(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0075350
(43) 공개일자 2025년05월28일

(51) 국제특허분류(Int. Cl.)
G06Q 50/26 (2024.01) *G06N 20/20* (2019.01)
G06T 11/00 (2006.01) *G06T 11/40* (2006.01)
G06T 7/55 (2017.01)
(52) CPC특허분류
G06Q 50/26 (2024.01)
G06N 20/20 (2021.08)
(21) 출원번호 10-2023-0162720
(22) 출원일자 2023년11월21일
심사청구일자 2023년11월21일

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
최수미
서울특별시 광진구 능동로 209, 대양AI센터 720호(군자동)
아볼가셈
서울특별시 광진구 능동로 209, 대양AI센터 433호(군자동)
(뒷면에 계속)
(74) 대리인
양성보

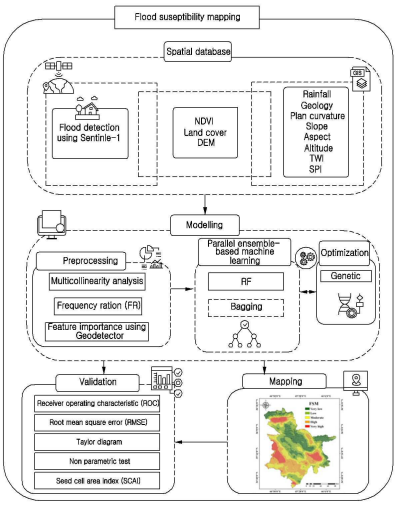
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **레이더 위성 영상을 이용한 홍수 민감도 매핑을 위한 머신러닝 알고리즘을 최적화하는 방법 및 장치**

(57) 요 약

레이더 위성 영상을 이용한 홍수 민감도 매핑을 위한 머신러닝 알고리즘의 최적화 기술이 개시된다. 홍수 민감도 매핑 방법은, 대상 지역의 홍수 상태 인자(flood conditioning factor)에 해당되는 데이터를 이용하여 병렬 앙상블 기반 머신러닝(parallel ensemble-based machine learning) 알고리즘을 통해 상기 대상 지역의 FSM(flood susceptibility map)을 생성하는 단계를 포함할 수 있다.

대 표 도 - 도2



(52) CPC특허분류
G06T 11/001 (2024.01)
G06T 11/40 (2013.01)
G06T 7/55 (2017.01)
G06T 2207/10032 (2013.01)
G06T 2207/20081 (2013.01)
(72) 발명자
바히드 라자비 테르메
서울특별시 광진구 능동로 209, 광개토관 1001호(군자동)

서명배
서울특별시 광진구 능동로 209, 광개토관 1001호(군자동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179366
과제번호	00156354
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	대학ICT연구센터지원사업
연구과제명	실-가상 연계 메타버스를 위한 초실감 XR 기술 연구
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1415183506
과제번호	P0016038
부처명	산업통상자원부
과제관리(전문)기관명	한국산업기술진흥원
연구사업명	산업기술국제협력(R&D)
연구과제명	가상·증강현실을 위한 지능형 콘텐츠 제작도구 개발
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2022.12.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

적어도 하나의 프로세서를 포함하는 컴퓨터 장치의 홍수 민감도 매핑 방법에 있어서,

상기 적어도 하나의 프로세서에 의해, 대상 지역의 홍수 상태 인자(flood conditioning factor)에 해당되는 데이터를 이용하여 병렬 앙상블 기반 머신러닝(parallel ensemble-based machine learning) 알고리즘을 통해 상기 대상 지역의 FSM(flood susceptibility map)을 생성하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 2

제1항에 있어서,

상기 생성하는 단계는,

상기 대상 지역의 레이더 위성 영상(radar satellite imagery)을 이용한 홍수 탐지(flood detection)를 통해 홍수 인벤토리 맵(flood inventory map)을 작성하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 3

제2항에 있어서,

상기 작성하는 단계는,

홍수 전 레이더 위성 영상과 홍수 후 레이더 위성 영상을 비교하여 홍수를 탐지하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 4

제2항에 있어서,

상기 작성하는 단계는,

궤도 파일 적용(orbital file application), 저감도 노이즈 감소(low-intensity noise reduction), 열 노이즈 제거(thermal noise removal), 방사선 측정 보정(radiometric calibration), 지형 보정(terrain correction), 및 후방 산란 계수(backscatter coefficient) 중 적어도 하나를 사용하여 상기 레이더 위성 영상을 전처리하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 5

제2항에 있어서,

상기 작성하는 단계는,

후방 산란 히스토그램(backscatter histogram)과 시행착오(trial and error)를 기반으로 선정된 임계값을 기준으로 상기 레이더 위성 영상에서 홍수 지역을 분류하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 6

제1항에 있어서,

상기 홍수 상태 인자는 토지 피복(land cover), 강우량(rainfall), 고도(altitude), 강까지의 거리(distance), TWI(topographic wetness index), 경사면(slope aspect), NDVI(normalized difference vegetation index), 지질(geology), 경사(slope), SPI(stream power index), 및 평면 곡률(plan curvature)을 포함하는 것

을 특징으로 하는 홍수 민감도 매핑 방법.

청구항 7

제1항에 있어서,

상기 생성하는 단계는,

과거의 홍수와 홍수를 야기하는 인자 간의 관계를 파악하기 위한 FR(Frequency ratio), 홍수에 영향을 미치는 매개변수 간의 관계를 파악하기 위한 다중공선성 분석(multicollinearity analysis), 및 변수의 공간적 층화 이질성(spatial stratified heterogeneity)을 기초로 이벤트 유발 인자(event's triggering factor)를 파악하기 위한 지리탐지기(Geodetector) 중 적어도 하나의 방법을 이용하여 상기 데이터를 전처리하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 8

제1항에 있어서,

상기 생성하는 단계는,

RF(Random forest) 알고리즘과 배깅(Bagging) 알고리즘의 최적화를 통해 상기 FSM을 모델링하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 9

제8항에 있어서,

상기 모델링하는 단계는,

유전 메타휴리스틱 알고리즘(genetic metaheuristic algorithm)을 이용하여 상기 RF 알고리즘과 상기 배깅 알고리즘의 하이퍼파라미터(hyperparameter)에 대한 최적 값을 선정하는 단계

를 포함하는 홍수 민감도 매핑 방법.

청구항 10

제1항에 있어서,

상기 홍수 민감도 매핑 방법은,

상기 적어도 하나의 프로세서에 의해, RMSE(root mean square error), ROC(receiver-operator characteristic) 곡선, SCAI(seed cell area index), 비모수 테스트(non-parametric test), 및 테일러 다이어그램(Taylor diagram) 중 적어도 하나의 지표를 이용하여 상기 FSM을 검증하는 단계

를 더 포함하는 홍수 민감도 매핑 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항의 홍수 민감도 매핑 방법을 상기 컴퓨터 장치에 실행시키기 위해 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

청구항 12

컴퓨터 장치에서 판독 가능한 명령을 실행하도록 구현되는 적어도 하나의 프로세서

를 포함하고,

상기 적어도 하나의 프로세서는,

대상 지역의 홍수 상태 인자(flood conditioning factor)에 해당되는 데이터를 이용하여 병렬 앙상블 기반 머신 러닝(parallel ensemble-based machine learning) 알고리즘을 통해 상기 대상 지역의 FSM(flood susceptibility map)을 생성하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 13

제12항에 있어서,

상기 적어도 하나의 프로세서는,

상기 대상 지역의 레이더 위성 영상(radar satellite imagery)을 이용한 홍수 탐지(flood detection)를 통해 홍수 인벤토리 맵(flood inventory map)을 작성하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 14

제12항에 있어서,

상기 적어도 하나의 프로세서는,

RF(Random forest) 알고리즘과 배깅(Bagging) 알고리즘의 최적화를 통해 상기 FSM을 모델링하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 15

제14항에 있어서,

상기 적어도 하나의 프로세서는,

유전 메타휴리스틱 알고리즘(genetic metaheuristic algorithm)을 이용하여 상기 RF 알고리즘과 상기 배깅 알고리즘의 하이퍼파라미터(hyperparameter)에 대한 최적 값을 선정하는 것

을 특징으로 하는 컴퓨터 장치.

발명의 설명

기 술 분 야

[0001] 아래의 설명은 홍수 관리를 위한 FSM(flood susceptibility map)을 생성하는 기술에 관한 것이다.

배 경 기 술

[0002] 지난 수십 년간 기후변화에 대한 증거는 광범위하게 존재해 왔으며, 이러한 변화는 홍수와 같은 기상 및 기후관련 현상들이 발생함에 따라 더욱 분명해지고 있다. 매년 전 세계적으로 1억 7천만 명이 홍수에 시달리고 있으며, 전체 자연재해 관련 사망자의 60%를 초과하고 있다. 부적절한 토지이용, 삼림벌채, 인구증가, 하천에 대한 인간의 과도한 개입 등으로 인한 환경 악화로 이러한 사건들이 증가하고 있다.

[0003] 이라크는 지형이 다양하여 다양한 자연재해에 취약하다. 홍수는 이라크의 생명과 재산에 가장 큰 위협이 되는 자연재해이다. 최근 이라크에서는 고온과 잦은 가뭄이 복합적으로 작용하여 극심한 돌발 홍수(flash flood)가 문제가 되고 있으며, 이는 마른 땅에 폭우가 쏟아지면 더욱 심해지는 현상이다. 2019년 3월과 4월에 내린 폭우로 인해 이라크의 여러 행정 단위에서 홍수가 발생하였다. 특히, 이라크 북동부 지역은 홍수가 발생하기 쉬우며, 특히 집중호우로 인한 돌발 홍수가 발생한다. 2019년 홍수로 인해 술라이마니야(Sulaymaniyah)의 기반 시설 및 농업지역에 광범위한 피해가 발생하였다.

