



텍스트 마이닝을 통한 우리나라의 벼 도열병 발생 개황 분석

Analysis of Rice Blast Outbreaks in Korea through Text Mining

*Corresponding author

Tel: +82-61-750-5193

E-mail: kitaekim@scnu.ac.kr

ORCID

<https://orcid.org/0000-0001-9816-1723>

<https://orcid.org/0000-0001-7258-6977>

<https://orcid.org/0000-0003-3265-5899>

<https://orcid.org/0000-0002-8249-5283>

송성민¹ · 정현정² · 김광형³ · 김기태^{1,4*}

¹순천대학교 식물의학과, ²국립식량과학원 작물기초기반과,

³서울대학교 농생명공학부, ⁴순천대학교 농생명과학과

Sungmin Song¹, Hyunjung Chung², Kwang-Hyung Kim³, and Ki-Tae Kim^{1,4*}

¹Department of Plant Medicine, Suncheon National University, Suncheon 57922, Korea

²Crop Foundation Research Division, National Institute of Crop Science, Rural Development Administration, Wanju 55365, Korea

³Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea

⁴Department of Agricultural Life Science, Suncheon National University, Suncheon 57922, Korea

Rice blast is a major plant disease that occurs worldwide and significantly reduces rice yields. Rice blast disease occurs periodically in Korea, causing significant socio-economic damage due to the unique status of rice as a major staple crop. A disease outbreak prediction system is required for preventing rice blast disease. Epidemiological investigations of disease outbreaks can aid in decision-making for plant disease management. Currently, plant disease prediction and epidemiological investigations are mainly based on quantitatively measurable, structured data such as crop growth and damage, weather, and other environmental factors. On the other hand, text data related to the occurrence of plant diseases are accumulated along with the structured data. However, epidemiological investigations using these unstructured data have not been conducted. The useful information extracted using unstructured data can be used for more effective plant disease management. This study analyzed news articles related to the rice blast disease through text mining to investigate the years and provinces where rice blast disease occurred most in Korea. Moreover, the average temperature, total precipitation, sunshine hours, and supplied rice varieties in the regions were also analyzed. Through these data, it was estimated that the primary causes of the nationwide outbreak in 2020 and the major outbreak in Jeonbuk region in 2021 were meteorological factors. These results obtained through text mining can be combined with deep learning technology to be used as a tool to investigate the epidemiology of rice blast disease in the future.

Keywords: Korea, *Magnaporthe oryzae*, Outbreak, Rice blast, Text mining

Received July 6, 2022

Revised September 1, 2022

Accepted September 13, 2022

서론

4차 산업혁명에 의한 정보기술의 발달로 인해 식물병역학을

포함한 다양한 농업 분야에서 빅데이터의 활용이 대두되고 있다(Kim과 Lee, 2020). 빅데이터의 정의는 기존의 데이터베이스 도구로 관리가 어려운 방대한 양의 데이터를 의미하는데, 이것은 규칙적이고 값의 의미를 파악하기 쉬운 정형 데이터(structured data)뿐만 아니라, 정해진 규칙이 없어 의미 파악이 어려운 비정형의 데이터(unstructured data)도 포함한다. 식물병역

Research in Plant Disease

eISSN 2233-9191

www.online-rpd.org

© The Korean Society of Plant Pathology

© This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

학 분야에서 정형 데이터의 예로는 각종 센서에 의해 얻어지는 농업 기상데이터(온도, 습도, 강우, 풍속, 일조량 등)와 관찰과 계산으로 얻을 수 있는 식물병 발병률(incidence), 발병도(severity) 등이 있다. 비정형 데이터에 대한 예제의 경우 기존의 분자 유전학적 진단 대비 빠르게 식물병을 동정하기 위한 식물병 이미지 데이터가 딥러닝 및 머신러닝 기법과 함께 활용되고 있다 (Ferentinos, 2018; Mohanty 등, 2016). 이 외의 비정형 데이터로는 텍스트, 비디오, 오디오 같은 멀티미디어 관련 데이터가 대표적이다.

현재까지 우리나라에서 개발된 식물병 예찰 시스템 및 역학, 개황 분석은 전반적으로 작물의 생육상태와 피해정도 측정, 기상 및 시공간과 같은 정형의 시계열 데이터를 기반으로 수행되었다(Kim, 2002; Kim과 Lee, 2020; Lee와 Kim, 2018). 최근 딥러닝을 활용한 인공지능의 발달로 축적된 시계열 데이터를 분석하여 더 정확한 식물병 예찰 모델들이 개발되고 있다(Fenu와 Mallocci, 2021; Kaundal 등, 2006). 한편으로는, 드론 및 위성을 통한 노지 이미지, 카메라를 통한 식물의 병징 이미지 및 현미경을 통한 병원체의 이미지 등 일부 비정형 데이터가 방대한 양으로 축적되고 있다(Ahmad 등, 2022). 이외에도 병 발생과 관련한 뉴스 기사와 논문 등 텍스트 데이터가 많이 축적되었으나, 식물병 관리에 활용하는 데에는 간과되고 있다. 그 이유는 쏟아지는 정보의 홍수 속에서 연구자가 모든 활자 정보의 내용을 파악하기 어렵기 때문이다. 따라서 텍스트 데이터를 식물병 관리에 활용하기 위해서는 필요한 핵심 내용을 선별하고 유의미한 정보를 추출할 수 있는 기술이 필요하다.

텍스트 마이닝(text mining) 기술은 무수한 비정형의 텍스트 데이터에서 패턴 또는 관계를 분석하여 유의미한 정보를 추출함으로써 위와 같은 어려움을 해결할 수 있다. 의학 연구 분야에서는 텍스트 마이닝 기술을 이용하여 유전자와 병 발생의 연관성을 밝히는 데에 일부 적용하고 있다(Pletscher-Frankild 등, 2015; Van Driel 등, 2006). 그러나 식물병역학 분야에서의 활용은 아직 미미하다(Ristaino 등, 2021). 본 연구는 텍스트 마이닝을 통해 우리나라에서 발생하는 도열병 관련 기사들을 조사하고 개황 분석을 수행하였다. 도열병은 식물병원성 곰팡이인 자낭균 *Magnaporthe oryzae*에 의해 화본과 작물에 발생하는 전 세계적으로 가장 중요한 식물병으로(Dean 등, 2012), 여름철 온도가 낮고 비가 많이 올 때 심하게 발생한다. 또한 질소질 비료를 많이 시비하고 감수성 벼 품종을 오랜 시간 재배한 포장에서 심각하게 발생한다. 우리나라에서는 주로 6월 하순에 잎도열병이 발생하고, 7-8월에 이삭도열병이 발생하여 연평균 약 117,000톤의 추정피해수량을 발생시킨다(Kim, 2014). 뉴스 기사에 의하면 도열병을 예방하기 위하여 각 시군에서는 매년 수

천만 원에서 수십억 원의 방제 비용을 지출하고, 병에 의한 수확량 감소로 경제적 손실이 크다(Shim 등, 2003).

