithms, ensemble, optimized, and hybridized models in both training (goodness-of-fit) and validation (predictive performance) steps. Accuracy assessment was carried out using cutoff-dependent metrics and cutoff-independent metrics (Receiver Operating Characteristic-Area Under the Curve [ROC-AUC]), and Root Mean Square Error (RMSE). Models employed in the current study are presented in the Fig. 6.

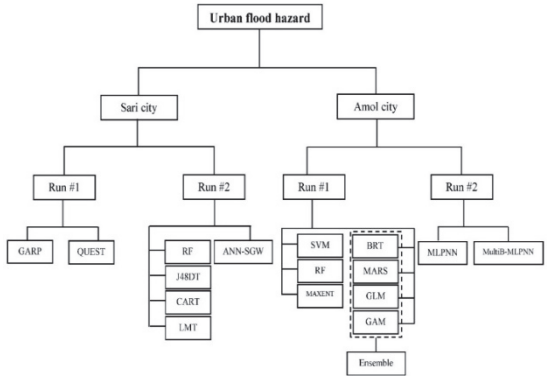

Fig. 6. Machine learning algorithms employed in the study.

3.3.2 Risk prediction

In this study, the urban flood risk map was produced through flood hazard and vulnerability maps (risk is a function of hazard and vulnerability) by using equation (Dewan, 2013)

$$\text{Risk} = \text{Hazard} \times \text{Vulnerability}, \tag{20}$$

where the flood hazard maps were determined from the following influencing factors: precipitation, LULC, CN, elevation, slope percent, groundwater level, distance to river and channel using the standalone, ensemble, hybridized and optimized ML algorithms and the flood vulnerability map was based on the following influencing factors: population density, building density, building age, quality of buildings, and socio-economic conditions.

4 Results and discussions

4.1 Flood hazard maps for Sari city

The first run of the machine learning algorithms for Sari city urban flood hazard maps was produced using the Genetic Algorithm Rule-Set Production (GARP) and Quick Unbiased Efficient Statistical Tree (QUEST) models. The flood hazard probability maps over the Sari city have been illustrated in Fig. 7. Both models (Fig. 7a and 7b) demonstrated that areas with high hazard probability are mainly located in the north and center of the study area and close to the along the Tajan River (riversides). As Fig. 7 shows, both algorithms were similar in terms of flood hazard prediction. They provided valuable information for flood hazard mapping, even though no detailed hydrological and hydraulic data were available.

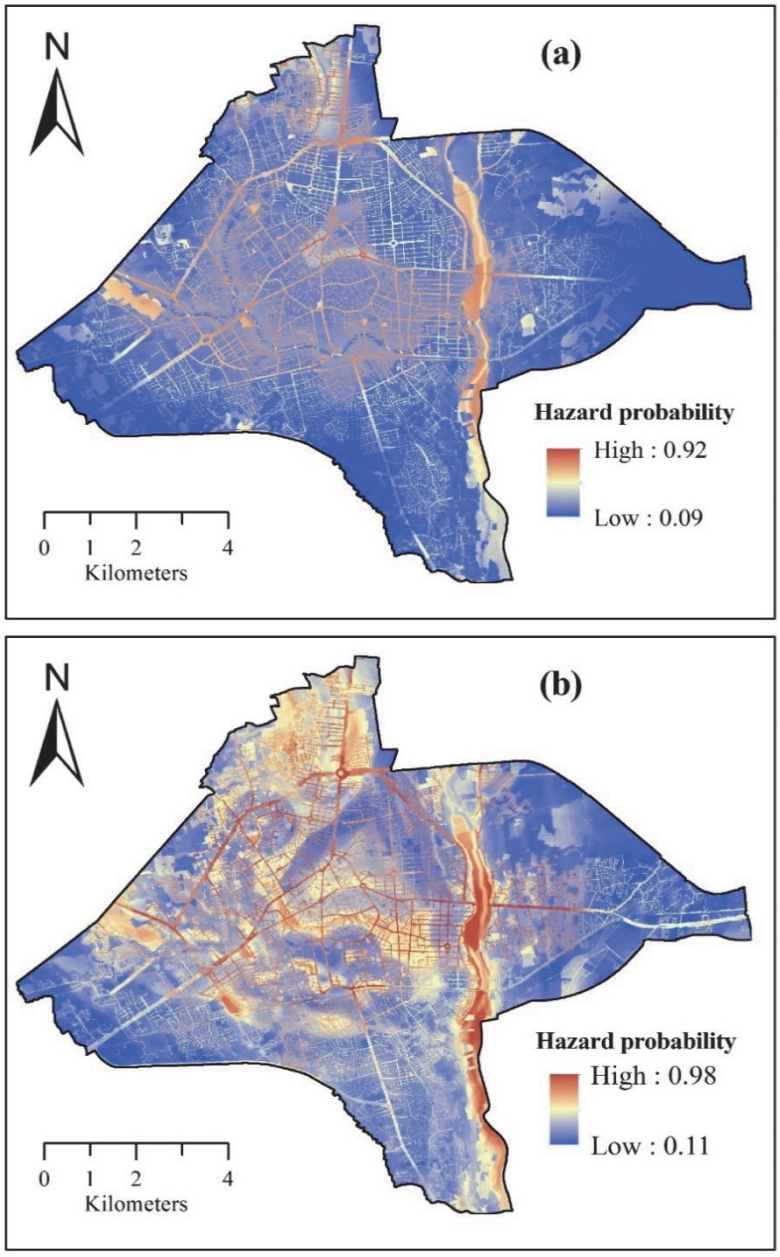

Fig. 7. Urban flood hazard probability maps using a) the GARP and b) the QUEST models for the Sari city (Reprinted, with permission, from Paper I © 2018 Elsevier B.V.).

The efficiency of the GARP and QUEST algorithms were measured using AUC-ROC and Kappa evaluation metrics. Based on the validation results GARP and QUEST algorithms reached 93.50% and 89.20% prediction rates, respectively. Also, the Kappa values were measured, which shows 0.86 for GARP and 0.79 for QUEST with excellent and good rates, respectively. Therefore, the efficiency of the GARP algorithm was somewhat better than the efficiency of the QUEST algorithm.

For the second round of the application of the ML algorithms of the Sari city, the urban flood hazard maps were produced by the random forest, J48 decision tree model, classification and regression trees (CART) and the logistic model tree LMT as benchmark models and new hybridized neural network-swarm intelligence-grey wolf (ANN-SGW) algorithm (Fig. 8). Upon preliminary review, distribution of high flood hazard zones seemed to be clearly distinguished over the Sari city. Especially, all the proposed benchmark and ANN-SGW algorithms revealed a comparatively similar pattern of flood susceptibility over the Sari city. Central, Northwestern, and areas quite close the Tajan River (riversides) parts of Sari city were indicated to be highly flood-susceptible areas, while eastern, southern, and southwestern parts showed low probability of flood hazard.

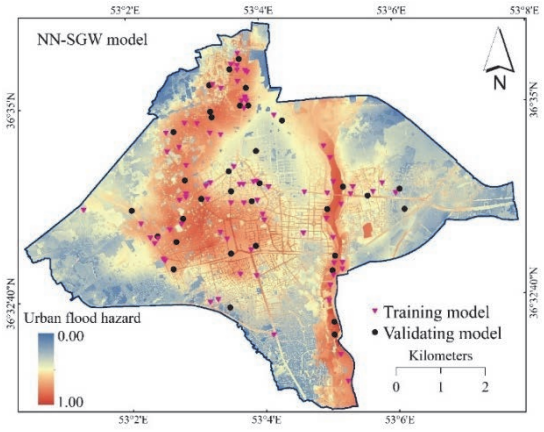

Fig. 8. Urban flood hazard maps generated by the ANN-SGW model for the Sari city (Adapted under CC BY 4.0 license from Paper IV © 2021 Authors).

4.1.1 Accuracy assessment of ANN-SGW model

The results for goodness-of-fit (training step) of the ANN-SGW model show that the value obtained for PPV, NPV, sensitivity, specificity, and efficiency (as cutoff-dependent evaluation metrics) was 93.7%, 87.3%, 88.1%, 93.2%, and 90.5%, respectively. The results of the confusion matrix metrics indicated significant arrangement between the trained model and observed data. AUC (as cutoff-independent metric) had a value of 96.3% in the training step, which shows high fitting skill of the hybridized ANN-SGW model. There was therefore agreement between the cutoff-dependent and -independent metrics as regards the performance of the model. However, goodness-of-fit individually estimates how well the model fits the training dataset and cannot be used to determine the model's ability in prediction. A testing step is needed to determine the predictive performance. In the predictive performance (validation step), the ANN-SGW model showed considerably maximum predictive performance based on both the cutoff-independent metrics (PPV=85.3%, NPV=73.5%, sensitivity=76.3%, specificity=83.3%, efficiency=79.4%) and the cutoff-independent metric (AUC=88.2%) and evaluation criteria (Table 7). The validation results demonstrated a robust arrangement between observed and predicted values by the ANN-SGW model. which can be classified as very good (80%<AUC<90%).

Table 7. Accuracy of the hybridized ANN-SGW model in the training and validation steps (Adapted under CC BY 4.0 license from Paper IV © 2021 Authors).