[0004] 홍수의 영향을 줄이고 피해 위험을 낮추기 위한 가장 중요한 방법 중 하나는 홍수 관리이다. 홍수 관리 및 예방의 첫 번째 단계는 홍수의 영향을 받을 가능성이 있는 지역을 매핑하는 것이다. 홍수를 예방하는 것이 어렵다고 생각되더라도 홍수 취약성에 대한 정확한 지도를 통해 잠재적인 홍수 발생 장소에 대한 예측이 가능하다. 홍수 위험 매핑의 주요 목표 중 하나는 지리적 지역을 다양한 홍수 위험 지역으로 나누는 것이다. 사람들이 홍수에

반응하기 보다는 적극적으로 대응할 수 있게 함으로써 이 지도를 기반으로 한 분석은 홍수로 인한 비용과 손실을 완화하는 데 도움이 될 수 있다.

[0005] 최근 위성 영상(satellite image), 데이터 분석(data analytic), 인공지능(artificial intelligence), GIS(geographic information science) 및 기타 관련 과학의 공간 및 시간 해상도 개선을 통해 홍수 취약성에 대한 심층 연구가 이루어지고 있다.

발명의 내용

해결하려는 과제

[0006] 레이더 위성 영상을 이용하여 침수 지역을 찾아 홍수의 인벤토리 맵(inventory map)을 작성할 수 있다.

[0007] 고위험 홍수 지역과 홍수에 기여하는 가장 중요한 인자를 식별하여 홍수 관리에 유용한 자원으로 활용할 수 있다.

[0008] 유전 알고리즘을 이용한 병렬 앙상블 기반 머신러닝 알고리즘의 최적화를 통해 FSM을 생성할 수 있다.

과제의 해결 수단

[0009] 적어도 하나의 프로세서를 포함하는 컴퓨터 장치의 홍수 민감도 매핑 방법에 있어서, 상기 적어도 하나의 프로세서에 의해, 대상 지역의 홍수 상태 인자(flood conditioning factor)에 해당되는 데이터를 이용하여 병렬 앙상블 기반 머신러닝(parallel ensemble-based machine learning) 알고리즘을 통해 상기 대상 지역의 FSM(flood susceptibility map)을 생성하는 단계를 포함하는 홍수 민감도 매핑 방법을 제공한다.

[0010] 일 측면에 따르면, 상기 생성하는 단계는, 상기 대상 지역의 레이더 위성 영상(radar satellite imagery)을 이용한 홍수 탐지(flood detection)를 통해 홍수 인벤토리 맵(flood inventory map)을 작성하는 단계를 포함할 수 있다.

[0011] 다른 측면에 따르면, 상기 작성하는 단계는, 홍수 전 레이더 위성 영상과 홍수 후 레이더 위성 영상을 비교하여 홍수를 탐지하는 단계를 포함할 수 있다.

[0012] 또 다른 측면에 따르면, 상기 작성하는 단계는, 궤도 파일 적용(orbital file application), 저감도 노이즈 감소(low-intensity noise reduction), 열 노이즈 제거(thermal noise removal), 방사선 측정 보정(radiometric calibration), 지형 보정(terrain correction), 및 후방 산란 계수(backscatter coefficient) 중 적어도 하나를 사용하여 상기 레이더 위성 영상을 전처리하는 단계를 포함할 수 있다.

[0013] 또 다른 측면에 따르면, 상기 작성하는 단계는, 후방 산란 히스토그램(backscatter histogram)과 시행착오(trial and error)를 기반으로 선정된 임계값을 기준으로 상기 레이더 위성 영상에서 홍수 지역을 분류하는 단계를 포함할 수 있다.

[0014] 또 다른 측면에 따르면, 상기 홍수 상태 인자는 토지 피복(land cover), 강우량(rainfall), 고도(altitude), 강까지의 거리(distance), TWI(topographic wetness index), 경사면(slope aspect), NDVI(normalized difference vegetation index), 지질(geology), 경사(slope), SPI(stream power index), 및 평면 곡률(plan curvature)을 포함할 수 있다.

[0015] 또 다른 측면에 따르면, 상기 생성하는 단계는, 과거의 홍수와 홍수를 야기하는 인자 간의 관계를 파악하기 위한 FR(Frequency ratio), 홍수에 영향을 미치는 매개변수 간의 관계를 파악하기 위한 다중공선성 분석(multicollinearity analysis), 및 변수의 공간적 층화 이질성(spatial stratified heterogeneity)을 기초로 이벤트 유발 인자(event's triggering factor)를 파악하기 위한 지리탐지기(Geodetector) 중 적어도 하나의 방법을 이용하여 상기 데이터를 전처리하는 단계를 포함할 수 있다.

[0016] 또 다른 측면에 따르면, 상기 생성하는 단계는, RF(Random forest) 알고리즘과 배깅(Bagging) 알고리즘의 최적화를 통해 상기 FSM을 모델링하는 단계를 포함할 수 있다.

[0017] 또 다른 측면에 따르면, 상기 모델링하는 단계는, 유전 메타휴리스틱 알고리즘(genetic metaheuristic algorithm)을 이용하여 상기 RF 알고리즘과 상기 배깅 알고리즘의 하이퍼파라미터(hyperparameter)에 대한 최적값을 선정하는 단계를 포함할 수 있다.

[0018] 또 다른 측면에 따르면, 상기 홍수 민감도 매핑 방법은, 상기 적어도 하나의 프로세서에 의해, RMSE(root mean

square error), ROC(receiver-operator characteristic) 곡선, SCAI(seed cell area index), 비모수 테스트(non-parametric test), 및 테일러 다이어그램(Taylor diagram) 중 적어도 하나의 지표를 이용하여 상기 FSM을 검증하는 단계를 더 포함할 수 있다.

[0019] 상기 홍수 민감도 매핑 방법을 상기 컴퓨터 장치에 실행시키기 위해 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램을 제공한다.

[0020] 컴퓨터 장치에서 판독 가능한 명령을 실행하도록 구현되는 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 대상 지역의 홍수 상태 인자(flood conditioning factor)에 해당되는 데이터를 이용하여 병렬 앙상블 기반 머신러닝(parallel ensemble-based machine learning) 알고리즘을 통해 상기 대상 지역의 FSM(flood susceptibility map)을 생성하는 것을 특징으로 하는 컴퓨터 장치를 제공한다.

발명의 효과

[0021] 본 발명의 실시예들에 따르면, 레이더 위성 영상을 이용하여 침수 지역을 찾아 홍수의 인벤토리 맵을 작성함으로써 데이터 부족 영역에서 효과적인 홍수 모니터링을 제공할 수 있다.

[0022] 본 발명의 실시예들에 따르면, 유전 메타휴리스틱 알고리즘을 사용하여 두 개의 병렬 앙상블 기반 머신러닝 알고리즘을 최적화하고 이를 통해 FSM을 생성함으로써 FSM 성능을 향상시킬 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 일실시예에 있어서 컴퓨터 장치의 내부 구성의 일례를 설명하기 위한 블록도이다.

도 2는 본 발명의 일실시예에 있어서 홍수 민감도 평가와 관련된 작업 흐름을 도시한 것이다.

도 3은 본 발명의 일실시예에 있어서 대상 지역의 지리적 위치를 도시한 것이다.

도 4는 본 발명의 일실시예에 있어서 SAR 영상을 통한 홍수 감지 과정을 도시한 것이다.

도 5는 본 발명의 일실시예에 있어서 SAR 영상에 의한 홍수 지역 예시를 도시한 것이다.

도 6 내지 도 7은 본 발명의 일실시예에 있어서 홍수와 관련된 상태 인자를 도시한 것이다.

도 8은 본 발명의 일실시예에 있어서 병렬 앙상블 기반 머신러닝 알고리즘의 최적화 과정을 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0024] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0026] 본 발명의 실시예들은 홍수 관리를 위한 FSM(flood susceptibility map)을 생성하는 기술에 관한 것이다.

[0027] 본 명세서에서 구체적으로 개시되는 것들을 포함하는 실시예들은 레이더 위성 영상을 이용하여 침수 지역을 찾아 홍수의 인벤토리 맵을 작성할 수 있고, 고위험 홍수 지역과 홍수에 기여하는 가장 중요한 인자를 식별하여 홍수 관리에 유용한 자원으로 활용할 수 있다.

[0028] 본 발명의 실시예들에 따른 홍수 민감도 매핑 장치는 적어도 하나의 컴퓨터 장치에 의해 구현될 수 있으며, 본 발명의 실시예들에 따른 홍수 민감도 매핑 방법은 홍수 민감도 매핑 장치에 포함되는 적어도 하나의 컴퓨터 장치를 통해 수행될 수 있다. 이때, 컴퓨터 장치에는 본 발명의 일실시예에 따른 컴퓨터 프로그램이 설치 및 구동될 수 있고, 컴퓨터 장치는 구동된 컴퓨터 프로그램의 제어에 따라 본 발명의 실시예들에 따른 홍수 민감도 매핑 방법을 수행할 수 있다. 상술한 컴퓨터 프로그램은 컴퓨터 장치와 결합되어 홍수 민감도 매핑 방법을 컴퓨터에 실행시키기 위해 컴퓨터 판독 가능한 기록매체에 저장될 수 있다.

[0029] 도 1은 본 발명의 일실시예에 따른 컴퓨터 장치의 예를 도시한 블록도이다. 예를 들어, 본 발명의 실시예들에 따른 홍수 민감도 매핑 장치는 도 1을 통해 도시된 컴퓨터 장치(100)에 의해 구현될 수 있다.

