

REVIEW ARTICLE

Harnessing machine learning approach for phenotyping and deciphering the plants biotic and abiotic stresses responsive molecular mechanisms

Bahman Panahi^{1*}, Bentolhoda Ghavidel², Poya Shahgoli³

¹Department of Genomics, Branch for Northwest & West region, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Tabriz, Iran.

²Department of Biotechnology, Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran.

³Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran.

Correspondence

Bahman Panahi

Email: panahi.lahroodi@gmail.com

How to cite

Panahi, B., Ghavidel, B., & Shahgoli, P. (2024). Harnessing machine learning approach for phenotyping and deciphering the plants biotic and abiotic stresses responsive molecular mechanisms. *Crop Biotechnology*, 13(46), 75-90.

ABSTRACT

Machine learning plays a crucial role in identifying specific stressors that impact plant species and provides a comprehensive understanding of the challenges plants face in natural environments. The use of machine learning algorithms has significantly enhanced our ability to classify and differentiate the types of stress. There are two main methodologies in machine learning: supervised learning and unsupervised learning. In supervised learning, the model is trained using input-output data pairs, while unsupervised learning involves training the model without access to output labels. Unsupervised learning is primarily used for data exploration and dimension reduction. This detailed classification helps us better understand the distinct characteristics associated with different stressors and provides a more nuanced view of the plant stress landscape. Machine learning also enables the quantitative assessment of stress intensity and extent, allowing for an accurate evaluation of its impact on plant health and productivity. This quantitative approach helps researchers measure the true extent of stressors and their effects on the overall health of plant ecosystems. By employing advanced algorithms, machine learning can make predictions about future occurrences of stress and their potential consequences on plant ecosystems. This foresight strengthens preventive measures for sustainable agricultural practices, as researchers and practitioners can anticipate and mitigate potential threats to plant health. The purpose of this review is to provide a comprehensive understanding of the applications and concepts of machine learning in uncovering the complexity of plant stress phenotyping and elucidating the involved molecular mechanisms.

KEYWORDS

Machine Learning, Stress, Biotic, Abiotic, Mechanism, Phenotyping.

نشریه علمی

زیست فناوری گیاهان زراعی

«مقاله مروری»

بهره‌گیری از یادگیری ماشین در فنوتایپینگ و درک مکانیسم‌های مولکولی مرتبط با تنش‌های زیستی و غیر زیستی گیاهان

بهمن پناهی^{۱*}، بنت‌الهدی قویدل^۲، پویا شاهگلی^۳

چکیده

یادگیری ماشینی محققان را قادر می‌سازد تا عوامل استرس‌زای خاصی را که بر گونه‌های گیاهی اثر می‌گذارند، مشخص کنند و درک دقیقی از چالش‌های متنوعی که گیاهان در محیط‌های طبیعی با آن مواجه هستند، ارائه کنند. با تشخیص الگوهای ظریف در مجموعه داده‌های گسترده، شناسایی نه‌تنها دقیق، بلکه روشنگر می‌شود و راه را برای مداخلات هدفمند هموار می‌کند. توانایی طبقه‌بندی و تمایز انواع استرس از طریق استفاده از الگوریتم‌های یادگیری ماشین بهبود یافته است. دو دسته اصلی رویکردهای مبتنی بر یادگیری ماشین شامل یادگیری نظارت شده و یادگیری بدون نظارت می‌باشند. در یادگیری نظارت شده، مدل با داده‌های زوج ورودی-خروجی آموزش داده می‌شود ولی در یادگیری بدون نظارت، مدل بدون دسترسی به برچسب‌های ورودی آموزش داده می‌شود و عمدتاً برای کاوش داده‌ها و کاهش ابعاد مورد استفاده قرار می‌گیرند. این طبقه‌بندی ظریف به درک کلی از ویژگی‌های متمایز مرتبط با عوامل استرس‌زای مختلف کمک می‌کند و نمای ظریف‌تری از چشم‌انداز تنش گیاه ارائه می‌دهد. یادگیری ماشینی ارزیابی کمی از شدت و میزان استرس را تسهیل می‌کند و امکان ارزیابی دقیق تأثیر بر سلامت و بهره‌وری گیاه را فراهم می‌کند. این رویکرد کمی محققان را قادر می‌سازد تا میزان واقعی عوامل استرس‌زا و پیامدهای آن‌ها را بر سلامت کلی اکوسیستم گیاهی بسنجند. با استفاده از الگوریتم‌های پیشرفته، پیش‌بینی‌هایی در مورد وقوع استرس در آینده و پیامدهای بالقوه آن‌ها بر اکوسیستم‌های گیاهی می‌توان انجام داد. این آینده‌نگاری اقدامات پیشگیرانه را برای شیوه‌های کشاورزی پایدار تقویت می‌کند، زیرا محققان می‌توانند تهدیدهای بالقوه برای سلامت گیاهان را پیش‌بینی کرده و کاهش دهند. هدف این بررسی ارائه یک درک جامع از کاربردها و مفاهیم یادگیری ماشینی در کشف اسرار فنوتیپ استرس گیاهی و تشریح مکانیسم مولکولی دخیل است.

واژه‌های کلیدی

یادگیری ماشین، تنش، زیستی، غیر زیستی، مکانیسم، فنوتایپینگ.

^۱گروه ژنومیکس، پژوهشکده بیوتکنولوژی شمال غرب و غرب کشور، پژوهشگاه بیوتکنولوژی کشاورزی ایران، سازمان تحقیقات، آموزش و ترویج کشاورزی، تبریز، ایران.
^۲گروه بیوتکنولوژی کشاورزی، دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، ایران.
^۳دانشکده فناوری اطلاعات و مهندسی کامپیوتر دانشگاه شهید مدنی آذربایجان، تبریز، ایران.

نویسنده مسئول:

بهمن پناهی

رایانامه: panahi.lahroodi@gmail.com

استناد به این مقاله:

پناهی، بهمن؛ قویدل، بنت‌الهدی و شاهگلی، پویا (۱۴۰۳). کاربرد یادگیری ماشین در فنوتایپینگ و درک مکانیسم‌های مولکولی مرتبط با تنش‌های زیستی و غیر زیستی گیاهان. فصلنامه علمی زیست فناوری گیاهان زراعی، ۱۳(۴۶)، ۷۵-۹۰.

<https://cropbiotech.journals.pnu.ac.ir/>

مقدمه

برای پاسخ به تقاضای آینده برای غذا، فیبر و سوخت، تولید محصولات کشاورزی باید تا سال ۲۰۵۰ دو برابر شود. عملکرد محصولات اساساً توسط تنش‌های گیاهی (زیستی و غیرزیستی) محدود می‌شود و اصلاح‌کنندگان گیاهی با ترکیب ژن‌های مقاومت و توسعه ارقام مقاوم به تغییرات اقلیمی، به بهبود عملکرد گیاهان در برابر تنش کمک می‌کنند. اصلاح‌کنندگان و محققان گیاهی، بر فنوتیپ گیاهی برای جمع‌آوری دقیق صفات و استفاده از ابزارهای ژنتیکی برای دستیابی به اهداف تحقیقاتی خود تکیه می‌کنند. فنوتایپینگ گیاهی به کاربرد روش‌ها و پروتکل‌ها برای اندازه‌گیری صفت خاص مرتبط با ساختار و عملکرد گیاه، از سطح سلولی تا کل گیاه یا سطح پوشش گیاه (canopy)، اطلاق می‌شود (Meraj *et al.*, 2024). در حال حاضر جمع‌آوری داده‌های فنوتیپی در مقیاس میکروسکوپی یک فرآیند دستی بوده که متخصصان را برای اندازه‌گیری فنوتیپ‌های مختلف درگیر می‌کند (van Dijk *et al.*, 2021). به منظور پیشرفت علم گیاهی، فنوتیپ دیجیتال گیاهی با هدف استخراج خودکار فنوتیپ‌ها از داده‌های حسگر به یک زمینه تحقیقاتی فعال تبدیل شده است (Kaiser *et al.*, 2024). در این زمینه یادگیری ماشین به‌عنوان رویکرد بین‌رشته‌ای بوده که از نظریه احتمال، آمار، تئوری تصمیم، تجسم و بهینه‌سازی برای تجزیه و تحلیل داده‌ها استفاده می‌کند. رویکردهای یادگیری ماشین معمولاً در موقعیت‌هایی مفید هستند که مقادیر زیادی از داده‌ها در دسترس هستند و ورودی‌ها (به‌عنوان مثال، داده‌های تصویر) را به مقادیر خروجی مورد علاقه (مثلاً فنوتیپ‌های استرس) مرتبط می‌کنند (Meraj *et al.*, 2024).

ظهور فناوری‌های توالی‌یابی سرتاسری و اطلاعات حاصل از آن، زیست‌شناسان را قادر می‌سازد تا به بررسی موضوعاتی بپردازند که زمانی بررسی آن‌ها دشوار یا غیرممکن بود. با این وجود داده‌های حاصل به دلیل خطاهای اندازه‌گیری و خطاهای ناشی از داده‌های ورودی قطعیت بالایی ندارند. لذا پردازش و تحلیل چنین مجموعه داده‌های بزرگی با استفاده از روش‌های آماری سنتی غیرممکن بوده و نیاز به رویکردهای یادگیری مدرن دارد (Panahi *et al.*, 2019; Sun *et al.*, 2020).