Evaluation approach	Evaluation metric	Training step or Goodness-of-fit	Validation step or predictive performance
Cutoff-dependent	True positive, TP	74	29
	True negative, TN	69	25
	False positive, FP	5	5
	False negative, FN	10	9
	Positive predictive value (PPV, %)	93.7	85.3
	Negative predictive value (NPV, %)	87.3	73.5
	Sensitivity (%)	88.1	76.3
	Specificity (%)	93.2	83.3
	Efficiency (%)	90.5	79.4
Cutoff-independent	Area under receiver operating curve (AUC, %)	96.3	88.2

4.1.2 Comparison of the ANN-SGW model with benchmark models

To confirm the performance of the ANN-SGW model, its predictive ability was compared with that of four commonly used benchmark models (RF, LMT, CART, J48DT). Among these, RF showed the highest accuracy in the training step in terms of both the cutoff-dependent (PPV=87.3%, NPV=72.2%, sensitivity=75.8%, specificity=85.1%, efficiency=79.7%) and cutoff-independent (AUC=89.4%) evaluation metrics. It was followed by LMT (PPV=84.8%, NPV=69.6%, sensitivity=73.6%, specificity=82.1%, efficiency=77.2%, AUC= 80%), and then J48DT (PPV=87.3%, NPV=63.3%, sensitivity=70.4%, specificity=83.3%, efficiency=75.3%, AUC=72.9%), and CART (PPV=77.2%, NPV=67.1%, sensitivity=70.1%, specificity=74.6%, efficiency=72.1%, AUC=72.2%). Therefore, agreement was obtained between observed values and values predicted by the benchmark models in the training step. The validation results show that RF outperformed the other benchmark models based on cutoff-dependent (PPV=94.12%, NPV=58.82%, sensitivity=69.57%, specificity=90.91%, efficiency=76.47%) and cutoff-independent (AUC=88.1%) evaluation metrics. It was followed by LMT (PPV=91.18%, NPV=55.88%, sensitivity=67.39%, specificity=86.39%, efficiency=73.53%, AUC=75.6%), and then CART (AUC= 69.8%) and J48DT (AUC= 69.7%). Therefore, the proposed ANN-SGW model had higher goodness-of-fit and predictive performance (AUC=96.3%; AUC=88.2%) than all benchmark models based on both cutoff-dependent and independent assessment metrics.

4.2 Flood hazard for Amol city

For the first run of the machine learning algorithms for the Amol city the seven models including Support Vector Regression (SVM), Random Forest (RF), Maximum Entropy (MAXENT), Boosted Regression Tree (BRT), Multivariate Adaptive Regression Spline (MARS), Generalized Linear Model (GLM) and Generalized Additive Model (GAM) have been applied to the Amol flood hazard mapping. All models demonstrated that zones with high hazard probability are mostly located in the north and center of Amol city and Haraz riverside which passes from Amol city (Fig. 9a–h). The areas with a high hazard probability were mainly identified by the GAM model (Fig. 9). Based on accuracy evaluation data, the ensemble model was built based on the BRT, MARS, GLM, and GAM

algorithms. The urban flood hazard map produced using the ensemble BRT, MARS, GLM, and GAM algorithms is shown in Fig. 9h.

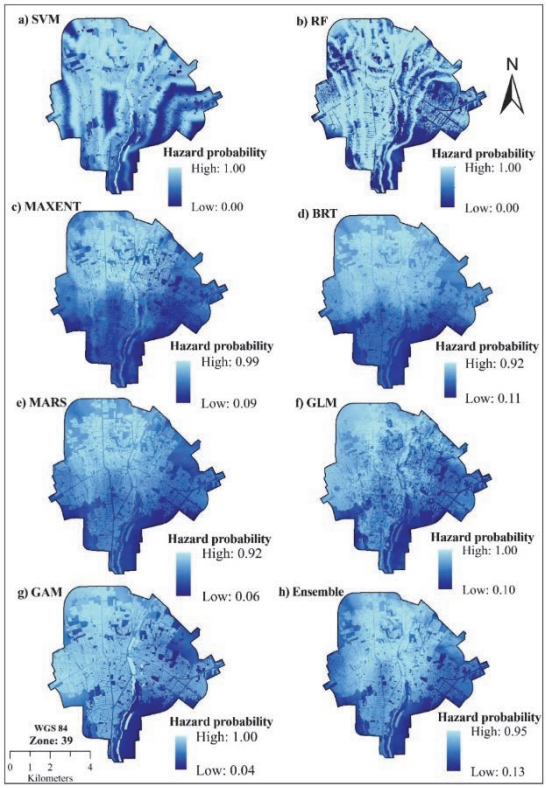

Fig. 9. Urban flood hazard probability maps using the individual and ensemble models for the Amol city (Adapted under CC BY 4.0 license from Paper II © 2020 Authors).

4.2.1 Accuracy assessment of the ensemble model

Accuracy assessment of the SVM, RF, MAXENT, BRT, MARS, GLM, and GAM models was carried out using cutoff-independent metric (ROC-AUC) in the training step (goodness-of-fit) and validation step (predictive performance). The ROC-AUC values during the training step were highest for the MARS (0.915), GAM (0.892), GLM (0.876), and BRT (0.838) models (Table 8). Also, the ROC-AUC values during the validation step were highest for the GAM (0.846), GLM (0.833), MARS (0.815), and BRT (0.824) models (Table 8). Based on the accuracy

assessment of the models in the validations step, the ensemble model was built from the BRT, MARS, GLM, and GAM algorithms. The accuracy of the ensemble model was 0.925 and 0.892 for the goodness-of-fit and predictive performance, respectively.

Table 8. Accuracy assessment of the ML algorithms and the ensemble model (Adapted under CC BY 4.0 license from Paper II © 2020 Authors).

ML algorithm	ROC-AUC	
	Training step or Goodness-of-fit	Validation step or predictive performance
SVM	0.818	0.774
RF	0.788	0.649
Maxent	0.806	0.764
BRT	0.838	0.824
MARS	0.915	0.815
GLM	0.876	0.833
GAM	0.892	0.846
Ensemble	0.925	0.892

While the employed ML algorithms may have contained some weaknesses, in this study, programming has tried to develop more sophisticated ensemble model to minimize the disadvantages of previous algorithms. The ensemble BRT, MARS, GLM, and GAM algorithms produced more fine output compared to standalone models. However, some little differences can be seen among their results regarding the flood hazard zones. Nevertheless, most employed algorithms generally had a good and excellent agreement between observed and predicted values over the Amol city.

For the second round of the application of the ML algorithms for the Amol city the urban flood hazard maps have been generated by MLPNN and MultiB-MLPNN models which demonstrated the same overall spatial pattern with high flood hazard probability in central and northern parts of Amol City (Figs. 10a and 10b). The differences between model outputs were generally the result of their structures. The mean flood susceptibility values (and standard deviations) produced by the models (Table 9) show that the MultiB-MLPNN flood risk map had the lower mean value (0.353), which means that the probability of flood occurrence in any cell in its flood hazard map was lower than its value in other models. Considering the greater accuracy of MultiB-MLPNN, this implies over-prediction by MLPNN, which yields higher probability of flood occurrence for any cell.

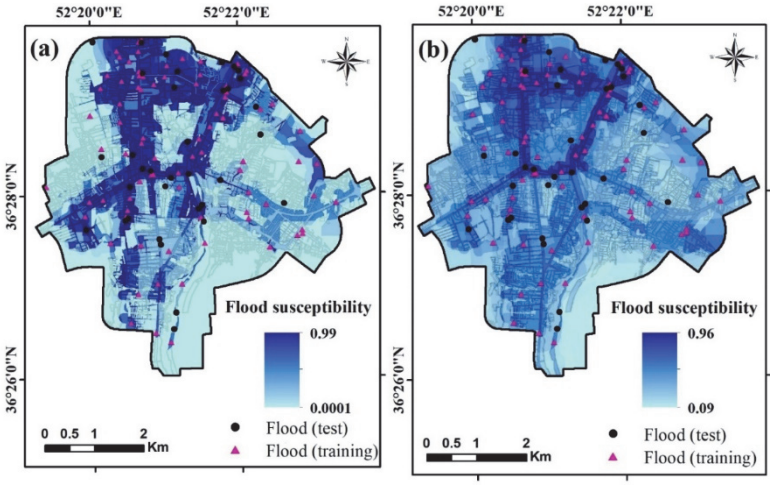

Fig. 10. Flood hazard maps using a) MultiB-MPLNN, and b) MLPNN for the Amol city (Reprinted under CC BY-NC-ND 4.0 license from Paper III © 2021 Authors).

Table 9. Flood hazard values produced by MultiB-MLPNN and MLPNN models (Adapted, with permission, from Paper III © 2021 Authors).

Model	Modeled susceptibility value	
	Mean	Standard deviation
MultiB-MLPNN	0.353	0.408
MLPNN	0.402	0.362

4.2.2 Accuracy assessment of the hybridized model

The goodness-of-fit of the models (i.e., model accuracy in the training system) based on the AUC cutoff-independent approach indicated that, in the training step, the RF model had the highest fitting skill (AUC=95%), followed by MultiB-MLPNN (87.8%), MLPNN (86.2%), and BRT (83.8%) (Fig. 11). According to all cutoff-dependent metrics, the RF model showed the best accuracy (E=0.81, TSS= 0.7, MCC=0.67, MR=0.34), followed by MultiB-MLPNN (E=0.78, TSS=0.58, MCC=0.58, MR=0.21), MLPNN (E=0.77, TSS=0.56, MCC=0.55, MR=0.22), and BRT (E=0.73, TSS=0.51, MCC=0.67, MR=0.26) (Table 10). Thus, the model ranking in the training step was similarly based on both cutoff-independent and cutoff-dependent metrics. However, goodness-of-fit only shows how well a model fits a training dataset, not the model's predictive ability (Paper I; Tehrany et al.,

2015). The prediction performance of the models was investigated using the validation dataset.