본 연구는 인터넷에서 자료를 수집하는 기술인 웹크롤링을 통해 '도열병'이란 단어가 포함된 뉴스 기사들을 수집하였고, 텍스트 마이닝을 통해 기사들의 핵심 내용을 분석하였다. 또한, 기존 문헌과 도열병 피해율, 발생 지역, 기상 및 벼 품종 공급량과 비교하여 우리나라에서 도열병 발생에 대한 종합적인 개황 분석을 수행하였다. 본 연구는 텍스트 마이닝 분석 결과와 기존에 축적된 정형 데이터를 분석하여 도열병 발생에 대한 다각화된 역학 결과를 도출하였다. 해당 결과는 추후 식물병 관리 및 방제를 위한 의사결정에 활용될 수 있을 것이다.

재료 및 방법

기사 자료 수집 및 전처리. 대한민국의 포털 사이트인 네이버에서 '도열병'을 키워드로 크롤링하여 1990년 1월 1일부터 2021년 12월 31일까지 뉴스 기사 데이터를 수집하였다. 수집된 데이터는 날짜, 기사 제목, 언론사, 내용 요약본, 링크, 기사 본문 내용이다. 병사들의 도열과 같이 해당 키워드가 다른 의미로 사용되는 기사들은 큐레이션을 통해 제외했다. 이를 통해 얻은 기사는 총 11,925건으로, 중복되거나 비슷한 내용의 기사들은 내용 요약본으로 딥러닝 자연어처리 기법인 코사인 유사도 측정을 통해 최신 기사이면서 텍스트 유사도가 0.5 이상인 기사들을 제거하였다. 이에 따라, 최종 수집된 기사는 총 5,085건이다 (Supplementary Table 1).

텍스트 마이닝. 텍스트 마이닝은 프로그래밍 언어인 Python의 자연어 처리 패키지인 KoNLPy v0.6.0에서 Open Korean Text (Okt) Class를 사용하여 수행하였다. 전처리로 한자가 포함된 내용은 마이크로소프트 엑셀의 한글/한자 변환 기능을 통해 한글로 치환하였다. 또한, 텍스트 정규화 함수를 통해 문장의 종결어미를 통일하였다. 키워드 '도열병'이 들어간 문장에 대한 핵심어 분석은 문장을 구(phrase)로 나누어 수행하였고, 단어의 동시 출현 빈도에 따른 연관성 분석은 Python의 네트워크 분석 패키지인 NetworkX v2.8을 사용하여 수행하였다. 핵심어 파악을 위해 의미가 없는 형태소(의미를 갖는 최소한의 말 단위)는 큐레이션을 통해 불용어 사전 데이터를 만들어 제거하였다.

도열병 피해율, 기상, 품종 데이터. 도열병 피해율 데이터는 국가농작물병해충관리시스템(National Crop Pest Management System, NCPMS; <https://ncpms.rda.go.kr/>)에서 2014년

부터 2021년까지 매 6월에서 9월까지 논벼 기준으로 제공되는 것을 사용하였다(Lee, 2012). 기후분석을 위한 데이터는 기상청 기상자료개방포털의 1990년에서 2021년까지의 월별 종관기상관측(automated synoptic observing system, ASOS) 데이터와 1990–2020년까지 30년 평년 데이터를 활용하였다(https://data.kma.go.kr/). 2020년 및 2021년 지역별 벼 품종 공급 비율에 대한 분석은 국립종자원의 품종별 생산 공급 현황 데이터 데이터를 사용하였다(https://www.seed.go.kr/).

결과 및 고찰

연도별 도열병 관련 기사와 핵심어 분석. 언론사의 개수가 증가함에 따라 도열병과 관련된 기사도 증가하는 추세를 보였으나($R^2=0.932$) (Supplementary Fig. 1), 약 5–6년 주기로 1993년, 1999년, 2004년, 2009년, 2014년, 2020–2021년에 관련 기사의 빈도수가 극적으로 증가하였다(Fig. 1A). 해당 연도들의 기

사 개수는 이동 평균보다 우위에 있고, 문헌에 기록된 우리나라에서 도열병이 대발생한 시기와 유사하다. 1993년에 경상남도 지역에서 잎도열병과 이삭도열병 발생면적은 평년 대비 각각 2.6배, 3.5배 증가하였다(Kang과 Kim, 1994). 1999년부터 2008년까지 조사된 우리나라 벼 도열병 발생 개황에 따르면(Lee 등, 2010), 잎도열병의 평균 병반면적률은 1999년과 2000년 0.28%와 0.84%로 조사되었고, 2004년에도 0.13%로 다른 연도에 비해 높았다. 반면, 2002년, 2003년, 2005–2008년에는 잎도열병 병반면적률이 0.1% 미만으로 거의 발생하지 않았다. 이삭도열병의 경우 이병수율을 기준으로 1999년과 2000년에 각각 1.67%, 2.36%로 높은 발병률을 보이다가 2001년, 2002년 발병률이 감소하였다. 이후 2003년, 2005년 이삭도열병은 다시 2.95%, 2.02%로 발병률이 증가하였으나 점차 감소하는 경향을 보였다. 2014년에는 경상남도 일부 지역과 전라남도에서 재배면적 대비 이삭도열병이 각각 11%, 14.6%로 다른 지역에 비교해 다발생하였고(Kang 등, 2019), 이는 NCPMS에 2014년 9월