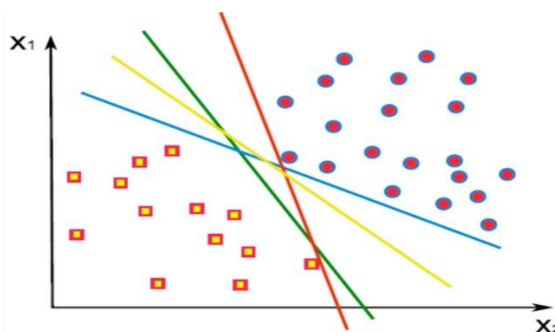
یادگیری ماشین علم برنامه‌نویسی جهت دریافت و شناخت الگوها و استخراج نتایج جدید از داده‌ها هست. این علم به‌طور گسترده در مطالعات اصلاح گیاهی و برای تفسیر داده‌های حاصل

از توالی‌یابی جهت مطالعه فنوتیپ گیاهان استفاده می‌شود (Niazian & Niedbala, 2020).

این رویکرد همچنین در زمینه‌های مرتبط با شناسایی تنش در گیاهان سودمند هست چراکه می‌تواند الگوها را با استفاده از ترکیب هم‌زمان عوامل متعدد به‌جای بررسی جداگانه صفات، تشخیص دهد. مفاهیم یادگیری ماشین (ML) در چهار دسته کلی در بررسی فنوتیپ تنش گیاهی مورد استفاده قرار می‌گیرد. این دسته‌بندی‌ها را به‌اصطلاح "ICQP" می‌گویند که مخفف (i) شناسایی، (ii) طبقه‌بندی، (iii) کمی‌سازی، و (iv) پیش‌بینی است. این چهار دسته شامل زنجیره‌ای از ویژگی‌ها هست که در آن اطلاعات به‌طور فزاینده‌ای از یک تصویر مشخص استنتاج می‌شود. شناسایی به تشخیص تنش خاص مانند سندرم مرگ ناگهانی سویا یا زنگ گندم اشاره دارد. طبقه‌بندی مرحله بعدی است، جایی که یادگیری ماشین برای طبقه‌بندی تصویر بر اساس علائم تنش و نشانه‌ها استفاده می‌شود (Anim-Ayeko *et al.*, 2018; Singh *et al.*, 2023). کمی‌سازی شامل توصیف کمی تنش از نظر بروز و شدت است. بروز بیماری به‌عنوان میزان رخ داد جدید بیماری در یک دوره زمانی یا در هر لحظه (معمولاً در زمان حداکثر بروز بیماری) تعریف می‌شود. در آسیب‌شناسی گیاهی، یک روش رایج برای توصیف بروز بیماری، درصد برگ‌های بیمار روی یک گیاه یا تعداد گیاهان بیمار از کل تعداد گیاهان در یک مزرعه یا قطعه است (Das Choudhury *et al.*). شدت بیماری یک معیار کمی دقیق‌تر است و به‌عنوان ناحیه بافت گیاهی تحت تأثیر بیماری (معمولاً به‌صورت درصد نشان داده می‌شود) روی یک برگ یا کل تاج پوشش گیاه گزارش می‌شود (Bock *et al.*, 2010). آخرین دسته، پیش‌بینی تنش گیاهی قبل از ظاهر شدن علائم تنش قابل مشاهده است که جهت کنترل به‌موقع و مقرون به‌صرفه تنش نقش مؤثری دارد (Gou *et al.*, 2024; Sarkar *et al.*, 2023; Singh *et al.*, 2018).

در مقاله مروری حاضر الگوریتم‌های مختلف یادگیری ماشین و نیز کاربردهای آن‌ها در ارتباط با فنوتایپینگ و درک مکانیسم‌های مرتبط با تنش‌های گیاهی مورد بررسی و بحث قرار می‌گیرد.

قرار می‌گیرند (Panahi *et al.*, 2019). آن‌ها با موفقیت در بسیاری از مطالعات مربوط به تشخیص تنش گیاهی، مانند شناسایی بیماری Huanglongbing (HLB) که با عنوان بیماری سبز شدن مرکبات شناخته می‌شود) و تنش‌های ناشی از کمبود مواد مغذی در برگ‌های مرکبات استفاده شده‌اند (Cen *et al.*, 2017). در عین حال که SVM ساده بوده و با ابعاد بالای داده‌ها (از جمله طیف‌ها و تصاویر) کاملاً خوب کار می‌کند، اما نمی‌تواند میزان تأثیر خطاها بر شناسایی کلاس واقعی را توضیح دهد. این امر به ویژه زمانی مشکل‌ساز است که مجموعه داده‌ها دارای داده‌های پرت بوده و تعریف یک مرز تصمیم‌گیری مشخص مشکل باشد (Sadeghi *et al.*, 2021; Ghahramani *et al.*, 2022; Tahmasebi *et al.*, 2023; Zubler & Yoon, 2020). همچنین ماشین بردار پشتیبان (SVM) برای تشخیص زودرس بیماری برگ قرمز بادام (*Prunus dulcis*) با استفاده از تصاویر هایپراسپکترال با وضوح بالا و تصاویر حرارتی استفاده شد. نتایج نشان داد که روش به کار رفته به صورت موثری توانایی تفکیک درختان سالم و بیمار را دارد (Peña *et al.*, 2015). در مطالعه‌ای دیگر روش طبقه‌بندی SVM برای شناسایی بیماری verticillium wilt در زیتون (*Olea europaea*) با استفاده از تصاویر حرارتی و هایپراسپکترال استفاده شد نتایج نشان داد که این روش با دقت طبقه‌بندی ۷۹٫۲٪ می‌تواند درختان آلوده به قارچ مذکور را از نمونه‌های سالم تشخیص دهد (Calderón *et al.*, 2015).



شکل ۱. یک مرز تصمیم برای داده‌های دو بعدی (Cervantes *et al.*, 2020).

شناسایی زودهنگام بیماری Huanglongbing (HLB) روش کارآمدی برای کنترل گسترش این بیماری به شمار می‌رود. یادگیری ماشین برای شناسایی و تخمین HLB بر اساس داده‌های طیف‌سنجی تصویربرداری فلورسانس نمونه‌های برگ

۱-۲- استفاده از یادگیری ماشین در فنوتایپینگ تنش‌های گیاهی

الگوریتم‌های یادگیری ماشین برای دسته‌بندی داده‌های فنوتایپ گیاهی استفاده می‌شود. بر اساس مطالعات صورت گرفته، برای دسته‌بندی تنش‌های زنده، عمدتاً از الگوریتم‌های یادگیری ماشین مانند درخت تصمیم، ماشین بردار پشتیبان (SVM)، و شبکه‌های عصبی استفاده می‌کنند تا با استفاده از ویژگی‌های مختلف مانند مورفولوژی گیاه، میزان رشد، و ویژگی‌های بیوشیمیایی، می‌توان تنش‌های زنده را شناسایی و دسته‌بندی کرد. تنش‌های غیرزنده که شامل عوامل محیطی مانند خشکسالی، شوری خاک، دما، و نور می‌باشند. برای دسته‌بندی تنش‌های غیرزنده، نیاز به تحلیل داده‌های محیطی و آماری داریم. الگوریتم‌های مبتنی بر رگرسیون خطی، مدل‌های مارکوف، و روش‌های مبتنی بر تشخیص الگو می‌توانند عمدتاً برای تفکیک و دسته‌بندی تنش‌های غیرزنده مفید می‌باشند (Gill *et al.*, 2022; Singh *et al.*, 2018). با استفاده از داده‌های جمع‌آوری شده میدانی، آزمایشگاهی، و حسگرهای مختلف، می‌توان الگوها و روندهای فعالیت گیاهان را در مقابل تنش‌های زنده و غیرزنده دسته‌بندی کرد. این دسته‌بندی‌ها می‌تواند به محققان و کشاورزان کمک کند تا بهترین راهکارها برای مدیریت تنش‌های گیاهی را پیدا کنند (Shoab *et al.*, 2023).

۱-۲-۱- ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) روش یادگیری نظارت‌شده بوده که می‌تواند داده‌ها را بر روی یک سیستم مختصات دوبعدی با کلاس‌های شناخته شده (تنش در مقابل سالم) به صورت خطی رسم کرده و به بهترین وجه داده‌ها را به دو کلاس تقسیم کند. به خط تقسیم کننده دو کلاس اصطلاحاً خط مرز تصمیم گفته می‌شود (شکل ۱) (Cervantes *et al.*, 2020). این روش همچنین می‌تواند برای داده‌های با سه بعد یا ابعاد بیشتر نیز استفاده شود. در حالی که SVM ذاتاً یک روش خطی است، جداسازی غیرخطی نیز با استفاده از هسته‌های غیرخطی امکان‌پذیر است. طبقه‌بندی به چند کلاس نیز با استفاده از مرزهای تصمیم‌گیری چندگانه امکان‌پذیر است (Zubler & Yoon, 2020). SVM ها یکی از رایج‌ترین الگوریتم‌های یادگیری ماشینی هستند که در برنامه‌های کشاورزی مورد استفاده

گزارشات صورت گرفته، این الگوریتم می‌تواند ژنوتیپ‌های ذرت را با دقت ۸۵ درصد از همدیگر تشخیص دهد (Seka et al., 2019). همسوسو با طیف متنوع کارکرد رویکردهای یادگیری ماشین، این الگوریتم همچنین برای ارزیابی میزان و شدت بیمار در ریشه گیاهان به کار رفته است. این رویکرد که در نرم افزاری با عنوان PlantCV ارائه شده است، می‌تواند به کمک تصاویر فراقمرز، ریشه‌های آلوده را از ریشه‌های سالم گیاهان زراعی تشخیص دهد. عملکرد این روش می‌تواند تاثیر بسزایی در کنترل بیماری‌های خاکزاد در گیاهان داشته باشد (Pierz et al., 2023). اخیراً، همچنین با استفاده از الگوریتم naïve bayes برای تشخیص گونه‌های گندم مورد استفاده قرار گرفته است (Arora et al., 2023). علاوه بر موارد مذکور کارکرد و توانایی الگوریتم naïve bayes در تشخیص ارقام گردوی مقاوم به سرمای دیررس بهاره (Panahi et al., 2022) و نیز تشخیص ژنوتیپ‌های مختلف انگور (Holasou et al., 2024) مورد تاکید قرار گرفته است.