According to the validation analysis, MultiB-MLPNN had the best predictive performance based on both cutoff-independent (AUC= 84.7%) (Fig. 11) and cutoff-dependent (E=0.8, TSS=0.61, MCC=0.6, MR=0.2) evaluation criteria (Table 10). RF was the second-best model (AUC=82.1%, E=0.7, TSS+0.54, MCC=0.32, MR=0.28), followed by MLPNN (AUC=80.7%, E=0.74, TSS=0.48, MCC=0.48, MR=0.25) and BRT (AUC=76.1%, E=0.68, TSS=0.37, MCC=0.37, MR=0.31). Therefore, both cutoff-dependent and cutoff-independent metrics demonstrated that the hybridized MultiB-MLPNN model outperformed the other models in terms of prediction performance.

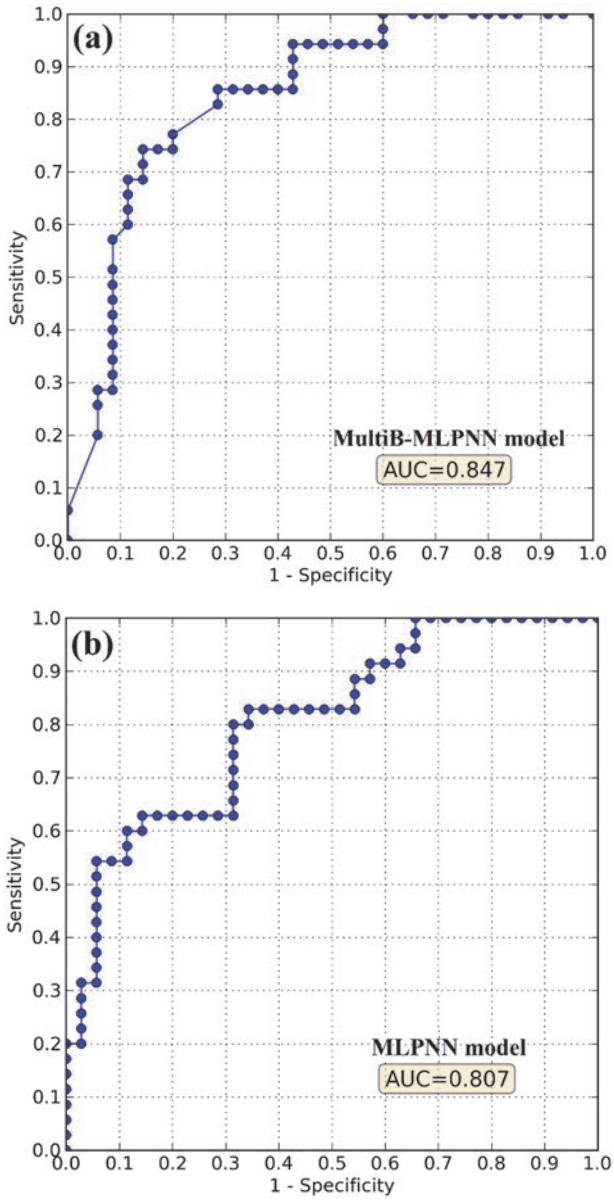

Fig. 11. AUC-ROC curves in the validation step for a) MultiB-MLPNN and b) MLPNN (Reprinted under CC BY-NC-ND 4.0 license from Paper III © 2021 Authors).

Table 10. Cutoff-dependent evaluation metrics of the MLPNN and MultiB-MLPNN (Adapted, with permission, from Paper III © 2021 Authors).

Metric	Goodness-of-fit		Predictive performance	
	MLPNN	MultiB-MLPNN	MLPNN	MultiB-MLPNN
Efficiency (E)	0.77	0.78	0.74	0.8
True skill statistic (TSS)	0.56	0.58	0.48	0.61
Matthews correlation coefficient (MCC)	0.55	0.58	0.48	0.6
Misclassification rate (MR)	0.22	0.24	0.25	0.2

Since MultiB-MLPNN exhibited the greater accuracy, its flood hazard map was re-classified into five classes using an equal-interval classification scheme: very low, low medium, high, and very high with 0–0.2, 0.21–0.4, 0.41–0.6, 0.61–0.8, and 0.81–1 flood probability respectively (Fig. 12). The MultiB-MLPNN assigned high and very high hazard-values to 12% and 14% of the study area, respectively (Table 11), and 54%, 9%, and 11% of the region was assigned very low, low, and medium hazard-values, respectively.

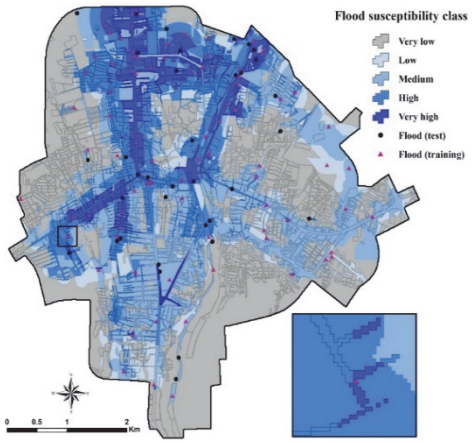

Fig. 12. Flood hazard classes generated by the hybridized MultiB-MLPNN model for the Amol city (Reprinted under CC BY-NC-ND 4.0 license from Paper III © 2021 Authors).

Table 11. Areas of flood hazard classes using the MultiB-MLPNN model over the Amol city (Adapted, with permission, from Paper III © 2021 Authors).

Number	Class	Area (%)
1	Very low	54
2	Low	9
3	Medium	11
4	High	12
5	Very high	14

4.3 Flood vulnerability map for Sari city

Urban flood vulnerability map was produced for Sari city based on the Fuzzy Analytical Network Process (FANP). Based on the expert knowledge using the FANP model, the weights assigned to each class of the urban density, quality of buildings, age of buildings, population density, and socio-economic factors are summarized in Table 12. Results on the relative importance of flood vulnerability influencing factors include population density, quality of buildings, and urban density as the most important factors with 0.370, 0.185, and 0.148 weights, respectively. Followed by the age of buildings and socio-economic conditions with 0.148 and 0.147 weights, respectively.

Table 12. Limit super-matrix of conditioning factors of urban flood vulnerability (Adapted, with permission, from Paper I © 2018 Elsevier B.V.).

Criteria	Class	Weights	Criteria	Class	Weights
Socio-economic conditions	A	0.0147	Quality of buildings	Very high	0.0271
	B	0.0234		High	0.0298
	C	0.0291		Medium	0.0321
	D	0.0350		Low	0.0347
	E	0.0392		Very low	0.0364
	Natural	0.0068		No building	0.0252
Population density	High	0.1343	Age of buildings	Very old	0.0470
	Medium	0.1089		old	0.0394
	Low	0.0741		Medium	0.0350
	Very low	0.0530		New	0.0224
Urban density	High	0.0482		Newest	0.0022
	Medium	0.0414		No building	0.0021
	Low	0.0321			
	Very low	0.0265			

Note: Population density: 0.370, Quality of buildings: 0.185, Urban density: 0.148, Age of buildings: 0.148 and Socio-economic conditions: 0.147.

The urban flood vulnerability value for each part of the city is shown in Fig. 13a. The results showed that the most vulnerable flooding zones are located in the center of Sari city. For better visual interpretation of urban flood vulnerability, the vulnerability map was classified into five classes (Fig. 13b): very low, low, moderate, high, and very high, occupying 35.49%, 15.25%, 13.52%, 22.61%, and 13.10% of the study area, respectively.

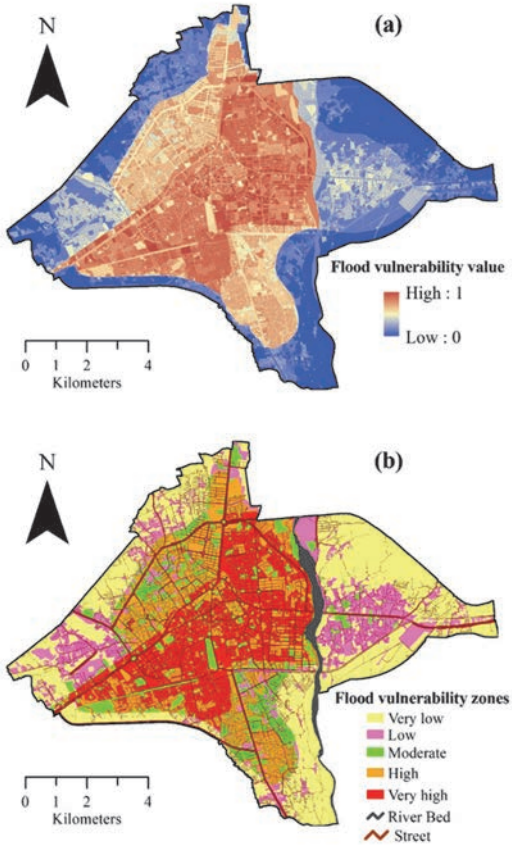

Fig. 13. Spatial distribution of a) flood vulnerability values and b) flood vulnerability zones for the Sari city (Reprinted, with permission, from Paper I © 2018 Elsevier B.V.).