[0030] 도 1에 도시된 바와 같이 컴퓨터 장치(100)는 본 발명의 실시예들에 따른 홍수 민감도 매핑 방법을 실행하기 위한 구성요소로서, 메모리(110), 프로세서(120), 통신 인터페이스(130) 그리고 입출력 인터페이스(140)를 포함할 수 있다.

[0031] 메모리(110)는 컴퓨터에서 판독 가능한 기록매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 여기서 ROM과 디스크 드라이브와 같은 비소멸성 대용량 기록장치는 메모리(110)와는 구분되는 별도의 영구 저장 장치로서 컴퓨터 장치(100)에 포함될 수도 있다. 또한, 메모리(110)에는 운영체제와 적어도 하나의 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 메모리(110)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 메모리(110)로 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 인터페이스(130)를 통해 메모리(110)에 로딩될 수도 있다. 예를 들어, 소프트웨어 구성요소들은 네트워크(160)를 통해 수신되는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 컴퓨터 장치(100)의 메모리(110)에 로딩될 수 있다.

[0032] 프로세서(120)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(110) 또는 통신 인터페이스(130)에 의해 프로세서(120)로 제공될 수 있다. 예를 들어, 프로세서(120)는 메모리(110)와 같은 기록 장치에 저장된 프로그램 코드에 따라 수신되는 명령을 실행하도록 구성될 수 있다.

[0033] 통신 인터페이스(130)는 네트워크(160)를 통해 컴퓨터 장치(100)가 다른 장치와 서로 통신하기 위한 기능을 제공할 수 있다. 일례로, 컴퓨터 장치(100)의 프로세서(120)가 메모리(110)와 같은 기록 장치에 저장된 프로그램 코드에 따라 생성한 요청이나 명령, 데이터, 파일 등이 통신 인터페이스(130)의 제어에 따라 네트워크(160)를 통해 다른 장치들로 전달될 수 있다. 역으로, 다른 장치로부터의 신호나 명령, 데이터, 파일 등이 네트워크(160)를 거쳐 컴퓨터 장치(100)의 통신 인터페이스(130)를 통해 컴퓨터 장치(100)로 수신될 수 있다. 통신 인터페이스(130)를 통해 수신된 신호나 명령, 데이터 등은 프로세서(120)나 메모리(110)로 전달될 수 있고, 파일 등은 컴퓨터 장치(100)가 더 포함할 수 있는 저장 매체(상술한 영구 저장 장치)로 저장될 수 있다.

[0034] 통신 방식은 제한되지 않으며, 네트워크(160)가 포함할 수 있는 통신망(일례로, 이동통신망, 유선 인터넷, 무선 인터넷, 방송망 등)을 활용하는 통신 방식뿐만 아니라 기기들 간의 근거리 유선/무선 통신 역시 포함될 수 있다. 예를 들어, 네트워크(160)는, PAN(personal area network), LAN(local area network), CAN(campus area network), MAN(metropolitan area network), WAN(wide area network), BBN(broadband network), 인터넷 등의 네트워크 중 하나 이상의 임의의 네트워크를 포함할 수 있다. 또한, 네트워크(160)는 버스 네트워크, 스타 네트워크, 링 네트워크, 메쉬 네트워크, 스타-버스 네트워크, 트리 또는 계층적(hierarchical) 네트워크 등을 포함하는 네트워크 토폴로지 중 임의의 하나 이상을 포함할 수 있으나, 이에 제한되지 않는다.

[0035] 입출력 인터페이스(140)는 입출력 장치(150)와의 인터페이스를 위한 수단일 수 있다. 예를 들어, 입력 장치는 마이크, 키보드, 카메라 또는 마우스 등의 장치를, 그리고 출력 장치는 디스플레이, 스피커와 같은 장치를 포함할 수 있다. 다른 예로 입출력 인터페이스(140)는 터치스크린과 같이 입력과 출력을 위한 기능이 하나로 통합된 장치와의 인터페이스를 위한 수단일 수도 있다. 입출력 장치(150)는 컴퓨터 장치(100)와 하나의 장치로 구성될 수도 있다.

[0036] 또한, 다른 실시예들에서 컴퓨터 장치(100)는 도 1의 구성요소들보다 더 적은 혹은 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 컴퓨터 장치(100)는 상술한 입출력 장치(150) 중 적어도 일부를 포함하도록 구현되거나 또는 트랜시버(transceiver), 카메라, 각종 센서, 데이터베이스 등과 같은 다른 구성요소들을 더 포함할 수도 있다.

[0037] 이하에서는 레이더 위성 영상을 이용한 홍수 민감도 매핑을 위한 머신러닝 알고리즘을 최적화하는 기술의 구체적인 실시예를 설명하기로 한다.

[0038] 다중 기준 결정 분석(MCDA), 물리적 기반 모델(physically based model), 통계 방법(statistical method), 그리고 머신러닝 방법(machine learning method)은 FSM을 만드는 데 가장 많이 사용되는 4가지 접근법이다.

[0039] AHP(analytical hierarchy process), ANP(analytic network process), TOPSIS(technique for order preference by similarity to ideal solution), WLC(weighted linear combination), 및 VIKOR(VlseKriterijumska Optimizacija I Kompromisno Resenje)과 같은 MCDA 방법이 FSM를 작성하는 데 사용된다.

[0040] 물리적 기반 방법이 홍수 모델링에 사용되기는 하나, 물리적 기반 방법의 단점은 전문가의 판단을 기반으로 하며 작은 영역에만 적합하다는 것이다.

[0041] 이변량 통계(bivariate statistical) 및 머신러닝 방법과 같은 데이터 기반(data-driven)을 사용한다. FR(frequency ratio), CF(certainty factor), WOE(weight of evidence), 및 EBF(evidential belief function)는 FSM을 만드는 데 사용되는 통계 방법이다. 홍수는 일반적으로 선형 구조가 아니나, 상기한 통계 방법은 변수가 선형이라는 가정을 기반으로 한다.

[0042] 머신러닝은 많은 이점 때문에 자연재해를 예측하는 데 효과적인 방법이다. 여기에는 트렌드와 패턴을 빠르고 쉽게 찾고, 반복 작업을 자동화하며, 다차원 및 다중 유형 데이터를 처리할 수 있는 능력이 포함된다. ANN(artificial neural network), ANFIS(adaptive neuro-fuzzy inference system), SVM(support vector machine), SVR(support vector regression), DT(decision tree), LR(logistic regression) 등의 앙상블 알고리즘을 활용하여 홍수 민감도 맵을 생성할 수 있다. 부트스트랩 집계(bootstrap aggregating)(Bagging), 부스팅(boosting), 스태킹(stacking), RF(random forest)의 앙상블 알고리즘은 모델링 정확도를 높이고 단일 머신러닝 알고리즘의 한계를 극복하기 위해 FSM을 준비하는 데 사용될 수 있다. 앙상블 알고리즘은 데이터 세트에 선형 및 비선형 데이터 유형이 모두 포함된 경우 모델의 조합이 필요하기 때문에 도움이 된다. 구성 설정(configuration setting)을 미세조정(fine-tuning)하는 절차는 앙상블 알고리즘의 주요 단점이다. 메타휴리스틱 알고리즘(Metaheuristic algorithm)은 머신러닝 알고리즘을 최적화하는 효과적인 방법이다. 다양한 홍수 모델링 연구에서는 메타 휴리스틱 알고리즘과 머신러닝을 결합하여 사용하고 있다.

[0043] 과거 홍수의 분포도는 FSM을 만드는 데 필요한 기본 데이터 중 하나이다. 대부분의 연구는 관계 기관에서 수집한 데이터의 통계적 분석이나 현장 조사를 통해 이러한 데이터를 수집한다. 현장 조사 방법은 비싸고 부정확하며 시간이 많이 걸리기 때문에 원격 탐지 이미지(remote-sensing image)를 사용하여 홍수 분포도를 생성한다. 자연재해의 가능성과 위험을 모니터링하고 평가하는 데 있어 원격 탐지 이미지는 점점 더 필수적인 도구가 되고 있다. 물리적으로 접근할 수 없는 지역에서도 추론된 원격 탐지 기술은 홍수 사건 공간 데이터를 수집하기 위한 빠르고 정확한 솔루션을 제공한다. SAR(synthetic aperture radar) 기술은 대규모 홍수를 효과적으로 모니터링할 수 있다. 홍수 매핑을 위한 SAR 이미지의 중요성에 기여하는 주요 특성 중 하나는 낮과 밤뿐만 아니라 어떤 기상 상태에서도 작동할 수 있다는 점이다. 홍수 매핑 접근 방식은 거의 실시간으로 홍수를 기록할 수 있어야 하며, 이는 특히 많은 이미지를 활용하여 홍수 진화 패턴을 분석할 때 최소 계산 비용으로 신속하게 작동해야 함을 의미한다.

[0044] 본 실시예에서는 반건조 지역(semi-arid region)의 홍수를 자동으로 신속하게 식별하기 위해 다중 시기(multi-temporal) 위성(Sentinel-1) GRD(ground range detected) 이미지를 활용할 수 있다. 본 발명의 목적은 Sentinel-1 SAR(synthetic aperture radar) 위성 영상을 사용하여 홍수 모니터링을 하고 두 개의 병렬 앙상블 기반 머신러닝 알고리즘의 최적화를 사용하여 FSM을 만드는 것이다. 이때, 유전 메타휴리스틱 알고리즘(genetic metaheuristic algorithm)을 사용하여 두 개의 병렬 앙상블 기반 머신러닝 알고리즘(Bagging과 RF)을 최적화할 수 있다. 다시 말해, 본 실시예들은 (1) 데이터 부족 영역에서 홍수 모니터링을 위해 Sentinel-1 SAR 위성 영상을 사용하는 것, (2) GA(genetic algorithm)에 의한 병렬 앙상블 기반 머신러닝 알고리즘의 최적화를 통해 FSM을 모델링하는 것을 포함한다.