۳-۲-۱-K-mean

K-mean محبوب‌ترین روش طبقه‌بندی خوشه‌بندی است که اولین بار در سال توسط MacQueen در سال ۱۹۶۷ پیشنهاد شد. K-mean روش خوشه‌بندی تکراری بدون نظارت، غیرقطعی و عددی است. در K-mean هر خوشه با مقدار میانگین اشیا در خوشه نشان داده می‌شود (Yadav & Sharma, 2013). این روش خوشه‌بندی نوعی یادگیری بدون نظارت است که بر روی داده‌های بدون برچسب اعمال می‌شود. هدف این الگوریتم شناسایی گروه‌های غیرهمپوشانی معینی در داده‌ها است و K متغیری است که تعداد گروه‌ها را نشان می‌دهد. الگوریتم به صورت تکراری کار می‌کند و و هر نقطه داده را به یکی از گروه‌های K اختصاص می‌دهد. بر اساس شباهت ویژگی داده‌ها، نقاط خوشه‌بندی می‌شوند (Paul et al., 2020).

هدف از خوشه‌بندی، گروه‌بندی نمونه‌ها به بخش‌هایی است که دارای ویژگی‌های مشابه بر اساس برخی معیارهای فاصله یا تشابه هستند. الگوریتم‌های مبتنی بر مرکز مانند K-means، یکی از متداول‌ترین الگوریتم‌های خوشه‌بندی است. مرکز K به‌عنوان مرکز خوشه‌های متناظر که دارای حداقل فاصله با نمونه‌های دیگر در همان خوشه‌ها هستند، در نظر گرفته می‌شود (Sun et al., 2020). این رویکرد به صورت موفق برای

استفاده شده است (Wetterich et al., 2013). تصاویر پیش‌پردازش شده (بخش بندی شده) و ویژگی‌های استخراج شده به‌عنوان ورودی SVM مورد استفاده قرار گرفتند (Singh et al., 2016). در پیش‌بینی بیماری بلاست برنج نیز SVM به‌عنوان یک مدل با استفاده از داده‌های مبتنی بر آب و هوا استفاده شده است (Kaundal et al., 2006). در پیش‌بینی بیماری سیب‌زمینی با استفاده از تصاویر برگ‌های سیب‌زمینی و ویژگی‌هایی که در این تصاویر مانند ویژگی‌های طیفی و زمینه‌ای وجود دارد با استفاده از الگوریتم SVM طبقه‌بندی می‌شود (Islam et al., 2017). از این رویکرد در تلفیق با تصاویر مادون قرمز برای شناسایی مناطقی از مزارع اسفناج که دچار تنش کم آبی شدند، استفاده گردید. نتایج بدست آمده نشان داد که الگوریتم SVM می‌تواند این نواحی را با دقت بالایی (۹۶ درصد) از نواحی دیگر که دچار خشکی نشده‌اند، تشخیص دهد (Raza et al., 2014). در مطالعه دیگری، SVM برای شناسایی تنش خشکی در گیاهان جو (*Hordeum vulgare*) با استفاده از مجموعه تصاویر هایپراسپکترال مورد استفاده قرار گرفت. در این مطالعه دقت تشخیص چهار روش پیش‌بینی نظارت شده شامل ماشین بردار پشتیبان (SVM)، one-vs.-all SVM، رگرسیون ماشین بردار پشتیبان (SVR)، رگرسیون ترتیبی ماشین بردار پشتیبان (SVORIM) برای استخراج سطوح استرس در جو مورد استفاده قرار گرفت. ارزیابی‌ها و نتایج بدست آمده نشان داد که رویکرد one-vs.-all SVM نسبت به روش‌های دیگر تنش خشکی را با دقت بالاتری (۸۳ درصد)

۲-۲-۱-Naïve Bayes

Naïve Bayes با فرض استقلال ویژگی‌ها مورد استفاده قرار می‌گیرد. این الگوریتم یک روش محبوب برای طبقه‌بندی متن می‌باشد (Naik et al., 2017b). مشخص شده است که الگوریتم naïve bayes می‌تواند بیماری میوه پایا را با دقت ۹۰٪ را تشخیص دهد، در این روش ابتدا تصاویر برگ این میوه را که شامل ۴۱ ویژگی هست و همچنین روی برگ‌هایی که دارای ۱۳ نوع بیماری می‌باشند استفاده شده است (Panahi et al., 2020; Sari et al., 2020). همچنین این رویکرد به صورت موفق برای ارزیابی شدت تنش در سویا مورد استفاده قرار گرفته است (Naik et al., 2017a). در مطالعه‌ای دیگر از این رویکرد برای تشخیص ژنوتیپ‌های مختلف ذرت استفاده شد. بر اساس

طریق یک یا چند لایه پنهان بعدی پردازش شده و در نهایت به لایه خروجی پیوند داده می‌شود. شبکه‌های عصبی در این زمینه از تکنیک‌های پیشرفته مانند طبقه‌بندی تصاویر، شیمی محاسباتی و فیزیک بهتر عمل کرده‌اند. در ژنومیک، برای تجزیه و تحلیل داده‌های بیان ژن، درک پروموتورها (promoters) و افزایش‌دهنده‌ها (enhancers) و شناسایی مکان‌های شروع ترجمه کاربرد دارد (Sun et al., 2020). در شبکه‌های عصبی مصنوعی، رویکردهای عمیق دارای لایه‌های پنهان بسیار بیشتری در شبکه هستند و از این رو قدرت تشخیص و پیش‌بینی بیشتری دارند (Pound et al., 2017). این روش یک مدل نظری است که رابطه بین سیگنال‌های ورودی و خروجی برگرفته از شبکه‌های عصبی بیولوژیکی را نشان می‌دهد (Paul et al., 2020).

در کشاورزی برای کاربردهای متعدد، مانند مدل‌سازی اطلاعات حرارتی برای ارزیابی تنش آب، نقشه برداری پوشش گیاهی، پیش‌بینی عملکرد و پیش‌بینی تنش نیتروژن کاربرد دارد. (Virnodkar et al., 2020). برای تشخیص بیماری گیاهان با استفاده از تصویر گیاهان که از شبکه عصبی مصنوعی و SVM و با استفاده از کاهش ویژگی‌ها اعم از کاهش رنگ که می‌تواند موجب افزایش سرعت الگوریتم شود استفاده شده است که دقت بسیار مناسب ۹۲٪ نیز را به دست آورده است (Pujari et al., 2016).

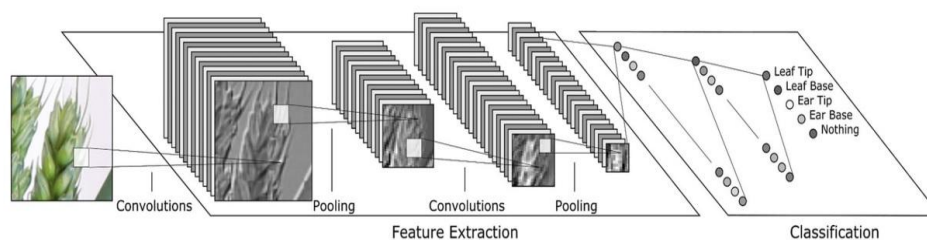
۱-۲-۴-۱- شبکه‌های عصبی پیچشی (CNN)

در شبکه‌های عصبی CNN نقشه‌های ویژگی‌های ممکن است شامل آرایش‌های پیچیده از ویژگی‌هایی باشد که اشیاء دنیای واقعی را نشان می‌دهند. توجه به این نکته ضروری است که این ویژگی‌ها الگوریتم‌های آموزشی CNN هستند و به صورت دستی کدگذاری نمی‌شوند. همان طور که در شکل ۲ نشان داده شده است، CNN‌های مدرن معمولاً از لایه‌های زیادی استفاده می‌کند که شبکه‌ها را پیچیده می‌کند و اغلب به صدها و گاهی هزاران تصویر برای آموزش با دقت مطلوب نیاز دارد (Pound et al., 2017).