4.4 Flood vulnerability map for Amol city

Using AHP and based on expert knowledge, weight and rank values were assigned to the flood vulnerability influencing factors and their classes according to the importance in the Amol study area. Results showed that the greatest weight of flood vulnerability influencing factors was population density with 0.38 weight, followed by building density, buildings age, and socio-economic conditions with 0.29, 0.19, and 0.14 weights respectively. The weights obtained for each class of the population density, building density, building age, and socio-economic conditions variables are shown in Table 13. In weighting factors and class ranking, total scores were

applied and then each pixel of the output map was assigned a value reflecting its factor and normalized weight. Fig. 13 shows urban flood vulnerability maps for Amol city as flood vulnerability value (Fig. 13a) and flood vulnerability zones (Fig. 13b).

Table 13. Weight and rank values assigned to the different classes of flood vulnerability factors (Adapted under CC BY 4.0 license from Paper II © 2020 Authors).

Factor	Weighting	Class	Ranking	Factor	Weighting	Class	Ranking
Building density	0.29	High	0.354	Population density	0.38	High	0.402
		Medium	0.283			Medium	0.281
		Low	0.202			Low	0.204
		Very low	0.152			Very low	0.111
		Open space	0.009			Open space	0.002
	Inconsistency ratio		0.017		Inconsistency ratio		0.014
Building age	0.19	Newest	0.049	Socio-economic conditions	0.14	Very good	0.002
		New	0.089			Good	0.041
		Moderate	0.234			Moderate	0.195
		Old	0.269			Poor	0.332
		Very Old	0.358			Very poor	0.429
		Open space	0.001			Open space	0.002
	Inconsistency ratio		0.019		Inconsistency ratio		0.021

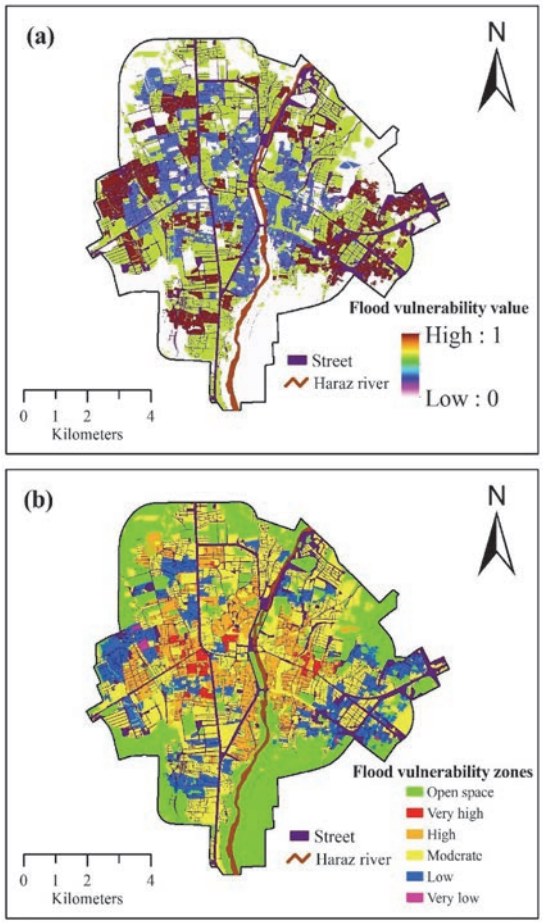

Fig. 14. Spatial distribution of a) flood vulnerability value and b) flood vulnerability zones for the Amol city (Reprinted under CC BY 4.0 license from Paper II © 2020 Authors).

4.5 Flood risk map for Sari city

In the flood risk index map, the risk value ranged from 0.05 to 0.76 (Fig. 15a). The flood risk map was classified into five classes using the natural break method: very low, low, moderate, high, and very high, occupying 32.90%, 21.71%, 21.98%, 7.70%, and 15.69% of the study area, respectively (Fig. 15b). The flood risk map indicated that central and northern sites in Sari city are most exposed to flood risk.

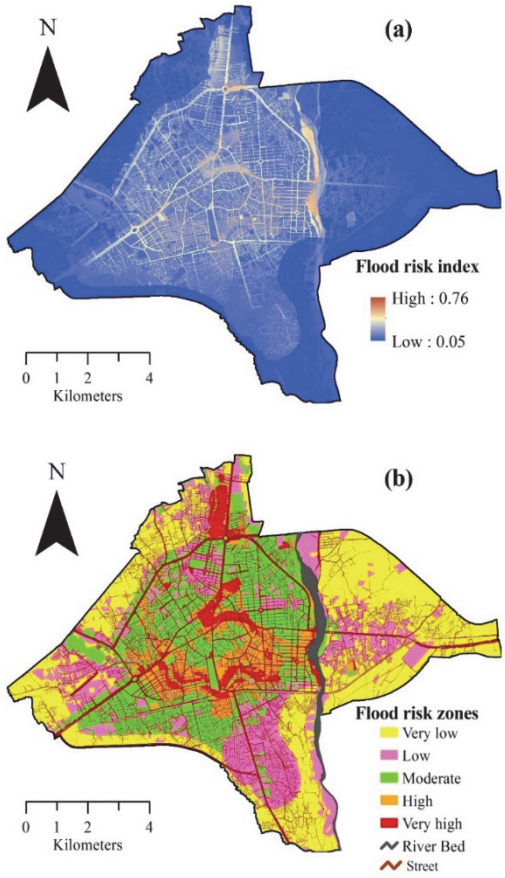

Fig. 15. Spatial distribution of a) flood risk index and b) flood risk zones for the Sari city (Adapted, with permission, from Paper I © 2018 Elsevier B.V.).

4.6 Flood risk map for Amol city

The spatial distribution of the ensemble model for the urban flood risk map is presented in Fig. 16a. Using the natural break method in ArcGIS 10.5, the risk index map was classified into very low (<0.20), low (0.21–0.40), moderate (0.41–0.60), high (0.61–0.80), and very high (>0.80), representing 3.40%, 16.41%, 14.17%, 18.58%, and 47.45% of total area, respectively (Fig. 16b). The risk index map confirmed that the northern and central areas of Amol city have the highest risk of

flooding. Accordingly, areas with a very high risk had the highest density of flood points, and areas with a very low risk had the lowest density (in a given area).

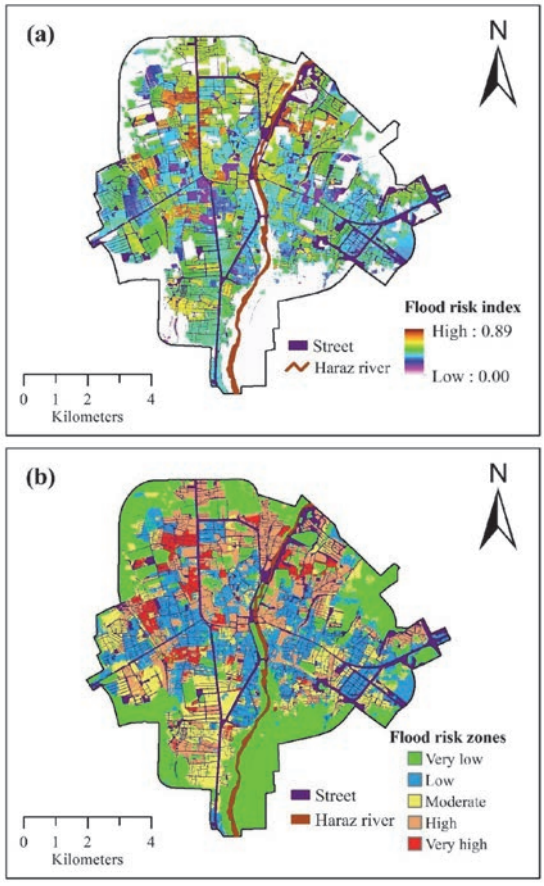

Fig. 16. Spatial distribution of a) flood risk values and b) flood risk zones based on the ensemble model for the Amol city (Adapted under CC BY 4.0 license from Paper II © 2020 Authors).

4.7 Contribution analysis of the variable importance

4.7.1 Variable importance over the Sari city

Since the ANN-SGW model showed the highest accuracy in both the training and testing dataset, importance of the variable has been calculated by the ANN-SGW model. Fig. 17 illustrates the relative importance of predictor variables in spatial modeling of urban flood hazard. The results indicated that distance to channel (1.00), land use (0.96), and elevation (0.89) were the most important factors, followed by curve number (0.83), distance to river (0.74), depth to groundwater (0.57), rainfall (0.38), and slope (0.22) (Fig. 17). Therefore, all influencing factor in urban flood hazard made a significant contribution in flood inundation modeling, and hence all were used as independent variables to generate the urban flood hazard map.

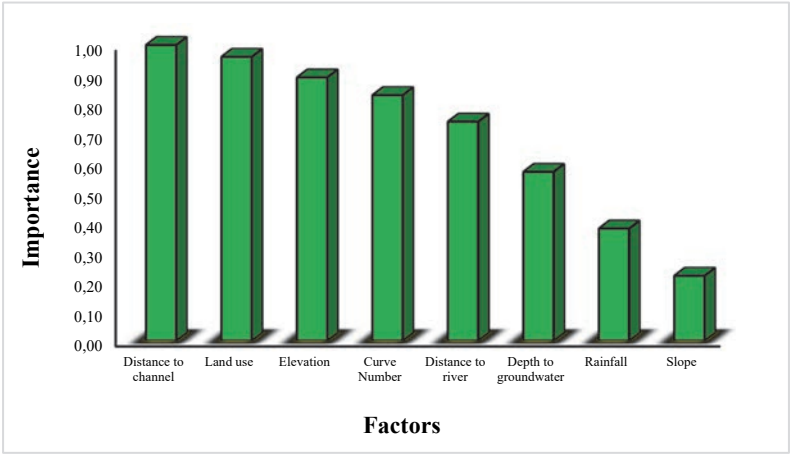

Fig. 17. Importance of flood hazard influencing factors in urban based on the GARP model for the Sari city (Data from Paper I).

