सांख्यिकी-विमर्श
2024

अंक
20

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली - 110012

सांख्यिकी-विमर्श

2024

अंक

20

संपादक मंडल	
प्रवीण आर्य	अध्यक्ष
मो. समीर फारुकी	
सुशील कुमार सरकार	
मो. हारुन	सदस्य
सुधीर श्रीवास्तव	
नितिन जोशी	

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान

लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली – 110012

सांख्यिकी-विमर्श

2024

अंक

20

द्वारा प्रकाशित

निदेशक

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान

लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली - 110012

मुद्रण : जुलाई 2025

रूपरेखा एवं मुद्रणः नक्षत्र आर्ट, ए 31, बेसमेंट, नारायणा इण्डस्ट्रिल एरिया फेस–2, नई दिल्ली–110028 फोन. .91.11.45636883

निदेशक का संदेश ....

प्रिय पाठकगण,

राजभाषा पत्रिका के इस नवीन संस्करण के प्रकाशन की आप सभी को हार्दिक शुभकामनाएं। इस नए संस्करण में आपको संस्थान की राजभाषा हिंदी में की गई विविध गतिविधियों, लेखन-प्रयासों, और तकनीकी-शैक्षणिक नवाचारों का संगठित रूप मिलेगा। बीते वर्ष हमने हिंदी को केवल प्रशासनिक भाषा ही नहीं, बल्कि अनुसंधान, संवाद, साहित्य और डिजिटल प्लेटफॉर्म में भी मजबूती से स्थापित किया।

इस वर्ष हमने कृषि-सांख्यिकी और सूचना विज्ञान हेतु हिंदी डिजिटल सामग्री जारी की, शोध-संकलन को सुदृढ़ किया तथा साहित्यिक संवादों में विविधता को समाहित किया। विशेष रूप से वैज्ञानिक लेखों के साथ-साथ कविताओं के सम्मेलन से यह पत्रिका और अधिक ज्ञानवर्धक एवं जीवंत बन पड़ी है।

मैं पूरे हृदय से उन सभी लेखकों, संपादकों, तकनीकी टीम और अनुवादकों का हार्दिक धन्यवाद ज्ञापन करता हूँ जिनके अथाह सहयोग एवं योगदान से यह पत्रिका सजीव, समृद्ध और प्रेरणादायक बनी है।

आइए, हम सभी मिलकर हिंदी को नवाचार, अनुसंधान और संवाद की भाषा के रूप में और अधिक पुष्ट करें। इससे हमारी राजभाषा की पहुँच तीव्र होगी और हमारी संस्थागत पहचान नई ऊँचाइयों को प्राप्त करेगी। यह प्रयास केवल एक पत्रिका नहीं, बल्कि राजभाषा हिंदी को नवोदय की तरफ ले जाने का प्रतिफल है। इस माध्यम से हम न केवल ज्ञान साझा करते हैं बल्कि प्रेरित करते हैं कि हिंदी में सोचें, लिखें और संवाद करें।

आपके बहुमूल्य सुझाव एवं प्रतिक्रियाएँ हमें और अधिक उत्कृष्ट दिशा की ओर ले जाएंगी।

कृपया अपनी प्रतिक्रिया हमें भेजिए और इस यात्रा के अनमोल साथी बनिए।

राजेन्द्र प्रसाद

निदेशक

संपादकीय

सांख्यिकी विज्ञान, गणित की एक अत्यंत महत्वपूर्ण और विश्लेषणपरक शाखा है, जिसका उपयोग जानकारी को सुव्यवस्थित करने, समझने, प्रस्तुत करने और विश्लेषण करने में किया जाता है। यह विज्ञान हमें आंकड़ों की गणना, उनके मूल्यांकन, व्याख्या, अनुमान तथा सारगर्भित प्रस्तुति की कला सिखाता है। इसके माध्यम से हम न केवल डेटा की सूक्ष्मताओं को पहचान सकते हैं, बल्कि उसमें छिपी नियमितताओं और संभावित अनियमितताओं को भी उजागर कर सकते हैं। सांख्यिकी का उपयोग आज जीवन के लगभग हर क्षेत्र जैसे कृषि, चिकित्सा, अर्थव्यवस्था या प्रौद्योगिकी मे हो रहा है। आंकड़ों का सार्थक विश्लेषण नीतियों के निर्माण से लेकर निर्णय-निर्धारण तक एक मजबूत आधार प्रदान करता है।

यह हर्ष का विषय है कि सांख्यिकी विमर्श का यह बीसवाँ अंक आपके समक्ष प्रस्तुत है। इस निरंतर यात्रा की सफलता में हमारे पाठकों एवं लेखकों का अमूल्य योगदान रहा है। हमारा लक्ष्य सांख्यिकी विमर्श की लोकप्रियता के साथ साथ इसकी गुणवत्ता, विषयवस्तु और प्रस्तुति को उच्चतम स्तर तक पहुंचाने के लिए प्रतिबद्ध हैं। हमें विश्वास है कि आने वाले समय में सांख्यिकी विमर्श और अधिक आकर्षक, ज्ञानवर्धक एवं विचारोत्तेजक स्वरूप में आपके समक्ष प्रस्तुत होता रहेगा।

इस अंक में सांख्यिकी एवं कृषि साहित्य से संबंधित अनेक समसामयिक एवं अनुसंधानपरक विषयों पर लेख प्रकाशित किए गए हैं। इनमें प्रमुख हैं - कृषि क्षेत्र में कृत्रिम बुद्धिमत्ता एवं एकाधिक फ्रेम सर्वेक्षण का उपयोग, फसलों में रोगों के लिए अर्ध-लैटिन आयत अभिकल्पनाएं, इम्प्यूटेशन विधियों की समालोचना, पूर्वानुमान हेतु मशीन लर्निंग तकनीकों का प्रयोग, खरपतवार वर्गीकरण में डीपनेट मॉडल, तथा उपज अनुमान हेतु मशीन लर्निंग विधियाँ। हमें पूर्ण विश्वास है कि यह अंक पाठकों के लिए न केवल ज्ञानवर्धक होगा, बल्कि उनकी अनुसंधान की दिशा को भी समृद्ध करेगा।

इन वैज्ञानिक लेखों के साथ-साथ इस अंक में कुछ साहित्यिक एवं सांस्कृतिक रचनाएँ भी सम्मिलित की गई हैं - जैसे कि संस्थान की राजभाषा यात्रा पर आलेख, डॉ. दरोगा सिंह स्मृति व्याख्यान तथा स्वरचित कविताएं। संपादक मंडल सांख्यिकी विमर्श के सभी लेखकों के प्रति आभार प्रकट करता है। पत्रिका की गुणवत्ता में सुधार हेतु कुमारी सुनीता, उपनिदेशक, राजभाषा के बहुमूल्य सुझाव सराहनीय है। आइए, इस ज्ञानयात्रा को निरंतर आगे बढ़ाएं और सांख्यिकी विमर्श को उत्कृष्टता की नई ऊंचाइयों तक पहुंचाएं।

प्रवीण आर्य

अध्यक्ष, संपादक मण्डल

अनुक्रमणिका

डॉ. दरोगा सिंह स्मृति व्याख्यान • श्री राकेश कुमार त्यागी, पूर्व महानिदेशक, राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय (NSSO)	1
अनुसंधान खण्ड	
कृषि में कृत्रिम बुद्धिमत्ता (एआई) और मशीन लर्निंग • हर्ष सचान, शालिनी कुमारी, मो. अशरफुल हक, चंदन कुमार देब, सपना निगम, संचिता नाहा, सुदीप मरवाह	7
फसलों के लिए कृत्रिम बुद्धिमत्ता आधारित रोग पहचान प्रणाली (AI-DISC): फसलों में जैविक तनाव की पहचान के लिए एक समाधान • सुदीप मरवाह, मो. अशरफुल हक, चंदन कुमार देब, सपना निगम, अक्षय धीरज, मधु, अंशु भारद्वाज और अलका अरोड़ा	13
खरपतवार वर्गीकरण के लिए एक ध्यान तंत्र संलग्न डेंसनेट मॉडल • अक्षय धीरज, सपना निगम, मधु, मो. अशरफुल हक, सुदीप मरवाह	19
डेयरी श्रिया : डेयरी पशुओं के स्वास्थ्य और उत्पादकता में सुधार के लिए एक कृत्रिम बुद्धिमत्ता चालित चैटबॉट • संचिता नाहा, रूपसी तिवारी, चन्दन कुमार देब, मधु, सपना निगम एवं सुदीप मरवाह	24
कृषि में छवि-आधारित खरपतवार पहचान के लिए कृत्रिम बुद्धिमत्ता • सपना निगम, वी. के. चौधरी, अक्षय धीरज, मधु, मो अशरफुल हक, रजनी जैन, सुदीप मरवाह	29
एएसएमआईआर: पौधों में अजैविक तनाव-विशिष्ट माइक्रो आरएनए के पूर्वानुमान हेतु मशीन लर्निंग की रूपरेखा • उपेन्द्र कुमार प्रधान, प्रविन कुमार मेहेर और एस.पी. सिंह	36
कृषि में आर्टिफिशियल इंटेलिजेंस के अनुप्रयोग • मधु, सपना निगम, अक्षय धीरज, मो. अशरफुल हक, सुदीप मरवाहा	53
पौधों की बीमारियों के परीक्षणों में अर्ध-लैटिन आयत अभिकल्पनाओं का उपयोग • कौशल कुमार यादव एवं सुकान्त दाश	56
कृषि में एकाधिक फ्रेम सर्वेक्षण • भारती, कौस्तव आदित्य, तौकीर अहमद, राहुल बनर्जी, दीपक सिंह, पंकज दास, अंकुर बिश्वास, राजू कुमार	60
फैमिली ऑफ लाइफटाइम डिस्ट्रिब्यूशन की विश्वसनीयता विशेषताओं के लिए प्रारंभिक परीक्षण अनुमानक और विश्वास अंतराल • हिमाद्री घोष एवं नितिन जोशी	65
मशीन लर्निंग तकनीकों का उपयोग करके फसल उपज का अनुमान • पंकज दास, भारती, राहुल बनर्जी, अंकुर बिश्वास, तौकीर अहमद, कौस्तव आदित्य	70

मशीन लर्निंग तकनीकों का उपयोग करके गेहूं की फसल पर लैंडस्केप डायग्नोस्टिक सर्वेक्षण डेटा का फीचर चयन और उपज का अनुमान • सौमेन पाल, अलका अरोडा, सुदीप मरवाह, रंजित कुमार पॉल, अजीत गुप्ता, कमलिका नाथ, सौरभ कुमार, रामकंवर मालिक, राजर्षि रॉय बर्मन एवं उधम सिंह गौतम	76
मशीन लर्निंग का उपयोग करके लघु-अणुओं की जैव-सक्रियता का पूर्वानुमान • स्नेहा मुर्मू, सुधीर श्रीवास्तव, दीपा भट्ट, सौम्या शर्मा, ऋत्विका दास, मोहम्मद समीर फारूकी, गिरीश कुमार झा	83
अर्जुन की छाल के अर्क का उपयोग करके रेशमी कपड़े की रंगाई • अभिलाषा रांगी, ललित जाजपुरा, राहुल बनर्जी	88
सिंगल-सेल सिक्वेंसिंग प्रक्रिया एवं विश्लेषण का संक्षिप्त विवरण • सुधीर श्रीवास्तव, मयंक रश्मि, दीपा भट्ट, द्विजेश चंद्र मिश्र, स्नेहा मुर्मू, मोहम्मद समीर फारूकी, कृष्ण कुमार चतुर्वेदी, गिरीश कुमार झा	93
खाद्य विज्ञान एवं पोषण में सांख्यिकीय विश्लेषण तकनीकें • डॉ. निधि जोशी	97
राजस्थान में प्रमुख तिलहन फसलों के क्षेत्रफल के लिए आपूर्ति प्रतिक्रिया फलन • दिनेश कुमार, रविन्द्र सिंह शेखावत एवं प्रवीण आर्य	100
रोगाणुरोधी प्रतिरोध: जैव सूचना विज्ञान-संचालित पहचान उपकरण, रुझान और संस्थागत पहल • रागिनी कुशवाहा, अनु शर्मा, द्विजेश चंद्र मिश्रा, मोहम्मद समीर फारूकी, सुधीर श्रीवास्तव, शशि भूषण लाल	106
मिसिंग डेटा को संभालने के लिए इम्प्यूटेशन विधियों की समीक्षा • राणा क्रिना दिव्येशभाई, पंकज दास, भारती, अंकुर बिश्वास, तौकीर अहमद, दीपक सिंह, राजू कुमार	112
राजभाषा खण्ड	
भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान	120
कविताएं	122
हिन्दी कार्यशाला-2024	132
हिन्दी पखवाड़ा-2024	137
कुछ यादगार लम्हें.....	138

डॉ. दरोगा सिंह स्मृति व्याख्यान

एन.एस. एस.ओ. (NSSO) के देशव्यापी सैम्पल सर्वे: परिवर्तन, विकास और सुधार की अंतर्निहित प्रक्रिया

श्री राकेश कुमार त्यागी

पूर्व महानिदेशक, राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय (NSSO)

नेशनल सैम्पल सर्वे

आजादी के बाद 18 दिसंबर 1949 को प्रधानमंत्री जवाहरलाल नेहरू ने आवश्यक जानकारी एकत्र करने के लिए पूरे देश में सैम्पल सर्वे आयोजित किये जाने की बात की। प्रो. महालनोबिस ने 25 दिसंबर 1949 तक, यानि एक सप्ताह में ही, 1930 के दशक के मध्य में भारतीय सांख्यिकी संस्थान (आईएसआई) द्वारा बड़े पैमाने पर किए गए सैम्पल सर्वेक्षणों से प्राप्त अनुभव के आधार पर नेशनल सैम्पल सर्वे (एनएसएस) आयोजित करने के लिए एक सार योजना तैयार कर दी। नेशनल सैम्पल सर्वे का पहला सर्वे 1 अक्टूबर 1950 को ग्रामीण भारत के लिए डेटा संग्रहण से शुरू हुआ। सर्वे के बढ़ते महत्व और माँग को देखते हुए भारत सरकार ने 1 जून 1972 को नेशनल सैम्पल सर्वे का कार्य एवं कार्य से संबद्ध कर्मचारियों को भारतीय सांख्यिकी संस्थान (आईएसआई) से सीधे अपने हाथ में ले लिया गया। तब से आज तक भारत सरकार में राष्ट्रीय महत्व के विषयों पर देशव्यापी सर्वे करने के लिए मुख्य रूप से एन.एस. एस.ओ.(NSSO) जिम्मेदार है। आज NSSO दुनिया का सबसे बड़ा सर्वेक्षण संगठन है जो नीति निर्माण के साथ-साथ शोध के लिए विभिन्न सामाजिक-आर्थिक विषयों पर महत्वपूर्ण डेटा प्रदान करता है।

सर्वे की बढ़ती माँग व क्षमता

देश के विकास के लिए योजना बनाने और नीति निर्माण में सर्वे द्वारा प्राप्त परिणामों की उपयोगिता और इस उपयोग के मद्देनजर देशव्यापी सैम्पल सर्वे करने की आवश्यकता बढ़ती गई और अब तक यह कई गुना बढ़ गई है। एन.एस. एस.ओ. अपने पहले से तय प्रोग्राम के अंतर्गत चल रहे सर्वेक्षण कार्यक्रमों तथा उपलब्ध सीमित संसाधनों के कारण अनेक बार अपने आप को इस बढ़ती माँग को पूरा करने में असमर्थ पाता है। अत: "निजी क्षेत्र की सर्वेक्षण क्षमताओं का आकलन" का अध्ययन करने का काम भारतीय कृषि सांख्यिकी अनुसंधान संस्थान (IASRI), नई दिल्ली को सौंपा गया, जिसका उद्देश्य निजी संगठनों/एजेंसियों की सर्वेक्षण क्षमताओं का आकलन करना था। इस आकलन में केवल कुछ ही संस्थाएँ थी जिन्होंने देशव्यापी सर्वे करने की क्षमता होने का दावा किया था लेकिन इनके बारे में भी इनके संसाधनों की पर्याप्तता और संसाधनों की भौगोलिक उपस्थिति में एकरूपता की कमी, विशेषतया पूर्वोत्तर राज्यों और दूर दराज स्थित यूनियन टेरिटरीज में, उनके देशव्यापी सर्वे करने की क्षमता के दावों की पूर्ण पुष्टि नहीं करती थी ।

एन.एस. एस.ओ. (NSSO) क्यों विशिष्ट है

देशव्यापी सर्वे करने की अद्वितीय क्षमता के अतिरिक्त अन्य कई संबंधित विशेषतयाएं हैं जो एन.एस. एस.ओ. को विशिष्ठ बनाती है। इन विशेषताओं में प्रमुख है एन.एस. एस.ओ. की सर्वे प्रणाली जो एक लंबे समय के दौरान विकसित हुई हैं। साथ ही इसके विकास में देश और विदेश के विशेषज्ञों का योगदान है। एनएसएस की पहली समीक्षा समिति की अध्यक्षता सर रोनाल्ड ए. फिशर ने की थी तथा श्री मॉरिस हैनसेन (अमेरिका), आर्थर लिंडर (स्विट्जरलैंड), टोसियो कितागावा (जापान) और फ्रैंक येट्स (यूके) इस समीक्षा समिति के सदस्य थे ।

सर्वेक्षण के डिजाइन और सर्वेक्षण के निष्पादन में शामिल विभिन्न चरणों और गतिविधियों के दौरान सैंपलिंग और गैर सैंपलिंग त्रुटियों, दोनों को, न्यूनतम करने पर ध्यान दिया जाता है। यद्यपि कुछ सर्वे में रेस्पोंडेंट का समय लेने वाली लंबी प्रश्नावली, परिणामों के प्रकाशन में विलम्ब, लगभग वही पुराना डिजाइन, संपन्न शहरी परिवारों से जानकारी लेने में समस्या जैसे मुद्दे आलोचना का विषय रहे हैं, फिर भी एन.एस. एस.ओ. का बेहतर विकल्प नहीं है यानि देशव्यापी बड़े सर्वे से जुड़ी इन जैसी स्वाभाविक समस्याओं का यथा संभव बेहतर जवाब भी एन.एस.एस. में ही है, क्योंकि इसके सिस्टम में अपने

अनुभवों और अपनी आलोचनाओं को लगातार अपने सुधार के लिए इस्तेमाल करने का एक स्वीकार्य और सक्षम प्रणाली निहित है ।

स्वतंत्रता और स्वायत्तता

एन.एस. एस.ओ. की गतिविधियों के समन्वय और सुपरविजन की जिम्मेदारी महानिदेशक पर है और यह राष्ट्रीय सांख्यिकी आयोग (एनएससी) के मार्गदर्शन में कार्य करता है। यह व्यवस्था एन.एस. एस.ओ. को इसके देश भर में फैले संसाधनों और मैनपॉवर से प्राप्त सर्वे क्षमता का विवेकपूर्ण उपयोग करने के साथ साथ इसे अपने कार्य करने और कार्यप्रणाली लागू करने के लिए पूर्ण स्वतंत्रता और स्वायत्तता प्रदान करती है जो किसी अन्य सरकारी या निजी सर्वे एजेंसी को प्राप्त नहीं है। एन.एस.एस.ओ. की स्वतंत्रता और स्वायत्तता राष्ट्रीय सांख्यिकी आयोग (NSC) द्वारा सुनिश्चित की जाती है। राष्ट्रीय सांख्यिकी आयोग में एक अंशकालिक अध्यक्ष, 4 अंशकालिक सदस्य और एक पदेन सदस्य होते हैं, जिनमें से प्रत्येक के पास निर्दिष्ट सांख्यिकीय क्षेत्र में विशेषज्ञता और अनुभव होता है। राष्ट्रीय सांख्यिकी आयोग (NSC) के अध्यक्ष और 4 अंशकालिक सदस्य, सभी नॉन-ऑफिसियल होते हैं ताकि यह और इसके मार्गदर्शन में NSSO स्वतंत्रता और स्वायत्तता के साथ काम कर सके। एन.एस. एस.ओ. के सर्वेक्षणों में शामिल किए जाने वाले विषय और उनकी आवधिकता का निर्णय राष्ट्रीय सांख्यिकी आयोग द्वारा विभिन्न मंत्रालयों की आवश्यकता/अनुरोधों पर विचार करते हुए किया जाता है। विषय विशेष पर सर्वे की मेथाडोलॉजी तैयार करने के लिए राष्ट्रीय सांख्यिकी आयोग एक वर्किंग ग्रुप गठित करती है जो केवल राष्ट्रीय सांख्यिकी आयोग के प्रति जवाबदेह है। वर्किंग ग्रुप में सर्वे विशेष के विषय से संबंधित विशेषज्ञ, सांख्यिकीविद्, सर्वे परिणाम के उपयोगकर्ता मंत्रालयों/संगठनों के अधिकारी और चयनित राज्यों के सांख्यिकीय अधिकारी शामिल होते हैं, जो सर्वेक्षण उपकरणों के विकास में उच्च स्तर बनाए रखने के साथ साथ उपयोग की दृष्टि से परिणामों की प्रासंगिकता सुनिश्चित करने में योगदान करते हैं। इस वर्किंग ग्रुप का अध्यक्ष भी नॉन-ऑफिसियल रखा जाता है और इसके सदस्यों में भी पर्याप्त संख्या में नॉन ऑफिशियल विशेषज्ञ सदस्य रखे जाते हैं ताकि वर्किंग ग्रुप के कार्य की स्वतंत्रता और स्वायत्तता बनाए रखी जा सके।

स्वतंत्रता और स्वायत्तता का योगदान

एन.एस. एस.ओ. के सर्वेक्षणों में शामिल किए जाने वाले विषयों और उनकी आवधिकता का निर्धारण सरकार द्वारा न होकर राष्ट्रीय सांख्यिकी आयोग द्वारा विभिन्न मंत्रालयों की आवश्यकता/अनुरोधों पर विचार करते हुए किया जाता है। विषय का चुनाव यह सुनिश्चित करता है कि सरकारी संसाधन उस सर्वे के लिए इस्तेमाल होंगे जिससे प्राप्त जानकारी देश के विकास के लिए बनाई जाने वाली योजनाओं और नीति निर्धारण में उपयोगिता की दृष्टि से अन्य विषयों पर सर्वे द्वारा प्राप्त जानकारी के मुकाबले ज्यादा महत्व रखती है। सर्वे प्रश्नावली में वे प्रश्न भी शामिल किये जा पाते हैं जिनके माध्यम से प्राप्त जानकारी मौजूदा नीतियों और कार्यक्रमों के कार्यान्वयन से अपेक्षित नतीजों/जानकारी के विरुद्ध हो सकती है। सर्वे के वे नतीजे भी प्रकाशित हो पाते हैं जो सरकार की मौजूदा नीतियों और कार्यक्रमों के कार्यान्वयन से अपेक्षित व इच्छित नतीजे प्राप्त होने की पुष्टि नहीं करते लेकिन जिनके आधार पर नीतियों, कार्यक्रमों और योजनाओं में आवश्यक संशोधन करते हुए इच्छित और बेहतर नतीजे प्राप्त किये जा सकते हैं।

पारदर्शिता का योगदान

न केवल संकलित परिणाम बल्कि सर्वेक्षण पद्धति के साथ सर्वेक्षण के इकाई स्तर के आंकड़े भी सार्वजनिक उपयोग के लिए जारी किए जाते हैं। अतः सर्वे पद्धति, इकाई स्तर के आंकड़े और परिणाम सभी सार्वजनिक जाँच के लिए उपलब्ध होते हैं। यह पारदर्शिता सुनिश्चित करती है कि आप एक स्तरहीन सर्वे मेथाडोलॉजी या फर्जी / असंगत डेटा के साथ विशेषज्ञों व उपयोगकर्ताओं की जायज और सीधी आलोचना से नहीं बच सकेंगे। पारदर्शी कार्यप्रणाली और सर्वे पद्धति प्राप्त सर्वे परिणामों की बेहतर गुणवत्ता सुनिश्चित करने के लिए एक सतत दवाब बनाए रखती है और उन्हें विश्वशनीयता और व्यापक स्वीकार्यता देती है। इकाई स्तर के आंकडों का सार्वजनिक उपयोग के लिए जारी होना सुनिश्चित करता है कि त्रुटिपूर्ण परिणाम विशेषज्ञों / उपयोगकर्ता से नहीं बच पाएंगे ।

पर्याप्त संसाधन

NSSO द्वारा किए जाने वाले सर्वो की संख्या, सर्वे में शामिल किए जाने वाले गांव और शहरों की संख्या आदि का निर्धारण एन.एस. एस.ओ. के पास उपलब्ध मैनपॉवर और कार्य के लिए तय नॉर्म के अनुसार किए जाते हैं। ये नॉर्म NSC (पहले गवर्निंग कॉन्सिल) द्वारा तय किये जाते हैं। पर्याप्त बजट होने के कारण NSSO को राष्ट्रीय सांख्यिकीय आयोग द्वारा अनुमोदित सर्वे पद्धति के अनुसार डेटा कलेक्शन कार्य करने और

स्थापित संबंधित नॉर्म के अनुसार सर्वे के लिए आवश्यक प्रशिक्षणों, निरीक्षणों आदि करने में किसी समझौते/कटौती की आवश्यकता नहीं पड़ती। नया सर्वे शुरू करने से पहले पायलट सर्वे करना, किसी सर्वे को शुरू करने से पहले उसकी सर्वे पद्धति और प्रश्नावली की प्री-टेस्टिंग और प्री-टेस्टिंग के आधार पर सर्वे उपकरण में आवश्यक संशोधन जैसे सभी कार्य, जो सर्वे की सफलता और गुणवत्ता सुनिश्चित करते हैं, अनिवार्यता पूर्वक किए जाते हैं। पायलट सर्वे के सफल होने के बाद नए सर्वे या एक अवधि के बाद किसी पुराने विषय पर देश व्यापक सर्वे को दोहराने का निर्णय लिए जाने के बाद वर्किंग ग्रुप के गठन और इसके द्वारा सर्वे पद्धति के विकास के लिए लगभग एक साल लंबा पर्याप्त समय दिया जाता है ताकि सर्वे की सफलता, पूर्ण उपयोगिता और गुणवत्ता निश्चित की जा सके।

अधिकारियों और कर्मचारियों का प्रशिक्षण

सर्वे गुणवत्ता सुनिश्चित करने के लिए अधिकारियों और कर्मचारियों का पर्याप्त और नियमित प्रशिक्षण NSSO के सर्वे कार्यक्रम का अहम हिस्सा है। ये प्रशिक्षण राष्ट्रीय, राज्य और क्षेत्रीय स्तर पर आयोजित किए जाते हैं। राज्य और क्षेत्रीय स्तर के प्रशिक्षकों में प्रशिक्षण यथा संभव ज्यादा से ज्यादा क्षेत्रीय भाषा में आयोजित किया जाता है। राज्य / क्षेत्रीय प्रशिक्षण में पर्याप्त समय की थियोरेटिकल ट्रेनिंग के बाद एक दिन का समय फील्ड में व्यावहारिक प्रशिक्षण को दिया जाता है और उस दौरान प्राप्त फीडबैक संबंधित क्षेत्रीय ऑफिस द्वारा सुझाव के साथ मुख्यालय भेजी जाती हैं ताकि सर्वे शुरू करने से पहले एक बार फिर सर्वे उपकरण और प्रशिक्षण सामग्री में आवश्यक संशोधन किए जा सके और प्रशिक्षण के दौरान कार्यालय विशेष के उपयोगी अनुभव का लाभ अन्य कार्यालयों को भी मिल सके। सर्वे शुरू होने के एक महीने बाद फिर क्षेत्रीय प्रशिक्षण कार्यक्रम आयोजित किए जाते हैं ताकि इस शुरुआती चरण के दौरान आई समस्याओं के संबंध में आवश्यक निर्देश और स्पष्टीकरण दिए जा सके। इसके बाद भी प्रशिक्षणों को समय समय पर आवश्यकता के अनुसार दोहराए जाने का भी प्रावधान रखा गया है।

टीम अप्रोच

पर्याप्त प्रशिक्षण के बावजूद एकत्र किये जाने वाले डेटा की गुणवत्ता सुनिश्चित करने के लिए अतिरिक्त सावधानी बरती जाती है। सर्वे के शुरू के दौर में डेटा एकत्रित करने जा रहे जूनियर इन्वेस्टिगेटर या नए इन्वेस्टिगेटर के साथ हैंड होल्डिंग के लिए एक वरिष्ठ इन्वेस्टिगेटर या सुपरवाइजर को भेजा जाता है ताकि वास्तविक सर्वे के दौरान आने वाली किसी अनपेक्षित समस्या का हल एवं प्रशिक्षण से ज्ञान अर्जित करने में रह गई किसी कमी की भरपाई वरिष्ठ अधिकारी के अनुभव और ज्ञान द्वारा की जा सके।

निगरानी और निरीक्षण

सर्वे करने के लिए NSSO के फील्ड ऑपरेशंस डिविजन (FOD) के देश भर में 6 जोनल कार्यालय, 53 क्षेत्रीय कार्यालय (आरओ) और 118 उप-क्षेत्रीय कार्यालय हैं जिन पर सर्वेक्षण के प्रबंधन और निगरानी की जिम्मेदारी है। निगरानी और निरीक्षणों के लिए एक स्थापित व्यवस्था और न्यूनतम कार्य मानक तय है। सर्वे के दौरान एकत्र किए जा रहे आंकड़ों के साथ साथ चलने वाली वेरिफिकेशन और वैलिडेशन की व्यवस्था है ताकि संदेहास्पद पहलुओं के संबंध में आवश्यक सुधारों के लिए समय से कार्यवाई की जा सके । डेटा के सत्यापन और वेलिडेशन के लिए उपयोग किए जा रहे सॉफ्टवेयर में व्यापक कंप्यूटर जांच बिंदु और अन्य वेलिडेशन चेक रखे जाते हैं। डेटा सत्यापन/ वेलिडेशन की कठोर प्रक्रिया के बाद प्राप्त परिणामों के आधार पर तैयार की गई रिपोर्ट के मसौदे पर एन.एस. एस.ओ. के विभिन्न प्रभाग के अधिकारियों और विषय से जुड़े कार्य समूह के अधिकारियों के बीच गहन चर्चा की जाती है ताकि अंतिम अनुमोदन के लिए रिपोर्ट का मसौदा राष्ट्रीय सांख्यिकी आयोग को प्रस्तुत करने से पहले हर तरह की संभावित त्रुटिया दूर की जा सके।

सतत विकास व सुधार

एन.एस. एस.ओ. लगातार अपने सर्वे के दौरान प्राप्त अनुभव और आई समस्याओं की जानकारी का उपयोग सतत अपने विकास और सुधार के लिए करता है। सर्वे के दौरान, सर्वे के विभिन्न कार्य से संबंधित प्रभाग, अपने अनुभव के साथ साथ भविष्य में सुधार के सुझाव नोट करते हैं । सर्वे संबंधित परिणाम के प्रकाशित होने के बाद सरकारी और गैर सरकारी उपयोगकर्ताओं और शोधकर्ताओं से परिणाम पर विश्लेषणात्मक /समीक्षात्मक शोध पत्र मांगे जाते हैं। इन शोध पत्रों सहित सर्वे के दौरान एन.एस. एस.ओ. के प्रभागों द्वारा पहले से संकलित किये गए अपने अनुभव और सुधार के सुझाव तथा परिणामों पर लिखे गए शोधपत्र 2 दिन के सालाना सेमिनार में प्रस्तुत किए जाते हैं। इन सेमिनारों के माध्यम से आने वाले सर्वो में, विशेषतया

उसी विषय पर दोहराए जाने वाले सर्वे में, किए जाने वाले सुधार पर चर्चा की जाती है। समय समय पर डेटा उपयोगकर्ताओं के सम्मेलन आयोजित किये जाते हैं ताकि भविष्य में किये जाने वाले सर्वे से उनकी अपेक्षाओं की पूर्ति को और बेहतर बनाया जा सके। अपने अनुभव और दूसरों की आलोचना/समीक्षा और सुझाव को सतत सुधार का आधार बनाते हुए विभिन्न स्तर पर ब्रेन-स्टॉर्मिंग सेशन आयोजित किए जाते हैं जिनकी अध्यक्षता अध्यक्ष, राष्ट्रीय सांख्यिकी आयोग/सचिव,सांख्यिकी एवं कार्यक्रम कार्यान्वन मंत्रालय/ महानिदेशक, एन.एस. एस.ओ. जैसे उच्च स्तर के अधिकारी द्वारा की जाती है। इन सभी चर्चाओं के परिणाम संस्थागत स्मृति/ज्ञान के रूप में दस्तावेजित किए जाते हैं।

हाल ही में किए गए कुछ मुख्य सुधार

डेटा की पूलीन्ग

एन.एस. एस.ओ. के अपने केन्द्रीय सैम्पल का साइज राज्य स्तर के भरोसेमंद अनुमान देने के लिए पर्याप्त है लेकिन इससे निचले, यानि जिला स्तर के भरोसेमंद अनुमान देने के लिए पर्याप्त नहीं है। अतः एनएसएस के सर्वेक्षण कार्यक्रम में राज्य सरकारों की भागीदारी की शुरुआत हुई ताकि सैम्पल साइज बढ़ाया जा सके और डेटा के दोनों सेटों को मिलाकर उप-राज्य स्तर पर, विशेष रूप से जिला स्तर पर बेहतर अनुमान लगाना संभव हो सके। इस भागीदारी में यह डेटा तो बहुत पहले से एकत्र किया जाता रहा है लेकिन राज्यों, विशेषतया छोटे राज्यों, के संसाधनों व प्रशिक्षण से संबंधित कई समस्याओं के कारण कई राज्यों में जिलेवार अनुमान जारी नहीं किये जा पा रहे थे या देरी से किये जाते थे। धीरे-धीरे यह भागीदारी सफल हो रही है और समय से जिलेवार अनुमान जारी करने वाले राज्यों की संख्या लगातार बढ़ रही है ।इस सफलता में एन.एस. एस.ओ. द्वारा जारी किये गए सॉफ्टवेयर और सॉफ्टवेयर के उपयोग में दिए गए प्रशिक्षण का बड़ा योगदान है।

फील्ड से डेटा की डायरेक्ट एंट्री और परिणाम जारी करने में लगने वाले समय में कमी

डेटा कलेक्शन कार्य पूरा होने के बाद परिणाम जारी होने तक 1 साल से ज्यादा लगता था। इस समय को घटाकर 1 साल किया गया । इसके बाद 2017-18 में पहली बार फील्ड कार्यालयों द्वारा अपने पर्सनल कंप्यूटर पर डेटा एंट्री करके भेजा गया और साथ ही नए पीरियोडिक लेबर फोर्स सर्वे में वर्ल्ड बैंक द्वारा विकसित CAPI सॉफ्टवेयर का इस्तेमाल करते हुए सीधे फील्ड से डेटा एंट्री करके डेटा प्रोसेसिंग के लिए भेजा गया। इस बदलाव से परिणाम जारी करने में लगने वाला समय 1 साल से घट कर 6 महीने पर आ गया। आज जब पूरी डेटा एंट्री सीधे फील्ड से करने के लिए सर्वे कर्मियों के पास पर्याप्त टेबलेट और अपना संबंधित सॉफ्टवेयर और सिस्टम उपलब्ध है तब लगभग 3 महीने के अंदर परिणाम जारी किए जा रहे हैं।

सांख्यिकी-विमर्श
2024

अंक
20

अनुसंधान खण्ड

कृषि में कृत्रिम बुद्धिमत्ता (एआई) और मशीन लर्निंग

हर्ष सचान, शालिनी कुमारी, मो. अशरफुल हक, चंदन कुमार देब, सपना निगम, संचिता नाहा, सुदीप मरवाहा

भा.कृ.अनु.प.–भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

शहरी और ग्रामीण दोनों ही अर्थव्यवस्थाओं का एक अनिवार्य पहलू खेती, वैश्विक जनसंख्या वृद्धि के संबंध में दक्षता और पर्यावरणीय स्थिरता में सुधार करने के लिए बढ़ते दबाव में है। रोग, कीट हमले, अनुचित उर्वरक उपयोग, सीमित सिंचाई और मिट्टी की गुणवत्ता में गिरावट जैसी चुनौतियों के कारण खेती में जटिलताएँ उत्पन्न होती हैं। हाल ही में, आर्टिफिशियल इंटेलिजेंस (AI) और मशीन लर्निंग (ML) स्थायी कृषि के लिए क्रांतिकारी उपकरण के रूप में उभरे हैं। AI/ML तकनीकों ने पौधों के स्वास्थ्य मूल्यांकन, पशुधन प्रबंधन, मृदा स्वास्थ्य मूल्यांकन, फसल पैदावार की भविष्यवाणी/पूर्वानुमान और आपूर्ति श्रृंखलाओं को अनुकूलित करने में सहायता करने की क्षमता दिखाई है। डेटा एक्सेस, बुनियादी ढांचे और उच्च लागत जैसी चुनौतियों के बावजूद, AI और ML में आगामी विकास में खेती को बदलने, इस बढ़ती आबादी की जरूरतों को पूरा करने और पर्यावरणीय स्थिरता को बढ़ावा देने की क्षमता है।

परिचय

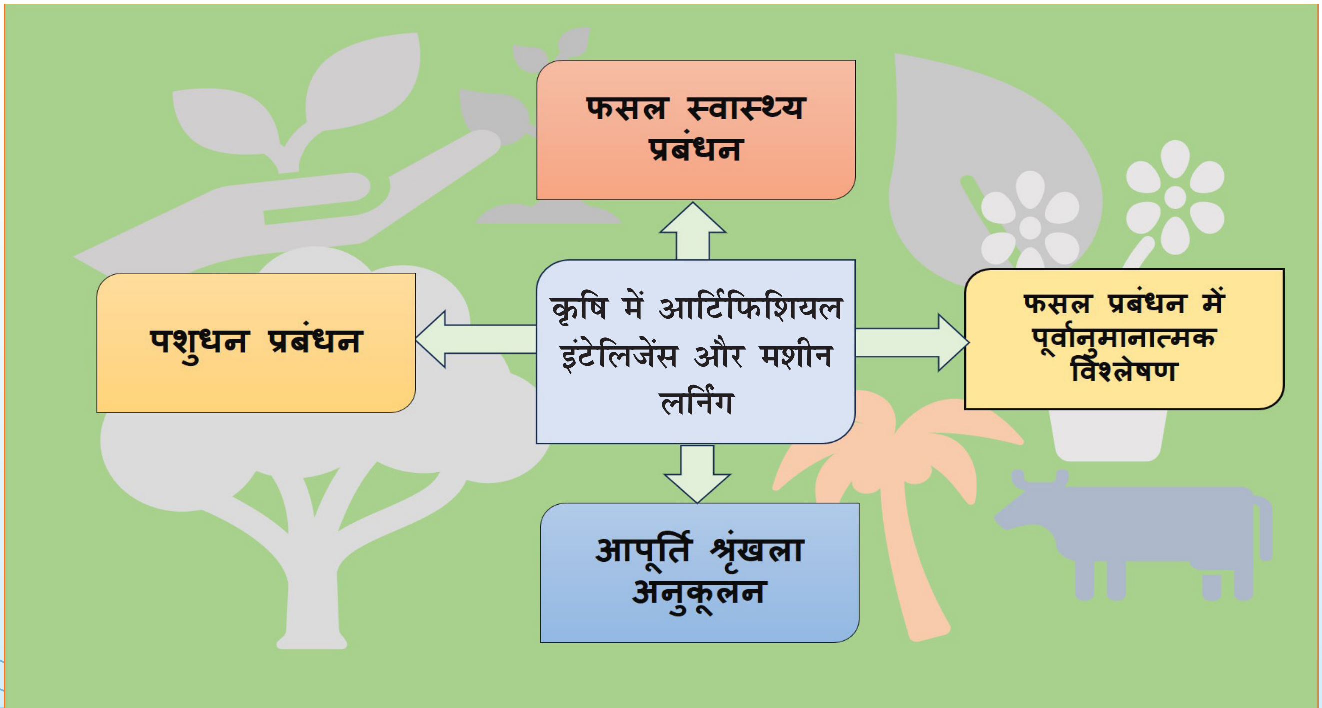
कृषि अर्थात खेती सबसे सम्मानजनक पेशों में से एक है, जो संपूर्ण मानवता को जीवन की सबसे बुनियादी आवश्यकता, भोजन प्रदान करता है। वैश्विक जनसंख्या कृषि और इससे जुड़े क्षेत्रों पर अत्यधिक निर्भर है, जो आर्थिक स्थिरता और सतत विकास सुनिश्चित करने में महत्वपूर्ण भूमिका निभाते हैं। संयुक्त राष्ट्र के अनुसार, वर्ष 2050 तक विश्व की जनसंख्या लगभग 10 अरब तक पहुँचने की संभावना है, जिसके लिए कुल खाद्य उत्पादन में 70% की वृद्धि की आवश्यकता होगी (मैककार्थी और अन्य, 2018)। इस बढ़ती जनसंख्या की खाद्य आवश्यकताओं को पूरा करने के लिए स्थायी रणनीतियों को अपनाना अनिवार्य है। हालाँकि, बीज बोने से लेकर फसल कटाई तक, किसानों को कई कठिनाइयों का सामना करना पड़ता है। इनमें बीमारियाँ और कीटों का प्रकोप, अपर्याप्त सिंचाई और जल निकासी व्यवस्था, मृदा स्वास्थ्य में गिरावट, खरपतवार प्रबंधन की समस्याएँ और कटाई के बाद होने वाले नुकसान शामिल हैं। इन चुनौतियों का समाधान करने के लिए फसलों और पशुधन की उत्पादकता बढ़ाना और गुणवत्ता बनाए रखना अत्यंत आवश्यक है। लेकिन, इससे कृषि तंत्र पर अधिक उत्पादन करने, पर्यावरणीय क्षति को कम करने और संसाधनों के संरक्षण का भारी दबाव पड़ता है।

इन समस्याओं का प्रभावी समाधान करने के लिए कृषि पारिस्थितिकी तंत्र की जटिलता और गतिशील प्रकृति को समझना जरूरी है। यह निरंतर निगरानी, मापन और विभिन्न भौतिक घटनाओं के विश्लेषण से संभव हो सकता है। हाल में, कंप्यूटर प्रौद्योगिकी के विकास से कृत्रिम बुद्धिमत्ता (AI) मानव जीवन के कई क्षेत्रों में, विशेष रूप से कृषि में, एक क्रांतिकारी उपकरण बन कर उभरा है (मरवाहा और अन्य, 2023a)। पिछले कुछ वर्षों में, AI ने कृषि और संबंधित क्षेत्रों में परिवर्तन लाने की अपनी संभावनाएँ सिद्ध की हैं। AI उन्नत गणनात्मक तकनीकों को समेकित करता है, जो मशीनों को मानव बुद्धि की नकल कर जटिल वास्तविक जीवन की समस्याओं को हल करने में

सक्षम बनाता है। मशीन लर्निंग (ML) और डीप लर्निंग (DL), AI के प्रमुख घटक हैं, जिन्होंने पारंपरिक कृषि को स्मार्ट और बुद्धिमान कृषि प्रणालियों में बदल दिया है। डेटा-संचालित जानकारी, पूर्वानुमान मॉडल और स्वचालन के माध्यम से, AI और ML फसल रोग और कीट प्रबंधन, फसल उत्पादन का पूर्वानुमान, पशुधन स्वास्थ्य निगरानी, आपूर्ति श्रृंखला का सरलीकरण और कृषि उत्पादकता एवं स्थिरता में सुधार करने का लक्ष्य रखते हैं (चित्र. 1)।

खेती में एआई और एमएल के अनुप्रयोग

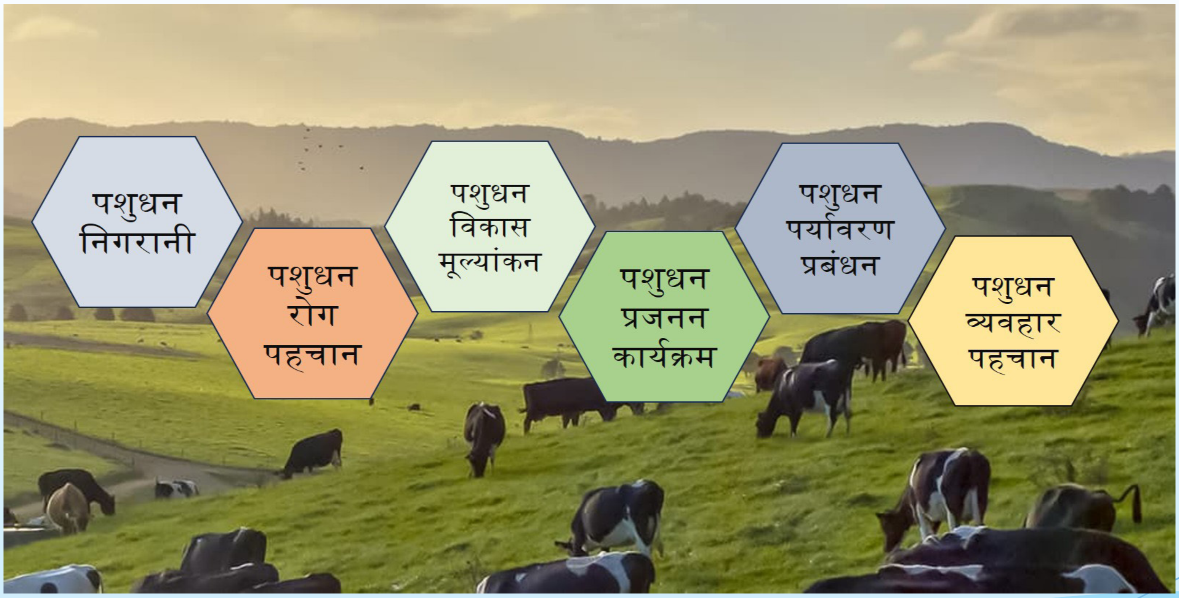
- **एआई/एमएल-संचालित फसल स्वास्थ्य प्रबंधनः** फसलें विभिन्न प्रकार के खतरों जैसे रोग संक्रमण, कीट-प्रकोप, खरपतवार, पोषक तत्वों की कमी आदि से गंभीर रूप से क्षतिग्रस्त होती हैं, जिससे उत्पादन और गुणवत्ता में भारी गिरावट आती है। अनुमान है कि रोग, कीट और खरपतवार सालाना फसल उत्पादन का 15-20 प्रतिशत तक नुकसान कर सकते हैं (गोयल और अन्य, 2022)। इन खतरों के खिलाफ पारंपरिक तरीकों की सीमाएँ हैं क्योंकि इनका उपयोग करने के लिए व्यापक ज्ञान की आवश्यकता होती है, जो कई किसानों के पास नहीं होता। इससे सटीक निदान करना चुनौतीपूर्ण हो जाता है। इसके अलावा, ये प्रक्रियाएँ अक्सर समय लेने वाली मैन्युअल जाँचों पर निर्भर होती हैं, जिससे समस्याओं को पहचानने और उनका समाधान करने में देरी हो सकती है। इस देरी से कीट या रोग फैल सकते हैं और अधिक नुकसान पहुँचा सकते हैं। इस संदर्भ में, AI (कृत्रिम बुद्धिमत्ता) और ML (मशीन लर्निंग) बेहतर समाधान प्रदान कर सकते हैं, जो कंप्यूटर विजन जैसी उन्नत तकनीकों का उपयोग करके कृषि खतरों की पहचान और नियंत्रण कर सकते हैं। कंप्यूटर विजन, AI का एक हिस्सा है, जिसका उद्देश्य कंप्यूटर को तस्वीरों और वीडियो का विश्लेषण कर अर्थपूर्ण जानकारी निकालने में सक्षम बनाना है, जैसे मानव दृष्टि करती है। कृषि में, कंप्यूटर विजन खेत से ली गई तस्वीरों को प्रोसेस करके रोगों के लक्षणों (हक और अन्य, 2022a; हक और अन्य, 2022b), कीट-प्रकोप, और खरपतवार (सचान और अन्य, 2023) की पहचान करता है। दृढ़ता से, AI-DISC: Artificial Intelligence-based Disease Identification System for Crops (मरवाहा और



चित्र 1: खेती में आर्टिफिशियल इंटेलिजेंस और मशीन लर्निंग के प्रमुख अनुप्रयोग

अन्य, 2023b) जैसे स्मार्ट मोबाइल एप्लिकेशन इन AI-आधारित समाधानों का सीधा उपयोग करते हैं और किसानों को त्वरित लाभ प्रदान करते हैं।

- **पशुधन निगरानी और प्रबंधन में एआई और एमएल:** AI और ML तकनीकों ने पशु स्वास्थ्य, उत्पादन, और समग्र फार्म प्रबंधन में सुधार करने की बड़ी क्षमता दिखायी है, जो पशुपालन निगरानी के क्षेत्र में क्रांति ला रही हैं (चित्र 2)। विभिन्न स्रोतों जैसे पहनने योग्य डिवाइस, कैमरे और पर्यावरणीय सेंसर से डेटा को प्रोसेस और विश्लेषण करके, पशुधन प्रबंधन में महत्वपूर्ण सुधार किया जा सकता है और जानवरों के स्वास्थ्य और व्यवहार की रियल-टाइम निगरानी की जा सकती है। AI-आधारित तकनीकों जैसे कंवल्यूशन न्यूरल नेटवर्क्स (CNNs), जटिल कार्यों को करने में सक्षम हैं, जैसे बीमारी के प्रारंभिक लक्षणों का पता लगाना, जानवरों की चेहरे की पहचान और जानवरों का वजन अनुमानित करना (देब और अन्य, 2023), जो प्रत्येक जानवर की वृद्धि और स्वास्थ्य की निगरानी में मदद करता है। AI और ML एल्गोरिदम का उपयोग विस्तृत आनुवंशिक डेटा प्रोसेस करने के लिए किया जा सकता है, जिससे प्रजननकर्ताओं को वांछनीय गुणों की पहचान करने और प्रजनन परिणामों का सही अनुमान लगाने में मदद मिलती है। इससे श्रेष्ठ जानवरों का चयन सुनिश्चित होता है, आनुवंशिक विविधता बनी रहती है और स्वस्थ व अधिक उत्पादक पशुधन को बढ़ावा मिलता है। इससे पशु कल्याण में सुधार, उत्पादकता में वृद्धि, और संसाधनों के अधिक कुशल उपयोग के परिणामस्वरूप, अंततः अधिक टिकाऊ और लाभकारी पशुपालन होता है। व्यापक रूप से, AI और ML तकनीक पशु और मुर्गी फार्म प्रबंधन और प्रजनन में सुधार लाने के लिए बड़ी उम्मीदें लेकर आई हैं, जो बेहतर निर्णय लेने, अधिक दक्षता और रियल-टाइम डेटा विश्लेषण के माध्यम से पशु कल्याण में मदद करती हैं।
- **एआई और एमएल का उपयोग करके फसल प्रबंधन के लिए पूर्वानुमानात्मक विश्लेषण:** फसल प्रबंधन के लिए पूर्वानुमान विश्लेषण AI और ML तकनीकों का उपयोग करता है, जो उन्नत डेटा विश्लेषण को लागू करके फसल प्रबंधन प्रथाओं को अनुकूलित करता है, जो पारंपरिक कृषि संचालन में महत्वपूर्ण परिवर्तन लाता है। AI और ML एल्गोरिदम का उपयोग फसल उपज का पूर्वानुमान, आदर्श फसल कटाई समय और मृदा स्वास्थ्य प्रबंधन के लिए किया जा सकता है, जिसमें विभिन्न प्रकार के डेटा जैसे उपग्रह चित्र, मौसम मानक,



चित्र 2: पशुधन प्रबंधन में एआई और एमएल का अनुप्रयोग

मृदा सेंसर, ऐतिहासिक फसल प्रबंधन पैरामीटर आदि शामिल हैं। कृषि खेतों से उत्पन्न विशाल और विविध डेटा सेट, जिसमें मौसम संबंधी डेटा, ऐतिहासिक उपज रिकॉर्ड, क्षेत्रीय सेंसर, उपग्रह डेटा शामिल हैं, को प्रोसेस और विश्लेषण किया जा सकता है ताकि फसल वृद्धि चरणों के दौरान त्वरित और प्रभावी निर्णय लिया जा सके। रिग्रेशन मॉडल, निर्णय वृक्ष और गहरे शिक्षण नेटवर्क कुछ AI और ML एल्गोरिदम के उदाहरण हैं, जो इस डेटा का विश्लेषण करते हैं और भविष्य की फसल उपज के बारे में अत्यधिक सटीक अनुमान प्रदान करते हैं (माजी और अन्य, 2022)। CNN मॉडल का भी फसल उपज के पूर्वानुमान और अनुमान को सुधारने में महत्वपूर्ण योगदान रहा है । इसके अतिरिक्त, AI और ML तकनीकों का उपयोग करके परिष्कृत प्रणालियाँ विकसित की जा सकती हैं, जो मृदा के स्वास्थ्य की निगरानी करती हैं, जैसे मृदा की नमी की मात्रा, pH स्तर, और पोषक तत्वों की उपलब्धता जैसे महत्वपूर्ण मृदा लक्षणों का विश्लेषण करके, और सिंचाई और उर्वरक के लिए अधिक प्रभावी वैज्ञानिक अनुशंसाएँ प्रदान करती हैं, जिससे अंततः मृदा की उर्वरता और फसल उपज में सुधार होता है । AI और ML में प्रगति निरंतर पूर्वानुमान विश्लेषण को सुधार रही है, जिससे संसाधन दक्षता और लाभप्रदता में सुधार होता है, और कृषि संचालन को अधिक टिकाऊ और लचीला बनाते हुए अधिक कुशल, टिकाऊ और उत्पादक कृषि प्रथाओं का वादा किया जा रहा है।

- **AI/ML संचालित आपूर्ति श्रृंखला अनुकूलनः** आपूर्ति श्रृंखला अनुकूलन कृषि में एक आवश्यक प्रक्रिया है, जिसका उद्देश्य उत्पादकता बढ़ाना, व्यय में कटौती करना और कृषि उद्योग में समग्र प्रदर्शन को मजबूत करना है। कृषि उद्योग ने AI और अत्याधुनिक तकनीकों के एकीकरण के परिणामस्वरूप जबरदस्त बदलाव देखा है, जो खेत से लेकर भोजन की थाली तक के संचालन को सरल बनाता है और जटिल रसद समस्याओं का समाधान प्रदान करता है (चित्र 3)। पूर्वानुमानात्मक विश्लेषण ऐतिहासिक डेटा और मशीन लर्निंग मॉडल का उपयोग करके बाजार मूल्य निर्धारण और कृषि उपज जैसे पैटर्न का पूर्वानुमान लगाता है, जिससे पूर्व-समायोजन संभव होता है और जोखिम कम होते हैं। IoT सेंसर और ब्लॉकचेन जैसी तकनीक के साथ, AI परिशुद्ध सूचना प्रवाह और पारदर्शिता सुनिश्चित करके आपूर्ति श्रृंखला दृश्यता में सुधार करता है। श्रेणीकरण और परिवहन जैसे कार्य स्वचालित रसद द्वारा संभाले जाते हैं, जिसमें रोबोट डिवाइस और चालक रहित कारें शामिल हैं, श्रम लागत कम होती है और उत्पादकता बढ़ती है । आपूर्ति रणनीतियों को बेहतर बनाने, अधिक उत्पादन से बचने और बर्बादी को कम करने के लिए, मांग पूर्वानुमान और अनुकूलन के लिए AI मॉडल ग्राहक व्यवहार और उद्योग के रुझानों की भी जांच करते हैं। कृषि आपूर्ति श्रृंखला अनुकूलन में एआई एक महत्वपूर्ण कारक है क्योंकि यह डेटा विश्लेषण, पूर्वानुमान कौशल और परिचालन दक्षता में सुधार करता है।

चित्र 3: कृषि में AI/ML संचालित आपूर्ति श्रृंखला अनुकूल

कृषि में AI और ML को अपनाने की सीमाएँ

क) डेटा की उपलब्धता और गुणवत्ताः AI और ML मॉडल के लिए विश्वसनीय और उच्च-गुणवत्ता वाले डेटा आवश्यक होते हैं, लेकिन कृषि डेटा अक्सर असंगत, अव्यवस्थित और प्राप्त करने में कठिन होता है।

ख) नेटवर्क और अवसंरचनाः अधिकांश कृषि कार्य ग्रामीण क्षेत्रों में होते हैं, जहां उपयुक्त अवसंरचना और विश्वसनीय इंटरनेट कनेक्टिविटी का अभाव होता है, जिससे AI और ML का प्रभावी उपयोग बाधित होता

है।

ग) व्यय और पूंजी निवेशः AI और ML तकनीक को लागू करना महंगा हो सकता है, जिसमें सेंसर, ड्रोन और डेटा प्रसंस्करण प्रणाली शामिल हैं, जो छोटे और मध्यम आकार के खेतों के लिए वित्तीय रूप से कठिन हो सकता है।

घ) तकनीकी दक्षताः किसानों और कृषि श्रमिकों में AI सिस्टम को स्थापित करने और प्रबंधित करने के लिए आवश्यक तकनीकी कौशल की कमी होती है, जिसके लिए शिक्षा और प्रशिक्षण कार्यक्रमों की आवश्यकता होती है।

ड) डेटा की सुरक्षा और गोपनीयताः डेटा सुरक्षा और गोपनीयता को लेकर चिंता हो सकती है, जिससे किसानों का विश्वास प्रभावित हो सकता है। AI तकनीक को अपनाने के लिए मजबूत डेटा शासन और कानूनी आश्वासन आवश्यक हैं।

च) मौजूदा प्रणालियों से कनेक्टिविटीः नए AI और ML सिस्टम को मौजूदा कृषि प्रणालियों और तकनीकी प्लेटफार्मों के साथ एकीकृत करना चुनौतीपूर्ण हो सकता है और इसके लिए सावधानीपूर्वक योजना और निष्पादन की आवश्यकता होती है।

छ) नैतिक और पर्यावरणीय पहलूः AI तकनीक का उपयोग ऐसे तरीके से किया जाना चाहिए जो पर्यावरण को नकारात्मक रूप से प्रभावित न करे या किसानों के बीच सामाजिक-आर्थिक असमानताओं को न बढ़ाए।

कृषि में भविष्य की प्रवृत्तियाँ और नवाचारः

AI और ML के कृषि में संभावित उपयोग व्यापक और परिवर्तनीय हैं, जो कई कृषि चुनौतियों का समाधान करते हुए उत्पादकता और स्थिरता को बढ़ाते हैं। ये नवाचार उन्नत फसल निगरानी को सक्षम करेंगे, जो उपग्रहों, ड्रोन और सेंसर से वास्तविक समय के डेटा का उपयोग करेंगे, जिससे सटीक कृषि और संसाधन आवंटन में मदद मिलेगी। AI-संचालित रोबोट खेतों में श्रम लागत को कम कर सकते हैं और संचालन की दक्षता में सुधार कर सकते हैं। पूर्वानुमान विश्लेषण कृषि पूर्वानुमान और जोखिम प्रबंधन को बेहतर बनाएंगे और बेहतर फसल परिणाम सुनिश्चित करेंगे। AI और ML आपूर्ति श्रृंखलाओं को भी सुव्यवस्थित कर सकते हैं, संदूषण को कम कर सकते हैं और संसाधनों की खपत और उनके पर्यावरणीय प्रभाव का विश्लेषण करके स्थायी कृषि पद्धतियों को बढ़ावा दे सकते हैं। भविष्य में, जनरेटिव AI कृषि के कई पहलुओं में क्रांतिकारी परिवर्तन ला सकता है, जिससे उत्पादकता, स्थिरता और रचनात्मकता में वृद्धि होगी। यह स्वचालित मशीनों के लिए उन्नत एल्गोरिदम विकसित करेगा, जो बुवाई, कटाई और खरपतवार प्रबंधन जैसे कार्यों को सटीक और दक्षता से अनुकूलित करेंगे। AI और ML से लैस स्व-चालित ट्रैक्टर और हार्वेस्टर न्यूनतम मानवीय संपर्क के साथ कार्य कर सकते हैं, जिससे श्रम लागत कम होगी और खेतों में काम में अनुकूलन होगा। इसके अतिरिक्त, AI/ML-सक्षम जैव प्रौद्योगिकी और पौधों की प्रजनन में प्रगति से स्थायी रूप से संशोधित फसलों की खेती संभव हो सकती है। कृषि में AI और ML का समावेश असाधारण दक्षता, स्थिरता और तकनीकी विकास का वादा करता है, जो कृषि पद्धतियों को बदलने और महत्वपूर्ण कृषि समस्याओं का समाधान करने में सहायक होगा।

निष्कर्ष

AI और ML का कृषि में उपयोग एक नया दृष्टिकोण है जो निरंतर समस्याओं का समाधान प्रदान करता है और उद्योग को अधिक उत्पादक, कुशल, और स्थायी तरीकों की ओर ले जाता है। AI प्रौद्योगिकियाँ जैसे कि स्वचालन, कंप्यूटर विजन, और पूर्वानुमान विश्लेषण निर्णय लेने में सुधार करती हैं, संसाधनों का अधिकतम उपयोग करती हैं, और उत्पादन बढ़ाती हैं। उदाहरण के लिए, AI-संचालित कीट और रोग निगरानी प्रणालियाँ समस्याओं को जल्दी पहचानने में मदद करती हैं, जिससे फसल हानि और कीटनाशकों के उपयोग में कमी आती है। AI-आधारित प्रणालियाँ पशु कल्याण को सुधारने, आहार अनुकूलन, प्रजनन कार्यक्रमों को बेहतर बनाने और मवेशियों के स्वास्थ्य प्रबंधन में सहायक हो सकती हैं। हालांकि, डेटा की उपलब्धता और गुणवत्ता,

अवसंरचनात्मक बाधाएँ, उच्च लागत, तकनीकी विशेषज्ञता और डेटा सुरक्षा समस्याएँ AI और ML को कृषि में अपनाए जाने में प्रमुख अड़चनें हैं। फिर भी, AI और ML के निरंतर विकास और विवेकपूर्ण उपयोग से कृषि में बड़े अवसर हैं, जो बढ़ती जनसंख्या को भोजन देने, पर्यावरणीय स्थिरता को सुधारने और किसानों की आजीविका को सुरक्षित करने में सहायक हो सकते हैं।

संदर्भ

देब, सी. के., तरफदार, ए., हक, एम. ए., मरवाहा, एस., भोज, एस., गौर, जी. के., & दत्त, टी. (2023) ।नियंत्रित इमेजिंग प्लेटफार्म में लाइव सूअर का वजन अनुमान लगाने के लिए कंवलूशन न्यूरल नेटवर्क (CNN)-आधारित प्रणाली। इंटरनेशनल कांफ्रेंस ऑन कम्युनिकेशन एंड इंटेलिजेंट सिस्टम्स में (पृ. 95-101) । स्प्रिंगर नेचर, सिंगापुर।

गोयल, एस., भाटिया, एम., & कुमार, पी. (2022)। रोग पहचान के लिए पौधों की विशेषताओं का खनन टेन्सरफ्लो: कृषि के लिए एक वरदान। राइजिंग थ्रेट्स इन एक्सपर्ट एप्लिकेशन्स एंड सॉल्यूशंस में (पृ. 361-371)। स्प्रिंगर, सिंगापुर।

हक, एम. ए., मरवाहा, एस., अरोड़ा, ए., देब, सी. के., मिश्रा, टी., निगम, एस., & हूडा, के. एस. (2022b)। मक्का की पत्तियों में मेयडिस लीफ ब्लाइट रोग के गंभीरता के चरणों की पहचान के लिए एक हल्का कंवलूशन न्यूरल नेटवर्क। फ्रंटियर्स इन प्लांट साइंस, 13, 1077568।

हक, एम. ए., मरवाहा, एस., देब, सी. के., निगम, एस., अरोड़ा, ए., हूडा, के. एस., ... & अग्रवाल, आर. सी. (2022a)। मक्का की फसल के रोगों की पहचान के लिए गहरी शिक्षा-आधारित दृष्टिकोण। साइंटिफिक रिपोर्ट्स, 12(1), 6334।

माजी, ए. के., मरवाहा, एस., कुमार, एस., अरोड़ा, ए., चिनुसामी, वी., & इस्लाम, एस. (2022)। SlypNet: उन्नत पौधों की फिनोटाइपिंग और कंप्यूटर विजन तकनीकों का उपयोग करके गेहूं की स्पाइकेल-आधारित उपज भविष्यवाणी। फ्रंटियर्स इन प्लांट साइंस, 13, 889853।

मरवाहा, एस., देब, सी. के., हक, एम. ए., नाहा, एस., माजी, ए. के. (2023a)। कृषि में कृत्रिम बुद्धिमत्ता और मशीन लर्निंग का अनुप्रयोग। इन: हरोहल्ली मस्तिगौड़ा, एम., गोपालारेड्डी, के., खोबरा, आर., सिंह, जी., प्रताप सिंह, जी. (संपादक) फिजियोलॉजिकल टूल्स को फसल प्रजनन को बढ़ाने के लिए अनुवादित करना। स्प्रिंगर, सिंगापुर।

मरवाहा, एस., अग्रवाल, आर. सी., पारसद, आर. , रामासुब्रमणियन, वी., अरोड़ा, ए., भारद्वाज, ए., अजीत, दहिया, एस., इस्लाम, एस. एन., देब, सी. के., हक, ए., निगम, एस., रे, एम., लामा, ए., पाल, एस. (2023b)। AI-DISC (कृत्रिम बुद्धिमत्ता-आधारित फसल रोग पहचान प्रणाली), कृषि प्रौद्योगिकी कोड: 201697795066168, ICAR-मान्यता प्राप्त तकनीक। उपलब्ध है: https://krishi.icar.gov.in/Technology/DetailReport.jsp?id=201697795066168।

मैककार्थी, यू., उयसल, आई., बदिया-मेलिस, आर., मर्सियर, एस., ओश्डोनेल, सी., & क्टेनीउदाकी, ए. (2018)। वैश्विक खाद्य सुरक्षा ख्र समस्याएँ, चुनौतियाँ और प्रौद्योगिकी समाधान। ट्रेंड्स इन फूड साइंस & टेक्नोलॉजी, 77, 11-20।

सचान, एच., इस्लाम, एस. एन., मिश्रा, एस., मरवाहा, एस., हक, ए., कुमार, एम., & पाल, एस. (2023)। कृत्रिम बुद्धिमत्ता तकनीकों का उपयोग करके गेहूं की फसल में घासफूस की पहचान। इंटरनेशनल जर्नल ऑफ एनवायरनमेंट एंड क्लाइमेट चेंज, 13(11), 4077-4083।

फसलों के लिए कृत्रिम बुद्धिमत्ता आधारित रोग पहचान प्रणाली (AI-DISC): फसलों में जैविक तनाव की पहचान के लिए एक समाधान

सुदीप मरवाह, मो. अशरफुल हक, चंदन कुमार देब, सपना निगम, अक्षय धीरज, मधु, अंशु भारद्वाज और अलका अरोड़ा

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

परिचय

भारत में, कृषि सबसे महत्वपूर्ण क्षेत्रों में से एक है जो समग्र अर्थव्यवस्था में महत्वपूर्ण भूमिका निभाता है। देश के कुल सकल घरेलू उत्पाद में कृषि और संबद्ध क्षेत्रों का हिस्सा लगभग 18% है (चंद और सिंह, 2022)। वित्तीय वर्ष 2021-2022 में, भारत ने लगभग 315.7 मिलियन टन खाद्यान्न का उत्पादन किया, जो पिछले वर्षों से 1.6% की उल्लेखनीय वृद्धि को दर्शाता है और देश की कृषि उत्पादकता को बढ़ाता है (रॉय, 2022) । खाद्यान्न के इस विशाल उत्पादन के बावजूद, फसल की बढ़ती अवधि के दौरान कई जैविक और अजैविक तनावों के कारण इस देश की बढ़ती आबादी को खिलाने के लिए एक बड़ी चुनौती मौजूद है। अध्ययनों से पता चला है कि भारत में बीमारी और कीट संक्रमण के कारण सालाना 35% तक फसल उपज खराब हो जाती है, जिससे देश की समग्र खाद्य आपूर्ति श्रृंखला, देश की आर्थिक स्थिरता और कृषक समुदायों की दैनिक आजीविका के लिए गंभीर खतरा पैदा हो जाता है (सिंह एट अल., 2020) । किसी भी जैविक बाधा का प्रबंधन करने के लिए , कारक को सही ढंग से पहचानना और फिर उसके अनुसार प्रबंधन रणनीति लागू करना बहुत महत्वपूर्ण है। इसलिए, खेत की स्थितियों में फसलों की बीमारियों की सही पहचान करना रोग प्रसार के प्रबंधन के लिए एक महत्वपूर्ण कदम है। परंपरागत रूप से, डोमेन विशेषज्ञ/प्लांट पैथोलॉजिस्ट फसलों में रोग के लक्षणों को मैन्युअल रूप से देखकर उसका निदान करते थे। हालाँकि, समय, लागत, भौतिक पहुँच और संसाधन उपलब्धता जैसी बाधाओं के कारण यह तरीका काफी हद तक व्यवहार्य नहीं है। इसलिए, फसलों में रोगों के निदान के लिए एक सटीक, त्वरित और लागत प्रभावी तरीका वैज्ञानिक समुदाय के लिए एक चुनौती है। इस संबंध में, मक्का की फसल में एक स्वचालित रोग पहचान दृष्टिकोण खाद्य सुरक्षा और टिकाऊ कृषि के लक्ष्यों तक पहुँचने के लिए एक व्यावहारिक और व्यवहार्य समाधान होगा।

इस संबंध में, AI का कंप्यूटर विजन घटक कृषि के क्षेत्र में एक महत्वपूर्ण योगदानकर्ता के रूप में उभरा है, जो एक ऐसा वातावरण प्रदान करता है जो रोगों और उनकी गंभीरता के चरणों के निदान की प्रक्रिया को स्वचालित कर सकता है (कामिलारिस और प्रीनाफेटा-बोल्डू, 2018)। डीप लर्निंग जैसी आधुनिक AI-आधारित तकनीकों में रोग के लक्षणों की अलग-अलग विशेषताओं को स्वचालित रूप से पहचानने और केवल उपलब्ध छवियों का उपयोग करके उनका निदान करने की क्षमता है (हक एट अल., 2022)। पिछले कुछ वर्षों में, चावल, गेहूं, टमाटर , मक्का आदि प्रमुख फसलों की बीमारियों की पहचान करने के लिए डीप लर्निंग तकनीकों का उपयोग किया जा रहा है ।

हाल ही में, आईसीटी और इंटरनेट में प्रगति ने स्मार्टफोन के उपयोग से कहीं भी, किसी से भी संपर्क करना संभव बना दिया है। स्मार्टफोन तकनीक अपने किफायती मूल्य-टैग, उपयोगकर्ता के अनुकूल इंटरफेस, उपयोगिता

अनुप्रयोगों, इंटरनेट तक तेज पहुँच आदि के कारण एक उपयोगी उपकरण के रूप में उभरी है। इसलिए, एप्लिकेशन इंटरफेस के माध्यम से स्मार्टफोन के साथ डीप लर्निंग मॉडल को एकीकृत करने से कृषि समुदाय तक पहुँचना संभव हो जाएगा। इसलिए, यह शोध कार्य रोगों के स्वचालित निदान के लिए एक दृष्टिकोण विकसित करने के साथ-साथ एक मोबाइल एप्लिकेशन के माध्यम से इस दृष्टिकोण के कार्यान्वयन के लिए किया गया है। इसलिए, हमने फसल की बीमारियों और कीटों की पहचान के लिए डीप लर्निंग-आधारित मॉडल विकसित करने पर काम किया। हमने कई फसलों की बीमारियों और कीटों की छवियों की पहचान करने के लिए डीप कन्वोल्यूशनल न्यूरल नेटवर्क मॉडल विकसित किए। इसके बाद, हमने बीमारी और कीट पहचान प्रक्रिया को स्वचालित करने के लिए एक AI-सक्षम मोबाइल एप्लिकेशन विकसित किया।

AI-DISC मोबाइल एप्लिकेशन

AI-DISC का मतलब फसलों के लिए कृत्रिम बुद्धिमत्ता आधारित रोग पहचान प्रणाली है। यह एक AI-सक्षम एंड्रॉइड मोबाइल एप्लिकेशन है, जिसे खेत की स्थिति में फसल में रोगों और कीटों की स्वचालित पहचान के लिए विकसित किया गया है। AI-DISC ऐप को 11 राज्य कृषि विश्वविद्यालयों और ICAR-IIMR, लुधियाना के सहयोग से ICAR-IASRI, नई दिल्ली के कंप्यूटर अनुप्रयोग प्रभाग द्वारा विकसित किया गया है। एप्लिकेशन उन्नत AI-आधारित एल्गोरिदम विकसित करने के लिए अपने ज्ञान और छवि डेटाबेस भंडार के रूप में नेशनल इमेज बेस फॉर प्लांट प्रोटेक्शन (NIBPP) का उपयोग करता है। AI-DISC मोबाइल ऐप AI-आधारित तकनीकों का उपयोग करके छवि-आधारित स्वचालित पौधों की बीमारी और कीट पहचान की सुविधा देता है। इसका उद्देश्य किसानों को उनकी उंगलियों पर वास्तविक समय की फसल सुरक्षा समाधान और विशेषज्ञ स्तर की सलाह प्रदान करना है। AI-DISC मोबाइल एप्लिकेशन का मानक वर्कफ्लो चित्र 1 में दिया गया है।यह ऐप गूगल प्ले स्टोर पर निःशुल्क उपलब्ध है (Link: https://play.google.com/store/apps/details?id=com.ai.ai_disc)

छवि भंडार - NIBPP

आर्टिफिशियल इंटेलिजेंस आधारित मॉडल का विकास मुख्य रूप से डेटा आधारित प्रक्रिया है। इसलिए, मॉडल को प्रशिक्षित करने के लिए बड़ी मात्रा में डेटासेट की आवश्यकता होती है। AI-आधारित मॉडलिंग के लिए छवि डेटासेट NAHEP परियोजना के 11 भागीदार संस्थानों में