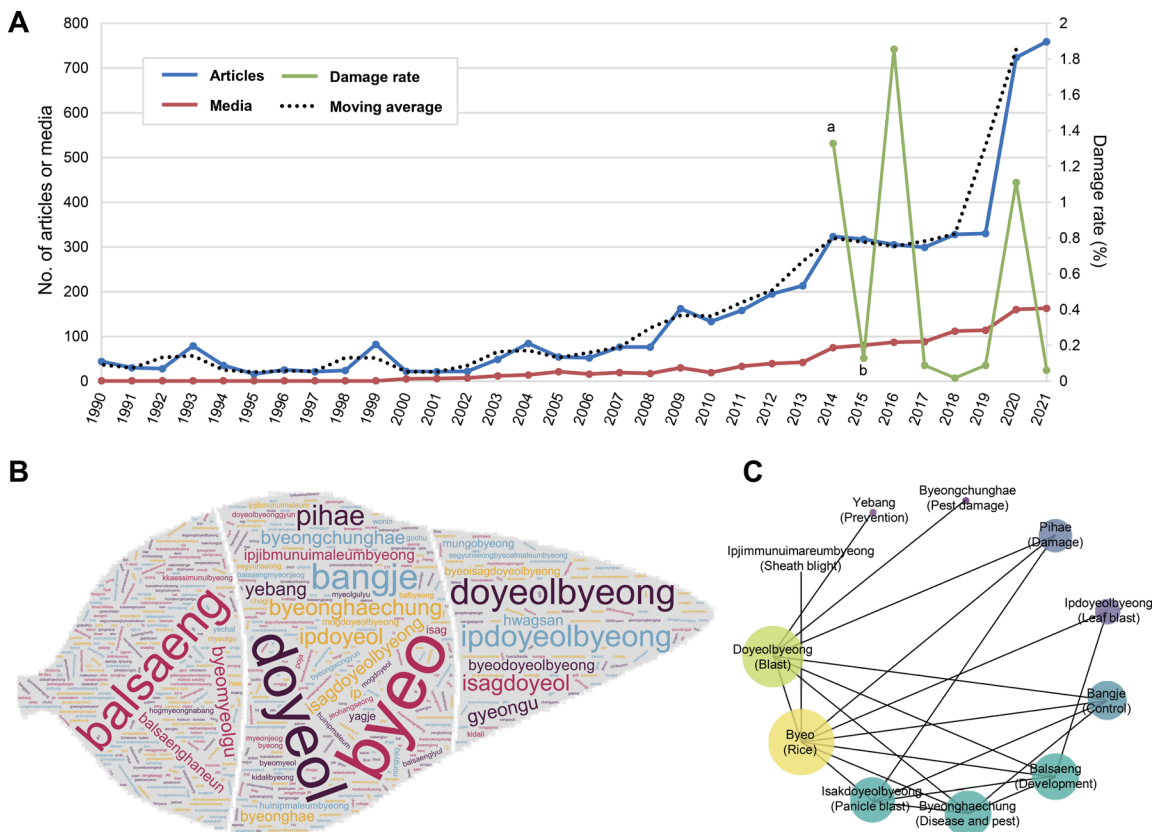


Fig. 1. Number of rice blast related articles published from 1990 to 2021 and the major keywords of articles. (A) The solid lines represent number of articles (blue) and press agencies (red) in each year. The dotted line represents moving average for the number of articles published each year (black). Sum of nation-wide rice blast damage rate from 2014 to 2021 is provided by National Crop Pest Management System (NCPMS) (green). ^aOnly September data for rice panicle blast was available. ^bOnly June to August rice leaf blast data was available. (B) A word cloud constructed from the phrases including the word “doyeolbyeong (rice blast)” from all collected articles. The Korean words are written in Roman characters. (C) A co-occurrence network of words from the phrases. The Korean words are written in Roman characters and its English meaning is in parenthesis.

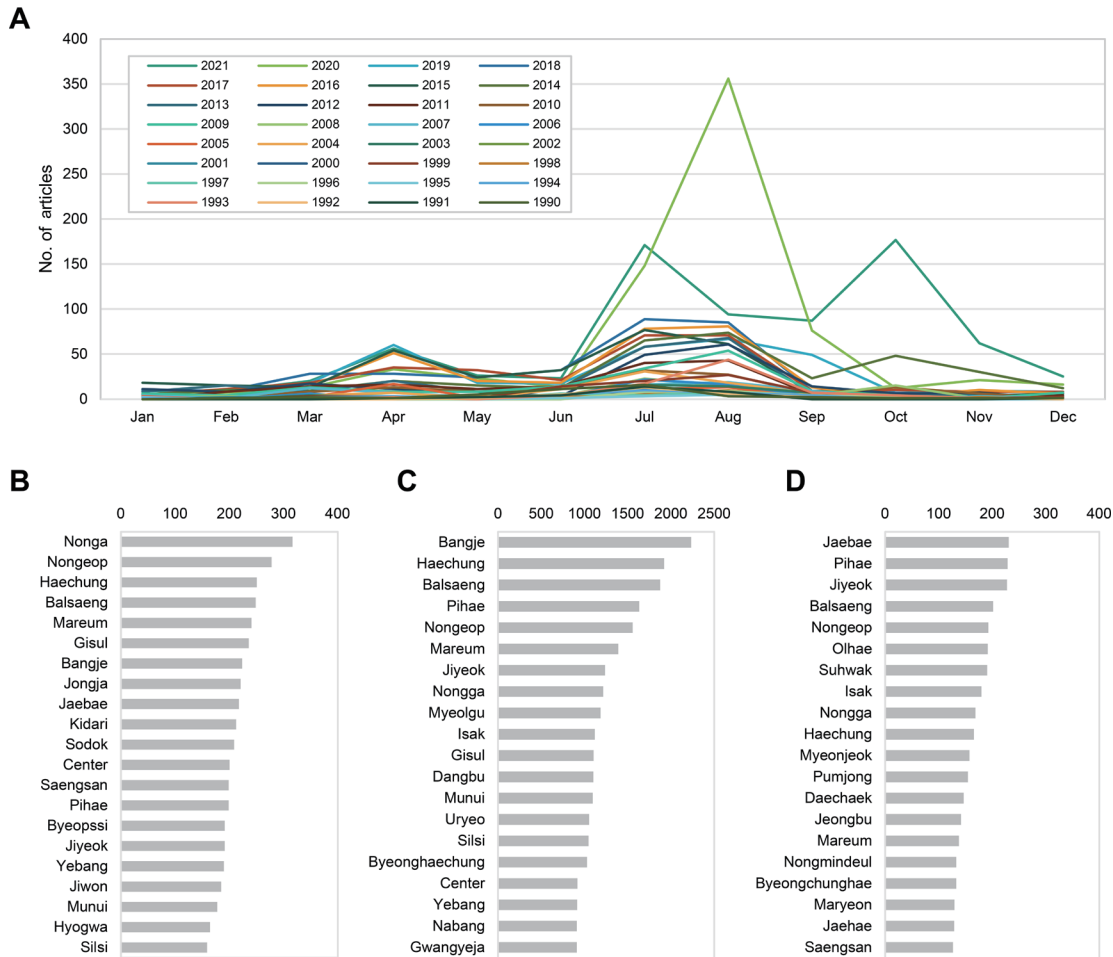


Fig. 2. Results derived from text mining of the articles. (A) Number of articles published in each month from 1990 to 2021. Each year is color-coded. (B) The top twenty most frequently appeared keywords in April, (C) July and August, and (D) October. The Korean words are written in Roman characters.