گروه‌بندی گیاهان آلوده به بیماری و نیز سالم با استفاده از تصاویر برگ مورد استفاده قرار گرفته است (Trivedi et al., 2022). همچنین کارکرد موفق این الگوریتم در ارزیابی ژنوتیپ‌های مختلف سویا در شرایط گلخانه گزارش شده است (Zhou et al., 2018). پیش‌بینی عملکرد برداشت، روش‌های مزرعه را قادر می‌سازد تا در طول فصل رشد اصلاح شوند. سنجش از دور مبتنی بر وسایل نقلیه هوایی بدون سرنشین (UAV) روشی امیدوارکننده برای تخمین عملکرد محصول است. در یک مطالعه، عملکرد برنج با تقسیم‌بندی مناطق دانه با استفاده از تصاویر RGB در ارتفاع پایین جمع‌آوری شده با استفاده از پهپاد نوع بال چرخشی برآورد شد. در این مطالعه، یک روش پردازش تصویر که ترکیبی از خوشه‌بندی K-means با یک الگوریتم برش نمودار (KCG) برای بخش‌بندی مناطق دانه برنج ارائه گردیده است. تصاویر RGB پیش‌زمینه به فضای رنگی Lab تبدیل و سپس از خوشه‌بندی K-means برای برجسب گذاری پیکسل‌ها بر اساس اطلاعات رنگ استفاده گردیده است. نتایج نشان داد که روش پیشنهادی می‌تواند نواحی دانه‌ها را با خطای نسبی ۶٪ تا ۳۳٪ تقسیم‌بندی کند و خطای نسبی روش قبلی را (۱٪ تا ۳۱٪) بهبود بخشد (Reza et al., 2019). عملکرد گندم تحت تأثیر تعداد خوشه در واحد سطح است و به طور سنتی از شمارش دستی برای تخمین عملکرد گندم استفاده می‌شود. در مطالعه‌ای، برای تحقق شمارش سریع و دقیق بلال گندم، از خوشه‌بندی K-means برای تقسیم خودکار تصاویر خوشه گندم گرفته شده توسط دستگاه‌های دستی استفاده شد. مجموعه داده‌های بخش بندی شده با ایجاد چهار دسته برجسب تصویر ساخته شد: بلال غیر گندم، یک خوشه گندم، دو خوشه گندم، و سه خوشه گندم، که سپس برای آموزش و آزمایش به مدل شبکه عصبی کانولوشن (CNN) ارسال شد. رویکرد به کار رفته توانست خوشه‌های گندم را با دقت بالا و قابل قبول دسته بندی و تعداد آنها را ارائه دهد (Xu et al., 2020).

۱-۲-۴-۱- شبکه‌های عصبی

شبکه عصبی (Neural network)، ساختار عصبی مغز انسان را با گره‌های به هم پیوسته که در لایه‌های مختلف سازماندهی شده است تقلید می‌کند. داده‌های آموزشی (Training data) ابتدا وارد لایه ورودی می‌شوند، سپس از



شکل ۲. یک مثال ساده از CNN که بر روی تصویر بخشی از خوشه گندم کار می‌کند. هر لایه به‌طور خودکار ویژگی‌های مفید مانند لبه‌ها یا گوشه‌ها را استخراج می‌کند و تعدادی نقشه ویژگی را خروجی می‌دهد. عملیات ادغام اندازه نقشه‌های ویژگی را برای بهبود کارایی کوچک می‌کند. تعداد نقشه‌های ویژگی عمیق‌تر در شبکه افزایش می‌یابد تا دقت طبقه‌بندی بهبود یابد. در نهایت، لایه‌های شبکه عصبی استاندارد، لایه‌های طبقه‌بندی را تشکیل می‌دهند که خروجی‌های احتمالی را برای هر کلاس ارائه می‌دهند.

نشان می‌دهد که مدل قادر به تشخیص و تمایز بین برگ‌های سالم و برگ‌های بیمار با دقت بالایی است (Shrestha et al., 2020).

۲-۴-۱- شبکه‌های عصبی پیچشی منطقه‌ای^۱

از آنجایی که CNN در تصاویری که از اشیا متفاوتی تشکیل شده است دچار مشکل می‌شود هم دقت پایینی دارد و هم سرعت کمتری دارد به همین دلیل بهبودهای زیادی برای این نوع الگوریتم ایجاد شده است که یکی از آن‌ها تقسیم کردن تصویر به بخش‌های کوچک‌تر به نام Region هست و از جستجوی انتخابی برای پیدا کردن این بخش‌ها استفاده می‌شود که در کل این رویکرد باعث می‌شود سرعت و دقت الگوریتم بهتر شود، بعد از این که این بخش‌ها به دست آمدند از الگوریتم CNN برای تشخیص اشیا موجود در تصویر استفاده می‌شود (Girshick et al., 2014).

۳-۴-۱- شبکه‌های عصبی پیچشی منطقه‌ای

سریع (Fast-RCNN)^۲

R-CNN اولین روشی بود که منطقه‌های مختلف تصویر را با استفاده از یک الگوریتم تشخیص انتخابی (Selective Search) شناسایی کرده و سپس هر منطقه را به یک شبکه CNN جهت استخراج ویژگی‌ها می‌فرستاد. این روش دقت بالایی در تشخیص اشیا داشت، اما بسیار کند بود و نیاز به محاسبات زیادی برای هر منطقه داشت. یک شبکه سریع R-CNN یک تصویر کامل و یک مجموعه را به‌عنوان ورودی می‌گیرد، ابتدا کل تصویر را با چندین لایه convolutional (conv) و max pooling پردازش می‌کند تا یک نقشه ویژگی conv تولید کند. سپس، برای هر پیشنهاد شی، یک لایه ادغام منطقه مورد علاقه (ROI) یک بردار ویژگی با طول ثابت را از نقشه ویژگی استخراج می‌کند. (Girshick, 2015).

ابتدا تصاویر به ورودی داده می‌شوند سپس پیش‌پردازش قطعه‌بندی شامل فرآیند تقسیم‌بندی تصویر، بهبود تصویر و تبدیل فضای رنگی است. ابتدا تصویر دیجیتالی تصویر با یک پالایه بهبود می‌یابد. سپس هر تصویر را به یک آرایه تبدیل می‌شود. با استفاده از نام علمی Binarizes Diseases، نام هر تصویر به یک فیلد باینری تبدیل می‌شود و در نهایت طبقه‌بندی کننده‌های CNN برای شناسایی بیماری‌ها در هر کلاس گیاهی آموزش دیده‌اند، از نتایج سطح ۲ برای فراخوانی یک طبقه‌بندی استفاده می‌شود که برای طبقه‌بندی بیماری‌های مختلف در آن گیاه آموزش دیده است، اگر وجود نداشته باشد، برگ‌ها به‌عنوان "سالم" طبقه‌بندی می‌شوند (Kumar, 2021). مطالعه شبیه‌سازی و تجزیه و تحلیل بر روی تصاویر نمونه از نظر پیچیدگی زمانی و مساحت منطقه آلوده با تکنیک پردازش تصویر انجام می‌شود. در مطالعه مورد نظر در مجموع ۱۵ مورد به مدل تغذیه شده است که از این تعداد ۱۲ مورد مربوط به برگ‌های گیاه بیمار به نام‌های لکه باکتریایی کاغذ زنگوله‌ای، پژمردگی زودرس سیب‌زمینی، پژمردگی دیررس سیب‌زمینی، لکه هدف گوجه‌فرنگی، ویروس موزاییک گوجه‌فرنگی، ویروس پیچ‌خوردگی برگ زرد گوجه‌فرنگی، لکه باکتریایی گوجه‌فرنگی، پژمردگی زودرس گوجه‌فرنگی، پژمردگی دیررس گوجه‌فرنگی، کپک برگ گوجه‌فرنگی، لکه برگ گوجه‌فرنگی سپتوریا و کنه عنکبوتی گوجه‌فرنگی و ۳ مورد برگ سالم به نام‌های کاغذ زنگی سالم، سیب‌زمینی سالم و گوجه‌فرنگی سالم هست. روش پیشنهادی پیش‌بینی بیماری گیاه ورودی را از تصاویر برگ‌های گیاه می‌گیرد. ابتدا داده‌ها با تغییر اندازه تصاویر ورودی پیش‌پردازش می‌شوند و سپس یک آرایه برای آن ایجاد می‌شود. سپس مجموعه داده‌ها و برچسب همه تصاویر از هم جدا می‌شوند. این مدل که بر روی مجموعه داده مذکور آموزش داده شده است، به‌طور متوسط در تشخیص برگ‌های بیمار با دقت ۸۸٪ عمل کرده است. این نتیجه

¹Region-based Convolutional Neural Network

²Fast Region-based Convolutional Neural Network

۵-۲-۱- درخت تصمیم^۳

یکی دیگر از الگوریتم‌های یادگیری نظارت‌شده است که در مسائل طبقه‌بندی استفاده می‌شود، درخت تصمیم هست. این الگوریتم نیازی به تنظیم پارامتر یا دانش دامنه ندارد، بنابراین مناسب برای دانش اکتشافی است و می‌تواند داده‌های چندبعدی را مدیریت کند. مراحل یادگیری و طبقه‌بندی درخت تصمیم سریع و ساده است (Paul et al., 2020). تشخیص به‌موقع و دقیق بیماری‌های برگ نقش عمده‌ای در جلوگیری از کاهش بهره‌وری و یا کاهش محصولات کشاورزی دارد. با استفاده از درخت تصمیم تشخیص به‌موقع بیماری برگ با دقت بالا و سرعت مناسب و همچنین نوع بیماری می‌تواند باعث بهبود کیفیت محصولات کشاورزی و همچنین میزان محصولات شود، که با استفاده از تصاویر برگ‌ها و درخت تصمیم دقت بسیار مناسب ۹۵٪ را در مدت‌زمان معقولی ارائه می‌کند (Rajesh et al., 2020). رویکرد بهبودیافته درخت تصمیم از یک فن انتخاب ویژگی (ویژگی‌هایی را در دیتاست انتخاب می‌کند که تأثیر بیشتری در عملکرد الگوریتم دارند و از اضافه کردن ویژگی‌هایی که تأثیر چندانی در دقت الگوریتم ندارند خودداری می‌کند که باعث افزایش سرعت الگوریتم می‌شود به شکلی که تأثیر منفی روی دقت الگوریتم ندارد.) و نمونه‌برداری (RO (Randor Over) برای متعادل کردن دیتاست استفاده می‌کند. هدف اصلی این رویکرد افزایش عملکرد الگوریتم درخت تصمیم برای تشخیص بیماری سوپا است که دقت مناسب در مدت‌زمان مناسبی دارد (Bhatia et al., 2020; Panahi et al., 2022).