चित्र 1: AI-DISC मोबाइल एप्लिकेशन का समग्र वर्कफ्ल

स्थित कई कृषि और किसान क्षेत्रों से एकत्र किए गए थे। डिजिटल छवियों को विभिन्न प्रकार के उपकरणों जैसे कम से उच्च कैमरा रिजॉल्यूशन वाले स्मार्टफोन, टैबलेट, डिजिटल कैमरा आदि का उपयोग करके खड़ी फसल से कैप्चर किया गया था। छवि कैप्चरिंग के दौरान फसलों/पौधों के हिस्सों को नष्ट न करने का कड़ाई से पालन किया गया। भारत में उगाई जाने वाली कई फसलों के रोगों और कीटों की छवियों के लिए एक व्यापक और अच्छी तरह से संरचित भंडार, नेशनल इमेज बेस फॉर प्लांट प्रोटेक्शन (NIBPP) बनाया गया था। छवियों के आसान संग्रह, संरक्षण और प्रसंस्करण की सुविधा के लिए, NIBPP मोबाइल ऐप का उपयोग किया गया था। वर्तमान में, NIBPP डेटाबेस में लगभग 66 कृषि संबंधी महत्वपूर्ण फसलों के रोगों और कीटों की लगभग चार लाख से अधिक छवियाँ हैं। इस लेख में, हमने चावल, गेहूँ, मक्का और टमाटर जैसी चार फसलों के AI-आधारित रोग पहचान मॉडलिंग को एक उपयोग के मामले के रूप में प्रस्तुत किया है।

चावल डेटासेट: NIBPP रिपोजिटरी के चावल डेटासेट में आठ प्रमुख बीमारियों जैसे बैक्टीरियल लीफ ब्लाइट, ब्राउन स्पॉट, फॉल्स स्मट, ब्लास्ट, शीथ ब्लाइट, नैरो ब्राउन लीफ स्पॉट, टंग्रो और शीथ रॉट के साथ-साथ स्वस्थ पौधों की तस्वीरें शामिल हैं। तस्वीरों का सारांश चित्र 2(A) में प्रस्तुत किया गया है।

गेहूं डेटासेट: एनआईबीपीपी रिपोजिटरी के गेहूं डेटासेट में तीन प्रमुख रस्ट रोगों जैसे स्टेम रस्ट, येलो रस्ट और ब्राउन रस्ट के साथ-साथ गेहूं की फसल के स्वस्थ रोगों की तस्वीरें शामिल हैं। चित्रों का सारांश चित्र 2(बB) में प्रस्तुत किया गया है।

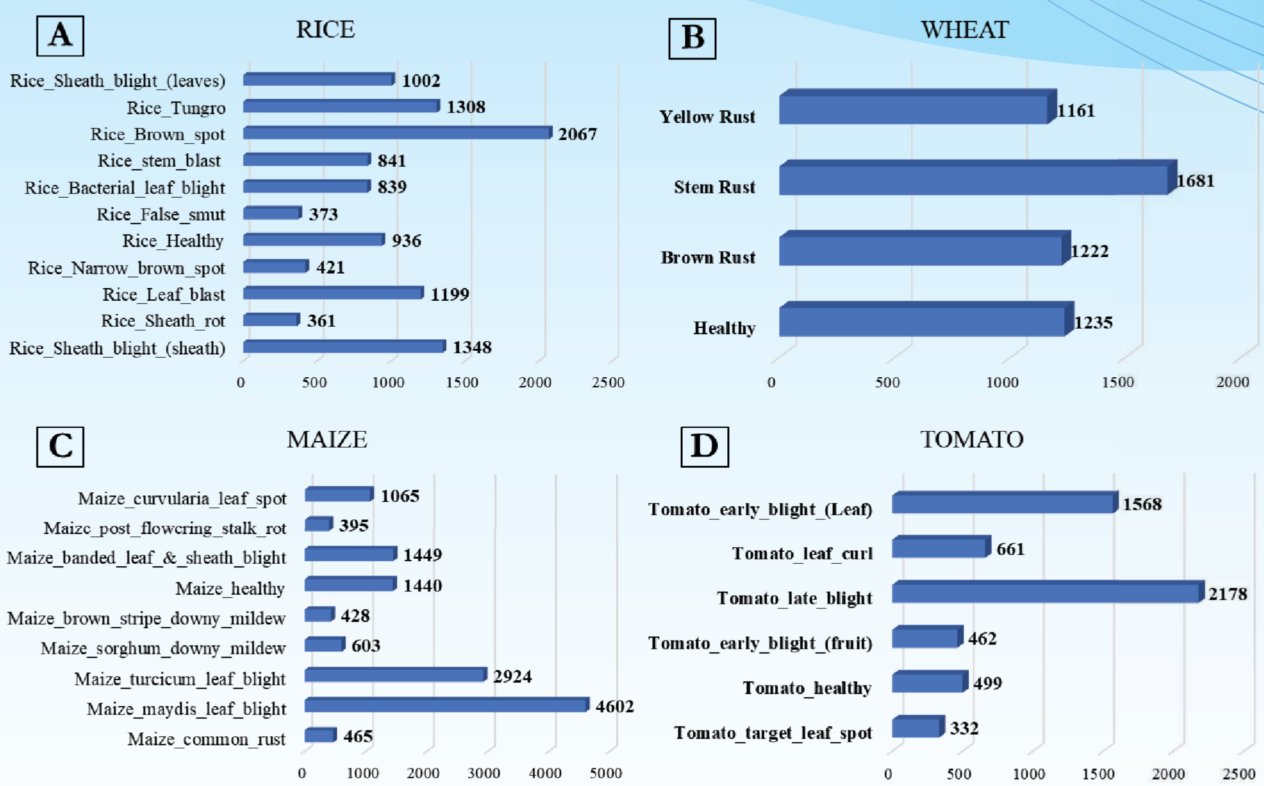
मक्का डेटासेट: NIBPP रिपोजिटरी के मक्का डेटासेट में आठ प्रमुख बीमारियों जैसे कि मेयडिस लीफ ब्लाइट, टर्किकम लीफ ब्लाइट, कॉमन रस्ट, ब्राउन स्पॉट, डाउनी मिल्ड्यू, कर्वुलरिया लीफ स्पॉट, बैंडेड लीफ, पोस्ट फ्लावरिंग स्टाक रॉट और शीथ ब्लाइट और सोरघम डाउनी मिल्ड्यू के साथ-साथ मक्का फसलों की स्वस्थ पत्तियों की तस्वीरें शामिल हैं। तस्वीरों का सारांश चित्र 2(सC) में प्रस्तुत किया गया है।

टमाटर डेटासेट: NIBPP रिपोजिटरी के टमाटर डेटासेट में चार प्रमुख रोगों जैसे लेट ब्लाइट, अर्ली ब्लाइट, लीफ कर्ल और टारगेट लीफ स्पॉट रोगों के साथ-साथ टमाटर की फसल के स्वस्थ पौधों की तस्वीरें शामिल हैं। चित्रों का सारांश चित्र 2(क) में प्रस्तुत किया गया है।

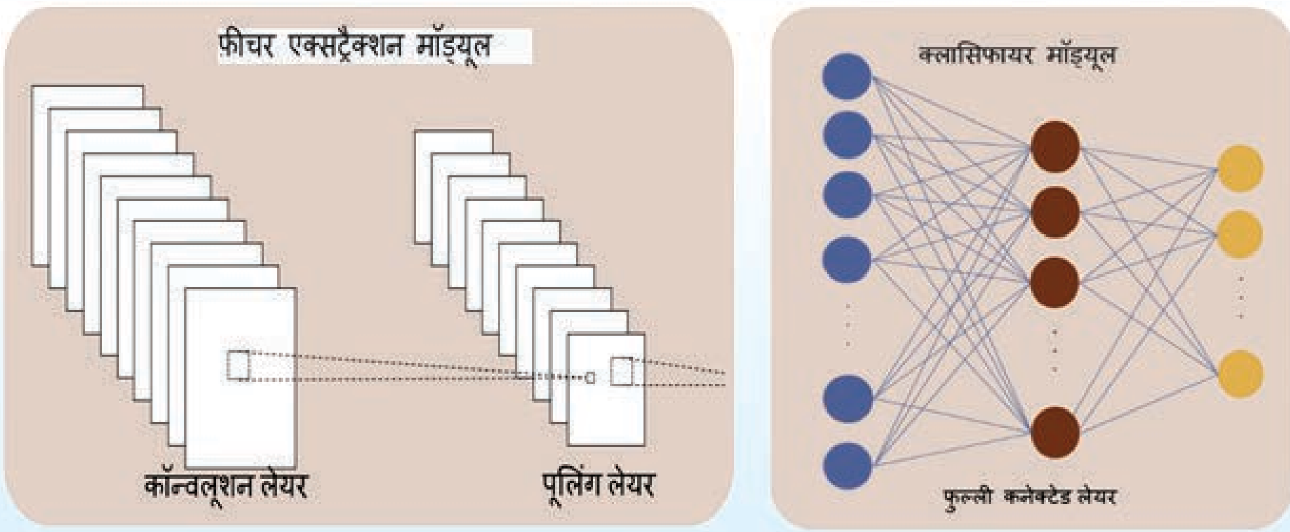
एआई-आधारित मॉडल विकास

पृष्ठभूमि: कंप्यूटर विजन कार्यों के मामले में, कन्वोल्यूशनल न्यूरल नेटवर्क (CNN) डीप लर्निंग तकनीकों की प्राथमिक पसंद हैं। हाल के वर्षों में, कन्वोल्यूशनल न्यूरल नेटवर्क मॉडल ने फसलों में रोग निदान में उल्लेखनीय उपलब्धियां दिखाई हैं। कन्वोल्यूशनल न्यूरल नेटवर्क फीड फॉरवर्ड ANN आधारित मॉडल हैं। CNN का उपयोग करने का लाभ यह है कि वे छवियों की अंतर्निहित विशेषताओं को स्वचालित रूप से कैप्चर कर सकते हैं। CNN मुख्य रूप से दो मॉड्यूल से बना होता है: एक फीचर एक्सट्रैक्शन मॉड्यूल और दूसरा क्लासिफायर मॉड्यूल जैसा कि चित्र 3 में दिखाया गया है। फीचर एक्सट्रैक्शन मॉड्यूल कन्वोल्यूशन लेयर और पूलिंग लेयर से बना होता है, जो अंतर्निहित विशेषताओं के स्वचालित रूप से निष्कर्षण के लिए जिम्मेदार होते हैं और बड़ी संख्या में छवियों से एक अच्छा प्रतिनिधि सक्रियण मानचित्र तैयार करते हैं। अगला, वर्गीकरण मॉड्यूल कृत्रिम न्यूरल नेटवर्क की वास्तुकला जैसा दिखता है जहां प्रत्येक इनपुट नोड नेटवर्क में हर दूसरे नोड के लिए क्लासिफायर होता है।

चावल रोग पहचान मॉडल: चावल रोगों की पहचान के लिए, हमने अत्याधुनिक इनसेप्शनV3 नेटवर्क आर्किटेक्चर को लागू किया और अपने चावल रोग डेटासेट के साथ स्क्रैच से प्रशिक्षित किया। इनसेप्शनV3 नेटवर्क 42-लेयर गहरा है जिसमें संयोजित कन्वोल्यूशन और पूलिंग लेयर हैं और कम्प्यूटेशनल लागत के मामले में GoogleNet से 2.5 गुना महंगा है। इस इनसेप्शनV3 नेटवर्क की मुख्य अवधारणा 'इनसेप्शन' मॉड्यूल है जो नेटवर्क को रोगग्रस्त छवियों से बारीक विशेषताओं को निकालने में सक्षम बनाता है।



चित्र 2: चार फसलों की बीमारी के अनुसार छवियों का सारांश A) चावल डेटासेट B) गेहूं डेटासेट C) मक्का डेटासेट और D) टमाटर डेटा से



चित्र 3: कन्वोल्यूशनल न्यूरल नेटवर्क (सीएनएन) का ग्राफिकल प्रतिनिधित्व

विकसित मॉडल को 64 के बैच आकार के साथ 100 युगों के लिए प्रशिक्षित किया गया था जबकि छवि का आकार 256 × 256 था। हाइपरपैरामीटर के बीच, हमने प्रशिक्षण समय के दौरान एडम ऑप्टिमाइजर और श्रेणीबद्ध क्रॉस एन्ट्रॉपी लॉस फंक्शन का उपयोग किया। प्रस्तावित मॉडल ने परीक्षण डेटासेट पर 98% सटीकता दिखाई।

गेहूं रोग पहचान मॉडलः गेहूं रोग पहचान मॉडल को अत्याधुनिक एफिशिएंटनेट नेटवर्क के आधार पर विकसित किया गया था। हमने एफिशिएंटनेट मॉडल के प्रीट्रेन्ड श्इमेजनेटश वेट का उपयोग करके ट्रांसफर लर्निंग दृष्टिकोण लागू किया। प्रशिक्षण समय के दौरान, मॉडल को श्इमेजनेटश वेट के साथ आरंभ किया गया और अंतिम कुछ परतों को हमारे गेहूं रोग डेटासेट पर प्रशिक्षित किया गया। हाइपरपैरामीटर के बीच, हमने प्रशिक्षण समय के दौरान एडम ऑप्टिमाइजर और श्रेणीबद्ध क्रॉस एन्ट्रॉपी लॉस फंक्शन का उपयोग किया। विकसित मॉडल ने परीक्षण डेटासेट पर लगभग 99% सटीकता हासिल की।

मक्का रोग पहचान मॉडलः मक्का रोगों की पहचान के लिए, हमने अत्याधुनिक इनसेप्शनV3 नेटवर्क आर्किटेक्चर को लागू किया और हमारे मक्का रोग डेटासेट का उपयोग करके मॉडल में मामूली वृद्धि के साथ स्क्रैच से प्रशिक्षित किया। हमने इनसेप्शनV3 नेटवर्क के शीर्ष पर एक वैश्विक औसत पूलिंग परत और साथ ही दो पूरी तरह से जुड़ी हुई परत को जोड़ा। विकसित मॉडल को 128 के बैच आकार के साथ 200 युगों के लिए प्रशिक्षित किया गया था जबकि छवि का आकार 384 × 384 था। हाइपरपैरामीटर के बीच, हमने प्रशिक्षण समय के दौरान एडम ऑप्टिमाइजर और श्रेणीबद्ध क्रॉस एन्ट्रॉपी लॉस फंक्शन का उपयोग किया। प्रस्तावित मॉडल ने परीक्षण डेटासेट पर लगभग 96% के f1-स्कोर के साथ लगभग 97% सटीकता हासिल की।

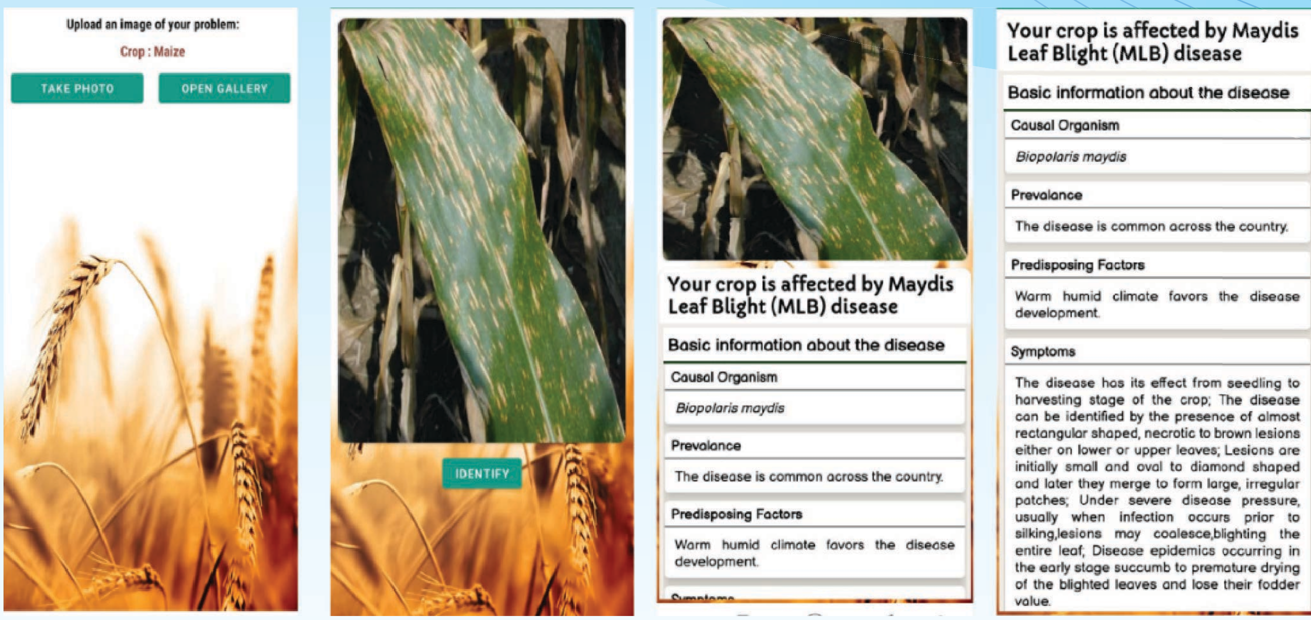
टमाटर रोग पहचान मॉडलः हमने पहले से प्रशिक्षित अत्याधुनिक मोबाइलनेट नेटवर्क आर्किटेक्चर के आधार पर टमाटर रोग पहचान मॉडल विकसित किया है। हमने अपने टमाटर रोग डेटासेट का उपयोग करके प्रस्तावित मॉडल को प्रशिक्षित करते समय स्थानांतरण सीखने के दृष्टिकोण को अपनाया। मोबाइलनेट नेटवर्क की अंतिम परत को वैश्विक औसत पूलिंग (GAP) और सॉफ्टमैक्स परत से बदल दिया गया था । विकसित मॉडल को 128 के बैच आकार के साथ 150 युगों के लिए प्रशिक्षित किया गया था जबकि छवि का आकार 224 × 224 था। प्रशिक्षण समय के दौरान एडम ऑप्टिमाइजर और श्रेणीबद्ध क्रॉस एन्ट्रॉपी लॉस फंक्शन का उपयोग किया गया था। मॉडल ने परीक्षण डेटासेट पर लगभग 98% सटीकता हासिल की।

AI-DISC में रोग और कीट पहचान मॉड्यूल (DPIM)

चार फसलों के विकसित मॉडलों को कृषिमेघ क्लाउड इंफ्रास्ट्रक्चर में कॉन्फिगर किए गए एआई इंफ्रेंस सर्वर में सहेजा और तैनात किया गया और आरईएसटी एपीआई का उपयोग करके एआई-डीआईएससी मोबाइल ऐप के साथ एकीकृत किया गया। फसलों की बीमारियों की स्वचालित पहचान की सुविधा के लिए एआई-डीआईएससी ऐप में एक रोग और कीट पहचान मॉड्यूल (डीपीआईएम) बनाया गया है। वर्तमान में, ऐप को चावल, गेहूं, मक्का और टमाटर की फसलों के रोगों और कीटों के लिए चार एआई-आधारित मॉडल के साथ एकीकृत किया गया है। उपयोगकर्ता ऐप के इस मॉड्यूल के माध्यम से खेतों से संक्रमित फसलों की छवि को कैप्चर और अपलोड करके कुछ सेकंड के भीतर रोगों और कीटों की पहचान कर सकते हैं। डीपीआईएम का वर्कफ्लो चित्र 4 में दिखाया गया है। समस्या की पहचान करने के साथ

निष्कर्ष

डिजिटल छवियों का उपयोग करके फसलों की स्वचालित बीमारी और कीट पहचान के लिए AI-क्ष्ैब् एप्लिकेशन विकसित किया गया है। कई कृषि संबंधी महत्वपूर्ण फसलों (जैसे चावल, गेहूं, मक्का और टमाटर) की बीमारियों के लिए डीप लर्निंग-आधारित मॉडल विकसित किए गए हैं। AI-आधारित मॉडल बनाने के लिए छप्ठच्च् इमेज डेटाबेस रिपॉजिटरी से इमेज डेटासेट लिए गए थे। विकसित AI-आधारित मॉडल ने परीक्षण डेटासेट की रोगग्रस्त छवियों की पहचान करने के लिए पहचान सटीकता के संदर्भ में महत्वपूर्ण प्रदर्शन की सूचना दी। AI-आधारित मॉडल को AI-इंफरेंस सर्वर में तैनात किया गया और REST API का उपयोग करके AI-क्ष्ैब् मोबाइल एप्लिकेशन के साथ एकीकृत किया गया। मोबाइल एप्लिकेशन क्षेत्र की स्थिति में रोग/कीट संक्रमित छवियों की स्वचालित पहचान के लिए इंटरफेस प्रदान करता है, जो उपयोगकर्ताओं को फसल-सुरक्षा से संबंधित व्यापक सलाह देता है।



चित्र 4: AI-DISC में रोग और कीट पहचान मॉड्यूल (DPIM)

संदर्भ

चंद, आर., और सिंह, जे. (2022)। कार्यबल में परिवर्तन और रोजगार पीएलएफएस डेटा श्रृंखला से कुछ निष्कर्ष। *मैनपावर जर्नल*, 56।

रॉय, डी. (2022)। भारत में कृषि विपणन: सुधार और किसानों की आय दोगुनी करने के दृष्टिकोण। *जर्नल ऑफ मार्केटिंग डेवलपमेंट एंड कॉम्पिटिटिवनेस*, 16(3)।

सिंह, डी., जैन, एन., जैन, पी., कयाल, पी., कुमावत, एस., और बत्रा, एन. (2020)। प्लांटडॉक : दृश्य पौधों की बीमारी का पता लगाने के लिए एक डेटासेट। 7वीं ACM IKDD CoDS और 25वीं COMAD की कार्यवाही में (पृष्ठ 249-253)।

कामिलारिस, ए., और प्रीनाफेटा-बोल्डू, एफएक्स (2018)। कृषि में गहन शिक्षा: एक सर्वेक्षण। *कंप्यूटर्स एंड इलेक्ट्रोनिक्स इन एग्रीकल्चर*, 147, 70-90।

हक, एम.ए., मरवाह, एस., देब, सी.के., निगम, एस., अरोड़ा, ए., हुड्डा, के.एस., ... और अग्रवाल, आर.सी. (2022)। मक्का की फसल के रोगों की पहचान के लिए गहन शिक्षण-आधारित दृष्टिकोण। साइंटिफिक रिपोर्ट, 12(1), 6334।

खरपतवार वर्गीकरण के लिए एक ध्यान तंत्र संलग्न डेंसनेट मॉडल

अक्षय धीरज, सपना निगम, मधु, मो. अशरफुल हक, सुदीप मरवाह
भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

खरपतवार प्रबंधन एक महत्वपूर्ण कारक है जो फसल की उपज और समग्र फसल उत्पादकता पर महत्वपूर्ण प्रभाव डालता है। खरपतवारों की सही पहचान करना किसानों के लिए पिछले कुछ दशकों से एक बड़ी समस्या रही है क्योंकि खरपतवारों का स्वरूप फसलों के समान होता है। यह आलेख मक्के की फसल के लिए कन्वोल्यूशनल तंत्रिका नेटवर्क (सीएनएन) और ध्यान तंत्र का उपयोग करते हुए एक विजन-आधारित खरपतवार वर्गीकरण प्रणाली प्रस्तुत करता है। डेंसेनेट121 मॉडल और कन्वोल्यूशनल ब्लॉक ध्यान तंत्र का उपयोग करके एक मॉडल प्रस्तुत किया गया है। प्रस्तावित मॉडल मक्के की फसल में खरपतवार की पहचान के लिए 99.58% सटीकता के साथ अच्छा प्रदर्शन करता है। प्रस्तावित कार्य को मक्के की फसल के 6000 छवियों और मक्के की चार खरपतवार प्रजातियों के छवि डेटासेट के साथ लागू किया गया। इसके अतिरिक्त, इस कार्य में कुछ प्रसिद्ध सीएनएन मॉडलों जैसे इफीशियंटनेटबी0, डेंसेनेट121, मोबाइलनेटवी2, इंसेप्शनरेसनेटवी2 और रेसनेट50वी2 को भी लागू किया गया और उनके प्रदर्शन की तुलना प्रस्तावित कार्य से की गई। सभी परखे गए सीएनएन मॉडलों में, हमारे प्रस्तावित मॉडल ने सभी पांच सीएनएन मॉडलों को पीछे छोड़ दिया। प्रस्तुत मॉडल को मक्के की फसल में खरपतवार वर्गीकरण के लिए मोबाइल उपकरणों में एकीकृत किया जा सकता है।

मुख्य शब्द: ध्यान तंत्र, खरपतवार वर्गीकरण, डीप लर्निंग, कन्वोल्यूशनल ब्लॉक ध्यान तंत्र, स्मार्ट कृषि

परिचय

मक्का दुनिया भर में सबसे व्यापक रूप से उगाए और खपत किए जाने वाले वस्तुओं में से एक है। इसकी व्यापक लोकप्रियता इसके असाधारण उत्पादकता और औद्योगिक एवं कृषि क्षेत्रों में इसके बहुप्रकारी अनुप्रयोगों से उत्पन्न होती है, जिसमें पशु आहार, बायोफ्यूल उत्पादन, और खाद्य मिठास उत्पादों का उत्पादन शामिल हैं। मक्का उत्पादकता को बढ़ाने में एक मुख्य बाधा खरपतवारों की प्रभावी नियंत्रण और प्रबंधन है। खरपतवार फसल उत्पादन पर सबसे नुकसानदायक प्रतिबंधों में से एक हैं और कृषि उत्पादकता के लिए एक महत्वपूर्ण खतरा उत्पन्न करते हैं क्योंकि वे फसलों से पानी, पोषक तत्वों और सूर्य के प्रकाश के लिए प्रतिस्पर्धा करते हैं। खरपतवार हानिकारक बैक्टीरिया, वायरस और अन्य सूक्ष्मजीवों को भी प्रस्तुत करते हैं, जिसके कारण उपज में भारी हानि होती है। इस प्रकार, खरपतवार नियंत्रण प्रभावी रूप से फसल और खाद्य उत्पादन को बढ़ाने के लिए आवश्यक है, ताकि बढ़ती हुई विश्व जनसंख्या की आवश्यकताएँ पूरी की जा सकें। वर्तमान में फसल के खेतों से खरपतवारों को नष्ट करने के लिए सबसे व्यापक रूप से उपयोग किया जाने वाला तरीका रासायनिक तरीका है, इसकी प्रभावशीलता और व्यवहारिकता के कारण। हालांकि, पारंपरिक खरपतवार नियंत्रण कार्यक्रमों के पर्यावरणीय प्रभाव नकारात्मक हो सकते हैं। यह तरीका गंभीर पर्यावरणीय प्रदूषण का कारण बन रहा है, क्योंकि इसे आमतौर पर खेतों में समान रूप से छिड़का जाता है। आधुनिक तकनीकों जैसे की मशीन लर्निंग

और कंप्यूटर विजन का उपयोग करके लक्षित हर्बिसाइडल आवेदन के लिए स्वचालित खरपतवार नियंत्रण अनुसंधान समुदाय में रुचि प्राप्त कर रहा है। स्वचालित खरपतवार नियंत्रण एक संभावित समाधान के रूप में उभरा है, जिसने खरपतवार लक्ष्यों के अधिक प्रभावी चयनात्मक आवेदन के माध्यम से कृषि उत्पादकता को बढ़ाया जबकि हर्बिसाइड्स को कम किया। कृषि में, डीप लर्निंग मॉडल प्रचलित था, जो छवि विश्लेषण के गहरे अध्ययन पर आधारित है, और खरपतवारों के प्रसार को नियंत्रित करने के लिए एक अपेक्षाकृत तेज, गैर-आक्रामक, और गैर-विनाशकारी तरीका प्रदान करता है। डीप लर्निंग का उपयोग पौधों के रोगों का पता लगाने और खरपतवार की पहचान करने के लिए किया गया है और यह कई विशिष्ट मामलों में सफल रहा है।

इसलिए, हमने मक्के की फसल में खरपतवार वर्गीकरण के लिए डेंसेनेट121 मॉडल को सीबीएअम मॉड्यूल के साथ उपयोग करके एक डीप लर्निंग मॉडल विकसित किया है। हमारा प्रस्तावित मॉडल खरपतवार वर्गीकरण के लिए 99.58% की कुल सटीकता स्कोर प्राप्त करने में सफल रहा है।

सामग्री और विधि

डेटासेट

इस अनुसंधान कार्य में मक्का खरपतवार डेटासेट का उपयोग किया गया है, जो सार्वजनिक रूप से उपलब्ध है। इस डेटासेट में खरपतवार और मक्का की पाँच श्रेणियाँ हैं और कुल 6000 छवियाँ हैं। डेटासेट में मक्का फसल से संबंधित चार खरपतवार प्रजातियाँ हैं, जिनके नाम हैं ब्लूग्रास (bluegrass), चेनोपोडियम एल्बम (chenopodium album), सर्सियम सेटोसम (cirsium setosum) और सेज (sedge)। डेटासेट को विभिन्न श्रेणियों में वितरित किया गया है, जैसा कि तालिका 1 में दिखाया गया है।

डेटा पूर्व-प्रसंस्करण

सीएनएन मॉडल में इनपुट के रूप में फीड करने से पहले सभी छवियों का आकार 800 × 600 के मूल छवि आकार से 224 × 224 कर दिया गया था। यहां प्रयोग किए गए सभी मॉडलों के लिए 224 × 224 आयाम की छवि का उपयोग किया गया है। पांच वर्गों के बीच डेटासेट का वितरण तालिका 1 में दिखाया गया है।

प्रस्तावित मॉडल की वास्तुकला

प्रस्तावित ध्यान-संलग्न डेंसेनेट121 मॉडल में डेंसेनेट121 मॉडल को सीबीएअम मॉड्यूल और कुछ जोड़े गए लेयर्स के साथ एकीकृत किया गया है। मक्का/खरपतवार की छवि को मॉडल में इनपुट के रूप में दिया जाता है। डेंसेनेट मॉडल फीचर मैप उत्पन्न करता है, जिसे फिर सीबीएअम मॉड्यूल को भेजा जाता है। सीबीएअम मॉड्यूल फिर डेंसेनेट121 मॉडल के फीचर मैप पर लागू किया जाता है, जिसके बाद कुछ लेयर्स का उपयोग किया जाता है ताकि मॉडल की जटिलता और ओवरफिटिंग को कम किया जा सके। सीबीएअम मॉड्यूल के बाद, वैश्विक औसत पूलिंग (global average pooling) लेयर का उपयोग किया जाता है, जिसके बाद 512 यूनिट्स का डेंस लेयर आता है। ओवरफिटिंग को कम करने के लिए 0.5 मान के साथ ड्रॉपआउट (dropout) का उपयोग किया गया है। सेलयू

तालिका 1. उपयोग किए गए डेटासेट का वितरण

क्रम संख्या	खरपतवार/फसल का नाम	कुल छवियाँ	कुल प्रशिक्षण छवियाँ	कुल परीक्षण छवियाँ	कुल सत्यापन छवियाँ
1.	ब्लूग्रास	1200	840	240	120
2.	चेनोपोडियम एल्बम	1200	840	240	120
3.	सर्सियम सेटोसम	1200	840	240	120
4.	मक्का	1200	840	240	120
5.	सेज	1200	840	240	120

(SeLU) को सक्रियण फंक्शन (activation function) के रूप में मॉडल में उपयोग किया गया है, जो सामान्यत: छवि वर्गीकरण कार्यों के लिए अच्छा प्रदर्शन करता है। अंतिम लेयर में मक्का और खरपतवार प्रजातियों को वर्गीकृत करने के लिए पाँच यूनिट्स के साथ डेंस लेयर और सॉफ्टमैक्स (softmax) क्लासीफायर का उपयोग किया गया है। ये अतिरिक्त लेयर्स ओवरफिटिंग को कम करती हैं और मॉडल के समग्र प्रदर्शन को बढ़ाती हैं। मॉडल के आर्किटेक्चर को चित्र 1 में दिखाया गया है।

हाइपरपैरामीटर्स फाइन-ट्यूनिंग

इस अनुसंधान अध्ययन में, ट्रांसफर लर्निंग दृष्टिकोण का उपयोग किया गया है। यहाँ उपयोग किए गए सभी सीएनएन आर्किटेक्चर के ऊपर, अतिरिक्त लेयर्स जोड़े गए हैं। अधिकतम ईपोक्स (epochs) की संख्या 30 निर्धारित की गई है और बैच साइज (batch size) 16 है। एडम (Adam) और श्रेणीबद्ध क्रॉस-एंट्रॉपी (categorical cross entropy) को ऑप्टिमाइजर और लॉस फंक्शन के रूप में उपयोग किया गया है। लर्निंग रेट का मान 0.0001 निर्धारित किया गया है और ड्रॉपआउट दर 0.5 निर्धारित की गई है ताकि ओवरफिटिंग को कम किया जा सके। सभी मॉडलों को समान समूह के हाइपरपैरामीटर्स के साथ प्रशिक्षित किया गया है।

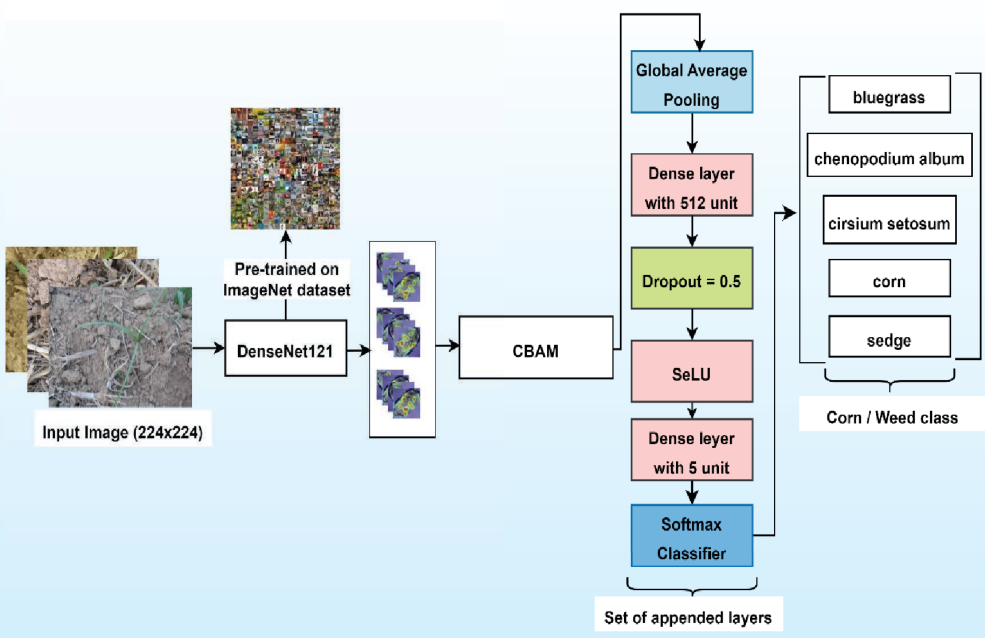
प्रयोग परिणाम

डेटासेट तैयारी

डेटासेट तैयारी में कई चरण होते हैं। पहले चरण में, डेटासेट को तीन अलग-अलग समूह में विभाजित किया गया, जैसे कि प्रशिक्षण, सत्यापन और परीक्षण डेटासेट, जिनका अनुपात क्रमशः 70%, 10% और 20% है। दूसरे चरण में, सभी छवियों का आकार 224 × 224 के आयाम में बदला गया है। तीसरे चरण में, सभी छवियों को सामान्यीकृत किया गया जिसमे प्रत्येक पिक्सल की तीव्रता मान को 255 से विभाजित किया गया है ताकि गणना की जटिलता को कम किया जा सके।

मॉडल्स के प्रदर्शन की तुलना

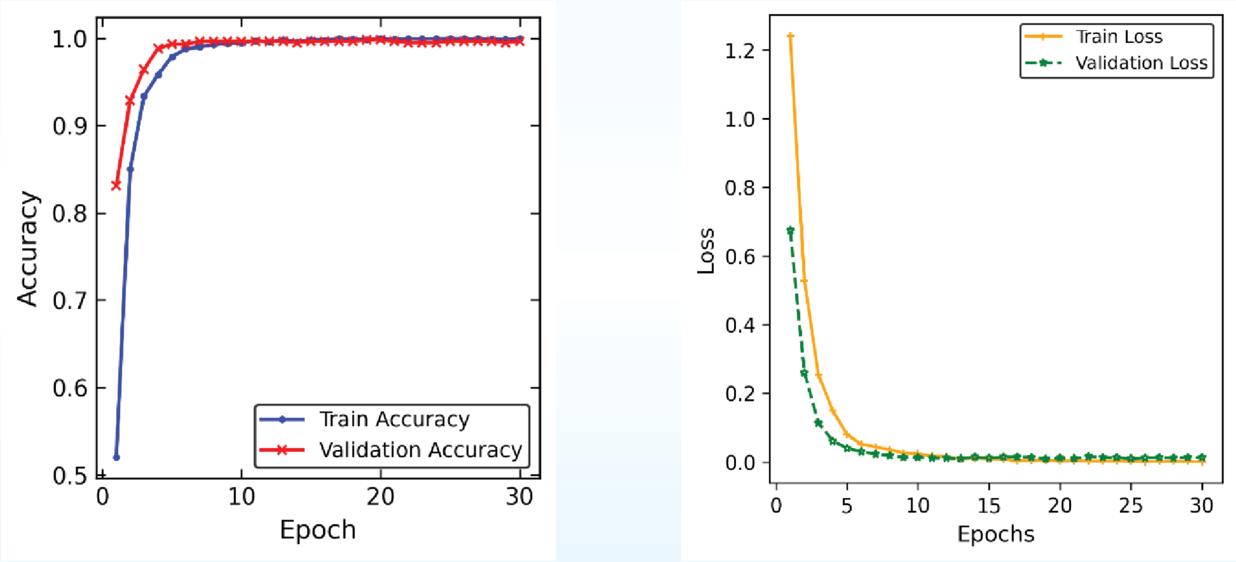
प्रस्तावित मॉडल के प्रदर्शन की तुलना पांच पूर्व प्रशिक्षित सीएनएन (CNN) मॉडल्स जैसे इफीशियंटनेटबी0, डेंसेनेट121, मोबाइलनेटवी2, इंसेप्शनरेसनेटवी2 और रेसनेट50वी2 से की गई है, और यह तालिका 2 में दर्शाया गया है। सभी परीक्षण किए गए मॉडल्स में, प्रस्तावित मॉडल



चित्र 1. प्रस्तुत मॉडल का आर्किटेक्चर

का समग्र प्रदर्शन सबसे अच्छा है, जिसमें 99.58% सटीकता प्राप्त हुई है, इसके बाद डेंसेनेट121 मॉडल है। प्रस्तावित मॉडल में डेंसेनेट121 के समान ही पैरामीटर और फ्लोप्स (FLOPs) की संख्या है। इसका मतलब है कि सीबीएअम थोड़ा पैरामीटर की संख्या को बढ़ाता है, लेकिन डेंसेनेट121 मॉडल के समग्र प्रदर्शन को बढ़ाता है। मोबाइलनेटवी2 सबसे कम पैरामीटर और फ्लोप्स का उपयोग करता है, लेकिन यहाँ सभी परीक्षण किए गए मॉडल्स में से इसका दूसरा सबसे खराब प्रदर्शन है। रेसनेट50वी2 गणना की दृष्टि से महंगा है, जिसमें 24 मिलियन पैरामीटर और 6.99 गीगाफ्लोप्स हैं, और यह सभी में सबसे खराब प्रदर्शन करने वाला मॉडल है। प्रस्तावित मॉडल का प्रिसीजन, रिकॉल और F1 स्कोर क्रमश: 99.59%, 99.58% और 99.58% हैं, जो अध्ययन के संतोषजनक परिणामों को दर्शाते हैं।

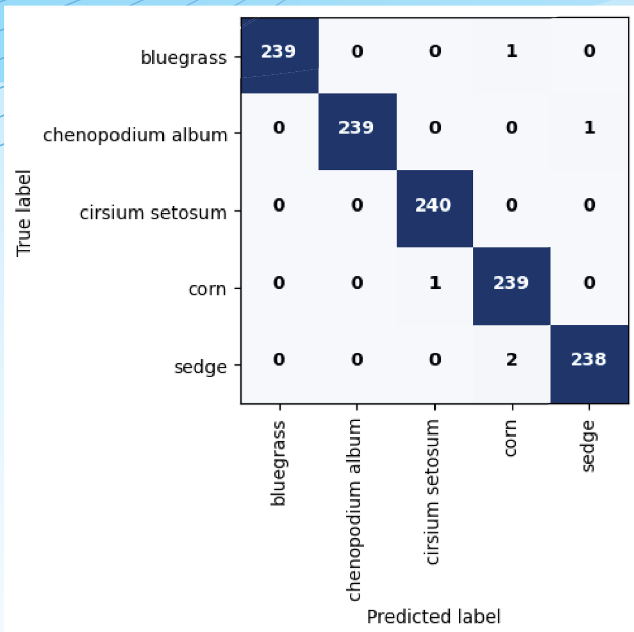
चित्र 2 में प्रस्तावित मॉडल की सटीकता और हानि कर्व को दिखाया गया है। जैसा कि चित्र से स्पष्ट है, प्रशिक्षण और सत्यापन सटीकता दोनों में समान पैटर्न दिखता है क्योंकि दोनों मानों में वृद्धि होती है और संबंधित हानि मानों में कमी होती है। यह दर्शाता है कि मॉडल को ठीक से प्रशिक्षित किया गया है और यह ओवरफिट नहीं है। चित्र 3 में प्रस्तावित मॉडल के लिए कन्फ्यूजन मैट्रिक्स को दिखाया गया है। मॉडल ने परीक्षण डेटासेट में पांच नमूनों को गलत वर्गीकृत किया है। एक छवि जो ब्लूग्रास की है, उसे मक्का के रूप में वर्गीकृत किया गया है। एक छवि जो चेनोपोडियम एल्बम की है, उसे गलत तरीके से सेज के रूप में वर्गीकृत किया गया है। मक्का की एक छवि को सर्सियम सेटोसम के रूप में गलत वर्गीकृत किया गया है और सेज की दो छवियों को गलत तरीके से मक्का के रूप में वर्गीकृत किया गया है।



चित्र 2. प्रस्तावित मॉडल का (a) सटीकता कर्व (b) हानि कर्व

तालिका 2. सभी परखे गए सीएनएन मॉडलों की तुलना

मॉडल का नाम	सटीकता (%)	प्रिसीजन (%)	रिकॉल (%)	F1 स्कोर (%)	कूल पैरामीटर्स	गीगाफ्लोप्स
डेंसेनेट121	99.50	99.51	99.50	99.50	7,564,869	5.7
इफीशियंटनेटबी0	99.42	99.42	99.42	99.42	4,708,008	0.796
इंसेप्शनरेसनेटवी2	99.08	99.09	99.08	99.08	54,528,741	13
रेसनेट50वी2	98.75	98.76	98.75	98.75	24,066,565	6.99
मोबाइलनेटवी2	99.17	99.17	99.17	99.17	2,571,589	0.613
प्रस्तावित मॉडल	99.58	99.59	99.58	99.58	7,631,655	5.7



चित्र 3. मॉडल की कन्फ्युशन मैट्रिक्स

निष्कर्ष

कम सटीकता और जटिल कृषि वातावरण में खरपतवार प्रजातियों और फसलों की पहचान में सीमित सामान्यीकरण क्षमता से संबंधित चुनौतियों का समाधान करने के लिए, इस अध्ययन में डेंसेनेट का एक उन्नत संस्करण उपयोग करते हुए एक खरपतवार पहचान विधि प्रस्तुत की गई है। डेंसेनेट121 मॉडल, सीबीएअम मॉड्यूल और कुछ अतिरिक्त परतों का संयोजन किया गया है ताकि समग्र प्रदर्शन को बढ़ाया जा सके। यह सीबीएअम मॉड्यूल छवियों से महत्वपूर्ण खरपतवार विशेषताओं को निकालने में सक्षम है। अतिरिक्त परतों का समूह पैरामीटर की संख्या को कम कर सकता है और ओवरफिटिंग को कम कर सकता है। प्रस्तावित मॉडल मौजूदा पूर्व प्रशिक्षित सीएनएन मॉडल्स को 99.58% समग्र सटीकता स्कोर और 7.6 मिलियन पैरामीटर के साथ मक्का खरपतवार डेटासेट पर बेहतर प्रदर्शन करता है। भविष्य में, प्रस्तावित मॉडल को अन्य खरपतवार/फसल डेटासेट के साथ तैनात किया जा सकता है ताकि मॉडल की दृढ़ता का मूल्यांकन किया जा सके। इस मॉडल को मोबाइल उपकरणों या अन्य किसी भी स्मार्ट खरपतवार नियंत्रण उपकरण पर तैनात किया जा सकता है, जिससे लक्षित खरपतवार नियंत्रण के लिए सटीक खेत की पहचान और समग्र मक्का फसल उत्पादन दक्षता को बढ़ाया जा सके।

संदर्भ

कामिलारिस, ए. और प्रेनेफेटा-बोल्डू, एफ.एक्स., 2018। कृषि में डीप लर्निंग: एक सर्वेक्षण। कंप्यूटर और इलेक्ट्रॉनिक्स इन एग्रीकल्चर, 147, पृ.70-90।

मणिकंदकुमार, एम. और कार्तिकेयन, पी., 2023। कण झुंड अनुकूलन और डीप लर्निंग मॉडलों का उपयोग करके खरपतवार वर्गीकरण। कंप्यूट. सिस्टम. साइ. इंज., 44(1), पृ.913-927।

हसन, ए.एम., डाइपेवीन, डी., लागा, एच., जोन्स, एम. जी. और सोहेल, एफ., 2023। फसल और खरपतवार की पहचान के लिए छवि पैच-आधारित डीप लर्निंग दृष्टिकोण। इकोलॉजिकल इंफॉर्मेटिक्स, 78, पृ.102361।

रेहमान, एम.यू., ईसार, एच., अब्बास, जेड., सेनविरत्न, एल., हुसैन, आई. और चोंग, के.टी., 2024। फीचर-संपन्न डीप लर्निंग दृष्टिकोण का उपयोग करके उन्नत ड्रोन-आधारित खरपतवार पहचान। नॉलेज-बेस्ड सिस्टम्स, 305, पृ.112655।

धीरज, ए. और चंद, एस., 2023। प्रारंभिक चरण में फसल और खरपतवार वर्गीकरण के लिए डीप लर्निंग मॉडलों का उपयोग। सेंटिमेंट एनालिसिस एंड डीप लर्निंग: आईसीएसएडीएल 2022 के कार्यवाही में (पृ. 931-942)। सिंगापुर: स्प्रिंगर नेचर सिंगापुर।

सकीब, एम.ए., अकीब, एम., ताहिर, एम.एन. और हफीज, वाई., 2023। फसलों में स्मार्ट खरपतवार प्रबंधन के लिए डीप लर्निंग-आधारित स्मार्ट फार्मिंग की दिशा में। फ्रंटियर्स इन प्लांट साइंस, 14, पृ.1211235।

अधिनाता, एफ.डी. और सुमिहार्तो, आर., 2024। मशीन लर्निंग और डीप लर्निंग का उपयोग करके खरपतवार और फसल वर्गीकरण पर एक व्यापक सर्वेक्षण। आर्टिफिशियल इंटेलिजेंस इन एग्रीकल्चर।

मुराद, एन.वाई., महमूद, टी., फोर्कान, ए.आर.एम., मर्शद, ए., जयरामन, पी.पी. और सिद्दीकी, एम.एस., 2023। डीप लर्निंग का उपयोग करके खरपतवार पहचान: एक व्यवस्थित साहित्य समीक्षा। सेंसर्स, 23(7), पृ.3670।

डेयरी श्रिया: डेयरी पशुओं के स्वास्थ्य और उत्पादकता में सुधार के लिए एक कृत्रिम बुद्धिमत्ता चालित चैटबॉट

संचिता नाहा[1], रूपसी तिवारी[2], चन्दन कुमार देब[1], मधु[1], सपना निगम[1] एवं सुदीप मरवाहा[1]

[1]भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

[2]भा.कृ.अनु.प.-भारतीय पशु चिकित्सा अनुसंधान संस्थान, इज्जतनगर

सार

डेयरी श्रिया, एक कृत्रिम बुद्धिमत्ता (एआई) संचालित चैटबॉट, भारतीय पशुपालन क्षेत्र में डेयरी पशुओं की देखभाल और उत्पादकता में सुधार के उद्देश्य से विकसित किया गया है। यह चैटबॉट बहुभाषी समर्थन प्रदान करता है, जिसमें हिंदी, अंग्रेजी, और अन्य भारतीय स्थानीय भाषाओं में पशुपालकों के प्रश्नों का उत्तर देने की क्षमता है। डेयरी श्रिया विभिन्न प्राकृतिक भाषा प्रसंस्करण (एनएलपी) तकनीकों, जैसे BERT, RoBERTa, DistilBERT, और XLNet का उपयोग करती है, ताकि उपयोगकर्ता के प्रश्नों को समझा जा सके और उनके लिए डेटाबेस में से सबसे प्रासंगिक उत्तर प्रदान किया जा सके। चैटबॉट की प्रशिक्षण प्रक्रिया के लिए पशु चिकित्सा वैज्ञानिकों द्वारा पशु स्वास्थ्य, पोषण, प्रजनन, और रोग प्रबंधन जैसे विषयों पर प्रश्न-उत्तर तैयार किये गए हैं। डेयरी श्रिया वेब-आधारित इंटरफेस [1] है और मोबाइल [2] के माध्यम से उपलब्ध है, जिससे इसका उपयोग सरल और व्यापक रूप से किया जा सकता है। उपयोगकर्ता की प्रतिक्रियाओं के आधार पर BERT मॉडल प्रश्नों को समझने में सबसे प्रभावशाली पाया गया है। यह चौटबॉट नवाचार पशुपालकों को वैज्ञानिक और वास्तविक समय की जानकारी प्रदान करके उनके जीवनस्तर और पशु पालन की उत्पादकता को सुधारने में योगदान देगा।

मुख्य शब्द: चैटबॉट; डेयरी पशु; एनएलपी; प्रश्न-उत्तर

परिचय

भारतीय पशुधन क्षेत्र देश की कृषि अर्थव्यवस्था में महत्वपूर्ण भूमिका निभाता है, जिसमें डेयरी फार्मिंग, ग्रामीण आजीविका और पोषण सुरक्षा की आधारशिला है। भारत दुनिया भर में दूध का सबसे बड़ा उत्पादक है, तथापि कम उत्पादकता, रोग प्रबंधन, अपर्याप्त आहार प्रथाएं और पशु चिकित्सा सेवाओं तक सीमित पहुंच जैसी चुनौतियां आज तक बनी हुई हैं। इस संदर्भ में, प्रौद्योगिकी-संचालित समाधानों को एकीकृत करने से डेयरी उद्योग में बदलाव आ सकता है। कृत्रिम बुद्धिमत्ता (एआई) से संचालित चैटबॉट किसानों को वास्तविक समय, सुलभ और लागत प्रभावी सलाहकार सेवाएं प्रदान करके इन चुनौतियों का समाधान करने की अपार संभावनाएं रखता हैं [3]। ऐसे चैटबॉट पशु स्वास्थ्य, पोषण, प्रजनन और बीमारी की रोकथाम पर अनुकूलित मार्गदर्शन प्रदान कर सकते हैं, किसानों को वैज्ञानिक अंतर्दृष्टि के साथ सशक्त बना सकता हैं जो पहले पहुंच योग्य नहीं था। एआई का लाभ उठाकर, ये सिस्टम ज्ञान के अंतर को घटा सकता हैं, पशु उत्पादकता बढ़ा सकता हैं और दूध की गुणवत्ता में सुधार कर सकता है, जिससे किसानों की आर्थिक स्थिरता और डेयरी क्षेत्र के समग्र विकास में योगदान मिलेगा।

यह नवीन दृष्टिकोण भारत में विशेष रूप से प्रासंगिक है, जहां पशुधन आबादी का एक महत्वपूर्ण हिस्सा छोटे

और सीमांत किसानों के पास है, जिनके पास अक्सर समय पर और विश्वसनीय पशु चिकित्सा विशेषज्ञता तक पहुंच का अभाव होता है। भारतीय डेयरी फार्मिंग की अनूठी जरूरतों के अनुरूप एआई-संचालित चैटबॉट डेयरी उत्पादों की गुणवत्ता और सुरक्षा में सुधार करते हुए पशुधन के सतत विकास को सुनिश्चित करते हुए इस क्षेत्र में क्रांति ला सकता है।

यह लेख डेयरी श्रिया (स्मार्ट अनुमानी प्रतिक्रिया आधारित बुद्धिमान एजेंट) का परिचय देता है, जो एक एआई-संचालित चैटबॉट है जिसे डेयरी जानवरों, विशेष रूप से गायों और भैंसों के स्वास्थ्य और उत्पादन प्रबंधन में सहायता के लिए डिजाइन किया गया है। डेयरी श्रिया हिंदी, अंग्रेजी, बंगाली, तमिल, मराठी, गुजराती, कन्नड़, मलयालम, तेलुगु और उर्दू सहित 10 अलग-अलग भारतीय भाषाओं में उपयोगकर्ताओं के साथ बातचीत कर सकता है, जिससे यह देश भर में उपयोगकर्ताओं की एक विस्तृत श्रृंखला के लिए सुलभ हो जाता है। यह अत्याधुनिक शैक्षिक चौटबॉट अपने उपयोगकर्ताओं को वास्तविक समय, प्रासंगिक जानकारी प्रदान करने के लिए पूर्व-प्रशिक्षित ट्रांसफार्मर-आधारित एआई मॉडल जैसी उन्नत प्राकृतिक भाषा प्रसंस्करण तकनीकों का उपयोग करता है।

कार्यप्रणाली

चैटबॉट के विकास के लिए दो अलग-अलग प्रारूपों में डेटा संग्रह की आवश्यकता होती है, एक में विभिन्न प्रकार के प्रश्नों को सूचीबद्ध किया गया है जिनका बॉट सामना कर सकता है, एवं दूसरे में संभावित प्रश्नों के जवाब वाले ज्ञान कोष को संकलित किया गया है। एआई-संचालित वर्चुअल एजेंट का कार्य उपयोगकर्ता के प्रश्नों को समझना और व्याख्या करना है। इसके बाद, प्रतिक्रिया पीढ़ी मॉड्यूल इनपुट क्वेरी के लिए सबसे उपयुक्त उत्तर की पहचान करता है। इस अध्ययन में, हमने उपयोगकर्ता की क्वेरी के जवाब में सबसे प्रासंगिक प्रश्न-उत्तर जोड़ी निर्धारित करने के लिए क्वेरी और प्रश्न समानता पद्धति को अपनाया है [4]।

डेटा संग्रह और विश्लेषण

इस चैटबॉट को विकसित करने के लिए, प्रशिक्षण डेटा में प्रश्न-उत्तर का एक व्यापक संग्रह शामिल है। ये प्रश्नोत्तर जोड़े विशेषज्ञ पशु चिकित्सा वैज्ञानिकों और शिक्षाविदों द्वारा विभिन्न स्रोतों द्वारा सावधानीपूर्वक तैयार किए गए हैं। गायों और भैंसों से संबंधित 80,000 से अधिक स्वास्थ्य और उत्पादन आधारित प्रश्नोत्तर जोड़े एकत्र किए गए हैं, जो अक्सर विभिन्न प्लैटफॉर्म पर पशुपालकों द्वारा पूछे जाते हैं। इन प्रश्नों का पशु चिकित्सा विशेषज्ञों और पशुधन वैज्ञानिकों ने दस्तावेजीकरण किया है और इनके सटीक उत्तर प्रदान किए हैं, जो अन्य ऑनलाइन प्लेटफार्मों पर आमतौर पर उपलब्ध नहीं होते। इसके अतिरिक्त, किसान कॉल सेंटर के पोर्टल से भी पशुधन के स्वास्थ्य से जुड़े विभिन्न पहलुओं का डेटा एकत्र किया गया है। डेटासेट को प्रत्येक प्रविष्टि के लिए संदर्भ, क्वेरी और उत्तर के ट्रिप्लेट प्रारूप में संरचित किया गया है। इन प्रश्नों में एस्ट्रस, स्त्री रोग संबंधी विकार, स्तनपान, दूध उत्पादन, कृत्रिम गर्भाधान, प्रजनन, भ्रूण स्थानांतरण, एफएमडी, गांठदार त्वचा रोग जैसे कई प्रबंधन श्रेणियां शामिल हैं। तालिका 1 में ऐसे कुछ डेटा ट्रिपलेट्स के उदाहरण प्रस्तुत किए गए हैं, जिनका उपयोग चैटबॉट के प्रशिक्षण के लिए किया गया है।

सिस्टम वर्कफ्लो

डेयरी श्रिया बॉट कई प्रमुख घटकों को एकीकृत करता है: यूजर इंटरफेस, वार्तालाप प्रबंधन मॉड्यूल, प्राकृतिक भाषा प्रसंस्करण (एनएलपी) मॉड्यूल और डेटा स्टोरेज मॉड्यूल। यह चैटबॉट दो प्लेटफार्मों के माध्यम से उपलब्ध है - एक मोबाइल एप्लिकेशन और एक वेब-आधारित एप्लिकेशन - जो उपयोगकर्ताओं के लिए उपयोग में आसानी और पहुंच सुनिश्चित करता है। उपयोगकर्ता इंटरफेस वर्चुअल एजेंट के साथ सहज बातचीत की सुविधा प्रदान करता है, जबकि वार्तालाप प्रबंधन मॉड्यूल बातचीत के प्रवाह को प्रबंधित करता है और पिछले प्रश्नों और प्रतिक्रियाओं सहित बातचीत के इतिहास को ट्रैक करता है। उपयोगकर्ता के इरादे को निर्धारित करने और इनपुट प्रश्नों को वेक्टर

तालिका 1: मवेशियों पर डेटाबेस में सूचीबद्ध नमूना प्रश्न और उत्तर

संदर्भ	प्रश्न	प्रतिक्रिया
इस्ट्रस	गाय में गर्मी के लक्षण क्या हैं?	गाय में गर्मी के लक्षण / इस्ट्रस के लक्षण हैं:अन्य गायों के साथ खड़े होने के लिए, अन्य गायों पर चढ़ने का प्रयास करना, योनि से रेशेदार श्लेष्मा का लटकना, नितंबों पर बलगम का जमा होना, बेचौनी और बार-बार चिल्लाना, दूध की पैदावार में गिरावट और भोजन का सेवन कम होना, ठुड्डी अन्य गायों द्वारा गाय की दुम पर आराम करना, पूंछ उठाना, योनि में सूजन, बार-बार पेशाब आना।
मादारोग संबंधी विकार	गायों में टीट ट्यूमर के लक्षण क्या हैं?	टीट ट्यूमफैक्शन से तात्पर्य किसी डेयरी पशु, आमतौर पर गाय के टीट की सूजन या वृद्धि से है। गायों में थन की सूजन के लक्षणों में थन की सूजन या वृद्धि, लालिमा, गर्मी, दर्द और थन से स्राव शामिल हो सकते हैं।
कृत्रिम गर्भाधान	मवेशियों में कृत्रिम गर्भाधान के लिए इष्टतम समय क्या है?	मद/गर्मी चक्र लगभग हर 21 दिन में एक बछिया या गाय को गर्भवती होने का मौका देता है। गर्मी के लक्षण प्रकट होने के 12 से 18 घंटे बाद अधिकतम गर्भाधान होता है। गर्मी के लक्षण दिखने के 12 से 18 घंटे के भीतर गाय को कृत्रिम गर्भाधान/एआई के लिए ले जाना चाहिए
कृत्रिम गर्भाधान	कृत्रिम गर्भाधान के लिए इष्टतम सफलता दर क्या होगी?	कृत्रिम गर्भाधान (एआई) में गर्भधारण दर/सफलता दर 40-50% है।
एफएमडी	मवेशियों में एफएमडी के इलाज का प्रोटोकॉल क्या है?	एफएमडी से प्रभावित जानवरों को अलग करके अलग रखा जाए। मुंह के छाले या पुटिका और पैर के घावों को 1% $KMNO_4$ (पोटेशियम परमैंगनेट घोल) या 2% सोडियम बाइकार्बोनेट घोल से अच्छी तरह धोना चाहिए। मुंह के घावों पर दिन में दो बार बोरोग्लिसरीन पेस्ट (बोरिक एसिड पाउडर+ग्लिसरॉल) या गुड़ लगाएं। द्वितीयक जीवाणु संक्रमण को नियंत्रित करने के लिए एंटीबायोटिक दवाओं का एक कोर्स दिया जाना चाहिए। बुखार और दर्द को कम करने के लिए ज्वरनाशक और दर्दनिवारक इंजेक्शन भी लगाए जाने चाहिए।

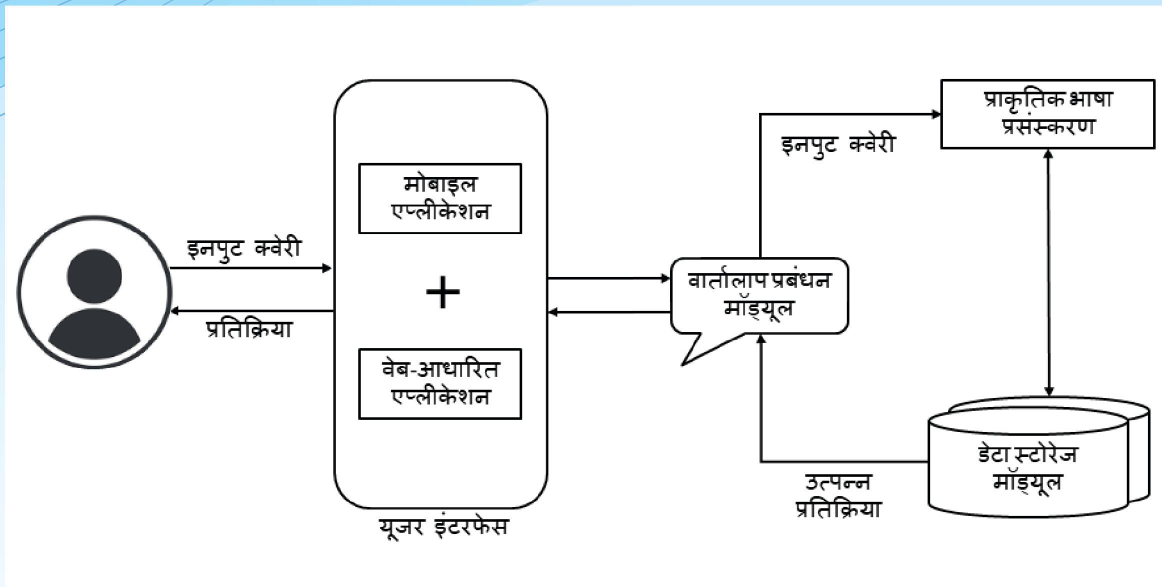
अभ्यावेदन में एन्कोड करने के लिए एनएलपी मॉड्यूल BERT [5], DistilBERT, XLNet और RoBERTa जैसे पूर्व-प्रशिक्षित भाषा मॉडल का उपयोग करता है। नॉलेजबेस में सबसे प्रासंगिक क्वेरी की पहचान करने के लिए कोसाइन समानता (समीकरण संख्या 2.1) उपायों का उपयोग करके इन वैक्टरों की तुलना बॉट के डेटाबेस से की जाती है।

$$\text{cosine_similarity}\ (A/B) = (A.B)/(||A||.||B||) \quad (2.1)$$

जहां, A एबं B डेटाबेस में इनपुट क्वेरी और प्रश्नों की सूची के लिए क्रमशः दो वैक्टर को वर्णन करना हैं। जुड़ी प्रतिक्रिया तब उपयोगकर्ता को दी जाती है। चित्र 1 डेयरी श्रिया चैटबॉट का सिस्टम वर्कफ्लो एबं आर्किटेक्चर को दर्शाता है। पशुधन किसानों और ऑन-फील्ड उपयोगकर्ताओं से प्रतिक्रिया इकट्ठा करने के लिए बॉट का क्षेत्रीय परीक्षण किया गया है। परीक्षण के दौरान, उपयोगकर्ताओं द्वारा पूछे गए प्रश्नों का मूल्यांकन बॉट-जनरेटेड प्रतिक्रियाओं के आधार पर किया गया है। प्रदर्शन का आकलन करने के लिए, इनपुट क्वेरी और मॉडल के अनुमानित आउटपुट की तुलना करते हुए एक समानता मीट्रिक नियोजित की गई थी। यह मीट्रिक दो वाक्यों के बीच समानता की डिग्री को मापता है, मिलान प्रश्नों में मॉडल की सटीकता और प्रभावशीलता में अंतर्दृष्टि प्रदान करता है और उचित प्रतिक्रियाएं उत्पन्न करता है।

परिणाम एवं चर्चा

चैटबॉट विकसित करने के लिए, गायों और भैंसों के विभिन्न स्वास्थ्य और उत्पादन पहलुओं पर सावधानीपूर्वक



चित्र 1: डेयरी श्रिया चैटबॉट का सिस्टम आर्किटेक्चर

तैयार किए गए 80,000 से भी अधिक प्रश्न-उत्तर जोड़ियों का एक डेटासेट संकलित किया गया है। इस डेटासेट का उपयोग विभिन्न पूर्व-प्रशिक्षित सेंटेन्स ट्रांसफार्मर मॉडल को प्रशिक्षित करने के लिए किया गया है, जिसमें BERT और इसके प्रकार जैसे RoBERTa, DistilBERT और XLNet शामिल है, जिनका उपयोग करके डेटाबेस क्वेरी को वेक्टर अभ्यावेदन में एन्कोड किया गया है। निकटतम मिलान की पहचान करने के लिए इन वैक्टरों की तुलना कोसाइन समानता मेट्रिक का उपयोग करके एक व्यक्तिगत इनपुट क्वेरी के एन्कोडेड वेक्टर के साथ की जाती है। सबसे समान क्वेरी के अनुरूप प्रतिक्रिया को चौटबॉट के उपयोगकर्ता इंटरफेस के माध्यम से अनुमानित उत्तर के रूप में दिखाया जाता है। इस बॉट का वेब-आधारित एप्लिकेशन पायथन के Django वेब फ्रेमवर्क का उपयोग करके विकसित किया गया है। चित्र 2 गाय पर एक उपयोगकर्ता के साथ चौटबॉट की सक्रिय बातचीत को दर्शाता है। डेयरी श्रिया बॉट के मोबाइल एप्लिकेशन को अब तक 1 हजार से अधिक उपयोगकर्ताओं द्वारा डाउनलोड किया जा चुका है।

निष्कर्ष

यह अध्ययन डेयरी पशुओं के स्वास्थ्य और उनकी उत्पादकता बढ़ाने में पशुपालकों की मदद के लिए एक

चित्र 2: सक्रिय बातचीत में शामिल डेयरी श्रिया चौटबॉट का स्नैपशॉट

स्वचालित चैटबॉट बनाने में पूर्व-प्रशिक्षित ट्रांसफार्मर-आधारित NLP मॉडलों के उपयोग और उनकी दक्षता को दिखाता है। हम डेयरी SHRIA नामक चैटबॉट पेश करते हैं, जो बहुभाषी उपयोगकर्ता प्रश्नों को समझने और स्वचालित उत्तर देने में सक्षम है। मॉडलों के तुलनात्मक विश्लेषण में पाया गया कि BERT मॉडल उपयोगकर्ता के इनपुट को समझने और डेटाबेस से सबसे प्रासंगिक प्रश्न ढूंढने में सबसे प्रभावी है। यह चौटबॉट पशुपालकों को उनके स्वास्थ्य और उत्पादन

से जुड़ी समस्याओं पर तुरंत सहायता प्रदान करता है, लेकिन यह केवल उन्हीं विषयों तक सीमित है जो डेटाबेस में शामिल हैं। यदि चैटबॉट को डेटाबेस में संबंधित प्रश्न नहीं मिलता है, तो यह उपयोगकर्ता के इनपुट के आधार पर एक Google लिंक प्रदान करता है, जो कई बार उपयोगी साबित हुआ है। भविष्य में, इस चैटबॉट को और बेहतर बनाने के लिए ऐसे फीचर्स विकसित किए जाएंगे, जिससे यह पीडीएफ या वर्ड फाइलों जैसे ज्ञान दस्तावेजों से सीधे प्रश्नों के उत्तर दे सके।

संदर्भ

[1]. https://dairyshria.icar.gov.in

[2]. https://play.google.com/store/apps/details?id=com.ivriapp.ivri_chatbot.ivri_chatbot&hl=en।

[3]. जैन, एन., जैन, पी., कायल, पी., साहित, जे., पचपांडे, एस., और चौधरी, जे. (2019)। एग्रीबॉट: कृषि-विशिष्ट प्रश्न उत्तर प्रणाली।

[4]. सकटा, डब्ल्यू., शिबाता, टी., तनाका, आर., और कुरोहाशी, एस. (2019)। क्वेरी-प्रश्न समानता और BERT-आधारित क्वेरी-उत्तर प्रासंगिकता का उपयोग करके FAQ पुनर्प्राप्ति। सूचना पुनर्प्राप्ति में अनुसंधान और विकास पर 42वें अंतर्राष्ट्रीय एसीएम एसआईजीआईआर सम्मेलन की कार्यवाही में, पृष्ठ 1113-1116।

[5]. केंटन, जे.डी.एम.डब्ल्यू.सी., और टुटानोवा, एल. के. (2019)। बर्ट: भाषा समझ के लिए गहरे द्वि दिश ट्रांसफार्मर का पूर्व-प्रशिक्षण। NaacL-HLT की कार्यवाही में, खंड 1, पृष्ठ 2।

कृषि में छवि-आधारित खरपतवार पहचान के लिए कृत्रिम बुद्धिमत्ता

सपना निगम[1], वी. के. चौधरी[2], अक्षय धीरज[1], मधु[1], मो अशरफुल हक[1], रजनी जैन[3], सुदीप मरवाह[1]
[1]भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली
[2]भा.कृ.अनु.प.-खरपतवार अनुसंधान निदेशालय, जबलपुर, मध्य प्रदेश
[3]भा.कृ.अनु.प.-राष्ट्रीय कृषि अर्थशास्त्र और नीति अनुसंधान संस्थान, नई दिल्ली

सार

खरपतवार वे अवांछित पौधे होते हैं जो उन स्थानों पर उगते हैं जहाँ इन्हें जानबूझकर उगाया नहीं जाता, ये वांछित वनस्पतियों से संसाधनों के लिए प्रतिस्पर्धा करते हैं और अक्सर आर्थिक या पारिस्थितिकीय नुकसान का कारण बनते हैं। इसलिए, पानी, पोषक तत्वों और सूर्यप्रकाश के लिए फसलों के साथ प्रतिस्पर्धा करने के कारण खरपतवार कृषि उत्पादन के लिए एक महत्वपूर्ण खतरा बन जाते हैं। साथ ही, पारंपरिक खरपतवार नियंत्रण तकनीकों और दृष्टिकोणों के पर्यावरणीय दुष्प्रभाव होते हैं। इसलिए, कृषि प्रणालियों की उत्पादकता और स्थिरता बनाए रखने के लिए खरपतवारों के संक्रमण को न्यूनतम करने के लिए प्रभावी खरपतवार प्रबंधन रणनीतियों की आवश्यकता है, जिसमें सभी सांस्कृतिक प्रथाएँ, यांत्रिक नियंत्रण, रासायनिक और जैविक नियंत्रण विधियाँ शामिल हैं। टिकाऊ खरपतवार नियंत्रण विधियों की बढ़ती मांग ने ऐसी नवाचार तकनीकों के विकास को प्रेरित किया है, जो हर्बीसाइड्स के उपयोग को कम करने का उद्देश्य रखती हैं। स्वचालित खरपतवार नियंत्रण इस संदर्भ में एक प्रमुख समाधान के रूप में उभरा है, जो खरपतवारों को अधिक सटीक और कुशल तरीके से लक्षित करने में मदद करता है, जिससे हर्बीसाइड्स का उपयोग कम होता है। कंप्यूटिंग हार्डवेयर में हालिया प्रगति, विशेष रूप से GPUs में, कृत्रिम बुद्धिमत्ता (AI) और डीप लर्निंग (DL) को एक परिवर्तनकारी दृष्टिकोण के रूप में उभारा है, जो कृत्रिम बुद्धिमत्ता और मशीन लर्निंग के व्यापक क्षेत्रों में पैटर्न वर्गीकरण के लिए सहायक सिद्ध हुआ है। कृत्रिम बुद्धिमत्ता आधारित खरपतवार पहचान ने खेतों से एकत्र की गई छवि डेटा से खरपतवारों की सटीक पहचान और वर्गीकरण में अद्वितीय क्षमता दिखाई है। इस प्रकार, स्वचालित AI आधारित एल्गोरिदम हर्बीसाइड्स के उपयोग को कम करते हुए फसल उत्पादकता और कृषि स्थिरता को बढ़ाते हैं।

मुख्य शब्द: कृत्रिम बुद्धिमत्ता, मशीन लर्निंग, छवि वर्गीकरण, सटीक कृषि, खरपतवार पहचान

परिचय

कृषि में प्रभावी खरपतवार नियंत्रण फसलों के स्वास्थ्य को बनाए रखने, स्थिरता सुनिश्चित करने, प्राकृतिक संसाधनों का संरक्षण करने और अंततः वैश्विक खाद्य सुरक्षा को बड़े पैमाने पर बढ़ाने में महत्वपूर्ण भूमिका निभाता है। खेतों में खरपतवार फसलों के साथ संसाधनों जैसे पानी, पोषक तत्वों और सूर्यप्रकाश के लिए प्रतिस्पर्धा करते हैं, जिससे फसल की उपज और गुणवत्ता में महत्वपूर्ण कमी आती है। फसलों पर इसका नकारात्मक प्रभाव न केवल किसानों के लिए नुकसान का कारण बनता है, बल्कि यह वैश्विक खाद्य उत्पादन के लिए भी खतरा उत्पन्न करता है। इसके अतिरिक्त, खरपतवारों के आर्थिक प्रभाव केवल उपज हानि तक सीमित नहीं होते; इसमें श्रम, हर्बीसाइड्स, मशीनरी से संबंधित खर्च और फसल की गुणवत्ता तथा बाजार मूल्य