16일에 집계된 이삭도열병 피해율과 유사하다(Supplementary Fig. 2). 2016년의 경우, 기사 개수는 이동 평균과 크게 차이나지 않았지만, NCPMS에 집계된 도열병 합계 피해율은 1.855%로 다른 해에 비교해 높게 나타났다(평균 합계 피해율 0.067%) (Fig. 1A). 그 이유는 2016년 6월 1일 특정 지역인 부산(0.325%), 대구(0.45%), 광주(0.86%)에서 집계된 도열병 피해율이 NCPMS에 높게 기록되어있기 때문이다(Supplementary Fig. 2). 그러나, 2016년 도열병이 다발생 했다는 학술적 근거는 찾을 수 없었다. 2020년 이후에는 도열병 관련 기사가 눈에 띄게 증가했는데 (Fig. 1A), 이는 우리나라에서 전국적인 도열병 다발생과 연관 있어 보인다(Chung 등, 2022). 실제로 2020년의 NCPMS 예찰 결과에 의하면 전국의 도열병 합계 피해율(damage rate)은 1.11%로 평균보다 높았다. 뉴스 기사의 경우 일반 농가 또는 지역에서 보고된 도열병 발생 여부를 전달하는 반면, NCPMS는 관찰포에서 포장검정법에 의해 계산된 피해율 수치를 제공해주기 때문에 이들의 직접적인 비교는 불가하지만, 거시적 관점에서 우리

나라 도열병 발생 개황을 판단할 수 있는 지표가 될 수 있다.

기사에서 ‘도열병’이 사용된 의미와 전체적인 주제를 파악하기 위해 ‘도열병’이 사용된 구를 추출하였고, 띄어쓰기로 분해하여 조사를 제거 후 워드클라우드를 생성하였다(Fig. 1B). 가장 빈도가 높은 단어는 ‘도열병’과 ‘벼’이며, 이삭도열병, 잎도열병, 목도열병과 같은 병 관련 파생어도 높은 빈도로 출현하였다. 그 외 워드클라우드를 통해 도열병과 관련된 다양한 내용을 파악할 수 있다. 빈도수가 높은 순서대로 ‘도열병’과 같이 사용된 단어는 ‘벼’, ‘병해충’, ‘발생’, ‘방제’, ‘피해’, ‘예방’이고, 이는 기사의 내용이 병 발생 및 피해 상황 보고 또는 방제 및 예방에 관한 내용을 나타낸다고 할 수 있다(Fig. 1C). ‘벼’의 경우 이삭도열병과 잎도열병 둘 다 연결되어 있지만, ‘피해’가 이삭도열병과 연결되어 있어, 이삭도열병에 의한 피해가 더 크다는 것을 알 수 있다. 또한, 벼와 또 다른 곰팡이병인 ‘잎집무늬마름병’도 도열병과 함께 빈번히 발생한다는 것을 나타낸다.

월별 기사의 빈도수를 조사했을 때, 4월에 빈도수가 소폭 증

가하고, 도열병이 많이 발생하는 시기인 매해 7월, 8월에 증가하며, 2014년과 2021년 10월에 기사의 빈도가 높은 것이 관찰되었다(Fig. 2A). 4월에 빈도수가 올라가는 원인을 핵심어를 통해 파악한 결과, 종자 소독을 통해 병 발생 예방을 선전하는 기사가 이유이고(Fig. 2B), 7-8월에는 도열병 및 멸구, 나방 등 해충 방제 권고, 발생 보고에 관한 내용이 많기 때문이다(Fig. 2C). 2020년 8월에는 유난히 기사의 빈도가 높는데, 그 이유는 후술할 전국적인 도열병 대발생이 원인으로 나타난다. 10월에는 농민의 피해에 대한 정부의 방안에 대한 내용 때문인 것으로 보인다(Fig. 2D). 식물병 관리에 텍스트 데이터를 활용하기 위해서는 뉴스의 내용이 분류되어야 한다. 분류는 크게 세 가지 범주로 (1) 방제 당부 또는 권고에 대한 내용(미래), (2) 병 발생 보고에 대한 내용(과거), (3) 정치, 사회, 경제, 문화 등 기타 내용이 될 수 있다. 본 연구에서는 병 발생 보고에 대한 내용을 분리 후, 이를 기반으로 우리나라에서 도열병 발생에 대한 지리적 개황 조사를 수행하였다.

도열병 발생 개황 및 역학 조사. 우리나라의 도열병 발생 개황을 조사하고자 도열병이 대발생한 연도별(1993, 2014, 2020, 2021) 기사에서 도열병 발생 지역과 상황에 대해 텍스트 마이닝 분석을 수행하였다. 1993년 기사에서 처음으로 잎도열병이 보고된 지역은 전라북도 부안, 고창, 진안으로 6월 24일경 발표되었다(Fig. 3A). 7월 초 전라남도 여천(현 여수) 및 인근의 경상남도 사천, 거제에서 잎도열병이 발견되었고, 중순쯤 거창, 하동에서도 보고되었으며, 8월에는 경상북도와 전라남도에서의 발생도 보고되었다. 이 결과는 1993년 경상남도에서의 7월 도열병 대발생 개황과 일치한다(Kang과 Kim, 1994). 2014년에는 8월 중순 이삭도열병이 경상남도에서 발생하기 시작하여 9월에는 전라남도에서도 발생한 것으로 나타났고(Fig. 3B), 이 또한 기존 결과와 일치한다(Kang 등, 2019). 2020년에는 우리나라 전국에서 7-9월에 걸쳐 도열병이 보고되었다(Fig. 3C). 7월 중하순경 경상북도 상주, 안동 등에서 기사가 나오기 시작하여, 7월 말에는 충청북도 청주, 경상남도 함안, 경기도 여주, 경상북도 영천 등에서 보고되었다. 특히, 7월 말 경기도 여주에서 발생한 도열병은 적절한 기후 조건, 감수성 품종의 재배, 질소 비료의 과다 시비에 의한 것으로 추정되었다(Chung 등, 2022). 8월에는 전라남도와 서해에 인접한 지역에서 보고되었으며, 9월에는 강원도 철원에서 보고된 사례도 있다. 2021년에는 7월초 경상북도 영덕에서 발생한 보고가 있었고, 중하순에 안동, 상주, 칠곡에서도 나타났으며, 8-9월에는 부안, 고창, 정읍, 군산, 익산 김제 등 주로 전라북도에서 도열병 피해가 보고되었다(Fig. 3D).