۶-۲-۱- جنگل‌های تصادفی^۴

جنگل‌های تصادفی طرحی است که در سال ۲۰۰۰ توسط Leo Breiman برای ساخت یک مجموعه پیش‌بینی‌کننده با مجموعه‌ای از درختان تصمیم‌گیری که در زیر فضاهای انتخابی تصادفی رشد می‌کنند، پیشنهاد شد (Biau, 2012). جنگل‌های تصادفی ترکیبی از پیش‌بینی‌کننده‌های درختی هستند به طوری که هر درخت به مقادیر یک بردار تصادفی که به‌طور مستقل و با توزیع یکسان برای همه درختان جنگل نمونه‌برداری شده است، بستگی دارد (Breiman, 2001). جنگل تصادفی (RF) یا جنگل تصمیم تصادفی یک مجموعه یادگیرنده است که با ساختن

۴-۲-۱- شبکه‌های عصبی پیچشی منطقه‌ای سریع‌تر^۱

هرچند شبکه‌های fast RCNN سریع‌تر و دقیق‌تر هستند نسبت به RCNN و CNN، ولی در دیتا ست‌های بزرگ باز هم عمل می‌کند و از روش جستجوی انتخابی استفاده می‌کند که سرعت الگوریتم را به نسبت کم می‌کند درحالی‌که در شبکه‌های Faster RCNN^۲ از روش RPN استفاده می‌کند، روش RPN به این صورت هست که یک تصویر به‌عنوان ورودی داده می‌شود و مجموعه‌ای از مستطیل‌ها که شامل اشیاء درون تصویر ورودی هست را به‌عنوان خروجی برمی‌گرداند، که باعث می‌شود که الگوریتم سریع‌تر عمل کرده و در دیتا ست‌های بزرگ که نیاز به زمان آموزش زیادی دارد به مشکل برنخورده و در مدت‌زمان کم به بهترین نتیجه دست یابد (Ren et al., 2015).

تشخیص سریع بیماری گیاهان با استفاده از تصاویر برگ‌های گیاهان می‌توان از ضررهای زیاد در زمینه کشاورزی جلوگیری کرد. از الگوریتم Faster RCNN جهت تشخیص بیماری گیاه با استفاده از تصاویر برگ‌های گیاهان که به‌صورت دستی جمع‌آوری شده است استفاده می‌شود و در زمان معقولی می‌تواند تشخیص دهد که آیا گیاه موردنظر دچار بیماری خاصی هست یا نه و اگر دچار بیماری هست نام بیماری چیست؟ که موجب می‌شود سریع‌تر بیماری را تشخیص داده و گیاه موردنظر را درمان کند و از آسیب‌های جدی و ضررهای مالی جلوگیری نماید (Cynthia et al., 2019).

بیماری‌ها چالش بزرگی برای بخش کشاورزی است. تشخیص دقیق و سریع بیماری‌ها در گیاهان می‌تواند زیان‌های اقتصادی را تا حد زیادی کاهش دهد. روشی مبتنی بر Faster RCNN برای تشخیص بیماری‌های گوجه‌فرنگی ارائه شده که RCNN سریع‌تر را با شبکه‌های عصبی پیچیده عمیق، از جمله resnet50 و resnet101 ترکیب کرده و مجموعه داده‌های گوجه‌فرنگی مبتلابه بیماری و گوجه‌فرنگی سالم جمع‌آوری شده را آموزش و آزمایش کرده‌اند، که شامل شش تصویر گوجه‌فرنگی بیمار و تصاویر گوجه‌فرنگی سالم هست. نتایج تجربی نشان می‌دهد که سیستم پیشنهادی می‌تواند به‌طور مؤثر انواع مختلف بیماری‌های گوجه‌فرنگی را شناسایی کند (Wang & Qi, 2019).

1Faster Region-based Convolutional Neural Network

2Region Proposal Network

3Decision tree

4Radom Forest

زیرمجموعه‌ای از ژن‌های بیان شده متفاوت طبقه‌بندی کند (Shaik & Ramakrishna, 2014).

۲-۷-۲-۱- تجزیه و تحلیل تشخیص خطی (LDA)

یک روش ثابت شده برای انجام مشکلات کاهش ابعاد است که در پیش‌پردازش برای برنامه‌هایی مانند یادگیری ماشین، طبقه‌بندی الگوها، و غیره استفاده می‌شود. LDA بسته به هدف مطالعه، برای ترکیبی خطی از متغیرها دو کلاس را از هم جدا می‌کند (Paul et al., 2020).

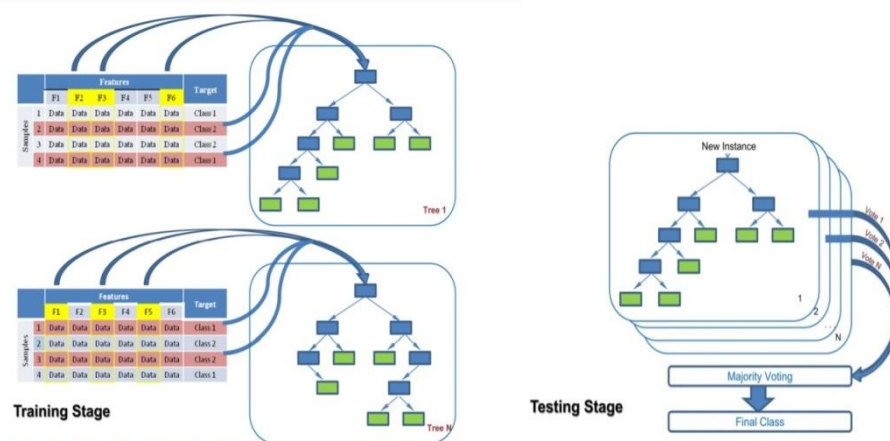
رگرسیون خطی به دلیل سادگی و قابلیت تفسیر در ژنومیک رایج است. از رگرسیون خطی برای ارزیابی اثرات core promoter element بر روی رونویسی و برای پیش‌بینی جهش‌های ژنی در نمونه‌های سرطانی بر اساس ویژگی‌های ژن استفاده می‌شود. با این حال، رگرسیون خطی هنگامی که برای تجزیه و تحلیل داده‌های با ابعاد بالا استفاده می‌شود، اغلب دارای خطای پیش‌بینی و واریانس بالایی است. معمولاً از الگوریتم‌های منظم‌سازی مانند رگرسیون رنج (ridge)، رگرسیون کمند (lasso) و رگرسیون خالص الاستیک (elastic net) استفاده می‌شود. این روش‌ها در مجموع عملکرد بهتری نسبت به رگرسیون خطی نشان می‌دهند و در مطالعات گسترده ژنوم محبوبیت پیدا کرده‌اند. محققان از آن‌ها برای پیش‌بینی ارزش‌های اصلاح ژنومیک برای حیوانات و گیاهان بر اساس همه نشانگرهای مولکولی و برای شناسایی SNP‌های مربوط به صفات مورد علاقه استفاده کرده‌اند (Sun et al., 2020).

بسیاری از درختان تصمیم ضعیف برای طبقه‌بندی و رگرسیون ساخته شده است. نمونه‌های آموزش (Bootstrap) به صورت تصادفی از یک مجموعه داده اصلی برای ساختن درختان متعدد با جایگزینی نمونه‌ها انتخاب شده است. این احتمال وجود دارد که اصلاً هیچ نمونه‌ای را انتخاب نکنید یا نمونه‌ای را بیش از یک بار انتخاب کنید. درختان به بهترین شکل ممکن رشد می‌کنند، یعنی هرس انجام نمی‌شود. جنگل به حداکثر اندازه خود رشد می‌کند تا زمانی که هر گره خالص شود. اکثریت رای پیش‌بینی‌های همه درختان، تصمیم نهایی گروه را تعیین می‌کند. برای تست کردن داده‌های جدید، از میان تمام درختان تولید شده عبور می‌کند و هر درخت به یک کلاس رأی می‌دهد. کلاسی که حداکثر رأی را دریافت کند، کلاس انتخابی نهایی خواهد بود. شکل ۳ مراحل آموزش و آزمایش الگوریتم جنگل تصادفی را نشان می‌دهد (Virnodkar et al., 2020). با روش "کیسه" و ادغام آن‌ها برای به دست آوردن پیش‌بینی بهتر نتایج. مشابه SVM، random forest می‌تواند هم برای طبقه‌بندی و هم برای رگرسیون استفاده شود (Sun et al., 2020). در شکل ۴ انواع الگوریتم‌های یادگیری ماشین به صورت شماتیک نشان داده شده است (Silva et al., 2019).