पर प्रभाव भी शामिल होते हैं । इसके अलावा, पारंपरिक खरपतवार नियंत्रण विधियाँ जैसे हर्बीसाइड्स का उपयोग पारिस्थितिकी तंत्र के स्वास्थ्य और जैव विविधता को बनाए रखने में समस्या उत्पन्न कर सकती हैं, क्योंकि इससे मृदा अपरदन, जल प्रदूषण और जैव विविधता की हानि जैसी समस्याएँ हो सकती हैं । साथ ही, हर्बीसाइड प्रतिरोधी खरपतवारों की बढ़ती संख्या ने एकीकृत खरपतवार प्रबंधन रणनीतियों की आवश्यकता को उजागर किया है, ताकि प्रतिरोधी विकास को रोका जा सके और हर्बीसाइड्स की प्रभावशीलता सुनिश्चित की जा सके। परिणामस्वरूप, इन चुनौतियों का समाधान स्वचालित खरपतवार नियंत्रण विधियों के माध्यम से किया जा सकता है, जो उत्पादकता बढ़ाने, खाद्य आपूर्ति की बढ़ती मांग को पूरा करने और खाद्य संकट और अकाल से संबंधित जोखिमों को कम करने में मदद कर सकती हैं। कृत्रिम बुद्धिमत्ता (AI) और कृषि प्रौद्योगिकियों में प्रगति के माध्यम से स्वचालित खरपतवार प्रबंधन प्रथाएँ ऊपर उल्लेखित खरपतवार संक्रमणों द्वारा उत्पन्न समस्याओं के समाधान के रूप में एक आशाजनक विकल्प साबित हो सकती हैं। AI-प्रेरित खरपतवार पहचान कंप्यूटर दृष्टि और मशीन लर्निंग एल्गोरिदम का उपयोग करती है ताकि कृषि क्षेत्रों में खरपतवारों का स्वचालित रूप से पता लगाया जा सके और उन्हें वर्गीकृत किया जा सके। ये तकनीकें किसानों को खरपतवारों की उपस्थिति और वितरण के बारे में वास्तविक समय में जानकारी प्रदान करेंगी, जिससे उन्हें सूचित निर्णय लेने और अधिकतम संसाधन उपयोग के साथ लक्षित अनुशंसाएँ लागू करने में मदद मिलेगी।

स्वचालित खरपतवार पहचान की आवश्यकता

स्वचालित खरपतवार पहचान आधुनिक कृषि के लिए कई महत्वपूर्ण कारणों से एक अनमोल संपत्ति बन गई है। सबसे पहले और महत्वपूर्ण, स्वचालित विधियाँ मैन्युअल समाधानों की तुलना में अधिक दक्षता और सटीकता प्रदान करती हैं; ये अवांछित वनस्पतियों के प्रबंधन के लिए तेज और अधिक सटीक रणनीतियाँ प्रदान करती हैं । स्वचालित खरपतवार पहचान श्रम घंटों और हर्बीसाइड्स की लागत को घटाकर लागत में कमी लाती है, और लक्षित हर्बीसाइड अनुप्रयोग के माध्यम से पर्यावरणीय स्थिरता को बढ़ाती है, जो रासायनिक उपयोग को न्यूनतम करता है और पर्यावरण प्रभाव को कम करता है। यह आश्चर्यजनक रूप से तुरंत हस्तक्षेप सुनिश्चित करता है क्योंकि यह खरपतवार की उपस्थिति और वितरण पर वास्तविक समय में डेटा प्रदान करता है; किसान तब infestations के खिलाफ त्वरित कार्रवाई कर सकते हैं और उन्हें जल्दी से प्रबंधित कर सकते हैं। स्वचालित प्रणालियों की स्केलेबिलिटी उन्हें बड़े कृषि क्षेत्रों को प्रभावी ढंग से कवर करने के लिए आदर्श बनाती है - विशेष रूप से बड़े पैमाने पर कृषि संचालन में यह अत्यंत उपयोगी है। कुल मिलाकर, AI और कंप्यूटर विजन तकनीकों का उपयोग खरपतवार पहचान के लिए फसल उत्पादन बढ़ाता है, साथ ही पर्यावरणीय प्रभाव को कम करता है और कृषि संचालन की आर्थिक स्थिरता को बढ़ाता है। सफल खरपतवार प्रबंधन आवश्यक संसाधनों के लिए प्रतिस्पर्धा को कम करके फसल उत्पादन को अधिकतम करता है, कृषि संचालन की आर्थिक व्यवहार्यता को बढ़ाता है, उत्पादन लागत और नुकसान को न्यूनतम करता है, पारिस्थितिकी तंत्र को विघटन से बचाता है और खरपतवारों की वृद्धि को नियंत्रित करके प्रदूषण को रोकता है; नियंत्रण रणनीतियों में विविधता हर्बीसाइड प्रतिरोध को प्रबंधित करने में मदद करती है और इसकी प्रभावशीलता को बढ़ाती है; सतत कृषि प्रथाएँ संसाधनों के कुशल उपयोग को प्राथमिकता देती हैं, जबकि पारिस्थितिकी तंत्र की लचीलापन को बढ़ाती हैं - इस प्रकार खरपतवार नियंत्रण दीर्घकालिक उत्पादकता और पर्यावरणीय कल्याण के लिए अनिवार्य बन जाता है।

पारंपरिक विधियों की चुनौतियाँ

- **श्रम प्रधान:** पारंपरिक खरपतवार पहचान विधियाँ मैन्युअल श्रम पर अत्यधिक निर्भर करती हैं, जो बड़ी मात्रा में मानव संसाधन की आवश्यकता होती है और इसे पूरा करने में समय और श्रम की महत्वपूर्ण खपत होती है।
- **विषयगतता:** मानव पर्यवेक्षकों द्वारा की गई पहचान

में विषयगत निर्णय और गलत पहचान हो सकती है, जिसके कारण असंगत परिणाम प्राप्त होते हैं।

- **विशेषज्ञता की आवश्यकता:** खरपतवारों की सटीक पहचान के लिए अक्सर विशिष्ट वनस्पति ज्ञान और प्रशिक्षण की आवश्यकता होती है; यह इस कार्य को प्रभावी ढंग से करने वाले व्यक्तियों की संख्या को सीमित कर देता है।
- **समय की सीमा:** मैन्युअल पहचान की लंबी प्रक्रिया और हस्तक्षेप उपायों से जुड़े संभावित देरी के कारण, खरपतवार संक्रमण अक्सर अप्रबंधित बढ़ते हैं, जिससे फसलों को उम्मीद से कहीं अधिक नुकसान होता है।
- **लागत:** बड़े कृषि संचालन के लिए खरपतवार पहचान महंगी हो सकती है, जो भारी मात्रा में मैन्युअल पहचान प्रक्रियाओं पर निर्भर होते हैं।
- **बड़े पैमाने पर कृषि संचालन की अक्षमता:** पारंपरिक विधियाँ बड़े पैमाने पर कृषि संचालन के लिए उपयुक्त नहीं हो सकतीं, जहाँ विशाल खेत मैन्युअल खरपतवार पहचान को समय-साध्य और असुविधाजनक बना देते हैं।
- **सीमित कवरेज:** मैन्युअल पहचान विधियाँ केवल खेतों के विशिष्ट हिस्सों को कवर कर सकती हैं, जिससे कुछ स्थानों पर खरपतवार संक्रमण का अनुमान कम या नजरअंदाज किया जा सकता है।
- **मौसमी परिवर्तनशीलता:** चूंकि खरपतवार प्रजातियों की संरचना और प्रचुरता मौसमी रूप से बदलती रहती है, इससे पूरे वर्ष के दौरान स्थिर प्रबंधन रणनीतियाँ लागू करना कठिन हो जाता है।

इन चुनौतियों का समाधान करने में AI की भूमिका

कृत्रिम बुद्धिमत्ता (AI) पारंपरिक खरपतवार पहचान विधियों से जुड़ी चुनौतियों को हल करने में एक महत्वपूर्ण भूमिका निभाती है और यह खरपतवार प्रबंधन के कई क्षेत्रों में हमारे दृष्टिकोण को बदलने में सक्षम रहेगी। AI खरपतवार पहचान प्रक्रियाओं को सरल और तेज बना सकती है, जिससे मैन्युअल श्रम पर निर्भरता में भारी कमी आती है, जबकि समग्र दक्षता में वृद्धि होती है। AI खरपतवार पहचान के प्रयासों के लिए स्केलेबिलिटी को भी बढ़ावा देती है, जिससे यह संभव हो पाता है कि बड़े कृषि क्षेत्रों को प्रभावी रूप से कवर किया जा सके। AI-संचालित प्रणालियाँ वास्तविक समय में पहचानने की क्षमता प्रदान करती हैं, जिससे संभावित खरपतवार संक्रमण के जवाब में त्वरित हस्तक्षेप उपाय तुरंत किए जा सकते हैं और उनके नुकसान को वास्तविक समय में कम किया जा सकता है। AI तकनीकों में प्रारंभिक निवेश की आवश्यकता होती है; हालांकि, लंबी अवधि में श्रम और हर्बीसाइड्स की लागत में कमी आने के कारण AI-आधारित खरपतवार पहचान किसानों के लिए लागत-कुशल विकल्प बन जाती है ।

छवि-आधारित खरपतवार पहचान के लिए कार्यप्रणाली

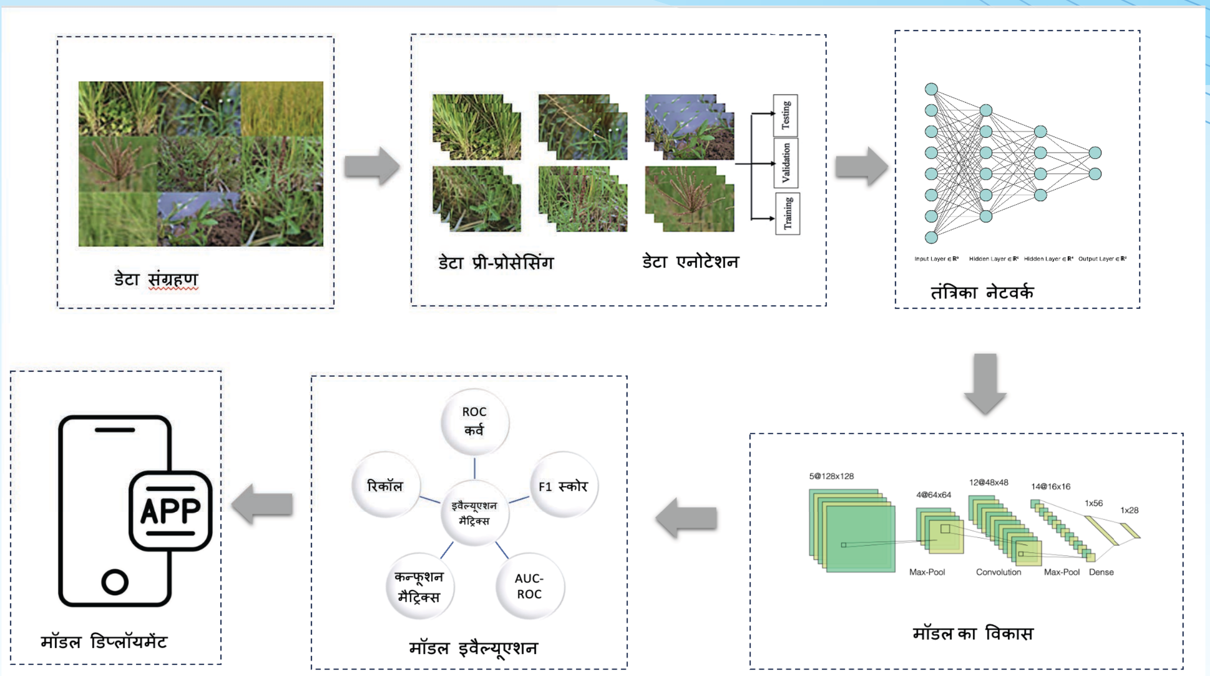
AI-संचालित खरपतवार पहचान में डेटा संग्रहण से लेकर मॉडल वैलिडेशन तक कई चरण शामिल होते हैं, जिनमें से प्रत्येक चरण सटीक और विश्वसनीय पहचान प्रणालियों को बनाने के लिए आवश्यक होता है। निम्नलिखित में इन चरणों को क्रमबद्ध तरीके से विस्तार से बताया गया है (चित्र 1):

a) डेटा संग्रहण (Data Acquisition)

इस चरण में, पहले चरण के तहत विभिन्न खरपतवार प्रजातियों, उनके विकास चरणों और पर्यावरणीय परिस्थितियों की छवियाँ एकत्र की जाती हैं, जिन्हें एक AI मॉडल को प्रशिक्षित करने के लिए उपयोग किया जाएगा, ताकि यह वास्तविक कृषि सेटिंग्स में मौजूद भिन्नताओं को पहचान सके। यह छवियाँ जमीन पर कैमरे, ड्रोन या उपग्रह चित्रण द्वारा प्राप्त की जा सकती हैं, जो पर्याप्त प्रतिनिधित्व प्रदान करती हैं।

b) डेटा पूर्व-प्रसंस्करण (Data Pre-processing)

एक बार जब छवियाँ एकत्र कर ली जाती हैं, तो उन्हें पूर्व-प्रसंस्करण से गुजरना पड़ता है ताकि गुणवत्ता बढ़ाई जा सके और उनके संग्रह का प्रारूप मानकीकरण किया जा सके। इसमें आकार बदलना, क्रॉपिंग करना और प्रकाश स्थिति को समायोजित करना शामिल हो सकता है ताकि डेटा सेट में एकरूपता बनी रहे। पूर्व-प्रसंस्करण AI मॉडल



चित्र 1: कृत्रिम बुद्धिमत्ता-आधारित खरपतवार पहचान में शामिल समग्र चरण

प्रशिक्षण के प्रदर्शन को अनुकूलित करने में महत्वपूर्ण भूमिका निभाता है।

c) एनोटेशन (Annotation)

अगला कदम यह है कि डेटा सेट में छवियों को इस प्रकार से एनोटेट किया जाए कि यह दर्शाए कि छवि में कहां और कब खरपतवार दिखाई दे रहे हैं। आमतौर पर बाउंडिंग बॉक्सेस या पिक्सल सेगमेंटेशन तकनीकों का उपयोग करके सभी उदाहरणों को सही तरीके से चिह्नित किया जाता है। सटीक एनोटेशन AI मॉडल को छवियों में खरपतवारों को अन्य वस्तुओं से पहचानने और उन्हें सही तरीके से अलग करने में सक्षम बनाता है।

d) मॉडल प्रशिक्षण (Model Training)

जब एक एनोटेटेड डेटा सेट उपलब्ध होता है, तो एक AI मॉडल को मशीन लर्निंग तकनीकों जैसे कन्वोल्यूशन न्यूरल नेटवर्क्स (CNNs) का उपयोग करके प्रशिक्षित किया जाता है। प्रशिक्षण सत्रों के माध्यम से, मॉडल यह सीखता है कि विभिन्न खरपतवार प्रजातियों से जुड़े पैटर्न और विशेषताएँ एनोटेटेड छवियों पर आधारित कैसे पहचानी जाती हैं। इसके बाद, इसके पैरामीटर को धीरे-धीरे समायोजित किया जाता है ताकि भविष्यवाणी की त्रुटियों को कम किया जा सके और प्रदर्शन को बेहतर बनाया जा सके।

e) मॉडल मूल्यांकन (Model Evaluation)

प्रशिक्षण पूरा होने के बाद, एक अलग मान्यकरण डेटा सेट का उपयोग किया जाता है, जिसे AI मॉडल ने पहले कभी नहीं देखा है, ताकि इसके प्रदर्शन का मूल्यांकन किया जा सके। मूल्यांकन मीट्रिक्स जैसे सटीकता (Accuracy), शुद्धता (Precision), पुनः प्राप्ति (Recall) और F1 स्कोर यह मापने के लिए उपयोग किए जाते हैं कि यह खरपतवारों की पहचान कितनी सटीकता से करता है। मॉडल मूल्यांकन यह सुनिश्चित करने में मदद करता है कि इसका प्रदर्शन अज्ञात डेटा पर भी सामान्यीकृत हो और वास्तविक जीवन की परिस्थितियों में विश्वसनीय रूप से कार्य करता है।

f) मॉडल तैनाती (Model Deployment)

सफल मूल्यांकन के बाद, AI मॉडल को कृषि क्षेत्रों में खरपतवार पहचान के उद्देश्य से तैनात किया जाता है। इसके उद्देश्य के अनुसार, इसे वास्तविक समय में पहचान के लिए ड्रोन या स्मार्टफोन जैसे उपकरणों के माध्यम से लागू किया जा सकता है, या फिर इसे मौजूदा कृषि मशीनरी

या सॉफ्टवेयर प्लेटफॉर्म में एकीकृत किया जा सकता है।

g) मॉडल निगरानी और अद्यतन (Model Monitoring and Updates)

AI मॉडल के प्रदर्शन पर नियमित समीक्षा आवश्यक होती है ताकि समय के साथ किसी भी बदलाव या गिरावट का पता लगाया जा सके। इसके अलावा, प्रजातियों की संरचना, पर्यावरणीय परिस्थितियाँ, या पहचान सटीकता पर प्रभाव डालने वाले अन्य कारकों में बदलाव के कारण अद्यतन या फिर से प्रशिक्षण की आवश्यकता हो सकती है। यह सुनिश्चित करता है कि मॉडल लगातार उच्च प्रदर्शन करता रहे और कृषि क्षेत्र में खरपतवार पहचान के लिए प्रभावी बना रहे।

खरपतवार पहचान में उपयोग किए जाने वाले मशीन लर्निंग एल्गोरिदम का अवलोकन

मशीन लर्निंग एल्गोरिदम खरपतवार पहचान में एक महत्वपूर्ण भूमिका निभाते हैं, जो छवियों से निकाली गई विशेषताओं के आधार पर स्वचालित वर्गीकरण प्रदान करते हैं। इन कार्यों में कई प्रकार की मशीन लर्निंग तकनीकों का उपयोग किया गया है; SVMs, रैंडम फॉरेस्ट, CNNs, DCNNs और ट्रांसफर लर्निंग ने प्रत्येक ने कृषि क्षेत्र में खरपतवारों के स्वचालित वर्गीकरण में विशिष्ट लाभ प्रदान किए हैं, जैसा कि नीचे बताया गया है:

a) **सपोर्ट वेक्टर मशीन (SVM):** यह एक सुपरवाइज्ड लर्निंग मॉडल है जो वर्गीकरण कार्यों के लिए उपयोग किया जाता है। SVMs एक आदर्श हाइपरप्लेन (hyperplane) खोजने का कार्य करती हैं जो उनके फीचर स्पेस में विभिन्न वर्गों को सर्वोत्तम तरीके से अलग करता है। खरपतवार पहचान के लिए, इसे छवियों से निकाली गई विशेषताओं जैसे रंग, बनावट और आकार का उपयोग करके छवियों को विशिष्ट खरपतवार प्रजातियों वाले समूहों में वर्गीकृत करने के लिए लागू किया गया है।

b) **रैंडम फॉरेस्ट (Random Forest):** यह एक एन्सेम्बल लर्निंग विधि है जो भविष्यवाणियाँ करने के लिए कई निर्णय वृक्षों का उपयोग करती है। प्रत्येक पेड़ को प्रशिक्षण डेटा के रैंडम उपसमूहों पर प्रशिक्षित किया जाता है, फिर छवियों से निकाली गई विशेषताओं के आधार पर इसका वर्गीकरण निर्णय लिया जाता है। रैंडम फॉरेस्ट को छवियों से निकाली गई विशेषताओं के माध्यम से खरपतवार प्रजातियों का वर्गीकरण करने में सफलतापूर्वक लागू किया गया है।

c) **कन्वोल्यूशन न्यूरल नेटवर्क्स (CNNs):** CNNs एक डीप लर्निंग आर्किटेक्चर है जिसे विशेष रूप से दृश्य डेटा प्रसंस्करण के लिए विकसित किया गया है। CNNs में कई परतों के कन्वोल्यूशन ऑपरेशन्स होते हैं, इसके बाद पूरी तरह से जुड़े हुए परतों का उपयोग वर्गीकरण के लिए किया जाता है। जब इसे खरपतवार पहचान कार्यों में उपयोग किया जाता है, तो CNNs ने शानदार प्रदर्शन किया है और फसल और खरपतवार के बीच अंतर करने और प्रजाति वर्गीकरण कार्यों में अत्याधुनिक परिणाम प्राप्त किए हैं।

d) **डीप कन्वोल्यूशन न्यूरल नेटवर्क्स (DCNNs):** यह डीप लर्निंग आर्किटेक्चर हैं जो CNNs को अतिरिक्त परतों और पैरामीटर्स के साथ बढ़ाते हैं, जिससे यह खरपतवार पहचान कार्यों में स्वचालित रूप से कच्चे छवि डेटा से पदानुक्रमिक विशेषताओं को सीखने में बेहतर प्रदर्शन करते हैं DCNNs का सफलतापूर्वक उपयोग प्रजाति वर्गीकरण, घनत्व अनुमान, और कृषि क्षेत्रों में फसल-खरपतवार विभाजन के लिए किया गया है।

e) **ट्रांसफर लर्निंग (TL):** ट्रांसफर लर्निंग एक उन्नत मशीन लर्निंग तकनीक है, जिसमें एक मौजूदा मॉडल को एक कार्य के लिए पहले से प्रशिक्षित किया जाता है और फिर इसे सीमित प्रशिक्षण डेटा के साथ दूसरे कार्य के लिए फाइन-ट्यून (fine-tune) किया जाता है। ट्रांसफर लर्निंग का उपयोग खरपतवार पहचान कार्यों में सफलतापूर्वक किया गया है, जहां बड़े छवि डेटा सेट्स पर पहले से प्रशिक्षित CNN मॉडल्स को

लिया जाता है और उन्हें खरपतवार प्रजाति पहचान कार्यों के लिए डेटा के साथ फाइन-ट्यून किया जाता है, जिससे इन पहचान मॉडल्स में सामान्यीकरण और मजबूती में सुधार होता है।

AI-आधारित खरपतवार पहचान प्रणालियों का कार्यान्वयन

कृषि परिवेश में AI-संचालित खरपतवार पहचान प्रणालियों को प्रभावी, विश्वसनीय और किसानों के कामकाजी तरीकों में एकीकृत करने के लिए कुछ व्यावहारिक मुद्दों पर विचार करना आवश्यक है। AI मॉडल्स को बड़े खेतों का समर्थन करने में सक्षम होना चाहिए, जैसे कृषि में पाए जाते हैं, जबकि ये प्रकाश की स्थिति, मौसम के पैटर्न या खरपतवार प्रजातियों की संरचना में बदलावों के प्रति प्रतिरोधी बने रहें। AI प्रणालियाँ इतनी मजबूत होनी चाहिए कि वे किसी भी परिस्थितियों में खरपतवारों की पहचान कर सकें, चाहे उपयोगकर्ता इंटरफेस में कोई समस्या हो। यह उपयोगकर्ता-मित्र तकनीकी इंटरफेस अपनी तकनीक को सरल बनाता है, जबकि यह हितधारकों को तकनीकी निवेश से अधिक लाभ प्राप्त करने में मदद करता है। कार्यान्वयन से संबंधित लागतों जैसे कि हार्डवेयर अधिग्रहण और सॉफ्टवेयर रखरखाव को संभावित लाभों के खिलाफ संतुलित किया जाना चाहिए, जैसे कि खरपतवार प्रबंधन दक्षता में सुधार और फसल की पैदावार, जिनमें हार्डवेयर आवश्यकताएँ, डेटा संग्रहण, एनोटेशन प्रक्रियाएँ, और मौजूदा कृषि प्रथाओं के साथ एकीकरण शामिल हैं। AI-आधारित खरपतवार नियंत्रण प्रणालियों के उपयोग के लिए निम्नलिखित बिंदुओं को ध्यान में रखना आवश्यक है:

हार्डवेयर आवश्यकताएँ

- **संवेदनशील उपकरण:** ड्रोन, ट्रैक्टर या अन्य कृषि मशीनरी में अक्सर उच्च-रिजॉल्यूशन कैमरे या मल्टीस्पेक्ट्रल सेंसर होते हैं जो कृषि क्षेत्रों की छवियाँ कैप्चर करते हैं।
- **संगणन संसाधन:** वास्तविक-समय में उपयोग के लिए, AI एल्गोरिदम को प्रभावी ढंग से और निर्बाध रूप से चलाने के लिए पर्याप्त कंप्यूटर शक्ति उपलब्ध होनी चाहिए। कनेक्टिविटी और विलंबता की आवश्यकताओं के आधार पर, उपकरण पर या क्लाउड प्रोसेसिंग समाधान का उपयोग किया जा सकता है।
- **ऊर्जा स्रोत:** क्षेत्र में तैनाती के दौरान, संवेदनशील उपकरणों और कंप्यूटिंग डिवाइसों के उचित संचालन को बनाए रखने के लिए बैटरियाँ या सोलर पैनल जैसे विश्वसनीय ऊर्जा आपूर्ति स्रोत उपलब्ध होने चाहिए।

डेटा संग्रहण और एनोटेशन

- **विविध और प्रतिनिधित्वकारी डेटासेट्स:** शक्तिशाली AI मॉडल्स को प्रशिक्षित करने के लिए, फसलें, खरपतवार प्रजातियाँ, विकास चरण और पर्यावरणीय परिस्थितियाँ को कवर करने वाले विविध और प्रतिनिधित्वकारी डेटासेट्स का संग्रह करना आवश्यक है।
- **एनोटेटेड डेटा:** छवियों को मैन्युअल रूप से एनोटेट (लेबल) करना समय लेने वाला हो सकता है, लेकिन यह पर्यवेक्षित शिक्षा दृष्टिकोण में आवश्यक है, इसलिए स्वचालित या अर्द्ध-स्वचालित एनोटेशन उपकरण इस प्रक्रिया को सुविधाजनक बना सकते हैं।
- **गुणवत्ता नियंत्रण:** AI मॉडल्स को पक्षपाती या गलतियों से बचाने के लिए, एनोटेटेड डेटा को AI कार्यक्रम में शामिल किए जाने से पहले गुणवत्ता नियंत्रण जांच से गुजरना चाहिए।

मौजूदा कृषि प्रथाओं के साथ एकीकरण

- **संगतता:** AI-संचालित खरपतवार पहचान प्रणालियाँ मौजूदा कृषि उपकरणों, सॉफ्टवेयर और मशीनरी के साथ संगत होनी चाहिए, जो किसान उपयोग करते हैं।
- **निर्बाध एकीकरण:** इसमें AI सिफारिशों को सटीक कृषि तकनीकों जैसे GPS-निर्देशित ट्रैक्टरों या परिवर्तनीय दर अनुप्रयोगों के साथ निर्बाध रूप से एकीकृत करना शामिल हो सकता है, ताकि AI सिफारिशों के आधार पर लक्षित खरपतवार नियंत्रण क्रियाएँ की जा सकें।

- **प्रशिक्षण और समर्थन:** दैनिक कृषि संचालन में सफल एकीकरण के लिए, इन प्रणालियों के उपयोगकर्ताओं जैसे किसानों और कृषि वैज्ञानिको को प्रशिक्षण और समर्थन प्रदान किया जाना चाहिए।

निष्कर्ष

कृत्रिम बुद्धिमत्ता (AI) में कृषि को क्रांतिकारी बनाने की क्षमता है, विशेष रूप से स्वचालित खरपतवार पहचान के द्वारा। यह दृष्टिकोण आर्थिक और पारिस्थितिकीय दृष्टिकोण से कई फायदे प्रदान करता है। कृत्रिम बुद्धिमत्ता प्रौद्योगिकियाँ उस प्रक्रिया को सुधार सकती हैं जिसके द्वारा कृषि विशेषज्ञ खरपतवारों की पहचान करते हैं और उनका प्रबंधन करते हैं, जिससे अधिक उत्पादकता और प्रभावशीलता प्राप्त होती है। AI-संवर्धित वास्तविक-समय की जानकारी सटीक कृषि तकनीकों को बेहतर बनाती है। यह संसाधन उपयोग, फसल की पैदावार को अनुकूलित करता है और खरपतवार प्रबंधन के लिए ऐसे स्थिर तरीके को बढ़ावा देता है जो रासायनिक उपयोग को घटाता है। कृत्रिम बुद्धिमत्ता को खरपतवार-प्रबंधन तकनीकों में शामिल करना किसानों और समाज के लिए महत्वपूर्ण आर्थिक, पर्यावरणीय और सामाजिक लाभ प्रदान करता है। AI-आधारित खरपतवार नियंत्रण उपकरण स्थिर कृषि को बढ़ावा देने और वैश्विक खाद्य सुरक्षा में योगदान देने का विशाल वादा रखते हैं। इन प्रौद्योगिकियों की दक्षता को बढ़ाकर, लागत को घटाकर और पर्यावरणीय प्रभाव को न्यूनतम करके, ये स्थिर खेती को बढ़ावा दे सकते हैं और वैश्विक खाद्य सुरक्षा को बेहतर बना सकते हैं। हितधारकों को कृत्रिम बुद्धिमत्ता (AI) द्वारा संचालित प्रणालियों के विकास में व्यावहारिक पहलुओं पर विचार करना चाहिए, जैसे कि इसके विस्तार की क्षमता और तनाव के प्रति प्रतिरोध, हार्डवेयर विनिर्देश, डेटा संग्रहण और एनोटेशन के तरीके, और वर्तमान कृषि प्रथाओं के लिए संगतता। AI के पास खरपतवार-प्रबंधन तकनीकों को उन्नत करने की क्षमता है, जबकि यह भविष्य में कृषि को पर्यावरण के अनुकूल और अधिक उत्पादक परिणामों की ओर प्रगति करने में मदद करेगा।

एएसएमआईआर: पौधों में अजैविक तनाव-विशिष्ट माइक्रो आरएनए के पूर्वानुमान हेतु मशीन लर्निंग की रूपरेखा

उपेन्द्र कुमार प्रधान, प्रविन कुमार मेहर और एस.पी. सिंह

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

हाल के वर्षों में अजैविक तनाव अपनी व्यापक प्रकृति और पौधों की वृद्धि, विकास और गुणवत्ता पर चौंकाने वाले प्रभावों के कारण एक बड़ी चुनौती बन गया है। माइक्रो आरएनए (एमआईआरएनएस) विभिन्न अजैविक तनावों के प्रति पौधों की प्रतिक्रिया में महत्वपूर्ण भूमिका निभाते हैं। इस प्रकार, विशिष्ट अजैविक तनाव-उत्तरदायी माइक्रो आरएनए की पहचान अजैविक तनावों के प्रति प्रतिरोधी किस्मों को विकसित करने के लिए फसल प्रजनन कार्यक्रमों में अत्यधिक महत्व रखती है। इस अध्ययन में, हमने ठंड, सूखा, गर्मी और नमक जैसे चार विशिष्ट अजैविक तनावों से जुड़े माइक्रो आरएनए की पूर्वानुमान के लिए एक मशीन लर्निंग-आधारित कम्प्यूटेशनल मॉडल विकसित किया। केमर आकार 1 से 5 की छद्म के-ट्यूपल न्यूक्लियोटाइड संरचना संबंधी विशेषताओं का उपयोग माइक्रो आरएनए को संख्यात्मक रूप में दर्शाने के लिए किया गया था। महत्वपूर्ण विशेषताओं का चयन करने के लिए फीचर चयन रणनीति को नियोजित किया गया था। चयनित फीचर सेट के साथ, सपोर्ट वेक्टर मशीन (एसवीएम) ने सभी चार अजैविक तनाव स्थितियों में उच्चतम क्रॉस-वैलिडेशन सटीकता हासिल की। परिशुद्धता-रिकॉल वक्र के तहत क्षेत्र के संदर्भ में उच्चतम क्रॉस-मान्य पूर्वानुमान सटीकता ठंड, सूखा, गर्मी और नमक के लिए क्रमशः 90.15, 90.09, 87.71 और 89.25% पाई गई। अजैविक तनावों के लिए स्वतंत्र डेटासेट के लिए समग्र पूर्वानुमान सटीकता क्रमशः 84.57, 80.62, 80.38 और 82.78% देखी गई। एसवीएम को अजैविक तनाव प्रतिक्रियाशील माइक्रो आरएनए की पूर्वानुमान के लिए विभिन्न गहन शिक्षण मॉडलों से बेहतर प्रदर्शन करते हुए भी देखा गया। हमारी पद्धति को आसानी से लागू करने के लिए, https://iasri-sg.icar.gov.in/asmir/ पर एक ऑनलाइन पूर्वानुमान सर्वर "एएसएमआईआर" स्थापित किया गया है। माना जाता है कि प्रस्तावित कम्प्यूटेशनल मॉडल और विकसित पूर्वानुमान उपकरण पौधों में विशिष्ट अजैविक तनाव-उत्तरदायी माइक्रो आरएनए की पहचान के लिए मौजूदा प्रयास के पूरक हैं।

मुख्य शब्द: अजैविक तनाव; एमआईआरएनएस; यंत्र अधिगम; ध्यान लगा के पढ़ना या सीखना; कम्प्यूटेशनल बायोलॉजी

परिचय

लगातार बढ़ती मानव आबादी की मांगों को पूरा करने के लिए सतत खाद्य उत्पादन आवश्यक है (मोचिदा और शिनोजाकी 2013)। इसके विपरीत, फसल के पौधे लगातार प्रतिकूल पर्यावरणीय गड़बड़ी के संपर्क में रहते हैं, जिसके परिणामस्वरूप महत्वपूर्ण कृषि फसलों में 70% उपज हानि होने की पूर्वानुमान की जाती है (बॉयर 1982; विज और त्यागी 2007; जुरब्रिगेन एट अल। 2010)। ठंड, सूखा, गर्मी, नमक और पोषक तत्वों की कमी जैसे अजैविक तनाव, फसल की उपज और उत्पादकता को सीमित करने में एक प्रमुख कारक रहे हैं (अकपनार एट अल. 2013; बुडक एट अल. 2015)। पौधों की वृद्धि, विकास और गुणवत्ता पर अजैविक तनाव की व्यापकता और चौंकाने वाले प्रभावों

ने उन्हें हाल के वर्षों में एक महत्वपूर्ण चिंता का विषय बना दिया है (अनवर और किम 2020)। अजैविक तनाव के जवाब में रक्षा तंत्र को सक्रिय करने के लिए, पौधे आनुवंशिक विनियमन के एक नेटवर्क को सक्रिय करते हैं, जिसमें ट्रांसक्रिप्शनल और/या पोस्ट-ट्रांसक्रिप्शनल विनियमन (कू एट अल 2015) के माध्यम से काफी संख्या में जीन में परिवर्तित जीन अभिव्यक्ति शामिल होती है। पौधों में सुरक्षात्मक जीन की अभिव्यक्ति विशेष रूप से बढ़ जाती है, जबकि नकारात्मक नियामकों की अभिव्यक्ति कम हो जाती है। कई प्रोटीन-कोडिंग जीन जो नियंत्रित करते हैं कि पौधे अजैविक तनावों पर कैसे प्रतिक्रिया करते हैं, हाल के वर्षों में खोजे गए हैं (झांग और वांग 2015)।

हाल के निष्कर्षों से पता चलता है कि पौधे अजैविक तनाव (झांग 2015) के तहत पौधों की वृद्धि और विकास को रोकने के लिए प्रमुख पोस्ट-ट्रांसक्रिप्शनल जीन-अभिव्यक्ति नियामक के रूप में माइक्रोआरएनए (एमआईआरएनए) नामक छोटे (20-24 एनटी) अंतर्जात आरएनए का उपयोग करते हैं। एमआरएनए क्लीवेज, ट्रांसलेशनल सप्रेशन, क्रोमैटिन रीमॉडलिंग, और/या डीएनए मिथाइलेशन कुछ ऐसे तरीके हैं जो माइक्रो आरएनए जीन अभिव्यक्ति को नियंत्रित करते हैं (वांग एट अल। 2019)। आमतौर पर, अजैविक तनाव के जवाब में अपग्रेड किए गए माइक्रो आरएनए अपने लक्ष्य माइक्रो आरएनए को डाउनरेगुलेट करते हैं, जबकि जो दबाए जाते हैं वे सकारात्मक नियामकों को जमा करने और सक्रिय होने का कारण बनते हैं (चिन्नुसामी एट अल। 2007)। कई शोधों के अनुसार, अजैविक तनाव पौधों में असंगत माइक्रो आरएनए अभिव्यक्ति की ओर ले जाता है। उदाहरण के लिए, विंटर और डिडेरिच (2011) और इवाकावा और तोमारी (2013) ने पाया कि जटिल जीन नेटवर्क के महत्वपूर्ण तत्वों को नियंत्रित करने के माध्यम से, अजैविक तनाव के प्रति पौधों की प्रतिक्रिया में माइक्रो आरएनए की भूमिका होती है। जैविक और अजैविक तनावों (नोमन और अकील 2017) के जवाब में पौधों की माइक्रो आरएनए अभिव्यक्ति में परिवर्तन का विश्लेषण करने के लिए कई अध्ययन किए गए हैं। माइक्रो आरएनए -167, माइक्रो आरएनए -169, माइक्रो आरएनए -171, माइक्रो आरएनए -319, माइक्रो आरएनए -393, माइक्रो आरएनए -394, और माइक्रो आरएनए -396 माइक्रो आरएनए के कुछ उदाहरण हैं जो विभिन्न अजैविक तनाव-संबंधी गतिविधियों में काम करते हैं (वांग एट अल 2014) ; गाओ एट अल. 2016)।

अजैविक तनावों के प्रति माइक्रो आरएनए की प्रतिक्रिया काफी हद तक जीनोटाइप, तनाव, ऊतक और माइक्रो आरएनए प्रकार (झांग 2015) द्वारा तय की जाती है। उदाहरण के लिए, माइक्रो आरएनए 408 एक्सप्रेशन को डाउन-रेगुलेट किया गया है सूखे के तनाव के दौरान चावल (झोउ एट अल. 2010), कपास (जी एट अल. 2015), और आड़ू (एल्डेम एट अल. 2012), जबकि अरेबिडोप्सिस (लियू एट अल. 2008), मेडिकागो (ट्रिनडेडे) में इसे विनियमित किया गया है। एट अल. 2009), और जौ (कांतार एट अल. 2011)। माइक्रो आरएनए की ऊतक-निर्भर प्रतिक्रिया के संदर्भ में, वांग एट अल। (2013) ने कपास में सूखे और लवणता के तनाव के जवाब में जड़ों बनाम पत्तियों में माइक्रो आरएनए की एक परिवर्तित अभिव्यक्ति प्रोफाइल की खोज की। एमआईआर 169 को एराबिडोप्सिस में लवणता उपचार से प्रेरित पाया गया लेकिन सूखे के तनाव से बाधित हुआ (ली एट अल. 2008), यह दर्शाता है कि अजैविक तनाव तनाव-निर्भर तरीके से माइक्रो आरएनए की अभिव्यक्ति को प्रेरित करते हैं। एराबिडोप्सिस में माइक्रो आरएनए 398 के समान, माइक्रो आरएनए 398 को यूवीबी प्रकाश द्वारा सक्रिय किया गया था, लेकिन लवणता, ठंड और ऑक्सीडेटिव तनाव (सुकर एट अल। 2006; जिया एट अल। 2009) द्वारा दबा दिया गया था। लवणता तनाव के तहत एराबिडोप्सिस में, माइक्रो आरएनए 397 की अभिव्यक्ति को काफी प्रेरित किया गया था, लेकिन माइक्रो आरएनए 398 की अभिव्यक्ति को काफी हद तक बाधित किया गया था, यह दर्शाता है कि अजैविक तनाव के प्रति पौधों की प्रतिक्रिया माइक्रो आरएनए -निर्भर है (लियू एट अल। 2008)। ऊपर उल्लिखित अध्ययनों से संकेत मिलता है कि माइक्रो आरएनए इसमें महत्वपूर्ण भूमिका निभाते हैं कि पौधे विभिन्न अजैविक तनावों पर कैसे प्रतिक्रिया करते हैं और पौधों को ऐसे अजैविक तनावों

के प्रति अधिक लचीला बनाने के लिए आनुवंशिक लक्ष्य के रूप में उपयोग किया जा सकता है। माइक्रो आरएनए की महत्वपूर्ण भूमिका के कारण, उन्हें प्लांटमिरनाटी (री एट अल. 2015), एमआईआरप्लांट (एन एट अल. 2014), पीएमआरडी (झांग एट अल. 2010), एमआईआरएनईएसटी (स्जेस्निएक एट अल) सहित विभिन्न डेटाबेस में शामिल किया गया है। 2012), और माइक्रोबेस (कोजोमारा एट अल. 2014)। अजैविक तनाव-उत्तरदायी माइक्रो आरएनए के लिए सबसे हालिया संसाधन पीएनसीस्ट्रेस (वू एट अल. 2020) है, जिसमें विविध अजैविक और जैविक तनावों से जुड़े प्रयोगात्मक रूप से मान्य माइक्रो आरएनए अनुक्रम शामिल हैं।

अजैविक तनाव से संबंधित माइक्रो आरएनए को खोजने के लिए आरटी-पीसीआर, क्लोनिंग, आरएनए-माइक्रोएरे और उत्तरी ब्लॉट सहित तकनीकों को बड़े पैमाने पर नियोजित किया गया है। इन संसाधन-गहन गीले प्रयोगों में सटीकता, रैखिक सीमा और पता लगाने की सीमा (कू एट अल. 2015; श्रीराम एट अल. 2016) सहित कमजोर विश्लेषणात्मक गुण भी हैं। यद्यपि एनजीएस और गहन अनुक्रमण प्रौद्योगिकियों (त्रिपाठी एट अल.2015) का उपयोग करके अजैविक तनाव-उत्तरदायी माइक्रो आरएनए की पहचान की गई है, अनुक्रमण विधियाँ प्रजाति-विशिष्ट हैं। इसलिए, मौजूदा प्लांट माइक्रो आरएनए अनुक्रम डेटा को नियोजित करना, मशीन लर्निंग-आधारित कम्प्यूटेशनल दृष्टिकोण अजैविक तनाव से संबंधित माइक्रो आरएनए की पूर्वानुमान के लिए एक बेहतर विकल्प हो सकता है। पादप माइक्रो आरएनए अनुक्रमों से अजैविक तनाव-संबंधी माइक्रो आरएनए की पूर्वानुमान करने के लिए, हमने पहले से ही एएसआरमाइक्रो आरएनए (मेहर एट अल. 2022) नामक एक मशीन लर्निंग-आधारित तकनीक विकसित की है। विकसित मॉडल अपने अनुक्रम से अजैविक प्रतिक्रियाशील माइक्रो आरएनए की पूर्वानुमान करता है। लेकिन, पौधों के माइक्रो आरएनए अनुक्रमों से विशिष्ट अजैविक तनाव के लिए माइक्रो आरएनए की पूर्वानुमान करना अभी भी आवश्यक है। अजैविक तनावों के प्रति पौधों की प्रतिक्रिया में माइक्रो आरएनए महत्व और ऐसे अजैविक तनाव विशिष्ट माइक्रो आरएनए की पूर्वानुमान के लिए कम्प्यूटेशनल तरीकों की कमी को देखते हुए, इस अध्ययन का उद्देश्य अजैविक तनाव-विशिष्ट (ठंडा, सूखा) की पूर्वानुमान के लिए एक मशीन लर्निंग-आधारित कम्प्यूटेशनल मॉडल विकसित करना है। गर्मी, और नमक) माइक्रो आरएनए अनुक्रमों से प्राप्त सुविधाओं का उपयोग करके माइक्रो आरएनए। कुछ अजैविक तनावों के तहत माइक्रो आरएनए की खोज के उद्देश्य से, वर्तमान अध्ययन को वेट-लैब तकनीकों और अन्य अनुक्रमण दृष्टिकोणों का पूरक माना जाता है।

सामग्री और तरीके

डेटासेट का संग्रह, प्रसंस्करण और निर्माण

23 अगस्त, 2022 को, अजैविक तनाव के लिए विशेष रूप से परिपक्व माइक्रो आरएनए अनुक्रम एकत्र करने के लिए पीएनसीस्ट्रेस डेटाबेस (वू एट अल। 2020) का उपयोग किया गया था। इस डेटाबेस में 114 पौधों से 4227 तनाव-उत्तरदायी गैर-कोडिंग आरएनए (माइक्रो आरएनए, लनसीआरएनए, और सीआईआरसीआरएनए) शामिल हैं जिन्हें 48 जैविक और 91 अजैविक तनावों का जवाब देने के लिए प्रयोगात्मक रूप से सत्यापित किया गया है। हमने 4 अलग-अलग अजैविक तनावों के लिए 2110 माइक्रो आरएनए अनुक्रम एकत्र किए, जिनमें सूखा (862), गर्मी (241), नमक (559) और ठंड (448) शामिल हैं। इसके अतिरिक्त, हमने 376 माइक्रो आरएनए अनुक्रमों को ध्यान में रखा जिनका उपयोग पिछले अध्ययन में एक नकारात्मक सेट के रूप में किया गया था (मेहर एट अल 2022)। हमने एक विशेष अजैविक तनाव के लिए विशिष्ट माइक्रो आरएनए की पूर्वानुमान के लिए मशीन लर्निंग एल्गोरिदम के प्रदर्शन का मूल्यांकन करने के लिए डेटासेट-I और डेटासेट-II नामक दो अलग-अलग डेटासेट बनाए।

डेटा सेट-I

प्रत्येक तनाव श्रेणी के लिए एकत्रित अजैविक तनाव अनुक्रमों में से तीस प्रतिशत (ठंड के लिए 128 अनुक्रम, सूखे के लिए 267 अनुक्रम, गर्मी के लिए 68 और नमक के लिए 167 अनुक्रम) को एक सकारात्मक स्वतंत्र परीक्षण सेट के रूप में उपयोग करने के लिए अलग रखा गया

था। प्रशिक्षण डेटासेट का सकारात्मक सेट अजैविक तनाव की प्रत्येक श्रेणी से शेष माइक्रो आरएनए अनुक्रमों से बना था। पूर्वानुमान सटीकता में समरूप पूर्वाग्रह को रोकने के लिए, प्रत्येक तनाव सेट के भीतर किसी भी अन्य अनुक्रम के लिए 60% से अधिक अनुक्रम समरूपता वाले अनुक्रमों को सीडी-एचआईटी एल्गोरिदम (हुआंग एट अल 2010) का उपयोग करके समाप्त कर दिया गया था। अनावश्यक अनुक्रमों को हटाने के बाद, ठंड, सूखा, गर्मी और नमक तनाव के लिए क्रमशः कुल 216, 350, 114 और 249 अनुक्रम प्राप्त हुए, जिनका उपयोग सकारात्मक प्रशिक्षण सेट बनाने के लिए किया गया था। समरूपता में कमी को सकारात्मक स्वतंत्र सेट पर भी लागू किया गया, जिससे कुल परिणाम प्राप्त हुआ

तनाव के प्रत्येक वर्ग के लिए नकारात्मक प्रशिक्षण सेट का निर्माण एकत्रित 376 माइक्रो आरएनए अनुक्रमों से समान मात्रा में अवलोकनों का उपयोग करके किया गया था, जबकि सकारात्मक प्रशिक्षण सेट डेटासेट-I के समान ही रहा। सकारात्मक स्वतंत्र सेट डेटासेट-I के समान ही रहा, और प्रत्येक मामले में शेष नकारात्मक अनुक्रम (नकारात्मक प्रशिक्षण सेट को छोड़कर) का उपयोग नकारात्मक स्वतंत्र परीक्षण सेट बनाने के लिए किया गया। बड़ी संख्या में अवलोकन वाले वर्ग के प्रति पूर्वानुमान पूर्वाग्रह को रोकने के लिए मॉडल को प्रशिक्षित करने के लिए प्रत्येक श्रेणी के लिए एक संतुलित डेटासेट पर विचार किया गया था। तालिका 1 सकारात्मक, नकारात्मक और स्वतंत्र डेटासेट का सारांश देती है।

संख्यात्मक विशेषताओं का निर्माण और सुविधा चयन

चूंकि न्यूक्लियोटाइड्स की छद्म संरचना लंबी दूरी के अनुक्रम क्रम प्रभाव के लिए जिम्मेदार है, इसलिए हमने प्रत्येक माइक्रो आरएनए अनुक्रम को परिवर्तित करने के लिए इस अध्ययन में छद्म के-ट्यूपल न्यूक्लियोटाइड कंपोजिशन (पीएसईकेएनसी) (गुओ एट अल। 2014; चेन एट अल। 2014) सुविधाओं का उपयोग किया। एक संख्यात्मक फीचर वेक्टर. पीएसईकेएनसी डिस्क्रिप्टर का उपयोग कई कम्प्यूटेशनल जीव विज्ञान क्षेत्रों में प्रभावी ढंग से किया गया है, जिसमें न्यूक्लियोसोम प्लेसमेंट की

तालिका 1. वर्तमान अध्ययन में प्रयुक्त सकारात्मक और नकारात्मक डेटासेट का सारांश

डेटासेट	डेटा प्रकार	क्लास	ठंडा	सूखा	ताप	नमक
डेटासेट-I*	प्रशिक्षण सेट	सकारात्मक	216	350	114	249
		नकारात्मक	216	350	114	249
	स्वतंत्र परीक्षण सेट	सकारात्मक	79	149	36	90
		नकारात्मक	79	149	36	90
डेटासेट-II*	परीक्षण सेट	सकारात्मक	216	350	114	249
		नकारात्मक	216	350	114	249
	स्वतंत्र परीक्षण सेट	सकारात्मक	79	149	36	90
		नकारात्मक	160	26	262	127

*तनाव श्रेणी के अनुक्रमों को लेकर नकारात्मक सेट तैयार किए गए माइक्रोबेस के माइक्रो आरएनए अनुक्रम लेकर रुनकारात्मक सेट तैयार किए गए

पूर्वानुमान (गुओ एट अल. 2014), अजैविक तनाव के प्रति उत्तरदायी माइक्रो आरएनए की पूर्वानुमान (मेहर एट अल. 2022) और अन्य (चेन एट अल.) शामिल हैं। . 2015; यांग एट अल. 2018)। पीएसईकेएनसी सुविधाओं को उत्पन्न करने के लिए, सबसे पहले सहसंबंध के स्तर (λ), वजन कारक (w), और केमर आकार (K) की पहचान करना आवश्यक है। चूंकि माइक्रो आरएनए अनुक्रम केवल 20-24 न्यूक्लियोटाइड लंबे होते हैं, इसलिए 2-स्तरों तक के सहसंबंध को ध्यान में रखा गया था। इस कार्य में, 0.2 के डिफॉल्ट वजन कारक ॒ मान का उपयोग किया गया था। 5 अलग-अलग केमर आकारों (K = 1, 2, 3, 4, और 5) का उपयोग करके बनाई गई सुविधाओं की संख्या क्रमशः 7, 19, 67, 259 और 1027 थी। प्रत्येक माइक्रो आरएनए अनुक्रम से कुल 1379 सुविधाएँ प्राप्त हुई। उत्पन्न विशेषताएं प्रकृति में विरल हैं क्योंकि माइक्रो आरएनए लंबाई में छोटे हैं। इसके अलावा, चूंकि डेटासेट छोटा है, इसलिए संभावना है कि बड़ी संख्या में सुविधाओं का उपयोग करने से अतिशयोक्ति हो जाएगी। हालाँकि, फीचर चयन दृष्टिकोण अनावश्यक और अप्रासंगिक सुविधाओं को हटाने, कम्प्यूटेशनल बोझ को कम करने और वर्गीकरण सटीकता को बढ़ाने में सहायता करता है (अक्सू एट अल. 2010; हुआंग एट अल.2014)। इस प्रकार,

एसवीएम-पुनरावर्ती फीचर उन्मूलन का उपयोग करके प्रमुख विशेषताओं को चुना गया था (एसवीएम-आरएफई) विधि (गायोन एट अल. 2002)। पीएसई-इन-वन सॉफ्टवेयर (लियू एट अल. 2015) का उपयोग पीएसईकेएनसी सुविध ाओं को उत्पन्न करने के लिए किया गया था, और " सिगफीचर " आर-पैकेज का उपयोग एसवीएम-आरएफई दृष्टिकोण (दास एट अल. 2020) को लागू करने के लिए किया गया था।

मशीन लर्निंग एल्गोरिदम के साथ पूर्वानुमान

मशीन लर्निंग दृष्टिकोण का उपयोग जैव सूचना विज्ञान के विभिन्न क्षेत्रों में सफलतापूर्वक किया गया है, जैसे जीन खोज और जीनोम एनोटेशन (गुओ एट अल. 2017), प्रोटीन वर्ग पूर्वानुमान (प्रधान एट अल. 2022), जीन अभिव्यक्ति विश्लेषण (अब्बास और ईएल-मंजालावी 2020) , जैविक प्रणालियों में जटिल अंत:क्रिया मॉडलिंग (प्रधान एट अल. 2021), और अन्य। इस अध्ययन में, हमने सात अलग-अलग मशीन लर्निंग तकनीकों का उपयोग किया, जिनमें सपोर्ट वेक्टर मशीन (एसवीएम) (वैपनिक 1963), एक्स्ट्रीम ग्रेडिएंट बूस्टिंग (एक्सजीबी) (चेन और गेस्ट्रिन 2016), रैंडम फॉरेस्ट (आरएफ) (ब्रेइमन 2001), शामिल हैं। लाइट-ग्रेडिएंट बूस्टिंग मेथड (एलजीबीएम) (के एट अल. 2017), बैगिंग (बीएजी) (ब्रेइमन 1996), एडाप्टिव बूस्टिंग (एडीबी) (फ्रायंड और शापिरे 1999) और ग्रेडिएंट ट्री बूस्टिंग (जीबीडीटी) (फ्रीडमैन 2001) . आर-सॉफ्टवेयर को शिक्षण एल्गोरिदम के निष्पादन के लिए लागू किया गया था।

क्रॉस सत्यापन और प्रदर्शन मेट्रिक्स

विभिन्न शिक्षण मॉडलों के प्रदर्शन का आकलन करने के लिए पांच गुना क्रॉस-सत्यापन दृष्टिकोण का उपयोग किया गया था। पांच-गुना क्रॉस-सत्यापन (जियांग और वांग 2017) करने के लिए सकारात्मक और नकारात्मक दोनों डेटासेट को समान आकार के पांच उपसमूहों में यादृच्छिक रूप से अलग किया गया था। क्रॉस-वैलिडेशन के प्रत्येक फोल्ड में, प्रत्येक वर्ग से एक यादृच्छिक रूप से चयनित उपसमुच्चय का उपयोग परीक्षण सेट के रूप में किया गया था, और दोनों वर्गों के शेष चार उपसमूहों को प्रशिक्षण सेट के रूप में काम करने के लिए विलय कर दिया गया था। वर्गीकरण प्रक्रिया के दौरान प्रत्येक तह के लिए अलग-अलग प्रशिक्षण और परीक्षण सेट का उपयोग किया गया था। प्रदर्शन माप प्रदान करने के लिए सभी पांच परीक्षण सेटों में सटीकता का औसत किया गया था। चित्र 1 में, प्रस्तावित दृष्टिकोण के प्रत्येक चरण को दर्शाने वाला एक फ्लोचार्ट प्रस्तुत किया गया है।

पूर्वानुमान मॉडल की प्रभावशीलता का मूल्यांकन करने के लिए, निम्नलिखित मैट्रिक्स का उपयोग किया गया था: सटीकता, रिसीवर ऑपरेटिंग विशेषता वक्र (एयूआरओसी) के तहत क्षेत्र, और सटीक रिकॉल वक्र (एयूपीआरसी) के तहत क्षेत्र।

$$\text{Accuracy} = \frac{1}{2}\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right)$$

$$\text{AUROC} = \int_0^1 \frac{TP}{P}\, d\left(\frac{FP}{N}\right)$$

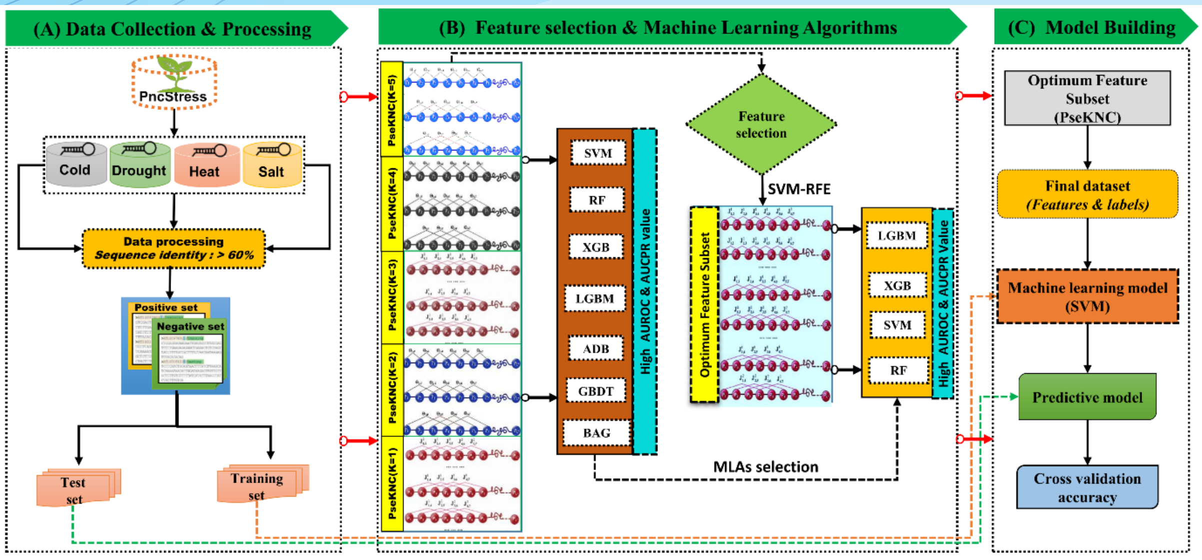
$$\text{AUPRC} = \int_0^1 \frac{TP}{TP+FP}\, d\left(\frac{TP}{P}\right)$$

यहां क्रमश: टीपी, एफपी, टीएन और एफएन, सकारात्मक नमूनों की संख्या का प्रतिनिधित्व करते हैं जिनके सकारात्मक होने की पूर्वानुमान की गई है, नकारात्मक नमूनों की सकारात्मक होने की पूर्वानुमान की गई है, नकारात्मक नमूनों की नकारात्मक होने की पूर्वानुमान की गई है, और सकारात्मक नमूनों के नकारात्मक होने की पूर्वानुमान की गई है।

परिणाम

पीएसईकेएनसी सुविधाओं के साथ एमएलए का प्रदर्शन विश्लेषण

प्रशिक्षण डेटासेट-I का उपयोग करके 5 अलग-अलग पीएसईकेएनसी फीचर सेट के साथ 7 मशीन लर्निंग एल्गोरिदम की पूर्वानुमान सटीकता का मूल्यांकन किया गया था। क्रमश: केमर आकार 4 और 5 के साथ, SVM को ठंड (60.4%) और सूखे (55.02%) के लिए उच्चतम एयूपीआरसी दिखाया गया (चित्र 2a)। जब गर्मी की बात आई, तो केमर आकार 4 के साथ पीएसईकेएनसी फीचर



चित्र 1. प्रस्तावित दृष्टिकोण की संक्षिप्त रूपरेखा का चित्रण। आरेख प्रत्येक अजैविक तनाव के लिए माइक्रो आरएनए पूर्वानुमान मॉडल विकसित करने के लिए अपनाई गई संपूर्ण कम्प्यूटेशनल रणनीतियों के समग्र डिजाइन को दर्शाता है। (ए) पीएनसीस्ट्रेस डेटाबेस से प्रयोगात्मक रूप से मान्य अजैविक प्रतिक्रियाशील माइक्रो आरएनए अनुक्रमों की पुनर्प्राप्ति और अनुक्रम डेटा का प्रसंस्करण; (बी) अनुक्रम-व्युत्पन्न पीएसईकेएनसी सुविधा निर्माण और एयूसी -आरओसी और एयूसी -पीआर पर आधारित सबसे महत्वपूर्ण सुविधाओं और मशीन लर्निंग एल्गोरिदम (एमएलए) का चयन; (सी) चयनित सुविधाओं और क्रॉस-वैलिडेशन सटीकता के मूल्यांकन के साथ मशीन लर्निंग तकनीकों का उपयोग करके मॉडल निर्माण।

सेट का उपयोग किया गया, और एलजीबीएम ने 60.9% एयूपीआरसी की उच्चतम सटीकता प्राप्त की, इसके बाद बीएजी (60.39%) और एक्सजीबी (59.43%) (छवि 2a) प्राप्त की। K=5 के साथ, बीएजी और जीबीडीटी को नमक तनाव के लिए उच्च सटीकता (56% एयूपीआरसी) प्राप्त करने के लिए देखा गया था, हालांकि एक्सजीबी और जीबीडीटी को केमर आकार 4 के लिए नमक तनाव (55% एयूपीआरसी) के लिए लगभग समान सटीकता तक पहुंचने के लिए देखा गया था (चित्र 2a) केमर आकार 4 और 5 के साथ बनाए गए फीचर सेट में अक्सर केमर आकार 1 से 3 के लिए बनाए गए फीचर सेट की तुलना में अधिक पूर्वानुमान सटीकता होती है, जो फीचर सेट के बड़े आकार के कारण हो सकता है।

चयनित सुविधाओं का उपयोग करके प्रशिक्षण सेट-I के लिए पूर्वानुमान विश्लेषण

हमने पाया कि केमर आकार 4 और 5 के साथ उत्पन्न सुविधाओं का उपयोग करने से पूर्वानुमान सटीकता में वृद्धि हुई है (चित्र 2a)। हालाँकि, माइक्रोआरएनए अनुक्रमों (20-24nt) की छोटी लंबाई के कारण सुविधाओं का एक महत्वपूर्ण हिस्सा प्रकृति में विरल है, जो सटीकता में महत्वपूर्ण पूर्वाग्रह पैदा कर सकता है। इसलिए, सभी केमर सुविधाओं को एकीकृत करने के बाद, एसवीएम-आरएफई दृष्टिकोण का उपयोग करके सुविधाओं को रैंक किया गया। यह पता चला कि सटीकता की सर्वोत्तम डिग्री प्राप्त करने के लिए प्रत्येक तनाव श्रेणी के लिए अलग-अलग संख्या में सुविधाओं का चयन किया गया था (छवि 2 बी)। यह भी दिखाया गया कि, जब चयनित सुविधाओं का उपयोग करके विश्लेषण किया गया, तो एसवीएम ने सभी तनाव श्रेणियों (छवि 2 बी) के लिए उच्चतम सटीकता प्राप्त की। एसवीएम और शीर्ष 246 चुनी गई विशेषताओं के साथ, शीत तनाव के लिए 60.01% का उच्चतम एयूपीआरसी प्राप्त किया गया। इसी प्रकार, क्रमशः 230, 310, और 240 सुविधाओं का उपयोग करते हुए, एसवीएम उच्चतम एयूपीआरसी (छवि 2बी) के साथ सूखे (53.54%), गर्मी (78.34%), और नमक (66.78%) तनाव की पूर्वानुमान करने में सक्षम था। पूर्वानुमान के लिए सभी सुविधाओं

का उपयोग करते समय सटीकता में भी गिरावट देखी गई। एकल पीएसईकेएनसी सुविधा सेट के साथ प्राप्त उनकी उच्चतम सटीकता की तुलना में, यानी, गर्मी के लिए केमर आकार 4 और नमक तनाव के लिए 5, चयनित सुविधाओं के साथ सटीकता को क्रमशः ~17% और ~10% तक बढ़ाया गया दिखाया गया था (चित्र 2ए और 2बी)। इसके विपरीत चुने गए फीचर सेट के साथ ठंड और सूखे की सटीकता में वृद्धि नहीं हुई।

स्वतंत्र परीक्षण सेट-I के साथ पूर्वानुमान

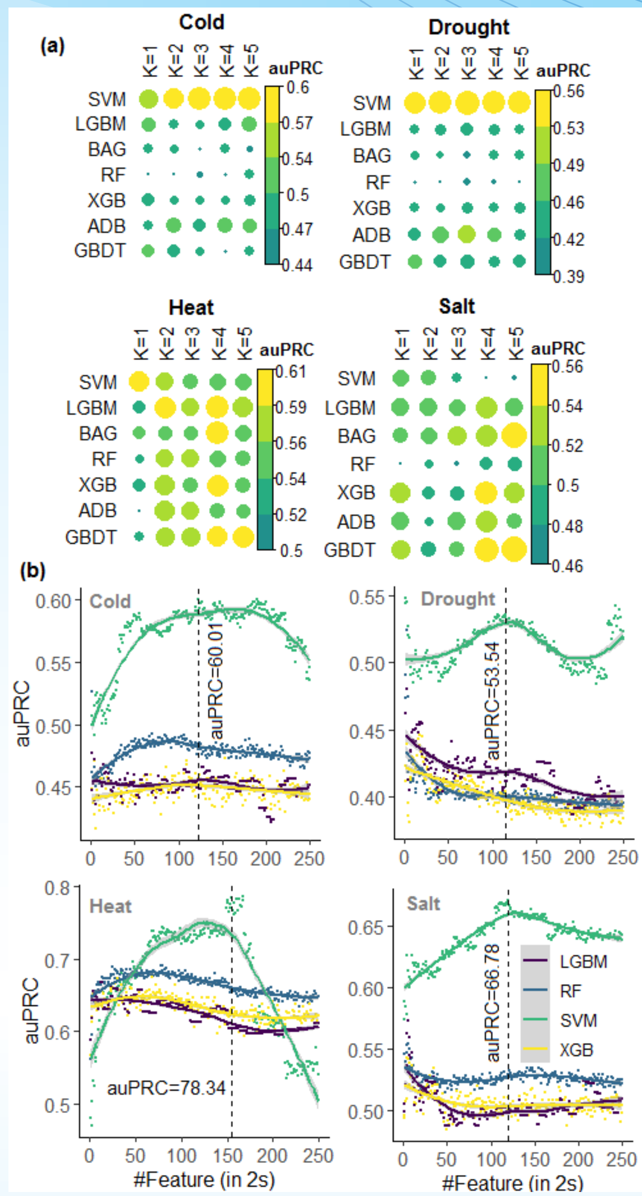
संबंधित प्रशिक्षण सेट-I के साथ प्रशिक्षित मॉडल का उपयोग स्वतंत्र डेटासेट-I की पूर्वानुमान करने के लिए किया गया था। ठंड, सूखा, गर्मी और नमक तनाव के लिए, एयूपीआरसी के संदर्भ में सटीकता क्रमशः 59.63, 66.94, 72.88 और 69.57 प्रतिशत पाई गई (तालिका 2)। क्रॉस-वैलिडेशन सटीकता के समान, उच्चतम पूर्वानुमान सटीकता गर्मी के लिए और सबसे कम ठंड के लिए देखी गई। यह भी देखा गया कि, उनकी संबंधित क्रॉस-वैलिडेशन सटीकता के सापेक्ष, स्वतंत्र डेटासेट की सटीकता सूखे और नमक के लिए बेहतर थी, और ठंड और गर्मी के लिए कम थी (तालिका 2)। ठंड, सूखा, गर्मी और नमक के लिए, समग्र सटीकता क्रमशः 62.02%, 61.40%, 77.78% और 66.67% निर्धारित की गई थी (तालिका 2)।

तालिका 2. स्वतंत्र परीक्षण सेट-I के लिए पूर्वानुमान सटीकता। पूर्वानुमान संबंधित चयनित फीचर सेट के साथ प्रशिक्षण सेट-I के साथ प्रशिक्षित मॉडल का उपयोग करके की गई थी।

तनाव	रुविशेषता*	सटीकता	एयूआरओसी	एयूपीआरसी
सर्दी	246	62.02	64.79	59.63
सूखा	230	61.40	67.72	66.94
ताप	310	77.78	83.25	72.88
नमक	240	66.67	73.20	69.57

*प्रशिक्षण सेट-I का उपयोग करके एसवीएम-आरएफई पद्धति के माध्यम से चयनित सुविधा

प्रशिक्षण सेट-II के साथ पूर्वानुमान विश्लेषण



चित्र 2. (ए) किमी आकार 1 से 5 के लिए पीएसईकेएनसी सुविधाओं के साथ विभिन्न मशीन लर्निंग एल्गोरिदम के लिए एयूपीआरसी के हीट मैप। (बी) एसवीएम-आरएफई विधि के माध्यम से चयनित रैंक सुविधाओं के साथ एयूपीआरसी का प्लॉट। दोनों ही मामलों में पूर्वानुमान विश्लेषण के लिए प्रशिक्षण डेटासेट-I का उपयोग किया गया था।

सुविधाओं के चुने हुए सेट का उपयोग करते हुए, प्रशिक्षण डेटासेट- II पर एसवीएम का उपयोग करके पांच गुना क्रॉस-वैलिडेशन पूर्वानुमान विश्लेषण भी किया गया था। ठंड, सूखा, गर्मी और नमक के लिए, अधिकतम सटीकता प्राप्त करने के लिए सुविधाओं की आदर्श संख्या क्रमशः

380, 272, 174 और 340 थी (तालिका 3)। एयूपीआरसी के संदर्भ में क्रॉस-वैलिडेटेड पूर्वानुमान सटीकता चुने गए फीचर सेट (तालिका 4) के साथ 90.15, 90.09, 87.71 और 89.25% पाई गई। स्वतंत्र परीक्षण सेट II के लिए पूर्वानुमान भी प्रशिक्षण सेट प्प के साथ सीखे गए मॉडल का उपयोग करके की गई थी। कुल मिलाकर, यह पाया गया कि पूर्वानुमान की सटीकता क्रमशः 84.57, 80.62, 80.38, और 82.78% थी (तालिका 4)। जब क्रमशः प्रशिक्षण सेट-I और स्वतंत्र परीक्षण सेट-I की तुलना की गई, तो प्रशिक्षण सेट-II और स्वतंत्र परीक्षण सेट-II की क्रॉस-सत्यापन सटीकता काफी अधिक पाई गई।

तालिका 3. प्रशिक्षण सेट-II और स्वतंत्र परीक्षण सेट-II के लिए प्रदर्शन मेट्रिक्स। पूर्वानुमान के लिए चयनित फीचर सेट वाले एसवीएम का उपयोग किया गया था। स्वतंत्र परीक्षण सेट-II के लिए पूर्वानुमान संबंधित प्रशिक्षण सेट-II के साथ प्रशिक्षित मॉडल का उपयोग करके की गई थी।

डेटा	तनाव	#फीचर*	सटीकता	एयूआरओसी	एयूपीआरसी
प्रशिक्षण सेट-II	कोल्ड	380	83.72	90.31	90.15
	सूखा	272	81.01	87.13	90.09
	ताप	174	81.36	87.52	87.71
	नमक	340	84.28	89.41	89.25
स्वतंत्र टेस्ट सेट- II	कोल्ड	380	84.57	91.67	92.03
	सूखा	272	80.62	88.01	89.67
	ताप	174	80.38	88.15	89.03
	नमक	340	82.75	89.27	88.22

*एसवीएम-आरएफई पद्धति का उपयोग करके चयनित सुविधाओं की संख्या प्रशिक्षण सेट की सटीकता 5-गुना क्रॉस-सत्यापन सटीकता है

व्यक्तिगत पूर्वानुमान विश्लेषण

इसके अतिरिक्त, दो अलग-अलग तनाव सेटों को वर्गीकृत करके बाइनरी वर्गीकरण भी किया गया था। पूर्वानुमान करने के लिए दोनों वर्गों के समान संख्या में अवलोकनों के साथ एक संतुलित डेटासेट का उपयोग किया गया था। सर्दी-सूखा, सर्दी-गर्मी, सर्दी-नमक, सूखा-गर्मी, सूखा-नमक और गर्मी-नमक के वर्गीकरण के लिए, सुविधाओं की इष्टतम संख्या क्रमशः 116, 80, 60, 194, 166, और 440 थी (चित्र 3). ठंड-गर्मी और सूखे-गर्मी की पहचान के लिए, समग्र क्रॉस-सत्यापन सटीकता क्रमशः 60.91% और 60.45% थी (तालिका 4)। शेष चार संयोजनों के लिए वर्गीकरण सटीकता 60% से कम पाई गई। (तालिका 4)। इसके अतिरिक्त, यह देखा गया कि, कुछ उल्लेखनीय अपवादों के साथ, गिरावट शुरू होने से पहले एक निश्चित बिंदु तक सटीकता में वृद्धि हुई (चित्र 3)।

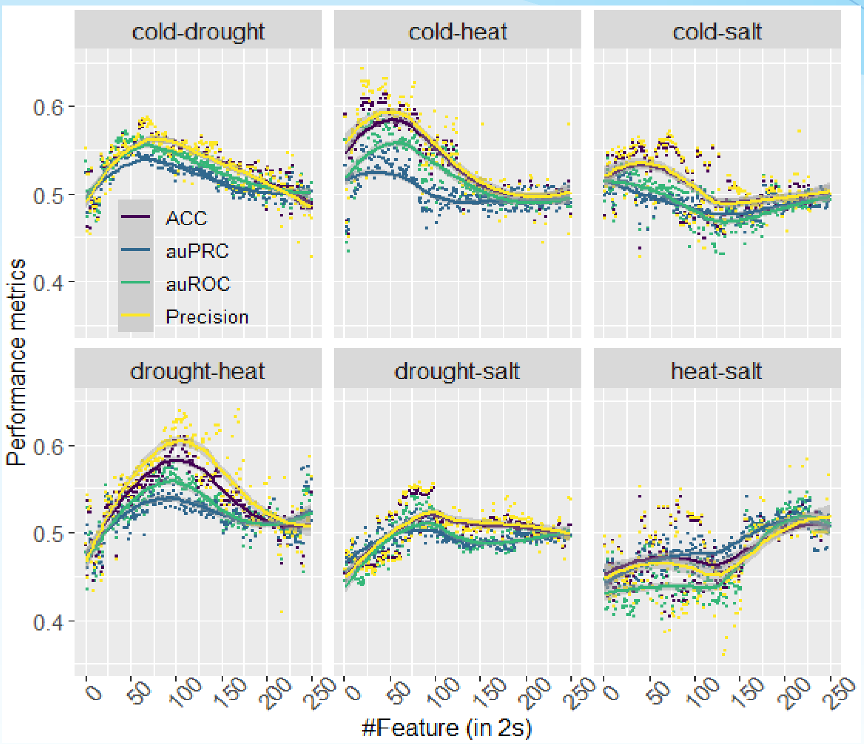
चयनित फीचर सेट में गहन शिक्षण मॉडल का प्रदर्शन विश्लेषण

चार अत्याधुनिक गहन शिक्षण मॉडल का प्रदर्शन, जिसमें एक-आयामी कन्वेन्शनल न्यूरल नेटवर्क (सीएनएन) (किम 2014), अटेंशन-बेस्ड कन्वोल्यूशनल न्यूरल नेटवर्क (एबीसीएनएन) (यिन एट अल। 2016), लॉन्ग शॉर्ट-टर्म मेमोरी (एलएसटीएम) शामिल हैं। (होक्रेइटर और शिमडहुबर 1997) और ऑटो-एनकोडर (एई) (लिउ एट अल 2014) की तुलना एसवीएम से भी की गई थी। पांच-गुना क्रॉस सत्यापन के माध्यम से प्रशिक्षण डेटासेट-II का उपयोग करके पूर्वानुमान विश्लेषण किया गया था,