4.7.2 Variable importance over the Amol city

Urban flood maps over the Amol city were constructed from the ML models for regions with a high and low hazard of urban flooding. Since the ensemble model showed the highest performance in both the training and testing dataset, the importance of the conditioning factors was determined based on ensemble

functions and the impact of the variables from the flooded points (flood inventory). Only the four most important factors were included in the ensemble model. These were: distance to channel (0.92), LCLU (0.88), CN (0.84), and distance to river (0.81) followed by elevation (0.68), depth to groundwater (0.048), rainfall (0.44) and slope (0.33) (Fig. 18).

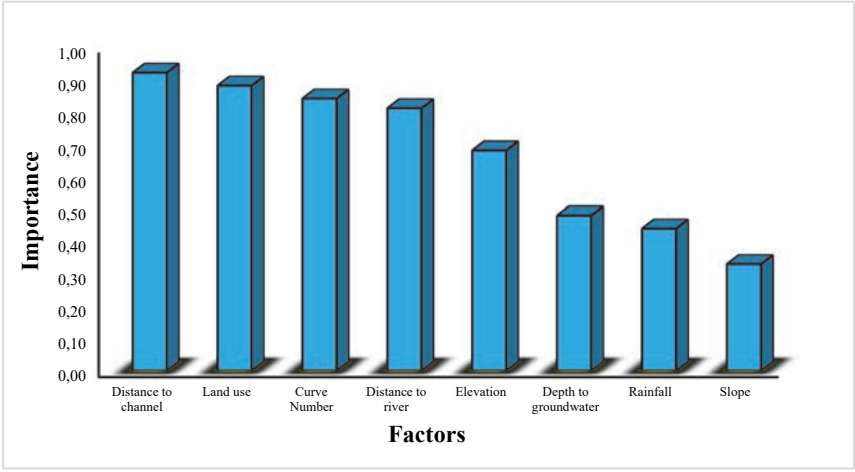

Fig. 18. Importance of the flood influencing factors based on the ensemble model for the Amol city (Data from Paper II).

5 Conclusions and directions for future studies

5.1 Conclusions

Floods are complex phenomena, which can have a detrimental impact on the environment, communities, buildings, and other infrastructure. Since accurate urban flood modeling has an effective role in surface runoff management and sustainable urban development, this study set out to use standalone ML algorithms to develop robust ensemble, optimized and hybridized models for spatial prediction of urban flood hazard, vulnerability, and risk mapping – using the cities of Sari and Amol in Mazandaran, Iran as case study sites. Urban floods can be influenced by different variables, which in developing countries are often related to unplanned urban development along lowlands and riverbanks, poor maintenance, and clogging of urban drains. The findings of this thesis can help decision-makers attendant investments out of dangerous zones such as Riverian zones, save property and lives, as well as guarantee that investments consider flood events in urban environments with high density population as well as multiple infrastructure (such as hospitals, schools, and service networks). The main contribution of this thesis is to provide a better understanding of the causes of urban floods by considering the importance of the different hydrological and physical characteristics of data-scarce urban environments. We have considered the essential role of urban flood conditioning factors by employing ML modelling for different datasets. To the best of our knowledge, the role of urban flood conditioning factors by employing ML has not been incorporated in previous literature. Therefore, the results of this thesis highlight the importance of urban flood conditioning factors in coping with the mapping of the flood risks over the cities. The more general outputs were:

- The results obtained from this research show how ML algorithms can be employed in urban flood mapping with some small-scale field surveys and without complex hydrodynamic models.
- Accurate predicting of flood inundation areas, where high population concentrations and critical infrastructures often make the economic and social impacts of a flood event very high.
- The proposed robust ensemble machine learning model, which has high prediction accuracy of flood inundation areas, is recommended for flood risk mapping in other regions.

The specific conclusions that can be summarized from this thesis are as follow:

- In developing countries, hydraulic data are typically not available in urban areas and data scarcity is a barrier to flood modeling. To overcome input data limitations, several ML algorithms were employed to assess flood risk. Thus, the methodology presented in this thesis is mainly useful in regions with little data, to quickly define and expose flooding conditions.
- Since ML algorithms application have shown promising/reliable results for flood hazard assessment, same framework can replace the current expensive field surveys or complex and sophisticated hydrodynamic modelling.
- Hazard and vulnerability mapping should serve as a first step in developing flood risk reduction strategies and allocating resources for flood defenses and warning systems across the urban environment.
- According to the validation results for Sari City, 96.3% and 93.50% prediction rates (AUC-ROC) were achieved for the ANN-SGW and GARP models, respectively. Therefore, the efficiency of these models was somewhat better than the others (QUEST, RF, J48DT, CART, and LMT).
- The accuracy of flood hazard mapping between the individual ML algorithms (SVM, RF, MAXENT, BRT, MARS, GLM and GAM) differed for Amol City. The MARS, GAM, GLM, and BRT models provided the most accurate results for both training and testing dataset. Therefore, these four individual algorithms were integrated into an ensemble model to derive a final flood hazard map, which was more accurate (AUC-ROC=92.5%) than the individual algorithms.
- Our analysis revealed that in flood hazard and vulnerability mapping, distance to channel and population density were the most? important influencing variables respectively. Therefore, proper design and maintenance of drainage systems is essential for sustainable urban management.
- Based on both evaluation metrics, the hybridized multilayer perceptron neural network (MLPNN) model using the multi-boosting (MultiB-MLPNN) technique outperformed the standalone MLPNN for flood hazard mapping across Amol City. Therefore, using MultiB-MLPNN is highly recommended for future studies, as the proposed methodological framework is transferable to other regions around the world.
- A novel optimized model (ANN-SGW) was developed and evaluated using statistical evaluation metrics and compared with several benchmark machine learning models (RF, LMT, CART, and J48DT Model). ANN-SGW was an

important scientific contribution to the development of a more powerful model for spatial prediction of flood hazard.

5.2 Recommendation and directions for future studies

Climate change may have long-term impacts on urban floods and urbanization (e.g. construction of new infrastructure, zoning or development strategies and adaptation plans). This has important implications for many developing countries, which are still characterized by rapid urban development and inadequate planning.

- In this study, several heavy rainfall events which resulted in devastating floods were considered in our analyses. The applied ML algorithms helped us to predict the probability of flood occurrence based on data from historical extreme events. We recognize that the modeling results in this thesis still come with some uncertainties due to dynamic influencing variables such as rainfall characteristics (e.g., rainfall amount, intensity, time, and frequency) and data limitations. Future studies should consider the dynamic conditions of influencing variables to produce more accurate results.
- In the future, the methodology and results obtained here could be extended with necessary precautions, to provide a detailed investigation of the impacts of urbanization (i.e., human activities) and climate change on urban floods through integrating tools on detailed GIS mapping and evaluation, hydrodynamic modeling risk-based flood mapping.
- Flood mapping across some city environments is challenging because of a lack of hydraulic and hydrological data, but accurate prediction of flood maps is necessary for urban planning and management. In this work, eight influencing factors were considered for flood inundation mapping. Further studies focusing on the role of other factors, such as rainfall intensity, are required.

List of references

Ahmad, S. S., & Simonovic, S. P. (2013). Spatial and temporal analysis of urban flood risk assessment. *Urban Water Journal*, *10*(1), 26–49.

Albano, R., Sole, A., Adamowski, J., Perrone, A., & Inam, A. (2018). Using FloodRisk GIS freeware for uncertainty analysis of direct economic flood damages in Italy. *International journal of applied earth observation and geoinformation*, *73*, 220–229.

Aleotti, P., & Chowdhury, R. (1999). Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the environment*, *58*(1), 21–44.

Amendola, A., Ermoliev, Y., Ermolieva, T. Y., Gitis, V., Koff, G., & Linnerooth-Bayer, J. (2000). A systems approach to modeling catastrophic risk and insurability. *Natural Hazards*, *21*(2), 381–393.

Anderson, R. P. (2003). Real vs. artefactual absences in species distributions: tests for Oryzomys albigularis (Rodentia: Muridae) in Venezuela. *Journal of Biogeography*, *30*(4), 591–605.

Arabameri, A., Chen, W., Loche, M., Zhao, X., Li, Y., Lombardo, L., & Bui, D. T. (2020). Comparison of machine learning models for gully erosion susceptibility mapping. *Geoscience Frontiers*, *11*(5), 1609–1620.

Arunakranthi, G., Rajkumar, B., Rao, V. C. S., & Harshavardhan, A. (in press). Advanced patterns of predictions and cavernous data analytics using quantum machine learning. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2020.11.062

Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education*, *7*(3), 112–118.

Barredo, J. I. (2007). Major flood disasters in Europe: 1950–2005. *Natural Hazards*, *42*(1), 125–148.

Blöschl, G., Nester, T., Komma, J., Parajka, J., & Perdigão, R. A. (2013). The June 2013 flood in the Upper Danube basin, and comparisons with the 2002, 1954 and 1899 floods. *Hydrology & Earth System Sciences Discussions*, *10*(7), 5197–5212.

Boeckmann, M., & Joyner, T. A. (2014). Old health risks in new places? An ecological niche model for I. ricinus tick distribution in Europe under a changing climate. *Health & place*, *30*, 70–77.