[0045] 도 2는 본 발명의 일실시예에 있어서 홍수 민감도 평가와 관련된 작업 흐름을 도시한 것이다.

[0046] 도 2를 참조하면, 첫 번째 단계는 홍수 인벤토리 맵과 홍수 상태 인자(flood condition factor)를 포함하는 기존 정보의 데이터베이스를 컴파일한다. SAR 영상은 홍수를 모니터링하고 홍수 인벤토리 맵을 준비하기 위해 사용된다. 두 번째 단계에서는 데이터 전처리를 수행한다. 전처리 단계에서는 다중공선성 접근법(multicollinearity approach)을 사용하여 기준(criterion) 간의 공선성을 결정하고, FR 방법을 사용하여 인자의 가중치를 결정한 후 지리탐지기(Geodetector)가 변수의 중요성(significance)을 계산한다. 세 번째 단계에서는 GA 알고리즘이 병렬 앙상블 기반 머신러닝 알고리즘(Bagging 및 RF)을 최적화한다. 다음 단계에서는 FSM을 생성하기 위해 Bagging-GA, RF-GA, Bagging, RF 등 네 가지 모델이 사용된다. 마지막으로 결과는 SCAI(seed cell area index), ROC(receiver-operator characteristic) 곡선, 테일러 다이어그램(Taylor diagram), 비모수 테스트(non-parametric test)(Friedman과 Wilcoxon 순위 테스트), RMSE(root mean square error)의 다섯 가지 다른 지표를 사용하여 평가된다.

[0047] 구체적인 실시예는 다음과 같다.

[0048] 대상 지역

[0049] FSM을 생성하기 위한 대상 지역을 이라크의 술라이마니야로 한다. 도 3은 대상 지역인 술라이마니야의 지리적

위치를 도시한 것이다. 이라크 쿠르디스탄 지역에서 술라이마니야는 가장 큰 지방이며, 이라크의 북동쪽 부분에 위치하고 있다. 술라이마니야는 위도 35° 04'와 36° 30' 사이, 경도 44° 50'과 46° 16' 사이에 위치하고 있다. 술라이마니야 지역은 해발 0미터에서 3422m 사이의 고도에 위치하며, 이 지역의 북부와 북동부는 더 높은 고도를 가지고 있다. 술라이마니야 지방은 총 17,023km^2를 차지한다. 술라이마니야의 여름(6월부터 8월)은 건조하고 덥고 평균 기온은 31.5° C이다. 겨울(12월부터 2월)은 평균 기온 7.6° C로, 기온이 현저히 낮아지고 강수량과 바람이 증가하며 산발적인 눈이 내린다. 연평균 강우량은 약 400mm에서 600mm이며, 10월에 약하게 시작하여 11월에 강도가 세지고, 그 후 5월까지 계속된다. 폭우와 변화하는 기후 패턴은 이 지역의 홍수의 주요 원인이다. 대상 지역의 토지 면적은 대부분 목초지와 농업지이다.

[0050] SAR 영상을 이용한 홍수 탐지(Flood detection using SAR imagery)

[0051] 미래에 홍수에 취약할 수 있는 세계 지역을 예측하기 위한 머신러닝 방법을 구현하기 위해 기존의 홍수 기록을 훈련 기준으로 사용한다. 본 실시예에서는 위성 영상을 사용하여 홍수 인벤토리 맵을 작성할 수 있다. 홍수의 영향을 받는 지역은 홍수 전후에 찍은 영상을 비교하여 탐지할 수 있다. 2019년에 이라크의 광대한 지역은 전례 없는 비로 황폐화되었다. 이라크 민방위 기관에서 보고한 바와 같이 홍수의 정점은 2019년 1월 27일과 28일에 발생했다. SAR 기반 홍수 모니터링을 위해서는 홍수 전(홍수 한 달 전)과 홍수 후(홍수 한 달 후) 영상이 필요하다. Sentinel-1 레벨 1 GRD 제품은 2018년 12월의 홍수 전 침수 매핑과 2019년 2월의 홍수 후 침수 매핑에 모두 사용된다. Sentinel-1 영상의 시간 해상도는 6일이다. 이를 위해 이러한 영상에는 IW(interferometry-wide) 스와트 센서(swath sensor), VV 및 VH 양극화(polarization), 상승 경로(ascending path)가 사용된다. Sentinel-1을 사용하여 홍수를 모니터링하기 위해 GEE(Google Earth Engine) 플랫폼(https://earthengine.google.com)을 사용한다. 도 4는 SAR 영상을 통한 홍수 감지 과정을 도시한 것이다. 도 4를 참조하면, 첫 번째 단계에서는 궤도 파일 적용(orbital file application), 저감도 노이즈 감소(low-intensity noise reduction), 열 노이즈 제거(thermal noise removal), 방사선 측정 보정(radiometric calibration), 지형 보정(terrain correction), 및 후방 산란 계수(backscatter coefficient)(σ°)를 사용하여 영상을 전처리한다. 그런 다음, 레이더 영상 얼룩(radar image speckle)의 영향을 줄이기 위해 평활 필터(smoothing filter)가 적용된다. 홍수 전 모자이크에서 홍수 후 모자이크를 빼서 픽셀당 변화 정도(degree of change)를 반영하는 래스터 레이어(raster layer)를 생성한다. 위성 영상 처리에서 두 가지 특징을 구별하는 방법 중 하나는 히스토그램 임계값(histogram threshold)이다. 홍수 지역은 VV 및 VH 양극화에서 별개의 지구 물체 사이에서 낮은 후방 산란 값을 나타내기 때문에 정의된 임계값은 홍수 지역을 분류하는 데 사용될 수 있다. 후방 산란 히스토그램(backscatter histogram)과 시행착오(trial and error)를 기반으로 임계값을 1.25로 선택한다. 1.25 이상의 모든 값에는 1 값이 할당되고, 1.25 미만의 모든 값에는 0 값이 할당된다. 잘못 분류된 영역을 제거하기 위해 DEM(digital elevation model) 및 GSW(global surface water)를 포함한 여러 추가 데이터 세트가 홍수 범위 계층(flood extent layer)에 적용된다. 먼저, 임계값을 초과한 픽셀은 물을 보유할 수 없고 대신 낮은 고도 영역으로 배수되기 때문에 5%를 초과하는 모든 장소가 제거된다. 이를 위해 SRTM(shuttle radar topography mission)에 의해 수집된 DEM이 사용된다. 다음으로 10개월 이상 물이 고인 곳은 상설수역과 반영구적 수역이므로 제외한다. GSW 데이터세트는 이러한 침수 지역을 개선하는 데 사용된다. 마지막으로, 연결된 픽셀이 8개 미만인 지역 또한 제거된다. 상기한 작업을 통해 홍수 범위에 대한 노이즈를 최소화할 수 있다. 도 5는 홍수 모니터링 지도를 도시한 것이다.

[0052] 먼저 홍수 발생 지역(폴리곤)을 래스터 지도(raster map)로 변환한 후, 각 픽셀을 홍수 지점(flood point)으로 변환하여 모델링할 수 있다. 마지막으로 모니터링된 홍수지역으로부터 160개의 홍수점을 도출하여 모델링에 사용한다(도 2 참조). 앙상블 모델의 음성 샘플(negative sample)로 침수되지 않은 나머지 지역의 포인트 중에서 무작위로 160개 포인트를 선택한다. 침수된 장소와 침수되지 않은 장소는 각각 1과 0으로 지정된다. 마지막으로 각 샘플 포인트에는 무작위 할당이 주어지며, 그 중 70%는 훈련 포인트로, 30%는 검증 포인트로 사용된다.

[0053] 상태 인자(Conditioning factor)

[0054] 홍수 민감도 매핑의 첫 번째 단계는 홍수 상태 매개변수(flood condition parameter)를 사용하여 공간 데이터베이스(spatial database)를 구축하는 것이다. 그러나 홍수에 영향을 미치는 요인은 다른 장소의 특성에 따라 다르고, 동일한 인자가 여러 지역에서 매우 다른 영향을 미친다. 본 실시예에서는 지역에 대한 철저한 분석, 홍수 모델링 작업 검토, 및 데이터 가용성을 거쳐 11개의 기준을 선택한다. 본 실시예에서 사용된 데이터에는 토지 피복(land cover), 강우량(rainfall), 고도(altitude), 강까지의 거리(distance), TWI(topographic wetness index), 경사면(slope aspect), NDVI(normalized difference vegetation index), 지질(geology), 경사

(slope), SPI(stream power index), 평면 곡률(plan curvature) 등이 포함된다. 지형 기반 기준(TWI, 고도, SPI, 평면 곡률, 양상, 경사)은 DEM을 기반으로 작성된다. 본 발명에서 사용된 DEM은 ASTER(Advanced Spaceborne Thermal Emission and Reflection Radiometer) 이미지의 데이터를 사용하여 작성된다. DEM 맵은 픽셀 크기가 30X30m인 GEE 플랫폼에서 작성된다. 지형 기준은 DEM에서 추출되어 30X30m인 픽셀로 ArcGIS 10.3 및 SAGA GIS 8.2.1 소프트웨어에서 작성된다. TWI 및 SPI 기준은 수학식 1, 수학식 2를 사용하여 계산된다.

[0055] [수학식 1]

[0056] $$TWI = ln\frac{A_s}{\tan\beta}$$

[0057] [수학식 2]

[0058] $$SPI = A_s \times \tan\beta$$

[0059] 여기서, A는 집수 면적(catchment area)을 나타내는 지표이고, β는 경사각(slope angle)이다.