तालिका 4. ट्रेनिंग सेट-II का उपयोग करके एक-से-एक पूर्वानुमान के लिए प्रदर्शन मेट्रिक्स। पूर्वानुमान के लिए संबंधित चयनित फीचर सेट वाले एसवीएम का उपयोग किया गया था।

सकारात्मक	नकारात्मक	रुविशेषता*	एसीसी	एयूआरओसी	एयूपीआरसी
शीत	सूखा	116	56.74	56.66	54.68
ठंडी	गर्मी	80	60.91	58.85	55.92
ठंडी	नमक	60	55.11	52.39	51.56
सूखा	गर्मी	194	60.45	57.97	55.46
सूखा	नमक	166	55.11	53.28	51.63
गर्मी	नमक	440	53.18	53.38	53.37

*एसवीएम-आरएफई पद्धति का उपयोग करके चयनित सुविधाओं की संख्या



चित्र 3. चयनित सुविधाओं के लिए प्रशिक्षण डेटासेट-II का उपयोग करके एसवीएम मॉडल के लिए प्रदर्शन मेट्रिक्स का प्लॉट।

जहां विश्लेषण के लिए चयनित सुविधाओं की संख्या (तालिका 4) का उपयोग किया गया था। गहन शिक्षण मॉडल में एई ने सभी चार अजैविक तनावों (ठंडा: 77.07%; सूखा: 77.57%; गर्मी: 77.47%; नमक: 79.91%) (तालिका 6) में उच्च सटीकता हासिल की। एबीसीएनएन को गहन शिक्षण मॉडल (तालिका 6) में सबसे कम प्रदर्शन करने वाला पाया गया। एसवीएम को सभी अजैविक तनावों के लिए अजैविक तनाव प्रतिक्रियाशील माइक्रो आरएनए की पूर्वानुमान करने के लिए सभी गहन शिक्षण एल्गोरिदम से बेहतर प्रदर्शन करते हुए देखा गया (तालिका 5)। विशेष रूप से, एसवीएम ने सबसे अच्छा प्रदर्शन करने वाले गहन शिक्षण मॉडल एई की तुलना में 5-6% अधिक सटीकता हासिल की।

पूर्वानुमान उपकरण ऐएसएमआईआर

हमने ठंड, सूखा, गर्मी और नमक में अजैविक तनाव-उत्तरदायी माइक्रो आरएनए की पूर्वानुमान के लिए एक ऑनलाइन पूर्वानुमान सर्वर ऐएसएमआईआर (https://iasri-sg.icar.gov.in/asmir/) की स्थापना की।

तालिका 5. गहन शिक्षण मॉडल के साथ एसवीएम के तुलनात्मक प्रदर्शन मेट्रिक्स। पूर्वानुमान के लिए चयनित फीचर सेट के साथ प्रशिक्षण डेटासेट-II का उपयोग किया गया था।

लर्निंग एल्गोरिदम	मेट्रिक्स	ठंडा	सूखा	गर्मी	नमक
एसवीएम	सटीकता	83.72	81.01	81.36	84.28
	एयूआरओसी	90.31	87.13	87.52	89.41
	एयूपीआरसी	90.15	90.09	87.71	89.25
सीएनएन	सटीकता	50.00	63.29	71.57	62.27
	एयूआरओसी	66.96	68.29	76.82	66.19
	एयूपीआरसी	68.14	70.87	71.66	67.57
एबीसीएनएन	सटीकता	50.00	50.00	73.62	50.00
	एयूआरओसी	54.23	55.19	78.16	45.03
	एयूपीआरसी	74.05	61.58	76.59	57.56
एलएसटीएम	सटीकता	51.16	50.57	72.17	55.03
	एयूआरओसी	61.88	55.95	79.16	65.08
	एयूपीआरसी	60.77	56.91	75.66	66.15
एई	सटीकता	77.07	77.57	77.47	79.91
	एयूआरओसी	82.01	82.31	81.61	83.77
	एयूपीआरसी	80.69	83.50	77.96	85.56

एसवीएम: सपोर्ट वेक्टर मशीन; सीएनएन: कन्वेन्शनल न्यूरल नेटवर्क; एबीसीएनएन: ध्यान-आधारित कन्वेन्शनल न्यूरल नेटवर्क; एलएसटीएम: दीर्घकालिक अल्पकालिक मेमोरी; एई: ऑटो-एनकोडर

सर्वर का फ्रंट एंड एचटीएमएल का उपयोग करके डिजाइन किया गया था, जबकि विकसित आर-कोड पीएचपी की मदद से बैक एंड पर चलाया जाता है। डेटासेट-II का उपयोग करके विकसित एसवीएम मॉडल को सभी चार अजैविक तनावों के लिए इसकी उच्च सटीकता के कारण इस सर्वर में लागू किया गया है। चार प्रकार के अजैविक तनावों का उपयोग करके पूर्वानुमान की जा सकती है। उपयोगकर्ता को माइक्रो आरएनए अनुक्रमों को एफएएसटीए प्रारूप में पेस्ट या अपलोड करना होगा। परिणाम सारणीबद्ध प्रारूप में प्रस्तुत किए जाते हैं, जहां संभावनाएं प्रदान की जाती हैं जिनके साथ प्रत्येक माइक्रो आरएनए अनुक्रम एक विशिष्ट तनाव श्रेणी की पूर्वानुमान करता है।

चर्चा

जलवायु परिवर्तन, जो हाल के वर्षों में तेज हो गया है, अजैविक तनाव का एक प्रमुख कारण है, जिससे सेलुलर होमियोस्टैसिस को नुकसान होता है और पौधों की वृद्धि और विकास पर नकारात्मक प्रभाव पड़ता है (मिकेलबार्ट एट अल 2015)। अजैविक तनाव के कारण पौधों की वृद्धि बाधित होती है क्योंकि पौधों में कोशिका विभाजन और वृद्धि के लिए आदर्श पर्यावरणीय परिस्थितियों का अभाव होता है। उदाहरण के लिए, सूखे का तनाव पौधों की वृद्धि को रोकता है क्योंकि कोशिका स्फीति के लिए पानी की आवश्यकता होती है, जो कोशिका विस्तार को बढ़ावा देता है (सेलेमैन एट अल. 2021); इसी तरह, ठंड का तनाव पौधों की वृद्धि को कम कर देता है क्योंकि एंजाइम और अन्य प्रोटीन गतिविधियां ठंडे तापमान में सीमित होती हैं (संघेरा एट अल. 2011)।

पौधे इन अजैविक तनावों के खिलाफ विभिन्न प्रकार के रक्षा तंत्र विकसित करते हैं, उनमें से अजैविक तनाव-उत्तरदायी जीन की अभिव्यक्ति को नियंत्रित करने के लिए माइक्रो आरएनए का उपयोग करना शामिल है। विभिन्न अजैविक तनावों के जवाब में, जहां जीन अभिव्यक्ति को ट्रांसलेशनल अवरोध द्वारा नियंत्रित किया जाता है, माइक्रो आरएनए अनुक्रम-विशिष्ट तरीके से जीन अभिव्यक्ति के पोस्ट-ट्रांसक्रिप्शनल नियामकों के रूप में कार्य करते हैं (श्रीराम, 2016)। आर्गनॉट प्रोटीन को माइक्रो आरएनए द्वारा बेस-पेयरिंग के माध्यम से विशेष रूप से माइक्रो आरएनए को लक्षित करने के लिए फिर से सुधारा जाता है ताकि इसके अनुवाद और स्थिरता को दबाया जा सके (चिपमैन और पास्क्विनेली 2019; यान एट अल। 2018)। ट्रांसलेशनल दमन की पूरी प्रक्रिया लक्ष्य क्षेत्र के साथ माइक्रो आरएनए की विशिष्ट आधार जोड़ी से शुरू होती है, जहां माइक्रो आरएनए में न्यूक्लियोटाइड का क्रम महत्वपूर्ण होता है। लक्ष्यीकरण विशेष रूप से माइक्रो आरएनए के बीज क्षेत्र की आधार जोड़ी पर निर्भर करता है, जिसमें एमआरएनए के 3'यूटीआर में साइटों के लिए न्यूक्लियोटाइड्स (एनटीएस) 2-7 होते हैं। इसके अतिरिक्त, यह पता चला है कि माइक्रो आरएनए के 3' सिरे लक्ष्य विशिष्टता और विनियमन (यान एट अल.,2018) को नियंत्रित करने में भूमिका निभाते हैं, जहां माइक्रो आरएनए 3' सिरे पर बेस पेयरिंग की डिग्री माइक्रो आरएनए की स्थिरता को प्रभावित कर सकती है। स्वयं (चिपमैन और पास्क्विनेली 2019)। इस प्रकार, पहचानें

पर्यावरणीय तनाव का संबंध है। इस दिशा में, हमने पहली बार पादप माइक्रो आरएनए अनुक्रमों (मेहर एट अल., 2022) से अजैविक तनाव-संबंधी माइक्रो आरएनए की पूर्वानुमान करने के लिए ASR माइक्रो आरएनए (मेहर एट अल., 2022) नामक एक मशीन लर्निंग-आधारित विधि विकसित की है। हालाँकि यह विधि अधिक सामान्यीकृत है और तनाव-विशिष्ट माइक्रो आरएनए की पूर्वानुमान नहीं कर सकती है। विशिष्ट अजैविक तनावों के प्रति पौधों की प्रतिक्रिया में माइक्रो आरएनए के महत्व को देखते हुए, इस अध्ययन में अजैविक तनाव-विशिष्ट (ठंड, सूखा, गर्मी और नमक) माइक्रो आरएनए की पूर्वानुमान के लिए एक मशीन लर्निंग-आधारित कम्प्यूटेशनल मॉडल विकसित करने पर ध्यान केंद्रित किया गया।

उपयुक्त डेटासेट का निर्माण पूर्वानुमानित मॉडल की गुणवत्ता निर्धारित करने वाले प्रमुख कारकों में से एक है और मशीन लर्निंग एल्गोरिदम सीखने की आधारशिला है, जो सीधे मॉडल सटीकता को प्रभावित करता है (शर्मा एट अल।, 2021)। इस अध्ययन में, हमने अजैविक

तनाव-विशिष्ट माइक्रो आरएनए की पूर्वानुमान के लिए मशीन सीखने के तरीकों के मूल्यांकन के लिए डेटासेट-I और डेटासेट-II नामक दो अलग-अलग डेटासेट तैयार किए। डेटासेट-I की तुलना में डेटासेट-II की सटीकता बहुत अधिक देखी गई। सटीकता में सुधार दोनों डेटासेट में विभिन्न नकारात्मक सेटों के उपयोग के कारण हो सकता है। जैसा कि हम जानते हैं कि एक ही माइक्रो आरएनए एक से अधिक अजैविक तनाव से जुड़ा हो सकता है, इसलिए बाकी तनाव श्रेणियों के अवलोकनों का उपयोग करके तैयार किए गए नकारात्मक डेटासेट कम सटीकता उत्पन्न कर सकते हैं। यह संभावित कारण हो सकता है कि एक-से-एक पूर्वानुमान के मामले में पूर्वानुमान की सटीकता कम है। हालाँकि, डेटासेट-II के नकारात्मक सेट का निर्माण माइक्रो बेस (कोजोमारा और ग्रिफिथ-जोन्स 2014) से एकत्र किए गए गैर-अजैविक तनाव माइक्रो आरएनए अनुक्रमों से किया गया था, जो डेटासेट-II के मामले में उच्च भेदभाव सटीकता के संभावित कारणों में से एक हो सकता है।

संख्यात्मक फीचर वैक्टर में माइक्रो आरएनए की एन्कोडिंग आवश्यक है, क्योंकि मशीन लर्निंग एल्गोरिदम केवल संख्यात्मक इनपुट को समायोजित कर सकते हैं (झांग एट अल। 2006; मेहर एट अल। 2018; एसेफपोर 2020)। लक्ष्य पहचान के लिए माइक्रोआरएनए का अनुक्रम क्रम महत्वपूर्ण है। यह पाया गया है कि कुछ स्थिति में उत्परिवर्तन माइक्रो आरएनए को उनके मूल लक्ष्य जीन (भट्टाचार्य और कुई 2017) से बांधने में बाधा डाल सकते हैं। इसलिए, हमने माइक्रो आरएनए में अनुक्रम क्रम को पकड़ने के लिए माइक्रो आरएनए को संख्यात्मक फीचर वैक्टर में एन्कोड करने के लिए छद्म K-ट्यूपल न्यूक्लियोटाइड कंपोजिशनल (पीएसईकेएनसी) सुविधाओं का उपयोग किया। जैविक अनुक्रम डेटा का उपयोग करके पूर्वानुमान के लिए पीएसईकेएनसी का पहले के अध्ययनों (गुओ एट अल. 2014; यांग एट अल. 2018; मेहर एट अल. 2022) में भी सफलतापूर्वक उपयोग किया गया है।

यहां, हमने केमर आकार 1 से 5 पर विचार किया, और कुल 1379 संख्या में सुविधाएँ उत्पन्न हुई। चूंकि माइक्रो आरएनए अनुक्रम केवल 20-24 न्यूक्लियोटाइड लंबे होते हैं, बड़ी संख्या में 0 वाले जेनरेट किए गए फीचर्स की अधिक संभावना होती है, जो फीचर सेट में अतिरेक ला सकता है। दूसरे शब्दों में, क्योंकि सभी सुविधाएँ पीएसईकेएनसी डिस्क्रिप्टर से ली गई हैं, अनावश्यक या अप्रासंगिक सुविधाएँ मौजूद होने पर पूर्वानुमान सटीकता भ्रामक हो सकती है। इसलिए, उत्पन्न सुविधाओं में से महत्वपूर्ण सुविधाओं को चुनना महत्वपूर्ण है। इस अध्ययन में, अजैविक तनाव के लिए विशिष्ट माइक्रो आरएनए की पूर्वानुमान के लिए निर्धारित आदर्श सुविधा को एसवीएम-आरएफई (वांग एट अल 2011) का उपयोग करके चुना गया था। कई अन्य अनुप्रयोगों, जैसे जीनोमिक्स (तांग एट अल. 2008), प्रोटिओमिक्स (दाओ एट अल. 2017), और मेटाबोलॉमिक्स (लिन एट अल. 2012) ने सफलतापूर्वक एसवीएम-आरएफई पद्धति को अपनाया है। अलग-अलग तनाव श्रेणी के लिए चयनित सुविधाओं की संख्या अलग-अलग थी।

हमने अजैविक तनाव-उत्तरदायी माइक्रो आरएनए की पूर्वानुमान के लिए सात अलग-अलग मशीन लर्निंग विधियों जैसे एसवीएम, आरएफ, एक्सजीबी, एडीबी, बीएजी, एलजीबीएम और जीबीडीटी का उपयोग किया। केमर आकार 4 और 5 के साथ उत्पन्न सुविधाओं के साथ पूर्वानुमान की सटीकता आम तौर पर अधिक पाई गई, जो कि केमर आकार 1 से 3 की तुलना में फीचर सेट के बड़े आकार के कारण हो सकती है। लेकिन सभी अजैविक के लिए चयनित फीचर सेट में तनाव के कारण, एसवीएम ने दोनों डेटासेट में अन्य शिक्षण एल्गोरिदम की तुलना में उच्च सटीकता हासिल की। बड़े और शोर वाले डेटा को संभालने की अपनी क्षमता के कारण, एसवीएम को कई कम्प्यूटेशनल अध्ययनों (ब्राउन एट अल. 2000; गुओ एट अल. 2014; चेन एट अल. 2014) में व्यापक रूप से और सफलतापूर्वक लागू किया गया है। एसवीएम के प्रदर्शन की तुलना संबंधित चयनित फीचर सेट के साथ प्रशिक्षण

डेटासेट- II का उपयोग करके गहन शिक्षण एल्गोरिदम के चार प्रकारों, जैसे सीएनएन, एबीसीएनएन, एलएसटीएम और एई के साथ की गई थी। एसवीएम ने सभी चार गहन शिक्षण एल्गोरिदम से बेहतर प्रदर्शन किया। उथले और गहन शिक्षण मॉडल के लिए पूर्वानुमान की कम सटीकता एसवीएम-आरएफई का उपयोग करके चयनित सुविधाओं के कारण हो सकती है जो अन्य गहन शिक्षण विधियों के साथ उच्च सटीकता प्राप्त करने के लिए उपयुक्त नहीं हो सकती हैं।

निष्कर्ष

प्रस्तावित उपकरण एएसएमआईआर (https://iasri-sg.icar.gov.in/asmir/) माइक्रो आरएनए अनुक्रमों से प्राप्त सुविधाओं का उपयोग करके अजैविक तनाव-विशिष्ट (ठंडा, सूखा, गर्मी और नमक) माइक्रो आरएनए की पूर्वानुमान के लिए एक वैकल्पिक दृष्टिकोण प्रदान करता है। उत्साहजनक परिणामों के कारण, एएसएमआईआर का उपयोग केवल अनुक्रम जानकारी का उपयोग करके अजैविक तनाव विशिष्ट माइक्रो आरएनए की बड़े पैमाने पर पूर्वानुमान के लिए प्रभावी ढंग से किया जा सकता है। अजैविक तनावों के प्रति पौधों की प्रतिक्रिया में माइक्रो आरएनए के महत्व और कम्प्यूटेशनल तरीकों की कमी को देखते हुए यह अनुमान लगाया गया है कि प्रस्तावित दृष्टिकोण अजैविक तनाव विशिष्ट माइक्रो आरएनए की पूर्वानुमान के लिए मौजूदा प्रयोगात्मक तकनीकों का पूरक होगा।

स्वीकृतिया

लेखक अनुसंधान कार्य को पूरा करने के लिए आवश्यक सुविधाएं प्रदान करने के लिए निदेशक, भा.कृ.अनु.प. - भा. कृ. सा. अ. सं., नई दिल्ली को ईमानदारी से स्वीकार करते हैं। लेखक भा.कृ.अनु.प. - भा. कृ. सा. अ. सं., नई दिल्ली में उपलब्ध अशोक सुपरकंप्यूटिंग सुविधाओं को भी स्वीकार करते हैं।

संदर्भ

अब्बास एम, ईएल-मंजालावी वाई (2020) मशीन लर्निंग बैस्ड रिफाइन्ड डिफरेन्शल जीन इक्स्प्रेशन अनैलिसिस ऑफ पीडीऐट्रिक सेप्सिस। बीएमसी मेड जीनोमिक्स 13:122। https://doi.org/10.1186/s12920-020-00771-4।

अक्पिनर बीए, लुकास एसजे, बुडक एच (2013) जीनोमिक्स अप्रोचिज फॉर क्राप इम्प्रूव्मन्ट अगेन्स्ट एबाइआटिक स्ट्रेस। साइंटिफिक वर्ल्ड जर्नल 2013:361921। https://doi.org/10.1155/2013/361921।

अक्सू वाई, मिलर डीजे, केसिडिस जी, यांग क्यूएक्स (2010) मार्जिन-मैक्सिमाइजिंग फीचर एलिमिनेशन मेथड्स फॉर लीनियर ऐन्ड नॉनलीनियर कर्नेल- बैस्ड डिस्क्रिमनन्ट फंगक्शन । आईईईई ट्रैन्जैक्शन्ज ऑन न्युरल नेटवर्क 21:701-717। https://doi.org/10.1109/TNN.2010.2041069।

अल्फारो, ई., गेमेज, एम. गार्सिया, एन. (2013)। एडबैग: एन आर पैकेज फॉर क्लैसफकेशन विद बूस्टिंग ऐन्ड बैगिंग। जर्नल ऑफ स्टैटिस्टिकल सॉफ्टवेयर, 54(2), 1-35। http://www.jstatsoft.org/v54/i02/।

एन जे, लाई जे, सज्जनहार ए, एट अल (2014) एमआईआरप्लांट: एन इन्टग्रेटिड टूल फॉर आइडेन्टफकेशन ऑफ प्लांट एमआईआरएनए फ्राम आरएनए सीक्वन्सिंग डेटा । बीएमसी बायोइनफॉर्मेटिक्स 15:275। https://doi.org/10.1186/1471-2105-15-275।

अनवर ए, किम जे-के (2020) ट्रांसजेनिक ब्रीडिंग अप्रोचिज फॉर इम्प्रूविंग एबाइआटिक स्ट्रेस टालरेन्स: रीसेन्ट प्रोग्रेस ऐन्ड फ्यूचर पर्स्पेक्टिव। इंट जे मोल साइअन्स 21:2695। https://doi.org/10.3390/परडे21082695।

एसेफपुर वैकिलियन के (2020) मशीन लर्निंग इम्प्रूवस आउर नालिज अबाउट माइक्रोआरएनए फंगक्शन टवॉर्ड्ज प्लैन्ट एबाइआटिक स्ट्रेस । साइअन्स रेप्रीजेटटिव 10:3041। https://doi.org/10.1038/s41598-020-59981-6।

बेगम वाई (2022) रेग्यलटॉरी रोल ऑफ माइक्रोआरएनए (एमआईआरएनए) इन द रीसन्ट डिवेलप्मन्ट ऑफ एबाइआटिक स्ट्रेस टालरेन्स ऑफ प्लैन्ट। जीन 821:146283. https://doi.org/10.1016/j.gene।2022.146283।

भट्टाचार्य ए, कुई वाई (2017) सिस्टमैटिक प्रीडिक्शन ऑफ द इम्पैक्टस ऑफ म्यूटेशन इन माइक्रोआरएनए सीड सीक्वन्स। जे इंटीग्र बायोइनफॉर्म 14:20 170001, /j/jib.2017.14.issue-1/jib-2017-0001/ jib-2017–0001.xml. https://doi.org/10.1515/ jib-2017-0001।

बॉयर जेएस (1982) प्लैन्ट प्रोडक्टिविटी ऐन्ड इन्वाइरन्मन्ट। साइअन्स 218:443-448. https://doi.org/10.1126/ science.218.4571.443।

ब्रिमन एल (1996) बैगिंग प्रेडिक्टर्स। मैक लर्न 24:123-140। https://doi.org/10.1007/BF00058655।

ब्रिमन एल (2001) रैंडम फॉरेस्ट। मशीन लर्निंग 45:5-32. https://doi.org/10.1023/।:1010933404324।

ब्राउन एमपी, ग्रुंडी डब्ल्यूएन, लिन डी, एट अल (2000) नालिज बैस्ड अनैलिसिस ऑफ माइक्रोएरे जीन इक्स्प्रेशन डेटा बाइ यूजिंग सपोर्ट वेक्टर मशीन । प्रोक नेटल एकेड साइंस यू एस ए 97:262-267। https://doi.org/10.1073/pnas.97.1.262।

बुडक एच, कांतार एम, बुलुट आर, अक्पिनर बीए (2015) स्ट्रेस रिस्पान्सिव माइक्रोआरएनए ऐन्ड आइसोमाइक्रोएरे इन सीरियल्स प्लैन्ट साइअन्स 235:1-13. https://doi.org/10.1016/j.plansci.2015.02.008।

चेन टी, गेस्ट्रिन सी (2016) एक्सजीबूस्ट: ए स्केलेबल ट्री बूस्टिंग सिस्टम। इन: प्रोसीडिंगज ऑफ द 22 एसीएम एसआईजीकेडीडी इन्टर्नैशनल कान्फरन्स आन नॉलेज डिस्कवरी ऐन्ड डेटा माइनिंग। कंप्यूटिंग मशीनरी एसोसिएशन, न्यूयॉर्क, एनवाई, यूएसए, पीपी 785-794।

चेन डब्ल्यू, लेई टी-वाई, जिन डी-सी, एट अल (2014) पीएसईकेएनसी: ए फ्लेक्सबल वेब सर्वर फॉर जेनरैटिंग सूडो के-ट्यूपल न्यूक्लियोटाइड काम्पजिशन। ऐनल बायोकेम 456:53-60। https://doi.org/10.1016/j.ab.2014.04.001।

चेन डब्ल्यू, लिन एच, चाउ के-सी (2015) सूडो के-ट्यूपल न्यूक्लियोटाइड काम्पजिशन ऑर पीएसईकेएनसी: एन इफेक्टिव फॉर्म्यूलैशन फॉर ऐनलाइज जीनोमिक सीक्वन्स। मोल बायोसिस्ट 11:2620-2634। https://doi.org/10.1039/C5MB00155B।

चिन्नुसामी वी, झू जे, झोउ टी, झू जे-के (2007) स्मॉल आरएनए: बिग रोल इन एबाइआटिक स्ट्रेस टालरेन्स ऑफ प्लैन्ट। इन: जेन्क्स एमए, हसेगावा पीएम, जैन एसएम (संस्करण) अड्वैन्स इन मलेक्यलर ब्रीडिंग टोवर्ड ड्राउट ऐन्ड साल्ट टालरन्ट क्रापस । स्प्रिंगर नीदरलैंड्स, डॉर्ड्रेक्ट, पीपी 223-260।

चिपमैन एलबी, पास्क्विनेली एई (2019) माइक्रोआरएनए टार्गिट: ग्रोइंग बिआन्ड द सीड । ट्रेन्ड जेनेट 35:215-222। https://doi.org/10.1016/j.tig.2018.12.005।

दाओ एफ-वाई, यांग एच, सु जेड-डी, एट अल (2017) रीसेन्ट ऐड्वैन्स इन कोनोटॉक्सिन क्लैसफकेशन बाइ यूजिंग मशीन लर्निंग मेथड्ज । मोलेक्यूल 22:1057. https://doi.org/10.3390/ मोलेक्यूल 22071057।

दास पी, रॉयचौधरी ए, दास एस, एट अल (2020) सिगफीचर: नावल सिग्निफकन्ट फीचर सिलेक्शन मेथड्ज फॉर क्लैसफकेशन ऑफ जीन इक्स्प्रेशन डेटा यूजिंग सपोर्ट वेक्टर मशीन ऐन्ड टी स्टेटिस्टिक । फ्रंटियर्स इन जेनेटिक्स 11:247। https://doi.org/10.3389/fgene.2020.00247।

एल्डेम वी, अक्के यूल, ओझुनेर ई, एट अल (2012) जीनोम-वाइड आइडेन्टफकेशन ऑफ माइक्रोआरएनए रिस्पान्सिव टू ड्राउट इन पीच बाइ हाई-थ्रूपुट डीप सीक्वेंसिंग (प्रूनस पर्सिका) । प्लोस वन 7:म50298. https://doi.org/10.1371/journal.pone.0050298।

फ्रायंड वाई, शापिरे आरई (1999) ए शॉर्ट इन्ट्रोडक्शन टू बूस्टिंग। जर्नल ऑफ जापानी सोसायटी फॉर आर्टिफिशियल इंटेलिजेंस। 14(5):771-780.।

फ्रीडमैन जेएच (2001) ग्रीडी फंक्शन अप्राक्समेशन: ए ग्रेडिएंट बूस्टिंग मशीन। द ऐनल्ज ऑफ स्टटिस्टिक्स 29:1189-1232। https://doi.org/10.1214/aos/1013203451।

गाओ एस, यांग एल, जेंग एच क्यू, एट अल (2016) ए काटन माइक्रोआरएनए इज इन्वाल्व्ड इन रेग्युलेशन ऑफ प्लैन्ट रिस्पान्स टू साल्ट स्ट्रेस। साइअन्स रेप 6:19736। https://doi.org/10.1038/srep19736।

गुओ एफ-बी, डोंग सी, हुआ एच-एल, एट अल (2017) ऐक्यरिट प्रीडिक्शन ऑफ ह्यूमन इसेन्शल जीन यूजिंग ओन्ली न्यूक्लियोटाइड काम्पजिशन ऐन्ड एसोसिएशन इन्फर्मेशन। बायोइन्फर्मेटिक्स 33:1758-1764।

गुओ एस-एच, डेंग ई-जेड, जू एल-क्यू, एट अल (2014) आईएनयूसी-पीएसईकेएनसी: ए सीक्वन्स बैस्ट प्रिडिक्टर फॉर प्रिडिक्ट न्यूक्लियोसोम इन जीनोम विद काम्पजिशन सूडो के-ट्यूपल न्यूक्लियोटाइड । बायोइन्फर्मेटिक्स 30:1522–1529. https://doi.org/10.1093/bioinformatics/btu083।

गयोन I, वेस्टन जे, बार्नहिल एस, वापनिक वी (2002) जीन सिलेक्शन फॉर कैन्सर क्लैसफकेशन यूजिंग सपोर्ट वेक्टर मशीन। मशीन लर्निंग 46:389–422. https://doi.org/10.1023/।:1012487302797।

होक्रेइटर एस, शिमधुबर जे (1997) लॉंग शॉर्टख्रटर्म मेमरी। न्युरल कम्प्यूट 9:1735ख्र1780 https://doi.org/10.1162/neco.1997.9.8.1735।

हुआंग एम-एल, हंग वाई-एच, ली डब्लूएम, एट अल (2014) एसवीएम-आरएफई बैस्ट फीचर सिलेक्शन ऐन्ड टैगुची पैरामीटर्स ऑप्टिमाइजेशन फॉर मल्टीक्लास एसवीएम क्लासिफायरियर। साइंटिफिकवर्ल्डजर्नल2014:795624. https://doi.org/10.1155/2014/795624।

हुआंग वाई, नीयू बी, गाओ वाई, एट अल (2010) सीडी-हिट सुइट: ए वेब सर्वर फॉर क्लस्टरिंग ऐन्ड कम्पेरिंग बाइअलाजिकल सीक्वन्स । बायोइन्फर्मेटिक्स 26:680–682।https://doi.org/10.1093/bioinformatics/btq003।

इवाकावा एच, तोमारी वाई (2013) मोलेक्यलर इन्साइट माइक्रोआरएनए- मीडीएट ट्रांसलेशन रिप्रेशन इन प्लैन्ट । मोलेक्यलर सेल 52:591–601. https://doi.org/10.1016/j.molcel.2013.10.033।

जिया एक्स, वांग डब्ल्यू-एक्स, रेन एल, एट अल (2009) डिफरेन्शल ऐन्ड डाइनैमिक रेग्युलेशन ऑफ माइक्रोआर 398 इन रिस्पान्स टू एबीए ऐन्ड साल्ट स्ट्रेस इन पॉपुलस्ट्रेमुला ऐन्ड अरेबिडोप्सिस्टलियाना। प्लांट मोल बायोल 71:51–59. https://doi.org/10.1007/s11103-009-9508-8।

जियांग जी, वांग डब्ल्यू (2017) एरर एस्टमेशन बैस्ट आन वेरीअन्स अनैलिसिस ऑफ के-फोल्ड क्रॉस वैलिडेशन। पैटर्न रेकग्निशन 69:94–106. https://doi.org/10.1016/j.patcog.2017.03.025।

के जी, मेंग क्यू, फिनले टी, एट अल (2017) लाइटजीबीएम: ए हाइली इफिशन्ट ग्रेडिएंट बूस्टिंग डिसीजन ट्री। इन प्रोसीडिंगज ऑफ द 31 इन्टर्नैशनल कान्फरन्स आन न्यूरल इन्फर्मेशन प्रासेसिंग सिस्टम। कुरेन एसोसिएट्स इंक., रेड हुक, एनवाई, यूएसए, पीपी 3149–3157।

किम वाई (2014) कन्वेन्शनल न्यूरल नेटवर्क्स फॉर सेन्टन्स क्लैसफकेशन। इन प्रोसीडिंगज ऑफ द 2014 कान्फरन्स एम्पिरिकल मेथड्ज इन नैचरल लैंगग्विज प्रासेसिंग (ईएमएनएलपी) । एसोसिएशन फॉर कम्प्यूटेशनल लिंगग्विस्टिक्स, दोहा, कतर, पीपी 1746–1751।

कोजोमारा ए, ग्रिफिथ-जोन्स एस (2014) माइक्रोबेस: एनोटेट हाइ कान्फडन्स माइक्रोआरएनए यूजिंग डीप सीक्वेंसिंग डेटा करना। न्यूक्लिक एसिड रेस 42 डी68-73। https://doi.org/10.1093/nar/gkt1181।

कू वाई-एस, वोंग जेडब्ल्यू-एच, मुई जेड, एट अल (2015) स्मॉल आरएनए इन प्लैन्ट रिस्पान्स टू एबाइआटिक स्ट्रेस रेग्यलटॉरी रोल्स ऐन्ड स्टडी मेथड्ज । इंट जे मोल साइअन्स 16:24532–24554. https://doi.org/10.3390/ijms161024532।

ली डब्ल्यू-एक्स, ओनो वाई, झू जे, एट अल (2008) द अरेबिडोप्सिस एनएफवाईए5 ट्रांसक्रिप्शन फैक्टर इज रेग्यलैटिड ट्रांसक्रिप्शनल ऐन्ड पोस्ट ट्रांसक्रिप्शनल टू प्रमोट ड्राउट रिजिस्टन्स । द प्लैन्ट सेल 20:2238ख्र 2251. https://doi.org/10.1105/tpc.108.059444।

लियाव ए, वीनर एम (2002)। क्लैसफकेशन ऐन्ड रिग्रेशन बाइ रैंडमफॉरेस्ट । आर न्यूज 2(3), 18-22।

लिउ सी-वाई, चेंग डब्ल्यू-सी, लिउ जे-डब्ल्यू, लिउ डी-आर (2014) ऑटोएनकोडर फॉर वर्ड्ज । न्यूरोकंप्यूट 139:84-96।. https://doi.org/10.1016/j.neucom.2013.09.055।

लिन एक्स, यांग एफ, झोउ एल, एट अल (2012) ए सपोर्ट वेक्टर मशीन रिकर्सिव फीचर इलिमनेशन फीचर सिलेक्शन मेथड बैस्ट आन आर्टफिशल कंट्रास्ट वेरीअबल्ज म्यूचूअल इन्फॉर्मैशन । जे क्रोमैटोग्र बी एनालिट टेक्नोल बायोमेड लाइफ साइंस 910:149–155. https://doi.org/10.1016/j.jchromb.2012.05.020।

लियू बी, लियू एफ, वांग एक्स, एट अल (2015) पीएसई-इन-वन वेब सर्वर फॉर जेनरैटिंग वेरीअस माड ऑफ सूडो कम्पोनन्ट्स ऑफ डीएनए, आरएनए ऐन्ड प्रोटीन सीक्वन्स । न्यूक्लिक एसिड रेस 43:W65–W71. https://doi.org/10.1093/nar/gkv458।

लियू एच-एच, तियान एक्स, ली वाई-जे, एट अल (2008) माइक्रोएरे- बैस्ट अनैलिसिस ऑफ स्ट्रेस-रेग्यलैटड माइक्रोआरएनए इन एराबिडोप्सिस थालियाना। आरएनए 14:836–843. https://doi.org/10.1261/rna.895308।

मेहर पीके, बेगम एस, साहू टीके, एट अल (2022) एएसआरएमआईआरएनए: एबाइआटिक स्ट्रेस रिस्पान्सिव माइक्रोआरएनए प्रीडिक्शन इन प्लैन्ट बाय यूजिंग मशीन लर्निंग एल्गोरिदम विद सूडो के-टुपल न्यूक्लियोटाइड काम्पजिशनल फीचर्ज।इंट जे मोल साइअन्स 23:1612. https://doi.org/10.3390/ijms23031612।

मेहर पीके, साहू टीके, मोहंती जे, एट अल (2018) एनआईएफप्रेड: प्रोटीम वाइड आइडेन्टफकेशन ऐन्ड कैटगरिजैशन ऑफ नाइट्रोजन-फिक्सेशन प्रोटीन ऑफ डायजट्रॉफ बैस्ट आन काम्पजिशन ट्रैन्जिशन डिस्ट्रब्यूशन फीचर्ज यूजिंग सपोर्ट वेक्टर मशीन। फ्रंट माइक्रोबायोल9:1100. https://doi.org/10.3389/fmicb.2018.01100।

मिकेलबार्ट एमवी, हसेगावा पीएम, बेली-सेरेस जे (2015) जनेटिक मेकनिजम्ज ऑफ एबाइआटिक स्ट्रेस टालरन्स देट ट्रैन्स्लैट टू क्राप यील्ड स्टबिलिटी। नेट रेव जेनेट16:237–251. https://doi.org/10.1038/nrg3901।

मोचिदा के, शिनोजाकी के (2013) अन्लाकिग ट्रिटिसिया जीनोमिक्स टू सस्टैनबली फीड द फीचर। प्लांट सेल फिजियोल 54:1931–1950. https://doi.org/10.1093/pcp/pct163।

नोमान ए, अकील एम (2017) । माइक्रोआरएनए-बैस्ट हेवी मेटल होमियोस्टैसिस ऐन्ड प्लांट ग्रोथ। एनवायरन साइंस पोलुट रेस इंट 24:10068–10082. https://doi.org/10.1007/11356-017-8593-5।

प्रधान यूके, मेहर पीके, नाहा एस, एट अल (2022) पीएलडीबीरेड: ए नावल कम्प्यूटेशनल मॉडल फॉर डिस्कवरी ऑफ डीएनए बाइंडिंग प्रोटीन इन प्लांट । ब्रीफिंग इन बायोइन्फर्मेटिक्स bbac483. https://doi.org/10.1093/इपइ/इइंब483।

प्रधान यूके, शर्मा एनके, कुमार पी, एट अल (2021) माइक्रोबायोम: मशीन-लर्निंग आन बायेसियन कॉजल नेट ऑफ आरबीपी- माइक्रोआरएनए इंटरैक्शन सक्सेस्फली प्रिडिक्ट माइक्रोआरएनए प्रोफाइल। प्लस वन 16:e0258550. https://doi.org/10.1371/journal.pone.0258550।

री एस, चाए एच, किम एस (2015) प्लांटमिरनाटी: माइक्रोआरएनए ऐन्ड एमआरएनए इन्टग्रेटिड अनैलिसिस, फुली यूटलाइज केरक्टरिस्टिक्स ऑफ प्लांट सीक्वन्सिंग डेटा मेथड्ज । 83:80–87. https://doi.org/10.1016/j.ymeth.2015.04.003।

संघेरा जीएस, वानी एसएच, हुसैन डब्ल्यू, सिंह एनबी (2011) एन्जनिरिंग कोल्ड स्ट्रेस टालरन्स इन क्राप प्लांट । करन्ट जीनोमिक्स 12:30-43। https://doi.org/10.2174/138920211794520178।

सेलेमन एमएफ, अल-सुहैबानी एन, अली एन, एट अल (2021) ड्राउट स्ट्रेस इम्पैक्ट आन प्लांट ऐन्ड डिफरन्ट अप्रोचिज टू अलीवीऐट इट्स ऐड्वर्स इफेक्ट्स। प्लांट (बेसल) 10:259। https://doi.org/10.3390/plans10020259।

श्रीराम वी, कुमार वी, देवरुमठ आरएम, एट अल (2016) माइक्रोआरएनए ऐज पटेन्शल टार्गट फॉर एबाइआटिक स्ट्रेस टालरन्स इन प्लांट। फ्रन्टीर इन प्लांट साइअन्स 7:817। https://doi.org/10.3389/fpls.2016.00817।

सुंकर आर, कपूर ए, झू जे-के (2006) पोस्टट्रांसक्रिप्शनल

इंडक्शन ऑफ टू ब्न/द सुपरऑक्साइड डिसम्यूटेज जीन इन एराबिडोप्सिस मीडीएट बाइ डाउनरेगुलेशन ऑफ माइक्रोआर398 ऐन्ड इम्पॉर्टन्ट फॉर ऑक्सीडेटिव स्ट्रेस टालरन्स । प्लांट सेल 18:2051-2065। https://doi.org/10.1105/tpc.106.041673।

स्जेस्निएक मेगावाट, डेओरोविक्ज एस, गैपस्की जे, एट अल (2012) माइक्रोआरनेस्ट डेटाबेस: एन इन्टग्रेविट अप्रोच इन माइक्रोआरएनए सर्च ऐन्ड एनोटेशन। न्यूक्लिक एसिड रेस 40:D198-204. https://doi.org/10.1093/nar/gkr1159।

टैंग वाई, झांग वाई-क्यू, हुआंग जेड (2007) डिवेलप्मन्ट ऑफ टू स्टेज एसवीएम-आरएफई जीन सिलेक्शन स्ट्रैटिजी फॉर माइक्रोएरे इक्स्प्रेशन डेटा अनैलिसिस। आईईईई/एसीएम ट्रांस कंप्यूट बायोल बायोइनफॉर्म 4:365-381। https://doi.org/10.1109/TCBB.2007.70224।

ट्रिनडेड आई, कैपिटो सी, डाल्मे टी, एट अल (2010) माइक्रोआर398 ऐन्ड माइक्रोआर408 इन रिस्पान्स आर अप रेग्यलैटिड टू वाटर डेफिसिट इन मेडिकैगो ट्रंकैटुला। प्लांटा 231:705-716. https://doi.org/10.1007/s00425-009-1078-0।

त्रिपाठी ए, गोस्वामी के, सानन-मिश्रा एन (2015) रोल ऑफ बायोइन्फर्मेटिक्स इन ईस्टैब्लिश माइक्रोआरएनए एज माजलेटर ऑफ एबाइआटिक स्ट्रेस रिस्पान्स: न्यू रिवोल्यूशन । फ्रंटियर्स इन फिजियोलॉजी 6:286। https://doi.org/10.3389/fphys.2015.00286।

वाप्निक वी (1963) पैटर्न रेकग्निशन यूजिगं जेनरलाइज्ड पोर्ट्रेट मेथड । ऑटमैशन ऐन्ड रिमोट कंट्रोल 24:774-780।

विज एस, त्यागी एके (2007) इमर्जिंग ट्रेन्ड इन द फंगक्शनल जीनोमिक्स ऑफ द एबाइआटिक स्ट्रेस रिस्पान्स इन क्राप प्लांट । प्लांट बायोटेक्नोल जे 5:361-380। https://doi.org/10.1111/र.1467-7652.2007.00239.x।

वांग बी, सन वाई-एफ, सॉन्ग एन, एट अल (2014) माइक्रोआरएनए इन्वाल्व इन कोल्ड, वून्डिंग ऐन्ड साल्ट स्ट्रेस इन ट्रिटिकम एस्टिवम। प्लांट फिजियोल बायोकेम 80:90-96। https://doi.org/10.1016/j.plaphy.2014.03.020।

वांग जे, मेई जे, रेन जी (2019) प्लांट माइक्रोआरएनए: बायोजेनेसिस, होमोस्टेसिस ऐन्ड डिग्रेडेशन। फ्रंटियर्स इन प्लांट साइअन्स 10:360। https://doi.org/10.3389/fpls.2019.00360।

वांग जे, शान जी, डुआन एक्स, वेन बी (2011) इम्प्रूवड एसवीएम-आरएफई फीचर सिलेक्शन मेथड फॉर मल्टी-एसवीएम क्लासिफायरियर। इन: 2011 इलेक्ट्रिकल ऐन्ड कन्ट्रोल इंजीनियरिंग कान्फरन्स । पीपी 1592-1595।

वांग एम, वांग क्यू, झांग बी (2013) रिस्पान्स ऑफ माइक्रोआरएनए ऐन्ड देयर टार्गट टू साल्ट ऐन्ड ड्राउट स्ट्रेस इन काटन (गॉसिपियम हिर्सुटम एल.)। जीन 530:26-32. https://doi.org/10.1016/j.gene.2013.08.009।

विंटर जे, डिडेरिच एस (2011) अर्गोनॉट प्रोटीन रेग्यलेट माइक्रोआरएनए स्टबिलिटी: इन्क्रीस्ट अबन्डन्स माइक्रोआरएनए बाय अर्गोनॉट प्रोटीन इज ड्यू टू माइक्रोआरएनए स्टैबलजैशन । आरएनए बायोल 8:1149-1157। https://doi.org/10.4161/rna.8.6.17665।

वू डब्ल्यू, वू वाई, हू डी, एट अल (2020) पीएनसीस्ट्रेस: मैन्युअली क्यूरेटेड डेटाबेस ऑफ इक्स्पेरमेन्टली वैलडैटड स्ट्रेस रिस्पान्सिव नान-कोडिंग आरएनए इन प्लांट। डेटाबेस 2020:baaa001. https://doi.org/10.1093/database/baaa001।

झी एफ, वांग क्यू, सन आर, झांग बी (2015) डीप सीक्वन्सिंग रिवील इम्पॉर्टन्ट रोल ऑफ माइक्रोआरएनए इन रिस्पान्स टू ड्राउट ऐन्ड सलिनिटी स्ट्रेस इन काटन। जे एक्सप बॉट 66:789-804। https://doi.org/10.1093/jxb/eru437।

जू क्यू, हे क्यू, ली एस, तियान जेड (2014) मोलेक्यलैर कैरिक्टरिजेशन ऑफ StNAC2 इन पटेटो ऐन्ड इट्स ओवरएक्सप्रेसन कन्फर ड्राउट ऐन्ड साल्ट टालरन्स । एक्टा फिजियोल प्लांट 36:1841-1851। https://doi.org/10.1007/s11738-014-1558-0।

यान वाई, एसेवेडो एम, मिग्नाका एल, एट अल (2018) द सीक्वन्स फीचर्ज देट डिफाइन इफिशन्ट ऐन्ड स्पसिफिक hAGO2- डिपेन्डन्ट माइक्रोआरएनए साइलेंसिंग गाइड । न्यूक्लिक एसिड रेस 46:8181-8196। https://doi.org/10.1093/nar/gky546।

यांग एच, किउ डब्ल्यू-आर, लियू जी, एट अल (2018) आइर स्पॉट-पीएसई6एनसी: आइडेन्टफाइइंग रीकाम्बनेशन स्पॉट इन सैक्रोमाइसेस सेरेविसिया स्पॉट बाय इन्कॉर्परिटगं हेक्सामर काम्पजिशन इनटू जेनरल पीएसईकेएनसी । इंटरनेशनल जर्नल ऑफ बायोलॉजिकल साइंसेज 14:883-891। https://doi.org/10.7150/ijbs.24616।

यिन डब्ल्यू, एबर्ट एस, शुट्ज एच (2016) अटेन्शन-बैस्ट कन्वेन्शनल न्यूरल नेटवर्क फॉर मशीन कॉम्प्रिहेंशन। इन: प्रोसीडिंगज ऑफ द वर्क्शाप आन मानव-कंप्यूटर क्वेस्चन ऐन्सरिंग। एसोसिएशन फॉर कम्प्यूटेशनल लिंगग्विस्टिक्स, सैन डिएगो, कैलिफोर्निया, पीपी 15-21।

झांग बी (2015) माइक्रोआरएनए: ए न्यू टार्गिट फॉर इम्प्रिवगं प्लांट टालरन्स टू एबाइआटिक स्ट्रेस । जे एक्सप बॉट 66:1749-1761। https://doi.org/10.1093/jxb/erv013।

झांग बी, पैन एक्स, कैनन सीएच, एट अल (2006) कान्सर्वेशन ऐन्ड डाइवर्जन्स ऑफ प्लांट माइक्रोआरएनए जीन । प्लांट जे 46:243-259। https://doi.org/10.1111/j.1365-313X.2006.02697.x।

झांग बी, वांग क्यू (2015) माइक्रोआरएनए-बैस्ट बाइओटेक्नालजी फॉर प्लांट इम्प्रूव्मन्ट। जे सेल फिजियोल 230:1-15. https://doi.org/10.1002/jcp.24685।

झांग एफ, यांग जे, झांग एन, एट अल (2022) रोल ऑफ माइक्रोआरएनए इन एबाइआटिक स्ट्रेस रिस्पान्स ऐन्ड केरक्टरिस्टिक्स रेग्यलैशन ऑफ प्लांट। फ्रंटियर्स इन प्लांट साइअन्स 13:919243 https://doi.org/10.3389/fpls.2022.919243।

झोउ एल, लियू वाई, लियू जेड, एट अल (2010) जीनोम-वाइड आइडेन्टफकेशन ऐन्ड अनैलिसिस ऑफ ड्राउट रिस्पान्सिव माइक्रोआरएनए इन ओरिजा सैटिवा। जे एक्सप बॉट 61:4157-4168। https://doi.org/10.1093/jxb/erq237।

झांग जेड, यू जे, ली डी, एट अल (2010) पीएमआरडी: प्लांट माइक्रोआरएनए डेटाबेस। न्यूक्लिक एसिड रेस 38:D806-813. https://doi.org/10.1093/nar/gkp818।

जुरब्रिगेन एमडी, हाजीरेजेई एम-आर, कैरिलो एन (2010) इंजीनियरिंग द फ्यूचर। डिवेलप्मन्ट ऑफ ट्रांसजेनिक प्लांट विद एन्हैन्स्ट टालरन्स टू ऐड्वर्स इन्वाइरन्मन्ट। बायोटेक्नॉल जेनेट इंजी रेव 27:33-56। https://doi.org/10.1080/02648725.2010.10648144।

कृषि में आर्टिफिशियल इंटेलिजेंस के अनुप्रयोग

मधु, सपना निगम, अक्षय धीरज, मो. अशरफुल हक, सुदीप मरवाहा
भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

भारतीय अर्थव्यवस्था मुख्य रूप से कृषि और उससे जुड़े क्षेत्रों पर निर्भर करती है। चूँकि जनसंख्या दिन-प्रतिदिन बढ़ती जा रही है, इसलिए किसानों द्वारा उपयोग की जाने वाली खेती की विधियाँ इतनी जनसंख्या की भोजन की आवश्यकता को पूरा करने के लिए पर्याप्त नहीं थीं। कृषि के क्षेत्र में आर्टिफिशियल इंटेलिजेंस एक उभरती हुई तकनीक है। आर्टिफिशियल इंटेलिजेंस की मदद से फसल की उपज को मौसम की समस्याओं, जलवायु परिवर्तन, कीटनाशकों और कीटनाशकों के कम छिड़काव आदि जैसे विभिन्न कारकों से बचाया गया है। इस तकनीक का उपयोग करके हम आपूर्ति और मांग के बीच अंतर का अनुमान लगा सकते हैं और फसल की उपज और मौसम की स्थिति के भविष्य के रुझान का पूर्वानुमान लगा सकते हैं।

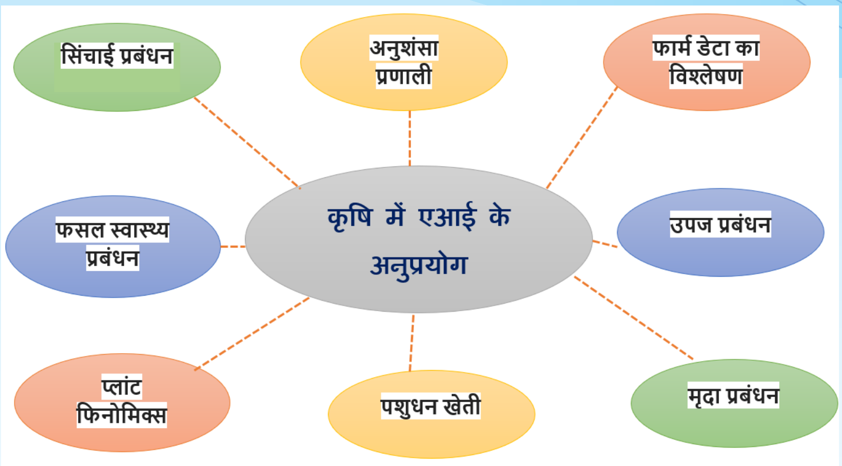
आर्टिफिशियल इंटेलिजेंस

आर्टिफिशियल इंटेलिजेंस (एआई) अध्ययन का एक बहुत पुराना क्षेत्र है और इसका एक समृद्ध इतिहास है। आधुनिक एआई को जॉन मैक्कार्थी द्वारा औपचारिक रूप दिया गया था स जिन्हें एआई का जनक माना जाता है। यह कंप्यूटर विज्ञान की एक शाखा है जिसकी स्थापना 1950 के दशक की शुरुआत में हुई थी। मुख्य रूप से आर्टिफिशियल इंटेलिजेंस शब्द तकनीकों के एक समूह को संदर्भित करता है जो कंप्यूटर या मशीन को समस्या सुलझाने के कार्यों में मनुष्यों के व्यवहार की नकल करने में सक्षम बनाता है। एआई का मुख्य उद्देश्य कंप्यूटर को कुछ कार्यों को मानवीय तरीके से करने के लिए प्रोग्राम करना है जैसे कि नॉलेजबेस, तर्क, सीखना, योजना बनाना, समस्या समाधान आदि। मशीन लर्निंग (एमएल) तकनीक एआई का सबसेट है जो कंप्यूटर/मशीन/प्रोग्राम को स्पष्ट रूप से प्रोग्राम किए बिना सीखने और कार्य करने में सक्षम बनाता है। एमएल तकनीक सिर्फ मानव व्यवहार की नकल करने का तरीका नहीं है बल्कि इंसान चीजों को कैसे सीखता है इसकी नकल करने का तरीका भी है। किसी भी प्रकार की समस्या को हल करने के लिए मशीन लर्निंग की मुख्य विशेषता "अनुभव से सीखना" हैस सीखने के तरीकों को तीन प्रकारों में वर्गीकृत किया जा सकता है:- ;ए. सुपरवाइज्ड लर्निंग :-जिसमें सीखने के लिए लेबल किए गए डेटा के साथ एल्गोरिदम प्रदान किया जाता है और हमें वांछित आउटपुट पैटर्न मिलते हैं ;बी. अनसुपरवाइज्ड लर्निंग:- जहां एल्गोरिदम को बिना लेबल वाला डेटा प्रदान किया जाता है और यह इनपुट डेटा से पैटर्न की पहचान करता है और ;सी. रीइन्फोर्स्मेन्ट लर्निंग:-जहां एल्गोरिदम एमएल तकनीकों को पुरस्कार या सुदृढीकरण के आधार पर सीखने योग्य चीजों को पकड़ने की अनुमति देता है।

कृषि में आर्टिफिशियल इंटेलिजेंस के अनुप्रयोग

सिंचाई प्रबंधन

सिंचाई को अनुकूलित करने के लिए फसल उत्पादकता बढ़ाने के लिए एआई आधारित स्मार्ट सिंचाई प्रणाली का उपयोग किया गया है। इन प्रणालियों में इंटरनेट ऑफ थिंग्स सेंसर का उपयोग फसलों को पानी देने और कीटनाशकों और उर्वरकों के उपयोग के लिए मूल्यवान जानकारी देने के लिए किया जा सकता है। रिमोट सेंसिंग प्रौद्योगिकियों और मृदा सेंसर का उपयोग करके किसान पूरे क्षेत्र में मिट्टी की नमी के स्तर तक पहुंच सकते हैं। यह जानकारी किसानों को ठीक वहीं पानी लगाने में मदद करती है जहां इसकी आवश्यकता होती है, पर्याप्त नमी वाले क्षेत्रों में



आकृति 1. कृषि में आर्टिफिशियल इंटेलिजेंस के अनुप्रयोग

अति-सिंचाई से बचने और अधिक पानी की मांग वाले क्षेत्रों में सिंचाई को लक्षित करने में मदद मिलती है।

फसल स्वास्थ्य प्रबंधन

फसल स्वास्थ्य निगरानी में फसल स्वास्थ्य, विकास पैटर्न और पोषक तत्वों की आवश्यकताओं के बारे में वास्तविक समय की जानकारी प्रदान करने के लिए विभिन्न एआई संचालित प्रणालियों जैसे उपग्रह इमेजरी, ड्रोन, मौसम सेंसर और आईओटी उपकरणों का उपयोग शामिल है। विभिन्न कंप्यूटर दृष्टि तकनीकों का उपयोग करके पत्ती रोग और कीट का पता लगाया जा सकता है । यह तकनीक किसानों को फसल वृद्धि पर वास्तविक समय का डेटा प्रदान कर सकती है और बड़ी समस्या बनने से पहले प्रमुख मुद्दों की पहचान करने में उनकी मदद कर सकती है।

प्लांट फिनोमिक्स

पौधों की डिजिटल छवियों को संसाधित करने के लिए कंप्यूटर विजन सिस्टम का उपयोग किया जाता है। इन तकनीकों का उपयोग करके वस्तु पहचान उद्देश्यों के लिए पौधों की विशिष्ट विशेषताओं का पता लगाया जा सकता है। फसल फेनोटाइप के बड़े संग्रह से नए रुझानों की भविष्यवाणी करने के लिए अद्वितीय डेटा को वर्गीकृत करने, नए पैटर्न और सुविधाओं की पहचान करने के लिए एआई प्रौद्योगिकियों का उपयोग किया जा रहा है। फेनोमिक्स का उपयोग कई फेनोटाइपिक लक्षणों जैसे स्पाइक का पता लगाना और गिनती उपज का पूर्वानुमान, पौधे में बुढ़ापे की मात्रा का ठहराव, पत्ती का वजन और गिनती आदि का अध्ययन करने के लिए किया गया है।

अनुशंसा प्रणाली

मशीन लर्निंग विधियों और एल्गोरिदम का उपयोग करते हुए, एआई-आधारित अनुशंसा प्रणाली ऐतिहासिक डेटा और वैज्ञानिक विशेषज्ञता को शामिल करके कृषि समस्याओं के समाधान प्रदान कर सकती है। यह न केवल वैज्ञानिक कार्यबल की दक्षता बढ़ाने में मदद करेगी, बल्कि उत्पादित खाद्य सामग्री की मात्रा का सटीक अनुमान लगाने में भी सुधार करेगी।

पशुधन खेती

पशुधन खेती में पशुधन के स्वास्थ्य व्यवहार और भोजन पैटर्न की निगरानी के लिए एआई संचालित सेंसर का उपयोग शामिल है। यह तकनीक किसानों को बड़ी समस्या बनने से पहले किसी भी स्वास्थ्य समस्या की पहचान करने और बीज और पानी के उपयोग को अनुकूलित करने में मदद कर सकती है। पशुधन के व्यवहार की निगरानी करके किसान भोजन या पानी व्यवस्था से संबंधित किसी भी समस्या की पहचान कर सकते हैं। एआई प्रौद्योगिकियों का उपयोग चारा प्रबंधन में अपशिष्ट प्रबंधन बनाने और जानवरों को दूध देने के लिए भी किया जा सकता है ।

मृदा प्रबंधन

मृदा विश्लेषण में पोषक तत्वों के स्तर को निर्धारित

करने और इष्टतम उर्वरकों के उपयोग की सिफारिश करने के लिए मिट्टी के नमूनों पर डेटा का विश्लेषण करने के लिए एआई एल्गोरिदम का उपयोग शामिल है। यह तकनीक किसानों को उनकी फसल की पैदावार को अनुकूलित करने और उपयोग के लिए सर्वोत्तम प्रकार और उर्वरक की मात्रा की पहचान करके बर्बादी को कम करने में मदद कर सकती है। मिट्टी की गुणवत्ता पर डेटा का विश्लेषण करके किसान यह निर्णय ले सकते हैं कि कौन सी फसलें बोई जाएं और अपने खेतों का प्रबंधन कैसे किया जाए।

एआई का उपयोग करके उपज प्रबंधन

एआई क्लाउड कंप्यूटिंग सैटेलाइट इमेजरी और उन्नत एनालिटिक्स ने आज स्मार्ट कृषि के लिए एक पारिस्थितिकी तंत्र बनाया है। इन प्रौद्योगिकियों का संयोजन किसानों को उच्च औसत उपज और खेती पर बेहतर नियंत्रण प्राप्त करने की अनुमति देता है।

एआई का उपयोग करके फार्म डेटा का विश्लेषण

मौसम, तापमान, पानी की खपत, मिट्टी की स्थिति आदि जैसे हजारों फील्ड डेटा बिंदुओं को सेंसर की मदद से हर दिन जमीन पर कैप्चर किया जाता है। इस डेटा का विश्लेषण किया जा सकता है, और एआई तकनीकों का उपयोग करके इन मापदंडों की अग्रिम भविष्यवाणी के लिए उपयोगी पैटर्न निकाले जा सकते हैं। इससे खेती की सटीकता और उत्पादकता में सुधार होगा।

पौधों की बीमारियों के परीक्षणों में अर्ध-लैटिन आयत अभिकल्पनाओं का उपयोग

कौशल कुमार यादव एवं सुकान्त दाश
भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान

सार

अर्ध-लैटिन आयत (SLR) का व्यापक रूप से कृषि और औद्योगिक परीक्षणों में उपयोग किया जाता है, जहाँ एक प्रभाव को पंक्तियों और दूसरे को स्तंभों में आवंटित किया जाता है, और प्रत्येक अंतथप्रतिच्छेदन (सेल) मे कई परीक्षणात्मक इकाइयाँ होती हैं। इस लेख में SLR अभिकल्पनाओं के अनुपरीक्षण, विशेष रूप से पौधों की बीमारियों के परीक्षणों में, का अध्ययन किया गया है और 2 के सेल आकार वाले SLR अभिकल्पनाओं के निर्माण के तरीकों को प्रस्तुत किया गया है। प्रस्तावित पद्धति परीक्षण ात्मक उपचार की व्यवस्था के लिए एक व्यवस्थित ढांचा प्रदान करती है, जो परीक्षणात्मक सेटअप को अनुकूलित करते हुए उपचारो के प्रभावों के प्रभावी अध्ययन को सक्षम बनाती है।

मुख्य शब्द: ओर्थोगोनालिटी, पंक्ति-स्तंभ अभिकल्पनाएं, अर्ध-लैटिन आयत

परिचय

कृषि परीक्षणों में, परीक्षणात्मक इकाइयों के बीच विषमता अक्सर उपचार के प्रभाव को प्रभावित करती है, जिसके कारण ऐसे अभिकल्पनाओं की आवश्यकता होती है जो इस भिन्नता को ध्यान में रखें। ऐसे चुनौतियों का समाधान करने के लिए, विभिन्न पंक्ति-स्तंभ अभिकल्पनाएँ विकसित किए गए हैं, जिनमें लैटिन स्क्वायर अभिकल्पनाएँ, यूडेन स्क्वायर अभिकल्पनाएँ, सामान्यीकृत यूडेन अभिकल्पनाएँ और अन्य संबंधित विधियाँ शामिल हैं। इन अभिकल्पनाएँ पर विस्तृत चर्चा शाह और सिन्हा (1996) में उपलब्ध है।

पारंपरिक पंक्ति-स्तंभ अभिकल्पनाएँ आमतौर पर प्रत्येक पंक्ति और स्तंभ के अंतथप्रतिच्छेदन पर एक ही परीक्षणात्मक इकाई शामिल करते हैं। हालांकि, व्यावहारिक परिस्थितियों में, अक्सर प्रत्येक अंतथप्रतिच्छेदन पर कई परीक्षणात्मक इकाइयों की आवश्यकता होती है ताकि उपचार के प्रभावों का पर्याप्त अध्ययन किया जा सके। ऐसी स्थितियों के लिए, अर्ध-लैटिन स्क्वायर अभिकल्पनाओं की अवधारणा प्रस्तुत की गई। अर्ध-लैटिन स्क्वायर अभिकलपानों को एक $r \times r$ सरणी के रूप में परिभाषित किया गया है, जिसमें प्रत्येक पंक्ति-स्तंभ अंतथप्रतिच्छेदन पर $k > 1$ समान संख्या में परीक्षणात्मक इकाइयाँ होती हैं, और प्रत्येक उपचार को प्रत्येक पंक्ति और स्तंभ में ठीक एक बार आवंटित किया जाता है। अर्ध-लैटिन स्क्वायर पर बुनियादी कार्य बेली (1988, 1992), बेली और चिग्बु (1997), बेली और रॉयल (1997), बेडफोर्ड और व्हिटेकर (2001), चिग्बु (2003), परसाद (2006), सोइचर (2012, 2013), और बेली और सोइचर (2021) के अध्ययनों में उपलब्ध हैं।

अर्ध-लैटिन स्क्वायर अभिकल्पनाएं उन परीक्षणों में विशेष रूप से लाभकारी होते हैं जहाँ दोनों अवांछनीय कारकों में समान स्तर होते हैं। लेकिन जब अवांछनीय कारकों के स्तर असमान होते हैं, तो अर्ध-लैटिन स्क्वायर उपयुक्त नहीं रहते। इस कमी को दूर करने के लिए, अर्ध-लैटिन स्क्वायर का विस्तार करते हुए अर्ध-लैटिन आयत अभिकल्पनाएँ विकसित किए गए। अर्ध-लैटिन आयत अभिकल्पनाएँ एक पंक्ति-स्तंभ अभिकलनाएँ है जिसमें v उपचार, R पंक्तियाँ,

और C स्तंभ होते हैं, और प्रत्येक अंतथप्रतिच्छेदन (सेल) में k > 1 परीक्षणात्मक इकाइयाँ होती हैं। प्रत्येक उपचार को पंक्तियों में r_R बार और स्तंभों में r_C बार पुनरावृत्त किया जाता है, जिससे यह अभिकल्पनाएँ परीक्षणात्मक के लिए एक बहुमुखी ढाँचा प्रदान करता है। अर्ध-लैटिन आयत अभिकल्पनाएँ की लचीलापन इसे उन परीक्षणों के लिए उपयुक्त बनाता है जिनमें अवांछनीय कारकों के स्तर असमान होते हैं। इन अभिकल्पनाओं के सैद्धांतिक चर्चा और व्यावहारिक अनुपरीक्षणों के लिए, इच्छुक पाठक बेली और मोनोड (2001), यूटो और बेली (2020, 2022), यादव इत्यादि (2024a, 2024b, 2024c) और अन्य संबंधित साहित्य का संदर्भ ले सकते हैं।

अर्ध-लैटिन आयत अभिकल्पनाएँ विभिन्न प्रकार के परीक्षणों में उपयोगी होते हैं, जिनमें कृषि अध्ययन जैसे पौधों की बीमारियों के परीक्षण, खाद्य संवेदन परीक्षण, ग्रीनहाउस परीक्षण, और उपभोक्ता परीक्षण परीक्षण शामिल हैं। उदाहरण के लिए, पौधों की बीमारियों के परीक्षण में, दस पौधे (C = 10) और आधी पत्तियों के जोड़े (k = 2) को पाँच पत्ती ऊँचाइयों (R = 5) पर व्यवस्थित किया जा सकता है, जिसमें पाँच उपचार (v =5) शामिल होते हैं, जो विभिन्न प्रकार के वायरस या वायरस की अलग-अलग खुराक हो सकती हैं। इनका उपयोग इस प्रकार किया जा सकता है कि प्रत्येक उपचार प्रत्येक पंक्ति में चार बार (r_R = 4) और प्रत्येक स्तंभ में दो बार (r_C = 2) प्रकट हो। परीक्षणात्मक अभिकल्पना तालिका 1 मे प्रस्तुत है:

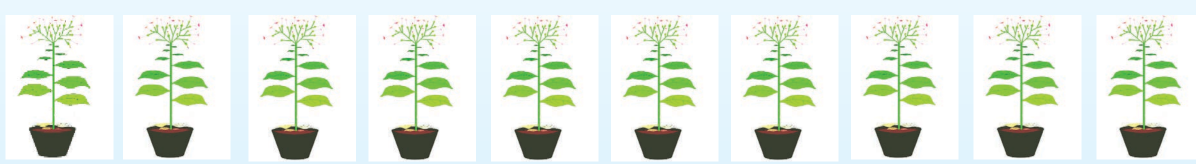
अर्ध-लैटिन आयत अभिकल्पनाओं मे यादृच्छिकीकरण

अर्ध-लैटिन आयतों (SLR) के लिए यादृच्छिकीकरण प्रक्रिया यह सुनिश्चित करती है कि पंक्तियों, स्तंभों या उपचारों के आयोजन में किसी भी संभावित पक्षपाती प्रभाव को न्यूनतम किया जाए, जिससे परीक्षणात्मक परिणामों की वैधता में वृद्धि होती है। यह प्रक्रिया SLR की सभी पंक्तियों को यादृच्छिकीकरण करने से शुरू होती है, जिससे यह सुनिश्चित होता है कि उनकी व्यवस्था केवल संयोग द्वारा निर्धारित की जाए। इसके बाद, स्तंभों को पंक्तियों से स्वतंत्र रूप से यादृच्छिकीकरण किया जाता है, जिससे दोनों आयामों की स्वतंत्रता बनी रहती है। अंत में, प्रत्येक सेल के भीतर, उपचारों की व्यवस्था को स्वतंत्र रूप से यादृच्छिकीकरण किया जाता है। यह बहु-चरण ीय यादृच्छिकीकरण प्रक्रिया यह सुनिश्चित करती है कि परीक्षणात्मक इकाइयों को उपचार आवंटित करते समय कोई प्रणालीगत नमूना नहीं होगा, जिससे परीक्षणात्मक अभिकल्पना की अखंडता बनी रहती है और परिणामों के बिना पक्षपाती सांख्यिकीय विश्लेषण की अनुमति मिलती है।

परीक्षणों के लिए अर्ध-लैटिन आयत अभिकल्पनाएं प्राप्त करना

वर्तमान लेख में, हम अर्ध-लैटिन आयत (SLR) अभिकल्पनाओं के लिए संरचना विकसित करने पर ध्यान केंद्रित कर रहे हैं, जिसमें प्रत्येक सेल का आकार दो होता

तालिका 1: पौधों की बीमारियों के परीक्षणों में 5 उपचारों के लिए अर्ध-लैटिन आयत अभिकल्पना



	पौधा 1	पौधा 2	पौधा 3	पौधा 4	पौधा 5	पौधा 6	पौधा 7	पौधा 8	पौधा 9	पौधा 10
प.ऊँ. 5	1, 2	2, 3	3, 4	4, 5	5, 1	1, 3	2, 4	3, 5	4, 1	5, 2
प.ऊँ. 4	2, 3	3, 4	4, 5	5, 1	1, 2	2, 4	3, 5	4, 1	5, 2	1, 3
प.ऊँ. 3	3, 4	4, 5	5, 1	1, 2	2, 3	3, 5	4, 1	5, 2	1, 3	2, 4
प.ऊँ. 2	4, 5	5, 1	1, 2	2, 3	3, 4	4, 1	5, 2	1, 3	2, 4	3, 5
प.ऊँ. 1	5, 1	1, 2	2, 3	3, 4	4, 5	5, 2	1, 3	2, 4	3, 5	4, 1

है। प्रस्तावित विधि लचीली है और इसे किसी भी संख्या के उपचारों v के लिए लागू किया जा सकता है। इन अभिकल्पनाओं को बनाने की प्रक्रिया निम्नलिखित चरणों का पालन करती है:

चरण-1: यादृच्छिक जोड़ियों का चयन

पहला कदम 1 से v तक के समुच्चय से यादृच्छिक जोड़ियों का चयन करना है। यह जोड़ी पहले पंक्ति और पहले स्तंभ दोनों में पहला सेल बनाएगी। इसके बाद, जोड़ियों के तत्वों में एक को जोड़कर अगले v-1 सेल बनाए जाते हैं। उदाहरण के लिए, यदि हम v = 5 के लिए जोड़ियां (1, 2) चुनते हैं, तो एक को जोड़ने से उत्पन्न जोड़ियां इस प्रकार होंगी: (2, 3), (3, 4), (4, 5), (5, 1)। ये जोड़ियाँ SLR अभिकल्पना के लिए पहली पंक्ति और स्तंभ बनाएंगी।

चरण-2: अतिरिक्त सेल के लिए चक्रीय घुमाव

एक बार जब प्रारंभिक v सेल पहली पंक्ति और स्तंभ में रख दिए जाते हैं, तो अगला कदम चक्रीय घुमाव करना होता है। इस ऑपरेशन के तहत, सेल्स को चक्रीय रूप से घुमाया जाता है ताकि अभिकल्पना की संरचना तैयार किया जा सके। परिणामी व्यवस्था अर्ध-लैटिन आयतों (SLR) का एक सेट उत्पन्न करेगी, जिसमें प्रत्येक सेल्स का आकार दो होगा।

इस पद्धति को लागू करके, $(v \times v)/2$ सेल्स का एक अभिकल्पना तैयार किया जा सकता है, जो परीक्षणात्मक संरचनाओ की बाधाओं को समायोजित करता है।

उदाहरण के लिए, v=5 के लिए SLR अभिकल्पना का संरचना होगा:

1 5	5 4	4 3	3 2	2 1
2 1	1 5	5 4	4 3	3 2
3 2	2 1	1 5	5 4	4 3
4 3	3 2	2 1	1 5	5 4
5 4	4 3	3 2	2 1	1 5

$(5 \times 5)/2$

v = 7 के लिए SLR अभिकल्पना का संरचना होगा:

1 7	7 6	6 5	5 4	4 3	3 2	2 1
2 1	1 7	7 6	6 5	5 4	4 3	3 2
3 2	2 1	1 7	7 6	6 5	5 4	4 3
4 3	3 2	2 1	1 7	7 6	6 5	5 4
5 4	4 3	3 2	2 1	1 7	7 6	6 5
6 5	5 4	4 3	3 2	2 1	1 7	7 6
7 6	6 5	5 4	4 3	3 2	2 1	1 7

$(7 \times 7)/2$

उपसंहारात्मक टिप्पणी

प्रस्तावित अभिकल्पनाओं की संरचना उन परीक्षणात्मक परिस्थितियों के लिए अत्यंत लाभकारी है, जो दो-तरफा भिन्नताओं को प्रदर्शित करते हैं, विशेष रूप से ऐसे संदर्भों में जहाँ प्रत्येक सेल मे दो परीक्षणात्मक इकाइयाँ शामिल होती हैं। यह अभिकल्पना विशेष रूप से पौधों की बीमारियों के परीक्षणों में उपयोगी है, जहाँ बीमारियों के प्रसार पर कई कारकों के प्रभाव को समझना महत्वपूर्ण होता है। इसके अलावा, ये अभिकल्पना उन वैज्ञानिक परीक्षणों के लिए भी उपयुक्त हैं, जो विषम परिस्थितियों वाले क्षेत्रों, जैसे पहाड़ी इलाकों में सीढ़ीनुमा खेती (टेरस फार्मिंग), में किए जाते हैं। ऐसे परिवेशों में, विभिन्न सीढ़ीदार खंडों में मिट्टी की बनावट, उर्वरता, और अन्य पर्यावरणीय कारकों में भिन्नताएँ परीक्षणात्मक परिणामों को महत्वपूर्ण रूप से प्रभावित कर सकती हैं। प्रस्तावित अभिकल्पना की लचीलापन शोध कर्ताओं को इन भिन्नताओं को ध्यान में रखने की अनुमति देती है, जिससे अधिक सटीक और कुशल विश्लेषण संभव हो पाता है। कुल मिलाकर, यह दृष्टिकोण शोधकर्ताओं को विभिन्न कारकों के आपसी प्रभावों का व्यवस्थित रूप से अध्ययन करने के लिए एक प्रभावी साधन प्रदान करता है, विशेष रूप से जटिल वातावरण और प्राकृतिक विविधता वाले कृषि और पारिस्थितिक अध्ययनों में।

संदर्भ

बेली, आर.ए. (1988)। सेमी-लैटिन स्क्वेयर्स। जर्नल ऑफ स्टैटिस्टिकल प्लानिंग एंड इन्फरेंस, 18(3), 299-312।