Booij, M. J. (2005). Impact of climate change on river flooding assessed with different spatial model resolutions. *Journal of hydrology*, *303*(1–4), 176–198.

Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(6), 493–507.

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

Bui, D. T., Ngo, P. T. T., Pham, T. D., Jaafari, A., Minh, N. Q., Hoa, P. V., & Samui, P. (2019). A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena*, *179*, 184–196.

Bui, D. T., Pradhan, B., Nampak, H., Bui, Q. T., Tran, Q. A., & Nguyen, Q. P. (2016). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology*, *540*, 317–330.

Büyüközkan, G., & Çifçi, G. (2012). A novel hybrid MCDM approach based on fuzzy DEMATEL, fuzzy ANP and fuzzy TOPSIS to evaluate green suppliers. *Expert Systems with Applications*, *39*(3), 3000–3011.

Carrera, L., Standardi, G., Bosello, F., & Mysiak, J. (2015). Assessing direct and indirect economic impacts of a flood event through the integration of spatial and computable general equilibrium modelling. *Environmental Modelling & Software*, *63*, 109–122.

Centre for Research on the Epidemiology of Disasters. (2020). EM-DAT: The OFDA/CRED International Disaster Database. Centre for Research on the Epidemiology of Disasters, Université catholique de Louvain, Brussels, Belgium. https://www.emdat.be/

Chen, B. J., & Chang, M. W. (2004). Load forecasting using support vector machines: A study on EUNITE competition 2001. *IEEE transactions on power systems*, *19*(4), 1821–1830.

Chen, J., Hill, A. A., & Urbano, L. D. (2009). A GIS-based model for urban flood inundation. *Journal of Hydrology*, *373*(1–2), 184–192.

Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, *8*(6), 866–883.

Chen, W., Pourghasemi, H. R., Kornejady, A., & Zhang, N. (2017). Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma*, *305*, 314–327.

Cherqui, F., Belmeziti, A., Granger, D., Sourdril, A., & Le Gauffre, P. (2015). Assessing urban potential flooding risk and identifying effective risk-reduction measures. *Science of the Total Environment*, *514*, 418–425.

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., & Mosavi, A. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*, *651*, 2087–2096.

Crisci, C., Ghattas, B., & Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, *240*, 113–122.

Cui, H., & Bai, J. (2019). A new hyperparameters optimization method for convolutional neural networks. *Pattern Recognition Letters*, *125*, 828–834.

Daniel, E., Anitha, J., & Gnanaraj, J. (2017). Optimum laplacian wavelet mask based medical image using hybrid cuckoo search–grey wolf optimization algorithm. *Knowledge-Based Systems*, *131*, 58–69.

Das, S., Hazra, S., Haque, A., Rahman, M., Nicholls, R. J., Ghosh, A., & de Campos, R. S. (2021). Social vulnerability to environmental hazards in the Ganges-Brahmaputra-Meghna delta, India and Bangladesh. *International Journal of Disaster Risk Reduction*, *53*, 101983.

Dayal, K. S., Deo, R. C., & Apan, A. A. (2018). Spatio-temporal drought risk mapping approach and its application in the drought-prone region of south-east Queensland, Australia. *Natural Hazards*, *93*(2), 823–847.

Delgado, J. M., Merz, B., & Apel, H. (2014). Projecting flood hazard under climate change: an alternative approach to model chains. *Natural Hazards and Earth System Sciences (NHESS)*, *14*(6), 1579–1589.

Demolli, H., Dokuz, A. S., Ecemis, A., & Gokcek, M. (2019). Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Conversion and Management*, *198*, 111823.

Dewan, A. (2013). *Floods in a megacity: geospatial techniques in assessing hazards, risk and vulnerability* (pp. 119–156). Springer.

Dimiduk, D. M., Holm, E. A., & Niezgoda, S. R. (2018). Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integrating Materials and Manufacturing Innovation*, *7*(3), 157–172.

Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American journal of epidemiology*, *156*(3), 193–203.

Dzyabura, D., & Yoganarasimhan, H. (2018). Machine learning and marketing. In *Handbook of Marketing Analytics*. Edward Elgar Publishing.

Eini, M., Kaboli, H. S., Rashidian, M., & Hedayat, H. (2020). Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *International Journal of Disaster Risk Reduction*, 101687.

Erdal, H. I., & Karakurt, O. (2013). Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. *Journal of Hydrology*, *477*, 119–128.

Fernández, D. S., & Lutz, M. A. (2010). Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Geology*, *111*(1–4), 90–98.

Few, R. (2003). Flooding, vulnerability and coping strategies: local responses to a global threat. *Progress in Development studies*, *3*(1), 43–58.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 1–67.

Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., & Blöschl, G. (2012). Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. *Water Resources Research*, *48*(4).

Gerl, T., Bochow, M., & Kreibich, H. (2014). Flood damage modeling on the basis of urban structure mapping using high-resolution remote sensing data. *Water*, *6*(8), 2367–2393.

Gong, P., Pu, R., & Chen, J. (1996). Elevation and forest-cover data using neural networks. *Photogrammetric Engineering & Remote Sensing*, *62*, 1249–1260.

Guisan, A., Edwards Jr, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*, *157*(2–3), 89–100.

Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P. L., Fragkias, M., Li, X., & Seto, K. C. (2017). Global scenarios of urban density and its impacts on building energy use through 2050. *Proceedings of the National Academy of Sciences*, *114*(34), 8945–8950.

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, *61*(4), 5–14.

Hand, D. J., & Adams, N. M. (2015). Data Mining. In *Wiley StatsRef: Statistics Reference Online* (pp. 1–7). https://doi.org/10.1002/9781118445112.stat06466.pub2

Hastie T. J., & Tibshirani R. J. (1990). *Monographs on Statistics and Applied Probability: Vol. 43. Generalized Additive Models* (1st ed.). Chapman and Hall/CRC.

Hinkel, J., Lincke, D., Vafeidis, A. T., Perrette, M., Nicholls, R. J., Tol, R. S., & Levermann, A. (2014). Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences*, *111*(9), 3292–3297.

Hu, Y., Scavia, D., & Kerkez, B. (2018). Are all data useful? Inferring causality to predict flows across sewer and drainage systems using directed information and boosted regression trees. *Water research*, *145*, 697–706.

Ierodiaconou, D., Monk, J., Rattray, A., Laurenson, L., & Versace, V. L. (2011). Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. *Continental Shelf Research*, *31*(2), S28–S38.

Ikeuchi, H., Hirabayashi, Y., Yamazaki, D., Muis, S., Ward, P. J., Winsemius, H. C., & Kanae, S. (2017). Compound simulation of fluvial floods and storm surges in a global coupled river-coast flood model: Model development and its application to 2007 C yclone S idr in B angladesh. *Journal of Advances in Modeling Earth Systems*, *9*(4), 1847–1862.

Irawan, A. M., Marfai, M. A., Nugraheni, I. R., Gustono, S. T., Rejeki, H. A., Widodo, A., & Faridatunnisa, M. (2021). Comparison between averaged and localised subsidence measurements for coastal floods projection in 2050 Semarang, Indonesia. *Urban Climate*, *35*, 100760.

Jalayer, F., De Risi, R., De Paola, F., Giugni, M., Manfredi, G., Gasparini, P., Topa, M. E., Yonas, N., Yeshitela, K., Nebebe, A., Cavan, G., Lindley, S., Printz, A., & Renner, F. (2014). Probabilistic GIS-based method for delineation of urban flooding risk hotspots. *Natural hazards*, *73*(2), 975–1001.

Jia, W., Qin, S., & Xue, X. (2019). A generalized neural network for distributed nonsmooth optimization with inequality constraint. *Neural Networks*, *119*, 46–56.

Jiang, P., Wu, H., Wang, W., Ma, W., Sun, X., & Lu, Z. (2007). MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Research*, *35*(suppl_2), W339–W344.

Karmakar, S., Simonovic, S. P., Peck, A., & Black, J. (2010). An information system for risk-vulnerability assessment to flood. *Journal of Geographic Information System*, *2*(03), 129.

Karthiban, K., & Raj, J. S. (2020). An efficient green computing fair resource allocation in cloud computing using modified deep reinforcement learning algorithm. *Soft Computing*, 1–10.

Kaspersen, P. S., & Halsnæs, K. (2017). Integrated climate change risk assessment: A practical application for urban flooding during extreme precipitation. *Climate services*, *6*, 55–64.

Khosronejad, A., Flora, K., Zhang, Z., & Kang, S. (2020). Large-eddy simulation of flash flood propagation and sediment transport in a dry-bed desert stream. *International Journal of Sediment Research*, *35*(6), 576–586.

Kjeldsen, T. R. (2010). Modelling the impact of urbanization on flood frequency relationships in the UK. *Hydrology Research*, *41*(5), 391–405.

Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Machine Learning*, *30*, 271–274. https://doi.org/10.1023/A:1017181826899

Koks, E. E., Bočkarjova, M., de Moel, H., & Aerts, J. C. (2015). Integrated direct and indirect flood risk modeling: development and sensitivity analysis. *Risk analysis*, *35*(5), 882–900.

Koks, E. E., Jongman, B., Husby, T. G., & Botzen, W. J. (2015). Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. *Environmental Science & Policy*, *47*, 42–52.

Kong, D., Lin, Z., Wang, Y., & Xiang, J. (2021). Natural disasters and analysts' earnings forecasts. *Journal of Corporate Finance*, *66*, 101860.