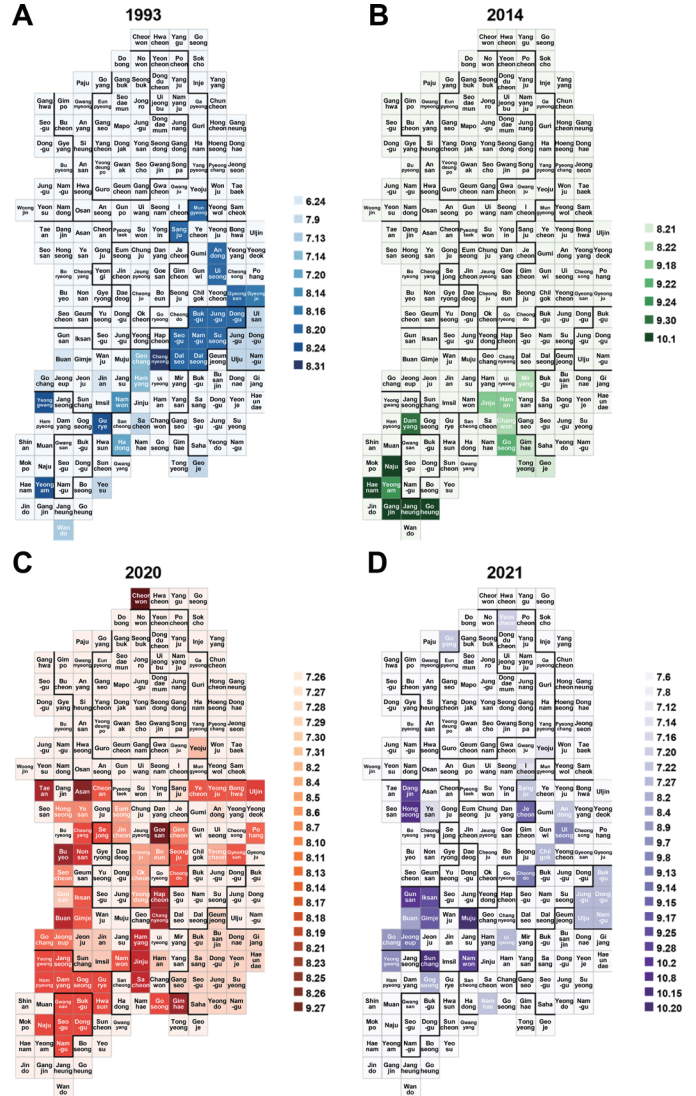


Fig. 3. Rice blast epidemic incidences in the provinces of Korea. (A) Epidemics in 1993 (blue). (B) Epidemics in 2014 (green). (C) Epidemics in 2020 (orange). (D) Epidemics in 2021 (purple). The dates of article published are color-coded.

2020년 및 2021년 도열병 발생 지역의 기상 조사. 1993년 경상남도의 도열병 대발생 원인은 일조 부족, 저온, 다우 등 기상요인이 크게 작용했으며, 감수성 품종의 재배면적 증가가 원인으로 밝혀졌다(Kang과 Kim, 1994). 1999년과 2000년 잎도열병의 발생 원인으로서는 7월의 평소보다 많은 강수량과 낮은 일조량에 기인한 것으로 추정되었고, 이삭도열병의 원인으로서는 8월의 많은 강수량과 9월의 높은 강수일수 및 낮은 일조량으로 추정되었다(Lee 등, 2010). 2014년의 남부지방에서의 이삭도열병 발생 원인은 평년 대비 낮은 기온, 많은 강수량, 적은 일조량으로 인해 발병에 좋은 조건이 갖춰졌기 때문으로 나타났다(Kang 등, 2019).

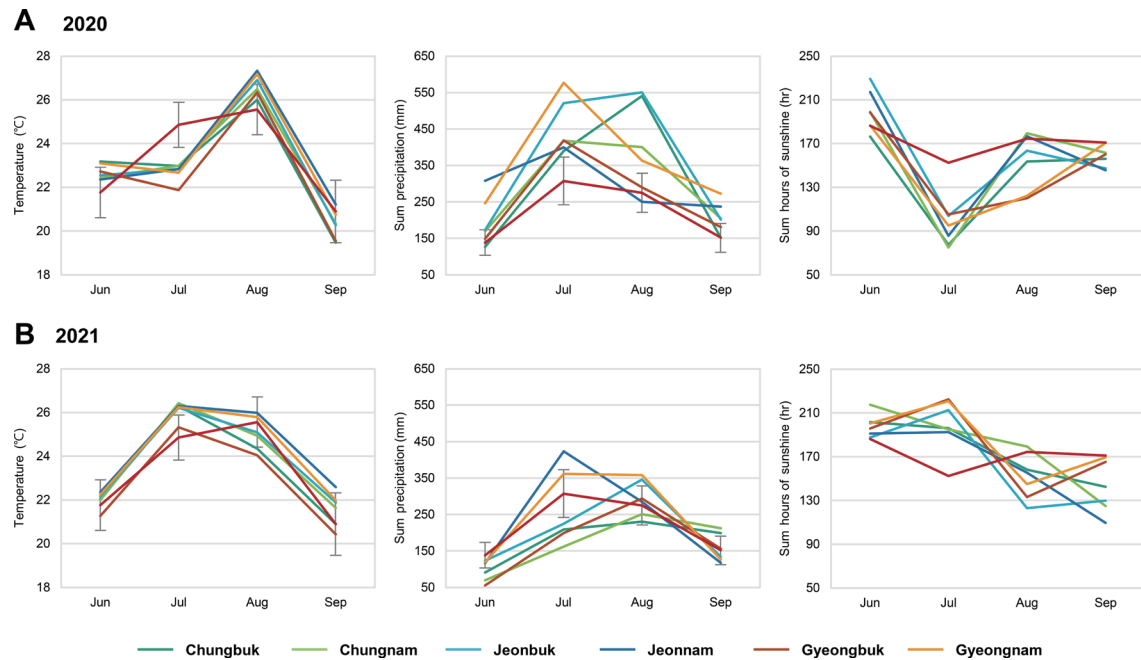


Fig. 4. Possible climate factors affecting the rice blast incidences through June to September in 2020 (A) and 2021 (B). Climate factors include average temperature ($^{\circ}\text{C}$) (left), sum precipitation (mm) (middle), and sum hours of sunshine (hr) (right). The red line represents nation-wide, climatological average temperature and sum precipitation for each month, and the other lines represent provinces in Korea.