[0060] Sentinel-2 영상의 분석은 2020년 토지피복도를 작성하기 위해 GEE 플랫폼에서 수행된 것으로, 하천지도로부터의 거리는 ArcGIS 10.3의 ArcHydro 도구를 사용하여 DEM으로부터 하천망을 획득하여 작성된다. 대상 지역의 강우맵(rainfall map)은 이라크 전역의 18개 측우기 자료에 30년(1991-2020) 동안 적용된 크리깅 보간법(kriging interpolation method)을 사용하여 작성된 것이다. NDVI 맵은 Landsat 8에서 획득한 영상을 사용하여 GEE 플랫폼에서 생성된 것이다. NDVI 지수는 수학식 3을 사용하여 계산된다.

[0061] [수학식 3]

[0062] $$NDVI = \frac{(NIR - R)}{(NIR + R)}$$

[0063] 여기서, R은 band 4 적색 스펙트럼이고 NIR은 band 5 근적외선 스펙트럼이다. 이라크 지질지도(geological map)는 해당 국가의 지질도를 1:100,000 축척으로 구성된다. 상기한 11개 기준 맵은 도 6과 도 7에 도시한 바와 같다.

[0064] FSM 모델링

[0065] (1) 전처리 방법(Preprocessing method)

[0066] FR(Frequency ratio) 방법: FR 방법은 통계 분석(statistical analysis)을 위한 방법이다. FR 방법은 과거의 홍수와 홍수를 야기하는 인자들 간의 관계를 설명한다. FR 방법의 첫 번째 단계는 자연파괴(natural break) 및 수동 분류(manual classification) 방법을 바탕으로 기준을 구분한 후 이러한 분류에서 홍수 지점을 추출한다. FR 값은 기준의 각 하위 클래스에 속한 홍수 수(NF)를 해당 지역의 총 홍수 수(AF)로 나눈 값과 각 하위 클래스의 픽셀 수(NP)를 해당 지역의 총 픽셀 수(AP)로 나눈 값으로 얻을 수 있다. FR 값이 높을수록 홍수 발생 시 해당 등급의 중요성을 나타낸다.

[0067] [수학식 4]

[0068] $$FR = \frac{NF/AF}{NP/AP}$$

[0069] 다중공선성 분석(multicollinearity analysis): 홍수에 영향을 미치는 매개변수 간의 관계를 파악하기 위해 다중공선성 분석이 사용된다. 다중공선성은 조건화 요인들 사이에 유의미한 상관관계가 있을 때 나타나는 문제이다. 기준 간의 공선성이 강하면 모델링 결과에 오류가 발생할 수 있다. 결과적으로, 다중공선성 발생을 찾아 기술하기 위해 VIF(variance inflation factor) 지수를 적용할 수 있다. VIF 점수가 10보다 크면 다중 공선성에 심각한 문제가 있다는 것을 의미한다. VIF 지수를 계산하기 위해 IBM SPSS Statistics 26 소프트웨어가 사용될 수 있다.

[0070] 지리탐지기 방법을 이용한 특징 중요도(Feature importance using the Geodetector method): 환경에서 지리적으로 계층화된 이질성의 존재는 계층 내 분산이 계층 간 분산보다 작을 때 발생하며, 이는 해당 관계가 지리적으로 변한다는 것을 나타낸다. 지리탐지기는 변수의 공간적 층화 이질성(spatial stratified heterogeneity)을

파악하여 이벤트 유발 인자(event's triggering factor)를 발견하기 위한 도구이다. 종속 매개변수(Y)를 측정하고 Y와 X(독립 매개변수) 사이의 관계를 테스트할 뿐만 아니라 두 독립 변수가 어떻게 상호 작용하여 Y에 영향을 미치는지 살펴본다. q-값은 독립 변수가 관심 종속 변수의 결정 요인인 정도(degree)를 결정하기 위해 요인 탐지기(factor detector)가 사용하는 측정항목이다.

[0071] [수학식 5]

[0072] $$q = 1 - \frac{SSW}{SST}$$

[0073] [수학식 6]

[0074] $$SSW = \sum_{h=1}^{L}\sum_{i=1}^{N_h}(Y_{hi} - \bar{Y}_h)^2 = \sum_{h=1}^{L} N_h\sigma_h^2$$

[0075] [수학식 7]

[0076] $$SST = \sum_{i=1}^{N}(Y_i - \bar{Y})^2 = N\sigma^2$$

[0077] 여기서, h=1,…L은 독립 변수 X의 계층을 나타내고, 계층 h에 있는 표본 단위의 수는 N_h로 표시되며, 전체 영역에 걸쳐 있는 표본 단위의 수는 N으로 표시된다. Y_i와 Y_{hi}는 각각 커뮤니티 및 계층 h에 있는 단위 i의 값을 나타낸다. 계층 및 지역의 분산은 각각 σ_h^2 및 σ^2로 표시된다. SST는 모든 제곱의 합이고, SSW는 주어진 집합 내 제곱의 합이다. 이 지수의 값은 0과 1 사이이며, 값이 클수록 매개변수의 중요도가 높음을 나타내며, 이를 계산하기 위해 지리탐지기 소프트웨어가 사용된다.

[0078] (2) 홍수 모델링 기법(Flood modeling techniques)

[0079] GA(Genetic algorithm): 유전 알고리즘은 최적화에 사용되며 자연 진화 개념을 기반으로 한다. GA는 모집단을 기반으로 하는 알고리즘이다. 각 염색체는 솔루션을 나타내고 각 매개변수는 유전자를 의미한다. GA는 적합도 함수의 조건과 얼마나 잘 일치하는지에 따라 모집단의 각 구성원에게 "적합도" 점수를 부여한다. 유전 알고리즘은 일반적으로 두 단계를 포함한다. 첫 번째 단계는 다음 세대를 생성할 개체를 선택하는 것이고, 두 번째 단계는 다음 세대를 생성하기 위해 교차 및 돌연변이와 같은 기술을 통해 선택된 개체를 조작하는 것이다. 유전 알고리즘의 첫 번째 단계는 모집단을 무작위로 생성하는 것이다. 가우스 무작위 분포(Gaussian random distribution)를 사용함으로써 생성된 모집단의 다양성을 높일 수 있다. GA 알고리즘에서는 적합도(또는 목적) 값을 활용하여 어떤 개체가 다음 세대에 더 높은 포함 확률을 부여받을 것인지를 결정한다. 선택 연산자(selection operator)를 사용하여 적합한 개체를 선택한 후 다음 세대를 생성하는 작업에 투입한다. 기존의 두 솔루션을 결합하여 두 개의 새로운 솔루션을 생성함으로써 GA 알고리즘은 이를 복제한다. 자손이 실행 가능한 솔루션을 생성한 후 돌연변이를 최종 진화 연산자로 사용하여 하나 이상의 유전자를 변경할 수 있다. 이 알고리즘은 종료 기준(termination criterion)에 도달할 때까지 위의 세 가지 연산자를 사용하여 모집단을 세분화한다.

[0080] RF(Random forest) 알고리즘: RF 알고리즘은 많은 의사결정 트리로 구성된 앙상블 머신러닝 알고리즘이다. RF 알고리즘은 데이터 세트를 훈련 집합(in-Bag)과 검증 집합(out-of-Bag)으로 무작위로 분할함으로써 시작된다. 그런 다음, 데이터 세트로부터 "부트스트랩 표본"을 무작위로 추출하여 많은 수의 의사결정 트리를 생성한다. 각 트리의 분기는 각 노드에서 무작위로 선택된 예측 변수에 의해 설정된다. 마지막으로, RF 추정(RF estimate)은 모든 개별 트리의 예측의 가중 평균(weighted mean)이다. RF 알고리즘의 장점 중 하나는 많은 양의 연속적이고 이산적인 변수를 관리할 수 있다는 것이다. RF 알고리즘의 예측 성능은 주로 1) 상관 관계를 줄여야 할 요구, 2) 각 기본 트리의 성능 향상, 3) 트리에 높은 계산 제약이 적용되는 세 가지 특징에 의해 영향을 받는다.

[0081] 부트스트랩 집계(Bootstrap aggregating)(Bagging) 알고리즘: Bagging은 훈련 데이터의 다양한 하위 집합을 사용하여 동일한 알고리즘을 여러 번 훈련하는 것을 포함하는 앙상블 접근법의 한 유형이다. 배깅의 버전은 훈련 집합의 복제본을 부트스트래핑하고 이를 새로운 학습 세트로 사용하여 생성된다. 배깅은 훈련 하위 집합이 교체를 통해 무작위로 선택되어 분류기(classifier)가 중복 데이터에 대해 훈련되는 것을 방지한다. 배깅의 목적은

학습된 분류기의 예측을 단일 예측으로 결합하는 복합 분류기(composite classifier)를 생성하여 기본 분류기의 정확도를 향상시키는 것이다. 투표(voting)는 배깅에서 사용하는 조합 메커니즘(combination mechanism)을 사용하여 수행된다. 이 기술은 의사결정 트리에서 자주 사용되지만 모든 모델에 적용 가능하다. 배깅 알고리즘의 계산은 수학식 8과 같다.

[0082] [수학식 8]

[0083] $$f_{bag} = f_1(x) + f_2(x) + \cdots + f_n(x)$$

[0084] 여기서, x는 목표 예측(target prediction)을 나타내고, f_{bag}은 배깅 예측을 나타내고, $f_1(x)+f_2(x)+\cdots+f_n(x)$는 개별 개체 학습자의 예측(individual base learners' prediction)을 나타낸다.