Koubbi, P., Moteki, M., Duhamel, G., Goarant, A., Hulley, P. A., O'driscoll, R., & Hosie, G. (2011). Ecoregionalization of myctophid fish in the Indian sector of the Southern Ocean: results from generalized dissimilarity models. *Deep Sea Research Part II: Topical Studies in Oceanography*, *58*(1–2), 170–180.

Kuang, D., & Liao, K. H. (2020). Learning from Floods: Linking flood experience and flood resilience. *Journal of Environmental Management*, *271*, 111025.

Kundzewicz, Z. W., Hirabayashi, Y., & Kanae, S. (2010). River floods in the changing climate-observations and projections. *Water Resources Management*, *24*(11), 2633–2646.

Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, *59*, 161–205.

Leathwick, J. R., Elith, J., & Hastie, T. (2006). Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological modelling*, *199*(2), 188–196.

Lee, S., & Jun, C. H. (2018). Fast incremental learning of logistic model tree using least angle regression. *Expert Systems with Applications*, *97*, 137–145.

Lee, S., & Park, I. (2013). Application of decision tree model for the ground subsidence hazard mapping near abandoned underground coal mines. *Journal of Environmental Management*, *127*, 166–176.

Leijnse, T., van Ormondt, M., Nederhoff, K., & van Dongeren, A. (2021). Modeling compound flooding in coastal systems using a computationally efficient reduced-physics solver: Including fluvial, pluvial, tidal, wind-and wave-driven processes. *Coastal Engineering*, *163*, 103796.

Littke, K. M., Cross, J., Harrison, R. B., Zabowski, D., & Turnblom, E. (2017). Understanding spatial and temporal Douglas-fir fertilizer response in the Pacific Northwest using boosted regression trees and linear discriminant analysis. *Forest Ecology and Management*, *406*, 61–71.

Lohi, S., & Tiwari, N. (in press). Preliminary study of embedding two-level reinforcement learning to enhance the functionality of setting objectives compared with machine learning. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.01.508

Lotfi, S., Shahabi Shahmiri, M., & Nikbakht, E. (2016). The feasibility study of applying creative multicenter network metropolitan approach in the metropolitan area of the Central Mazandaran. *Geography and Development Iranian Journal*, *14*(43), 1–18.

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. *International Journal of Research in Marketing*, *37*(3), 481–504.

Maddox, I. (2014, October 31). Three common types of flood explained. Intermap Technologies. https://www.intermap.com/risks-of-hazard-blog/three-common-types-of-flood-explained

Madsen, A. L., Jensen, F., Salmerón, A., Langseth, H., & Nielsen, T. D. (2017). A parallel algorithm for Bayesian network structure learning from large data sets. *Knowledge-Based Systems*, *117*, 46–55.

Mafarja, M. M., & Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, *260*, 302–312.

Marfai, M. A., & King, L. (2008). Coastal flood management in Semarang, Indonesia. *Environmental Geology*, *55*(7), 1507–1518.

Masood, M., & Takeuchi, K. (2012). Assessment of flood hazard, vulnerability and risk of mid-eastern Dhaka using DEM and 1D hydrodynamic model. *Natural Hazards*, *61*(2), 757–770.

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article "Assessment of economic flood damage". *Natural Hazards and Earth System Sciences (NHESS)*, *10*(8), 1697–1724.

Meyer, V., Haase, D., & Scheuer, S. (2009). A multicriteria flood risk assessment and mapping approach. *Flood risk management research and practice*, *4*, 1687–1694.

Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Cumulative hazard: The case of nuisance flooding. *Earth's Future*, *5*(2), 214–223.

Mohseni-Bandpei, A., & Yousefi, Z. (2013). Status of water quality parameters along Haraz river. *International Journal of Environmental Research*, *7*(4), 1029–1038.

Mosa, M. A., Anwar, A. S., & Hamouda, A. (2019). A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms. *Knowledge-Based Systems*, *163*, 518–532.

Naimi, B., & Araújo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, *39*(4), 368–375.

Nasiri, H., Yusof, M. J. M., & Ali, T. A. M. (2016). An overview to flood vulnerability assessment methods. *Sustainable Water Resources Management*, *2*(3), 331–336.

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons B*, *4*, 51–62.

Nied, M., Pardowitz, T., Nissen, K., Ulbrich, U., Hundecha, Y., & Merz, B. (2014). On the relationship between hydro-meteorological patterns and flood types. *Journal of Hydrology*, *519*, 3249–3262.

Niu, P., Niu, S., & Chang, L. (2019). The defect of the Grey Wolf optimization algorithm and its verification method. *Knowledge-Based Systems*, *171*, 37–43.

Noh, S. J., Lee, S., An, H., Kawaike, K., & Nakagawa, H. (2016). Ensemble urban flood simulation in comparison with laboratory-scale experiments: Impact of interaction models for manhole, sewer pipe, and surface flow. *Advances in Water Resources*, *97*, 25–37.

Ouma, Y. O., & Tateishi, R. (2014). Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment. *Water*, *6*(6), 1515–1545.

Parker, D. J., Green, C. H., & Thompson, P. M. (1987). *Urban flood protection benetits: A project appraisal guide*. Gower technical press.

Peterson, A. T., Ball, L. G., & Cohoon, K. P. (2002). Predicting distributions of Mexican birds using ecological niche modelling methods. *Ibis*, *144*(1), E27–E32.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259.

Pielke, R. (2019). Tracking progress on the economic costs of disasters under the indicators of the sustainable development goals. *Environmental Hazards*, *18*(1), 1–6.

Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational Intelligence: A logical approach*. Oxford University Press.

Qasim, T., & Bhatti, N. (2019). A hybrid swarm intelligence-based approach for abnormal event detection in crowded environments. *Pattern Recognition Letters*, *128*, 220–225.

Qin, Z., Zhang, J. E., DiTommaso, A., Wang, R. L., & Wu, R. S. (2015). Predicting invasions of Wedelia trilobata (L.) Hitchc. with Maxent and GARP models. *Journal of Plant Research*, *128*(5), 763–775.

Raju, B., & Bonagiri, R. (in press). A cavernous analytics using advanced machine learning for real world datasets in research implementations. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2020.11.089

Saharia, M., Kirstetter, P. E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping flash flood severity in the United States. *Journal of Hydrometeorology*, *18*(2), 397–411.

Sahin, S. (2012). An aridity index defined by precipitation and specific humidity. *Journal of Hydrology*, *444*, 199–208.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229.

Samui, P. (2013). Multivariate adaptive regression spline (Mars) for prediction of elastic modulus of jointed rock mass. *Geotechnical and Geological Engineering*, *31*(1), 249–253.

Schubert, J. E., & Sanders, B. F. (2012). Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency. *Advances in Water Resources*, *41*, 49–64.

Sedaghat, M., Solaimani, K., & Rashidpour, M. (2009). Assessment of flood susceptibility in Amol city using GIS technique. In *Proceedings of the 3th National Conference on Advanced Studies and Research in Geography, Architecture and Urban Science of Iran NICONF03_250* (p. 11).

Şenel, F. A., Gökçe, F., Yüksel, A. S., & Yiğit, T. (2019). A novel hybrid PSO–GWO algorithm for optimization problems. *Engineering with Computers*, *35*(4), 1359–1373.

Shirwaikar, R. D., Acharya, D., Makkithaya, K., Surulivelrajan, M., & Srivastava, S. (2019). Optimizing neural networks for medical data sets: A case study on neonatal apnea prediction. *Artificial Intelligence in Medicine*, *98*, 59–76.

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom* (pp. 1310–1315). IEEE.

Smith, K., & Ward, R. (1998). Mitigating and managing flood losses. In *Floods: Physical Processes and Human Impacts*. John Wiley & Sons.

Suriya, S., & Mudgal, B. V. (2012). Impact of urbanization on flooding: The Thirusoolam sub watershed–A case study. *Journal of Hydrology*, *412*, 210–219.

Sut, N., & Simsek, O. (2011). Comparison of regression tree data mining methods for prediction of mortality in head injury. *Expert Systems with Applications*, *38*(12), 15534–15539.

Tanaka, T., Kiyohara, K., & Tachikawa, Y. (2020). Comparison of fluvial and pluvial flood risk curves in urban cities derived from a large ensemble climate simulation dataset: A case study in Nagoya, Japan. *Journal of Hydrology*, *584*, 124706.

Tanim, A. H., & Goharian, E. (2020). Developing a hybrid modeling and multivariate analysis framework for storm surge and runoff interactions in urban coastal flooding. *Journal of Hydrology*, 125670.

Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, *125*, 91–101.

Termeh, S. V. R., Kornejady, A., Pourghasemi, H. R., & Keesstra, S. (2018). Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment*, *615*, 438–451.

Thieken, A. H., Müller, M., Kreibich, H., & Merz, B. (2005). Flood damage and influencing factors: New insights from the August 2002 flood in Germany. *Water resources research*, *41*(12). https://doi.org/10.1029/2005WR004177

Tingsanchali, T. (2012). Urban flood disaster management. *Procedia Engineering*, *32*, 25–37.

Trigo, R. M., Ramos, C., Pereira, S. S., Ramos, A. M., Zêzere, J. L., & Liberato, M. L. (2016). The deadliest storm of the 20th century striking Portugal: Flood impacts and atmospheric circulation. *Journal of Hydrology*, *541*, 597–610.

Ture, M., Tokatli, F., & Kurt, I. (2009). Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4. 5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications*, *36*(2), 2017–2026.

Turkington, T., Breinl, K., Ettema, J., Alkema, D., & Jetten, V. (2016). A new flood type classification method for use in climate change impact studies. *Weather and climate extremes*, *14*, 1–16.