2020년 우리나라 전역과 2021년 전라북도 지역에서 도열병이 대발생한 원인을 찾기 위해 기상 요인(평균 기온, 합계 강수량, 합계 일조시간)을 분석하였다(Fig. 4A, B). 우리나라의 1990–2020년까지 6–9월 전국 평균 월평균 기온은 각각 21.8°C , 24.9°C , 25.6°C , 20.9°C 로 나타났고, 평년 월 강수량은 각각 138.5 mm, 307.7 mm, 275.2 mm, 151.7 mm로 나타났으며, 일조시간은 186.3시간, 152.6시간, 174.4시간, 171.0시간으로 나타났다(평년 일조시간의 경우 기온과 강수량보다 적은 지점의 ASOS에서 관측되었다). 조사된 지역(충청북도, 충청남도, 전라북도, 전라남도, 경상북도, 경상남도)의 2020년 6월, 8월, 9월 기온은 대부분 평년 표준편차 범위 안에 속해있지만, 7월의 경우 표준편차 범위 이하로 낮았고, 7–9월의 일조시간도 평년보다 낮은 것으로 나타났다(Fig. 4A). 모든 지역에서 7월 일조시간이 평년 대비 약 50시간 정도 줄어들었다. 반면 강수량의 경우 대부분의 지역에서 6–9월 평년보다 많은 양의 비가 내렸고, 특히 경상남도와 전라북도에는 7월, 충청북도에 8월에 500 mm 이상의 강수량을 보여주었다(Fig. 4A). 따라서, 2020년의 전국적 도열병 대발생에는 평년보다 서늘했던 7월 기온, 전국적으로 높은 강수량, 낮은 일조시간이 주로 기여했을 것으로 추정된다(Fig. 3C).

2021년 6–9월 각 지방의 평균 기온은 2020년 대비 평년과 비슷한 양상을 보였으나, 7월 도열병이 발생 보고가 많았던 경상북도는 타지방보다 더 낮은 평균 기온을 보였다(Figs. 3D, 4B). 강수량의 경우 2020년보다 전반적으로 낮았으나, 7월 전라남도

와 8월 경상남도, 전라북도에서 평년의 표준편차보다 높게 관측되었다(Fig. 4B). 9월에는 충청도에서 평년의 표준편차보다 높은 강수량이 기록되었다(Fig. 4B). 일조시간의 경우 2020년과 반대로 6–7월 모든 지역에서 평년보다 높았지만, 8–9월에는 작년과 같이 대부분 지역이 평년보다 낮았고, 특히 8월 전라북도의 일조시간이 가장 낮았다. 따라서, 2021년 7월 경상북도에서 도열병 다발생의 기상학적 요인은 낮은 온도 및 일조시간으로 추정되고, 8–9월 전라북도의 다발생은 8월의 높은 강수량과 낮은 일조시간과 연관이 있어 보인다. 그러나 같은 전라북도 내에서도 도열병 발생이 보고된 곳과 아닌 곳이 있으므로, 이에 대한 기상학적 요인을 조사하였다.

전라북도의 경우 군산, 부안, 고창 등 서해안 지방에서 도열병에 의한 피해가 기사화됐지만, 전주, 임실, 장수 등 내륙에서는 기사화되지 않았다(Fig. 3D). 해당 차이의 조사를 위해 기상요인을 분석한 결과(Supplementary Fig. 3), 서해안 지방의 기온은 전라북도 평균 기온의 표준편차 안에 있었지만, 내륙(전주, 임실, 장수)은 범위 밖으로 조사되었다. 합계 강수량의 경우 8월에 7월 강수량이 많았던 군산을 제외한 타 전라북도 지역에서 전반적으로 높았다. 그러나, 9월에는 전주를 제외한 임실과 장수는 전라북도 평균보다 낮았다. 8–9월 일조시간의 경우 전라북도가 다른 도에 비해 전반적으로 낮았음에도 불구하고, 오히려 도열병이 기사화되지 않은 임실과 장수는 다른 지역에 비교해 낮은 수치를 보여주었다. 이를 종합하면 전라북도 내에서 도열병 발생이 보고되지

Table 1. Characteristics of government supplied cultivars in 2021, adapted from the Korea Seed and Variety Service (2020)

Maturation period	Variety name	Development year	Height (cm)	Yield (kg/10a)	Rice blast resistance index	Adapted region
Early -Maturing	Odae	1983	77	460	M	North central mountains, northern inland plains, south central highlands, and north central east coast
	Ungwang	2004	60	586	S	North central midlands, southern mountains, and highlands
	Haedamssal	2014	67	548	S	Yeongnam plains
Medium -Maturing	Haiami	2008	83	538	M	Central plains and central west coast
	Haepum	2013	73	526	M	Plains below Chungnam and southwest coast
Mid-Late -Maturing	Mipum	2010	77	564	W	Inland plains below Chungnam
	Samgwang	2003	87	569	MS	Central plains and southern midlands
	Saenuli	2007	78	571	M	Plains below Chungnam and southwest coast
	Saeilmi	2011	77	585	MS	Coast and plains below south central Chungnam and northern Gyeongnam plains
	Saecheongmu ^a	2017	79	560	MS	Jeonnam
	Sugwang	2011	89	549	MS	Plains below Chungnam
	Sindongjin	1999	80	596	MW	Inland plains below Chungnam
	Yeonghojinmi	2009	70	544	W	Southern plains
	Ilmi	1995	79	522	M	Southern plains
	Ilpum	1990	79	534	W	Central plains and southern inland midlands
	Jinsumi ^a	2008	85	555	S	Yeongnam inland plains, southern Gyeonggi, and central Chungbuk plains
	Chucheong	1970	84	547	W	Central plains and southern midlands
	Chindeul	2012	83	561	M	Southern west coast and plains below Pyeongtek
Dongjinchal	1998	80	549	W	Plains below Daejeon	
Baegogchal	2008	81	526	M	Yeongnam, Honam, and central inland plains	

Note that height and yield can be different due to agricultural practice and environmental factors.

Rice blast resistance index: weak (W), medium (M), strong (S).

^aNewly supplied variety in 2021.