बेली, आर.ए. (1992)। एफिशिएंट सेमी-लैटिन स्क्वेयर्स।

स्टैटिस्टिका सिनीका, 2, 413-437।

बेली, आर.ए. एंड चिग्बू, पी.ई. (1997)। एनुमरेशन ऑफ सेमी-लैटिन स्क्वेयर्स। डिस्क्रीट मैथमेटिक्स, 167, 73-84।

बेली, आर.ए. एंड मोनोड, एच. (2001)। एफिशिएंट सेमी-लैटिन रेक्टैंगल्स: डिजाइन्स फॉर प्लांट डिजीज एक्सपेरिमेंट्स। स्कैंडिनेवियन जर्नल ऑफ स्टैटिस्टिक्स, 28(2), 257-270।

बेली, आर.ए. एंड रॉयल, जी. (1997)। ऑप्टिमल सेमी-लैटिन स्क्वेयर्स विथ साइड सिक्स एंड ब्लॉक साइज टू। प्रोसीडिंग्स ऑफ द रॉयल सोसाइटी ऑफ लंदन। सीरीज ए: मैथमेटिकल, फिजिकल एंड इंजीनियरिंग साइंसेस, 453(1964), 1903-1914।

बेली, आर.ए. एंड सोइचर, एल.एच. (2021)। यूनिफॉर्म सेमी-लैटिन स्क्वेयर्स एंड देयर पेयरवाइज-वैरिएंस अबरेशन्स। जर्नल ऑफ स्टैटिस्टिकल प्लानिंग एंड इन्फरेंस, 213, 282-291।

बेडफोर्ड, डी. एंड व्हिटेकर, आर.एम. (2001)। अ न्यू कंस्ट्रक्शन फॉर एफिशिएंट सेमी-लैटिन स्क्वेयर्स। जर्नल ऑफ स्टैटिस्टिकल प्लानिंग एंड इन्फरेंस, 98(1-2), 287-292।

चिग्बू, पी.ई. (2003)। द "बेस्ट" ऑफ द थ्री ऑप्टिमल (4×4)/4 सेमी-लैटिन स्क्वेयर्स। सांख्य: द इंडियन जर्नल ऑफ स्टैटिस्टिक्स, 65(3), 641-648।

परसाद, आर. (2006)। अ नोट ऑन सेमी-लैटिन स्क्वेयर्स। जर्नल ऑफ इंडियन सोसाइटी एंड एग्रीकल्चरल स्टैटिस्टिक्स, 60(2), 131-133।

सोइचर, एल.एच. (2012)। यूनिफॉर्म सेमी-लैटिन स्क्वेयर्स एंड देयर शूर-ऑप्टिमैलिटी। जर्नल ऑफ कॉम्बिनेटोरियल डिजाइन्स, 20(6), 265-277।

सोइचर, एल.एच. (2013)। ऑप्टिमल एंड एफिशिएंट सेमी-लैटिन स्क्वेयर्स। जर्नल ऑफ स्टैटिस्टिकल प्लानिंग एंड इन्फरेंस, 143(3), 573-582।

उटो, एन.पी. एंड बेली, आर.ए. (2020)। बैलेंस्ड सेमी-लैटिन रेक्टैंगल्स: प्रॉपर्टीज, एक्सिस्टेंस एंड कंस्ट्रक्शंस फॉर ब्लॉक साइज टू। जर्नल ऑफ स्टैटिस्टिकल थ्योरी एंड प्रैक्टिस, 14(3), 1-11।

उटो, एन.पी. एंड बेली, आर.ए. (2022)। कंस्ट्रक्शंस फॉर रेग्युलर-ग्राफ सेमी-लैटिन रेक्टैंगल्स विथ ब्लॉक साइज टू। जर्नल ऑफ स्टैटिस्टिकल प्लानिंग एंड इन्फरेंस, 221, 81-89।

यादव, के.के., दाश, एस., मंडल, बी.एन., एंड परसाद, आर. (2024a)। कंस्ट्रक्शन ऑफ बैलेंस्ड सेमी-लैटिन रेक्टैंगल्स इन ब्लॉक साइज फोर: एन अल्गोरिदमिक अप्रोच। जर्नल ऑफ स्टैटिस्टिकल थ्योरी एंड प्रैक्टिस, 18(3), 29।

यादव, के.के., दाश, एस., पारसाद, आर., मंडल, बी.एन., कुमार, ए., एंड कुमार, एम. (2024b)। कंस्ट्रक्शन ऑफ पार्टियली बैलेंस्ड सेमी-लैटिन रेक्टैंगल्स विथ ब्लॉक साइज 4। जर्नल ऑफ द इंडियन सोसाइटी ऑफ एग्रीकल्चरल स्टैटिस्टिक्स, 78(2), 115-123।

यादव, के.के., दाश, एस., कुमार, ए., मंडल, ए., एंड जाट, एस.एल. (2024c)। एप्लिकेशन ऑफ सेमी-लैटिन रेक्टैंगल्स डिजाइन इन मक्का एक्सपेरिमेंट्स। मक्का जर्नल, 13(2), 124-127।

कृषि में एकाधिक फ्रेम सर्वेक्षण

भारती, कौस्तव आदित्य, तौकीर अहमद, राहुल बनर्जी, दीपक सिंह, पंकज दास, अंकुर बिश्वास, राजू कुमार

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

सर्वेक्षणों का उद्देश्य किसी समष्टि के बारे में जानकारी एकत्रित करना होता है, लेकिन जब समष्टि की सभी इकाइयाँ अप्राप्य होती हैं या उनकी पूरी सूची तैयार नहीं की जा सकती, तो इस स्थिति में प्रतिदर्श (सैम्पल) सर्वेक्षण का प्रयोग किया जाता है। इस तरह, एक छोटे से प्रतिनिधि समूह से जानकारी प्राप्त की जाती है, जो समष्टि का सटीक और प्रभावी अनुमान प्रस्तुत कर सकता है। इस प्रक्रिया में, अध्ययन के लिए सही और सटीक प्रतिनिधिक प्रतिदर्श चुनने के लिए एक अच्छे प्रतिचयन फ्रेम का उपयोग किया जाता है, जिसमें समष्टि की सभी इकाइयाँ शामिल होती हैं। हालांकि, जब एकल फ्रेम से समष्टि का पूरा प्रतिनिधित्व नहीं हो पाता है या ढांचा पूरी तरह से अद्यतन नहीं होता, तो एकाधिक फ्रेम सर्वेक्षण विधि का उपयोग किया जाता है। यह विधि दो या दो से अधिक फ्रेम्स के संयोजन से अधिक सटीकता और व्याप्ति में सुधार करने में मदद करती है। इन सर्वेक्षणों का मुख्य उद्देश्य विभिन्न फ्रेमों से डेटा एकत्र करके समष्टि के प्रत्येक भाग का सही प्रतिनिधित्व सुनिश्चित करना है, जिससे परिणाम अधिक विश्वसनीय और सटीक हो सकें।

मुख्य शब्दः एकाधिक फ्रेम, प्रतिदर्श सर्वेक्षण, कृषि

परिचय

सर्वेक्षणों का मुख्य उद्देश्य, समष्टि के किसी चर विशेष के बारे में जानकारी एकत्रित करना है। 'समष्टि' से तात्पर्य है - सर्वेक्षण के उद्देश्यों के अनुसार परिभाषित इकाइयों का एक समूह। समष्टि की समस्त इकाइयों पर किसी चर विशेष पर सूचना के एकत्रीकरण को संपूर्ण परिगणना या जनगणना कहा जाता है। जनगणना करने के लिए अधिकतर ज्यादा अपेक्षित धनराशि और समय की आवश्यकता होती है और ऐसी अनेक परिस्थितियाँ होती हैं जहाँ सीमित संसाधनों के साथ संपूर्ण परिगणना संभव नहीं होती है। कई ऐसे उदाहरण हैं जहाँ समष्टि की सभी इकाइयाँ अप्राप्य होती हैं, जिससे सभी इकाइयों की परिगणना करना संभव नहीं हो पाता। ऐसे मामलों में अनुसंधानकर्ता के पास प्रतिदर्श सर्वेक्षण के अलावा कोई अन्य विकल्प नहीं बचता। जैसे कि, जब रक्त परीक्षण के लिए डॉक्टर के पास जाते हैं, तो वे शरीर से सारा रक्त नहीं निकालते है। इसके स्थान पर, वे आमतौर पर रक्त का एक छोटा सा प्रतिदर्श लेते हैं। इस प्रतिदर्श के आधार पर, डॉक्टर शरीर के समग्र स्वास्थ्य के बारे में निष्कर्ष निकालते हैं, यह मानते हुए कि प्रतिदर्श पूरे शरीर (यानी, समष्टि) का प्रतिनिधि है। प्रतिदर्श समष्टि से लिए गए प्रतिनिधिक उपसमूह को कहा जाता है, जिसके आधार पर समष्टि के प्राचल का आकलन किया जाता है।

समष्टि से प्रतिदर्श चयन करने के लिए एक प्रतिचयन फ्रेम/ढांचा (सैंपलिंग फ्रेम) की आवश्यकता होती है। अध्ययन की जाने वाली समष्टि की सभी प्रतिदर्शी इकाइयों की सूची को प्रतिचयन फ्रेम/ढांचा कहा जाता है। किसी क्षेत्र की पंजीकरण सूची, किसान रजिस्ट्री, फार्मों की सूची, गांवों या राज्य में उपयुक्त क्षेत्र खण्डों की सूची इत्यादि प्रतिचयन फ्रेम के उदाहरण हैं। एक अच्छी तरह से परिभाषित प्रतिचयन फ्रेम से सटीक और प्रतिनिधिक प्रतिदर्श चयन करना संभव होता है, जिससे सर्वेक्षण के परिणाम विश्वसनीय होते हैं।

एक अच्छी तरह से परिभाषित प्रतिचयन फ्रेम की निम्नलिखित विशेषताएँ होनी चाहिएः

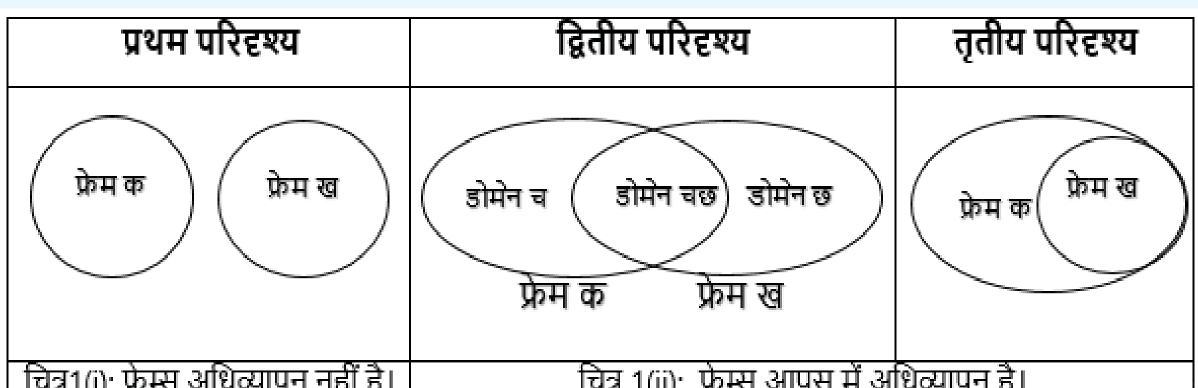
- **सटीकता:** यह अद्यतन और वर्तमान समष्टि को सही ढंग से दर्शाना चाहिए।
- **समावेशन:** इसमें लक्षित समष्टि के सभी इकाइयाँ शामिल होनी चाहिए।
- **विशिष्टता:** प्रत्येक इकाई को विशिष्ट होना चाहिए और उसमें कोई पुनरावृत्ति (डुप्लिकेट) नहीं होनी चाहिए।
- **विश्वसनीयता:** यह सटीक होना चाहिए और इसमें बहुत अधिक यादृच्छिक त्रुटि या पक्षपाती नहीं होना चाहिए।
- **गोपनीयता और व्यक्तिगत जानकारी की सुरक्षा:** इसमें नैतिक मानदंड का उल्लंघन नहीं होना चाहिए और सहभागी की जानकारी की सुरक्षा सुनिश्चित होनी चाहिए।

आमतौर पर यह देखा गया है कि प्रतिचयन ढाँचे पूर्ण और अद्यतित नहीं होते है और यदि यह पूर्ण हो, तो उस ढाँचे का उपयोग करना बहुत महंगा और जटिल हो सकता है। समय-समय पर इन ढाँचों को अद्यतन करना और सुनिश्चित करना कि वे पूरी समष्टि का सही प्रतिनिधित्व करते हैं, एक चुनौतीपूर्ण कार्य होता है। इसके अलावा, कभी-कभी इन ढाँचों में जानकारी की कमी या पुरानी जानकारी हो सकती है, जिससे सर्वेक्षण में पक्षपाती परिणाम प्राप्त हो सकते हैं। इसलिए इन परिस्थितियों में, कम लागत में समष्टि की बेहतर व्याप्ति प्राप्त करने के लिए एकाधिक ढाँचे का उपयोग किया जा सकता है। एकाधिक फ्रेम सर्वेक्षण में दो या दो से अधिक प्रतिचयन ढांचों का उपयोग किया जाता है। एकाधिक फ्रेम सर्वेक्षण का उपयोग लक्षित समष्टि की बेहतर व्याप्ति, व्याप्ति त्रुटियों के कारण होने वाली अभिनति को कम करने, और कम लागत में सटीकता के वांछित स्तर को प्राप्त करने के लिए किया जाता है।

एकाधिक फ्रेम सर्वेक्षण का अवलोकन

यह एक सर्वेक्षण विधि है जिसमें एक ही अध्ययन के लिए दो या दो से अधिक विभिन्न प्रतिचयन ढाँचे/फ्रेम्स का उपयोग किया जाता है। यह विधि तब उपयोगी होती है जब एकल फ्रेम से लक्षित समष्टि के सभी हिस्सों का सही प्रतिनिधित्व प्राप्त करना मुश्किल हो, या जब कुछ समष्टि समूहों तक पहुँचने में कठिनाई हो। इसमें विभिन्न प्रकार के फ्रेम, जैसे लिस्ट फ्रेम (सूची आधारित) और क्षेत्रीय फ्रेम (भौगोलिक आधार पर) का संयोजन किया जाता है, ताकि अधिक से अधिक समष्टि समूहों को कवर किया जा सके। प्रत्येक फ्रेम से स्वतंत्र प्रतिदर्श लिए जाते हैं, और प्राप्त डेटा को मिलाकर पूरी लक्षित जनसंख्या के बारे में निष्कर्ष निकाले जाते हैं। एकाधिक फ्रेम सर्वेक्षणों की अवधारणा को समझने के लिए, एक उदाहरण पर विचार करते है जहाँ दो फ्रेम्स (दोहरे फ्रेम सर्वेक्षण) का उपयोग पूरे लक्षित समष्टि को कवर करने के लिए किया जाता है।

दोहरे फ्रेम सर्वेक्षणों में, जो दो अलग-अलग प्रतिचयन ढाँचे/फ्रेम्स का उपयोग करते हैं, तीन विशिष्ट परिदृश्य उत्पन्न हो सकते हैं। पहला परिदृश्य तब उत्पन्न होता है जब फ्रेम्स आपस में अधिव्यापन नहीं करते और मिलकर



चित्र1(i): फ्रेम्स अधिव्यापन नहीं है।   चित्र 1(ii): फ्रेम्स आपस में अधिव्यापन है।

पूरी लक्षित समष्टि को कवर करते हैं, यह सुनिश्चित करते हुए कि समष्टि की प्रत्येक इकाइ को दो फ्रेम्स में से एक में बिना पुनरावृत्ति के शामिल किया जाता है (चित्र 1)।

दूसरा परिदृश्य ओवरलैप करने वाले फ्रेम्स से संबंधित है, जो मिलकर पूरी लक्षित समष्टि को कवर करते हैं। मान लीजिए क और ख दो अधिव्यापन वाले फ्रेम्स हैं, जो मिलकर लक्षित समष्टि को कवर करते हैं। इसके परिणामस्वरूप तीन पारस्परिक रूप से विशिष्ट डोमेन उत्पन्न होते हैं: (i) डोमेन च: उन इकाइयों को शामिल करता है जो फ्रेम क में हैं लेकिन फ्रेम ख में नहीं हैं; (ii) डोमेन छ: उन इकाइयों को शामिल करता है जो फ्रेम ख में हैं लेकिन फ्रेम क में नहीं हैं; और (iii) डोमेन चछ: ओवरलैप करने वाली इकाइयाँ जो दोनों फ्रेम क और फ्रेम ख में हैं। इस अधिव्यापन को सावधानीपूर्वक संभालने की आवश्यकता होती है ताकि उन इकाइयों को पुन: गिनने से बचा जा सके जो दोनों फ्रेम्स में दिखाई देती हैं।

एकाधिक फ्रेम सर्वेक्षण के विभिन्न उदाहरण

उदाहरण के लिए, गन्ना उत्पादन में, गन्ना सामान्यत: विभिन्न संग्रह केंद्रों के माध्यम से चीनी मिलों को आपूर्ति किया जाता है और उन सभी उत्पादकों की एक सूची मिल में उपलब्ध होती है जो गन्ना आपूर्ति करते हैं। मिल को आपूर्ति किए गए गन्ने के अलावा, उत्पादन का एक हिस्सा गाँव स्तर पर अन्य गुड़ उत्पादन के लिए भी उपयोग किया जाता है। इस प्रकार, इन दोनों प्राप्त सूची को लिस्ट फ्रेम के रूप में उपयोग किया जा सकता है।

- उदाहरण के तौर पर, एक राज्य में गेहूं पर सर्वेक्षण में, एक क्षेत्रीय फ्रेम (जैसे सैटिलाइट इमेज आधारित फ्रेम) सभी गेहूं उगाने वाले क्षेत्रों को शामिल कर सकता है, लेकिन इस फ्रेम से प्रतिदर्श चुनने से लागत और जटिलता बढ़ जाएगी क्योंकि पहले से एक मौजूदा प्रणाली (फसल क्षेत्र की गणना के लिए टाइमली रिपोर्टिंग स्कीम) मौजूद है, जो सूची फ्रेम के आधार पर प्रतिदर्श एकत्र करती है। एक बेहतर विकल्प यह होगा कि सूची फरेम और क्षेत्रीय फ्रेम दोनों से यादृच्छिक प्रतिदर्श लेकर फसल क्षेत्र का अनुमान अधिक सटीकता से किया जाए।
- मान लीजिए कि एक शोधकर्ता भारतीय ग्रामीण क्षेत्रों में स्वास्थ्य पर आधारित सर्वेक्षण करना चाहता है। यदि वह केवल सरकारी स्वास्थ्य केंद्रों से डेटा लेता है, तो वह निजी अस्पतालों में इलाज करवाने वाले लोगों की स्थिति से अनजान रहेगा। ऐसे में वह एकाधिक फ्रेम सर्वेक्षण कर सकता है, जिसमें सरकारी स्वास्थ्य केंद्र, निजी अस्पताल, और अन्य समुदाय आधारित डेटा स्रोतों को शामिल किया जाए।
- किसी राज्य में उपभोक्ता के खरीदारी पैटर्न का एक व्यापक और स्पष्ट चित्र प्राप्त करने के लिए एकाधिक फ्रेम का उपयोग किया जा सकता है । इस सर्वेक्षण में, एक फ्रेम ऑनलाइन शॉपिंग करने वाले उपभोक्ताओं से जानकारी प्राप्त करने पर ध्यान केंद्रित करेगा, जैसे कि वे कौन सी वेबसाइट्स या ऐप्स पर खरीदारी करते हैं, उनके खरीदारी निर्णयों में कौन से कारक प्रमुख होते हैं (जैसे मूल्य, सुविधाएं, समीक्षाएं)। वहीं दूसरा फ्रेम पारंपरिक दुकानों पर शॉपिंग करने वाले उपभोक्ताओं का होगा। दोनों फ्रेम्स के डेटा का विश्लेषण करने से उपभोक्ता के खरीदारी पैटर्न का एक व्यापक और स्पष्ट चित्र सामने आता है, जिससे यह समझने में मदद मिलती है कि वे किस प्रकार के उत्पाद और सेवाएं ऑनलाइन और ऑफलाइन में अधिक पसंद करते हैं, और उनके फैसलों को प्रभावित करने वाले प्रमुख तत्व कौन से हैं।

एकाधिक फ्रेम सर्वेक्षण के उद्देश्य

एकाधिक फरेम सर्वेक्षणों का उद्देश्य विभिन्न सैंपलिंग फ्रेम को मिलाकर डेटा संग्रह की सटीकता और व्यापकता को सुधारना है। यह विधि तब उपयोग की जाती है जब कोई एकल सैंपलिंग फ्रेम पूरी समष्टि का सही तरीके से प्रतिनिधित्व नहीं करता है, या जब समष्टि के विभिन्न उपसमूहों को विभिन्न फ्रेम्स के माध्यम से बेहतर तरीके से कवर किया जा सकता है। इसके मुख्य उद्देश्य निम्नलिखित हैं:

- **प्रतिनिधित्व में सुधारः** जब एकल फ्रेम का उपयोग किया जाता है, तो उसमें कुछ समष्टि इकाइयों का प्रतिनिधित्व नहीं हो पाता है। एकाधिक फ्रेम सर्वेक्षण का उद्देश्य यह सुनिश्चित करना होता है कि सभी इकाइयों का उचित प्रतिनिधित्व हो, ताकि परिणाम अधिक सटीक और विश्वसनीय हों।
- **सटीकता में वृद्धिः** एकाधिक फ्रेम सर्वेक्षण विभिन्न फ्रेमों से डेटा संग्रहित करने के कारण अधिक सटीक और विश्वसनीय परिणाम प्रदान करने में सक्षम होते हैं।
- **लागत की दक्षताः** कुछ परिस्थितियों में, एकाधिक फ्रेम्स का उपयोग करने से लागतों में कमी आ सकती है, क्योंकि उपलब्ध संसाधनों (जैसे कि विभिन्न डेटाबेस या रजिस्टर) का लाभ उठाया जा सकता है।
- **जटिलता को कम करनाः** जब एकल फ्रेम सर्वेक्षण में किसी समूह या डेटा की कमी हो, तो इसे हल करने के लिए एकाधिक फ्रेमों का उपयोग किया जाता है। इससे सर्वेक्षण की जटिलता कम होती है और अधिक विविध डेटा एकत्रित करने में मदद मिलती है।

अतः, एकाधिक फ्रेम सर्वेक्षण का उपयोग लक्षित समष्टि की बेहतर व्याप्ति, व्याप्ति त्रुटियों के कारण होने वाली अभिनति को कम करने, और कम लागत में सटीकता के वांछित स्तर को प्राप्त करने के लिए किया जाता है।

एकाधिक फ्रेम सर्वेक्षण की प्रक्रिया

एकाधिक फ्रेम सर्वेक्षण की प्रक्रिया में मुख्यतः निम्नलिखित चरण होते हैं:

- **फ्रेमों का चयनः** अध्ययन का उद्देश्य निर्धारित करने के बाद, यह तय किया जाता है कि अध्ययन में उपयोग के लिए कितनी फ्रेम उपलब्ध हैं। उदाहरण के लिए, एक सर्वेक्षण में राष्ट्रीय जनसंख्या रजिस्टर, नागरिकों की सूची, स्वास्थ्य केंद्रों से प्राप्त डेटा आदि को विभिन्न फ्रेम के रूप में इस्तेमाल किया जा सकता है। लिस्ट फ्रेम (सूची आधारित) और क्षेत्रीय फ्रेम (भौगोलिक आधार पर) का संयोजन किया जाता है, ताकि अधिक से अधिक समष्टि इकाइयों को कवर किया जा सके।
- **प्रतिदर्श चयनः** प्रत्येक फ्रेम से स्वतंत्र रूप से प्रतिदर्श चयन किये जाते हैं। जब अलग-अलग फ्रेमों से डेटा एकत्रित कर लिया जाता है, तो उसे संयोजित किया जाता है और उसके बाद सम्पूर्ण लक्षित जनसंख्या के बारे में निष्कर्ष निकाले जाते हैं।
- **विश्लेषणः** एक बार डेटा संयोजित हो जाने के बाद, उसका विश्लेषण किया जाता है। सबसे पहले हार्टले ने 1962 में सर्वेक्षणों में दो या दो से अधिक फ्रेम्स के संयुक्त उपयोग के लिए एक व्यवस्थित पद्धति दी। उन्होंने अमेरिकन स्टैटिस्टिकल एसोसिएशन की मिनियापोलिस बैठक में मल्टीपल फ्रेम सर्वेक्षण शीर्षक से एक शोध पत्र प्रस्तुत किया। उन्होंने पहले केवल दो फ्रेम (ड्यूल फ्रेम) पर विचार किया और यह मान लिया कि प्रतिदर्श दोनों फ्रेम्स से स्वतंत्र रूप से लिए जाते हैं। इस दृष्टिकोण का एक दिलचस्प पहलू यह था कि अधिव्यापन फ्रेम्स से आने वाले प्रतिदर्श के हिस्सों के लिए प्रतिदर्श आकार और वजन 'p' और '1–p' का अनुकूलन करके एक संयुक्त अनुमान प्राप्त किया गया। इस प्रकार, हार्टले (1962) द्वारा दोहरे फ्रेम सर्वेक्षणों के लिए समष्टि कुल का अनुमान प्रस्तुत किया जब दोनों फ्रेम्स में साधारण यादृच्छिक प्रतिदर्श चयन का उपयोग किया गया, जैसा कि नीचे दिया गया है:

$$\hat{Y} = N_a\bar{y}_a + N_{ab}(p\bar{y}_{ab} + q\bar{y}_{ba}) + N_b\bar{y}_b \quad (p+q=1)$$

जहाँ $\alpha = \frac{N_{ab}}{N_A}$; $\beta = \frac{N_{ab}}{N_B}$; $\bar{y}_a$; = डोमेन 'च' का प्रतिदर्श माध्य; $\bar{y}_{ab}$ = फ्रेम 'क' से प्रतिदर्श लिया जाने पर, डोमेन 'चछ' का प्रतिदर्श माध्य; $\bar{y}_{ba}$ = फ्रेम 'ख' से प्रतिदर्श लिया जाने पर, डोमेन 'चछ' का प्रतिदर्श माध्य; $\bar{y}_b$ = डोमेन 'छ' का प्रतिदर्श माध्य ।

इसके बाद, शोधकर्ताओं ने एकाधिक फ्रेम सर्वेक्षणों में स्वतंत्र प्रतिदर्श से जानकारी एकत्र करने के लिए विभिन्न विधियाँ प्रस्तावित की हैं, ताकि समष्टि के प्राचल जैसे कुल, माध्य, और अन्य प्राचल अनुमानित किए जा सकें। इनमें फुलर और बर्माइस्टर (1972), हार्टले (1974), सक्सेना एट अल (1984), बैंकीयर (1986), काल्टन और

एंडरसन (1986), स्किनर एट अल (1994), स्किनर और राव (1996), रानाली एट अल (2016), रूएडा एट अल (2017), दासगुप्ता (2019), कुमार एट अल (2021) आदि शामिल हैं।

एकाधिक फ्रेम सर्वेक्षणों की चुनौतियाँ

1. **जटिलता:** बहु-फ्रेम सर्वेक्षण जटिल हो सकते हैं क्योंकि इसमें विभिन्न फ्रेमों से डेटा एकत्र करना और उसे संयोजित करना पड़ता है। यह प्रक्रिया समय लेने वाली और श्रमसाध्य हो सकती है।
2. **डेटा समन्वय:** विभिन्न फ्रेमों से एकत्रित डेटा का समन्वय करना चुनौतीपूर्ण हो सकता है। यदि विभिन्न फ्रेमों से प्राप्त आंकड़े आपस में मेल नहीं खाते, तो इसका परिणाम सटीक नहीं हो सकता।
3. **प्रभावित निर्णय:** कभी-कभी विभिन्न फ्रेमों से डेटा एकत्र करते समय, एक फ्रेम से अधिक आंकड़े प्राप्त हो सकते हैं, जबकि अन्य फ्रेम से कम। इससे डेटा का चयन पक्षपाती हो सकता है।

निष्कर्ष

एकाधिक फ्रेम सर्वेक्षण एक प्रभावी तरीका है, जो विशेष रूप से तब उपयोगी होता है जब एकल फ्रेम से समष्टि का सही प्रतिनिधित्व प्राप्त करना मुश्किल हो। यह विधि समय, लागत और जटिलता को कम करते हुए, अधिक सटीक और विस्तृत जानकारी एकत्र करने की क्षमता प्रदान करती है। हालांकि, इस विधि को लागू करने में जटिलताएँ भी हो सकती हैं, जैसे कि विभिन्न फ्रेम्स से डेटा का समन्वय और डेटा में किसी प्रकार का पक्षपात। बावजूद इसके, सही ढंग से योजना बनाकर और विभिन्न फ्रेमों के डेटा का समग्र विश्लेषण करके एकाधिक फ्रेम सर्वेक्षण एक मूल्यवान उपकरण सिद्ध हो सकता है, जो सर्वेक्षणों की सटीकता और विश्वसनीयता में वृद्धि करता है।

संदर्भ

काल्टन, जी., एंडरसन, डी.डब्ल्यू.। सैम्पलिंग रियर पॉपुलेशन्स। जर्नल ऑफ द रॉयल स्टैटिस्टिकल सोसाइटी, ए, 1986, 149: 65-82।

कुमार, आर., राय, ए., अहमद, टी., बिश्वास, ए., मूरी, पी. के.। रेस्केलिंग बूटस्ट्रैप टेक्निक फॉर वैरिएंस एस्टिमेशन इन ड्यूल फ्रेम सर्वे। जर्नल ऑफ द इंडियन सोसाइटी ऑफ एग्रीकल्चरल स्टैटिस्टिक्स, 2021, 75(2), 117-125।

दासगुप्ता, पी., अहमद, टी., राय, ए., बिश्वास, ए.। बूटस्ट्रैप वैरिएंस एस्टिमेशन टेक्निक अंडर ड्यूल फ्रेम रैंक सेट सैम्पलिंग। जर्नल ऑफ द इंडियन सोसाइटी ऑफ एग्रीकल्चरल स्टैटिस्टिक्स, 2019, 73(3), 197-206।

फुलर, डब्ल्यू.ए., बर्मेस्टर, एल.एफ। एस्टिमेटर्स फॉर सैम्पल्स सेलेक्टेड फ्रॉम टू ओवरलैपिंग फ्रेम्स। प्रोसीडिंग्स ऑफ द सोशल स्टैटिस्टिक्स सेक्शन, अमेरिकन स्टैटिस्टिकल एसोसिएशन, 1972, 245-249।

बैंकियर, एम.डी.। एस्टिमेटर्स बेस्ड ऑन सेवरल स्ट्रेटिफाइड सैम्पल्स विद एप्लिकेशंस टू मल्टीपल फ्रेम सर्वे। जर्नल ऑफ द अमेरिकन स्टैटिस्टिकल एसोसिएशन, 1986, 81, 1074-1079।

रनाली, एम.जी., आर्कोस, ए., रुएडा, एम.डी.एम., थियोडोरो, ए. । कैलिब्रेशन एस्टिमेशन इन ड्यूल फ्रेम सर्वे। स्टैट. मेथड्स एप्लाई., 2016, 25(3), 321-349।

रुएडा, एम.डी.एम., आर्कोस, ए., मोलिना, डी., रनाली, एम.जी.। एस्टिमेशन टेक्निक्स फॉर ऑर्डिनल डेटा इन मल्टीपल फ्रेम सर्वे विद कॉम्प्लेक्स सैम्पलिंग डिजाइन। इंटरनेशनल स्टैटिस्टिकल रिव्यू, 2017, 86(1), 51-67।

सक्सेना, बी.सी., नारायण, पी., श्रीवास्तव, ए.के. । मल्टीपल फ्रेम सर्वे इन टू स्टेज सैम्पलिंग। सांख्या बी, 1984, 46(1): 75-82।

स्किनर, सी.जे., राओ, जे.एन.के.। एस्टिमेशन इन ड्यूल फ्रेम सर्वे विद कॉम्प्लेक्स डिजाइन्स। जर्नल ऑफ द अमेरिकन स्टैटिस्टिकल एसोसिएशन, 1996, 91, 349-356।

स्किनर, सी.जे., होल्म्स, डी.जे., होल्ट, डी.। मल्टीपल फ्रेम सैम्पलिंग फॉर मल्टीवेरिएट स्ट्रेटिफिकेशन। इंटरनेशनल स्टैटिस्टिकल रिव्यू, 1994, 62, 333-347।

हार्टले, एच.ओ.। मल्टीपल फ्रेम मेथोडोलॉजी एंड सेलेक्टेड एप्लिकेशंस। सांख्या सी, 1974, 36, 99-118।

हार्टले, एच.ओ.। मल्टीपल फ्रेम सर्वे। प्रोसीडिंग्स ऑफ द सोशल स्टैटिस्टिक्स सेक्शन, अमेरिकन स्टैटिस्टिकल एसोसिएशन, 1962, 203-206।

फैमिली ऑफ लाइफटाइम डिस्ट्रिब्यूशन की विश्वसनीयता विशेषताओं के लिए प्रारंभिक परीक्षण अनुमानक और विश्वास अंतराल

हिमाद्री घोष एवं नितिन जोशी

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

हम विश्वसनीयता कार्य के दो मापों पर विचार करते हैं, अर्थात् $R(t) = P(X > t)$ और $P = (X > Y)$, जो कि जीवनकाल वितरण के परिवार को कवर करते हैं। इस परिवार के वितरण से रिकॉर्ड डेटा के आधार पर, हम न्यूनतम परिवर्तनशीलता वाले असंपूर्ण अनुमानक (UMVUE) और अधिकतम संभाव्यता अनुमानक (MLE) पर आधारित प्रारंभिक परीक्षण अनुमानक (PTEs) और प्रारंभिक परीक्षण विश्वास अंतराल (PTCIs) प्राप्त करते हैं।इसके अतिरिक्त, हमने MLE पर आधारित विश्वास अंतराल को अधिक सटीक पाया। लेखकों के ज्ञान के अनुसार, R(t)और P के लिए कोई PTEs रिकॉर्ड डेटा पर आधारित नहीं थे। इस पेपर में, हम MLE और UMVUE का उपयोग करके R(t)और P के लिए बेहतर PTEs को परिभाषित करते हैं। विभिन्न विधियों के तुलनात्मक अध्ययन से पता चलता है कि PTEs, सामान्य UMVUE और MLE की तुलना में बेहतर प्रदर्शन करते हैं।

मुख्य शब्द: फैमिली ऑफ लाइफटाइम डिस्ट्रिब्यूशन, रिकॉर्ड मूल्य, प्रारंभिक परीक्षण अनुमानक, प्रारंभिक परीक्षण विश्वास अंतराल, सिमुलेशन अध्ययन।

परिचय

सांख्यिकीय अनुमान में, हम अक्सर ऐसी समस्याओं का सामना करते हैं जहां मापदंडों पर कुछ पूर्व जानकारी (अक्सर बाधाओं के रूप में मानी जाती है) उपलब्ध होती है, जो प्रतिबंधित मॉडल को जन्म देती है। जैसा कि नाम से पता चलता है, प्रतिबंधित (अप्रतिबंधित) मॉडल से प्राप्त अनुमानकों को प्रतिबंधित (अप्रतिबंधित) अनुमानक के रूप में जाना जाता है। स्वाभाविक रूप से, एक प्रतिबंधित अनुमानक की वैधता संदिग्ध होगी और इसलिए प्रतिबंधों पर प्रारंभिक परीक्षण करना आवश्यक है। पिछले ज्ञान या अनुभव के कारण, प्रयोगकर्ता रुचि के कुछ मापदंडों पर प्रारंभिक अनुमान लगाने की स्थिति में हो सकता है। ऐसे मामलों में, हम मापदंडों पर पूर्व सूचना को शामिल करके एक बेहतर अनुमानक प्रदान कर सकते हैं। सांख्यिकीय वितरण के कुछ या सभी मापदंडों पर इस पूर्व जानकारी का उपयोग आमतौर पर एक बेहतर अनुमानात्मक अध्ययन की ओर ले जाता है। प्रतिबंधित मॉडल विश्लेषण की दक्षता और वैधता बाधाओं से प्रेरित प्रतिबंधित पैरामीट्रिक स्थान से अधिक है, जबकि संपूर्ण पैरामीट्रिक स्थान पर अप्रतिबंधित मॉडल विश्लेषण के लिए भी यही बात लागू होती है। इस प्रकार, प्रतिबंधित या अप्रतिबंधित मॉडल के परिणामस्वरूप होने वाला विश्लेषण दो अनुमानात्मक तकनीकों के बीच चयन करते समय बाधाओं की वैधता के मुकाबले दक्षता में कमी के अधीन हो सकता है। इसलिए जब हमें पूर्व सूचना पर पर्याप्त भरोसा हो तो प्रतिबंधित अनुमान प्रक्रियाओं को अपनाना बुद्धिमानी है। बैनक्रॉफ्ट (1944), सालेह और सेन (1978), सालेह और किब्रिया (1993), किब्रिया (2004), सालेह (2006), किब्रिया और सालेह (1993; 2004; 2005; 2006; 2010) और बेलाघी, अरशी और तबताबे (2014; 2015) ।

अब तक, अनुमानित विश्वसनीयता के साहित्य में, शोध कर्ताओं ने विभिन्न वितरणों के मापदंडों के पीटीई विकसित किए हैं। हालाँकि, लेखकों की सर्वोत्तम जानकारी के अनुसार, PTE विश्वसनीयता फंक्शन R(t) और P के लिए उपलब्ध नहीं हैं। वर्तमान पेपर में, हम पैरामीटर की शक्तियों और विश्वसनीयता फंक्शन के दो मापों के लिए PTE प्राप्त करते हैं, अर्थात् R(t) और P. विश्वसनीयता फंक्शन R(t) को समय t तक विफलता-मुक्त संचालन की संभावना के रूप में परिभाषित किया गया है। इस प्रकार, यदि यादृच्छिक चर (आरवी) एक्स किसी वस्तु या सिस्टम के जीवनकाल को दर्शाता है, तो $R(t) = P(X > t)$। तनाव-शक्ति सेटअप के तहत विश्वसनीयता का एक अन्य उपाय प्रायिकता पी = पी (एक्सझ वाई) है, जो यादृच्छिक तनाव वाई के अधीन किसी वस्तु या यादृच्छिक ताकत एक्स की प्रणाली की विश्वसनीयता का प्रतिनिधित्व करता है। साहित्य में बहुत काम किया गया है रिकॉर्ड मूल्यों के आधार पर आर(टी) और पी के बिंदु अनुमान और परीक्षण के लिए। संक्षिप्त समीक्षा के लिए, कोई भी चतुर्वेदी और मल्होत्रा (2016,2017ए) का संदर्भ ले सकता है। अब तक, विभिन्न वितरणों के मापदंडों के पीटीईएस को अनुमानित विश्वसनीयता के साहित्य में कई शोधकर्ताओं द्वारा विकसित किया गया है। इस पेपर में, विश्वसनीयता के दो उपाय फंक्शन, R(t) और P, UMVUE और MLE पर आधारित हैं, वहां PTES विकसित किए गए हैं। साहित्य में कुछ काम किया गया है जहां बूर XII मॉडल बेलाघी, अरशी और तबताबे (2014; 2015) के लिए रिकॉर्ड मूल्यों के आधार पर प्रारंभिक परीक्षण अनुमानक और प्रारंभिक विश्वास अंतराल विकसित किए गए थे, तीन मापदंडों के लिए विश्वसनीयता विशेषताओं के रिकॉर्ड प्रारंभिक परीक्षण अनुमानकों के आधार पर बूर XII वितरण (चतुर्वेदी, ए. और मल्होत्रा, ए.(2017बी)। इस पेपर में, आजीवन वितरण के परिवार की शुरुआत की गई है चतुर्वेदी और कुमारी द्वारा (2017सी) पर विचार किया गया है।

विधि

इस खंड में, हम रिकॉर्ड मूल्य के आधार पर प्रारंभिक परीक्षण अनुमान के प्रदर्शन का आकलन करने का प्रयास कर रहे हैं। इसलिए हम $G(x;a,\underline{\theta}) = x^P$ को प्रतिस्थापित करके वेइबुल वितरण को आजीवन वितरण के परिवार के एक विशेष मामले के रूप में मानते हैं। चूंकि सापेक्ष दक्षता $\left(\hat{\lambda}^p_{PT-ML} \mid \hat{\lambda}^p_{ML}\right)$ और $e\left(\hat{\lambda}^p_{PT-U} \mid \hat{\lambda}^p_U\right)$d नमूना आकार (n+1) और महत्व के स्तर α पर निर्भर करता है। तालिका 1 $\hat{\lambda}^p_{PT-ML}$ के ऊपर $\hat{\lambda}^p_{ML}$ की सापेक्ष दक्षता दिखाती है और हम देखते हैं कि एक अंतराल δ मौजूद है जिसके लिए दक्षता 1 से अधिक है इसी प्रकार तालिका 2 λ^p_U के ऊपर λ^p_U की सापेक्ष दक्षता दिखाती है और हम देखते हैं एक अंतराल मौजूद है δ जिसके लिए दक्षता 1 से अधिक है।

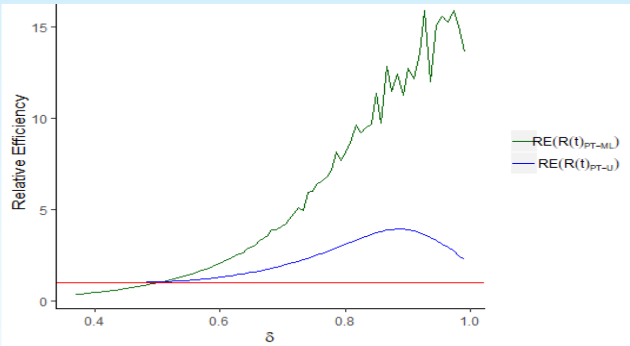
तालिका 1: n के विभिन्न नमूना आकार और महत्व के स्तर α के लिए λ^p के MLE पर MLE के आधार पर λ^p की PTE की सापेक्ष दक्षता जब $\lambda = \lambda_0$ =25 और p=6

α n	0.01	0.05	0.10	0.20
5	34.2381	9.3839	5.4899	3.2609
10	49.4901	12.5478	7.0294	3.9653
15	62.4767	15.2044	8.3343	4.5863
20	79.5035	17.9214	9.6983	5.2316
40	135.7626	30.8198	16.3142	8.6420
60	207.0365	46.1964	24.2528	12.7360
90	325.3012	71.6876	37.4159	19.5292

तालिका 2: n के विभिन्न नमूना आकार और महत्व के स्तर α के लिए λ^p के UMVUE पर UMVUE के आधार पर λ^p की PTE की सापेक्ष दक्षता जब $\lambda = \lambda_0$ = 25 और p=6

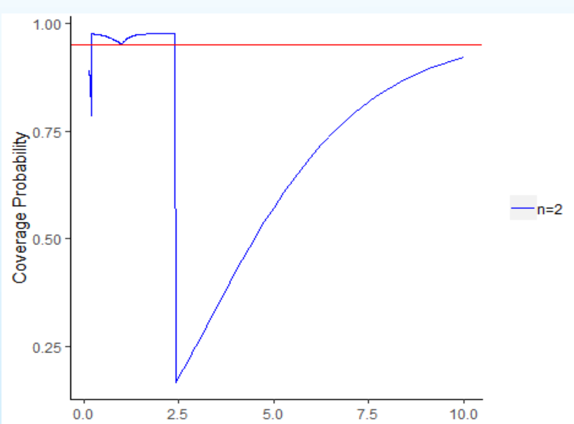
α n	0.01	0.05	0.10	0.20
5	18.4327	6.1332	2.1010	1.4670
10	20.7466	6.1742	2.1169	1.7306
15	23.1546	6.2083	2.1302	1.8549
20	24.7819	6.2371	2.1414	1.9272
40	27.7466	6.3332	2.1788	1.0079
60	29.1546	6.3875	2.1997	10983
90	30.7819	6.5656	3.0010	1.1227

तालिका 1 और 2 से हम देखते हैं कि नमूना आकार और महत्व के स्तर के बावजूद, एमएलई और यूएमवीयूई पर आधारित λ^p का पीटीई हमेशा अधिक कुशल होता है।

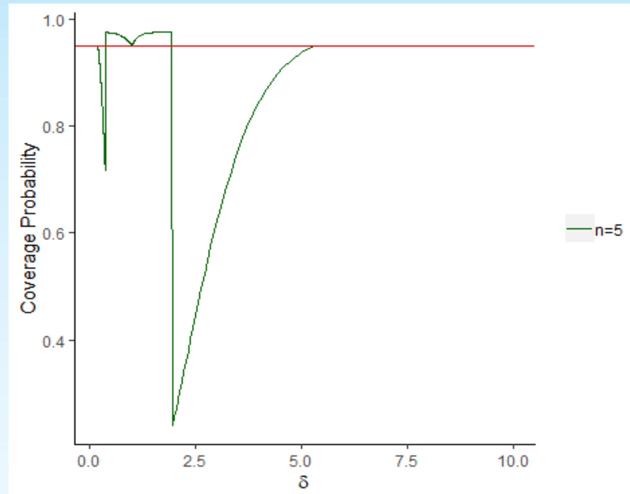


चित्र 1: सापेक्ष दक्षता $\hat{R}(t)_{PT-ML}$ की $\hat{R}(t)$ और $\tilde{R}(t)_{PT-U}$ की $\tilde{R}(t)$ के उपर $\delta=\frac{\lambda_0}{\lambda}$ के संबंध में

चित्र 1: $\hat{R}(t)_{PT-ML}$ की $\hat{R}(t)$ के सापेक्ष दक्षता और $\tilde{R}(t)_{PT-U}$ की $\tilde{R}(t)_{PT-U}$ के सापेक्ष दक्षता को विभिन्न महत्व स्तर α और बिंदु अनुमान के लिए दर्शाता है। दक्षता का विश्लेषण $\delta=\frac{R_0(t)}{R(t)}$ के संदर्भ में किया गया है। परिणाम यह इंगित करते हैं कि $\tilde{R}(t)$ के प्रारंभिक परीक्षण प्राक्कलन, जो अधिकतम संभावना प्राक्कलन (MLE) और समान रूप से न्यूनतम विचलन निष्पक्ष प्राक्कलन (UMVUE) का उपयोग करके प्राप्त किए गए हैं, λ के विशिष्ट अंतरालों में $\tilde{R}(t)$ की तुलना में बेहतर प्रदर्शन करते हैं।



चित्र 2: α=0.05 और n=2 के लिए $\delta=\frac{\lambda_0}{\lambda}$ के संबंध में λ के PTCI की कवरेज संभावना और

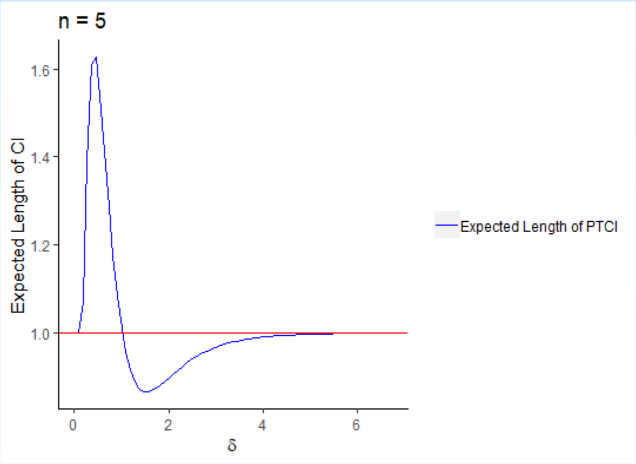


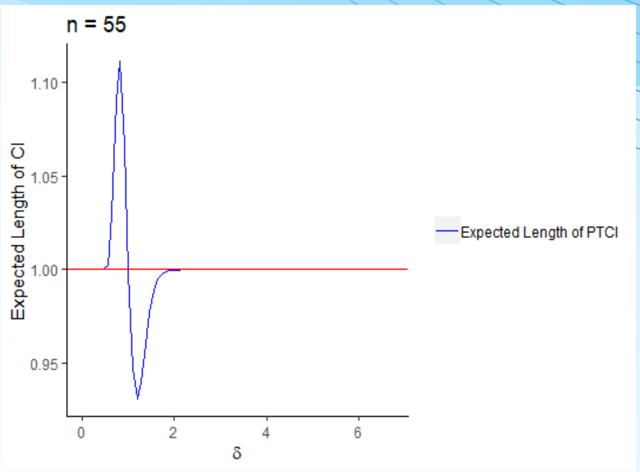
चित्र 3: α=0.05 और n=5 के लिए $\delta=\frac{\lambda_0}{\lambda}$ के संबंध में λ के PTCI की कवरेज संभावना क्रमशः।

चित्र 2 और 3 में, हमने α=0.05 और n=2 और n=5 के लिए $\delta=\frac{\lambda_0}{\lambda}$ के संबंध में λ के PTCI की कवरेज संभावना दिखाई है, हम देखते हैं कि δ का मान 0 से ∞ तक जाता है। PTCI का CP हमेशा $1-\alpha$ से अधिक होता है। यह वर्चस्व अंतराल छोटे नमूना आकारों के लिए बड़ा है। इस प्रकार, हम यह निष्कर्ष निकाल सकते हैं कि 1 के आसपास एक विशिष्ट अंतराल में δ के कुछ मूल्यों के लिए λ के PTCI का CP, ETCI के CP से अधिक है। चित्र 4 और 5 नीचे दिए गए हैं,

चित्र 4 और 5 में, हमने α=0.05 और n=5 और n=55 के लिए $\delta=\frac{\lambda_0}{\lambda}$ के संबंध में ETCI बनाम λ की PTCI की अपेक्षित लंबाई दिखाई है। δ के संबंध में PTCI बनाम ETCI के स्केल किए गए EL पर विचार किया गया है, जिसके लिए λ के PTCI का EL ETCI से कम है। δ का अंतराल जिसके लिए λ के PTCI का EL, λ के ETCI की तुलना में कम है, नमूना आकार में वृद्धि के साथ घटता जाता है। इसके अलावा, हम यह भी ध्यान देते हैं कि जैसे-जैसे अंतराल की ऊपरी सीमा ∞ की ओर बढ़ती है, λ के PTCI का EL λ के ETCI के EL के करीब होता जाता है।

वास्तविक डेटा पर एक उदाहरण





चित्र 4: α=0.05 और n=5 के लिए $\delta=\frac{\lambda_0}{\lambda}$ के संबंध में λ के PTCI के साथ ETCI की अपेक्षित लंबाई और चित्र 5: α=0.05 और n=5 के लिए $\delta=\frac{\lambda_0}{\lambda}$ के संबंध में λ के PTCI के साथ ETCI की अपेक्षित लंबाई α=0.05 और n=55 क्रमशः ।

हमने भारतीय कृषि अनुसंधान परिषद, केंद्रीय समुद्री मत्स्य पालन अनुसंधान संस्थान के वास्तविक आंकड़ों पर विचार किया है। हम यह जांचने के लिए कोलमोगोरोव-स्मिरनोव (के-एस) परीक्षण लागू करते हैं कि क्या समुद्री मछली डेटा में वेइबुल वितरण है या नहीं, संपूर्ण डेटा के लिए लाइफटाइम मॉडल से वेइबुल वितरण पर विचार करते हुए, इसका के-एस आँकड़ा 0.087 है जिसका पी-वैल्यू 0.8213 है जो इंगित करता है कि डेटा लाइफटाइम वितरण के परिवार में फिट बैठता है। . पुनरावृत्त एल्गोरिदम का उपयोग करके, हमने अधिकतम संभावना अनुमानक $\hat{\lambda}_{ML}$=0.3558 प्राप्त किया। परिकल्पना पर विचार करते हुए

$$H_0{:}\lambda = 0.35$$
$$H_1{:}\lambda \neq 0.35$$

परिकलित परीक्षण-सांख्यिकी $\frac{2G(R_n;a,\underline{\theta})}{\lambda_0}$=5.7142 जो विश्वास अंतराल [5.62862, 6.1189] में निहित है। इस प्रकार, हम 5% महत्व के स्तर पर शून्य परिकल्पना को अस्वीकार नहीं करते हैं जो इंगित करता है कि $\hat{\lambda}_{PT-ML} = 0.35$ ।

निष्कर्ष

इस पेपर में, हमने आजीवन वितरण के परिवार के लिए मापदंडों, आर (टी) और पी की शक्तियों का आकलन करने के लिए विभिन्न पीटीई का प्रदर्शन किया। सभी पीटीई के पूर्वाग्रह और माध्य वर्ग त्रुटि अभिव्यक्ति प्राप्त की जाती हैं। संख्यात्मक निष्कर्षों से यह पता चलता है कि प्रस्तावित पीटीई शून्य परिकल्पना के पड़ोस में संबंधित सामान्य अनुमानकों जैसे यूएमवीयूई और एमएलई से बेहतर प्रदर्शन करते हैं। ईटीसीआई की तुलना में पीटीसीआई की कवरेज संभावना अधिक है और अपेक्षित लंबाई कम है। इस प्रकार, हम मापदंडों के बेहतर अनुमानक और विश्वास अंतराल स्थापित करने में सक्षम थे।

संदर्भ

अहसानुल्लाह, एम. (1980)। दो पैरामीटर घातीय वितरण के लिए रिकॉर्ड मानों की रैखिक भविष्यवाणी। सांख्यिकी गणित संस्थान के इतिहास, 32(3), 363-368।

अरशी, एम., और इमादी, एम. (2008)। रिकॉर्ड डेटा और अंतर-रिकॉर्ड समय के आधार पर साक्ष्यात्मक अनुमान। सांख्यिकीय कागजात, 49(2), 291-301।

अर्नोल्ड, बी.सी., बालाकृष्णन, एन., और नागराजा, एच.एन. (1992)। ऑर्डर स्टैटिस्टिक्स में पहला कोर्स, खंड 54। जॉन विली संस।

बालाकृष्णन, एन., अहसानुल्लाह, एम., और चान, पी. एस. (1995)। लॉजिस्टिक रिकॉर्ड मूल्यों और संबंधित अनुमान पर। एप्लाइड स्टैटिस्टिकल साइंस जर्नल,

2(3), 233-248।

बैनक्रॉफ्ट, टी.ए. (1944)। महत्व के प्रारंभिक परीक्षणों के उपयोग के कारण अनुमान में पूर्वाग्रह पर। गणितीय सांख्यिकी का इतिहास, 15, 190-204।

बेलाघी, आर.ए., अरशी, एम., और तबताबेय, एस.एम.एम. (2014)। रिकॉर्ड मूल्यों के आधार पर बूर XII मॉडल के स्केल पैरामीटर के लिए बेहतर विश्वास अंतराल। कम्प्यूटेशनल सांख्यिकी, 29, 1153-1173।

बेलाघी, आर.ए., अरशी, एम., और तबताबेय, एस.एम.एम. (2015)। बूर XII मॉडल के लिए रिकॉर्ड मूल्यों के आधार पर प्रारंभिक परीक्षण अनुमानक के निर्माण पर सांख्यिकी-सिद्धांत और विधियों में संचार, 44(1), 1-23।

बूर, आई.डब्ल्यू. (1942)। संचयी आवृत्ति फंक्शंस। गणितीय सांख्यिकी का इतिहास, 13, 215-232।

बूर, आई.डब्ल्यू., और सिस्लाक, पी.जे. (1968)। वितरण की एक सामान्य प्रणाली पर ५ इसकी वक्र-आकार विशेषताएँ II. नमूना माध्यिका. जर्नल ऑफ द अमेरिकन स्टैटिस्टिकल एसोसिएशन, 63(322), 627-635।

चांडलर, के.एन. (1952)। रिकॉर्ड मानों का वितरण और आवृत्ति. रॉयल स्टैटिस्टिकल सोसाइटी का जर्नल: सीरीज बी (मेथडोलॉजिकल), 14, 220-228।

मशीन लर्निंग तकनीकों का उपयोग करके फसल उपज का अनुमान

पंकज दास, भारती, राहुल बनर्जी, अंकुर बिश्वास, तौकीर अहमद, कौस्तव आदित्य
भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली

सार

फसल उपज का अनुमान आधुनिक कृषि का एक महत्वपूर्ण पहलू है, और मशीन लर्निंग मॉडल्स का संघीकरण इन अनुमानों की सटीकता और दक्षता को बढ़ाने में प्रभावी साबित हुआ है। फसल की पैदावार को प्रभावित करने वाले कई कारकों का विश्लेषण करने के लिए विभिन्न मशीन लर्निंग (एमएल) मॉडल का उपयोग किया जाता है, जिससे किसानों को सूचित निर्णय लेने और अपनी कृषि पद्धतियों को सुधारने में मदद मिलती है। मशीन लर्निंग मॉडल्स ऐतिहासिक डेटा, सैटेलाइट इमेजरी, और रियल-टाइम सेंसर डेटा का उपयोग करके फसल उपज का अनुमान लगाने में महत्वपूर्ण भूमिका निभाते हैं। इन मॉडल्स का एकीकरण किसानों को सटीक भविष्यवाणियाँ प्रदान करता है, जिससे वे समय पर उचित हस्तक्षेप कर सकते हैं और संसाधनों का अधिकतम उपयोग कर सकते हैं, जो अंततः कृषि उत्पादकता और स्थिरता में सुधार करता है।

मुख्य शब्दः मशीन लर्निंग; फसल उपज अनुमान; कृषि उत्पादकता

परिचय

कृषि वैश्विक अर्थव्यवस्था का एक महत्वपूर्ण आधार है, जो अरबों लोगों के लिए आजीविका प्रदान करती है और विश्व भर में खाद्य आपूर्ति सुनिश्चित करती है। भारत जैसे कृषि प्रधान देश में लगभग 60% ग्रामीण परिवार कृषि पर अपनी आय का मुख्य स्रोत निर्भर करते हैं। दुनिया की जनसंख्या वृद्धि के साथ-साथ खाद्य खपत में तेजी से वृद्धि होने की संभावना है, जिससे कृषि प्रणालियों पर सततता और उत्पादकता में सुधार के लिए दबाव और बढ़ रहा है। फसल उपज का सटीक अनुमान लगाना, इन चुनौतियों से निपटने के लिए एक महत्वपूर्ण आवश्यकता बन गया है, क्योंकि यह हितधारकों को अवगत निर्णय लेने और प्रभावी रूप से योजना बनाने में मदद करता है।

फसल उपज और फसल उपज अनुमान का महत्व

फसल उपज (Crop Yield) का मतलब है किसी निश्चित कृषि भूमि पर उगाई गई फसल का उत्पादन, जिसे आमतौर पर एक निर्धारित क्षेत्र (जैसे एक हेक्टेयर) से मापा जाता है। यह कृषि उत्पादकता का एक महत्वपूर्ण मापदंड है और खाद्य सुरक्षा, आर्थिक स्थिरता, और पर्यावरणीय सततता का एक आवश्यक संकेतक होता है। उच्च फसल उपज से भूख और गरीबी में कमी आती है, किसानों की आय बढ़ती है, और वैश्विक व्यापार को समर्थन मिलता है। इसके विपरीत, कम उपज से खाद्य संकट, आर्थिक अस्थिरता, और प्राकृतिक संसाधनों पर दबाव बढ़ सकता है, क्योंकि अधिक भूमि को कृषि कार्यों के लिए उपयोग में लाने की कोशिश की जाती है। मौसम, मिट्टी, और कृषि प्रबंधन तकनीकों में भौगोलिक भिन्नताएँ फसल उपज में बदलाव का कारण बनती हैं। इसलिए, सीजन के दौरान फसल उपज का अनुमान किसानों को उत्पादकता बढ़ाने के लिए समय पर कदम उठाने में मदद कर सकता है और प्रभावी योजना विकसित करने के लिए सरकारी संगठनों के लिए सहायक हो सकता है। सटीक फसल उपज अनुमान से विभिन्न हितधारकों को लाभ होता है, जिनमें शामिल हैं:

1. **आर्थिक विकासः** कृषि कई देशों के जी.डी.पी. में महत्वपूर्ण योगदान देती है। उच्च उपज से किसानों

की आय बढ़ती है और कृषि अर्थव्यवस्था मजबूत होती है। किसानों की आय में वृद्धि से ग्रामीण क्षेत्रों में जीवन स्तर में सुधार होता है, गरीबी में कमी आती है। एक समृद्ध कृषि क्षेत्र न केवल खेतों पर बल्कि परिवहन, खाद्य प्रसंस्करण, और कृषि मशीनरी जैसे संबंधित उद्योगों में भी रोजगार सृजित करता है। उच्च कृषि उत्पादकता वाले देश अपनी अतिरिक्त फसलें निर्यात कर सकते हैं, जिससे उनकी अर्थव्यवस्था और विदेशी मुद्रा भंडार में वृद्धि होती है।

2. **खाद्य सुरक्षाः** सटीक उपज अनुमान पर्याप्त खाद्य आपूर्ति सुनिश्चित करता है, खाद्य पदार्थों की कमी को रोकने में मदद करता है, और संसाधनों के प्रभावी वितरण को सक्षम बनाता है। बढ़ी हुई पैदावार वर्तमान और भविष्य में आबादी के लिए अधिक भोजन उपलब्ध कराती है। सटीक उपज अनुमान अचानक खाद्य संकट या मूल्य वृद्धि को रोकने में मदद करता है, जो अप्रत्याशित घटनाओं जैसे सूखा, रोग या कीट संक्रमण के कारण हो सकता है।

3. **पर्यावरणीय सततताः** फसल की पैदावार का अनुकूलन पानी, उर्वरक और कीटनाशकों के अत्यधिक उपयोग को कम कर सकता है, जिससे पर्यावरण प्रदूषण और पारिस्थितिकी तंत्र में गिरावट को रोका जा सकता है। पहले से मौजूद फसल भूमि पर पैदावार बढ़ाकर जंगलों, घास के मैदानों और आर्द्रभूमि को कृषि भूमि में बदलने से बचा जा सकता है। पैदावार में सुधार करने वाली सतत कृषि पद्धतियाँ ग्रीनहाउस गैस उत्सर्जन को कम करने और मिट्टी में कार्बन संधारण को बढ़ाने में भी मदद कर सकती हैं। यह जैव विविध ता और महत्वपूर्ण पारिस्थितिक तंत्र की रक्षा करता है।

4. **नीति योजनाः** सरकारें और संगठन उपज अनुमानों पर निर्भर करते हैं ताकि वे आपातकालीन स्थितियों के लिए योजना बना सकें, संसाधनों का आवंटन कर सकें, और कृषि नीतियाँ विकसित कर सकें। सटीक उपज डेटा सरकारों को कृषि सब्सिडी, व्यापार नीतियों, और खाद्य भंडारों के बारे में अवगत निर्णय लेने की अनुमति देता है। उपज अनुमान प्राकृतिक आपदाओं के कारण संभावित खाद्य संकटों के लिए योजना बनाने में मदद करते हैं, जिससे समय पर हस्तक्षेप और सहायता वितरण संभव होता है।

संक्षेप में, फसल उपज एक महत्वपूर्ण कारक है जो न केवल हमारी खाद्य आपूर्ति बल्कि हमारी अर्थव्यवस्था, पर्यावरण और समग्र कल्याण को भी प्रभावित करती है। फसल की पैदावार बढ़ाने और सटीक अनुमान लगाने पर ध्यान केंद्रित करके, सभी के लिए अधिक सुरक्षित, समृद्ध और टिकाऊ भविष्य बनाया जा सकता है ।

मशीन लर्निंग

मशीन लर्निंग (एमएल) कृत्रिम बुद्धिमत्ता (एआई) का एक प्रमुख घटक है, जो कंप्यूटर को बिना किसी विशेष प्रोग्रामिंग के डेटा से सीखने और अनुभव के आधार पर निर्णय लेने की क्षमता प्रदान करता है। इसका अर्थ है कि मशीन लर्निंग एल्गोरिदम कंप्यूटर को बड़े पैमाने पर डेटा से पैटर्न और जानकारी पहचानने में सक्षम बनाते हैं, जिससे वे समय के साथ और अधिक सटीक और प्रभावी बन सकते हैं। मशीन लर्निंग जटिल समस्याओं को हल करने के लिए मानव मस्तिष्क की तरह नकल करने की शक्ति प्रदान करके किसी मॉडल के प्रदर्शन में सुधार करती है। मशीन लर्निंग मॉडल्स ज्यादातर डेटा-आधारित, स्व-लवणशील और गैर-रैखिक होते हैं। छिपे हुए डेटा पैटर्न की पहचान करने और अनुभव के साथ सुधार करने की क्षमता के कारण इन मॉडलों ने चिकित्सा, डेटा विज्ञान आदि जैसे विभिन्न क्षेत्रों में महत्वपूर्ण महत्व प्राप्त किया। मशीन लर्निंग एल्गोरिदम पैटर्न की पहचान करते हैं, अनुमान लगाते हैं और अनुभव तथा अधिक डेटा के संपर्क के माध्यम से समय के साथ अपने प्रदर्शन में सुधार करते हैं। इस सीखने की प्रक्रिया में एक डेटासेट पर एक मॉडल को प्रशिक्षित करना शामिल है, जो इसे अंतर्निहित संबंधों को पहचानने और नए, अनदेखे डेटा पर सही निर्णय लेने की अनुमति देता है। मशीन लर्निंग मॉडल्स इमेज पहचान, भाषा प्रसंस्करण, और धोखाधड़ी पहचान जैसी कार्यों को स्वचालित करते हैं। इसके प्रमुख प्रकारों में सुपरवाइज़्ड लर्निंग (लेबल डेटा का उपयोग),

अनसुपरवाइज्ड लर्निंग (बिना लेबल डेटा में पैटर्न की पहचान), और रिइंफोर्समेंट लर्निंग (परीक्षण और त्रुटि के माध्यम से सीखना) शामिल हैं।

एमएल मॉडल निर्माण की प्रक्रिया

मशीन लर्निंग मॉडल निर्माण एक बहु-चरणीय प्रक्रिया है। यह समस्या को स्पष्ट रूप से परिभाषित करने और मापनीय उद्देश्यों को सेट करने से शुरू होती है। इसके बाद, संबंधित डेटा एकत्रित किया जाता है, डेटा क्लीनिंग की जाती है, अन्वेषण किया जाता है और अप्रत्याशितताओं को संभालने के लिए पूर्व-प्रसंस्कृत किया जाता है। इसके बाद फीचर इंजीनियरिंग होती है, जहां संबंधित फीचर्स का चयन और रूपांतरण किया जाता है ताकि मॉडल के प्रदर्शन में सुधार किया जा सके। एक उपयुक्त मशीन लर्निंग मॉडल चुना जाता है, जो डेटा के एक भाग पर प्रशिक्षित होता है, और उसके हाइपरपैरामीटर को सर्वोत्तम परिणामों के लिए समायोजित किया जाता है। प्रत्येक मॉडल के विभिन्न हाइपरपैरामीटर होते हैं। फिर मॉडल की सटीकता और सामान्यीकरण सुनिश्चित करने के लिए मॉडल के प्रदर्शन का अलग-अलग डेटा पर मूल्यांकन और सत्यापन किया जाता है। अंत में, प्रशिक्षित मॉडल को वास्तविक परिस्थिति में उपयोग में लाया जाता है और समय के साथ इसकी प्रभावशीलता बनाए रखने के लिए नियमित रूप से निगरानी की जाती है। इसके लिए अक्सर नए डेटा और अपडेट के साथ पुनः प्रशिक्षण की आवश्यकता होती है।

फसल उपज अनुमान के लिए प्रमुख विशेषताएँ: मशीन लर्निंग मॉडल्स जो फसल उपज अनुमान के लिए उपयोग किए जाते हैं, विभिन्न विशेषताओं पर निर्भर करते हैं, जिनमें शामिल हैं:

1. मौसम डेटा: तापमान, वर्षा, आर्द्रता, इत्यादि ।
2. मिट्टी के गुण: pH, जैविक पदार्थ, पोषक तत्व स्तर, मिट्टी का प्रकार, इत्यादि।
3. फसल प्रबंधन पद्धतियाँ: सिंचाई, उर्वरक का उपयोग, कीट नियंत्रण विधियाँ, इत्यादि।
4. रिमोट सेंसिंग डेटा: सैटेलाइट इमेजरी और ड्रोन डेटा जो NDVI (नॉर्मलाइज्ड डिफरेंस वेजिटेशन इंडेक्स) जैसे वेजिटेशन इंडेक्स प्रदान करता है।
5. ऐतिहासिक उपज डेटा: प्रवृत्ति विश्लेषण के लिए पिछली फसल की उपज की जानकारी।

फसल उपज अनुमान के लिए मशीन लर्निंग तकनीकें

फसल की उपज फसल की पैदावार भौतिक, आर्थिक और तकनीकी जैसे कई कारकों से प्रभावित होती है। ये कारक उपज में वृद्धि और कमी दोनों में महत्वपूर्ण भूमिका निभाते हैं (दास एट अल., 2023)। एक अच्छा पूर्वानुमान मॉडल विभिन्न कारकों और उपज के बीच जटिल संबंधों का पता लगाता है। यह प्रबंधन तकनीकों को सुधारने और वास्तविक उपज को बढ़ाने में मदद करता है। एक अच्छा पूर्वानुमान मॉडल विश्वसनीय, सुसंगत, उद्देश्यपूर्ण, लागत-प्रभावी और चरम घटनाओं के प्रति संवेदनशील होना चाहिए (दास एट अल., 2023; भारती एट अल., 2023)। फसल उपज का अनुमान एक जटिल प्रक्रिया है, क्योंकि इसका व्यवहार कई कारकों पर निर्भर करता है। कई शोधकर्ताओं ने सरल सहसंबंध, पाथ एनालिसिस, एकाधिक रैखिक प्रतिगमन, चरणबद्ध प्रतिगमन, फकटोरियल एनालिसिस और परिंसिपल कम्पोनेंट एनालिसिस जैसे विभिन्न मॉडलों का उपयोग करके फसल की उपज का मॉडल बनाने का प्रयास किया है। ये अध्ययन इस धारणा पर आधारित हैं कि पौधों के गुणों और फसल की उपज के बीच संबंध रैखिक है। हालांकि, ये मॉडल फसल की उपज और उसके कारकों के बीच गैर-रैखिक संबंध को समझने में सफल नहीं रहे हैं। परंपरागत तरीके जो फसल उपज का अनुमान लगाने के लिए उपयोग किए जाते हैं, अक्सर खेतों में सर्वेक्षण और विशेषज्ञों के आकलन पर निर्भर करते हैं, जो समय-साध्य, श्रम-गहन और त्रुटि से प्रभावित हो सकते हैं। मशीन लर्निंग (एमएल) इन चुनौतियों को दूर करने के लिए एक शक्तिशाली उपकरण के रूप में उभरा है, जो विशाल डेटा सेट का विश्लेषण करके सटीक पूर्वानुमान करता है। मौसम पैटर्न, मिट्टी की स्थिति और रिमोट सेंसिंग इमेजरी जैसे विभिन्न डेटा स्रोतों का उपयोग करके, एमएल तकनीकें

उन जटिल पैटर्न और संबंधों को उजागर कर सकती हैं जो पहले अप्राप्य थे।

फसल उपज अनुमान में उपयोग किए जाने वाले कुछ सामान्य एमएल मॉडल्स

1. **रैंडम फॉरेस्ट (Random Forest):** रैंडम फॉरेस्ट एक एन्सेम्बल लर्निंग तकनीक है, जो प्रशिक्षण के दौरान कई डिसिशन ट्री बनाती है। अंतिम पूर्वानुमान प्रत्येक ट्री के अनुमान का औसत लेकर किया जाता है। रैंडम फॉरेस्ट मॉडल बड़े डेटा सेट्स को संभालने और विभिन्न कारकों के बीच जटिल संबंधों को पहचानने में सक्षम होते हैं, जैसे मौसम की स्थिति, मिट्टी की गुणवत्ता, ऐतिहासिक फसल डेटा, आदि (प्रसाद एट अल., 2021; ढिल्लों एट अल., 2023)।
2. **सपोर्ट वेक्टर मशीन (SVM):** सपोर्ट वेक्टर मशीन एक सुपरवाइज्ड लर्निंग मॉडल है जो विशेष रूप से वर्गीकरण और रिग्रेशन विश्लेषण के लिए डेटा का विश्लेषण करता है। यह हाइपरप्लेन को ढूंढता है जो एक डेटासेट को सबसे प्रभावी तरीके से वर्गीकृत करता है या सतत परिणामों का अनुमान लगाता है। SVM का उपयोग फसल उपज अनुमान में तापमान, वर्षा और पोषक तत्वों के स्तर जैसे विभिन्न इनपुट फीचर्स को मानचित्रित करने के लिए किया जा सकता है (दास एट अल., 2023; डांग एट अल., 2021)।
3. **न्यूरल नेटवर्क्स (Neural Networks):** न्यूरल नेटवर्क्स, विशेष रूप से डीप लर्निंग मॉडल्स, इंटरकनेक्टेड नोड्स के लेयर से मिलकर बनता है जो मानव मस्तिष्क की संरचना की नकल करते हैं। ये मॉडल डेटा में जटिल पैटर्न और संबंधों को पहचानने और सीखने में सक्षम होते हैं। गैर-रेखीय संबंधों को संभालने और बड़ी मात्रा में डेटा को संसाधित करने की क्षमता के कारण न्यूरल नेटवर्क फसल उपज अनुमान के लिए मूल्यवान हैं। जब विभिन्न चरों के बीच जटिल अंत:क्रियाएँ होती हैं तो वे बेहतर प्रदर्शन करते हैं(दास एट अल., 2023; भारती एट अल., 2023)।
4. **डिसिशन ट्री (Decision Trees):** डिसिशन ट्री विभिन्न स्थितियों के आधार पर डेटासेट को छोटे उपसमूहों में तोड़ देते हैं। इनका उपयोग वर्गीकरण और प्रतिगमन दोनों कार्यों के लिए किया जाता है। विभिन्न कारकों जैसे कि मौसम पैटर्न, मिट्टी की स्थिति और फसल प्रकार आदि को ध्यान में रखते हुए, डिसिशन ट्री का उपयोग फसल उपज अनुमान में किया जा सकता है(भटनागर और गोहेन, 2020)।
5. **लीनियर रिग्रेशन** (Linear Regression)**:** लीनियर रिग्रेशन मॉडल्स इनपुट फीचर्स और लक्षित चर के बीच रैखिक संबंध स्थापित करते हैं। ये विशेष रूप से फसल उपज पर व्यक्तिगत कारकों के प्रभाव को समझने के लिए उपयोगी होते हैं। फसल की उपज, तापमान, वर्षा और ऐतिहासिक फसल उपज जैसे कारकों के बीच रैखिक सहसंबंध का विश्लेषण करके फसल की उपज का अनुमान लगाने के लिए रैखिक प्रतिगमन का उपयोग किया जा सकता है। (Das et al., 2023; Bharti et al., 2023)।
6. **ग्रेडिएंट बूस्टिंग** (Gradient Boosting)**:** ग्रेडिएंट बूस्टिंग एक एन्सेम्बल लर्निंग तकनीक है, जो क्रमिक रूप से वीक लर्नर (आमतौर पर डिसिशन ट्री) की एक श्रृंखला बनाती है, जिसमें प्रत्येक अपने पूर्ववर्ती की त्रुटियों को सुधारता है। जब कई कारकों के प्रभाव को समझने और प्रत्येक पुनरावृत्ति के साथ अनुमानों को परिष्कृत करने की आवश्यकता होती है, तो ग्रेडिएंट बूस्टिंग मॉडल फसल उपज अनुमान के लिए प्रभावी होते हैं (अरुमुगम एट अल, 2021)।
7. **K-Nearest Neighbors (KNN):** KNN एक सरल और सहज एल्गोरिदम है जो डेटा बिंदु को उसके नेबरस के वर्गीकृत के आधार पर श्रेणीबद्ध करता है। करंट ग्रोइंग परिस्थितियों और ऐतिहासिक आंकड़ों के बीच समानता पर विचार करके, KNN को फसल उपज अनुमान के लिए उपयोग किया जा सकता है (मेदार, और राजपुरोहित, 2014)।
8. **डीप लर्निंग मॉडल्स (Deep Learning Models):** डीप लर्निंग मॉडल ने डेटा में जटिल पैटर्न और