United Nations Development Programme. (2004). *Reducing Disaster Risk: A Challenge for Development-a Global Report*. United Nations. https://digitallibrary.un.org/record/515746/files/Reducing.PDF

Urbani, F., D'Alessandro, P., Frasca, R., & Biondi, M. (2015). Maximum entropy modeling of geographic distributions of the flea beetle species endemic in Italy (Coleoptera: Chrysomelidae: Galerucinae: Alticini). *Zoologischer Anzeiger-A Journal of Comparative Zoology*, *258*, 99–109.

Wang, S., Adhikari, K., Wang, Q., Jin, X., & Li, H. (2018). Role of environmental variables in the spatial distribution of soil carbon (C), nitrogen (N), and C: N ratio from the northeastern coastal agroecosystems in China. *Ecological Indicators*, *84*, 263–272.

Wang, X. S., Ryoo, J. H. J., Bendle, N., & Kopalle, P. K. (2020). The role of machine learning analytics and metrics in retailing research. *Journal of Retailing*, *97*(4), 658–675. https://doi.org/10.1016/j.jretai.2020.12.001

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, *527*, 1130–1141.

Wang, Z., Xia, J., Zhou, M., Deng, S., & Li, T. (2020). Modelling hyperconcentrated floods in the Middle Yellow River using an improved river network model. *CATENA*, *190*, 104544.

Watson, L. M. (2020). Using unsupervised machine learning to identify changes in eruptive behavior at Mount Etna, Italy. *Journal of Volcanology and Geothermal Research*, *405*, 107042.

Webb, G. I. (2000). Multiboosting: A technique for combining boosting and wagging. *Machine learning*, *40*(2), 159–196.

Witting, A. B., Bagley, L. A., Nelson, K. F., & Lindsay, T. (2020). Natural Disasters and the Relational Study of the Family: A 2-Decade Scoping Review. *International Journal of Disaster Risk Reduction*, 101990.

Wright, J. M. (Ed.) (2007). Types of Floods and Floodplains. In *Floodplain Management, Principles and Current Practices* [Course material]. https://training.fema.gov/hiedu/docs/fmc/chapter%202%20-%20types%20of%20floods%20and%20floodplains.pdf

Wu, H., Adler, R. F., Hong, Y., Tian, Y., & Policelli, F. (2012). Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *Journal of Hydrometeorology*, *13*(4), 1268–1284.

Wu, W. W., & Lee, Y. T. (2007). Developing global managers' competencies using the fuzzy DEMATEL method. *Expert Systems with Applications*, *32*(2), 499–507.

Wu, W., McInnes, K., O'grady, J., Hoeke, R., Leonard, M., & Westra, S. (2018). Mapping dependence between extreme rainfall and storm surge. *Journal of Geophysical Research: Oceans*, *123*(4), 2461–2474.

Xu, J., Wang, Z., Shen, F., Ouyang, C., & Tu, Y. (2016). Natural disasters and social conflict: a systematic literature review. *International Journal of Disaster Risk Reduction*, *17*, 38–48.

Xu, L., Cai, F., Hu, Y., Lin, Z., & Liu, Q. (2021). Using deep learning algorithms to perform accurate spectral classification. *Optik*, *231*, 166423.

Yang, Q., Zhang, S., Dai, Q., & Yao, R. (2020). Assessment of Community Vulnerability to Different Types of Urban Floods: A Case for Lishui City, China. *Sustainability*, *12*(19), 7865.

Yang, X. S. (2014). Swarm intelligence-based algorithms: a critical analysis. *Evolutionary Intelligence*, *7*(1), 17–28.

Yin, H., & Li, C. (2001). Human impact on floods and flood disasters on the Yangtze River. *Geomorphology*, *41*(2–3), 105–109.

Yin, J., Yu, D., Yin, Z., Liu, M., & He, Q. (2016). Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China. *Journal of Hydrology*, *537*, 138–145.

Yousuf, T., Nakhle, A., Rawal, H., Harrison, D., Maini, R., & Irimpen, A. (2020). Natural disasters and acute myocardial infarction. *Progress in Cardiovascular Diseases*, *63*(4), 510–517.

Zali, N., Ganji, R., & Hoseini, H. (2017). Population Balance Planning for Metropolitan Area Network (MAN) in North of Iran in the 1400 Horizon. *Journal of the Geographical Engineering of Territory*, *1*(1), 54–71.

Zedadra, O., Guerrieri, A., Jouandeau, N., Spezzano, G., Seridi, H., & Fortino, G. (2018). Swarm intelligence-based algorithms within IoT-based systems: A review. *Journal of Parallel and Distributed Computing*, *122*, 173–187.

Zhai, X., Zhang, Y., Zhang, Y., Guo, L., & Liu, R. (2020). Simulating flash flood hydrographs and behavior metrics across China: Implications for flash flood management. *Science of The Total Environment*, 142977.

Zhan, X., & Huang, M. L. (2004). ArcCN-Runoff: an ArcGIS tool for generating curve number and runoff maps. *Environmental modelling & software*, *19*(10), 875–879.

Zhang, W., & Goh, A. T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers*, *7*(1), 45–52.

Zhang, Y., Jin, Z., & Chen, Y. (2020). Hybrid teaching–learning-based optimization and neural network algorithm for engineering design optimization problems. *Knowledge-Based Systems*, *187*, 104836.

Zhao, B., Ren, Y., Gao, D., Xu, L., & Zhang, Y. (2019). Energy utilization efficiency evaluation model of refining unit Based on Contourlet neural network optimized by improved grey optimization algorithm. *Energy*, *185*, 1032–1044.

Zhao, G., Pang, B., Xu, Z., Yue, J., & Tu, T. (2018). Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of the Total Environment*, *615*, 1133–1142.

Zhao, Y., & Zhang, Y. (2008). Comparison of decision tree methods for finding active objects. *Advances in Space Research*, *41*(12), 1955–1959.

Zhao, Y., Li, Y., Zhang, L., & Wang, Q. (2016). Groundwater level prediction of landslide based on classification and regression tree. *Geodesy and Geodynamics*, *7*(5), 348–355.

Original publications

I Darabi, H., Choubin, B., Rahmati, O., Haghighi, A. T., Pradhan, B., & Kløve, B. (2019). Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *Journal of Hydrology*, *569*, 142–154. https://doi.org/10.1016/j.jhydrol.2018.12.002

II Darabi, H., Haghighi, A. T., Mohamadi, M. A., Rashidpour, M., Ziegler, A. D., Hekmatzadeh, A. A., & Kløve, B. (2020). Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. *Hydrology Research*, *51*(1), 127–142. https://doi.org/10.2166/nh.2019.090

III Darabi, H., Rahmati, O., Naghibi, S. A., Mohammadi, F., Ahmadisharaf, E., Kalantari, Z., Haghighi, A. T., Soleimanpour, A. M., Tiefenbacher, J. P., & Tien Bui, D. (2021). Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood. *Geocarto International*, 1–27. https://doi.org/10.1080/10106049.2021.1920629

IV Darabi, H., Haghighi, A. T., Rahmati, O., Shahrood, A. J., Rouzbeh, S., Pradhan, B., & Tien Bui, D. (2021). A hybridized model based on neural network and swarm intelligence-grey wolf algorithm for spatial prediction of urban flood-inundation. *Journal of Hydrology*, *603*, 126854. https://doi.org/10.1016/j.jhydrol.2021.126854

Original publications are not included in the electronic version of the dissertation.

[1] https://creativecommons.org/licenses/by/4.0/
[2] https://creativecommons.org/licenses/by-nc-nd/4.0/

802. Nellattukuzhi Sreenivasan, Harisankar (2021) Synthesis and alkali activation of Magnesium-rich aluminosilicates

803. Kallio, Johanna (2021) Unobtrusive stress assessment in knowledge work using real-life environmental sensor data

804. Meriö, Leo-Juhani (2021) Observations and analysis of snow cover and runoff in boreal catchments

805. Lovén, Lauri (2021) Spatial dependency in edge-native artificial intelligence

806. Törmänen, Matti (2021) Improved analysis of tube flow fractionation data for measurements in the pulp and paper industry

807. Rusanen, Annu (2021) Catalytic conversion of sawdust-based sugars into 5-hydroxymethylfurfural and furfural

808. Yaraghi, Navid (2021) Analyzing human impacts on the quality and quantity of river water

809. Tarakanchikova, Yana (2021) Multilayered polyelectrolyte assemblies as delivery system for biomedical applications

810. Shaheen, Rana Azhar (2021) Design aspects of millimeter wave multiband front-ends

811. Kodukula, Suresh (2021) Ridging in stabilized ferritic stainless steels : the effects of casting and hot-rolling parameters

812. Hietaharju, Petri (2021) Predictive optimization of heat demand utilizing heat storage capacity of buildings

813. Hultgren, Matias (2021) Control design for CFB boilers integrated with process design

814. Safarpour, Mehdi (2021) Energy efficient solutions for computing and sensing

815. Kupila, Riikka (2021) Catalytic conversion of furfural and glucose over activated carbon-supported metal catalysts

816. Okwuibe, Jude (2021) Software-defined resource management for industrial internet of things

817. Luttinen, Esko (2022) Spectrum as a resource, spectrum as an asset: a techno-economic study

Book orders:
Virtual book store
http://verkkokauppa.juvenesprint.fi

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

UNIVERSITY OF OULU

ISBN 978-952-62-3195-2 (Paperback)
ISBN 978-952-62-3196-9 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)