않은 곳의 기상요인은 보고된 지역과 비교해 전라북도 평균의 표준편차 범위 밖에 있었다. 이는 해당 지역들의 기상요인이 도열병 발생에 적합하지 않았다는 것을 나타낸다.

조사된 지역들의 2021년 기상요인은 2020년과 비교하여 평년과 더 유사했음에도 전라북도 서해안과 경상북도에서 집중적으로 도열병 발생 보고가 많았다는 것은 기상조건 이외에 병 발생에 영향을 주는 요인이 존재할 수 있다는 것을 시사한다. 따라서, 지역별 공급된 벼 품종을 조사하였다.

2020년과 2021년 지역별 공급된 벼 품종 조사. 동일 벼 품

종의 지속적인 재배는 벼 품종의 저항성 붕괴를 야기하여 다 발생의 원인이 될 수 있다(Baek 등, 2021). 기상요인 외에 2021년 경상북도 일부와 전라북도에서 집중적으로 도열병이 발생한 원인을 다각적으로 추측하기 위하여 국립종자원의 정부 보급종 품종특성표에서 제공하는 2020년 및 2021년 지역별 벼 종자 공급량을 분석하였다(Table 1, Fig. 5). 조사된 지역에서 두 해 모두 가장 많이 공급된 벼 품종은 중만생종으로 출수기가 늦어 잎도열병을 회피할 수 있는 장점이 있다(Shim 등, 2003). 그러나 품종특성표에 고시된 중만생종들의 도열병 저항성 지수는 2021년 새로 보급된 진수미(strong, S)를 제외하고 모

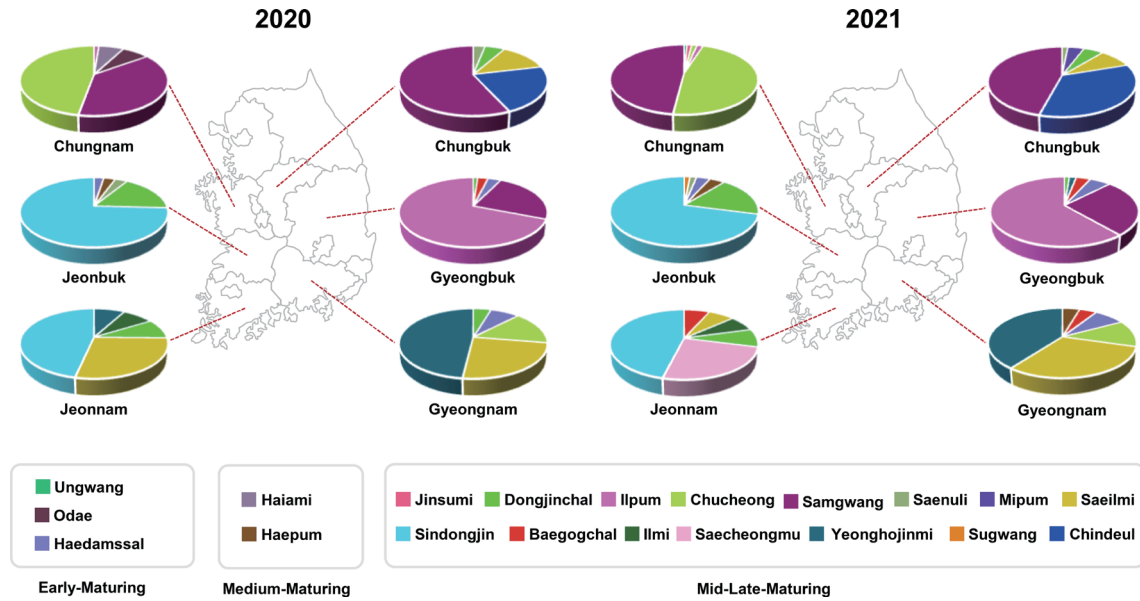


Fig. 5. Proportion of the top five rice varieties supplied to the provinces of Korea in 2020 and 2021.

두 중강(medium-strong, MS) 이하의 저항성을 갖는다(Table 1). 충청북도의 경우 2020년에 순서대로 삼광벼(MS)와 친들벼 (medium, M)가 가장 많이 보급되었고, 2021년에는 친들벼(M)의 공급량이 크게 증가하였다. 충청남도에서는 2020년 추청벼 (weak, W)가 삼광벼(MS)보다 더 많이 보급되었으나, 2021년에는 반대의 공급량을 보였다. 경상북도는 두 해 모두 일품벼(W)와 삼광벼(MS)가 비슷한 비율로 보급되었고, 경상남도는 2020년 영호진미벼(W)와 새일미벼(MS)가 보급되었지만 2021년 영호진미벼가 줄고 새일미벼의 공급량이 증가하였다. 전라남도는 2020년의 경우 신동진벼(medium-weak, MW)와 새일미벼 (MS)가 보급되었지만, 2021년에는 새일미벼(MS) 대신 새청무벼 (MS)가 보급되었다. 전라북도의 경우 2020년과 2021년 모두 신동진벼(MW)가 약 75%의 비율로 보급되었다. 다른 지방의 경우 2020년 대비 2021년 품종의 비율이 줄어들거나 교체된 반면 병 발생이 가장 많이 보고된 전라북도와 경상북도는 두 해 모두 비슷하게 유지되었다. 연작이 직접적인 도열병 다발생 원인이 아닐 수 있지만, 지속적인 재배에 의한 새로운 도열병균 레이스의 출현이 다발생에 기인할 수 있다(Han 등, 2001). 따라서, 도열병균 레이스에 대한 분포 조사와 현재 재배되고 있는 주요 품종의 저항성 검정이 시급하다. 또한 질소질 비료의 과용도 대부분 벼 품종의 도열병 발생에 기여할 수 있기 때문에 시비량에 대한 조사 및 통계치도 필요한 실정이다.