[0085] 최적화 알고리즘(Optimization algorithm): 머신러닝 알고리즘을 사용함에 있어 가장 어려운 작업 중 하나는 모델 최적화이다. 머신러닝에서 하이퍼파라미터 최적화의 목표는 특정 머신러닝 알고리즘의 어떤 하이퍼파라미터가 가장 높은 성능을 나타내는지를 결정하는 것이다. 본 발명의 경우 RF 및 배깅 알고리즘(RF-GA, Bagging-GA)의 하이퍼파라미터에 대한 최적 값을 선정하기 위해 GA 알고리즘을 활용한다. 모델 최적화를 위한 하이브리드 알고리즘의 작업 과정은 도 8과 같다.

[0086] GA 알고리즘에서 목적 함수(R^2)는 수학식 9와 같다.

[0087] [수학식 9]

[0088] $$Object\ function = Max(R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2})$$

[0089] 여기서, $\hat{y}$는 y의 예측 값, $\bar{y}$는 y의 평균 값, y는 실제 값(actual value)을 의미한다.

[0090] (3) 검증 방법(Validation method)

[0091] 홍수 취약지역을 파악하기 위해 사용된 모델의 효과를 평가하는 것이 모델화의 중요한 부분으로 RMSE, ROC, 테일러 다이어그램, SCAI 지수, Friedman, Wilcoxon 부호 순위 테스트를 사용한다.

[0092] RMSE 지수: 예측오차는 모델 성능을 평가하기 위한 정량적인 측정치로 실제 값과 예측값의 차이를 의미한다. 모델의 정확도를 평가하기 위해 수학식 10의 RMSE 지수를 사용한다.

[0093] [수학식 10]

[0094] $$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N}}$$

[0095] 여기서, N은 관측치(observation)의 수이고, y_i는 실제 값이며, $\hat{y}$는 예측된 값이다.

[0096] ROC 곡선: ROC 곡선은 예측을 위한 다양한 결정론적 모델과 확률론적 모델의 정확도를 비교하는 데 사용할 수 있다. 홍수 발생과 비발생을 정확하게 구별하기 위해 선택된 모델의 정확성은 ROC 곡선과 AUC(area under the curve) 값을 사용하여 평가될 수 있다. 수학식 11을 통해 AUC 값을 계산할 수 있다.

[0097] [수학식 11]

[0098] $$AUC = \sum_{i=1}^{n}(X_{i+1} - X_i) \times (S_{i+1} - S_i - \frac{S_i}{2})$$

[0099] 여기서, X_i는 특이성(specificity)을 나타내고, S_i는 민감성(sensitivity)을 나타낸다.

[0100] AUC의 값은 0.5에서 1 사이로 추정되고 있으며, AUC 값이 1에 가까울수록 예측 모델의 정확도가 높다는 것을 의미한다.

[0101] SCAI 지수: SCAI에서는 주어진 민감도 클래스(susceptibility class)에 속하는 추정 픽셀 수와 해당 클래스 내에 위치한 실제 픽셀 수 간의 비율이 고려된다. 가장 취약한 그룹의 SCAI 값은 다른 그룹의 SCAI 값보다 낮다.

SCAI 지수는 수학식 12와 같다.

[0102] [수학식 12]

[0103] $$SCAI = \frac{Number\ of\ pixels\ of\ susceptibility\ class(\%)}{Number\ of\ floods\ per\ class(\%)}$$

[0104] 비모수 테스트: 모델 결과 간의 통계적으로 유의한 편차를 평가하기 위해 비모수 테스트가 사용된다. Friedman 테스트를 실행하기 전에 모든 예측 모델은 5% 유의 수준(significance level)이어야 한다. 두 모델의 예측 간에 유의한 차이가 없다는 결론을 내리기 위해서는 두 모델 모두에 대해 Friedman 테스트의 P-값 임계값을 충족해야 한다. Wilcoxon 순위 테스트는 다양한 예측 모델 간의 차이의 통계적 유의성(statistical significance)을 결정하는 데 사용된다. z-값이 +1.96 또는 -1.96보다 높으면 대립 가설(alternative hypothesis)이 확인된다.

[0105] 테일러 다이어그램: 테일러 다이어그램은 패턴이 관측값과 얼마나 잘 일치하는지 그래픽으로 요약할 수 있다. 상관관계 측정항목(Metrics of correlation), 중앙의 RMSE 차이, 그리고 표준편차 진폭(standard deviation amplitude)은 두 패턴 간의 유사성을 평가하는 데 유용한 방법이다. 이러한 다이어그램은 여러 가지 모델의 상대적 품질을 결정하는 데 매우 유용하다.

[0106] 본 실시예들은 대상 지역에 대한 FSM을 구축하기 위해 4개의 머신러닝 알고리즘(RF, Bagging, RF-GA, Bagging-GA)이 사용될 수 있으며, 병렬 앙상블 기반 머신러닝 알고리즘(RF)과 부트스트랩 집계(Bagging)를 미세 조정하기 위해 유전 알고리즘(GA)을 사용할 수 있다. 그리고, 병렬 앙상블 기반 머신러닝 알고리즘에 입력을 제공하기 위해 11개 기준(토지 피복, 강우량, 고도, 강까지의 거리, TWI, 경사면, NDVI, 지질, 경사, SPI, 평면 곡률)의 데이터를 수집하여 처리할 수 있다. 또한, 본 실시예에서는 Sentinel-1 SAR 위성 영상을 활용하여 침수 지역을 찾아 홍수 인벤토리 맵을 작성할 수 있으며, FSM을 위한 데이터 전처리에 다중 공선성, FR 및 지리탐지기 방법이 사용될 수 있다.

[0107] 이처럼 본 발명의 실시예들에 따르면, 레이더 위성 영상을 이용하여 침수 지역을 찾아 홍수의 인벤토리 맵을 작성함으로써 데이터 부족 영역에서 효과적인 홍수 모니터링을 제공할 수 있다. 더욱이, 본 발명의 실시예들에 따르면, 유전 메타휴리스틱 알고리즘을 사용하여 두 개의 병렬 앙상블 기반 머신러닝 알고리즘을 최적화하고 이를 통해 FSM을 생성함으로써 FSM 성능을 향상시킬 수 있다.

[0108] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 어플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0109] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0110] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 이때, 매체는 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수 개의 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워

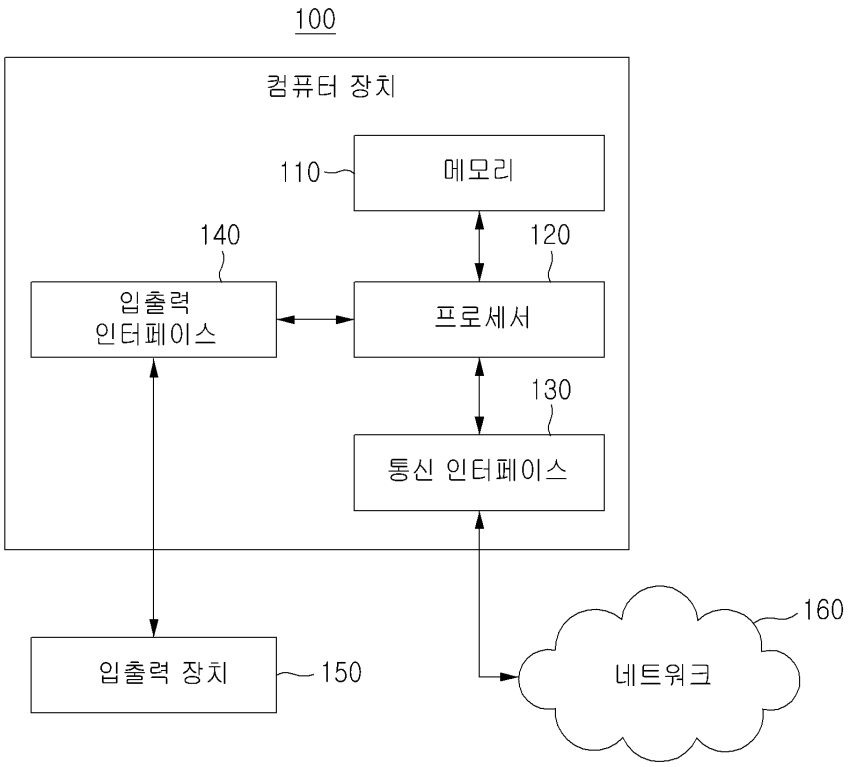
크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체의 예시로, 어플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다.