۲-۷-۱- سایر الگوریتم‌های استفاده شده در فنوتیپ تنش

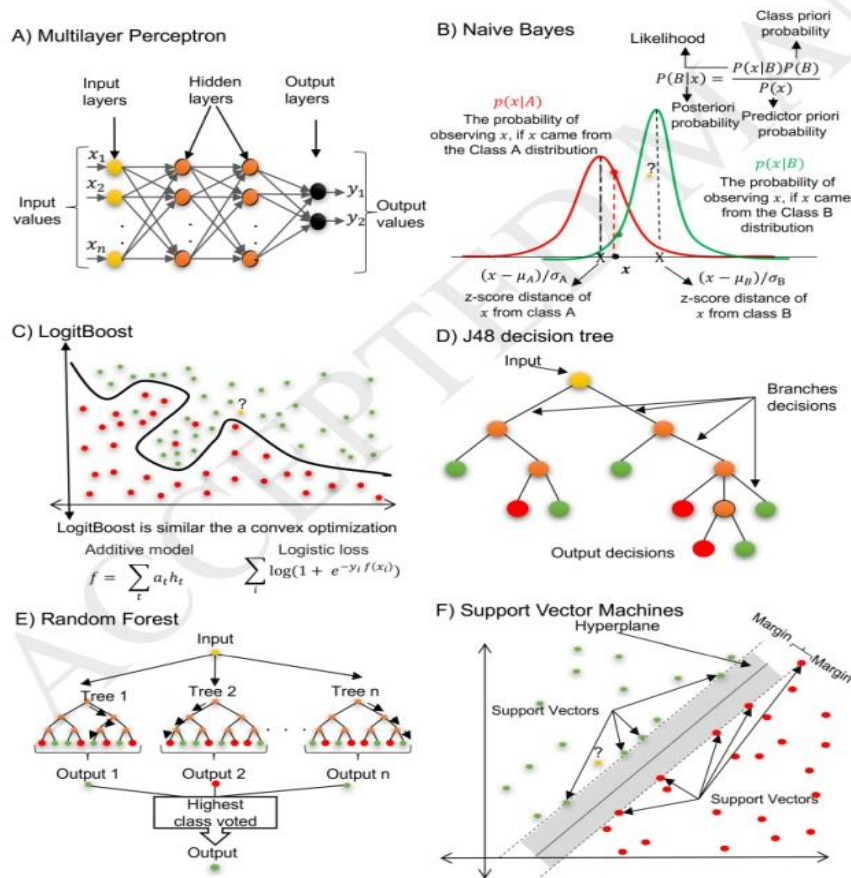
۱-۲-۷-۱- ماشین بردار پشتیبان بازگشتی^۱ (R-SVM)

یک مدل یادگیری ماشینی نظارت شده است و می‌تواند تنش‌های غیر زیستی و زیستی را با دقت ۱۰۰ درصد با استفاده از



شکل ۳. مراحل آموزش و آزمایش در الگوریتم جنگل تصادفی

1. Recursive-support vector machines
2. Linear Discriminate Analysis



شکل ۴. الگوریتم‌های یادگیری ماشین؛ A) شبکه عصبی مصنوعی. توپ‌های زرد نورون‌هایی هستند که لایه ورودی را تعریف می‌کنند. دایره‌های نارنجی نورون‌های حد واسط یا لایه‌های پنهان را مشخص می‌کنند. دایره‌های سیاه نورون‌هایی هستند که لایه‌های خروجی را مشخص می‌کنند. B) فرآیند تصمیم‌گیری احتمالی برای جداسازی کلاس‌های A (تراکم قرمز رنگ) و B (تراکم سبز رنگ) را نشان می‌دهد. نقاط قرمز و سبز یک نمونه داده از کلاس را طبقه‌بندی می‌کنند. نقطه زرد یک نمونه جدید را نشان می‌دهد که باید طبقه‌بندی شود. C) LogitBoost نوعی روش تقویتی است که یک رگرسیون لجستیک افزایشی را پیاده‌سازی می‌کند و احتمال a را به حداکثر می‌رساند. بهینه‌سازی محدب (Convex optimization)، طبقه‌بندی با روش‌های تقریبی است و نقاط قرمز نمونه‌هایی از کلاس‌های منفی را نشان می‌دهند، در حالی که نقاط سبز نشان‌دهنده کلاس‌های مثبت هستند. نقطه زرد نمونه جدیدی از داده‌ها را مشخص می‌کند که باید طبقه‌بندی شوند. D) درخت تصمیم. دایره زرد نشان‌دهنده داده‌های ورودی است. دایره نارنجی یک معیار تصمیم‌گیری میانی و دایره قرمز طبقه‌بندی نادرست را نشان می‌دهد. دایره سبز طبقه‌بندی درستی می‌دهد. E) جنگل تصادفی. نقطه زرد نشان‌دهنده داده‌های ورودی است. پس از چندین تصمیم تصادفی مختلف، درختان ساخته می‌شوند. دایره نارنجی معیارهای تصمیم‌گیری میانی را نشان می‌دهد. دایره قرمز نشان‌دهنده طبقه‌بندی نادرست است. دایره سبز نشان‌دهنده طبقه‌بندی صحیح است. F) ماشین بردار پشتیبانی. نقاط زرد نشان‌دهنده یک نمونه جدید برای طبقه‌بندی است (Silva et al., 2019).

چالش‌های مختلف حوزه کشاورزی به کار گرفته شده است. این روش دارای قدرت پیش‌بینی بیشتری است (Paul et al., 2020).

۴-۷-۲-۱- الگوریتم‌های مبتنی بر چگالی

مانند خوشه‌بندی فضایی مبتنی بر چگالی برنامه‌ها با نویز یا DBSCAN است. مزیت خوشه‌بندی مبتنی بر چگالی این است که شکل خوشه می‌تواند دلخواه باشد (Sun et al., 2020).

۳-۷-۲-۱- یادگیری عمیق

این یک زیر کلاس تنظیم‌شده از خانواده یادگیری ماشینی برای یادگیری و ارائه داده‌ها است که یک روش الگوریتمی مبتنی بر شبکه‌های عصبی مصنوعی با چندلایه است که از ساختار و عملکرد مغز انسان الهام گرفته شده است. اخیراً فن‌های یادگیری عمیق در

مورد عملکردهای miRNA نسبت به تنش‌های غیرزنده گیاهی بهبود می‌بخشد. NB برای پیش‌بینی تنش گیاه با داشتن غلظت miRNA به‌عنوان ورودی استفاده می‌شود (Vakilian, 2020; Yan & Wang, 2022).

۲-۳- K-mean

تجزیه و تحلیل خوشه‌ای K-mean برای کشف پاسخ‌های ژنوتیپ‌های ماش در شرایط مختلف شور استفاده شد و نتایج نشان داد افزایش سطح تنش شوری اثرات نامطلوب شدید بر پارامترهای مورفولوژیکی و بیوشیمیایی مختلف مانند طول ریشه، طول ساقه، وزن ریشه، وزن اندام هوایی، کلروفیل-a، کلروفیل-b و محتوای فنلی دارد (Aslam et al.).

ارقام مختلف برنج ممکن است واکنش و مکانیسم‌های متفاوتی نسبت به خشکی داشته باشند. آنالیزهای فیزیولوژیکی و مقایسه پروتئومی برنج یاس تایلندی و دو رقم برنج چک در پاسخ به تنش خشکی با توجه به نتایج حاصل از خوشه‌بندی K-mean نشان داد که بیان پروتئین را می‌توان با توجه به تجمع بالای آن در هر رقم، به سه گروه طبقه‌بندی کرد (Maksup et al., 2014).

۲-۴- شبکه‌های عصبی

ANN ها برای بهینه‌سازی توابع سیگموئید متعلق به نورون‌های لایه پنهان به داده‌های آموزشی زیادی نیاز دارند. بنابراین، در تحقیقاتی که تعداد نمونه‌های آموزشی کم است، نمی‌توان فرآیند بهینه‌سازی را به‌درستی انجام داد. در مطالعاتی از جمله روابط بین غلظت miRNA و تنش گیاهی، استفاده از شبکه‌های عصبی مصنوعی (ANN) به‌عنوان الگوریتم یادگیری امکان‌پذیر نیست زیرا شبکه‌های عصبی مصنوعی به داده‌های آموزشی زیادی نیاز دارند (Vakilian, 2020).

فرض شده است که با توضیح خواص پپتیدهای کوچک درگیر در تنش، شبکه‌های عصبی مصنوعی و روش‌های یادگیری ماشینی می‌توانند به طبقه‌بندی مؤثر این پپتیدهای کوچک کمک کنند (Tripathi et al., 2021).

۲-۵- جنگل‌های تصادفی

جنگل‌های تصادفی می‌توانند به‌طور مؤثر مجموعه‌های داده بزرگ را که دارای متغیرهای زیادی است مدیریت کنند و همچنین رتبه‌بندی نسبی ویژگی‌ها را ارائه می‌دهد که می‌تواند در

۲- کاربرد یادگیری ماشین در شناخت مکانیسم‌های

مولکولی

۲-۱- ماشین بردار پشتیبان

SVM و انواع آن پرکاربردترین الگوریتم‌ها برای پیش‌بینی ژن‌های مقاومت به بیماری‌ها هستند (Sun et al., 2020). همچنین SVM عملکرد خوبی برای پیش‌بینی پروتئین‌های مقاوم به بیماری در پنج گونه گیاهی داشت (Shaik & Ramakrishna, 2014).

ابزار PlantMiRNAPred طبقه‌بندی miRNA ها را بر اساس یک الگوریتم SVM برای پیش‌بینی پیش‌miRNA های گیاهی انجام می‌دهد و در کشف miRNA ها با توجه به ایمنی سیستم گیاهی کارآمد است (Daneshafrooz et al., 2022; Silva et al., 2019).

مقدار متوسط غلظت miRNA نسبت به سطوح مختلف تنش گیاه برای ساخت یک پایگاه داده مناسب برای یادگیری ماشین استفاده می‌شود. در تحقیقی که کیوان عاصف پور و کیلیان در سال ۲۰۲۰ انجام داد، عملکرد SVM نسبت به سایر روش‌های یادگیری ماشین در پیش‌بینی تنش گیاه بهتر بود. SVM توانست خروجی را با $R^2 = 0.96$ پیش‌بینی کند، به این معنی که اگر غلظت miRNA-396، miRNA-159، miRNA-169 و miRNA-393 را که به ترتیب بیشترین سهم را در پاسخ گیاه به خشکی، شوری، سرما و گرما داشتند در برگ گیاه Arabidopsis thaliana اندازه‌گیری کنیم، شانس خوبی برای پیش‌بینی تنش وجود دارد (Vakilian, 2020).

یادگیری ماشین همچنین برای انتخاب ژن‌های درگیر در تحمل شوری در برنج با استفاده از نوعی الگوریتم SVM (SVM-RFE) کارآمد است (Silva et al., 2019).