संबंधों को समझने में उल्लेखनीय क्षमताएं दिखाई हैं। जजब इन्हें फसल उपज अनुमान के लिए लागू किया जाता है, तो ये विभिन्न कारकों के बीच जटिल निर्भरताओं को पकड़ने में सक्षम होते हैं। कन्वेन्शनल न्यूरल नेटवर्क्स (CNNs), रिकरंट न्यूरल नेटवर्क्स (RNNs), लॉन्ग शॉर्ट-टर्म मेमोरी नेटवर्क्स (LSTMs), ऑटोएन्कोडर्स, जेनरेटिव एडवरसैरियल नेटवर्क्स (GANs) कुछ मॉडल हैं जिनका उपयोग फसल उपज अनुमान के लिए किया जा सकता है (खाकी और वांग, 2019; ह्यूबर एट अल., 2022)।

फसल उपज अनुमान में मशीन लर्निंग के अनुप्रयोग

1. **सटीक कृषि (Precision Agriculture):** मशीन लर्निंग कृषि क्षेत्र में क्रांति ला रहा है, जो विभिन्न कृषि कार्यों में डेटा-आधारित निर्णय-निर्माण और स्वचालन को संभव बनाता है। फसल की बीमारियों और कीटों की प्रारंभिक पहचान इमेज रिकग्निशन और प्रेडिक्टिव मॉडलिंग के माध्यम से संभव हो रही है, जो नुकसान को कम करती है। मशीन लर्निंग एल्गोरिदम विभिन्न डेटा स्रोतों जैसे मिट्टी की स्थिति और मौसम पैटर्न का विश्लेषण करके फसल की पैदावार का प्रभावी अनुमान लगा सकते हैं, जिससे संसाधनों का आवंटन अधिक कुशल हो जाता है। स्वचालित सिंचाई प्रणालियों में, वास्तविक समय का डेटा विश्लेषण सुनिश्चित करता है कि पानी का उपयोग सटीक रूप से किया जाए, जिससे अत्यधिक सिंचाई से बचाव होता है और संसाधनों की बचत होती है। इसके अलावा, मशीन लर्निंग लक्षित खरपतवार और कीट प्रबंधन को सरल बनाता है, जिससे व्यापक रासायनिक उपयोग की आवश्यकता कम होती है और पर्यावरणीय प्रभाव कम होता है। अंत में, मिट्टी डेटा विश्लेषण मिट्टी का विस्तृत मानचित्रण और उर्वरकों का परिवर्तनीय दर से आवेदन सुनिश्चित करता है, जिससे फसल की वृद्धि के लिए पोषक तत्वों का आदान-प्रदान अनुकूल होता है।
2. **आपदा प्रबंधन (Disaster Management):** प्राकृतिक आपदाओं, जैसे सूखा और बाढ़, के प्रभाव को फसल उपज पर अनुमानित करना।
3. **मार्किट पूर्वानुमान (Market Forecasting):** हितधारकों को आपूर्ति रुझानों का आकलन करने और मूल्य निर्धारण रणनीतियों को समायोजित करने में मदद करना।
4. **जलवायु परिवर्तन अध्ययन (Climate Change Studies):** बदलते मौसम पैटर्न के कृषि उत्पादकता पर प्रभाव का विश्लेषण करना।

फसल उपज अनुमान में मशीन लर्निंग मॉडल्स से संबंधित चुनौतियाँ

1. **डेटा उपलब्धता और गुणवत्ता:** उच्च गुणवत्ता वाले, उचित रूप से एनोटेटेड डेटा तक सीमित पहुँच मॉडल प्रशिक्षण में रुकावट डाल सकती है।
2. **मॉडल की व्याख्यायिता:** जटिल मॉडल, विशेष रूप से डीप लर्निंग एल्गोरिदम, अक्सर ष्ब्लैक बॉक्सष के रूप में काम करते हैं, जिससे उनके निर्णयों को समझना मुश्किल हो जाता है।
3. **स्केलेबिलिटी:** विभिन्न क्षेत्रों में विभिन्न जलवायु, मिट्टी और फसलों के साथ मॉडल को अनुकूलित करना चुनौतीपूर्ण हो सकता है।
4. **मौजूदा प्रणालियों के साथ एकीकरण:** पारंपरिक कृषि कार्यप्रवाह में एमएल मॉडल्स को शामिल करने के लिए महत्वपूर्ण प्रयास की आवश्यकता होती है।

भविष्य दिशा-निर्देश

1. **IoT और एज कम्प्यूटिंग का एकीकरण:** IoT उपकरणों जैसे मिट्टी सेंसर और वेदर स्टेशन से रीयल-टाइम डेटा का उपयोग करके मॉडल अनुमानों को बेहतर बनाया जा सकता है।
2. **फेडरेटेड लर्निंग:** कोलेबोरेटिव एमएल दृष्टिकोणों के द्वारा विभिन्न हितधारक बिना संवेदनशील डेटा साझा किए मॉडल प्रशिक्षित कर सकते हैं।
3. **एक्सप्लेनेबल एआई (XAI):** फसल उपज को

प्रभावित करने वाले कारकों को समझने के लिए एक्सप्लेनेबल मॉडल विकसित करना।

निष्कर्ष

फसल उपज अनुमान सतत और लाभकारी कृषि के लिए अत्यंत महत्वपूर्ण है। मशीन लर्निंग फसल उपज अनुमान को अधिक सटीक, कुशल और स्केलेबल बनाने की विशाल क्षमता प्रदान करता है। विविध डेटा सेट्स और उन्नत एल्गोरिदम का उपयोग करके, संबंधित पक्ष कृषि में आने वाली चुनौतियों को अधिक प्रभावी ढंग से हल कर सकते हैं। हालांकि, इन तकनीकों की सफलता डेटा की उपलब्धता, मॉडल की व्याख्या करने की क्षमता और मौजूदा प्रणालियों में एकीकरण से संबंधित बाधाओं को दूर करने पर निर्भर करती है। निरंतर प्रगति के साथ, मशीन लर्निंग वैश्विक खाद्य सुरक्षा और सतत कृषि प्रथाओं को सुनिश्चित करने में महत्वपूर्ण भूमिका निभाएगा। इसके अतिरिक्त, रिमोट सेंसिंग, आर्टिफिशियल इंटेलिजेंस और मशीन लर्निंग एल्गोरिदम का कृषि में एकीकरण अनुमान की सटीकता को बढ़ाएगा और खेती की प्रथाओं को क्रांतिकारी रूप से बदलते हुए वैश्विक खाद्य सुरक्षा का समर्थन करेगा।

संदर्भ

अरुमुगम, पी., केमुरा, ए., शाउबर्गर, बी., एंड गॉर्नोट, सी. (2021). रिमोट सेंसिंग बेस्ड यील्ड एस्टीमेशन ऑफ राइस (ओरिजा सैटिवा एल.) यूजिंग ग्रेडियंट बूस्टेड रिग्रेशन इन इंडिया। रिमोट सेंसिंग, 13(12), 2379।

भारती, दास, पी., बनर्जी, आर., अहमद, टी., देवी, एस., एंड वर्मा, जी. (2023). आर्टिफिशियल न्यूरल नेटवर्क बेस्ड एप्पल यील्ड प्रेडिक्शन यूजिंग मॉर्फोलॉजिकल कैरेक्टर्स। हॉर्टिकल्चरे, 9(4), 436।

भटनागर, आर., एंड गोहैन, जी. बी. (2020). क्रॉप यील्ड एस्टीमेशन यूजिंग डिसीजन ट्रीज एंड रैंडम फॉरेस्ट मशीन लर्निंग एल्गोरिदम्स ऑन डेटा फ्रॉम टेरा (EOS AM-1) एंड एक्वा (EOS PM-1) सैटेलाइट डेटा। मशीन लर्निंग एंड डेटा माइनिंग इन एरोस्पेस टेक्नोलॉजी, 107-124।

दांग, सी., लिउ, वाई., युए, एच., कियान, जे., एंड झू, आर. (2021). ऑटम क्रॉप यील्ड प्रेडिक्शन यूजिंग डेटा-ड्रिवन अप्रोचेज:- सपोर्ट वेक्टर मशीन, रैंडम फॉरेस्ट, एंड डीप न्यूरल नेटवर्क मेथड्स। कनाडियन जर्नल ऑफ रिमोट सेंसिंग, 47(2), 162-181।

दास, पी., झा, जी. के., लामा, ए., एंड पारसाद, आर. (2023). क्रॉप यील्ड प्रेडिक्शन यूजिंग हाइब्रिड मशीन लर्निंग अप्रोच: अ केस स्टडी ऑफ लेन्टिल (लेन्स क्यूलिनारिस मेडिक.)। एग्रीकल्चर, 13(3), 596।

ढिल्लों, एम. एस., डाह्म्स, टी., क्यूबर्ट-फ्लॉक, सी., रमलर, टी., आर्नोल्ट, जे., स्टेफान-डेवंटर, आई., एंड उल्मान, टी. (2023). इंटेग्रेटिंग रैंडम फॉरेस्ट एंड क्रॉप मॉडलिंग इम्प्रूव्स द क्रॉप यील्ड प्रेडिक्शन ऑफ विंटर व्हीट एंड ऑयल सीड रैप। फ्रंटियर्स इन रिमोट सेंसिंग, 3, 1010978।

ह्यूबर, एफ., युशचेंको, ए., स्ट्रैटमैन, बी., एंड स्टाइनहाज, वी. (2022). एक्सट्रीम ग्रेडियंट बूस्टिंग फॉर यील्ड एस्टीमेशन कंपेयरड विद डीप लर्निंग अप्रोचेज।। कंप्यूटर एंड इलेक्ट्रॉनिक्स इन एग्रीकल्चर, 202, 107346।

खाकी, एस., एंड वांग, एल. (2019). क्रॉप यील्ड प्रेडिक्शन यूजिंग डीप न्यूरल नेटवर्क्स। फ्रंटियर्स इन प्लांट साइंस, 10, 621।

मेडार, आर. ए., एंड राजपुरोहित, वी. एस. (2014). ए सर्वे ऑन डेटा माइनिंग टेक्नीकस फॉर क्रॉप यील्ड प्रेडिक्शन। इंटरनेशनल जर्नल ऑफ एडवांस रिसर्च इन कंप्यूटर साइंस एंड मैनेजमेंट स्टडीज, 2(9), 59-64।

प्रसाद, एन. आर., पटेल, एन. आर., एंड दानोडिया, ए. (2021). क्रॉप यील्ड प्रेडिक्शन इन कॉटन फॉर रीजनल लेवल यूजिंग रैंडम फॉरेस्ट अप्रोच। स्पैटियल इंफॉर्मेशन रिसर्च, 29, 195-206।

मशीन लर्निंग तकनीकों का उपयोग करके गेहूं की फसल पर लैंडस्केप डायग्नोस्टिक सर्वेक्षण डेटा का फीचर चयन और उपज का अनुमान

सौमेन पाल[1], अलका अरोडा[1], सुदीप मरवाहा[1], रंजित कुमार पॉल[1], अजीत गुप्ता[1], कमलिका नाथ[1], सौरभ कुमार[1], रामकंवर मालिक[2], राजर्षि रॉय बर्मर्न[3] एवं उधम सिंह गौतम[3]

[1]भा.कृ.अनु.प. – भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

[2]अंतर्राष्ट्रीय मक्का और गेहूं सुधार केंद्र, भारत

[3]भारतीय कृषि अनुसंधान परिषद, नई दिल्ली

सार

प्रत्येक जिले के कृषि विज्ञान केंद्र (केवीके) के माध्यम से दक्षिण एशिया के लिए अनाज प्रणाली पहल (सीएसआईएसए) के तहत भारत के बिहार और पूर्वी उत्तर प्रदेश के क्रमानुसार 31 और 9 जिलों में लैंडस्केप डायग्नोस्टिक सर्वेक्षण (एलडीएस) आयोजित किया गया था। इस अध्ययन का उद्देश्य उन कारकों का पता लगाना था जो गेहूं की उपज निर्धारित करने में प्रमुख रूप से योगदान देता है। सर्वेक्षण में किसानों द्वारा गेहूं की फसल में अपनाई गई उत्पादन पद्धतियों को शामिल किया गया। सर्वेक्षण में सभी जिलों से दो चरणों के नमूना डिजाइन (two stage sampling designs) के बाद कुल 7634 किसानों से जानकारी एकत्र की गई। उपयुक्त कारकों (जो उपज निर्धारित करने में महत्वपूर्ण योगदान देते हैं) का पता लगाने के लिए स्टेपवाइज रिग्रेशन (एसआर), लासो, रैंडम फॉरेस्ट रिग्रेशन (आरएफआर) और रिज रिग्रेशन (आरआर) का उपयोग किया गया। चयनित कारकों का उपयोग करके फसल उपज का अनुमान के लिए, मल्टीपल लीनियर रिग्रेशन (एमएलआर), आरएफआर, सपोर्ट वेक्टर रिग्रेशन (एसवीआर) और आरआर का उपयोग किया गया। मॉडलों की तुलना के लिए सांख्यिकीय निदान उपाय जैसे माध्य वर्ग त्रुटि (MSE), माध्य निरपेक्ष त्रुटि (MAE), मूल माध्य वर्ग त्रुटि (RMSE) और निर्धारण गुणांक (R^2) का उपयोग किया गया। यह देखा गया कि फीचर चयन के लिए आरएफआर सबसे उपयोगी तकनीक के रूप में सामने आई, जबकि एसवीआर (आरएफआर के माध्यम से चयनित कारकों का उपयोग करके) प्रशिक्षण और परीक्षण डेटासेट दोनों के मामले में उपज का अनुमान के लिए अन्य तकनीकों से बेहतर प्रदर्शन करती है। सिंचाई की संख्या, फसल की समय सीमा, किस्म, यूरिया की कुल मात्रा, खरपतवार की गंभीरता, मैन्युअल निराई की संख्या इत्यादि गेहूं की उपज का अनुमान के लिए महत्वपूर्ण कारकों के रूप में सामने आई।

मुख्य शब्द : गेहूँ, मल्टीपल लीनियर रिग्रेशन, रैंडम फॉरेस्ट रिग्रेशन, लैंडस्केप डायग्नोस्टिक सर्वेक्षण, सपोर्ट वेक्टर रिग्रेशन।

परिचय

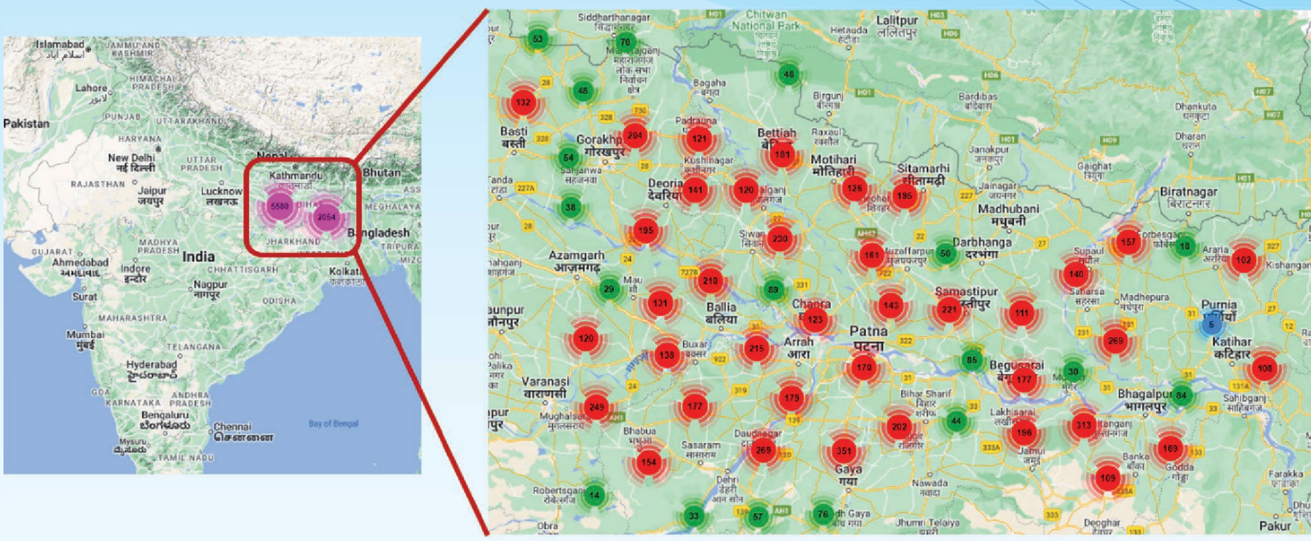
एलडीएस को सीएसआईएसए (अंतर्राष्ट्रीय मक्का और गेहूं सुधार केंद्र (सीआईएमएमवाईटी) के नेतृत्व में) और भारतीय कृषि अनुसंधान परिषद (आईसीएआर) के विस्तार प्रभाग द्वारा तैनात किया गया है। इस सर्वेक्षण का उद्देश्य हाल के डेटाबेस साक्ष्य तैयार करना, बेहतर अंतर्दृष्टि प्राप्त

करना और नीति निर्माताओं द्वारा सूचित निर्णय लेने की सुविधा प्रदान करना था [1]। सर्वेक्षण में चावल और गेहूं की फसलों में किसानों द्वारा लागू उत्पादन प्रथाओं को शामिल किया गया। दो प्रमुख फसल मौसमों (बरसात और सर्दी) के लिए प्रत्येक जिले में यादृच्छिक रूप (randomly) से चयनित किसानों की पूर्व निर्धारित संख्या का सर्वेक्षण किया गया। यह इलेक्ट्रॉनिक रूप से सक्षम सर्वेक्षण जिला स्तर पर कृषि विज्ञान केंद्र के माध्यम से लागू किया गया था। एलडीएस की पहली सर्वेक्षण फसल 2017 में चावल थी, उसके बाद 2018 की फसल गेहूं थी, जिसमें बिहार के 31 जिलों, पूर्वी उत्तर प्रदेश (ईयूपी) के 9 जिलों और ओडिशा के 3 जिलों को कवर किया गया था। प्रस्तुत शोधपत्र में, गेहूं की फसल पर एलडीएस डेटा पर विचार किया गया है। सर्वेक्षण में 226 चर दर्ज किए गए, जिसमें पिछली फसल, अवशेष प्रबंधन, फसल स्थापना विधि, किस्म और बीज स्रोत, पोषक तत्व प्रबंधन, सिंचाई प्रबंधन, खरपतवार और उनका प्रबंधन, कटाई विधि और किसान द्वारा बताई गई उपज जैसे प्रमुख फसल उत्पादन कारक शामिल हैं [2]। एलडीएस विश्लेषण विभिन्न क्षेत्रों में मेगा या बड़े रुझानों (उदाहरण के लिए फसल किस्म या संकर उपयोग), इंटरैक्शन और अधिक स्थानीय या पारिस्थितिकी विशिष्ट संदर्भ (उदाहरण के लिए विभिन्न पारिस्थितिकी में लंबी बनाम मध्यम अवधि की किस्मों का प्रदर्शन) की पहचान कर सकता है, और महत्वपूर्ण रूप से लक्ष्यीकरण और अपनाने के लिए नई श्रेणियों या समूहों की पहचान कर सकता है (जैसे कि सिंचाई के प्रकार और संख्या के अनुसार)। इस पारिस्थितिकी तंत्र को समस्याओं का तेजी से निदान करने और तदनुसार समाधान खोजने के लिए जैविक और सामाजिक वैज्ञानिकों के एकीकृत प्रयास की आवश्यकता है। इसे प्राप्त करने के लिए, उन चर या कारकों का पता लगाना आवश्यक है जो फसल की उपज निर्धारित करने में प्रमुख योगदान दे रहे हैं। मशीन लर्निंग एल्गोरिदम जैसे स्टेपवाइज मल्टी लीनियर रिग्रेशन (एसएमएलआर) [3], लासो (न्यूनतम निरपेक्ष संकोचन और चयन ऑपरेटर) [3], सपोर्ट वेक्टर रिग्रेशन (एसवीआर) [[3] [4] [6] [7]] रैंडम फॉरेस्ट रिग्रेशन (आरएफआर) [[4] [5] [6]] आदि पर्यवेक्षित शिक्षण (supervised learning) पर आधारित हैं और फीचर चयन के लिए व्यापक रूप से उपयोग किए जाते हैं।

कार्यप्रणाली

इस अध्ययन में ओपन डेटा किट (ओडीके) का उपयोग करके फसल उत्पादन प्रथाओं का नैदानिक सर्वेक्षण किया गया। ओडीके के 3 घटक हैं - निर्माण, संग्रह और एकत्रीकरण। ओडीके कलेक्ट एक एंड्रॉइड ऐप है, जिसका उपयोग इस सर्वेक्षण-आधारित डेटा संग्रहण में किया गया था। सर्वेक्षण का उद्देश्य चावल-गेहूं फसल प्रणाली से विस्तृत जानकारी प्राप्त करना था। बिहार और ईयूपी के प्रत्येक जिले से, आकार के अनुपात में संभावना (पीपीएस) पद्धति का उपयोग करके 30 गांवों का चयन किया गया था। सरल यादृच्छिक नमूने के माध्यम से प्रत्येक गांव से 7 घरों का चयन किया गया। इस प्रकार, प्रत्येक जिले से 210 (30×7) चयनित किसानों से उनकी वर्तमान फसल उत्पादन प्रथाओं के लिए गहराई से साक्षात्कार लिया गया। कुल 7634 डेटा बिंदु एकत्र किए गए। सर्वेक्षण चयनित किसान के सबसे बड़े गेहूं भूखंड पर आयोजित किया गया था। चयनित सबसे बड़े भूखंड में 2 स्थानों से क्रॉप-कट लिया गया। सर्वेक्षण किए गए गेहूं के खेतों के जीपीएस बिंदुओं को ओडीके कलेक्ट के माध्यम से कैप्चर किया गया और चित्र 1 में प्रदर्शित किया गया।

ओडीके कलेक्ट के माध्यम से सर्वेक्षण डेटा को सर्वेक्षक द्वारा ओडीके एग्रीगेट में एकत्रित किया गया। सर्वेक्षण डेटा को संग्रहीत करने, विश्लेषण करने और प्रस्तुत करने के लिए ओडीके एग्रीगेट का उपयोग किया गया। ओडीके एग्रीगेट से, डेटा को माइक्रोसॉफ्ट एसक्यूएल (MS SQL) सर्वर में निर्यात किया गया था। उपज के निर्धारण में महत्वपूर्ण योगदान देने वाले उपयुक्त चरों (स्वतंत्र) का पता लगाने के लिए, स्टेपवाइज रिग्रेशन (एसआर), लासो, रैंडम फॉरेस्ट रिग्रेशन (आरएफआर) और रिज रिग्रेशन (आरआर) का उपयोग किया गया। चयनित फीचर्स के साथ, फसल उपज की अनुमान के लिए, मल्टीपल लीनियर रिग्रेशन (एमएलआर), आरएफआर, सपोर्ट वेक्टर रिग्रेशन



चित्र 1. सर्वेक्षण डेटा बिंदुओं का भौगोलिक स्थान

(एसवीआर) और आरआर का उपयोग किया गया। मॉडलों की तुलना के लिए सांख्यिकीय निदान उपायों जैसे माध्य वर्ग त्रुटि (एमएसई), माध्य निरपेक्ष त्रुटि (एमएई), मूल माध्य वर्ग त्रुटि (आरएमएसई) और निर्धारण गुणांक (R^2) का उपयोग किया गया। इन तकनीकों का संक्षिप्त विवरण नीचे दिया गया है।

फीचर चयन और फसल उपज की अनुमान के लिए तरीके

मल्टीपल लीनियर रिग्रेशन (एमएलआर)रू एमएलआर मॉडल का सामान्य रूप निम्नलिखित है:

$$y = f(x_1, x_2, ..., x_p) + \varepsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + ... + \beta_p x_p + \varepsilon \quad (1)$$

यहाँ, y आश्रित या समझाया गया चर है और $x_1, x_2, ..., x_p$ स्वतंत्र या व्याख्यात्मक चर हैं।

न्यूनतम निरपेक्ष संकोचन और चयन ऑपरेटर लासो लासो प्रतिगमन (Regression) एक नियमितीकरण (Regularcation) तकनीक है जिसका उपयोग अधिक सटीक अनुमान के लिए प्रतिगमन विधियों पर किया जाता है। यह मॉडल सिकुड़न का उपयोग करता है (डेटा मान को औसत के रूप में एक केंद्रीय बिंदु की ओर सिकोड़ दिया जाता है)। लासो प्रक्रिया सरल, विरल मॉडल (यानी कम पैरामीटर वाले मॉडल) को प्रोत्साहित करती है। यह निरर्थक चर राशि के गुणांकों को शून्य तक कम कर देता है और इस प्रकार सीधे फीचर चयन भी करता है। लासो का लागत कार्य (Cost function) निम्नलिखित हैरू

$$\sum_{i=1}^{n}(y_i - x_j'\beta)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \quad (2)$$

यहाँ,

β = लासो अनुमानक का वेक्टर

x_j' = इनपुट चर राशि की jवां पंक्ति वेक्टर

y_i = iवां आउटपुट

λ = हाइपर पैरामीटर

रैंडम फॉरेस्ट रिग्रेशन (आरएफआर): इस पद्धति में, सामूहिक शिक्षण तकनीक का उपयोग किया जाता है जो कई पेड़-आधारित (tree-based) शिक्षार्थियों के परिणामों को जोड़ती है। फीचर्स (इनपुट) X और आउटपुट Y का एक प्रशिक्षण सेट दिए जाने पर बैगिंग बार-बार प्रशिक्षण सेट का एक यादृच्छिक नमूना चुनता है और पेड़ों को इन नमूनों में फिट करता है। प्रत्येक निर्णय पेड़ (Decision tree) के लिए, हम उदाहरणों का एक क्रम प्राप्त करते हैं जो प्रशिक्षण सेट से यादृच्छिक रूप से नमूना प्रतिस्थापन (मशीन लर्निंग में बैगिंग) होते हैं। उदाहरणों का प्रत्येक क्रम एक विशिष्ट पेड़ बनाने वाले यादृच्छिक वेक्टर ϕ_k से मेल खाता है। इनपुट X के लिए Kवां पेड़ की अनुमान को निम्न समीकरण द्वारा दर्शाया जा सकता है:

$$h(X) = h(X, \phi_k), \forall k \in \{1,2,3, ..., K\} \quad (3)$$

जहाँ K = पेड़ों की संख्या

समान निर्णय नियमों (Decision rules) का पालन करके, हम उदाहरणों के औसत आउटपुट के रूप में किसी भी उप-पेड़ को पूर्वानुमानित कर सकते हैं। अंत में, हम प्रत्येक पेड़ के उत्पादन के औसत (माध्य) के रूप में अंतिम अनुमान प्राप्त कर सकते हैं, जैसा कि निम्नलिखित समीकरण द्वारा दर्शाया गया है:

$$h(X)=\frac{1}{k}\sum_{i=1}^{k} h_k(X) \quad (4)$$

रिज रिग्रेशन (आरआर): रिज रिग्रेशन एक विशेष तकनीक है जिसका उपयोग मल्टीपल रिग्रेशन डेटा का विश्लेषण करने के लिए किया जाता है जो प्रकृति में मल्टी कॉलिनियर होता है। यह एक मौलिक नियमितीकरण तकनीक है, लेकिन इसके पीछे जटिल विज्ञान के कारण इसका बहुत व्यापक रूप से उपयोग नहीं किया जाता है। रिज रिग्रेशन का उपयोग पार्सिमोनियस मॉडल बनाने के उद्देश्य से किया जाता है जब किसी दिए गए सेट में पूर्वानुमान चर की संख्या अवलोकनों की संख्या से अधिक हो जाती है या जब डेटासेट में बहुसंरेखता होती है। रिज रीग्रेशन का लागत कार्य निम्नलिखित है:

$$L_{hridge}(\hat{\beta}) = \sum_{i-1}^{n}(y_i - x_j'\beta)^2 + \lambda\sum_{j=1}^{m}\beta_j^2 \quad (5)$$

यहाँ,

β = रिज प्रतिगमन अनुमानक

x_j' = इनपुट चर राशि की jवां पंक्ति वेक्टर

y_i = iवां आउटपुट

λ = नियमितीकरण गुणांक

समर्थन वेक्टर प्रतिगमन (एसवीआर): एसवीआर एक पर्यवेक्षित मशीन लर्निंग एल्गोरिदम है जिसका उपयोग आमतौर पर डेटा में गैर-रेखीय (non-linear) संबंध को उच्च आयामी रैखिक फीचर स्पेस में मॉडल करने के लिए किया जाता है ताकि एक रैखिक डोमेन में अनुमान प्राप्त की जा सके [8]। गैर-रेखीय एसवीआर का सामान्य रूप निम्नलिखित है:

$$f(x) = w^T \varphi(x) + b \quad (6)$$

यहाँ, $\varphi(.): R^n \rightarrow R^{n_h}$ जो मूल इनपुट स्पेस से उच्च आयामी फीचर स्पेस में एक गैर-रेखीय मैपिंग फंक्शन है। एसवीआर निम्नलिखित फंक्शन को न्यूनतम करता है:

$$(RC) = \frac{1}{2}\|w\|^2 + C\left[\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(y_i, f(x_i))\right] \quad (7)$$

C और ε दोनों उपयोगकर्ता द्वारा निर्धारित हाइपर-पैरामीटर हैं। यहां पहले पद को नियमित पद कहा जाता है जो फंक्शन की समतलता को मापता है। $\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(y_i, f(x_i))$ को अनुभवजन्य (Empirical) त्रुटि कहा जाता है जिसका अनुमान वापनिक (Vapnik) ε-असंवेदनशील लॉस (Loss) फंक्शन द्वारा किया जाता है।

फसल उपज का अनुमान के लिए मॉडल का प्रदर्शन विश्लेषण

फसल उपज का अनुमान के लिए उपयोग की जाने वाली कई तकनीकों में प्राप्त सटीकता की तुलना करने के लिए प्रदर्शन विश्लेषण किया गया है, और माध्य वर्ग त्रुटि (एमएसई), माध्य निरपेक्ष त्रुटि (एमएई), मूल माध्य वर्ग त्रुटि (आरएमएसई) और निर्धारण गुणांक (R^2) जैसे उपयुक्त प्रदर्शन मापदंडों के साथ मूल्यांकन किया गया है।

प्रदर्शन मेट्रिक्स

माध्य निरपेक्ष प्रतिशत त्रुटि (एमएपीई): इसे माध्य निरपेक्ष प्रतिशत विचलन (एमएपीडी) के रूप में भी जाना जाता है और निम्नलिखित समीकरण द्वारा दर्शाया जाता है:

$$MAPE = \sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (8)$$

यहाँ,

y_i = iवां प्रेक्षित मान है

$\hat{y}_i$ = संगत अनुमानित मान है

n = प्रेक्षणों की संख्या

निर्धारण का गुणांक (R^2): R^2 रिग्रेशन मॉडल में एक सांख्यिकीय माप है जो आश्रित चर राशि में भिन्नता के अनुपात को निर्धारित करता है जिसे स्वतंत्र चर राशि द्वारा समझाया जा सकता है।

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} \quad (9)$$

यहाँ

SS_{RES} = वर्ग का अवशिष्ट योग

SS_{TOT} = वर्ग का कुल योग

मूल माध्य वर्ग त्रुटि (आरएमएसई): यह सांख्यिकीय मॉडल के अनुमानित मूल्यों और वास्तविक मूल्यों के बीच औसत अंतर को मापता है। गणितीय रूप से, यह अवशेषों का मानक विचलन है। अवशेष प्रतिगमन रेखा और डेटा बिंदुओं के बीच की दूरी का प्रतिनिधित्व करते हैं।

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (10)$$

यहाँ,

y_i = iवां प्रेक्षित मान है

$\hat{y}_i$ = संगत अनुमानित मान है

n = प्रेक्षणों की संख्या

माध्य निरपेक्ष त्रुटि (एमएई): एमएई पूर्वानुमानों और प्रेक्षणों के एक समूह के लिए निरपेक्ष त्रुटियों (किसी प्रेक्षण के पूर्वानुमान और उस प्रेक्षण के वास्तविक मान के बीच अंतर का परिमाण) का औसत, सम्पूर्ण समूह के लिए त्रुटियों के परिमाण के माप के रूप में लेता है। यह इस प्रकार व्यक्त है:

$$MAE = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n} \qquad (11)$$

यहाँ,

y_i = iवां प्रेक्षित मान है

$\hat{y}_i$ = संगत अनुमानित मान है

n = प्रेक्षणों की संख्या

परिणाम और चर्चा

फीचर चयन और फसल उपज का अनुमान के लिए बिहार, ईयूपी और संयुक्त (बिहार+ईयूपी) एलडीएस डेटा पर विचार किया गया। फीचर चयन के लिए, स्टेपवाइज रिग्रेशन (एसआर), लासो, रैंडम फॉरेस्ट रिग्रेशन (आरएफआर) और रिज रिग्रेशन (आरआर) का उपयोग किया गया। उपयोग की गई विभिन्न तकनीकों में निकाली गई फीचर्स की संख्या भिन्न थी। एक विशेष तकनीक का उपयोग करके निकाली गई फीचर्स का उपयोग सभी फसल उपज का अनुमान मॉडलों अर्थात एमएलआर, आरएफआर, एसवीआर और आरआर में स्वतंत्र चर राशि के रूप में किया गया है। मॉडल में प्रति हेक्टेयर टन के संदर्भ में उपज को आश्रित चर राशि के रूप में उपयोग किया गया। संपूर्ण डेटा को प्रशिक्षण और परीक्षण सेट में विभाजित किया गया। 90 प्रतिशत डेटा मॉडल को प्रशिक्षित करने के लिए तथा बाकी डेटा का उपयोग मॉडल के सत्यापन के लिए किया गया था। सांख्यिकीय निदान उपायों का उपयोग करके मॉडलों के प्रदर्शन की तुलना की गई। इस प्रकार, चर चयन तकनीक +फसल उपज का अनुमान के संयोजन को प्रत्येक डेटासेट के लिए अंतिम माना गया जहां अधिकतम R^2 और न्यूनतम माध्य वर्ग त्रुटि (MSE), माध्य निरपेक्ष त्रुटि (MAE) एबं मूल माध्य वर्ग त्रुटि (RMSE) प्राप्त हुए। इस प्रकार पहचाने गए सर्वोत्तम मॉडल संयोजन, सभी डेटासेट के लिए तालिका 1 में रिपोर्ट किए गए हैं। तालिका 1 से यह देखा गया है कि आरएफआर सभी डेटासेट के लिए फीचर चयन के लिए सर्वत्तोम तकनीक के रूप में सामने आया है।

तालिका 1. गेहूं पर एलडीएस डेटासेट के लिए पहचाने गए सर्वोत्तम मॉडल संयोजन

एलडीएस डेटासेट		फीचर चयन विधि	फसल उपज का अनुमान के लिए मॉडल
फसल	राज्य		
गेहूँ	बिहार	आरएफआर	एसवीआर
	ईयूपी	आरएफआर	एसवीआर
	बिहार+ईयूपी	आरएफआर	एसवीआर

इस प्रकार अलग-अलग डेटासेट के लिए निकाली गई चयनित फीचर्स तालिका 2 में दर्शाई गई हैं।

व्यक्तिगत डेटासेट के लिए तालिका 2 में उल्लिखित चर के साथ, फसल उपज का अनुमान के लिए मॉडल का उपयोग किया गया। एसवीआर ने सभी डेटासेट के लिए प्रशिक्षण और परीक्षण दोनों के मामले में सांख्यिकीय निदान उपायों के संदर्भ में सर्वोत्तम सटीकता का उत्पादन किया। संबंधित परिणाम तालिका 3 और 4 में दिखाए गए हैं।

इसके अलावा, प्रत्येक परीक्षण डेटासेट के लिए अवशिष्ट निदान के हिस्से के रूप में, त्रुटि श्रृंखला में

तालिका 2. गेहूं पर एलडीएस डेटासेट के लिए निकाली गई चयनित फीचर्स

<table>
<tr><th colspan="2">एलडीएस डेटासेट</th><th rowspan="2">चयनित फीचर्स</th></tr>
<tr><th>फसल</th><th>राज्य</th></tr>
<tr><td rowspan="3">गेहूँ</td><td>बिहार</td><td>सिंचाई की संख्या, फसल की अवधि, पिछली फसल के लिए जुताई की प्रक्रिया, रोपाई की तारीख, किस्म का नाम, फसल की खेती के अंतर्गत क्षेत्र, फसल कटाई की तिथि, फसल के विकास का चरण जिस पर सिंचाई की गई, बीज की मात्रा, डीएपी की कुल मात्रा, मिट्टी की बनावट, मिट्टी की जल निकासी वर्ग, यूरिया की कुल मात्रा, कीड़ों की गंभीरता, डीएपी की कुल मात्रा, यूरिया की कुल मात्रा, एमओपी की कुल मात्रा, फसल वृद्धि अवस्था जब सूखा पड़ा, मिट्टी की गुणवत्ता (उच्च/मध्यम/निम्न), थ्रेशिंग विधि, हाथ से निराई-गुड़ाई की संख्या, एफवाईएम का प्रयोग (हां/नहीं), खरपतवार की गंभीरता, कीट की गंभीरता</td></tr>
<tr><td>ईयूपी</td><td>सिंचाई की संख्या, यूरिया की कुल मात्रा, फसल की अवधि, बीज बोने या रोपाई की तारीख, फसल की खेती के अंतर्गत क्षेत्र, फसल के विकास का चरण जिस पर सिंचाई की गई, कटाई की तिथि, बीज की मात्रा, किस्म का नाम, शाकनाशी के प्रयोग की संख्या, फसल गिरने का प्रतिशत, डीएपी की कुल मात्रा, भूमि तैयार करने की विधि, हाथ से निराई-गुड़ाई की संख्या, फसल वृद्धि अवस्था जब सूखा पड़ा, पिछली फसल के लिए जुताई की प्रक्रिया, मिट्टी की बनावट, मिट्टी की गुणवत्ता (उच्च/मध्यम/निम्न), उपयोग की गई नई प्रबंधन पद्धतियाँ, जिंक सल्फेट की मात्रा, मड़ाई की विधि, कटाई की विधि, सूखे की गंभीरता, कीट की गंभीरता, खरपतवार की गंभीरता</td></tr>
<tr><td>बिहार+ईयूपी</td><td>फसल कटाई की तिथि, सिंचाई की संख्या, किस्म का नाम, यूरिया की कुल मात्रा, खरपतवार की गंभीरता, मिट्टी की बनावट, मिट्टी की जल निकासी वर्ग, मिट्टी की गुणवत्ता (उच्च/मध्यम/निम्न), एफवाईएम का प्रयोग (हां/नहीं), फसल अवधि के दिन (Crop Duration Days), बीज की मात्रा, जिंक सल्फेट की कुल मात्रा, कीड़ों की गंभीरता, फसल के विकास का चरण जिस पर सिंचाई की गई, खरपतवार की गंभीरता, फसल अवशेष प्रतिशत, डीएपी की कुल मात्रा, सूखे की गंभीरता, रोग की गंभीरता, थ्रेशिंग की विधि, फसल गिरने का प्रतिशत, भूमि की तैयारी की विधि, उपयोग की गई नई प्रबंधन पद्धतियाँ, हाथ से निराई-गुड़ाई की संख्या, शाकनाशी के प्रयोग की संख्या</td></tr>
</table>

तालिका 3. गेहूं पर एलडीएस डेटासेट के लिए प्रशिक्षण डेटासेट पर प्रदर्शन मेट्रिक्स

डेटा	मॉडल (फीचर चयन, उपज का अनुमान)	एमएइ	एमएसई	आरएमएसई	R^2
बिहार	आरएफआर, एसवीआर	0.34	0.24	0.49	0.66
ईयूपी	आरएफआर, एसवीआर	0.30	0.21	0.45	0.64
बिहार+ईयूपी	आरएफआर, एसवीआर	0.36	0.27	0.51	0.63

तालिका 4. गेहूं पर एलडीएस डेटासेट के परीक्षण डेटासेट पर प्रदर्शन मेट्रिक्स

डेटा	मॉडल (फीचर चयन, उपज का अनुमान)	एमएइ	एमएसई	आरएमएसई	R^2
बिहार	आरएफआर, एसवीआर	0.51	0.24	0.49	0.38
ईयूपी	आरएफआर, एसवीआर	0.45	0.59	0.65	0.35
बिहार+ईयूपी	आरएफआर, एसवीआर	0.51	0.66	0.51	0.43

स्व-सहसंबंध (autocorrelation) की जांच करने के लिए डर्बिन-वाटसन (Durbin-Watson) परीक्षण आयोजित किया गया था। यह देखा गया कि प्रत्येक डेटासेट के लिए, अवशेषों के बीच कोई स्व-सहसंबंध नहीं पाया गया।

निष्कर्ष

वर्तमान अध्ययन में, गेहूं की फसल के एलडीएस डेटा पर फीचर चयन और फसल उपज का अनुमान के लिए मशीन लर्निंग तकनीकों का उपयोग किया गया था। रैंडम

फॉरेस्ट रिग्रेशन (फीचर चयन) – सपोर्ट वेक्टर रिग्रेशन (फसल उपज का अनुमान) का प्रदर्शन सभी तरीकों में सबसे अच्छा पाया गया। निकाली गई फीचर्स का आगे अध्ययन करके फसल की पैदावार पर एकल चर राशि के प्रभाव का पता लगाया जा सकता है। उत्पादन पद्धतियों के इष्टतम संयोजन का पता लगाने के लिए, जिससे फसल की उपज अधिकतम हो सके, कारणात्मक विश्लेषण (Causal Analysis) को अपनाया जा सकता है।

संदर्भ

राय, ए., अजय, ए., शर्मा, एस. और अरोड़ा, ए. (2019). भारत में परिदृश्य स्तर पर नैदानिक फसल उत्पादन सर्वेक्षण के लिए ओपन डेटा किट। *न्यू फ्रंटियर्स इन एग्रीकल्चरल एक्सटेंशन – खंड 1* (संस्करण सिंह, ए.के. एट अल), अंतर्राष्ट्रीय मक्का और गेहूं सुधार केंद्र (सीआईएमएमवाईटी), पीपी. 11–17।

अजय, ए., क्राउफर्ड, पी. और शर्मा, एस. (2021). पूर्वी भारत में गेहूं उत्पादन प्रथाओं और उपज का लैंडस्केप डायग्नोस्टिक सर्वेक्षण डेटा। *ओपन डाटा जर्नल फॉर एग्रीकल्चरल रिसर्च*, **7**, 20–26।

गुप्ता, एस., वशिष्ठ, ए., कृष्णन, पी., लामा, ए., प्रसाद, एस. और अरविंद, के. (2022). हाइब्रिड मशीन लर्निंग तकनीकों का उपयोग करके मल्टी-स्टेज गेहूं उपज का अनुमान। *जर्नल ऑफ एग्रोमेटोरोलॉजी*, **24**(4), 373–379।

बोंद्रे, डी. ए. और महागांवकर, एस. (2019). मशीन लर्निंग एल्गोरिदम का उपयोग करके फसल की उपज और उर्वरक अनुशंसा की अनुमान। *इंटरनेशनल जर्नल ऑफ इंजीनियरिंग एप्लाइड साइंसेज एंड टेक्नोलॉजी*, **4**(5), 371–376।

सुंदरी, वी., अनुश्री, एम. और स्वेता, यू. (2022). मशीन लर्निंग एल्गोरिदम का उपयोग करके फसल की सिफारिश और उपज की अनुमान। *वर्ल्ड जर्नल ऑफ एडवांस्ड रिसर्च एंड रिव्यूज*, **14**(3), 452–459।

वर्शिता, डी.एन. और चौधरी, एस. (2022). मशीन लर्निंग तकनीक का उपयोग करके कॉफी बागान की मिट्टी की उर्वरता और उपज की अनुमान। *रिसर्च जर्नल ऑफ एग्रीकल्चरल साइंसेज*, **13**(2), 514–518।

दास, पी., झा, जी.के., लामा, ए. और प्रसाद, आर. (2023). हाइब्रिड मशीन लर्निंग दृष्टिकोण का उपयोग करके फसल उपज की अनुमानरू दाल का एक केस स्टडी (लेंस क्यूलिनारिस मेडिक)। *एग्रीकल्चर*, **13**(3), 596।

गैराई, एस. और पॉल, आर. के. (2023). सीईईएमडीएएन विघटन और मशीन इंटेलिजेंस का उपयोग करके एमसीएस आधारित-एनसेंबल मॉडल का विकास। *इंटेल. सिस्टम. एप्लिकेशन*, **18**, 200202।

मशीन लर्निंग का उपयोग करके लघु-अणुओं की जैव-सक्रियता का पूर्वानुमान

स्नेहा मुर्मू, सुधीर श्रीवास्तव, दीपा भट्ट, सौम्या शर्मा, ऋत्विका दास, मोहम्मद समीर फारूकी, गिरीश कुमार झा

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

फंगल संक्रमण कृषि के लिए महत्वपूर्ण खतरा पैदा करते हैं,, जिसके चलते ऐसे नये टूल्स की आवश्यकता होती है जो फंगल -नाशक क्षमता वाले जैव-सक्रिय यौगिकों की पहचान कर सकें। यह अध्ययन मशीन लर्निंग (एमएल) का उपयोग करके फंगल लक्ष्य प्रोटीनों के विरुद्ध, यौगिक जैव-सक्रियता का पूर्वानुमान लगाने के लिए एक मजबूत प्रणाली विकसित करता है। इस शोध में चेमब्ल (ChEMBL) से एक क्यूरेटेड डेटासेट को एसएमआईएलईएस (SMILES) का उपयोग करके प्री-प्रोसेस और एनकोड किया गया, जिससे पडेल (PaDEL) के माध्यम से मॉलिक्यूलर डिस्क्रिप्टर प्राप्त हुए। फीचर चयन की प्रक्रिया के माध्यम से डिस्क्रिप्टर्स की संख्या को घटाकर 241 कर दिया गया, जिससे मॉडल की मजबूती में सुधार हुआ। कई मशीन लर्निंग एल्गोरिदम का परीक्षण किया गया, परीक्षण किए गए कई एमएल एल्गोरिदम में, रैंडम फॉरेस्ट (आरएफ) मॉडल ने दूसरों से बेहतर प्रदर्शन किया, सत्यापन सेट पर 97.04% और परीक्षण सेट पर 91.02% सटीकता हासिल की। फंगीपैड टूल के साथ तुलनात्मक विश्लेषण ने प्रस्तावित आरएफ मॉडल के बेहतर प्रदर्शन को प्रदर्शित किया।

मुख्य शब्द : मशीन लर्निंग, बायोएक्टिविटी, वर्गीकरण, आईसी 50

परिचय

प्राचीन काल से पौधों के अर्क अपने रोगाणुरोधी गुणों के लिए जाने जाते हैं, और कई अध्ययनों ने उनकी प्रभावशीलता की पुष्टि की है (मेशुक 2021; खान, एट अल., 2022; खान, एट अल., 2023)। ये प्राकृतिक अर्क अपनी उपलब्ध ता, लागत-प्रभावशीलता, सुरक्षा, बायोडिग्रेडेबिलिटी और पर्यावरण-मित्रता के कारण लाभप्रद हैं, जो उन्हें सूक्ष्मजीवीय खतरों से निपटने के लिए टिकाऊ विकल्प बनाते हैं। कम्प्यूटेशनल बायोलॉजी और फाइटोकेमिकल अनुसंधान का एकीकरण कृषि चुनौतियों के लिए आशाजनक संभावनाएं प्रदान करता है। कम्प्यूटेशनल विधियों ने ड्रग रीपर्पजिंग (सडेघ एट अल., 2021), सरकुलर आरएनए-रोग संबंध भविष्यवाणी और कोलोरेक्टल कैंसर में एक्सोसोमल एमआईआरएनए-एमआरएनए नेटवर्क विश्लेषण (मा एट अल., 2021) जैसे क्षेत्रों में महत्वपूर्ण प्रगति की है।

मशीन लर्निंग (एमएल) तकनीकों ने विशेष रूप से जटिल डेटा सेटों को संभालने और डेटा-आधारित पूर्वानुमान प्रदान करने में अपनी उपयोगिता सिद्ध की है। उदाहरण के लिए, प्रोटीन संरचना प्रिडिक्शन (सीनियर एट अल., 2020) के लिए डीप लर्निंग को नियोजित किया गया है, जीन अभिव्यक्ति-आधारित रोग निदान के लिए एनसेंबल विधियों (कोउरो एट अल., 2015), प्रोटीन-प्रोटीन इंटरैक्शन की प्रिडिक्शन करने के लिए रैंडम फॉरेस्ट (मुर्मू एट अल., 2023), और ड्रग-टारगेट इंटरैक्शन प्रिडिक्शन के लिए सपोर्ट वेक्टर मशीन (यामानिशी एट अल., 2010) का उपयोग किया गया है।

कीटनाशक की खोज में, क्लॉगपी, हाइड्रोजन बॉन्ड डोनर (एचबीडी) और एरोमैटिक बॉन्ड (एआरबी) जैसे प्रमुख आणविक गुणों को कवकनाशी गतिविधि की प्रिडिक्शन

करने के लिए महत्वपूर्ण माना गया है (हाओ एट अल., 2011,)। अवराम एट अल. (2014) जैसे अध्ययनों ने आणविक भार, लॉगपी, हाइड्रोजन बॉन्ड स्वीकर्ता (एचबीए), एचबीडी, रोटेटेबल बॉन्ड (आरओबी) और एआरबी का उपयोग करके कीटनाशक-समानता के मात्रात्मक अनुमान स्थापित किए हैं। हालांकि, ये निष्कर्ष फंगल -नाशक प्रिडिक्शन में भौतिक-रासायनिक गुणों के महत्व पर जोर देते हैं, लेकिन लिपिंस्की के पांच के नियम पर पारंपरिक निर्भरता रासायनिक विविधता अन्वेषण को सीमित करती है। इससे पूर्वानुमान क्षमताओं को बढ़ाने के लिए मजबूत और विविध मापदंडों की आवश्यकता होती है। इस शोध में, हम फंगल लक्ष्य प्रोटीनों के विरुद्ध यौगिकों की जैव-सक्रियता का मूल्यांकन करने के लिए एक मशीन लर्निंग आधारित प्रिडिक्टर प्रस्तुत करते हैं।

सामग्री और विधियां

डेटा संग्रह और प्री-प्रोसेसिंग

आईसी50 मानों द्वारा दर्शाए गए जैव-सक्रियता डेटा को चेमब्ल डेटाबेस (https://www.ebi.ac.uk/chembl/) से प्राप्त किया गया। यह डेटा एकल फंगल प्रोटीन प्रकार के साथ इंटरैक्ट करने वाले यौगिकों पर केंद्रित था। यौगिकों को "सक्रिय" (आईसी50 $\leq$ 1000 nM) या "निष्क्रिय" (आईसी50 $\geq$ 10,000 nM) के रूप में वर्गीकृत किया गया, जिसमें मध्यवर्ती आईसी50 मान वाले यौगिकों को स्पष्ट वर्गीकरण बनाए रखने के लिए हटा दिया गया। प्रत्येक श्रेणी के लिए 8,000 यौगिकों का संतुलित डेटासेट तैयार किया गया। इस डेटासेट का 90% मॉडल विकास के लिए आवंटित किया गया, जिसमें पांच-फोल्ड क्रॉस-वैलिडेशन का उपयोग किया गया, और शेष 10% स्वतंत्र परीक्षण डेटासेट के रूप में आरक्षित रखा गया। आणविक अभ्यावेदन को सरलीकृत आणविक-इनपुट लाइन-एंट्री सिस्टम (SMILES) का उपयोग करके एन्कोड किया गया। आईसी50 मानों को नकारात्मक लघुगणकीय स्केल (-log10(आईसी50)) में परिवर्तित किया गया, ताकि डेटा को एक समान रूप दिया जा सके। PaDEL सॉफ्टवेयर (https://pypi.org/project/padelpy/) के माध्यम से आणविक वर्णनकर्ता (descriptors) की गणना की गई, जिससे 800 से अधिक फीचर्स उत्पन्न हुए। कम-वैरिएंस वाले वर्णनकर्ताओं (थ्रेशोल्ड = 0.1) को हटाकर फीचर चयन प्रक्रिया के माध्यम से इनकी संख्या को कम किया गया, जिससे गणनात्मक दक्षता और मॉडल प्रदर्शन में सुधार हुआ।

मॉडल विकास

जैव-सक्रियता की प्रिडिक्शन के लिए पाँच मशीन लर्निंग एल्गोरिदम का उपयोग किया गया:

- रैंडम फॉरेस्ट (आरएफ): शोर (noise) के प्रति मजबूत, बड़े डेटासेट के लिए प्रभावी, और ओवरफिटिंग को कम करता है।
- सपोर्ट वेक्टर मशीन (एसवीएम): उच्च-आयामी डेटा स्पेस और स्पष्ट रूप से परिभाषित डेटा पृथक्करण के लिए उत्कृष्ट है।
- के- नियरेस्ट नेबर (केएनएन): उदाहरण-आधारित और गैर-पैरामीट्रिक तकनीक, जो नियरेस्ट नेबर के बहुमत मत (majority voting) के आधार पर वर्गीकरण करता है।
- लिनियर डिस्क्रिमिनेंट एनालिसिस (एलडीए): वर्ग पृथक्करण (class separation) को अधिकतम करता है और गॉसियन-वितरित तथा रेखीय रूप से पृथक डेटा के लिए उपयुक्त है।
- एडा बूस्ट: कमजोर वर्गीकरणकर्ताओं (weak classifiers) को संयोजित करके कठिन-से-वर्गीकृत नमूनों पर केंद्रित होता है, जिससे सटीकता में सुधार होता है।

सर्वश्रेष्ठ प्रदर्शन करने वाले मॉडल को हाइपरपैरामीटर ट्यूनिंग के लिए ग्रिड-खोज (grid-search) विधि का उपयोग करके अनुकूलित किया गया।

प्रदर्शन मूल्यांकन

प्रदर्शन का मूल्यांकन सटीकता (Accuracy), प्रीसिशन

(Precision), रिकॉल (Recall), और F1-स्कोर जैसे मेट्रिक्स का उपयोग करके किया गया। स्वतंत्र परीक्षण डेटासेट का उपयोग करते हुए इष्टतम मॉडल की तुलना मौजूदा टूल फंगीपैड (https://dpai.ccnu.edu.cn/CoPLE/) से की गई।

परिणाम और चर्चा

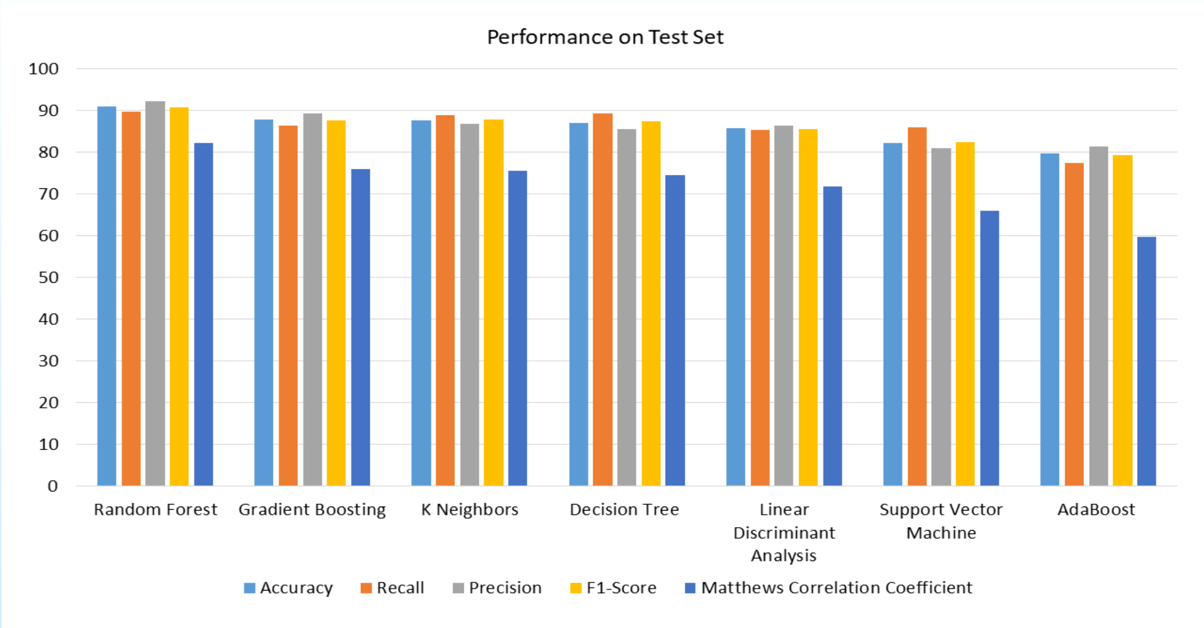
मशीन लर्निंग-आधारित जैव-सक्रियता पूर्वानुमान

ChEMBL से 4120 डेटासेट संकलित किए गए, जिन्हें प्री-प्रोसेसिंग के बाद 1454 सकारात्मक और नकारात्मक मामलों में परिष्कृत किया गया, जिससे जैव-सक्रियता वर्गों का संतुलन सुनिश्चित हुआ। आणविक संरचनाओं को एसएमआईएलईएस नोटेशन का उपयोग करके एन्कोड किया गया, और पडेल सॉफ्टवेयर का उपयोग करके 880 आणविक वर्णनकर्ता (molecular descriptors) उत्पन्न किए गए। फीचर चयन प्रक्रिया में 0.1 के निम्न-विविधता थ्रेशोल्ड को लागू करके वर्णनकर्ताओं को घटाकर 241 कर दिया गया, जिससे मॉडल की मजबूती में सुधार हुआ।

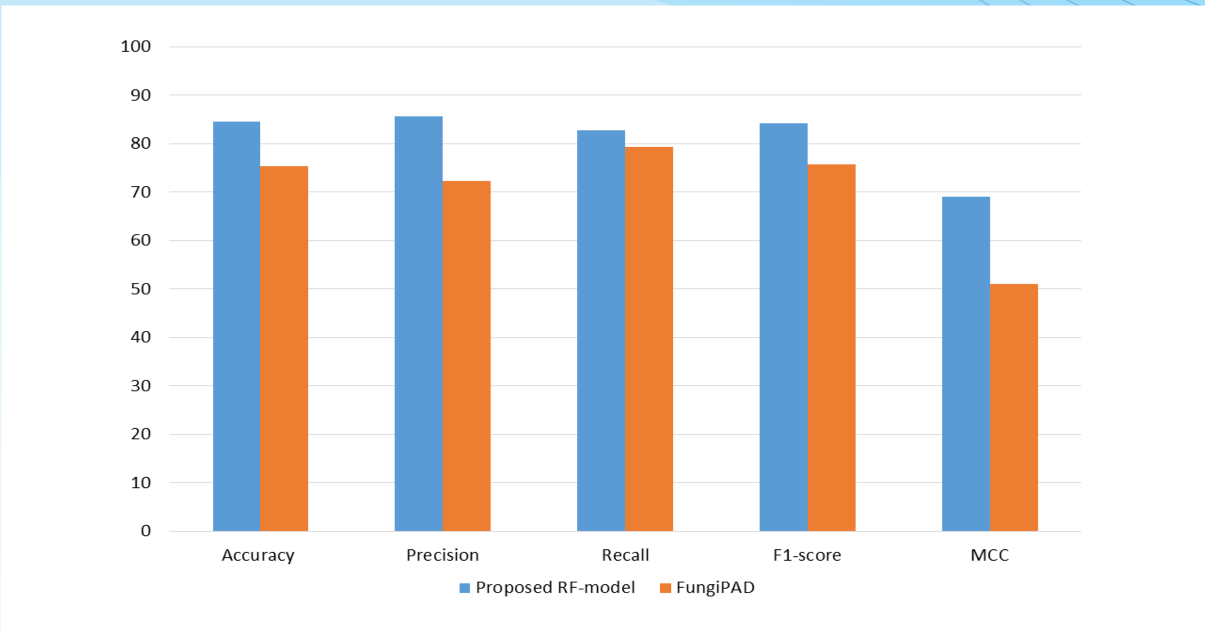
रैंडम फॉरेस्ट (RF) मॉडल ने सत्यापन (validation) सेट पर 97.04% और परीक्षण सेट पर 91.02% सटीकता प्राप्त की, जिससे यह सबसे अच्छा प्रदर्शन करने वाला वर्गीकरणकर्ता साबित हुआ (चित्र 1)। हाइपरपैरामीटर ट्यूनिंग से पता चला कि 'gini' मानदंड, 'sqrt' को max features के लिए और 20 n_estimators के संयोजन ने सर्वोत्तम प्रदर्शन प्रदान किया। इस सेटअप ने विभाजनों में विविध विशेषताओं का उपयोग सुनिश्चित किया, जिससे भिन्नता (variance) कम हुई और ओवरफिटिंग रोकी गई।

मौजूदा टूल्स के साथ तुलना

फंगल जैव-सक्रियता पूर्वानुमान के लिए उपलब्ध टूल सीमित हैं, जिनमें फंगीपैड (वांग एट अल., 2019) मुख्य टूल है। परीक्षण सेट पर रैंडम फॉरेस्ट (RF) मॉडल ने फंगीपैड को पछाड़ते हुए 84.5% सटीकता प्राप्त की, जबकि फंगीपैड ने 75.3% सटीकता हासिल की। फंगीपैड जहां एम डब्लू , लॉग पी, एच बी डी और एरोमैटिक रिंग्स जैसे पैरामीटरों का उपयोग करते हुए नियम-आधारित पूर्वानुमानों पर निर्भर करता है, वहीं प्रस्तावित आरएफ मॉडल मशीन लर्निंग, व्यापक फीचर चयन, और अनुकूलन (optimization) का उपयोग करता है। यह जटिल अंत:क्रियाओं को कैप्चर करता है, जिससे बेहतर प्रदर्शन सुनिश्चित होता है (चित्र 2)।



चित्र 1: परीक्षण सेट का उपयोग करके एमएल क्लासिफायर का प्रदर्शन।



चित्र 2. प्रस्तावित मॉडल की मौजूदा टूल के साथ तुलना।

निष्कर्ष

यह अध्ययन फंगल लक्ष्यों के विरुद्ध जैव-सक्रिय यौगिकों की प्रिडिक्शन के लिए एक नवीन मशीन लर्निंग-आधारित दृष्टिकोण प्रस्तुत करता है, जो फंगीपैड (FungiPAD) जैसे मौजूदा टूल्स की कमियों को दूर करता है। व्यापक फीचर चयन और हाइपरपैरामीटर अनुकूलन का उपयोग करके, प्रस्तावित रैंडम फॉरेस्ट मॉडल ने उत्कृष्ट सटीकता और विश्वसनीयता प्रदर्शित की। डॉकिंग और सिमुलेशन अध्ययनों के एकीकरण ने पूर्वानुमानित सक्रिय यौगिकों को और मान्यता प्रदान की, जिससे मॉडल की मजबूती को सुदृढ़ किया गया। यह पाइपलाइन कीटनाशक खोज के लिए संगणनात्मक विधियों में एक महत्वपूर्ण प्रगति का प्रतिनिधित्व करती है, जिसके कृषि में संभावित अनुप्रयोग हैं। भविष्य के कार्यों में रासायनिक विविधता का विस्तार करने और प्रिडिक्शन क्षमताओं को और बढ़ाने के लिए अतिरिक्त आणविक विशेषताओं को शामिल करने पर ध्यान केंद्रित किया जाएगा।

संदर्भ

अवराम, एस., फुनार-टिमोफी, एस., बोरोटा, ए., चेन्नामनेनी, एस. आर., मंचाला, ए. के., & मुरेसन, एस. (2014)। क्वांटिटेटिव एस्टीमेशन ऑफ पेस्टिसाइड-लाइसेंस फॉर एग्रोकेमिकल डिस्कवरी। जर्नल ऑफ केमइनफॉर्मेटिक्स, 6, 1-11।

हाओ, जी., डोंग, क्यू., & यांग, जी. (2011)। ए कम्पेरेटिव स्टडी ऑन द कोंस्टीटूटिव प्रॉपर्टीज ऑफ मार्केटेड पेस्टीसाइड्स। मॉलिक्यूलर इन्फॉर्मेटिक्स, 30(6, 7), 614-622।

जबीन, आर. (2006)। इवैल्यूएशन ऑफ बॉटनिकल्स अगेंस्ट जेनथोमोनास ओरिजा पी वी .ओरिजा कॉसिंग बैक्टीरियल लीफ ब्लाइट डिजीज ऑफ राइस एंड कैरेक्टराइजेशन ऑफ बायोएक्टिव कंपाउंड्स (डॉक्टोरल शोध, कायदे-आजम यूनिवर्सिटी इस्लामाबाद, पाकिस्तान)।

कुरू, के., एक्सार्कोस, टी. पी., एक्सार्कोस, के. पी., करामौजिस, एम. वी., & फोटियाडिस, डी. आई. (2015)। मशीन लर्निंग एप्लिकेशन्स इन कैंसर प्रोग्नोसिस एंड प्रिडिक्शन। कम्प्यूटेशनल एंड स्ट्रक्चरल बायोटेक्नोलॉजी जर्नल, 13, 8-17।

मा, जे., वांग, पी., हुआंग, एल., कियाओ, जे., & ली, जे. (2021)। बायोइन्फोर्मेटिक एनालिसिस रिवील्स ऑन एक्सोसोमल एमआईआरएनए- एमआरएनए नेटवर्क

इन कोलोरेक्टल कैंसर। बीएमसी मेडिकल जीनोमिक्स, 14, 1-18।

मुर्मू, एस., चौरसिया, एच., गुहा मजूमदार, एस., राव, ए. आर., राय, ए., - अर्चक, एस. (2023)। प्रिडिक्शन ऑफ प्रोटीनख्रप्रोटीन इंटरेक्शन्स बिटवीन एन्टी-क्रिस्पर एंड क्रिस्पर-कैस युसिंग मशीन लर्निंग टेक्नीक। जर्नल ऑफ प्लांट बायोकैमिस्ट्री एंड बायोटेक्नोलॉजी, 32(4), 818-830।

सादेघ, एस., स्केल्टन, जे., अनास्तासी, ई., बर्नेट, जे., ब्लूमेन्थल, डी. बी., गालिंदेज, जी., ... - काकप्रोव्स्की, टी. (2021)। नेटवर्क मेडिसिन फॉर डिजीज मॉडल आइडेंटिफिकेशन एंड ड्रग रेपुरपोसिंग विथ द नेड्रेक्स प्लेटफार्म। नेचर कम्युनिकेशन्स 12, नो. 1 (2021): 6848।

सीनियर, ए. डब्ल्यू., इवांस, आर., जंपर, जे., किर्कपैट्रिक, जे., सिफ्र, एल., ग्रीन, टी., ... & हसाबिस, डी. (2020)। इम्प्रूवड प्रोटीन स्ट्रक्चर प्रिडिक्शन युसिंग पोटेनटिअल्स फ्रॉम डीप लर्निंग। नेचर, 577(7792), 706-710।

वांग, एम. वाई., वांग, एफ., हाओ, जी. एफ., & यांग, जी. एफ. (2019)। फंगीपैडः ए फ्री वेब टूल फॉर कंपाउंड प्रॉपर्टी इवैल्यूएशन एंड फंजीसाइड -लाइकेनेस्स एनालिसिस। जर्नल ऑफ एग्रीकल्चरल एंड फूड केमिस्ट्री, 67(7), 1823-1830।

वांग, वाई., लियू, एक्स., शेन, वाई., सॉन्ग, एक्स., वांग, टी., शांग, एक्स., & पेंग, जे. (2023)। कोलेबोरेटिव डीप लर्निंग इम्प्रुव्स डिजीज-रिलेटेड सरकुलर आर एनए प्रिडिक्शन बेस्डऑन मल्टी -सोर्स फंक्शनल इनफार्मेशन। ब्रीफिंग्स इन बायोइंफॉर्मेटिक्स, 24(2), bbad069।

यामानिशी, वाई., अराकी, एम., गटरिज, ए., होंडा, डब्ल्यू., & केनेहिसा, एम. (2008)। प्रिडिक्शन ऑफ ड्रगख्र टारगेट इंटरेक्शन नेटवर्क्स फ्रॉम द इंटीग्रेशन ऑफ केमिकल एंड जीनोमिक स्पेसेस। बायोइन्फॉर्मेटिक्स, 24(13), i232-i240।

अर्जुन की छाल के अर्क का उपयोग करके रेशमी कपड़े की रंगाई

अभिलाषा रांगी[1], ललित जाजपुरा[2], राहुल बनर्जी[3]

[1]केन्द्रीय रेशम प्रौघोगिक अनुसंधान संस्थान, केन्द्रीय रेशम बोर्ड, वस्त्र मंत्रालय, भारत सरकार, बेंगलुरू

[2]वस्त्र प्रौद्योगिकी विभाग, डॉ. बी.आर.अम्बेडकर राष्ट्रीय प्रौद्योगिकी संस्थान, जालंधर

[3]भा.कृ.अनु.प.–भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, पूसा, नई दिल्ली

सार

वस्त्र रंगाई उद्योग जल निकायों में बड़ी मात्रा में रंगों और रसायनों के उत्सर्जन के कारण जल प्रदूषण में प्रमुख योगदानकर्ता है। वस्त्र उद्योग को पर्यावरण के अनुकूल बनाने के लिए प्राकृतिक रंगों ने शोधकर्ताओं का विशेष ध्यान आकर्षित किया है। पौधों के अर्क, जो पारंपरिक रूप से आयुर्वेद में उनके औषधीय गुणों के लिए उपयोग किए जाते रहे हैं, वस्त्रों को रंगने के लिए प्राकृतिक रंगों के रूप में उपयोग की क्षमता रखते हैं। टर्मिनलिया अर्जुन, जिसे सामान्यत: अर्जुन के रूप में जाना जाता है, एक ऐसा वृक्ष है जिसकी छाल जैव सक्रिय यौगिकों से भरपूर होती है। यह अध्ययन टर्मिनलिया अर्जुन की छाल के अर्क का उपयोग करके रेशमी वस्त्रों की रंगाई पर केंद्रित है। रंगाई की स्थितियों को प्रतिक्रिया सतह अभिकल्पनाओं का उपयोग करके अनुकूलित किया गया और रंगे हुए वस्त्रों का रंग स्थायित्व और गुणवत्ता मापदंडों के आधार पर मूल्यांकन किया गया। अध्ययन के परिणाम बताते हैं कि अर्जुन की छाल का अर्क, टैनिन और फेनोलिक यौगिकों से भरपूर होने के कारण, रेशमी वस्त्र पर आकर्षक रंग प्रदान करता है।

मुख्य शब्द : वस्त्र रंगाई उद्योग, जल प्रदूषण, प्राकृतिक रंग, पौधों के अर्क, जैव सक्रिय यौगिक, रेशमी वस्त्र, रंग स्थायित्व, गुणवत्ता मापदंड।

परिचय

ग्राहकों द्वारा उत्पादों के चयन में पर्यावरण–अनुकूलता एक महत्वपूर्ण कारक बन गई है। पर्यावरण संरक्षण के प्रति बढ़ती जागरूकता के कारण आजकल प्राकृतिक रंगों की मांग में तेजी आई है (हुड्डा और रांगी, 2015)। प्राकृतिक रंगों में कुछ सीमाएँ अवश्य हैं, लेकिन उनकी पर्यावरण–अनुकूल प्रकृति के कारण वे लोकप्रियता प्राप्त कर रहे हैं (जाजपुरा, पॉल और रांगी, 2016)। विश्व स्तर पर पर्यावरण–अनुकूल उत्पादों की बढ़ती मांग के कारण प्राकृतिक रंगों से रंगे हुए उत्पादों का निर्यात भी बढ़ रहा है। अधिकांश प्राकृतिक रंग पौधों के विभिन्न भागों से प्राप्त होते हैं। हालांकि, रंगों की गुणवत्ता और मात्रा से संबंधित अनुसंधानात्मक डेटा की कमी के कारण कई संभावित रंग देने वाले पौधे अनुपयोगी रह जाते हैं। यदि इन पौधों के बारे में वैज्ञानिक डेटा उपलब्ध हो, तो आयुर्वेद में औषधीय गुणों वाले कई पौधों को वस्त्रों के लिए प्राकृतिक रंगों के रूप में उपयोग किया जा सकता है। टर्मिनलिया अर्जुन (अर्जुन) एक ऐसा वृक्ष है जो बांग्लादेश, उत्तर प्रदेश, मध्य प्रदेश, पश्चिम बंगाल, दक्षिण और मध्य भारत में पाया जाता है। इस वृक्ष की छाल जैव सक्रिय यौगिकों से भरपूर होती है, जिससे प्राकृतिक रंग निकाला जा सकता है (मोर्शेद, 2011)। अर्जुन की छाल का अर्क प्राकृतिक रंग के रूप में कार्य करने की क्षमता रखता है। दिलचस्प बात यह है कि अर्जुन, एंथेरिया माइलिटा (वह रेशमकीट जो तसर रेशम का उत्पादन करता है) का मेजबान पौधा भी है (भवाने, 2007)। रेशमकीट के मेजबान पौधे की छाल से रंग निकालकर