[0111] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

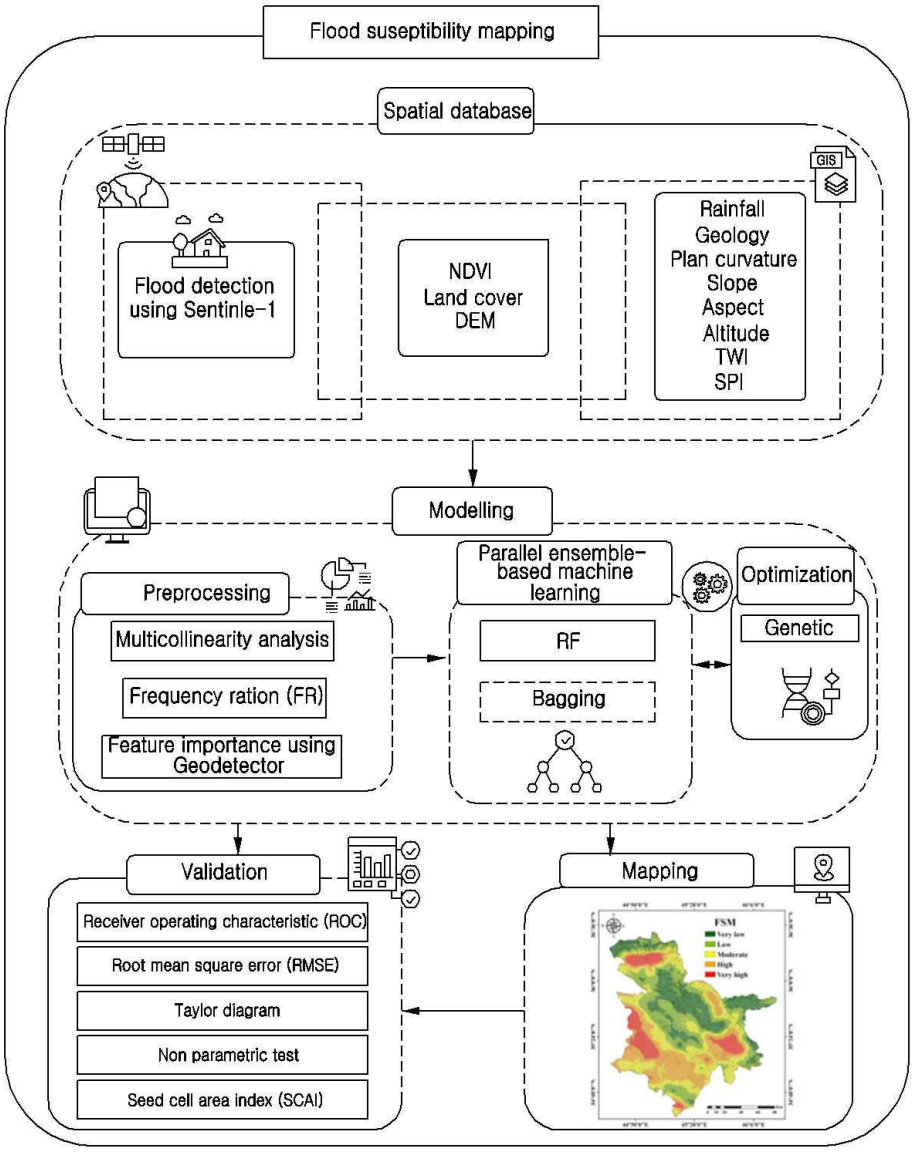
[0112] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

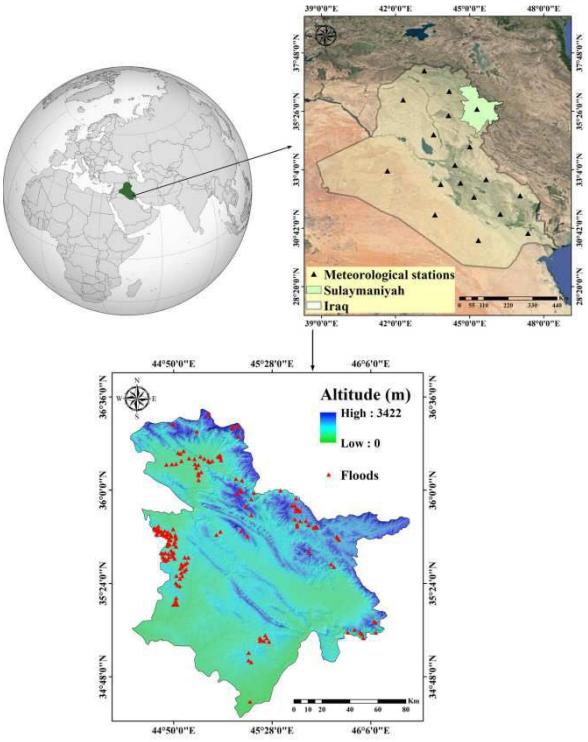
도면1



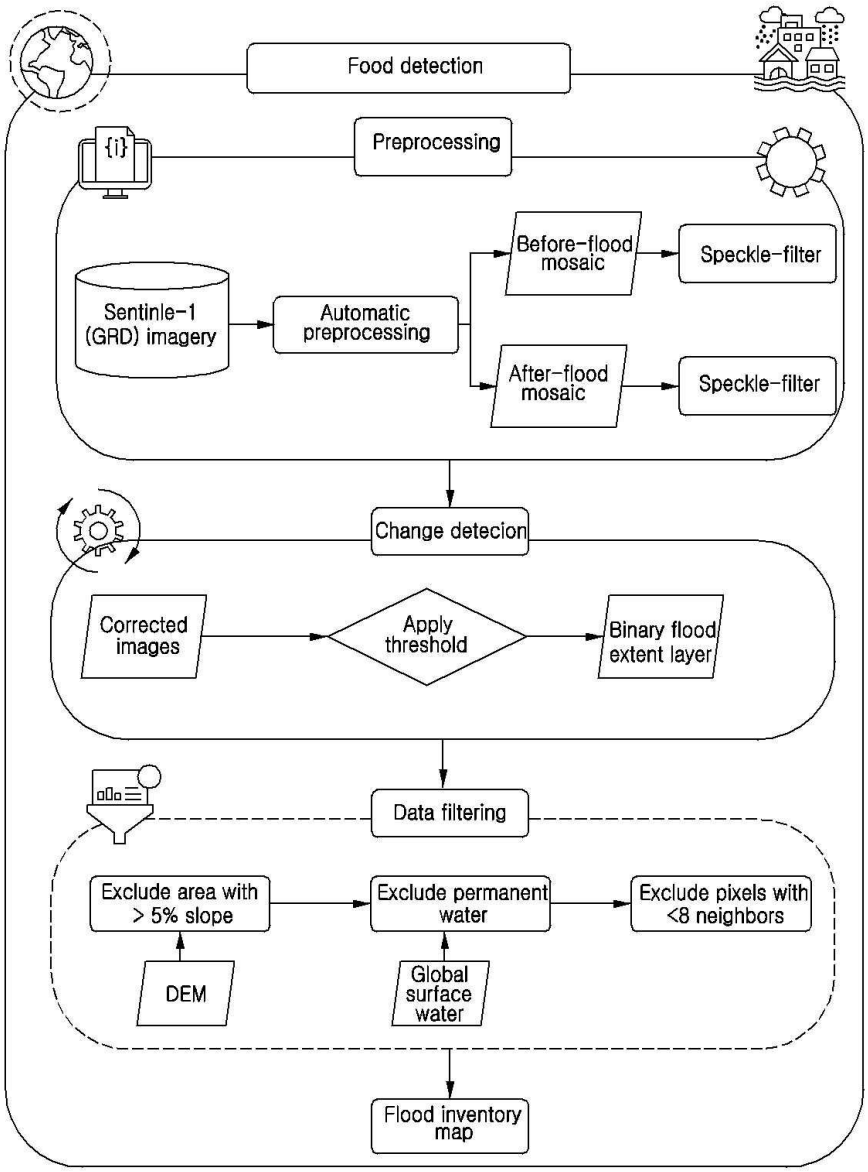
도면2


Flood suseptibility mapping
Spatial database
Flood detection using Sentinle-1
NDVI
Land cover
DEM
Rainfall
Geology
Plan curvature
Slope
Aspect
Altitude
TWI
SPI
GIS
Modelling
Preprocessing
Multicollinearity analysis
Frequency ration (FR)
Feature importance using Geodetector
Parallel ensemble-based machine learning
RF
Bagging
Optimization
Genetic
Validation
Receiver operating characteristic (ROC)
Root mean square error (RMSE)
Taylor diagram
Non parametric test
Seed cell area index (SCAI)
Mapping
FSM

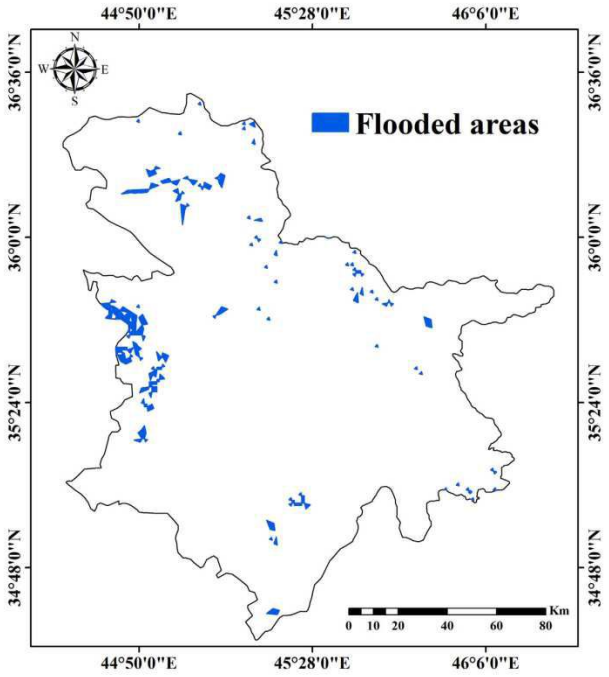
도면3


Meteorological stations
Sulaymaniyah
Iraq
Altitude (m)
High : 3422
Low : 0
Floods