انواع مختلف فراوانی‌های ترکیبی توالی (فراوانی اسیدآمین، فراوانی دی‌پپتید، فراوانی تری‌پپتید، فراوانی‌های متعدد، بار و ترکیب آب‌گریزی) با بردارهای مشخص برای مدل‌های آموزشی با الگوریتم SVM برای پیش‌بینی پروتئین NBS-LRR استفاده شده است (Silva et al., 2019).

۲-۲- Naïve Bayes

فراوانی اسیدهای آمینه و طبقه‌بندی کننده ساده بیز برای پیش‌بینی پروتئین‌های ترشح شده قارچ‌های بیماری‌زا ترکیب شدند (Silva et al., 2019). یادگیری ماشینی دانش ما را در

پروموترها و مکان‌های اتصال پروتئین اعمال شود (Silva et al., 2019).

۸-۲- یادگیری ماشین و تجزیه و تحلیل شبکه

روش‌های یادگیری ماشینی و آماری برای تحلیل شبکه، یک چارچوب قابل دسترس برای تجزیه و تحلیل ساختاری نمودارها با کنار هم قرار دادن رویکردهای شناخته شده و جدید در کلاس‌های گراف و معیارهای گراف برای طبقه‌بندی فراهم می‌کند. از مجموعه داده‌های واقعی برای نمایش روش‌ها و موضوعات مورد بحث استفاده می‌شود که بررسی رویکردهای محاسباتی برای بازسازی و تقسیم شبکه‌های بیولوژیکی، مقدمه‌ای بر شبکه‌های پیچیده، ویژگی‌های آماری و مدل‌ها، مدل‌سازی برای شبکه‌های زیستی در حال تکامل، ساختار یک نمودار دویخشی تصادفی در حال تکامل، شمارش بر اساس چگالی در داده‌های سازمان‌یافته را شامل می‌شوند (Dehmer & Basak, 2012; Hesami et al., 2022).

۳- نتیجه‌گیری

یادگیری ماشین به استفاده از رایانه‌ها برای شبیه‌سازی فعالیت‌های یادگیری انسان و روش‌های خود به‌سازی رایانه‌ها می‌پردازد که برای به دست آوردن دانش و مهارت‌های جدید به‌طور مداوم عملکرد و دستاوردها را بهبود می‌بخشد. در واقع یادگیری ماشینی یک تکنیک داده‌کاوی هوشمند است که یک مدل پیش‌بینی را بر اساس یادگیری قبلی ایجاد می‌کند. در مقایسه با یادگیری انسان، یادگیری ماشینی سریع‌تر یاد می‌گیرد. رایانه‌ها بر جامعه انسانی تأثیر می‌گذارند و هرگونه پیشرفت انسان در زمینه یادگیری ماشینی باعث توسعه تحقیقات جدید در زمینه‌های نظری خواهد شد، بنابراین توسعه مدل‌ها و الگوریتم‌ها از اهمیت ویژه‌ای برخوردار است. بر اساس بررسی منابع صورت گرفته، مشخص گردید الگوریتم‌های درخت تصمیم، ماشین بردار پشتیبان (SVM)، و شبکه‌های عصبی در تنش‌های زیستی نسبت به تنش‌های غیر زیستی کارآمدتر می‌باشند. این درحالی است که الگوریتم‌های مبتنی بر رگرسیون خطی، مدل‌های مارکوف، و روش‌های مبتنی بر تشخیص الگو در مطالعات مربوط به تنش‌های غیرزنده کارآمدتر می‌باشند. با این وجود مطالعه بیشتر و شناسایی الگوریتم‌های تشخیصی و روش‌های ترکیبی در تحقیقات آتی باید بیشتر مورد توجه قرار بگیرد.

تفسیر مدل و داده‌ها مفید باشد (Singh et al., 2017). بردار پشتیبان بازگشتی (R-SVM) و جنگل‌های تصادفی، ۱۹۶ ژن مشترک با دامنه پویا از سطوح بیان را در تنش‌های متعدد و مقاومت در برنج نشان دادند. ژن‌های شناسایی شده که به شدت تنش‌های غیر زیستی و زیستی را متمایز می‌کنند، به‌عنوان مؤلفه‌های کلیدی برای درک بیشتر ماهیت ذاتاً پیچیده پاسخ در تنش‌های متعدد در گیاهان کاربرد دارد (Shaik & Ramakrishna, 2014). در مطالعه‌ای، مدل‌های رگرسیون، یعنی درخت تصمیم (DT)، ماشین‌های بردار پشتیبان (SVMs)، و ساده بیز (NB) برای پیش‌بینی تنش گیاه با داشتن غلظت miRNAهای گیاهی استفاده شدند (Vakilian, 2020).

۶-۳- درخت تصمیم

مدل جنگل تصادفی چند کلاسه (MCRF) برای پیش‌بینی و تشخیص دسته‌های خاص پیتیدهای کوچک ترشح‌شده ناشی از استرس در گیاهان استفاده می‌شود. مشاهده شد که مدل پیشنهادی در زمان و هزینه صرفه‌جویی می‌کند (Tripathi et al., 2021). RNAهای غیر کد کننده طویل (lncRNAs) در انواع فرآیندهای بیولوژیکی از جمله تنش دخیل هستند. برای پیش‌بینی lncRNA در گیاهان با استفاده از داده‌های رونویسی، PLncPRO مبتنی بر یادگیری ماشین و استفاده از الگوریتم جنگل تصادفی برای طبقه‌بندی رونوشت‌های کدگذاری و رونوشت‌های طولانی غیر کدگذاری استفاده می‌شود. با استفاده از الگوریتم جنگل تصادفی، با اطمینان بالا، ۳۷۱۴ و ۳۴۵۷ lncRNA به ترتیب در برنج و نخود و ویژگی‌های مختلف و بیان متفاوت تحت شرایط تنش خشکی یا شوری بررسی شد و lncRNAها از طریق RTqPCR تأیید شد (Singh et al., 2017). در زمینه‌های مختلف ژنومیک از جمله شناسایی ارتباط بین SNPها و صفات، شناسایی مکان‌های اتصال DNA به پروتئین نیز استفاده شده است (Sun et al., 2020).

۷-۲- یادگیری عمیق

یادگیری عمیق یک روش جدید ML است و برای شناسایی محل‌های اتصال پروتئین جدید استفاده شده است. الگوریتم DeepBind یک الگوریتم یادگیری عمیق است که بر بسیاری از الگوریتم‌های سنتی ML برتری دارد. DeepBind یک ابزار همه‌کاره است که می‌تواند برای ریزآرایه‌ها، داده‌های توالی یابی،

References

- Anim-Ayeko, A. O., Schillaci, C., & Lipani, A. (2023). Automatic blight disease detection in potato (*Solanum tuberosum* L.) and tomato (*Solanum lycopersicum*, L. 1753) plants using deep learning. *Smart Agricultural Technology*, 4, 100178.
- Arora, A., Misra, T., Kumar, M., Marwaha, S., Kumar, S., & Chinnusamy, V. (2023). Computer Vision Approaches for Plant Phenotypic Parameter Determination. In *Digital Ecosystem for Innovation in Agriculture* (pp. 263-270). Springer.
- Aslam, M., Maqbool, M. A., Zaman, Q. U., & Akhtar, M. A. Uncovering the tolerance of mungbean (*Vigna radiata* L. Wilczek) genotypes under saline conditions using k-mean cluster analysis.
- Behmann, J., Schmitter, P., Steinrücken, J., & Plümer, L. (2014). Ordinal classification for efficient plant stress prediction in hyperspectral data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 29-36.
- Bhatia, A., Chug, A., & Singh, A. P. (2020). Plant disease detection for high dimensional imbalanced dataset using an enhanced decision tree approach. *International Journal of Future Generation Communication and Networking*, 13(4), 71-78.
- Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 13, 1063-1095.
- Bock, C., Poole, G., Parker, P., & Gottwald, T. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical reviews in plant sciences*, 29(2), 59-107.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Calderón, R., Navas-Cortés, J. A., & Zarco-Tejada, P. J. (2015). Early detection and quantification of *Verticillium* wilt in olive using hyperspectral and thermal imagery over large areas. *Remote Sensing*, 7(5), 5584-5610.
- Cen, H., Weng, H., Yao, J., He, M., Lv, J., Hua, S., Li, H., & He, Y. (2017). Chlorophyll fluorescence imaging uncovers photosynthetic fingerprint of citrus Huanglongbing. *Frontiers in plant science*, 8, 1509.
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215.
- Cynthia, S. T., Hossain, K. M. S., Hasan, M. N., Asaduzzaman, M., & Das, A. K. (2019). Automated detection of plant diseases using image processing and faster R-CNN algorithm. 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI),
- Daneshafrooz, N., Bagherzadeh Cham, M., Majidi, M., & Panahi, B. (2022). Identification of potentially functional modules and diagnostic genes related to amyotrophic lateral sclerosis based on the WGCNA and LASSO algorithms. *Scientific reports*, 12(1), 20144.
- Das Choudhury, S., Guadagno, C. R., Bashyam, S., Ewers, B. E., & Samal, A. Stress Phenotyping Analysis Leveraging Autofluorescence Image Sequences with Machine Learning. *Frontiers in plant science*, 15, 1353110.
- Dehmer, M., & Basak, S. C. (2012). *Statistical and machine learning approaches for network analysis*. Wiley Online Library.
- Ghahramani, N., Shodja, J., Rafat, S. A., Panahi, B., & Hasanpur, K. (2021). Integrative systems biology analysis elucidates mastitis disease underlying functional modules in dairy cattle. *Frontiers in Genetics*, 12, 712306.
- Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., & Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics*, 2(3), 156-183.
- Girshick, R. (2015). Fast r-cnn. Proceedings of the IEEE international conference on computer vision,
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition,
- Gou, C., Zafar, S., Hasnain, Z., Aslam, N., Iqbal, N., Abbas, S., Li, H., Li, J., Chen, B., & Ragauskas, A. J. (2024). Machine and Deep Learning: Artificial Intelligence Application in Biotic and Abiotic Stress Management in Plants. *Frontiers in Bioscience-Landmark*, 29(1), 20.
- Hesami, M., Alizadeh, M., Jones, A. M. P., & Torkamaneh, D. (2022). Machine learning: Its challenges and opportunities in plant system biology. *Appl Microbiol Biotechnol*, 106(9-10), 3507-3530.