उसे रेशम में उपयोग करने से इस प्रक्रिया में परिपत्रता (circularity) और स्थिरता (sustainability) को बढ़ावा मिलेगा। इसलिए, यह अध्ययन अर्जुन वृक्ष की छाल से प्राकृतिक रंग निकालने और उसे रेशमी वस्त्रों पर लागू करने पर केंद्रित है।

पद्धति

सामग्री

अर्जुन की छाल और रेशमी कपड़े स्थानीय बाजार से खरीदे गए। रंग निकालने के लिए धुली और सूखी हुई अर्जुन की छाल का उपयोग किया गया (रांगी और जाजपुरा, 2017)। अध्ययन में प्रयोग किए गए सभी कांच के बर्तन और रसायन प्रयोगशाला ग्रेड के थे, ताकि प्रयोग की शुद्धता और विश्वसनीयता सुनिश्चित की जा सके।

विधि

रेशम की रंगाई

सॉक्सलेट निष्कर्षण तकनीक का उपयोग करके, अर्जुन की छाल और पानी को 1:10 (सामग्री से पानी का अनुपात) को 8 पीएच पर 30 मिनट तक उबालने के बाद रंग निकाला गया। फिर अर्क का उपयोग सीधे रेशमी कपड़े को रंगने के लिए किया गया। रंगाई करने के समय की स्थिति जैसे समय, तापमान और पीएच कपड़े पर विकसित रंग पर बहुत महत्वपूर्ण भूमिका निभाते हैं। रंगाई स्थितियों को अनुकूलित करने के लिए 33 घटकीय बॉक्स बेनकेन (बीएनबी) प्रतिक्रिया सतह अभिकल्पना का उपयोग किया गया। कारक स्तर को समान रूप से दूरी पर रखकर निम्न, मध्यम एवं उच्च के रूप में कोडित किया गया जैसा कि तालिका 1 में दिखाया गया है। तापमान, पीएच और समय को तीन चरों में लिया गया और रंगे कपड़े के K/S मूल्य को प्रतिक्रिया कारक के रूप में उपयोग किया गया। सभी रंगाई प्रयोगों में 1:30 (MLR) सामग्री से पानी का अनुपात लेते हुए रंगाई की गई। डेटा का विश्लेषण करने के लिए डिजाइन विशेषज्ञ सॉफ्टवेयर संस्करण 12 का उपयोग किया गया। मॉडल समीकरण का सांख्यिकीय महत्व भी निर्धारित किया गया।

तालिका 1. रेशमी कपड़ों की रंगाई के लिए बॉक्स और बेनकेन डिजाइन में उपयोग की जाने वाली परिवर्तनशील स्थितियाँ

परिवर्तनशील स्थितियाँ	स्तर		
	निम्न −1	मध्यम 0	उच्च +1
A: तापमान (°C)	50	70	90
B: pH	4	6	8
C: अवधि (उपद)	60	90	120

रंगे कपड़े का निस्प्रण

रंग मान

रंगे कपड़ों की रंग मान को सी.आई.ई. (C.I.E.) प्रणाली के अनुसार कंप्यूटर रंग मिलान सॉफ्टवेयर पर मापा गया। नमूने द्वारा परावर्तित प्रकाश को उपकरण द्वारा मापा गया और रंग शक्ति मूल्यों में परिवर्तित किया गया। नीचे दिए गए सूत्र का उपयोग करके अनुपचारित और उपचारित नमूनों में रंग की गहराई (K/S) को मापने के लिए परावर्तन विधि (कटलर, 1965) का उपयोग किया गया।

$$\frac{K}{S}=\frac{(1-R)^2}{2R}$$

जहाँ,

R = नमूने से परावर्तन,

K = अवशोषण गुणांक,

S = रंगे कपड़ों का प्रकीर्णन गुणांक,

K/S मानों को C.I.E. रंग स्थान (L*, a* और b*) समन्वय के साथ मापा गया।

रंग स्थिरता

रंगे रेशम के नमूनों की धोने, रगड़ने और रोशनी में रंग स्थिरता की जाँच क्रमशः मानक परीक्षण विधियों IS/ISO 105-C10, IS/ISO 105 X12 और IS/ISO 105-B02 का उपयोग करके की गई। नमूनों को कंप्यूटर पर स्कैन कर रंग मिलान के बाद फास्टनेस रेटिंग दी गई।

परिणाम और चर्चा

रंगाई स्थितियों का परिणामों पर प्रभाव

जैसा कि खंड 2.2.1 में बताया गया है, रंगाई चर को तीन स्तर तीन घटकीय बॉक्स और बेनकेन डिजाइन का उपयोग करके अनुकूलित किया गया। K/S मान पर स्वतंत्र चर के प्रभाव को दर्शाते हुए प्राप्त परिणाम चित्र 1 में 3डी सतह प्लॉट के रूप में दिखाए गए हैं।

चित्र 1 से यह स्पष्ट है कि समय तीनों चरों में सबसे कम भूमिका निभाता है। यह देखा गया कि रंगाई प्रक्रिया को प्रभावित करने वाले महत्वपूर्ण कारक तापमान और पीएच हैं। जब समय 60 से 90 मिनट तक बढ़ाया गया तो रंग के अणु कपड़े में स्थानांतरित होते रहे, लेकिन 90 मिनट के बाद अवशोषण और विशोषण दोनों समान दर पर होने लगा, जिससे रंग की ताकत के मूल्यों में ज्यादा वृद्धि नहीं हुई (शेनाई, 1992)। जब तापमान बढ़ता है तो ऊर्जा भी बढ़ जाती है और इससे मौजूद अणुओं की गतिशीलता बढ़ जाती है। इसलिए जब रंग घोल का तापमान 500°C से 900°C तक बढ़ाया जाता है तो K/S में वृद्धि देखी गई। पीएच रंगाई प्रक्रिया को नियंत्रित करने वाला दूसरा सबसे महत्वपूर्ण कारक है जो नमूनों के K/S मूल्य पर भारी प्रभाव दिखाता है। पीएच मानक बड़े पैमाने पर रंगाई व्यवहार को प्रभावित करते हैं और यहां अम्लीय पीएच (4) पर अधिकतम K/S मानक पाया गया।

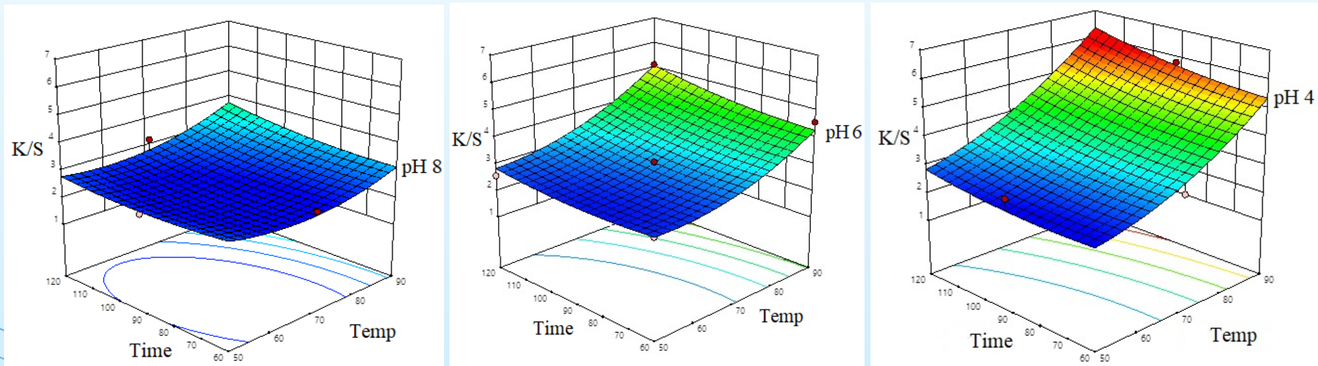
रंग मान और रंग स्थिरता रेटिंग

अनुकूलित परिस्थितियों का उपयोग करके रंगे गए नमूनों का रंग मान और स्थिरता रेटिंग तालिका 2 और 3 में दिखाई गई है। अर्जुन की छाल के अर्क से रेशम रंगों पर प्राप्त L (हल्का या गहरा) में 43.997 के L मान और a (11.274) और b (16.366) के सकारात्मक मूल्यों के साथ मध्यम गहराई वाली लाल-भूरे रंग की शेड प्राप्त हुई। धुआली के प्रति स्थिरता रेटिंग अच्छी पाई गई, लेकिन नम रगड़ स्थिरता रेटिंग खराब थी। प्रकाश के संपर्क में आने पर कुछ प्राकृतिक रंग ऑक्सीकृत होकर गहरे रंग के हो जाते हैं, जो रंगे हुए नमूनों में भी देखा गया है।

अर्जुन छाल अर्क से रंगे रेशम के नमूनों की रंग स्थिरता रेटिंग तालिका 3 में दी गई है। यह देखा जा सकता है कि

तालिका 2: रंगे रेशम के नमूने का रंग मान

	रंगाई की स्थितियाँ				रंग मान				
	तापमान (°C)	पीएच	समय (min-)	MLR	शेड	K/S	L	a	b
बिना रंगे रेशम का नमूना	-	-	-	-		0.36	84.03	0.97	13.46
रंगे रेशम का नमूना	90	4	90	1:30		5.80	44.00	11.27	16.36



चित्र 1: 3-डी सतह प्लॉट अर्जुन छाल के अर्क से रंगे रेशम की रंग शक्ति मूल्यों पर रंगाई स्थितियों के प्रभाव को दर्शाता है

तालिका 3: रंगे रेशम के नमूने की रंग स्थिरता रेटिंग

धोने के लिए रंग स्थिरता			रगड़ने पर रंग की स्थिरता		प्रकाश के प्रति रंग स्थिरता
फिका होना	मैला करना		शुष्क	नम	
	रेशम	सूत			
4	4	4-5	3-4	2-3	रंग गहरा होना

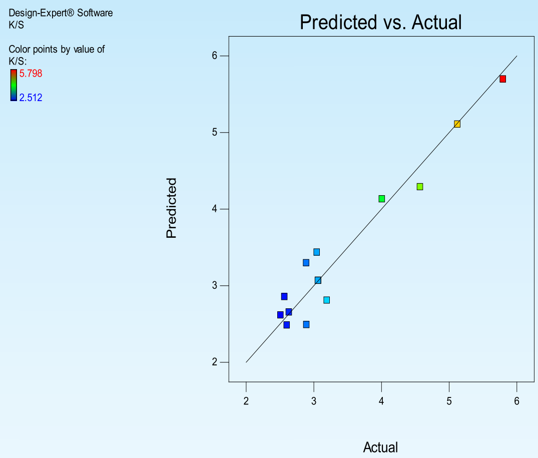
रंगे हुए नमूनों में धोने की स्थिरता अच्छी है किन्तु प्रकाश के संपर्क में आने पर रंगे हुए नमूनों का रंग फीका पड़ने के बजाय गहरा हो जाता है। कपड़े के नमूनों में मौजूद प्राकृतिक रंगों का ऑक्सीकरण प्रकाश स्थिरता परीक्षण के दौरान प्रकाश के संपर्क में आने पर कपड़े के गहरे रंग का होने का कारण हो सकता है।

प्रतिक्रिया सतह के लिए एनोवा

एनोवा मॉडल में प्राप्त r^2 0.9414 है जिसमें चर A, B, C, AB, A^2 के p-मान क्रमशः <0.001, 0.0024, 0.0461, 0.0083 और 0.0240 हैं, जो सभी 0.0500 से कम हैं। मॉडल का प्राप्त F मान 12.49 है जो मॉडल के महत्व को दर्शाते हुए काफी अधिक है क्योंकि त्रुटि के कारण इतना बड़ा मूल्य नहीं हो सकता है। चित्र 2 से पता चलता है कि के/एस रंग शक्ति के वास्तविक मूल्यों का प्लॉट अनुमानित मूल्यों के प्लॉट के पास बिखरा हुआ है।

निष्कर्ष

अर्जुन की छाल का अर्क टैनिन, फ्लेवेनॉइड्स और फेनोलिक यौगिकों से भरपूर होता है, जो इसे कपड़ों को रंगने के लिए एक प्रभावी प्राकृतिक डाई के रूप में उपयोग किए जाने की संभावनाएँ प्रदान करता है। इस अध्ययन में अर्जुन की छाल के अर्क का उपयोग करते हुए रेशम की रंगाई को प्रतिक्रिया सतह डिजाइन के माध्यम से अनुकूलित किया गया। अर्जुन की छाल से निकाला गया रंग रेशम को लाल-भूरे रंग का आकर्षक स्वरूप प्रदान करता है। स्थिरता परीक्षणों से यह पता चला कि रंगे हुए नमूने में धोने की स्थिरता अच्छी है, हालांकि प्रकाश के संपर्क में आने पर रंग का हल्का परिवर्तन देखा गया। इस अध्ययन के आधार पर यह निष्कर्ष निकाला जा सकता है कि अर्जुन की



चित्र 2: रंगे हुए नमूने के रंग शक्ति मानों के लिए अनुमानित बनाम वास्तविक मानों के लिए प्लॉट

छाल का अर्क प्राकृतिक और टिकाऊ रंग प्रदान करने में सक्षम है। यह जहरीले सिंथेटिक रंगों का एक सुरक्षित और पर्यावरण-अनुकूल विकल्प साबित हो सकता है।

सन्दर्भ

भवाने, जी., यालामाली, वी.डी., गायकवाड़, एस., वाघ, एस.आर., निकम, डी.एस., और भैसारे, एस.एस.। (2007). टर्मिनलिया कटप्पा; पोटेंशियल फूड प्लान्ट ऑफ ट्रॉपिकल तसर सिल्कवर्म, एन्थीरिया माइलिट्टा ड्रुरी, बुलेटिन ऑफ इण्डियन अकादमी ऑफ सेरीकल्चर, 11, 1-4।

कटलर, ए. ई. (1965). ए न्यू कलर मैचिंग कम्प्यूटर, जर्नल ऑफ द सोसाइटी ऑफ दयरस एंड कलॉरिस्ट, 81, 601।

हुड्डा, ए., और रांगी, ए. (2015). नैचरल डाइज: सस्टेनेबल वे फॉर डाइंग ऑफ टेक्सटाइल्स, जर्नल ऑफ बेसिक एंड एप्लाइड इंजीनियरिंग रिसर्च, 2(1), 43-46।

जाजपुरा, एल., पॉल, एस., और रांगी, ए. (2016). सस्टेनेबल डाइंग ऑफ कॉटन विथ पोमेग्रेनेट रिंड इन कंजंक्शन विथ नेचुरल मॉरडंट एंड बायोपॉलीमर काइटोसन, मैन मेड टेक्सटाइल्स इंडिया, 44, 180।

मोर्शेद, एम. ए., उद्दीन, एम. ए., हसन, टी., अहमद, टी., उद्दीन, एफ., जकारिया, एम., एट.ऑल. (2011). इवैल्यूएशन ऑफ एनाल्जेसिक एंड एंटी इन्फ्लैमटॉरी

इफेक्ट ऑफ टेर्मिनेलिया अर्जुना इथेनॉल एक्सट्रैक्ट, इंटरनेशनल जर्नल ऑफ फार्मास्यूटिकल साइंस एंड रिसर्च, 2(5), 1182-1187।

रांगी, ए., और जाजपुरा, एल. (2017). इफेक्ट ऑफ एमएलआर, टेम्परेचर एंड साल्वेंट ऑन द एक्सट्रैक्शन ऑफ कलॉरैशन फ्रॉम अर्जुन (टर्मिनलिया अर्जुना) बार्क, जर्नल ऑफ नैचरल रिसोर्सेज एंड एग्रीकल्चरल ईकोसिस्टम्स, 4, 87–89।

शेनाई, वी. के. (1992). प्रिंसिपल्स ऑफ डाइंग, सेवक पब्लिकेशन्स।

सिंगल-सेल सिक्वेंसिंग प्रक्रिया एवं विश्लेषण का संक्षिप्त विवरण

सुधीर श्रीवास्तव, मयंक रश्मि, दीपा भट्ट, द्विजेश चंद्र मिश्र, स्नेहा मुर्मू, मोहम्मद समीर फारूकी, कृष्ण कुमार चतुर्वेदी, गिरीश कुमार झा

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

सिंगल-सेल आरएनए-सिक एक शक्तिशाली विधि के रूप में विकसित हुआ है जो जटिल ऊतकों का वर्णन करने में सक्षम है और जिसे बल्क आरएनए-सिक द्वारा संबोधित नहीं किया जा सकता है। यह कोशिश की जैविक प्रक्रिया के ज्ञान को बढ़ाने में 5 सबसे अधिक इस्तेमाल की जाने वाली तकनीक है। यहां हम सेल आइसोलेशन और लाइब्रेरी प्रिपरेशन के लिए उपलब्ध एकल-कोशिका प्रौद्योगिकियों का अवलोकन और सिंगल-सेल आरएनए-सिक डेटा का विश्लेषण करने की प्रक्रिया को कवर करते हैं, जिसमें रीड मैपिंग, क्वालिटी कंट्रोल, जीन एक्सप्रेशन, सामान्यीकरण, फीचर चयन, आयामीता में कमी और सेल क्लस्टरिंग शामिल है। प्लांट साइंस में सिंगल-सेल ट्रांसक्रिप्टोम के विस्तार को देखते हुए इसे फसल विज्ञान में सेल की पहचान के लिए तेजी से लागू किया जाना चाहिए ताकि फसल उत्पादन को बढ़ाके और विस्तारित करके अधिक उत्पादन एवं नई किस्मों को विकसित किया जा सके।

मुख्य शब्द: क्लस्टरिंग, डिफरेंशियल एक्सप्रेशन विश्लेषण, नोर्मलाइजेशन, आरएनए-सिक डेटा, सिंगल-सैल

परिचय

सिंगल-सेल सिक्वेंसिंग एक विधि है जो एक कोशिका के ट्रांसक्रिप्टोम और जीनोम को अनुक्रमित करने के लिए उपयोग की जाती है। यह ट्रांसक्रिप्टोमिक्स, जीनोमिक्स और अन्य ओमिक्स जानकारी तक पहुंचने के लिए उपयोगी है, ताकि कोशिका समूहों, कोशिका प्रतिक्रियाओं और कोशिकाओं के बीच इवोल्यूशनरी संबंधों को समझा जा सके। सिंगल सेल आरएनए सिक्वेंसिंग (scRNA-Seq) का अध्यन पहली बार 2009 में टैंग और उनके समूह द्वारा किया गया था। अब तक, सिंगल सेल ट्रांसक्रिप्टोमिक अध्ययनों के लिए कई तकनीकें प्रस्तावित हैं और इन्हें सेल आइसोलेशन, सेल लाइसिस, रिवर्स ट्रांसक्रिप्शन, एम्पलीफिकेशन, ट्रांसक्रिप्ट कवरेज, स्ट्रैंड स्पेसिफिसिटी और यूनिक मॉलिक्यूलर आइडेंटिफायर (यूएमआई/UMI) जैसे कम से कम एक पहलू द्वारा विभेदित किया जाता है। UMI अद्वितीय ट्रांसक्रिप्ट की उपलब्धता का पता लगाता है और उसकी मात्रा निर्धारित करता है।

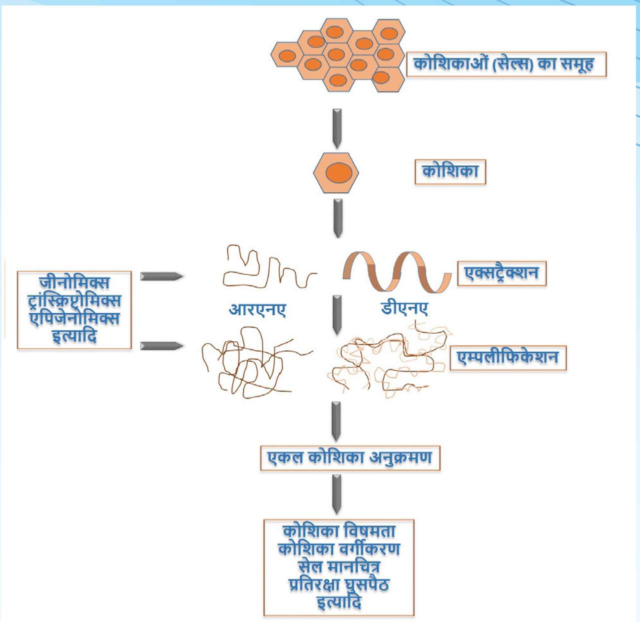
सिक्वेंसिंग प्रक्रिया

सिंगल सेल सिक्वेंसिंग एक ऐसी प्रक्रिया है जिसमें सिक्वेंसिंग के लिए एक कोशिका को आइसोलेट किया जाता है और फिर मॉलिक्यूलर मैपिंग, कोशिका हेट्रोजेनिटी, एपिजेनेटिक परिवर्तन और प्रतिरक्षा इंफिल्ट्रेशन का अध्ययन किया जाता है। सबसे पहले, रुचि के अनुसार लिए गए ऊतक सैम्पल से एकल कोशिकाओं को अलग किया जाता है, जिसमें माइक्रो-डाइसेक्शन, माइक्रोफ्लूइडिक प्लेटफॉर्म और ड्रॉपलेट-आधारित विधियों जैसे कुछ चरण शामिल होते हैं। इसके बाद, एक कोशिका का लाइसिस किया जाता है ताकि कोशिकीय एम-आरएनए अलग हो सके।

एम-आरएनए अणुओं को अन्य कोशिका अशुद्धियों से अलग करने के लिए पॉली(टी) सीक्वेंस प्राइमरों का उपयोग किया जाता है, जो एम-आरएनए के पॉली(ए) टेल्स से जुड़ते हैं। इसके बाद, पॉली(टी)-प्राइम एम-आरएनए को

रिवर्स ट्रांसक्रिप्शन का उपयोग करके सीडीएनए में परिवर्तित किया जाता है। यह रिवर्स ट्रांसक्राइब किया गया सीडीएनए आमतौर पर पीसीआर या इन-विट्रो ट्रांसक्रिप्शन के माध्यम से प्रवर्धित किया जाता है। सीडीएनए लाइब्रेरी तैयार करते समय उसमें इंडेक्स न्यूक्लियोटाइड बारकोड जोड़े जाते हैं, जो प्रत्येक लाइब्रेरी की पहचान करते हैं। इसके बाद, सीडीएनए लाइब्रेरी को पूल किया जाता है और नेक्स्ट जेनरेशन सीक्वेंसिंग तकनीक का उपयोग करके सिक्वेंसिंग लाइब्रेरी तैयार की जाती है (चित्र 1)।

एकल-कोशिका सिक्वेंसिंग लाइब्रेरी तैयार करने के लिए अब कई वाणिज्यिक किट और अभिकर्मक (Reagent) उपलब्ध हैं, जो कि संपूर्ण वेट-लैब प्रक्रिया पूर्ण करने में उपयोगी हैं। इनमें शुरुआत (कोशिका लाइसिस) से लेकर समाप्ति (सिक्वेंसिंग) तक सभी चरण हैं। इनमें से कुछ एमआरएनए आइसोलेशन, रिवर्स ट्रांसक्रिप्शन और सीडीएनए प्रवर्धन के लिए आरएनए टेम्पलेट के 5' छोर पर स्विचिंग मैकेनिज्म (स्मार्टर) प्रक्रिया पर आधारित हैं। हाल के समय में, कई ड्रॉपलेट-आधारित प्लेटफॉर्म जैसे 10x जीनोमिक्स का क्रोमियम, बायो-रेड लेबोरेटरीज का डीडीसिक, 1सेलबायो का इनड्रॉप तथा डोलोमाइट बायो/ब्लैकट्रेस होल्डिंग्स का μ-एनकैप्सुलेटर वाणिज्यिक रूप से उपलब्ध हैं।



चित्र 1. सिंगल-सेल सिक्वेंसिंग विधि का प्रवाह संचित्र

सिंगल-सेल आरएनए-सिक डेटा विश्लेषण

अनुक्रमित डेटा का विश्लेषण बायोइन्फॉर्मेटिक्स और कम्प्यूटेशनल पाइपलाइन का उपयोग करके विभिन्न सॉफ्टवेयर और टूल्स की मदद से किया जाता है। सिंगल-सेल आरएनए-सिक डेटा विश्लेषण के सभी चरण नीचे दिए गए हैं (चित्र 2):

चित्र 2. सिंगल-सेल सिक्वेंसिंग डेटा विश्लेषण प्रक्रिया के विभिन्न चरण

- लाइब्रेरी प्रीपरेशन विधि

3' सिरे या 5' सिरे के आरएनए सीक्वेंसेस रीड्स या टैग: 10X जीनोमिक्स, सेल-सिक2, ड्रॉप-सिक, इन-ड्रॉप्स पूर्ण -लंबाई के आरएनए ट्रांसक्रिप्ट: स्मार्ट-सिक

- सीक्वेंस रीड्स डेटाबेसेस

जी.इ.ओ. (GEO), एस.आर.ए. (SRA), एससीआरएनए -सिकडीबी (scRNASeqDB), प्लांट्सएससीआरएनएडीबी (plantscRNAdb)

- गुणवत्ता नियंत्रण

फास्ट क्यू सी (FastQC), सिन क्यू सी (SinQC), स्केटर (Scater)

- काउंट मैट्रिक्स जेनरेशन

सेल रेंजर (Cell Ranger) [10x Genomics], फीचरकाउंट्स (FeatureCounts) [Subread], एल्विन फ्राई (Alevin–fry) [Salmon], एचटी सिक-काउंट (HTSeq-count), यूएमआई टूल्स (umi-tools), स्टार सोलो (STARsolo) [STAR]

- बैच इफेक्ट करेक्शन

  - एम.एन.एन. म्यूच्यूअल नियरेस्ट नेबर (MNN-Mutual Nearest Neighbor) - विभिन्न बैच में समान कोशिकाओं के लिए प्रयोग किया जाता है
  - केबेट (kBET) - χ^2 विधि पर आधारित है

- नोर्मलाइजेशन

  - लॉग-नोर्मलाइजेशन/Log-Normalization [Seurat (NormalizeData function] method= "LogNormalize"), scanpy (normalize_total function)]
  - काउंट्स पर मिलियन/ CPM (Counts Per Million) [edgeR using the cpm function]
  - ट्रांसक्रिप्ट्स पर मिलियन/ TPM (Transcripts Per Million) [salmon and kallisto]
  - स्क्रैन का डीकॉनवोलुशन नोर्मलाइजेशन/ Scran's Deconvolution Normalization [scran (R package), Function: computeSumFactors],
  - SCTransform [Seurat (SCTransform function)]

- क्लस्टरिंग

  - पीसीए (Principal Component Analysis) - लीनियर डाइमेंशनलिटी रिडक्शन के लिए
  - टी-एसएनइ (t-distributed Stochastic Neighbor Embedding) - नॉन-लीनियर डाइमेंशनलिटी रिडक्शन के लिए

- मार्कर की पहचान, डिफरेंशियल एक्सप्रेशन विश्लेषण और अन्य विश्लेषण

महत्वपूर्ण मार्करों की पहचान के लिए विभिन्न टूल्स और तकनीकें उपलब्ध हैं। इनकी सहायता से एकल-कोशिका स्तर पर प्रोटीन-कोडिंग और नॉन-कोडिंग आरएनए दोनों के एक्सप्रेशन का व्यापक रूप में पता लगा सकते हैं।

सिंगल सेल आरएनए-सिक डेटा विश्लेषण के मुख्य उपयोग इस प्रकार हैं:

- कोशिकाओं की विविधता को जानने के लिए एक निश्चित स्थिति या ऊतक के भीतर कोशिका सब-पापुलेशन (अलग-अलग पापुलेशन में अलग-अलग प्रकार की कोशिका होती हैं) की पहचान करना।
- अलग-अलग सब-पापुलेशन या कोशिकाओं के समूहों

तालिका 1. सिंगल-सेल आरएनए-सिक डाटा विश्लेषण के लिए टूल्स / पैकेजेस की सूची

टूल्स / सर्वर्स / पैकेजेस	प्लेटफार्म
सिंगल सी ऐनेलाईजर (SingleCAnalyzer)	क्लाउड-बेस्ड सिस्टम
एस सी टाइपर (scTyper)	आर पैकेज
सेलेनिक्स (Cellenics)	वेब एप्लीकेशन
गैलेक्सी (Galaxy)	सिंगल-सेल ओमिक्स वर्कबेंच
सेलस्नेक (Cellsnake)	ओपन-सोर्स टूल
एस सी आरएनए- एक्स्प्लोरर (scRNA-Explorer)	एन्ड-यूजर ऑनलाइन टूल
स्युरॉट (Seurat)	आर पैकेज

के बीच महत्वपूर्ण एंव भिन्न रूप से व्यक्त जीन का पता लगाना।

- अल्टरनेटिव स्प्लाइसिंग के लिए, आम तौर पर पांच बुनियादी मोड होते हैं जिनमें एक्सॉन-स्किपिंग (कैसेट एक्सॉन), म्यूचुअल्ली एक्सक्लूसिव एक्सॉन, अल्टरनेटिव डोनर साइट, अल्टरनेटिव एक्सेप्टर साइट और इंट्रॉन रिटेंशन शामिल हैं। अल्टरनेटिव स्प्लिसिंग के लिए कुछ टूल्स जैसे सिंगलस्प्लिस, सेंसस, बीआरआईई और एक्सपीडिशन हैं।
- आरएनए-एडिटिंग डिटेक्शन डायनामिक्स का पता लगाना।
- एलीलिक एक्सप्रेशन एनालिसिस द्वारा पैरेंटल और मैटरनल जीनोम के बीच समान रूप से एक्सप्रेस जीन की पहचान करना।
- वेटेड जीन को-एक्सप्रेशन नेटवर्क विश्लेषण (डब्ल्यूजीसीएनए) का उपयोग करके कई सैम्पल्स में जीन रेगुलेटरी नेटवर्क इन्फेरेन्स का पता लगाना।

निष्कर्ष

पिछले 13 वर्षों में, विभिन्न प्रकार के सिंगल-सेल आरएनए-सिक प्रोटोकॉल विकसित किए गए हैं जो सेल एक्सप्रेशन वेरिएबिलिटी और सिंगल-सेल आरएनए-सिक डेटा की डायनामिक्स में सहायता प्रदान करते हैं। सिंगल-सेल आरएनए-सिक विधियाँ जिनका उपयोग फिक्सेशन प्रक्रिया और फ्रोजेन नमूनों द्वारा अत्यधिक हेटेरोजिनस क्लीनिकल नमूनों का अध्ययन करने के लिए किया जा सकता है। सिंगल सेल आरएनए-सिक डेटा का कम्प्यूटेशनल विश्लेषण सिंगल सेल ट्रांसक्रिप्टॉमिक्स के विकास को बढ़ावा देता है और जैविक और क्लिनिकल रिसर्च की सुविधा प्रदान करता है। यह जीन एक्सप्रेशन की हेट्रोजेनिटी और कोशिकाओं की डाइनेमिक्स में गहरी अंतर्दृष्टि भी प्रदान करता है।

संदर्भ

टैंग, एफ., बारबाशियोरू, सी., वांग, वाई., नॉर्डमैन, ई., ली, सी., जू, एन., एट आल. (2009) एमआरएनए-सेक होल-ट्रांस्क्रिप्टोम एनालिसिस ऑफ आ सिंगल सेल। नेचर मेथड्स. 6, 377-382।

इलिसिक, टी., किम, जे.के., कोलोडजीज्जिक, ए.ए., बैगर, एफ.ओ., मैक्कार्थी, डी.जे., मैरियोनी, जे.सी., एट अल, (2016) क्लासिफिकेशन ऑफ लो क्वालिटी सेल्स फ्रॉम सिंगल-सेल आरएनए-सेक डाटा। जीनोम बायोलॉजी. 17:29।

हाघवेर्डी, एल., लून, ए.टी.एल., मॉर्गन, एम.डी., मैरियोनी, जे.सी. (2018) बैच इफेक्ट्स इन सिंगल-सेल आरएनए-सिक्वेंसिंग डाटा आर करेक्टेड बाइ मैचिंग म्यूच्यूअल नीअरेस्ट नेइबोर्स। नेचर बायोटेक्नोलॉजी, 36, 421-427।

बटनर, एम., मियाओ, जेड., वोल्फ, एफ.ए., टीचमैन, एस.ए., थीस, एफ.जे. (2019) अ टेस्ट मीट्रिक फॉर अस्सेस्सिन्ग सिंगल-सेल आरएनए-सेक बैच करेक्शन। नेचर मेथड्स, 16, 43-49।

प्रीतो, सी., बाररियस, डी., विलवे,र्डे ए. (2022) सिंगल सी ऐनेलाईजर : इंटरैक्टिव एनालिसिस ऑफ सिंगल सेल आरएनए-सेक डाटा आन दी क्लाउड। फ्रंटियर्स इन बायोइन्फरमेटिक्स, 2, 793309।

चोइ, जे. एच., इन किम, एच., वो, एच. जी. (2020) एस सी टाइपर: अ कम्प्रेहैन्सिव पाइपलाइन फॉर दी सेल टाइपिंग एनालिसिस ऑफ सिंगल-सेल आरएनए-सेक डाटा। बी एम सी बायोइन्फरमेटिक्स, 21, 342।

टेकमन, एम., बटूट, बी., ओस्ट्रोवस्की, ए., अंतोनिएव्सकी, सी., क्लेमेंट्स, डी., रामिरेज, एफ., एथेरिंगटोन, जी. जे., होट्ज, एच. आर. (2020) अ सिंगल-सेल आरएनए -सिक्वेंसिंग ट्रेनिंग एंड एनालिसिस सुईट युसींग दी गैलेक्सी फ्रेमर्वा गीगाससिएन्स, 9(10), giaa102।

बालटसविआ, आइ., आउलास, ए., थेओडोसीओ, टी., लैविग्ने, एम.डी., अन्डरीकोस, इ., मव्रोथैलसिटिस, जी., इलीओपोलोस, आई. (2024) एस सी आरएनए-एक्स्प्लोरर : एन एन्ड-यूजर ऑनलाइन टूल फॉर सिंगल सेल आरएनए-सेक डाटा एनालिसिस फीचरिंग जीन कोरिलेशन एंड डाटा फिल्टरिंग. जर्नल ऑफ मॉलिक्यूलर बायोलॉजी, 436, 16865।

खाद्य विज्ञान एवं पोषण में सांख्यिकीय विश्लेषण तकनीकें

डॉ. निधि जोशी

विषय वस्तु विशेषज्ञ

कृषि विज्ञान केन्द्र, भा.कृ.अनु.प.-केंद्रीय कृषि अभियांत्रिकी संस्थान, भोपाल

सार

खाद्य विज्ञान और पोषण अनुसंधान डेटा-संचालित सांख्यिकीय विश्लेषण पर निर्भर करता है, जो खाद्य उत्पादों, आहार की आदतों और उपभोक्ता वरीयताओं की हमारी समझ को बढ़ाता है। खाद्य विज्ञान और पोषण से सम्बंधित सांख्यिकीय विश्लेषणों में, प्रतिगमन विश्लेषण (रिग्रेशन एनालिसिस), खोजपूर्ण कारक विश्लेषण (एक्सप्लोरैटरी फैक्टर एनालिसिस), पुष्टिकरण कारक विश्लेषण (कन्फर्मेटरी फैक्टर एनालिसिस), प्रमुख घटक विश्लेषण (प्रिंसिपल कंपोनेंट एनालिसिस) एवं क्लस्टर विश्लेषण जैसी तकनीकें अपरिहार्य हैं। यह विधियाँ रिश्तों को उजागर करती हैं, सर्वेक्षण उपकरणों को परिष्कृत करती हैं, आहार प्रणाली का विश्लेषण करती हैं, और उपभोक्ताओं को प्रभावी ढंग से समूहों में विभाजित करने में सहायता करती हैं।

यह लेख खाद्य विज्ञान और पोषण अनुसंधान में मल्टीवेरिएट सांख्यिकीय विश्लेषण के अनुप्रयोग पर संक्षिप्त जानकारी प्रदान करता है।

प्रतिगमन विश्लेषणः रिश्तों को समझना

प्रतिगमन विश्लेषण एक सांख्यिकीय तकनीक है जिसका उपयोग स्वतंत्र एवं आश्रित चरों के बीच संबंधों को पहचानने एवं निर्धारित करने के लिए किया जाता है। इसका व्यापक रूप से खाद्य विज्ञान और पोषण में उपयोग किया जाता है।

1. खाद्य सामग्री का उत्पाद विशेषताओं पर प्रभावः एकाधिक प्रतिगमन द्वारा वसा, चीनी या फाइबर की मात्रा में भिन्नता का खाद्य उत्पादों की बनावट, स्वाद एवं शेल्फ जीवन पर प्रभाव का विश्लेषण किया जा सकता है।
2. स्वास्थ्य परिणामों का अध्ययनः प्रतिगमन मॉडल द्वारा आहार संबंधी कारकों का स्वास्थ्य पर प्रभाव का अध्ययन किया जाता है। उदाहरण के लिए, रैखिक प्रतिगमन द्वारा फल और सब्जी के सेवन एवं बॉडी मास इंडेक्स के बीच संबंधों की जांच की जा सकती है।
3. पूर्वानुमानित मॉडलिंगः विशिष्ट चर के आधार पर उपभोक्ता संतुष्टि या पोषक तत्वों की मात्रा का पूर्वानुमान के द्वारा उत्पाद विकास एवं आहार हस्तक्षेप के लिए रणनीति तैयार करने में मदद मिलती है।

उदाहरण

उम्र एवं आहार वरीयताओं को नियंत्रित करते हुए, प्रतिगमन विश्लेषण के द्वारा नमक की मात्रा को कम करके विभिन्न सोडियम स्तर उपभोक्ता स्वीकृति को कैसे प्रभावित करते हैं, इसका अध्ययन किया जा सकता है।

खोजपूर्ण एवं पुष्टिकरण कारक विश्लेषणः प्रश्नावली परिशोधन

खाद्य विज्ञान और पोषण में सर्वेक्षण-आधारित अनुसंधान किए जाते हैं, विशेष रूप से उपभोक्ता व्यवहार, आहार की आदतों एवं खाद्य उत्पादों की धारणाओं को समझने के

लिए। प्रश्नावली को तैयार करने एवं मान्यता देने के लिए कारक विश्लेषण तकनीक का प्रयोग किया जाता है।

खोजपूर्ण कारक विश्लेषण (एक्सप्लोरैटरी फैक्टर एनालिसिस): अंतर्निहित कारकों की खोज

एक्सप्लोरैटरी फैक्टर एनालिसिस के द्वारा सर्वेक्षण प्रश्नावली में सम्बंधित प्रश्नों को समूहीकृत करके अंतर्निहित कारकों की पहचान की जाती है। उदाहरण के लिए, जैविक खाद्य पदार्थो के प्रति उपभोक्ता दृष्टिकोण का आकलन करने वाली प्रश्नावली स्वास्थ्य जागरूकता, पर्यावरण जागरूकता एवं मूल्य संवेदनशीलता से सम्बंधित कारकों को प्रकट कर सकती है।

पुष्टिकरण कारक विश्लेषण (कन्फर्मेटरी फैक्टर एनालिसिस): कारक संरचनाओं को मान्य करना

पुष्टिकरण कारक विश्लेषण के द्वारा सर्वेक्षण उपकरण की वैधता निर्धारित की जा सकती है।

- अनुप्रयोगों:
  - आहार स्मरण उपकरण को मान्य करना।
  - उपभोक्ता वरीयताओं को मापने वाले उपकरण की विश्वसनीयता का आकलन करना।

उदाहरण

खाद्य अपशिष्ट दृष्टिकोण से सम्बंधित अंतर्निहित कारकों पर्यावरणीय चिंता एवं वित्तीय जागरूकता की पहचान करने के लिए खोजपूर्ण कारक विश्लेषण का उपयोग किया जा सकता है एवं पुष्टिकरण कारक विश्लेषण द्वारा इन कारकों की विश्वसनीयता की पुष्टि की जा सकती है।

प्रधान घटक विश्लेषण: आहार प्रणाली का विश्लेषण

प्रधान घटक विश्लेषण (प्रिंसिपल कंपोनेंट एनालिसिस) आयामों को कम करने वाली तकनीक है जो बड़े डेटासेट को घटकों के एक छोटे सेट में बदलती है। पोषण अनुसंधान में प्रधान घटक विश्लेषण का उपयोग अक्सर किया जाता है।

1. आहार प्रणाली की पहचान करना: प्रधान घटक विश्लेषण जटिल आहार डेटा (जानकारी) को सरल करता है, "पश्चिमी आहार" (प्रसंस्कृत खाद्य पदार्थ) या "भूमध्य आहार" (फलों, सब्जियों और स्वस्थ वसा से भरपूर) जैसे पैटर्न की पहचान करता है।
2. पोषक तत्वों के सेवन का अध्ययन: पोषक तत्वों के सेवन का विश्लेषण करके, प्रधान घटक विश्लेषण संबंधित पोषक तत्वों के समूहों की पहचान करने में मदद करता है।
3. सार्वजनिक स्वास्थ्य रणनीतियों को सूचित करना: प्रधान घटक विश्लेषण पोषण शिक्षा अभियानों और नीति विकास का मार्गदर्शन कर सकता है।

क्लस्टर विश्लेषण: उपभोक्ता समूहों की पहचान करना

क्लस्टर विश्लेषण व्यक्तियों या वस्तुओं की विशेषताओं में समानता के आधार पर समूहों में विभाजित करता है। यह उपभोक्ताओं को विभाजित करने में विशेष रूप से उपयोगी है, जिससे उत्पाद विकास और विपणन योजना बनाने में सहायता मिलती है।

खाद्य विज्ञान में अनुप्रयोग

1. उपभोक्ता विभाजन: आहार वरीयताओं, खरीद की आदतों या सतत आहार के प्रति दृष्टिकोण के आधार पर समूहों की पहचान करना।
2. उत्पाद अनुकूलन: विशिष्ट समूहों के लिए उत्पाद अनुकूलन करना, उदाहरणार्थ स्वास्थ्य के प्रति जागरूक या बजट-संवेदनशील उपभोक्ता।
3. आहार अनुसंधान: जनसंख्या उपसमूहों के बीच पोषक तत्वों के सेवन का विश्लेषण करना।

क्लस्टर विश्लेषण के प्रकार

- के(K)- मीन्स क्लस्टरिंग: क्लस्टर के भीतर विचरण को कम करके k समूहों में डेटा को विभाजित करता है।
- पदानुक्रमित क्लस्टरिंग: समूहों के बीच संबंधों के अध्ययन के द्वारा पेड़ की आकर की संरचना का निर्माण करता है।

उदाहरण

पादप आधारित मांस के विकल्प के प्रति दृष्टिकोण का अध्ययन क्लस्टर विश्लेषण द्वारा किया जा सकता है, जोकि उपभोक्ताओं को समूहों में विभाजित करता है।

व्यापक अंतर्दृष्टि प्रदान करने के लिए तकनीकों को एकीकरण

मल्टीवेरिएट सांख्यिकीय विश्लेषण को संयोजित करके समृद्ध विश्लेषण जानकारी प्राप्त की जा सकती है। उदाहरणार्थः

1. प्रतिगमन एवं प्रधान घटक विश्लेषणः आहार प्रणाली की पहचान करने के लिए प्रधान घटक विश्लेषण का उपयोग करना तथा चिन्हित प्रणालियों को स्वास्थ्य परिणामों से जोड़ने के लिए प्रतिगमन विश्लेषण का उपयोग करना।
2. खोजपूर्ण एवं पुष्टिकरण कारक विश्लेषण तथा क्लस्टर विश्लेषणः खोजपूर्ण एवं पुष्टिकरण कारक विश्लेषण का उपयोग करके एक मान्य प्रश्नावली विकसित की जा सकती है। उपभोक्ताओं को उनके दृष्टिकोण या व्यवहार के आधार पर क्लस्टर विश्लेषण द्वारा विभाजित किया जा सकता है।
3. प्रधान घटक विश्लेषण एवं क्लस्टर विश्लेषणः बहुआयामी आहार डेटा के आधार पर प्रधान घटक एवं क्लस्टर विश्लेषण के द्वारा व्यक्तियों को समूहों में विभाजित किया जा सकता है।

चुनौतियां एवं भविष्य की संभावनाएँ

चुनौतियां

1. डेटा गुणवत्ताः सटीक एवं प्रतिनिधि डेटा संग्रह सुनिश्चित करना।
2. अंतःविषय ज्ञानः मल्टीवेरिएट सांख्यिकीय विश्लेषण के प्रभावी अनुप्रयोग के लिए सांख्यिकीविदों, खाद्य वैज्ञानिकों एवं पोषण विशेषज्ञों के बीच परस्पर सहयोग की आवश्यकता।
3. जटिलताः बहुआयामी डेटा का विश्लेषण करने के लिए सांख्यिकीय सॉफ्टवेयर की जानकारी आवश्यक है।

भविष्य की संभावनाएँ

1. मशीन लर्निंग इंटीग्रेशनः मशीन लर्निंग एल्गोरिदम के साथ पारंपरिक सांख्यिकीय तरीकों का संयोजन सटीकता बढ़ा सकता है और छिपे हुए पैटर्न को दर्शा सकता है।
2. निजीकृत पोषणः उन्नत विश्लेषण के द्वारा आनुवंशिक, जीवन शैली एवं पर्यावरणीय कारकों के आधार पर अनुकूलित आहार सिफारिशों को विकसित किया जा सकता है।
3. स्थिरता अंतर्दृष्टिः खाद्य अपशिष्ट को कम करने और सतत आहार प्रणाली को बढ़ावा देने के लिए, सांख्यिकीय विश्लेषण के द्वारा रुझानों की पहचान कर कार्यनीति तैयार की जा सकती है।

निष्कर्ष

प्रतिगमन विश्लेषण, खोजपूर्ण और पुष्टिकरण कारक विश्लेषण, प्रमुख घटक विश्लेषण एवं क्लस्टर विश्लेषण जैसी सांख्यिकीय तकनीकें खाद्य विज्ञान और पोषण अनुसंधान में क्रांति ला रही हैं। रिश्तों को उजागर करके, उपकरणों को परिष्कृत करके, प्रणालियों का विश्लेषण करके, और आबादी को विभाजित करके, यह विधियां कार्रवाई योग्य अंतर्दृष्टि प्रदान करती हैं, जो उपभोक्ताओं, शोधकर्ताओं एवं उद्योग के पेशेवरों को समान रूप से लाभान्वित करती हैं। नई तकनीकों के विकास से, बड़े डेटा और मशीन लर्निंग के साथ मल्टीवेरिएट सांख्यिकीय विश्लेषण का एकीकरण नवाचार को दिशा प्रदान करता है, जो हमारी क्षेत्र सम्बंधित जानकारी को बढ़ाता है।

राजस्थान में प्रमुख तिलहन फसलों के क्षेत्रफल के लिए आपूर्ति प्रतिक्रिया फलन

दिनेश कुमार[1], रविन्द्र सिंह शेखावत[1] एवं प्रवीण आर्य[2]

[1]भा.कृ.अनु.प.-केन्द्रीय शुष्क क्षेत्र अनुसंधान संस्थान, क्षेत्रीय अनुसंधान केन्द्र, बीकानेर, राजस्थान

[2]भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

तिलहन फसलें कृषि अर्थव्यवस्था का दूसरा सबसे महत्वपूर्ण निर्धारक हैं, जो खेत की फसलों के क्षेत्र में अनाज के बाद दूसरे स्थान पर हैं। तिलहनों में आत्मनिर्भरता 1990 के दशक की शुरुआत में "पीली क्रांति" के माध्यम से प्राप्त की गई थी और इनमें से प्रमुख हैं सोयाबीन, मूंगफली, सरसों, तिल, सूरजमुखी, अरंडी, कुसुम, अलसी और नाइजर। भारत में नौ तिलहन फसलों में, तिलहनों के कुल उत्पादन में सबसे अधिक औसत योगदान सोयाबीन का है, उसके बाद सरसों और मूंगफली का स्थान है। इसी तरह, कुल तिलहन क्षेत्र में सबसे अधिक औसत क्षेत्र योगदान सोयाबीन (44%) का है, उसके बाद सरसों (24%) और मूंगफली (20%) का स्थान है। अध्ययन का मुख्य उद्देश्य चयनित वर्ष में क्षेत्रफल की प्रतिक्रिया जानना है। इस पृष्ठभूमि में, वर्तमान अध्ययन राजस्थान में महत्वपूर्ण तिलहन फसलों के लिए नेरलोवियन लैग्ड एडजस्टमेंट फंक्शन को नियोजित करने का प्रयास करता है ताकि क्षेत्रफल /आपूर्ति प्रतिक्रियाशीलता की डिग्री का अनुमान लगाया जा सके। अध्ययन के लिए 1997-98 से 2020-21 तक के द्वितीयक आंकड़ों को क्रमशः बीकानेर, अलवर और बारां जिलों में मूंगफली, सरसों और सोयाबीन जैसी तिलहन फसलों की एकड़ प्रतिक्रियाशीलता के परिणाम प्राप्त करने के लिए एकत्र किया गया था। परिणामों से पता चला कि मूंगफली और सोयाबीन के तहत क्षेत्रफल में वृद्धि हुई है।

परिचय

भारत दुनिया का चौथा सबसे बड़ा तिलहन उत्पादक देश है, जो अमेरिका, चीन और ब्राजील के बाद दूसरे स्थान पर है। तिलहन के विश्व उत्पादन में भारत का हिस्सा लगभग 10 प्रतिशत रहा है। वैश्विक स्तर पर खेती के तहत कुल क्षेत्रफल का 20.8% हिस्सा भारत में है, जो वैश्विक उत्पादन का 10% है। देश में विविध कृषि-पारिस्थितिक परिस्थितियाँ 9 वार्षिक तिलहन फसलों को उगाने के लिए अनुकूल हैं, जिनमें 7 खाद्य तिलहन (मूंगफली, सरसों, सोयाबीन, सूरजमुखी, तिल, कुसुम और नाइजर) और दो गैर-खाद्य तिलहन (अरंडी और अलसी) शामिल हैं। (सांख्यिकीय सारांश, राजस्थान, 2021)

भारत का तिलहन उत्पादन 1980 में 94 मिलियन टन से बढ़कर 2021-22 में 376 मिलियन टन हो गया। अनुकूल मौसम स्थितियों और तिलहन उत्पादन/विकास कार्यक्रमों और नीतियों को भारत सरकार द्वारा दिए गए समर्थन के कारण देश ने 2017-18 के दौरान तिलहन उपज का उत्पादकता स्तर 1284 किलोग्राम/हेक्टेयर दर्ज किया, जबकि 2020-21 के दौरान यह 1254 किलोग्राम/हेक्टेयर रहा। (सांख्यिकीय सारांश, राजस्थान, 2021)

देश के कुल फसली क्षेत्र के लगभग 12 प्रतिशत हिस्से पर विभिन्न तिलहन उगाए जाते हैं। तिलहन क्षेत्र का लगभग 72% हिस्सा छोटे किसानों द्वारा की जाने वाली वर्षा आधारित खेती तक सीमित है, जिसके कारण उत्पादकता कम है। हालाँकि, नवीनतम फसल उत्पादन तकनीकों को शुरू करने से तिलहन उत्पादन में एक सफलता मिली है। भारत में तिलहन का उत्पादन पिछले पाँच वर्षों से बढ़ रहा है।

2020-21 में देश का उत्पादन 365.65 लाख टन था जो पिछले वर्ष की तुलना में 10% अधिक था। वर्ष 2015-16 से 2020-21 तक उत्पादन की चक्रवृद्धि वार्षिक वृद्धि दर (बढ़त) 7.7% थी। यह रबी के दौरान सरसों पर विशेष कार्यक्रमों और भारत सरकार द्वारा बेहतर तकनीक के क्लस्टर प्रदर्शनों जैसे विभिन्न कार्यक्रमों के कार्यान्वयन के कारण हासिल किया गया था (आरबीआई सांख्यिकी, 2022)। बीजों में सोयाबीन (34%), मूंगफली (27%), सरसों (27%) कुल तिलहन उत्पादन में 88% से अधिक और वनस्पति तेल में 80% से अधिक का योगदान करते हैं, जिसमें सरसों (35%), सोयाबीन (23%) और मूंगफली (25%) का प्रमुख हिस्सा है।

भारत में सबसे बड़े तिलहन उत्पादक राज्यों में आंध्र प्रदेश (मूंगफली), गुजरात (मूंगफली), हरियाणा (सरसों), कर्नाटक (मूंगफली), मध्य प्रदेश (सोयाबीन), महाराष्ट्र (सोयाबीन), राजस्थान (सरसों और सोयाबीन), तमिलनाडु (मूंगफली), उत्तर प्रदेश (सरसों) और पश्चिम बंगाल (सरसों) शामिल हैं। इन राज्यों में से राजस्थान, गुजरात, मध्य प्रदेश और महाराष्ट्र क्रमशः कुल उत्पादन में लगभग 20%, 20%, 19% और 16% की हिस्सेदारी के साथ शीर्ष उत्पादक हैं [जैन, *एट. एल.*, 2005]

भारत के तिलहन उत्पादन में राजस्थान राज्य का प्रमुख स्थान है। राजस्थान राज्य की महत्वपूर्ण तिलहन फसलें मूंगफली, सोयाबीन, सरसों, तिल और तारामीरा हैं। राज्य में इन फसलों का विकास पैटर्न समय के साथ और कृषि-जलवायु क्षेत्रों में जोखिम भरा रहा है, क्योंकि वर्षा का व्यवहार, लंबे समय तक सूखे की अवधि और राज्य में उपलब्ध सीमित जल संसाधन और सुविधाएँ हैं (2005)। वर्तमान अध्ययन राजस्थान की प्रमुख तिलहन फसलों के लिए नरलोवियन लेग्ड समायोजन फंक्शन का वर्णन करता है।

सामग्री और विधि

वर्तमान अध्ययन द्वितीयक आंकड़ों पर आधारित है। 1997-98 से 2020-21 तक के समय श्रृंखला आंकड़े आर्थिक और सांख्यिकी निदेशालय, राजस्थान से प्राप्त किए गए। उच्चतम उत्पादन के आधार पर तीन प्रमुख तिलहन फसलों यानी मूंगफली, सरसों और सोयाबीन का चयन किया गया था। राजस्थान में क्रमशः बीकानेर, अलवर और बारां जिलों में मूंगफली, सरसों और सोयाबीन जैसी प्रमुख तिलहन फसलों के उच्चतम उत्पादन के आधार पर चयन किया गया ।

चयनित फसलों के क्षेत्रफल / आपूर्ति प्रतिक्रिया विश्लेषण की डिग्री का अनुमान लगाने के लिए नरलोवियन लैग्ड एडजस्टमेंट फंक्शन का उपयोग किया गया था। इस पद्धति का उपयोग (अहमद और भौमिक, 1991; बापना, 1980; देवी, 1964; जाखड़े और मुजुमदार, 1964) ने विभिन्न आपूर्ति प्रतिक्रिया मॉडल में किया था। अध्ययन में, यह माना गया था कि इनपुट लागत या तो समान है या विभिन्न फसलों के लिए समय के साथ समान रूप से चलती है। मॉडल को 1997-98 से 2020-21 की अवधि के आंकड़ों पर फिट किया गया। मापदंडों का अनुमान लगाने के लिए, डबल लॉगरिदमिक मॉडल का उपयोग किया गया है। इसे प्राथमिकता दी गई है क्योंकि यह अल्पकालिक लोच का प्रत्यक्ष अनुमान प्रदान करता है और गुणांकों का बेहतर अनुमान देता है। किसान के व्यवहार को उसके सरलतम रूप में दर्शाने वाला नरलोवियन मॉडल नीचे दिया गया है:

$$A_t^* = b_0 + b_1 P_{t-1} + b_2 Z_{t-1} + b_3 R_t + b_4 GI_t + b_5 CVY_t + b_6 CVP_t + U_t \qquad (1)$$

$$A_t - A_{t-1} = B(A_t^* - A_{t-1})$$

(नरलोवियन समायोजन समीकरण) (2)

चूँकि अपेक्षित चर अवलोकनीय नहीं हैं, इसलिए अनुमान के प्रयोजनों के लिए, समीकरण (2) से At* का मान समीकरण (1) में प्रतिस्थापित करने के बाद केवल अवलोकनीय चरों वाले एक संक्षिप्त रूप को लिखा जाएगा, इस प्रकार:

$$A_t = C_0 + C_1 P_{t-1} + C_2 Z_{t-1} + C_3 R_t + C_4 GI_t + C_5 CVY_t + C_6 CVP_t + C_7 A_{t-1} + V_t \qquad (3)$$

पहला समीकरण एक व्यवहारिक समीकरण है, जो बताता है कि वांछित क्षेत्रफल (A_t^*) निम्नलिखित स्वतंत्र

चरों पर निर्भर करता है।

A_{t-1} = एक वर्ष का लेग्ड क्षेत्रफल

P_{t-1} = एक वर्ष का लेग्ड मूल्य

Z_{t-1} = एक वर्ष का लेग्ड उपज

R_t = बुवाई से पहले तीन महीनों के लिए औसत वर्षा

GI_t = चालू वर्ष का सकल सिंचित क्षेत्रफल

CVP_t = पिछले तीन वर्षों के मूल्य का भिन्नता गुणांक

CVY_t = पिछले तीन वर्षों की उपज का भिन्नता गुणांक

समीकरण (1) में सभी फसलों के लिए चर शामिल हैं। समीकरण (3) पिछले दो समीकरणों का कम किया हुआ रूप है, जो एक प्रेक्षित चर (A_t) द्वारा अनिर्धारित चर (A_t^*) का अनुमान लगाता है। समीकरण (3) के गुणांक और त्रुटि पद समीकरण (I) और समायोजन गुणांक से इस प्रकार संबंधित हैं:

$C_0 = b_0B$, $C_1 = b_1B$, $C_2 = b_2B$,

$C_3 = b_3B$, $C_4 = b_4B$, $C_5 = b_5B$,

$C_6 = b_6B$, $C_7 = (1–B)$, $V_t = BV_t$

उपरोक्त अभिगणक समीकरण के इन मापदंडों का अनुमान कम से कम वर्ग विधि द्वारा आपूर्ति/क्षेत्रफल प्रतिक्रिया मॉडल में शामिल संबंधित चर पर समय-श्रृंखला डेटा का उपयोग करके लगाया जा सकता है। विभिन्न चर के प्रतिगमन गुणांक का परिमाण स्पष्ट रूप से दर्शाता है कि आश्रित चर (अवधि में क्षेत्रफल) स्वतंत्र चर पर कितना निर्भर करता है। निर्धारण गुणांक (R_2) का मान आश्रित चर में भिन्नता का प्रतिशत दर्शाता है जिसे स्वतंत्र चर द्वारा समझाया जाता है। समायोजन गुणांक द्वारा लघु अवधि लोच को विभाजित करके दीर्घ अवधि लोच की गणना की गई। समायोजन गुणांक 1-C_7 यानी 1-(1-B) के बराबर है।

प्रयुक्त चरों की विशिष्टताएँ

i. चालू वर्ष का क्षेत्र (A_t)

क्षेत्रफल प्रतिक्रिया मॉडल अध्ययन ने कुल नियोजित उत्पादन में होने वाले परिवर्तनों को विभिन्न आर्थिक और पर्यावरणीय कारकों में होने वाले परिवर्तनों से जोड़ा है, लेकिन सिंह एट अल.(1974) ने बताया कि किसानों के निर्णय फसलों के अंतर्गत आने वाले क्षेत्र के संदर्भ में अनुमानित होते हैं, न कि उसकी उपज के संदर्भ में। क्योंकि फसल को आवंटित क्षेत्र किसान के भूमि आवंटन निर्णयों का बेहतर बैरोमीटर है। इसके अलावा, फसल के अंतर्गत आने वाला क्षेत्र कई अंतर्जात चर/कारकों का एक कार्य है, जबकि उपज कई बहिर्जात कारकों से बहुत प्रभावित होती है। इसके अलावा, चूंकि नियोजित उत्पादन के समय-श्रृंखला अनुमान उपलब्ध नहीं हो सकते हैं, इसलिए कुछ प्रतिनिधि का उपयोग किया जाना चाहिए। इस प्रकार संबंधित फसल के अंतर्गत लगाए गए क्षेत्र को प्रतिगमन मॉडल में आश्रित चर के रूप में लिया गया है। नरलोवियन (1958) ने खुद बताया कि आपूर्ति का अधिक उपयुक्त उपाय एकड़ लगता है क्योंकि कृषक के पास क्षेत्र पर अधिक नियंत्रण होता है।

ii. लेग्ड क्षेत्र (A_{t-1})

एकड़ प्रतिक्रिया फंक्शन में एक स्वतंत्र चर के रूप में लैग्ड एकड़ को शामिल करना समायोजन गुणांक तक पहुँचने के लिए एक वाहन के रूप में कार्य करता है, जिसे स्थिर माना जाता है और हमेशा शून्य और एक के बीच होता है। यह चर सभी गैर-मूल्य कारकों जैसे अर्ध-स्थिर कारकों, जोखिम और अनिश्चितताओं और ऐसे तकनीकी परिवर्तनों के प्रभाव को भी ध्यान में रखता है, जिन्हें मापना मुश्किल है लेकिन उत्पादन के परिवर्तन पर उनका एक विशिष्ट प्रभाव पड़ता है। यह भी माना जाता है कि सामान्य परिस्थितियों में, कृषक अपनी पारिवारिक आवश्यकताओं और अन्य जरूरतों को ध्यान में रखते हुए कम से कम अपने पिछड़े वर्ष का एकड़ रखता है।

iii. लेग्ड मूल्य (P_{t-1})

किसानों के एकड़ प्रतिक्रिया मॉडल में निर्णय लेने में मूल्य निर्माण का उपयोग बहुत महत्व रखता है। वैकल्पिक मूल्य विनिर्देशों को संसाधन आवंटन निर्णयों के संबंध में उत्पादक के अपेक्षित व्यवहार में सबसे अधिक प्रासंगिक माना जाता है। किसान पिछले वर्ष की कीमत देखते हैं, चाहे वह फसल के बाद की पूर्ण कीमत हो, सापेक्ष कीमत हो या सापेक्ष लाभप्रदता हो, ताकि वे चालू वर्ष के उत्पादन

के लिए निर्णय ले सकें। इस अध्ययन में कृषि फसल की कीमतों का उपयोग इस धारणा के साथ किया गया है कि उनके उत्पादों का बड़ा हिस्सा फसल कटाई के दो महीने के भीतर बाजार में बिक जाएगा।

iv. लेग्ड उपज (Y_{t-1})

समय के साथ फसलों की उपज में काफी भिन्नता रही है। गेहूं, धान, बाजरा और चना की फसलों में उच्च उपज वाली किस्मों के आने, कीटों और बीमारियों के हमलों और प्रकृति की अनियमितताओं के कारण उपज में भिन्नता देखी जाती है। इसलिए, लेग्ड वर्ष की उपज को एक स्वतंत्र चर के रूप में मॉडल में शामिल किया गया।

v. वर्षा (R_t)

सीजन के दौरान वर्षा को एक स्वतंत्र चर के रूप में लिया गया। फसल की उपज और उत्पादन वर्षा के वितरण से प्रभावित होता है। यह एक स्वतंत्र चर है और इसका कुछ फसलों जैसे बाजरा और चना की कीमतों पर गहरा प्रभाव पड़ता है। राजस्थान में इन फसलों की खेती की जाने वाली मिट्टी और नमी की स्थिति को देखते हुए इन फसलों और वर्षा के बीच संबंधों का पैटर्न उचित प्रतीत होता है।

vi. सकल सिंचाई (GI_t)

एकड़ आवंटन पर सिंचाई सुविधाओं के प्रभाव की जांच करने के लिए, यानी वह क्षेत्र जो किसान प्रत्येक फसल के तहत लगाने के लिए तैयार है, प्रत्येक फसल के बढ़ते मौसम के दौरान सकल सिंचित क्षेत्र को एक स्वतंत्र चर के रूप में शामिल किया गया था, क्योंकि ऐसी सुविधाओं से उपज और भूमि उपयोग की लचीलापन बढ़ता है। आम तौर पर, चालू वर्ष के सकल सिंचित क्षेत्र और चालू वर्ष के एकड़ के बीच एक सकारात्मक सहसंबंध देखा जाता है। इसलिए, सिंचित फसलों के लिए, चालू वर्ष के सकल सिंचित क्षेत्र को एक स्वतंत्र चर के रूप में शामिल किया गया है।

vii. जोखिम

जोखिम कारक किसानों के निर्णयों में एक महत्वपूर्ण तत्व हैं, विशेष रूप से बदलती प्रौद्योगिकी के साथ कृषि क्षेत्र में। आम तौर पर, कृषि में दो प्रकार के जोखिमों को ध्यान में रखा जाता है, अर्थात उपज जोखिम और मूल्य जोखिम। उपज जोखिम (Y) प्रकृति की अनिश्चितताओं, कीटों और कीटों के हमलों के कारण कृषि उत्पादन में प्रवेश करते हैं, जबकि मूल्य जोखिम (p) मांग और आपूर्ति की अनिश्चितताओं के कारण होते हैं। उपज परिवर्तनशीलता को उपज के परिवर्तन के गुणांक से तीन साल पहले और मूल्य परिवर्तनशीलता को मूल्य के परिवर्तन के गुणांक से तीन साल पहले के हिसाब से माना जाता है।

शोध परिणाम

बीकानेर जिले में मूंगफली की खेती के लिए एक महत्वपूर्ण क्षेत्र है और यह राजस्थान का सबसे अधिक मूंगफली उत्पादक जिला है। मूंगफली के लिए क्षेत्रफल प्रतिक्रिया मॉडल के प्रतिगमन परिणाम तालिका 1 में दिए गए हैं। मूंगफली की खेती के तहत क्षेत्र का निर्धारण करने में लैग्ड क्षेत्र, लैग्ड उपज, लैग्ड कृषि फसल की कीमतें और मूंगफली के तहत सिंचित क्षेत्र के लिए प्रतिगमन गुणांक सभी महत्वहीन थे।

अलवर में सरसों के अंतर्गत काफी क्षेत्र है और राजस्थान में सबसे ज्यादा सरसों का उत्पादन होता है। मॉडल से प्राप्त प्रतिगमन परिणाम तालिका 2 में प्रस्तुत किए गए हैं। मूंगफली की खेती के अंतर्गत क्षेत्र के निर्धारण में लैग्ड

तालिका 1. राजस्थान के बीकानेर जिले में मूंगफली के लिए क्षेत्रफल प्रतिक्रिया (1997-98 से 2020-21)

प्रतिगमन गुणांक	अनुमान	मानक त्रुटि
1. स्थिर	−0.6685	1.2109
2. मूंगफली का लैग्ड क्षेत्र	0.8620***	0.1522
3. मूंगफली की लैग्ड उपज	0.0285	0.1006
4. मूंगफली की लैग्ड फसल कीमत	0.1401	0.2032
5. जिले का सकल सिंचित क्षेत्र	0.2063	0.3893
6. वर्षा	−0.0989	0.0907
7. उपज जोखिम	0.02444	0.0532
8. मूल्य जोखिम	−0.0387	0.0554
समायोजित R^2(B)	0.96	

***महत्व को इंगित करें संभावना का स्तर 1 प्रतिशत

तालिका 2. राजस्थान के अलवर जिले में सरसों के लिए क्षेत्रफल प्रतिक्रिया (1997–98 से 2020–21)

प्रतिगमन गुणांक	अनुमान	मानक त्रुटि
1. स्थिर	−1.3555	2.8507
2. मूंगफली का लैग्ड क्षेत्र	0.269	0.2639
3. मूंगफली की लैग्ड उपज	0.0808	0.1840
4. मूंगफली की लैग्ड फसल कीमत	0.0476	0.0855
5. जिले का सकल सिंचित क्षेत्र	0.8224	0.5272
6. वर्षा	0.1541	0.1250
7. उपज जोखिम	0.0223	0.0674
8. मूल्य जोखिम	−0.0008	0.0533
समायोजित R^2(B)	0.31	

तालिका 3. राजस्थान के बारां जिले में सोयाबीन के लिए क्षेत्रफल प्रतिक्रिया (1997–98 से 2020–21)

प्रतिगमन गुणांक	अनुमान	मानक त्रुटि
1. स्थिर	0.7002	1.7794
2. मूंगफली का लैग्ड क्षेत्र	0.3603	0.2590
3. मूंगफली की लैग्ड उपज	0.5681*	0.2659
4. मूंगफली की लैग्ड फसल कीमत	0.3353*	0.1515
5. जिले का सकल सिंचित क्षेत्र	0.9629**	0.3632
6. वर्षा	−0.1008	0.2334
7. उपज जोखिम	−0.2509**	0.1068
8. मूल्य जोखिम	−0.0023	0.0788
समायोजित R^2(B)	0.66	

** &* 5 प्रतिशत और 10 प्रतिशत संभाव्यता स्तर पर महत्त्व दर्शाते हैं

क्षेत्र, लैग्ड उपज, लैग्ड कृषि फसल मूल्य और मूंगफली के अंतर्गत सिंचित क्षेत्र के लिए प्रतिगमन गुणांक उनके प्रभाव में महत्वहीन पाए गए।

राजस्थान के बारां जिले में सोयाबीन की खेती के अंतर्गत आने वाले क्षेत्र में महत्वपूर्ण हिस्सेदारी है और यह राजस्थान के जिलों में सबसे अधिक उत्पादन करता है। चरों के लिए गुणांक जैसे कि विलंबित उपज, विलंबित कृषि फसल मूल्य और सकल सिंचित क्षेत्र 5 प्रतिशत पर सकारात्मक और महत्वपूर्ण थे, जैसा कि तालिका 3 में दिया गया है। इसी प्रकार का परिणाम खाद्यान्न फसलों में कुमार एट अल, (2023) द्वारा प्रस्तुत किया गया था।

समायोजित R^2 0.66 का मान जो दर्शाता है कि आश्रित चर में होने वाले परिवर्तन को स्वतंत्र चर द्वारा बहुत अच्छी तरह से समझाया गया है ।

संक्षिप्त विवरण

क्षेत्रफल प्रतिक्रिया विश्लेषण में, यह निष्कर्ष निकाला गया है कि मूंगफली, सरसों, और सोयाबीन का एकड़ मुख्य रूप से वर्ष के दौरान लैग्ड उपज और सकल सिंचित क्षेत्र द्वारा नियंत्रित किया गया था, जबकि लैग्ड हुई कृषि फसल की कीमतें भी क्षेत्रफल आवंटन विश्लेषण का प्रमुख निर्धारक प्रतीत हुई। सरसों, मूंगफली और सोयाबीन जैसे तिलहनों के मामले में, यदि कीमत में एक प्रतिशत की वृद्धि होती है, तो किसानों द्वारा मूंगफली में 0.2 प्रतिशत और सोयाबीन के तहत 0.33 प्रतिशत एकड़ का प्रतिशत बढ़ाया जाएगा। यह निष्कर्ष निकाला गया है कि नेरलोवियन रकबा प्रतिक्रिया कार्य विभिन्न योगदान कारकों द्वारा बहुत अच्छी तरह से समझाया गया है और आपूर्ति प्रतिक्रिया मूल्य और लैग्ड हुई कीमत के अनुसार है।

संदर्भ

आर.बी.आई. सांख्यिकी, 2022. उपलब्धः https://www.findeasy.in/indianstates–by–oilseeds–production/ 15.10.2022 को मूल्यांकन किया गया।

जैन, पी.के., सिंह आई.पी., कुमार, अनिल. 2005. राजस्थान राज्य में तिलहन उत्पादन वृद्धि में जोखिम:एक नीतिगत परिप्रेक्ष्य,*कृषि अर्थशास्त्र अनुसंधान समीक्षा*, 18:115–133।

अहमद, ए.यू., भौमिक, बी.सी. 1991. असम में कुछ महत्वपूर्ण फसलों की कीमत, उत्पादकता और रकबा प्रतिक्रिया का व्यवहार, *भारत में कृषि की स्थिति*, 45(4): 225–229।

बापना, एस.एल.1980. विकासशील क्षेत्र में फसलों की समग्र रकबा प्रतिक्रिया, नई दिल्ली, *सुल्तान चंद एंड संस.*।

देवी, पी., कमला. 1964. कीमत में बदलाव के प्रति रकबा

की प्रतिक्रिया - मद्रास राज्य में एक केस स्टडी.द *इकोनॉमिक एंड पॉलिटिकल वीकली*, 16(38): 1535-1536।

जाखड़े, वी.एम., मुजुमदार, एन.ए. 1964. कृषि उत्पादकों की कीमतों पर प्रतिक्रिया- भारत में जूट और चावल का मामला. *इंडियन जर्नल ऑफ एग्रीकल्चरल इकोनॉमिक्स*, 21(1): 204-209।

सिंह, आर.डी., सिंह, डी., राव, पी.आर. 1974. कृषि रकबा प्रतिक्रिया संबंध का आकलन: कुछ पद्धतिगत मुद्दे *इंडियन जर्नल ऑफ एग्रीकल्चरल इकोनॉमिक्स*, 29(1):26-38।

नरलोवियन, एम.1958. आपूर्ति की गतिशीलता: मूल्य के प्रति किसानों की प्रतिक्रिया का अनुमान: प्रथम संस्करण. *जॉन हॉपकिंस यूनिवर्सिटी प्रेस*, बाल्टी-मोर।

कुमार, डी., शर्मा, एम., शेखावत, आर. एस., राठौर, वी.एस., मंजूनाथ, बी.एल., नाथावत, एन.एस. 2023. राजस्थान में प्रमुख खाद्यान्न फसलों की आपूर्ति प्रतिक्रिया कार्य. *एशियाई कृषि विस्तार, अर्थशास्त्र और समाजशास्त्र जर्नल*,;41 (3):1-6।

राजस्थान सरकार, 2021. *सांख्यिकीय सारांश*।

रोगाणुरोधी प्रतिरोधः जैव सूचना विज्ञान-संचालित पहचान उपकरण, रुझान और संस्थागत पहल

रागिनी कुशवाहा, अनु शर्मा, द्विजेश चंद्र मिश्रा, मोहम्मद समीर फारूकी, सुधीर श्रीवास्तव, शशि भूषण लाल