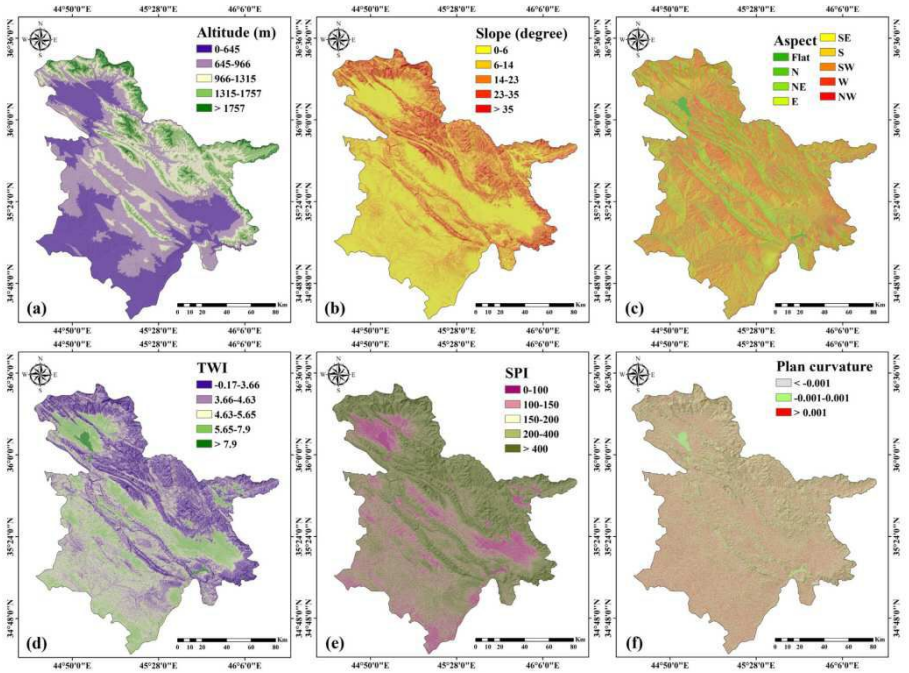
도면4


Food detection
Preprocessing
Sentinle-1 (GRD) imagery
Automatic preprocessing
Before-flood mosaic
Speckle-filter
After-flood mosaic
Speckle-filter
Change detecion
Corrected images
Apply threshold
Binary flood extent layer
Data filtering
Exclude area with > 5% slope
Exclude permanent water
Exclude pixels with <8 neighbors
DEM
Global surface water
Flood inventory map

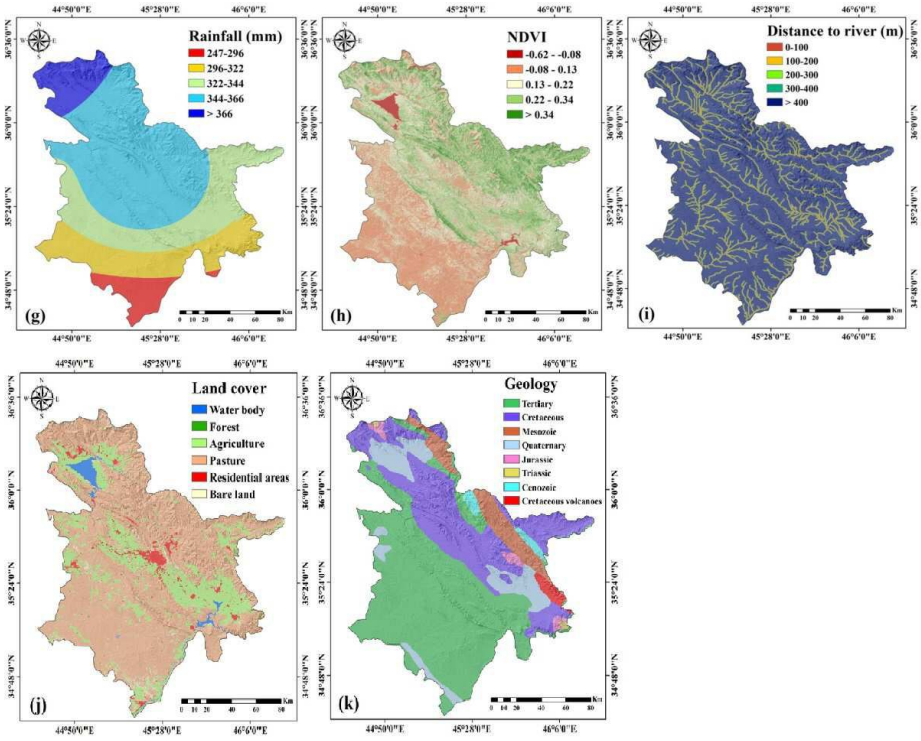
도면5


44°50'0"E
45°28'0"E
46°6'0"E
36°36'0"N
36°0'0"N
35°24'0"N
34°48'0"N
N
W
E
S
Flooded areas
0 10 20 40 60 80
Km

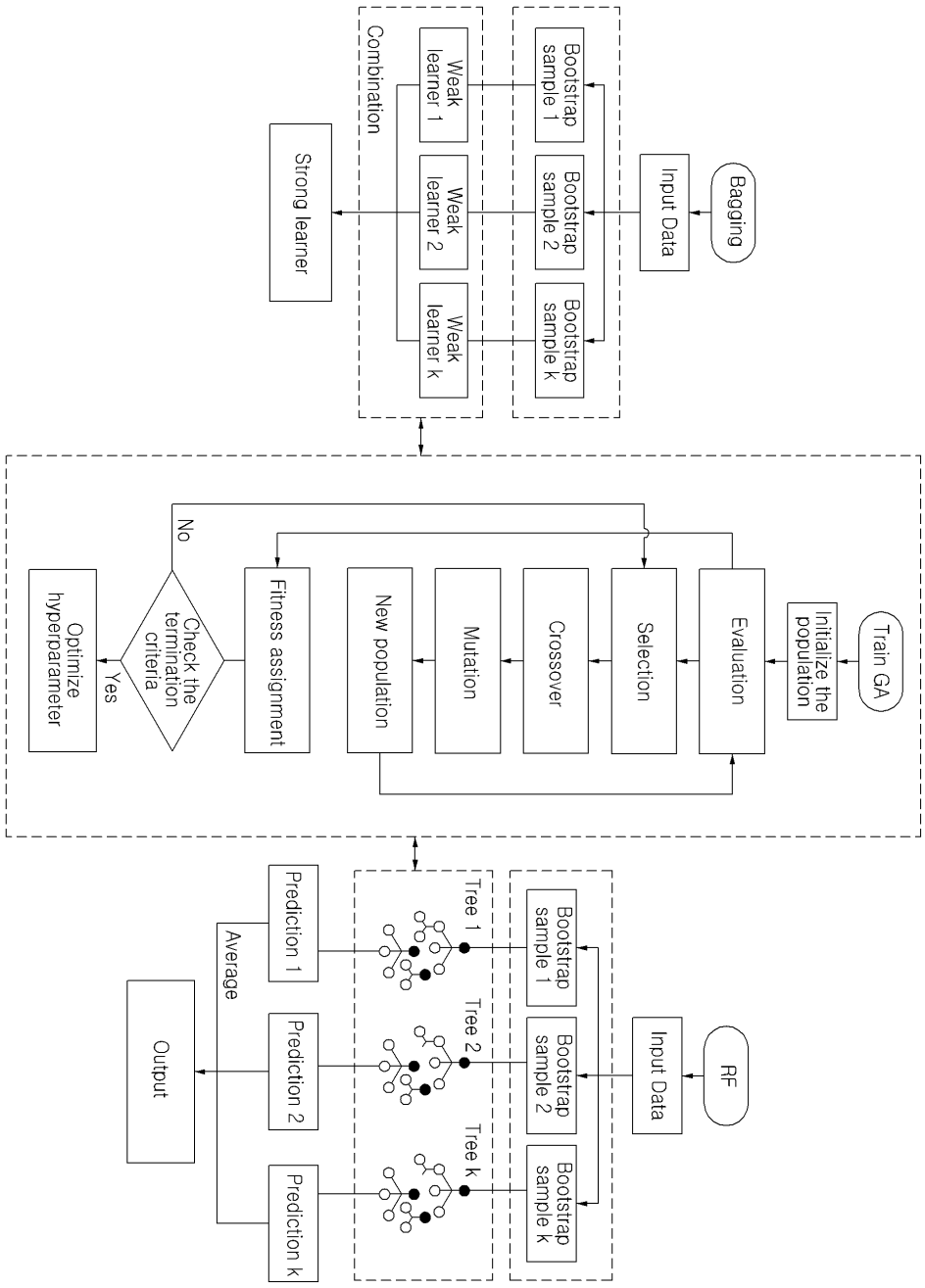
도면6


Altitude (m)
0-645
645-966
966-1315
1315-1757
> 1757
(a)
Slope (degree)
0-6
6-14
14-23
23-35
> 35
(b)
Aspect
Flat
N
NE
E
SE
S
SW
W
NW
(c)
TWI
-0.17-3.66
3.66-4.63
4.63-5.65
5.65-7.9
> 7.9
(d)
SPI
0-100
100-150
150-200
200-400
> 400
(e)
Plan curvature
< -0.001
-0.001-0.001
> 0.001
(f)

도면7


Rainfall (mm)
247-296
296-322
322-344
344-366
> 366
(g)
NDVI
-0.62 - -0.08
-0.08 - 0.13
0.13 - 0.22
0.22 - 0.34
> 0.34
(h)
Distance to river (m)
0-100
100-200
200-300
300-400
> 400
(i)
Land cover
Water body
Forest
Agriculture
Pasture
Residential areas
Bare land
(j)
Geology
Tertiary
Cretaceous
Mesozoic
Quaternary
Jurassic
Triassic
Cenozoic
Cretaceous volcanoes
(k)

도면8


Bagging
Input Data
Bootstrap sample 1
Bootstrap sample 2
Bootstrap sample k
Weak learner 1
Weak learner 2
Weak learner k
Combination
Strong learner
Train GA
Initialize the population
Evaluation
Selection
Crossover
Mutation
New population
Fitness assignment
Check the termination criteria
No
Yes
Optimize hyperparameter
RF
Input Data
Bootstrap sample 1
Bootstrap sample 2
Bootstrap sample k
Tree 1
Tree 2
Tree k
Prediction 1
Prediction 2
Prediction k
Average
Output