- Holasou, H. A., Panahi, B., Shahi, A., & Nami, Y. (2024). Integration of machine learning models with microsatellite markers: New avenue in world grapevine germplasm characterization. *Biochemistry and Biophysics Reports*, *38*, 101678.
- Islam, M., Dinh, A., Wahid, K., & Bhowmik, P. (2017). Detection of potato diseases using image segmentation and multiclass support vector machine. 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE),
- Kaiser, E., Von Gillhaussen, P., Clarke, J., & Schurr, U. (2024). IPPS 2022-plant phenotyping for a sustainable future. *Frontiers in plant science*, *15*, 1383766.
- Kaundal, R., Kapoor, A. S., & Raghava, G. P. (2006). Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC bioinformatics*, *7*(1), 1-16.
- Kumar, S. (2021). Plant disease detection using CNN. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(12), 2106-2112.
- Maksup, S., Roytrakul, S., & Supaibulwatana, K. (2014). Physiological and comparative proteomic analyses of Thai jasmine rice and two check cultivars in response to drought stress. *Journal of plant interactions*, *9*(1), 43-55.
- Meraj, T., Sharif, M. I., Raza, M., Alabrah, A., Kadry, S., & Gandomi, A. H. (2024). Computer vision-based plants phenotyping: A comprehensive survey. *Isience*, *27*(1).
- Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., Singh, A., Singh, A. K., & Ganapathysubramanian, B. (2017a). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant methods*, *13*, 1-12.
- Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., Singh, A., Singh, A. K., & Ganapathysubramanian, B. (2017b). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant methods*, *13*(1), 1-12.
- Niazian, M., & Niedbala, G. (2020). Machine learning for plant breeding and biotechnology. *Agriculture*, *10*(10), 436.
- Panahi, B., Frahadian, M., Dums, J. T., & Hejazi, M. A. (2019). Integration of cross species RNA-Seq meta-analysis and machine-learning models identifies the most important salt stress-responsive pathways in microalga *Dunaliella*. *Frontiers in Genetics*, *10*, 752.
- Panahi, B., Mohammadi, S. A., & Doulati-Baneh, H. (2020). Characterization of Iranian grapevine cultivars using machine learning models. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, *90*, 615-621.
- Panahi, B., Tajaddod, S., Mohammadzadeh Jallali, H., Hejazi, M. A., & Zeinalabedini, M. (2022). Variability and association among some pomological and physiochemical traits in spring frost tolerant genotypes of Persian walnut (*Juglans regia* L.) and selection of genotypes with superior traits based on machine learning algorithms. *Genetic Resources and Crop Evolution*, 1-13.
- Paul, A., Ghosh, S., Das, A. K., Goswami, S., Choudhury, S. D., & Sen, S. (2020). A review on agricultural advancement based on computer vision and machine learning. In *Emerging technology in modelling and graphics* (pp. 567-581). Springer.
- Peña, J. M., Torres-Sánchez, J., Serrano-Pérez, A., De Castro, A. I., & López-Granados, F. (2015). Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors*, *15*(3), 5609-5626.
- Pierz, L. D., Heslinga, D. R., Buell, C. R., & Haus, M. J. (2023). An image-based technique for automated root disease severity assessment using PlantCV. *Applications in Plant Sciences*, *11*(1), e11507.
- Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., Bulat, A., Tzimiropoulos, G., Wells, D. M., & Murchie, E. H. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience*, *6*(10), gix083.
- Pujari, D., Yakkundimath, R., & Byadgi, A. S. (2016). SVM and ANN based classification of plant diseases using feature reduction technique. *IJIMAI*, *3*(7), 6-14.
- Rajesh, B., Vardhan, M. V. S., & Sujihelen, L. (2020). Leaf disease detection and classification by decision tree. 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184),
- Raza, S.-e.-A., Smith, H. K., Clarkson, G. J., Taylor, G., Thompson, A. J., Clarkson, J., & Rajpoot, N. M. (2014). Automatic detection of regions in spinach canopies responding to soil moisture deficit using combined visible and thermal imagery. *Plos one*, *9*(6), e97612.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

- Reza, M. N., Na, I. S., Baek, S. W., & Lee, K.-H. (2019). Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. *Biosystems engineering*, 177, 109-121.
- Sadeghi, M., Panahi, B., Mazlumi, A., Hejazi, M. A., Komi, D. E. A., & Nami, Y. (2022). Screening of potential probiotic lactic acid bacteria with antimicrobial properties and selection of superior bacteria for application as biocontrol using machine learning models. *LWT*, 162, 113471.
- Sari, W. E., Kurniawati, Y. E., & Santosa, P. I. (2020). Papaya Disease Detection Using Fuzzy Naïve Bayes Classifier. 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI),
- Sarkar, C., Gupta, D., Gupta, U., & Hazarika, B. B. (2023). Leaf disease detection using machine learning and deep learning: Review and challenges. *Applied Soft Computing*, 110534.
- Seka, D., Bonny, B., Yoboué, A., Sié, S., & Adopo-Gourène, B. (2019). Identification of maize (*Zea mays* L.) progeny genotypes based on two probabilistic approaches: Logistic regression and naïve Bayes. *Artificial intelligence in agriculture*, 1, 9-13.
- Shaik, R., & Ramakrishna, W. (2014). Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. *Plant physiology*, 164(1), 481-495.
- Shoab, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in plant science*, 14, 1158933.
- Shrestha, G., Das, M., & Dey, N. (2020). Plant disease detection using CNN. 2020 IEEE Applied Signal Processing Conference (ASPCON),
- Silva, J. C. F., Teixeira, R. M., Silva, F. F., Brommonschenkel, S. H., & Fontes, E. P. (2019). Machine learning approaches and their current application in plant molecular biology: A systematic review. *Plant Science*, 284, 37-47.
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science*, 21(2), 110-124.
- Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends in plant science*, 23(10), 883-898.
- Singh, U., Khemka, N., Rajkumar, M. S., Garg, R., & Jain, M. (2017). PLncPRO for prediction of long non-coding RNAs (lncRNAs) in plants and its application for discovery of abiotic stress-responsive lncRNAs in rice and chickpea. *Nucleic acids research*, 45(22), e183-e183.
- Sun, S., Wang, C., Ding, H., & Zou, Q. (2020). Machine learning and its applications in plant molecular studies. *Briefings in functional genomics*, 19(1), 40-48.
- Tahmasebi, A., Niazi, A., & Akrami, S. (2023). Integration of meta-analysis, machine learning and systems biology approach for investigating the transcriptomic response to drought stress in *Populus* species. *Scientific reports*, 13(1), 847.
- Tripathi, A., Goswami, T., Trivedi, S. K., & Sharma, R. D. (2021). A multi class random forest (MCRF) model for classification of small plant peptides. *International Journal of Information Management Data Insights*, 1(2), 100029.
- Trivedi, V. K., Shukla, P. K., & Pandey, A. (2022). Automatic segmentation of plant leaves disease using min-max hue histogram and k-mean clustering. *Multimedia Tools and Applications*, 81(14), 20201-20228.
- Vakilian, K. A. (2020). Machine learning improves our knowledge about miRNA functions towards plant abiotic stresses. *Scientific reports*, 10(1), 1-10.
- van Dijk, A. D. J., Kootstra, G., Kruijer, W., & de Ridder, D. (2021). Machine learning in plant science and plant breeding. *Iscience*, 24(1), 101890.
- Virnodkar, S. S., Pachghare, V. K., Patil, V., & Jha, S. K. (2020). Remote sensing and machine learning for crop water stress determination in various crops: a critical review. *Precision Agriculture*, 21(5), 1121-1155.
- Wang, Q., & Qi, F. (2019). Tomato diseases recognition based on faster RCNN. 2019 10th International Conference on Information Technology in Medicine and Education (ITME),
- Wetterich, C. B., Kumar, R., Sankaran, S., Junior, J. B., Ehsani, R., & Marcassa, L. G. (2013). A comparative study on application of computer vision and fluorescence imaging spectroscopy for detection of citrus huanglongbing disease in USA and Brazil. *Laser Science*,
- Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., Shen, S., Jiang, B., & Ma, X. (2020). Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant methods*, 16, 1-13.

- Yadav, J., & Sharma, M. (2013). A Review of K-mean Algorithm. *Int. J. Eng. Trends Technol*, 4(7), 2972-2976.
- Yan, J., & Wang, X. (2022). Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. *The Plant Journal*, 111(6), 1527-1538.
- Zhou, J., Fu, X., Zhou, S., & Zhou, J. (2018). Evaluation of the performance of machine learning methods in soybean segmentation for image-based high-throughput phenotyping in greenhouse. 2018 ASABE Annual International Meeting,
- Zubler, A. V., & Yoon, J.-Y. (2020). Proximal Methods for Plant Stress Detection Using Optical Sensors and Machine Learning. *Biosensors*, 10(12), 193.