요 약

벼 도열병은 전 세계적으로 발병하여 쌀 수확량을 크게 감소

시키는 주요 식물병이다. 벼 도열병은 한국에서도 주기적으로 대발생하여 사회경제적으로 큰 피해를 입힌다. 이를 예방하기 위해서는 병 발생 예측 시스템이 필요하다. 또한 병 발생에 대한 역학 조사는 식물병 관리를 위한 의사결정을 내릴 때 도움을 줄 수 있다. 현재 도열병 예측 및 역학 조사는 주로 작물의 생육량, 기상 환경 요인 등 정량적으로 측정 가능한 정형 데이터를 기반으로 수행되고 있다. 정형 데이터와 함께 도열병 발생과 관련한 텍스트 자료들도 많이 축적되고 있다. 그러나 이러한 비정형 데이터를 이용한 역학 조사는 이루어지지 않고 있다. 비정형 데이터를 활용하여 유용한 정보를 추출한다면 도열병을 포함한 앞으로의 식물병 관리에 사용할 수 있을 것이다. 이 연구는 텍스트 마이닝을 통해 도열병 관련 뉴스 기사를 분석하여 우리나라에서 벼 도열병이 다발생한 연도와 지역을 조사하였고, 해당 지역의 평균 기온, 합계 강수량, 일조시간, 공급된 벼 품종을 분석하였다. 이를 통해 2020년 평년에 비해 낮은 기온과 일조시간 및 높은 강수량이 전국적인 도열병 다발생의 원인에 기여했고, 2021년 전라북도와 경상북도 일부 지역의 다발생은 비슷한 기상학적 요인에 의한 것으로 추측할 수 있었다. 더하여 같은 벼 품종의 연작에 의한 도열병 다발생 가능성과 질소 비료의 시비량이 병 발생에 미치는 영향에 대한 추후 연구가 필요하다. 결론적으로, 쏟아지는 정보의 홍수속에서 관련 기사를 종합적으로 보기 어렵다. 따라서, 텍스트 마이닝을 통해 얻은 결과로 특정 키워드들이 많이 관찰될 때 적극적 방제에 대한 의사결정을 할 수 있는 시스템이 구축될 필요가 있다. 이는 추후 딥러닝 기술과 접목되어 벼 도열병 역학 조사 도구로 사용될 수도 있을 것이다. 텍스트 마이닝을 통해 얻은 유의미한 정보를 기존의 정

형 데이터 기반의 모델과 결합한다면 농업현장에서 병발생 예측 또는 방제기술 개선에 필요한 고품질 정보를 제공해줄 수 있을 것이라고 예상된다.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by a Research promotion program of SCNU.

Electronic Supplementary Material

Supplementary materials are available at Research in Plant Disease website (<http://www.online-rpd.org/>) and Figshare (<https://doi.org/10.6084/m9.figshare.21200551.v1>).

References

- Ahmad, A., Saraswat, D. and El Gamal, A. 2022. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agric. Technol.* 3: 100083.
- Baek, M.-K., Park, H.-S., Lee, C.-M., Lee, H.-J., Jeong, J.-M., Ahn, E.-K. et al. 2021. Identification of stable resistance genes based on resistance evaluation to blast for monogenic lines and leading Japonica varieties in rice. *Korean J. Breed. Sci.* 53: 217-229. (In Korean)
- Chung, H., Jeong, D. G., Lee, J.-H., Kang, I. J., Shim, H.-K., An C. J. et al. 2022. Outbreak of rice blast disease at Yeosu of Korea in 2020. *Plant Pathol. J.* 38: 46-51.
- Dean, R., Van Kan, J. A. L., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D. et al. 2012. The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13:414-430.
- Fenu, G. and Mallocci, F. M. 2021. Forecasting plant and crop disease: an explorative study on current algorithms. *Big Data Cogn. Comput.* 5: 2.
- Ferentinos, K. P. 2018. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145:311-318.
- Han, S.-S., Ryu, J. D., Shim, H.-S., Lee, S.-W., Hong, Y.-K. and Cha, K.-H. 2001. Breakdown of resistance of rice cultivars by new race KI-1117a and race distribution of rice blast fungus during 1999~2000 in Korea. *Res. Plant Dis.* 7: 86-92. (In Korean)
- Kang, S. W. and Kim, H. K. 1994. Factors affecting unusually severe outbreak of rice blast in Gyeongnam province in 1993. *Korean J. Plant Pathol.* 10: 78-82. (In Korean)
- Kang, W. S., Seo, M.-C., Hong, S. J., Lee, K. J. and Lee, Y. H. 2019. Outbreak of rice panicle blast in southern provinces of Korea in 2014. *Res. Plant Dis.* 25: 196-204. (In Korean)
- Kaundal, R., Kapoor, A. S. and Raghava, G. P. S. 2006. Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC Bioinformatics* 7: 485.
- Kim C.-H. 2002. Current status and future prospect of plant disease forecasting system in Korea. *Res. Plant Dis.* 8: 84-91. (In Korean)
- Kim, K.-H. and Lee, J. 2020. Smart plant disease management using agrometeorological big data. *Res. Plant Dis.* 26: 121-133. (In Korean)
- Kim, Y. 2014. Investment analysis of insects and diseases prevention of rice in public using a real option approach. Ph.D. thesis. Seoul National University, Seoul, Korea. 59 pp. (In Korean)
- Korea Seed and Variety Service. 2020. List of variety characteristics for government-supplied species in 2021. URL <https://www.seed.go.kr/> [20 May 2022].
- Lee, S. and Kim, K.-H. 2018. Predicting potential epidemics of rice leaf blast disease using climate scenarios from the best global climate model selected for individual agro-climatic zones in Korea. *J. Clim. Change Res.* 9: 133-142. (In Korean)
- Lee, Y.-H. 2012. Pest monitoring, prediction: one-stop processing up to diagnosis. *Life Agrochem.* 277: 22-25. (In Korean)
- Lee, Y. H., Ra, D.-S., Yeh, W.-H., Choi, H.-W., Myung I.-S., Lee, S.-W. et al. 2010. Survey of major disease incidence of rice in Korea during 1999-2008. *Res. Plant Dis.* 16: 183-190. (In Korean)
- Mohanty, S. P., Hughes, D. P. and Salathé, M. 2016. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7: 1419.
- Pletscher-Frankild, S., Pallegà, A., Tsafou, K., Binder, J. X. and Jensen, L. J. 2015. DISEASES: text mining and data integration of disease-gene associations. *Methods* 74: 83-89.
- Ristaino, J. B., Anderson, P. K., Bebbler, D. P., Brauman, K. A., Cunniffe N. J., Fedoroff, N. V. et al. 2021. The persistent threat of emerging plant disease pandemics to global food security. *Proc. Natl. Acad. Sci. U. S. A.* 118: e2022239118.
- Shim, H.-S., Kim, Y.-K., Hong, S.-J., Han, S.-S. and Sung, J.-M. 2003. Assessments of yield and quality of rice affected by rice panicle blast. *Res. Plant Dis.* 9: 183-188. (In Korean)
- Van Driel, M. A., Bruggeman, J., Vriend, G., Brunner, H. G. and Leunissen, J. A. 2006. A text-mining analysis of the human phenome. *Eur. J. Hum. Genet.* 14: 535-542.