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

सार

रोगाणुरोधी प्रतिरोध [Antimicrobial Resistance (AMR)] मानव, पशु और पर्यावरणीय स्वास्थ्य को प्रभावित करने वाला एक बढ़ता हुआ वैश्विक खतरा है। यह अध्ययन AMR के प्रमुख कारणों, परिणामों और पूर्वानुमान रणनीतियों पर प्रकाश डालता है। पशुधन और फसलों में रोगाणुरोधी दवाओं का अत्यधिक उपयोग और दुरुपयोग प्रतिरोधी सूक्ष्मजीवों के उद्‌भव में महत्वपूर्ण योगदान देता है, जिससे उपचार प्रभावकारिता कम हो जाती है और खाद्य सुरक्षा को खतरा होता है। WHO और ICAR सहित राष्ट्रीय और अंतर्राष्ट्रीय पहल, AMR निगरानी और प्रबंधन में महत्वपूर्ण भूमिका निभाती हैं। प्रतिरोध पैटर्न की भविष्यवाणी करने के लिए इस अध्ययन में उन्नत जैव सूचना विज्ञान उपकरण और डेटाबेस सहित विभिन्न तरीकों पर चर्चा की गई है। इस अध्ययन में AMR जीन की भविष्यवाणी को प्रदर्शित करने के लिए CARD–RGI का उपयोग किया गया, जिससे जीनोमिक अनुक्रमों से प्रतिरोध निर्धारकों की पहचान करने में इसकी उपयोगिता प्रदर्शित हुई। लक्षित हस्तक्षेप विकसित करने और टिकाऊ पशु स्वास्थ्य और खाद्य प्रणालियों को सुनिश्चित करने के लिए ये एकीकृत दृष्टिकोण महत्वपूर्ण हैं।

परिचय

अपने पूरे जीवन में पौधे, जानवर और मनुष्य लगातार हानिकारक सूक्ष्मजीवों जैसे बैक्टीरिया, वायरस, कवक और परजीवी का सामना करते हैं। ये रोगजनक गंभीर बीमारियों का कारण बन सकते हैं और कुछ मामलों में, दुनिया भर में प्रकोप का कारण बन सकते हैं। इन संक्रमणों को संबोधित करने के लिए, रोग पैदा करने वाले सूक्ष्मजीवों को नियंत्रित करने और खत्म करने के साधन के रूप में रोगाणुरोधी एजेंटों को पेश किया गया था। रोगाणुरोधी युग ने संक्रामक रोगों के उपचार में क्रांति ला दी। रोगाणुरोधी एजेंट ऐसी दवाएँ हैं जो मनुष्यों, पौधों और जानवरों में संक्रमण से लड़ती हैं या तो सूक्ष्मजीव को मार देती हैं या उन्हें बढ़ने और गुणा करने में मुश्किल बनाती हैं। इनमें एंटीबायोटिक्स (बैक्टीरिया को लक्षित करना), एंटीवायरल (वायरस के खिलाफ), एंटीफंगल (फंगल संक्रमण के लिए) और एंटीपैरासिटिक दवाएँ शामिल हैं। हालाँकि, इन उपचारों की प्रभावशीलता को खतरे में डालने वाली एक बड़ी चुनौती रोगाणुरोधी प्रतिरोध [Antimicrobial Resistance (AMR)] है। AMR तब होता है जब सूक्ष्मजीव उन दवाओं के प्रभावों का सामना करने के लिए तंत्र विकसित करते हैं जो कभी उन्हें खत्म करने में प्रभावी थीं। AMR रोगाणुरोधी युग की उपलब्धियों के लिए एक गंभीर खतरा है क्योंकि यह मानक उपचारों को अप्रभावी बनाता है और संक्रमण के प्रसार में योगदान देता है, इस प्रकार यह वैश्विक सार्वजनिक स्वास्थ्य के लिए एक गंभीर खतरा है (1)।

AMR के लिए जिम्मेदार कारक

सूक्ष्मजीवों में कई कारकों के संयोजन के कारण AMR विकसित होता है, जो मुख्य रूप से रोगाणुरोधी एजेंटों

के अत्यधिक उपयोग और दुरुपयोग से प्रेरित होता है। एंटीबायोटिक्स, एंटीवायरल और एंटीफंगल के अत्यधिक और अनुचित नुस्खे। एक अन्य प्रमुख कारण रोगाणुरोधी उपचारों का अनुचित या अधूरा उपयोग है। कृषि और पशुधन में रोगाणुरोधी पदार्थों का व्यापक उपयोग भी AMR में महत्वपूर्ण भूमिका निभाता है। बड़े पैमाने पर पशु प्रोटीन उत्पादन के लिए विकास को बढ़ावा देने और बीमारियों को रोकने के लिए अक्सर पशु आहार में रोगाणुरोधी पदार्थों का उपयोग किया जाता है, जिससे भोजन के सेवन के माध्यम से मनुष्यों में प्रतिरोधी बैक्टीरिया का स्थानांतरण होता है (2)। इसके अतिरिक्त, प्रतिरोधी सूक्ष्मजीवों वाले खेतों से निकलने वाला कचरा मिट्टी और जल स्रोतों को दूषित करता है, जिससे पर्यावरण में प्रतिरोध फैलता है। AMR में शामिल आनुवंशिक तत्वों को एंटीमाइक्रोबियल रेजिस्टेंस जीन (ARG) के रूप में जाना जाता है। ये जीन प्रोटीन को एनकोड करते हैं जो प्रतिरोध प्रदान कर सकते हैं, जिससे सूक्ष्मजीव एंटीमाइक्रोबियल एजेंटों के संपर्क में जीवित रह सकते हैं। ARG प्राकृतिक रूप से हो सकते हैं या आनुवंशिक उत्परिवर्तन और जीन दोहराव जैसे आनुवंशिक तंत्र के कारण उत्पन्न हो सकते हैं। सूक्ष्मजीव क्षैतिज जीन स्थानांतरण [Horizontal Gene Transfer (HGT)] के माध्यम से भी प्रतिरोध प्राप्त करते हैं, जहाँ वे संयुग्मन, परिवर्तन या पारगमन के माध्यम से प्रतिरोध जीन का आदान-प्रदान करते हैं, जो सूक्ष्मजीव आबादी के बीच प्रतिरोध के प्रसार को सक्षम बनाता है (3)। सहज आनुवंशिक उत्परिवर्तन दवा के लक्ष्यों को भी बदल सकते हैं, जिससे एंटीमाइक्रोबियल एजेंटों की प्रभावशीलता कम हो जाती है।

कृषि में AMR के परिणाम

कृषि, खाद्य सुरक्षा और सार्वजनिक स्वास्थ्य सभी पौधों और जानवरों में AMR से गंभीर रूप से खतरे में हैं, जिसके परिणामस्वरूप उपचार की प्रभावशीलता कम हो सकती है, बीमारी का बोझ बढ़ सकता है और वित्तीय नुकसान हो सकता है। क्योंकि साल्मोनेला, स्टैफिलोकोकस ऑरियस और एस्चेरिचिया कोली सहित प्रतिरोधी जीव जानवरों में लंबे समय तक संक्रमण, उपचार विफलताओं और मृत्यु दर में वृद्धि का कारण बनते हैं, AMR अंततः पशु चिकित्सा व्यय को बढ़ाता है और पशुधन उत्पादन को कम करता है। चूंकि प्रतिरोधी रोगाणु सीधे संपर्क, खाद्य उपभोग या पर्यावरण प्रदूषण के माध्यम से जानवरों से लोगों में स्थानांतरित होते हैं, इसलिए जूनोटिक संचरण सार्वजनिक स्वास्थ्य के लिए खतरों को बढ़ाता है। इसी तरह, पौधों में एएमआर के परिणामस्वरूप फसल की गुणवत्ता और उत्पादन में कमी आती है क्योंकि फ्यूजेरियम, स्यूडोमोनास और जैंथोमोनास जैसे जीवाणु और फंगल रोगों को प्रभावी ढंग से नियंत्रित नहीं किया जाता है। कृषि मिट्टी में, टेट्रासाइक्लिन और स्ट्रेप्टोमाइसिन जैसे जीवाणुनाशकों का दुरुपयोग प्रतिरोधी बैक्टीरिया के उद्भव को प्रोत्साहित करता है, जिससे मानव और पशु संक्रमणों में HGT की संभावना बढ़ जाती है (4)। रोगाणुरोधी प्रतिरोध को संबोधित करने के लिए बहुआयामी दृष्टिकोण की आवश्यकता होती है, जिसमें संक्रमण की रोकथाम में सुधार, एंटीबायोटिक के उपयोग पर सख्त नियम, बेहतर स्वच्छता अभ्यास, जन जागरूकता अभियान, नई दवाओं के अनुसंधान और विकास में निवेश में वृद्धि और स्वास्थ्य पेशेवरों, पशु चिकित्सकों और नीति निर्माताओं के बीच सहयोग को बढ़ावा देना शामिल है (5)।

AMR की समस्या से निपटने की योजना

AMR के बढ़ते खतरे से निपटने के लिए राष्ट्रीय और अंतर्राष्ट्रीय संगठनों द्वारा कई परियोजनाएँ शुरू की गई हैं। विश्व स्वास्थ्य संगठन [World Health Organization (WHO)] ने AMR पर अपनी वैश्विक कार्य योजना के माध्यम से जागरूकता बढ़ाने, शोध और निगरानी को बढ़ावा देने, संक्रमण दर को कम करने और नए उपकरणों और दवाओं में निवेश करने का बीड़ा उठाया है। AMR को समग्र रूप से संबोधित करने के लिए, WHO एक स्वास्थ्य दृष्टिकोण का समर्थन करता है जो खाद्य और कृषि संगठन [Food and Agriculture Organization (FAO)] और विश्व पशु स्वास्थ्य संगठन [World Organization of Animal Health (WOAH)] के साथ साझेदारी में मानव,

पशु और पर्यावरणीय स्वास्थ्य क्षेत्रों को जोड़ता है। पशुओं और मत्स्य पालन में ।डत की निगरानी और नियंत्रण के लिए, भारतीय कृषि अनुसंधान परिषद [Indian Council of Agricultural Research (ICAR)] ने INFAAR नेटवर्क की स्थापना के लिए FAO के साथ सहयोग किया। इसमें शामिल महत्वपूर्ण ICAR संस्थान जलीय कृषि के लिए ICAR-CIFE, निदान और प्रशिक्षण के लिए ICAR-IVRI और निगरानी के लिए प्ब्।त्-छप्टम्क् हैं। फिलहाल, 19 ICAR संस्थान इस क्षेत्र में लगे हुए हैं, जो सुरक्षित रोगाणुरोधी उपयोग, AMR अनुसंधान और निगरानी को बढ़ावा दे रहे हैं।

AMR की भविष्यवाणी के लिए मौजूदा तरीके

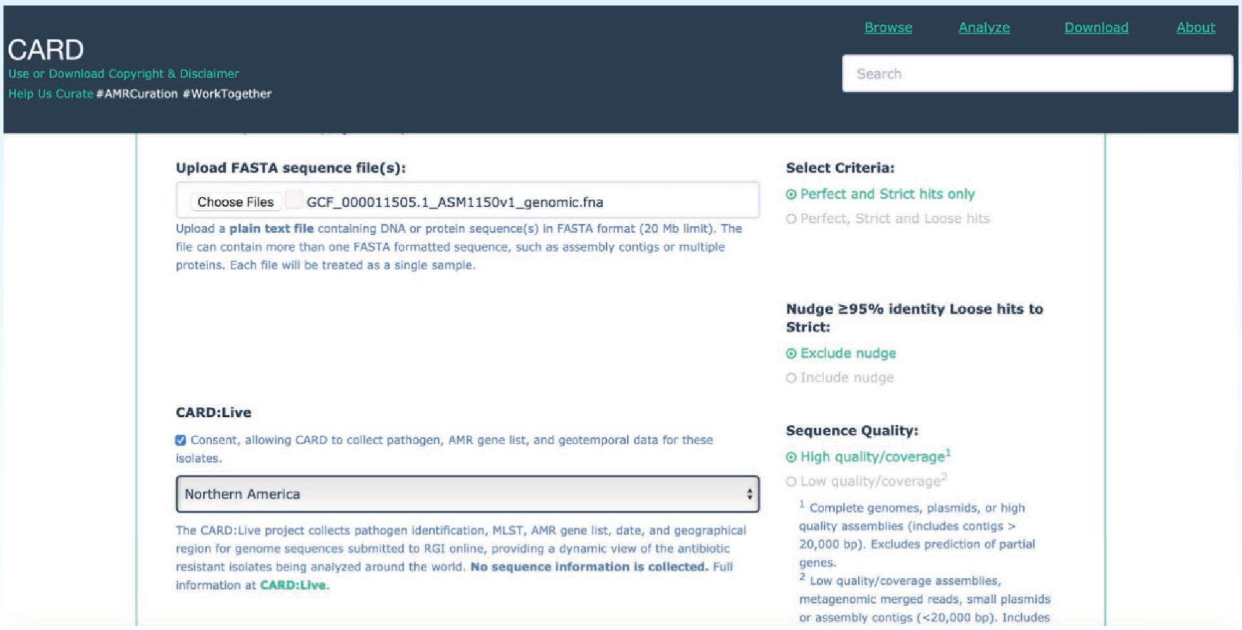
AMR की भविष्यवाणी कई तरह की पद्धतियों जैसे कि फेनोटाइपिक परीक्षण (Phenotypic Testing), आणविक प्रक्रिया, बायोइनफॉरमैटिक्स दृष्टिकोण और मशीन लर्निंग का उपयोग करके की जाती है। बैक्टीरिया के विकास के अवरोध का आकलन करके, ब्रोथ माइक्रोडिल्यूशन (Broth Microdilution), डिस्क डिफ्यूजन (Disk Diffusion), एंटीमाइक्रोबियल ससेप्टिबिलिटी टेस्टिंग [Antimicrobial Susceptibility Testing (AST)] और ई-टेस्ट सहित फेनोटाइपिक परीक्षण प्रतिरोध स्थापित करते हैं। प्रतिरोध जीन को आणविक विधियों जैसे कि MALDI-TOF MS, DNA माइक्रोएरे और पॉलीमरेज चेन रिएक्शन [Polymerase Chain Reaction (PCR)] का उपयोग करके आनुवंशिक स्तर पर पाया जाता है। संपूर्ण जीनोम अनुक्रमण [Whole Genome Sequencing (WGS)] ने बैक्टीरिया के रोगजनकों के संपूर्ण जीनोम की जांच करके, उत्परिवर्तनों का पता लगाकर और प्रतिरोध पैटर्न की निगरानी करके AMR की भविष्यवाणी में पूरी तरह से क्रांति ला दी है। बायोइनफॉरमैटिक्स डेटाबेस-संचालित विधियों, कम्प्यूटेशनल मॉडलिंग और WGS डेटा को जोड़ती है और प्रतिरोध निर्धारकों को खोजने में महत्वपूर्ण भूमिका निभाती है (6)। ARG-ANNOT, CARD-RGI, SRST2, MEGARes, ARIBA और ResFinder कुछ ऐसे बायोइन्फॉर्मेटिक्स उपकरण हैं जिनका उपयोग अक्सर AMR पूर्वानुमान के लिए किया जाता है (7)। ये प्रोग्राम अनुक्रमण डेटा से रोगाणुरोधी प्रतिरोध जीन (ARG) की पहचान करते हैं। जबकि CARD-RGI (व्यापक एंटीबायोटिक प्रतिरोध डेटाबेस-प्रतिरोध जीन पहचानकर्ता) ज्ञात उत्परिवर्तन और जीन अनुक्रमों के आधार पर प्रतिरोध का पूर्वानुमान लगाता है, ARG-ANNOT ज्ञात AMR जीन के लिए एक एनोटेशन सिस्टम प्रदान करता है। उपकरणों के अलावा, विभिन्न AMR डेटाबेस प्रतिरोध जीन पर आनुवंशिक जानकारी संग्रहीत करते हैं। ARDB (एंटीबायोटिक प्रतिरोध जीन डेटाबेस), ResFams, PATRIC(8), ARGMiner और CARD कुछ सबसे व्यापक रूप से उपयोग किए जाने वाले डेटाबेस हैं। ARDB प्रतिरोध तंत्र द्वारा वर्गीकृत AMR जीन का संग्रह प्रदान करता है, जबकि PATRIC प्रतिरोध प्रोफाइल वाले बैक्टीरिया उपभेदों की जीनोमिक तुलना के लिए एक प्लेटफॉर्म प्रदान करता है। ARGMiner मेटाजेनोमिक डेटा से प्रतिरोध जीन को खनन करने में सहायता करता है, और ResFams AMR से जुड़े क्यूरेटेड प्रोटीन परिवार प्रदान करता है। ये जैव सूचना विज्ञान संसाधन शोधकर्ताओं को AMR विकास को ट्रैक करने, उभरते प्रतिरोध तंत्रों की पहचान करने और लक्षित रोगाणुरोधी रणनीतियों को विकसित करने में मदद करते हैं। वर्तमान में रैंडम फॉरेस्ट, सपोर्ट वेक्टर मशीन (SVM) और डीप लर्निंग दृष्टिकोण जैसे मशीन लर्निंग मॉडल जीनोमिक डेटा का विश्लेषण करने और प्रतिरोध पैटर्न की भविष्यवाणी करने के लिए तेजी से उपयोग किए जा रहे हैं (5)।

AMR पूर्वानुमान

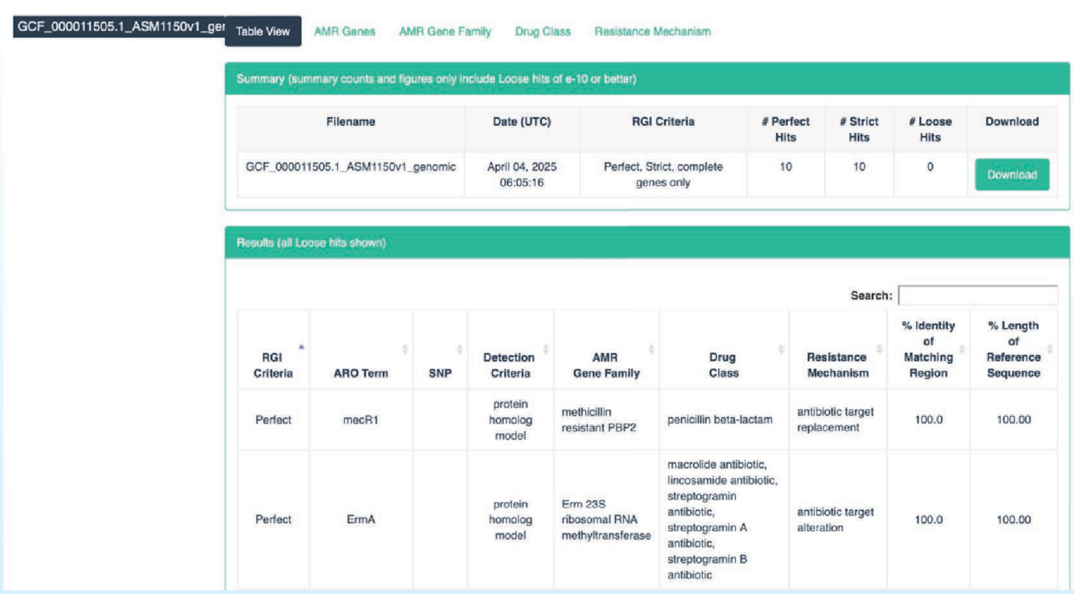
WGS और AI-आधारित मॉडल के साथ जैव सूचना विज्ञान का एकीकरण एएमआर पूर्वानुमान को बदल रहा है, जिससे पशु चिकित्सा और कृषि विज्ञान में तेजी से निदान, निगरानी और प्रभावी नियंत्रण रणनीतियां संभव हो रही हैं। जैसा कि पहले उल्लेख किया गया है, अब जैव सूचना विज्ञान उपकरणों की सहायता से WGS डेटा से सीधे एएमआर का पूर्वानुमान लगाना बहुत आसान है। इस

प्रयोग में, एएमआर पूर्वानुमान के लिए मार्गदर्शन प्रदान करने के लिए, अनुक्रम आईडी GCF_000011505.1_ ASM1150v1_genomic वाले न्यूक्लियोटाइड अनुक्रम डेटासेट को GenBank डेटाबेस से FASTA प्रारूप फाइल में डाउनलोड किया गया था। इसके बाद, CARD–RGI की सहायता से प्रतिरोधी जीन का पूर्वानुमान लगाया गया; प्रतिरोधी जीन की सूची चित्र 3ए में प्रदर्शित की गई है।

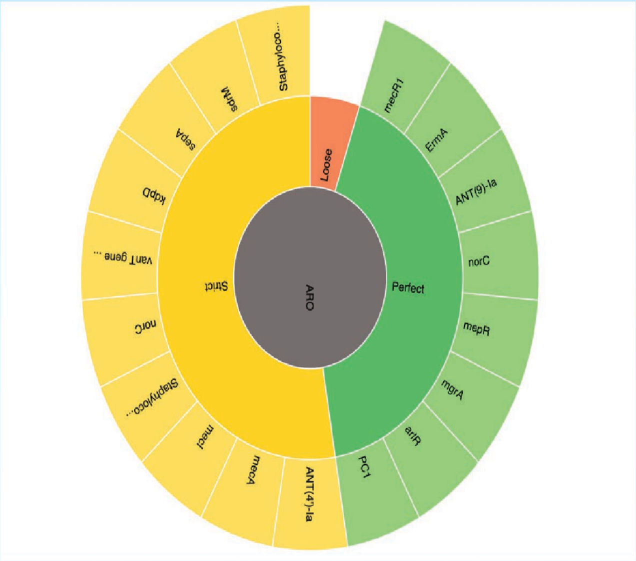
CARD–RGI के लिए प्रदर्शन मीट्रिक की गणना करने के लिए, प्रयोगात्मक प्रयोगशाला–पुष्टिकृत AMR डेटा (MIC या AST परीक्षणों से फेनोटाइपिक प्रतिरोध)

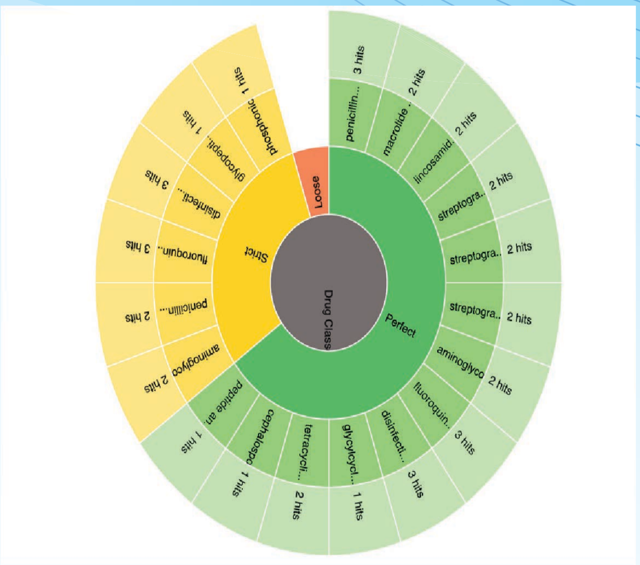


चित्र 1. CARD–RGI का होम पेज जिसमें फास्टा फाइल अपलोड की गई है।



चित्र 2. दिए गए अनुक्रम के लिए AMR पूर्वानुमान के परिणाम तालिका में दिखाए गए हैं, जो दर्शाते हैं कि प्रजाति मेथिसिलिन और अन्य एंटीबायोटिक के प्रति प्रतिरोधी है। मेथिसिलिन एंटीबायोटिक्स के बीटा-लैक्टम वर्ग से संबंधित है और यह अनुक्रम डेटाबेस में मौजूदा प्रतिरोधी अनुक्रमों के साथ 100% समानता दिखा रहा है।





चित्र 3. ए) जीन जो एंटीबायोटिक दवाओं के प्रति प्रतिरोध पैदा करने के लिए जिम्मेदार हैं। बी) एंटीबायोटिक्स जिनके लिए जीन प्रतिरोध दिखा रहा है। [हरा रंग क्वेरी अनुक्रम का CARD डेटाबेस में मौजूदा अनुक्रमों से पूर्ण मिलान दिखा रहा है जबकि लाल रंग अनुक्रम के उस क्षेत्र को दर्शाता है जो कोई मिलान नहीं दिखा रहा है।]

तालिका 1. प्रायोगिक परिणामों का प्रतिनिधित्व।

जीन;नाम	प्रतिरोध तंत्र	एंटीबायोटिक वर्ग	मिलान के प्रकार	पहचान (%)	भविष्यवाणी
mecA	लक्ष्य स्थल का परिवर्तन	Beta–lactams (Methicillin)	बहतरीन मैच	100%	Methicillin के प्रति प्रतिरोधी
blaZ	बीटा–लैक्टामेज एंजाइम उत्पादन	Beta–lactams (Penicillin)	बहतरीन मैच	100%	Penicillin के प्रति प्रतिरोधी
ermA	राइबोसोमल RNA का मिथाइलेशन	Macrolides, Lincosamides	बहतरीन मैच	100%	Erythromycin के प्रति प्रतिरोधी
tetK	इफ्लक्स पंप	Tetracyclines	बहतरीन मैच	100%	Tetracyclin के प्रति प्रतिरोधी

की आवश्यकता होती है, जिसकी तुलना CARD-RGI द्वारा अनुमानित प्रतिरोध जीन से की जाती है। आइए मान लें कि किसी दिए गए बैक्टीरियल आइसोलेट के लिए, CARD-RGI ने 4 एंटीबायोटिक दवाओं के लिए प्रतिरोध की भविष्यवाणी की है। लैब परीक्षण ने उनमें से 3 के लिए प्रतिरोध और 1 के लिए संवेदनशीलता की पुष्टि की। लैब ने CARD-RGI द्वारा अनुमानित नहीं किए गए 1 एंटीबायोटिक के लिए प्रतिरोध की भी पुष्टि की। फिर:

निष्कर्ष

पौधों और जानवरों में एएमआर के दूरगामी परिणाम होते हैं, जिनमें उपचार विफलताएं, आर्थिक नुकसान, पर्यावरण प्रदूषण और सार्वजनिक स्वास्थ्य जोखिम शामिल हैं। प्रारंभिक पहचान, एंटीबायोटिक प्रबंधन और वैकल्पिक उपचार विकसित करने के लिए AMR की भविष्यवाणी करना आवश्यक है। सिलिको में कई प्रयोगों के एकीकरण से एएमआर जीन की तेजी से पहचान और प्रतिरोध जीन की

Accuracy	Sensitivity	Specificity	Precision	F1 Score
0.81	0.75	0.83	0.75	0.75

निगरानी की अनुमति मिलती है, जिससे प्रभावी एएमआर प्रबंधन रणनीतियों के विकास में मदद मिलती है। उन्नत जीनोमिक्स और बायोइनफॉरमैटिक्स दृष्टिकोणों को लागू करने से एएमआर के प्रसार को कम करने, टिकाऊ खाद्य उत्पादन और वैश्विक स्वास्थ्य सुरक्षा सुनिश्चित करने में मदद मिल सकती है।

संदर्भ

टैंग, के. डब्ल्यू. के., मिलर, बी. सी., और मूर, जे. ई. (2023)। रोगाणुरोधी प्रतिरोध (एएमआर)। ब्रिटिश जर्नल ऑफ बायोमेडिकल साइंस, 80, 11387।

होसैन, एम. जेड., कबीर, एस. एल., और कमाल, एम. एम. (2021)। विकासशील देशों में पशुधन उत्पादन के लिए रोगाणुरोधी उपयोग। वेटरनरी वर्ल्ड, 14(1), 210।

डैडगोस्टार, पी. (2019)। एंटीमाइक्रोबियल प्रतिरोध: निहितार्थ और लागत। संक्रमण और दवा प्रतिरोध, 3903-3910।

एल्कॉक, बी.पी., हुइन्ह, डब्ल्यू., चालिल, आर., स्मिथ, के.डब्ल्यू., राफेन्या, ए.आर., वलोडास्र्की, एम.ए., ... और मैकआर्थर, ए.जी. (2023)। कार्ड 2023: व्यापक एंटीबायोटिक प्रतिरोध डेटाबेस में विस्तारित क्यूरेशन, मशीन लर्निंग के लिए समर्थन और रेसिस्टोम भविष्यवाणी। न्यूक्लिक एसिड अनुसंधान, 51(डी1), डी690-डी699।

ब्रांडा, एफ., और स्कार्पा, एफ. (2024)। एंटीमाइक्रोबियल प्रतिरोध को संबोधित करने में आर्टिफिशियल इंटेलिजेंस के निहितार्थ: नवाचार, वैश्विक चुनौतियाँ और स्वास्थ्य सेवा का भविष्य। एंटीबायोटिक्स, 13(6), 502।

होसैन, एम. जेड., कबीर, एस. एल., और कमाल, एम. एम. (2021)। विकासशील देशों में पशुधन उत्पादन के लिए रोगाणुरोधी उपयोग। वेटरनरी वर्ल्ड, 14(1), 210।

केल्ब्रिक, एम., हेस्से, ई., और ओश्ब्रायन, एस. (2023)। रोगाणुरोधी प्रतिरोध की खेती: कैसे गहन कृषि एंटीबायोटिक प्रतिरोध के लिए रास्ता बनाती है। माइक्रोबायोलॉजी, 169(8), 001384।

मोर, ए. वी., सोनवाने, एस. डी., और बैस, एस. के. दवा विकास में एक प्रमुख रसायन के रूप में प्राकृतिक उत्पादों पर एक समीक्षा।

सीओने, ए., और बौ, जी. (2021)। रोगाणुरोधी प्रतिरोध के अध्ययन के लिए जैव सूचना विज्ञान दृष्टिकोण। रेविस्टा एस्पानोला डी क्विमियोथेरेपी, 34(सप्ल 1), 15।

जंकारी, ई., हसमैन, एच., कॉसेंटिनो, एस., वेस्टरगार्ड, एम., रासमुसेन, एस., लुंड, ओ., ... और लार्सन, एम. वी. (2012)। अधिग्रहित रोगाणुरोधी प्रतिरोध जीन की पहचान। जर्नल ऑफ एंटीमाइक्रोबियल कीमोथेरेपी, 67(11), 2640-2644।

ओल्सन, आर. डी., असफ, आर., ब्रेटिन, टी., कॉनराड, एन., कुसिनेल, सी., डेविस, जे. जे., ... और स्टीवंस, आर. एल. (2023)। जीवाणु और वायरल जैव सूचना विज्ञान संसाधन केंद्र (बीवी-बीआरसी) का परिचय: पैट्रिक, आईआरडी और वीआईपीआर को मिलाकर एक संसाधन। न्यूक्लिक एसिड अनुसंधान, 51(डी1), डी678-डी689। सागा, टी., और यामागुची, के. (2009)। रोगाणुरोधी एजेंटों और प्रतिरोधी बैक्टीरिया का इतिहास। जेमज, 52(2), 103-108।

मिसिंग डेटा को संभालने के लिए इम्प्यूटेशन विधियों की समीक्षा

राणा क्रिना दिव्येशभाई, पंकज दास, भारती, अंकुर बिश्वास, तौकीर अहमद, दीपक सिंह, राजू कुमार
भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, लाइब्रेरी एवेन्यू, नई दिल्ली

सार

अनुपस्थित डेटा सांख्यिकीय विश्लेषण को महत्वपूर्ण रूप से प्रभावित कर सकता है, जिससे पक्षपाती परिण ााम और सटीकता में कमी आ सकती है। इस समस्या को हल करने के लिए विभिन्न इम्प्यूटेशन विधियों का उपयोग किया जाता है, जिनमें सरल इम्प्यूटेशन, रिग्रेशन इम्प्यूटेशन, मल्टीपल इम्प्यूटेशन और मशीन लर्निंग आधारित विधियाँ जैसे के-नियरिस्ट नेबरस, रैंडम फॉरेस्ट और एंसेंबल विधियाँ शामिल हैं। प्रत्येक विधि की अपनी विशेषताएं और सीमाएँ होती हैं, और उपयुक्त विधि का चयन करने में कम्प्यूटेशनल लागत, डेटा की जटिलता और अनुपस्थित डेटा की प्रकृति जैसी चीजें महत्वपूर्ण भूमिका निभाती हैं। मल्टीपल इम्प्यूटेशन और एंसेंबल विधियाँ आमतौर पर सबसे विश्वसनीय मानी जाती हैं, विशेष रूप से जटिल या बड़े डेटा सेटों के लिए, लेकिन इन्हें सावधानीपूर्वक लागू करने की आवश्यकता होती है।

परिचय

सांख्यिकीय विश्लेषण में अनुपस्थित मान एक महत्वपूर्ण समस्या है क्योंकि वे पक्षपाती अनुमान, कम सांख्यिकीय शक्ति और गलत निष्कर्ष का कारण बन सकते हैं। डेटासेट में विभिन्न अनुपस्थित डेटा पैटर्न (यूनीवेरिएट, मोनोटोन, नॉन मोनोटोन) और जैसे कि मिसिंग कंप्लीटली एट रैंडम (एमसीएआर), मिसिंग एट रैंडम (एमएआर), मिसिंग नॉट एट रैंडम (एमएनएआर) (मैक एट अल, 2018) हैं। डेटासेट में लुप्त मानों को संभालने के लिए इम्प्यूटेशन विधियों का उपयोग किया जाता है। इम्प्यूटेशन की प्रक्रिया में कुछ अनुमानित मूल्यों द्वारा अनुपस्थित मानों को बदलना शामिल है। गैर- अनुपस्थित मान डेटा सेट का उपयोग आमतौर पर अनुपस्थित मानों को बदलने के लिए उपयोग किए जाने वाले मूल्यों का अनुमान लगाने के लिए किया जाता है (डोंडर्स एट अल, 2006)।

इम्प्यूटेशन विधियाँ

इम्प्यूटेशन के कई प्रकार हैं, जिनमें से प्रत्येक अलग-अलग प्रकार के डेटा और परिदृश्यों के लिए उपयुक्त है:

- **सरल इम्प्यूटेशन**

सरल इम्प्यूटेशन में अनुपस्थित मानों को मात्रात्मक या गुणात्मक विशेषता से प्रतिस्थापित करना शामिल है, जैसे कि उपलब्ध डेटा का माध्य, माध्यिका या बहुलक। जबकि यह विधि अपनी सरलता और कार्यान्वयन में आसानी के लिए लोकप्रिय है, यह पूर्वाग्रह या अवास्तविक परिणाम पेश कर सकती है, विशेष रूप से उच्च-आयामी डेटासेट में। बड़े डेटा के उदय के साथ, सरल इम्प्यूटेशन कम प्रभावी हो गया है और बड़े, जटिल डेटासेट को संभालने के लिए अपर्याप्त माना जाता है।

- **प्रतिगमन (रिग्रेशन) इम्प्यूटेशन**

प्रतिगमन गुम डेटा को संभालने के लिए एक आम तौर पर इस्तेमाल की जाने वाली तकनीक है, जिसे अक्सर सशर्त माध्य प्रतिरूपण के रूप में संदर्भित किया जाता है। यह प्रक्रिया दो चरणों में की जाती है: सबसे पहले, संपूर्ण डेटा का उपयोग करके एक प्रतिगमन मॉडल (रिग्रेशन मॉडल) बनाया जाता है, और फिर इस मॉडल से पूर्वानुमानों के आधार पर गुम मूल्यों को प्रतिरूपित किया जाता है।

हालांकि, स्थिर परिणामों के लिए इसके लिए पर्याप्त रूप से बड़े डेटासेट की आवश्यकता होती है। एक सीमा यह है कि यह प्रतिरूपण के लिए एकल प्रतिगमन वक्र का उपयोग करता है, जो डेटा में वास्तविक परिवर्तनशीलता को प्रतिबिंबित नहीं कर सकता है। यदि कई विशेषताओं में गुम मूल्य हैं, तो एक बहुभिन्नरूपी प्रतिगमन मॉडल का उपयोग किया जाता है। यह मॉडल देखता है कि कई स्वतंत्र चर कई आश्रित चर से कैसे संबंधित हैं। यह विधि प्रभावी पाई गई लेकिन इसकी सीमाएँ थीं, जैसे कि जब गुम डेटा बहुत अधिक प्रचलित था या जब मॉडल गलत तरीके से निर्दिष्ट किया गया था तो यह अनुपयुक्त था । यह विधि डेटा संबंधों को बनाए रखती है लेकिन यदि मॉडल गलत तरीके से निर्दिष्ट किया गया हो तो सटीक परिणाम नहीं देती।

- **मल्टीपल इम्प्यूटेशन**

रुबिन (1987) द्वारा प्रस्तावित मल्टीपल इम्प्यूटेशन एक ऐसी तकनीक है जिसका उपयोग कई इंप्यूटेड डेटासेट बनाकर लापता डेटा को संबोधित करने के लिए किया जाता है जो लापता मूल्यों की अनिश्चितता को पकड़ते हैं। एकल इंप्यूटेशन के विपरीत, जो कम विश्वसनीय परिणामों को जन्म दे सकता है, मल्टीपल इम्प्यूटेशन में तीन मुख्य चरण शामिल हैं: इंप्यूटेड मूल्यों के साथ कई पूर्ण डेटासेट बनाना, प्रत्येक डेटासेट का विश्लेषण करना और फिर अंतिम अनुमान बनाने के लिए परिणामों को संयोजित करना। जबकि इसे व्यापक रूप से लापता डेटा को संभालने के लिए एक मानक दृष्टिकोण माना जाता है, इसकी प्रभावशीलता वास्तविक दुनिया की जटिलताओं से प्रभावित हो सकती है, जैसे कि लापता डेटा की उच्च दर और सर्वेक्षण, नैदानिक, या औद्योगिक डेटासेट में अक्सर पाए जाने वाले गैर-रैखिक संबंध। इसके अतिरिक्त, पारंपरिक मल्टीपल इम्प्यूटेशन विधियाँ उच्च-आयामी डेटा के साथ संघर्ष कर सकती हैं, जिससे बेहतर एल्गोरिदम की आवश्यकता होती है। निरंतर-आधारित इंप्यूटेशन विधियों को श्रेणीबद्ध डेटा पर लागू करते समय भी सावधानी बरतनी चाहिए, क्योंकि इससे परिणामों में पूर्वाग्रह आ सकता है। यह विधि डेटा में मौजूद अनिश्चितता को ध्यान में रखती है, लेकिन यह गणनात्मक रूप से अधिक जटिल होती है।

- **मशीन लर्निंग आधारित इम्प्यूटेशन**

मशीन लर्निंग पर आधारित इंप्यूटेशन विधियाँ परिष्कृत तकनीकें हैं, जिनमें अधिकतर अप्रशिक्षित या पर्यवेक्षित शिक्षण का उपयोग करके लापता मानों को संभालने के लिए एक पूर्वानुमानित दृष्टिकोण विकसित करना शामिल है।

- **के-निकटतम पड़ोसी (केएनएन):** के-नियरिस्ट नेबरस (केएनएन) प्रतिरूपण एल्गोरिथ्म यूक्लिडियन, मैनहट्टन या कोसाइन दूरी जैसे दूरी माप का उपयोग करके किसी मिसिंग मूल्य के निकटतम नेबरस की पहचान करके गुम डेटा को भरता है। यूक्लिडियन दूरी अपनी दक्षता के कारण सबसे अधिक उपयोग की जाती है। केएनएन असतत और निरंतर डेटा दोनों को संभाल सकता है और डेटासेट में कई गुम मूल्यों को प्रतिरूपित करने में सक्षम है। हालाँकि, इसकी सीमाएँ हैं, जिनमें संभावित कम सटीकता और नेबरस के लिए पूरे डेटासेट में खोज करने की आवश्यकता के कारण बढ़ा हुआ कम्प्यूटेशनल समय शामिल है। इन चुनौतियों के बावजूद, इसके प्रदर्शन को बढ़ाने के लिए विभिन्न सुधारों का सुझाव दिया गया है।
- **रैंडम फॉरेस्ट:** रैंडम फॉरेस्ट बैगिंग के माध्यम से बनाए गए डिसिशन ट्री का लाभ उठाकर मिसिंग मूल्यों को लागू करने के लिए एक लोकप्रिय विधि है, जो बहुमत या औसत के आधार पर पूर्वानुमान बनाने के लिए कई यादृच्छिक अनुमानों को जोड़ती है। इस प्रक्रिया में प्रतिदर्श का यादृच्छिक रूप से चयन करना, विभाजन के लिए चर चुनना और मिसिंग मूल्यों को लागू करने के लिए कई ट्री का निर्माण करना शामिल है। जबकि अध्ययनों से पता चला है कि यह प्रतिगमन और सपोर्ट वेक्टर मशीनों जैसे पारंपरिक तरीकों की तुलना में अच्छा प्रदर्शन करता है, लेकिन इसकी सीमाएँ हैं। रैंडम फॉरेस्ट कम्प्यूटेशनल रूप से महंगा हो सकता

है, गहरे या कई ट्री के साथ ओवरफिट होने का खतरा होता है, और इसमें व्याख्या की कमी होती है, जिससे यह समझना मुश्किल हो जाता है कि इंप्यूटेशन कैसे प्राप्त किए जाते हैं। यदि डेटा यादृच्छिक रूप से गुम नहीं है (एमएनएआर) तो यह पूर्वाग्रह का परिचय दे सकता है और छोटे डेटासेट, उच्च-कार्डिनैलिटी श्रेणीबद्ध डेटा या अत्यधिक सहसंबंधित विशेषताओं के साथ संघर्ष करता है, जिससे यह कुछ मामलों में कम प्रभावी हो जाता है। इसलिए, विशेष रूप से जटिल या सीमित डेटा के लिए, इंप्यूटेशन के लिए रैंडम फॉरेस्ट का उपयोग करते समय सावधानीपूर्वक विचार करने की आवश्यकता है।

- **एंसेंबल विधियाँ:** एनसेंबल विधियाँ कई इम्पुटेशन मॉडल से पूर्वानुमानों को जोड़ती हैं, विभिन्न एल्गोरिदम की पूरक शक्तियों का लाभ उठाती हैं, जैसे कि के- नियरिस्ट नेबरस (झांग, 2012), रैखिक प्रतिगमन, और रैंडम फॉरेस्ट-आधारित विधियाँ (स्टेकहोवेन और बुहलमैन, 2012)। ये विधियाँ विशेष रूप से गुम डेटा को संभालने के लिए उपयोगी हैं, क्योंकि अध्ययनों से पता चला है कि वे एकल एल्गोरिदम की तुलना में अधिक सटीक परिणाम प्रदान करते हैं। इसके अतिरिक्त, एनसेंबल तकनीकों को समानांतर कंप्यूटिंग वातावरण में लागू किया जा सकता है, जिससे वे बड़े डेटासेट के लिए उपयुक्त हो जाते हैं। एनसेंबल विधियों में सामान्य रणनीतियों में बैगिंग, बूस्टिंग और स्टैकिंग शामिल हैं। बैगिंग डेटा के विभिन्न यादृच्छिक उपसमूहों का उपयोग करके कई मॉडलों को प्रशिक्षित करके काम करता है, औसत के माध्यम से उनके अनुमानों को जोड़ता है। इसे लागू करना सरल और कुशल है। दूसरी ओर, बूस्टिंग, डेटा को फिर से सैंपल करके और अनुमानों को मिलाकर कमजोर सीखने के एल्गोरिदम को मजबूत में बदलने पर ध्यान केंद्रित करता है, लेकिन इसमें प्रशिक्षण का समय धीमा होता है। स्टैकिंग विभिन्न मॉडलों को जोड़ती है और अंतिम अनुमान करने के लिए मेटा-लेवल क्लासिफायर के लिए इनपुट के रूप में उनके अनुमानों का उपयोग करती है। जब उच्च सटीकता की आवश्यकता होती है तो एनसेंबल विधियाँ आम तौर पर सबसे अच्छी तरह से काम करती हैं और उन्हें हाथ में मौजूद विशिष्ट समस्या के अनुरूप विविधतापूर्ण होने के लिए डिजाइन किया जाता है। उदाहरण के लिए, खान एट अल. ने डेटा में उच्च मिसिंगनेस को संभालने के लिए बूटस्ट्रैपिंग एनसेंबल विधि का प्रस्ताव दिया, जिसमें दिखाया गया कि यह उच्च मिसिंगनेस अनुपात (30% तक) के लिए विशेष रूप से मजबूत है और एकल इम्प्यूटेशन विधियों की तुलना में बेहतर प्रदर्शन करता है।

इम्प्यूटेशन विधियों की तुलना

निष्कर्ष

इम्प्यूटेशन विधियाँ अनुपस्थित डेटा को संभालने में महत्वपूर्ण भूमिका निभाती हैं, जिनमें सरल इम्प्यूटेशन से लेकर उन्नत मशीन लर्निंग तकनीकें शामिल हैं। प्रत्येक

विधि	लाभ	सीमाएँ
सरल इम्प्यूटेशन	सरल, तेज	पक्षपातपूर्ण या अवास्तविक परिणाम दे सकता है
रिग्रेशन इम्प्यूटेशन	डेटा में संबंधों को पकड़ता है	मॉडल की गलत निर्दिष्टता पर निर्भर करता है
मल्टीपल इम्प्यूटेशन	अनिश्चितता को ध्यान में रखता है	गणनात्मक रूप से जटिल
केएनएन इम्प्यूटेशन	गैर-पैरामीट्रिक, लचीला	गणनात्मक रूप से महंगा
रैंडम फॉरेस्ट	जटिल संबंधों को संभाल सकता है	अधिक फिटिंग की समस्या, ट्यूनिंग की आवश्यकता
एंसेंबल विधि	पक्षपात को कम करता है, अधिक सटीकता	विविध और गतिशील डेटा के प्रति संवेदनशील

विधि की अपनी विशेषताएं और फायदे होते हैं, और सबसे उपयुक्त विधि डेटा सेट की जटिलता और अनुपस्थित डेटा तंत्र पर निर्भर करती है। मल्टीपल इम्प्यूटेशन और एसेंबल विधियाँ आमतौर पर सबसे विश्वसनीय मानी जाती हैं, विशेष रूप से जटिल या बड़े डेटा सेटों के लिए।

संदर्भ

गार्सिया-लांसीना, पी. जे., सांचो-गोमेज, जे. एल., फिगुएरास-विडाल, ए. आर., और वर्लेसेन, एम. (2010) के-नियरिस्ट नेबरस विद म्यूचुअल इंफॉर्मेशन फॉर सिमल्टेनियस क्लासिफिकेशन एंड मिसिंग डेटा इम्प्यूटेशन। न्यूरोकंप्यूटिंग, 72(7–9), 1483-1493।

झांग, एस. (2012) नियरिस्ट नेबर चयन फॉर इटरेटिवली kNN इम्प्यूटेशन। जर्नल ऑफ सिस्टम्स एंड सॉफ्टवेयर। जर्नल ऑफ सिस्टम्स एंड सॉफ्टवेयर, 85(11): 2541–2552।

ट्वाला, बी. (2009) एन एम्पिरिकल कम्पैरिजन ऑफ टेक्नीकस फॉर हैंडलिंग इनकम्प्लीट डेटा यूजिंग डिसीजन ट्रीज। एप्लाइड आर्टिफिशियल इंटेलिजेंस, 23(5), 373–405।

डोंडर्स एआरटी, वैन डेर हेजडेन जीजे, स्टिजनन टी, मून्स केजी. (2006) ए जेंटल इंट्रोडक्शन टू इम्प्यूटेशन ऑफ मिसिंग वैल्यूज। जेक्लिन एपिडेमियोल, 59(10):1087-91।

मैक सी, झाओहुई सु, और वेस्ट्रेइच डी. (2018) टाइप्स ऑफ मिसिंग डेटा। इन मैनेजिंग मिसिंग डेटा इन पेशेंट रेजिस्ट्रीज: अडेंडम टू रेजिस्ट्रीज फॉर ईवैल्यूएटिंग पेशेंट आउटकम्स: ए यूजरश्स गाइड, थर्ड एडिशन [इंटरनेट]। यूआरएल https://www.ncbi.nlm.nih.gov/books/NBK493614/।

रुबिन, डी. बी. (1987) मल्टिपल इम्प्यूटेशन फॉर नॉनरेस्पांस इन सर्वेज। जॉन वाइली एंड संस।

लिटिल र. एंड रुबिन डी. (2019) स्टैटिस्टिकल एनालिसिस विद मिसिंग डेटा। विल्ली इंडिया, थर्ड एडिशन द्य

वन बुरें, स. (2018) फ्लेक्सिबल इम्प्यूटेशन ऑफ मिसिंग डेटा। चौपमैन एंड हॉल/सीआरसी।

स्टेखोवन डी. और बुहलमैन पी. (2012) मिसफॉरेस्ट-नॉन-पैरामेट्रिक मिसिंग वैल्यू इम्प्यूटेशन फॉर मिक्स्ड-टाइप डेटा। बायोइन्फॉर्मेटिक्स, 28 (1):112–118।

"अधिकांश डेटा सेट में अनुपस्थित डेटा एक वास्तविकता है। सबसे अच्छा समाधान जरूरी नहीं कि अनुपस्थित डेटा को हटा दिया जाए, बल्कि आधुनिक इम्प्यूटेशन तकनीकों का सही तरीके से उपयोग करना चाहिए।"

पॉल डी. एलिसन, समाजशास्त्री और सांख्यिकीय विधियों के विशेषज्ञ

सांख्यिकी-विमर्श
2024

अंक
20

राजभाषा खण्ड

संस्थान की राजभाषा यात्राः वर्ष 2024

हिन्दी एकक

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान में दिन प्रतिदिन हिन्दी के प्रगामी प्रयोग में अभिवृद्धि हो रही है। राजभाषा नीति को संस्थान में सुचारु रूप से कार्यान्वित किया जा रहा है। भारत सरकार, गृह मंत्रालय, राजभाषा विभाग द्वारा जारी वार्षिक कार्यक्रम में निर्धारित लक्ष्यों को संस्थान में लगभग पूरा कर लिया गया है। संस्थान द्वारा समस्त प्रशासनिक कार्य शत-प्रतिशत हिन्दी में किया जाता है तथा धारा 3(3) का भी पूर्ण रूप से अनुपालन किया जा रहा है।

संस्थान में राजभाषा हिन्दी की प्रगति का जायजा लेने के लिए उपमहानिदेशक (कृषि शिक्षा), भारतीय कृषि अनुसंधान परिषद मुख्यालय द्वारा 01 जनवरी 2024, 21 मई 2024, 14 अक्टूबर 2024 तथा 20 दिसम्बर 2024 को संस्थान के राजभाषा संबंधी निरीक्षण किए गए। उपमहानिदेशक (कृषि शिक्षा), ने निरीक्षण रिपोर्ट में संस्थान में हिन्दी में हो रहे कार्यों की प्रगति पर संतोष व्यक्त करते हुए संस्थान की सराहना की। संस्थान के अलग-अलग प्रभागों/अनुभागों में हिन्दी में किए जा रहे कार्यों की समीक्षा करने के लिए हिन्दी एकक के अधिकारियों द्वारा कुल 19 प्रभागों/अनुभागों का निरीक्षण किया गया।

संस्थान में प्रशासनिक कार्य के साथ-साथ वैज्ञानिक प्रकृति के कार्यों में भी हिन्दी के उपयोग को प्रोत्साहित किया जाता है। संस्थान के वैज्ञानिक प्रभागों द्वारा आयोजित प्रशिक्षण कार्यक्रमों की ई-संदर्भ पुस्तिकाओं में आमुख एवं आवरण पृष्ठ द्विभाषी रूप में प्रस्तुत करने के साथ-साथ परियोजना रिपोर्ट के आवरण पृष्ठ, आमुख, एवं सारांश द्वि भाषी रूप में प्रस्तुत करने के साथ-साथ कुछ हिन्दी के व्याख्यान भी शामिल किए। संस्थान के वैज्ञानिकों द्वारा हिन्दी में वैज्ञानिक विषयों पर हिन्दी कार्यशालाओं का भी आयोजन किया गया। इसके अतिरिक्त, संस्थान में एम.एस. सी. तथा पी.एच.डी के विद्यार्थियों ने अपने शोध-प्रबन्धों के सार द्विभाषी रूप में प्रस्तुत किए। वैज्ञानिकों एवं तकनीकी कर्मियों द्वारा कुछ शोध-पत्र भी हिन्दी में प्रकाशित किए गए।

प्रतिवेदनाधीन अवधि के दौरान संस्थान में राजभाषा कार्यान्वयन समिति की 04 बैठकें क्रमशः मार्च 26, 2024; जून 27, 2024; सितम्बर 27, 2024 एवं दिसम्बर 30, 2024 को आयोजित की गई। इन बैठकों में राजभाषा नियम एवं अधिनियम को कारगर ढंग से लागू करने तथा इसमें दिए गए प्रावधानों के अनुसार वार्षिक कार्यक्रम में निर्धारित लक्ष्यों की प्राप्ति हेतु चर्चा की गई तथा आवश्यक कदम उठाए गए।

संस्थान के समस्त कर्मियों को 02 वर्ष की अवधि में कम से कम एक बार हिन्दी कार्यशाला में सहभागिता करने का अवसर मिले, इस अनिवार्यता के संबंध में भारत सरकार, गृह मंत्रालय, राजभाषा विभाग द्वारा समय-समय पर जारी कार्यालय ज्ञापन द्वारा निर्धारित लक्ष्य की प्राप्ति के लिए, संस्थान द्वारा अक्तूबर 10, 2022 से दिसम्बर 31, 2024 के दौरान राजभाषा हिन्दी के साथ-साथ संस्थान से संबंधित विभिन्न विषयों पर हिन्दी कार्यशाला का आयोजन कर संस्थान के समस्त कर्मियों को कम से कम एक बार हिन्दी कार्यशाला में प्रशिक्षित किया गया। इस प्रकार संस्थान ने राजभाषा विभाग द्वारा निर्धारित लक्ष्य को उक्त अवधि में पूरा किया।

प्रतिवेदनाधीन अवधि के दौरान संस्थान में विभिन्न वर्गों के कर्मियों एवं राष्ट्रीय कृषि अनुसंधान प्रणाली के अन्य संस्थानों के वैज्ञानिकों के लिए पाँच हिन्दी कार्यशालाएं आयोजित की गई। पहली कार्यशाला संस्थान में वैज्ञानिक एवं तकनीकी वर्ग के कर्मियों के लिए संगणक अनुप्रयोग प्रभाग के वैज्ञानिकों, डॉ. मधु दहिया, डॉ. संचिता नाहा एवं डॉ. सपना निगम द्वारा मार्च 06-08, 2024 (03 दिवसीय) के दौरान "कृषि

शिक्षा में डिजिटल पहल" विषय पर ऑफ-लाइन आयोजित की गई, जिसमें 07 वक्ताओं द्वारा विषय से संबन्धित 13 उप-विषयों पर व्याख्यान दिए गए। इस कार्यशाला में कुल 14 प्रतिभागियों ने सहभागिता की, जिसमें संस्थान के 09 वैज्ञानिक, 04 तकनीकी अधिकारी तथा 01 तकनीकी सहायक द्वारा सहभागिता की।

दूसरी कार्यशाला संस्थान में वैज्ञानिक एवं तकनीकी वर्ग के कर्मियों के लिए कृषि जैव सूचना विज्ञान प्रभाग के वैज्ञानिकों, डॉ. सुधीर श्रीवास्तव, डॉ. मोहम्मद समीर फारुकी एवं डॉ. स्नेहा मुर्मू द्वारा 27 जून 2024 को (एक पूर्ण दिवसीय) के दौरान "कृषि में ओमिक्स डेटा विश्लेषण का परिचय" विषय पर ऑन-लाइन आयोजित की गई, जिसमें 05 वक्ताओं द्वारा विषय से संबंधित 05 उप-विषयों पर व्याख्यान दिए गए। इस कार्यशाला में कुल 17 प्रतिभागियों ने सहभागिता की, जिसमें संस्थान के 11 वैज्ञानिक तथा 03 तकनीकी सहायक तथा भा.कृ.अनु.प. के बाह्य संस्थानों से 01 वैज्ञानिक और 01 सहायक प्रोफेसर के अलावा 01 तकनीकी अधिकारी द्वारा सहभागिता की।

तीसरी कार्यशाला संस्थान में वैज्ञानिक, तकनीकी एवं प्रशासनिक वर्ग के कर्मियों के लिए सूचना प्रौद्योगिकी प्रकोष्ठ के वैज्ञानिकों डॉ. एस. बी. लाल, डॉ. मुकेश कुमार एवं श्री संजीव कुमार द्वारा सितम्बर 11, 2024 (01 पूर्ण दिवसीय) के दौरान "परिषद में ई-गवर्नेंस का अनुप्रयोग" विषय पर ऑफ-लाइन आयोजित की गयी जिसमें 06 वक्ताओं द्वारा विषय से संबंधित 06 उप-विषयों पर व्याख्यान दिए गए। इस कार्यशाला में कुल 26 कर्मियों ने सहभागिता की, जिसमें संस्थान के 12 वैज्ञानिक, 01 वरिष्ठ तकनीकी अधिकारी, 01 सहायक मुख्य तकनीकी अधिकारी, 03 सहायक प्रशासनिक अधिकारी, 01 तकनीकी सहायक, 04 सहायक, 02 प्रवर लिपिक तथा 02 अवर लिपिक द्वारा सहभागिता की।

चौथी कार्यशाला संस्थान में वैज्ञानिक एवं तकनीकी वर्ग के कर्मियों के लिए परीक्षण अभिकल्पना प्रभाग के वैज्ञानिकों, डॉ. मोहम्मद हारून एवं डॉ. अनिंदीता दत्ता द्वारा दिसम्बर 19, 2024 (01 पूर्ण दिवसीय) के दौरान "परीक्षण अभिकल्पनाएं एवं विश्लेषण" विषय पर ऑन-लाइन आयोजित की गई जिसमें 03 वक्ताओं द्वारा विषय से संबंधित 06 उप-विषयों पर व्याख्यान दिए गए। इस कार्यशाला में कुल 23 कर्मियों ने सहभागिता की, जिसमें संस्थान के 16 वैज्ञानिक, 02 मुख्य तकनीकी अधिकारी, 01 वरिष्ठ तकनीकी अधिकारी, 01 तकनीकी अधिकारी, 01 सहायक मुख्य तकनीकी अधिकारी, तथा 02 तकनीकी सहायक द्वारा सहभागिता की।

पाँचवी कार्यशाला संस्थान में वैज्ञानिक, तकनीकी एवं प्रशासनिक वर्ग कर्मियों के लिए हिन्दी एकक द्वारा दिसम्बर 24, 2024 (11:00 बजे से 01:00 बजे) के दौरान "भारतीय संविधान" विषय पर ऑफ-लाइन आयोजित की गई। जिसमें 01 वक्ता द्वारा विषय से संबंधित 01 उप-विषय पर व्याख्यान दिया गया। इस कार्यशाला में कुल 34 कर्मियों ने सहभागिता की, जिसमें संस्थान के 13 वैज्ञानिक, 01 मुख्य प्रशासनिक अधिकारी, 01 मुख्य तकनीकी अधिकारी, 02 सहायक मुख्य तकनीकी अधिकारी, 04 सहायक प्रशासनिक अधिकारी, 01 निजी सहायक, 07 सहायक, 02 प्रवर लिपिक एवं 03 अवर लिपिक द्वारा सहभागिता की।

राजभाषा विभाग द्वारा जारी वार्षिक कार्यक्रम में निहित लक्ष्यों को पूरा करते हुए संस्थान के अधिकारियों/कर्मचारियों द्वारा समस्त पत्राचार हिन्दी में अथवा द्विभाषी रूप में किया गया। संस्थान के विभिन्न वैज्ञानिक प्रभागों तथा प्रशासनिक अनुभागों द्वारा आयोजित बैठकों की कार्यसूची तथा कार्यवृत्त हिन्दी अथवा द्विभाषी रूप में जारी किए गए। संस्थान में अपना कार्य शत-प्रतिशत हिन्दी में करने के लिए 11 अनुभागों को विनिर्दिष्ट किया गया है। गृह मंत्रालय, राजभाषा विभाग द्वारा जारी विभिन्न नकद पुरस्कार योजनाएँ संस्थान में लागू हैं तथा संस्थान के कर्मियों ने इन योजनाओं में बढ़-चढ़कर भाग लिया।

संस्थान में कार्यरत सभी हिन्दीतर अधिकारियों/कर्मचारियों द्वारा हिन्दी ज्ञान संबंधी प्रशिक्षण पूरा किया जा चुका है। आज तक की स्थिति के अनुसार, संस्थान में अब कोई ऐसा हिन्दीतर अधिकारी/कर्मचारी शेष नहीं रह गया है जिसे हिन्दी ज्ञान संबंधी प्रशिक्षण दिया जाना शेष हो।

संस्थान की वेबसाइट पर "हिन्दी सेवा लिंक" उपलब्ध है जिसमें सांख्यिकीय एवं प्रशासनिक शब्दावली के वर्ण क्रमानुसार कुछ शब्द, कुछ द्विभाषी प्रपत्र, दैनिक काम-काज के प्रयोग में आने वाली कुछ टिप्पणियाँ, द्विभाषी पदनाम, वाक्यांश इत्यादि उपलब्ध हैं। संस्थान के कर्मियों द्वारा अपना दैनिक कार्य हिन्दी में सरलता से करने के लिए इस सेवा का उपयोग किया जाता है।

संस्थान द्वारा प्रकाशित वार्षिक हिन्दी पत्रिका 'सांख्यिकी-विमर्श' का नियमित प्रकाशन किया जा रहा है तथा 19वें अंक का प्रकाशन प्रतिवेदनाधीन अवधि के दौरान किया गया।

संस्थान में 14 से 30 सितम्बर, 2024 के दौरान हिन्दी पखवाड़े का आयोजन किया गया। इस वर्ष हिन्दी पखवाड़े का शुभारम्भ अर्थात उदघाटन हिन्दी दिवस एवं चतुर्थ अखिल भारतीय राजभाषा सम्मेलन के अवसर पर 14 सितम्बर, 2024 को भारत मंडपम, नई दिल्ली में माननीय गृह राज्य मंत्री जी द्वारा किया गया था। हिन्दी पखवाड़े का आयोजन एवं इससे संबन्धित प्रतियोगताएं संस्थान में दिनांक 18 से 30 सितम्बर, 2024 के दौरान आयोजित की गयी। दिनांक 18 सितम्बर, 2024 को काव्य-पाठ प्रतियोगिता का आयोजन किया गया। हिन्दी पखवाड़ा के दौरान डॉ दरोगा सिंह स्मृति व्याख्यान के साथ-साथ वैज्ञानिक प्रभागों में हिन्दी में सर्वाधिक वैज्ञानिक कार्य करने के लिए प्रभागीय चल-शील्ड, डिजिटल हिन्दी शोध-पत्र प्रस्तुतिकरण प्रतियोगिता (वैज्ञानिक वर्ग के कर्मियों के लिए), हिन्दी श्रुतलेख प्रतियोगिता, हिंदीतर कर्मियों के लिए शब्दार्थ लेखन प्रतियोगिता, अंताक्षरी प्रतियोगिता तथा प्रश्न मंच प्रतियोगिता भी आयोजित की गई। सभी प्रतियोगिताओं में संस्थान के विभिन्न वर्ग के कर्मियों ने बढ़-चढ़कर हिस्सा लिया। संस्थान में प्रत्येक वर्ष हिन्दी दिवस के अवसर पर डॉ दरोगा सिंह स्मृति व्याख्यान का आयोजन किया जाता है। इस वर्ष इस कड़ी का तैंतीसवां व्याख्यान राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय (एन.एस.एस.ओ.) के पूर्व महानिदेशक श्री राकेश कुमार त्यागी द्वारा दिया गया और इस कार्यक्रम की अध्यक्षता संस्थान के निदेशक महोदय द्वारा की गई। दिनांक 30 सितम्बर, 2024 को हिन्दी पखवाड़ा के समापन समारोह के अवसर पर इस दौरान आयोजित प्रतियोगिताओं के सफल प्रतियोगियों को नकद पुरस्कारों की घोषणा की गयी। इसके अतिरिक्त इस अवसर पर अक्टूबर, 2023 से अगस्त, 2024 तक की अवधि के दौरान संस्थान में आयोजित हिन्दी कार्यशालाओं के वक्ताओं को प्रशस्ति-पत्र प्रदान किए जाने की भी घोषणा की गयी।

भारतीय कृषि सांख्यिकी अनुसंधान संस्थान – कृषि सांख्यिकी में विश्व का अग्रणी संस्थान

ब्रह्म प्रकाश
भा.कृ.अनु.प.–भारतीय गन्ना अनुसंधान संस्थान लखनऊ

नई दिल्ली स्थित भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
भारतीय कृषि अनुसंधान परिषद का है एक अग्रणी संस्थान।
यह संस्थान कृषि सांख्यिकी, कंप्यूटर अनुप्रयोग एवं जैव सूचना
विज्ञान में अनुसंधान, शिक्षण के साथ ही प्रशिक्षण करता प्रदान।

वर्ष 1930 में इंपीरियल कृषि अनुसंधान परिषद में एक छोटे से
सांख्यिकी अनुभाग के रूप में की गई थी संस्थान की स्थापना।
इस शोध संस्थान को वर्तमान में राष्ट्रीय तथा अंतर्राष्ट्रीय स्तर
पर कृषि सांख्यिकी में अग्रणी संस्थान के रूप में है जाता जाना।।

हमारे देश में कृषि सांख्यिकी और सूचना विज्ञान में नवीन शोध
करने के लिए यह शोध संस्थान सदा ही रहा है बहुत जिम्मेदार।
कृषि सांख्यिकी तथा सूचना विज्ञान में उच्चतम शिक्षा तथा
प्रशिक्षण प्रदान करने हेतु सम्पूर्ण भारत मानता है इसका आभार।।

संस्थान भारत में कृषि सांख्यिकी व सूचना विज्ञान में प्रशिक्षित
मानव संसाधन विकसित करने का प्रदान करता है यह योग।
नए उभरते क्षेत्रों में गुणवत्ता में सुधार तथा चुनौतियों का सामना
करने हेतु किया गया कृषि सांख्यिकी शोध व शिक्षा का उपयोग।।

भारतीय कृषि सांख्यिकी अनुसंधान संस्थान को देश की राष्ट्रीय
कृषि सांख्यिकी प्रणाली में भी प्राप्त है अत्यंत गौरवपूर्ण स्थान।
इस शोध संस्थान ने राष्ट्रीय नीतियों को प्रभावित करने के लिए
एनएएसएस को सुदृढ़ करने हेतु किए कई महत्वपूर्ण योगदान।।

संस्थान द्वारा कृषि सांख्यिकी तथा सूचना विज्ञान में देश में
एनएआरईएस को किया गया उत्कृष्ट मानव संसाधन प्रदान।
कृषि सांख्यिकी, कंप्यूटर अनुप्रयोग तथा जैवसूचना विज्ञान में
स्नातकोत्तर शिक्षण संचालित कर दिया एक महान योगदान।।

संस्थान का अधिदेश कृषि शोध हेतु सांख्यिकी, कंप्यूटर अनुप्रयोग
तथा जैव सूचना विज्ञान में अनुसंधान, शिक्षा एवं प्रशिक्षण देना।
राष्ट्रीय कृषि अनुसंधान तथा शिक्षा प्रणाली/राष्ट्रीय कृषि सांख्यिकी
प्रणाली को परामर्श सेवाएं व कम्प्यूटेशनल समाधान प्रदान करना।।

भारत में कृषि अनुसंधान की गुणवत्ता को और भी समृद्ध करने हेतु
कृषि सांख्यिकी एवं सूचना विज्ञान का विजन रखता यह संस्थान।
कृषि सांख्यिकी तथा सूचना विज्ञान में अनुसंधान, शिक्षा के साथ
प्रशिक्षण का संचालन करने का मिशन रखता है यह खास संस्थान।।

वर्ष 2022 में संस्थान ने फसलों की कृत्रिम बुद्धिमत्ता आधारित रोग
पहचान प्रणाली (एआई-डीआईएससी) का किया गया है विकास।
किसान कॉल सेंटर हाइपरटेक्स्ट यूजर-इंटरफेस सहित ऐतिहासिक
रूप से एकत्रित ज्ञान-आधारित प्रणाली (केसीसी चक्षु) का विकास।।

वर्ष 2021 में अट्ठारह कृषि विश्वविद्यालयों में वर्चुअल क्लासरूम
की स्थापना व कृषि-दीक्षा वेब शिक्षा चौनल का किया गया आरंभ।
किसान-सारथी- कृषि-सूचना संसाधनों की प्रणाली ऑटो-ट्रांसमिशन
और प्रौद्योगिकी हब इंटरफेस का संस्थान द्वारा किया गया आरंभ।।

किसान 2.0 (कृषि ऐप नेविगेशन के लिए कृषि एकीकृत समाधान)
ऐप के विकास हेतु जिम्मेदार कृषि सांख्यिकी अनुसंधान संस्थान।
प्रमुख पशुधन उत्पादों हेतु एकीकृत नमूना सर्वेक्षण समाधान हेतु
ईएलआईएसएस तथा मोबाइल ऐप विकसित किया यही संस्थान।।

आईसीएआर रिसर्च डेटा मैनेजमेंट हेतु भारत सरकार द्वारा ओपन डेटा
चौम्पियनशिप श्रेणी में गोल्ड आइकन अवार्ड का इसे मिला सम्मान।
'नवीन भारत' की डिजिटल कृषि की दिशा में 'पहल कृषि-मेघ' द्वारा
आगे बढ्ने की शुरूआत करने हेतु जाना जाता यही विशिष्ट संस्थान।।

शिक्षा पोर्टल - भाकृअनुप विकसित और लॉन्च करने वाला तथा
मोबाइल ऐप्स के विकास का आरंभ करने वाला भी यही संस्थान।
बागवानी फसलों, पशुधन व मछली व इसके उत्पादों के उत्पादन
उपरांत होने वाली क्षति का सटीक अनुमान भी लगाता संस्थान।।

कृषि विज्ञान केंद्र ज्ञान नेटवर्क और मोबाइल एप्लीकेशन विकसित
और लॉन्च करने के लिए जाना जाता यही कृषि सांख्यिकी संस्थान।
विभिन्न स्थितियों के लिए फसल क्षेत्र तथा उपज के आकलन हेतु
नमूनाकरण पद्धति विकसित करने वाला भी यही विशिष्ट संस्थान।।

इसी संस्थान ने वेबसर्वर और मोबाइल ऐप तथा वीआईएसटीए
(ट्रिटिकम एस्टिवम की किस्म पहचान प्रणाली) विकसित की।
संस्थान ने वैज्ञानिकगणों की कैडर स्ट्रेन्थ तथा स्थानांतरण के
प्रबंधन हेतु कार्मिक प्रबंधन प्रणाली विकसित कर है लागू की।।

मक्का एग्रीदक्ष व बीज मसालों पर विशेषज्ञ प्रणाली व भारतीय
एनएआरएस सांख्यिकीय कंप्यूटिंग पोर्टल तक हैं किए प्रदान।
तथा आई.ए.आर.आई., नई दिल्ली के पीजी स्कूल के सहयोग से
एम.एससी/पीएचडी डिग्री पाठ्यक्रम का संचालन करता संस्थान।।

हिम्मत और उम्मीद

प्रकाश कुमार, वैज्ञानिक

हमने जिंदगी के कई रंग देखे हैं, कभी धूप में चले, तो कभी छांव में भी बैठे हैं,
बारिशों के साथ भीगते हुए मौसमों के बदलते रंग देखे हैं।
जब-जब हालात बदले, लोगों के बदलते ढंग देखे हैं।
वो भी दिन थे जब हम मायूस हो जाया करते थे,
अपने दर्द और मायूसी के गीत दूसरों को सुनाया करते थे।
और कभी कभार तो जज्बात मैं आकर आँसू भी बहाया करते थे
और लोग अक्सर हमारे आसुओं को देखकर हमारी हँसी उड़ाया करते थे ।
फिर एक दिन जिंदगी ने नया रास्ता दिखाया,
हमने अपने दु:खों को जाहिर करना ही छोड़ दिया।
अब हम दूसरों की जिंदगी में उम्मीद के बीज बोते हैं,
और खुद को कभी अगर रोना भी पड़े तो हस्ते हस्ते रो लेते है ।
ये जिंदगी के अँधेरे भी अजीब है जो प्रभाव रवि का भी कमा देती है
तभी तो ये जिंदगी एक स्टेटिस्टीशियन से कवि भी बना देती है ।

सैनिक का परिवार

राखी सोनी, सहायक

तिरंगे में लिपटा वो बड़ी शान से घर आया था,
देश भक्ति, बहादुरी, राजकीय सम्मान सब उसने पाया था ।
पर अपने पीछे वो, सुनी आंखें सिसकती रातें छोड़ गया,
एक माँ की आन के लिए, दूसरी माँ से सारे रिश्ते तोड़ गया ।।
कहते है सरहद पर, बड़ी बहादुरी से उसने दुसमानों को मारा था,
खुद छलनी हो गया, पर वो हिम्मत नहीं हारा था
पर उसके छोटे छोटे बच्चे, इस बात को कहाँ समझते हैं
वो तो बस जानते है पापा आएंगे, हमेशा की तरह हँसते हुए
पत्नी की आंखो में, आसूं और बेबसी है
कैसे उनके बिना, पूरी कर पाएंगी जिम्मेदारियाँ, जो सामने खड़ी है
माँ बिलख बिलख कर रो रही है एक ही बेटा था मेरा
बड़ी मन्नतों से मांगा था मैंने, आखिर क्यों चला गया बेटा मेरा
बाप खामोश है परेशान है उदास है
कैसे बताए किसी को कि आखिर क्या बात है
उसे पता है देश आखिर इस बलिदान पर, उसे क्या देगा
राजकीय सम्मान के साथ, उसके हाथों में, मेडल थमा देगा
कुछ दिन अखबार, टीवी कि सुर्खियां बनेगा
फिर उसके बाद, कोई भी उसके परिवार की, सुध न लेगा
उसकी पेंशन कहाँ इतना बड़ा परिवार चलेगा
अपनी तो छोड़ो, उसके छोटे छोटे बच्चों का पेट, कैसे भरेगा
इतनी उम्र में अब नौकरी मिलना भी आसान नहीं
और मजदूरी करले, इन बूढी हड्डियों में उतनी जान नहीं
सोच रहा है इकलौते बेटे, को काश,, उसने रोक लिया होता
देश को तो बचा लेते दूसरे भी, पर अपना बेटा तो बचा लिया होता
पर अगले ही पल अपनी ही सोच पर वो खुद को ही शर्मिंदा पाता है
झूके हुए कंधें, कांपते हुए हाथों से सहारा ले, वो बेटे की अर्थी में शामिल हो जाता है
एक देश भक्त बेटे, को बाप होने के गर्व से उसकी आँखों में चमक आ जाती है
पर, फिर कैसे जिएगा परिवार, उसके बिना, सोच वो चमक फिर से कहीं खो जाती है

“शीशे की एक दीवार”

सौम्या शर्मा, वैज्ञानिक

उसकी जिंदगी की हकीकत और सपनों के संसार के बीच थी एक
शीशे की दीवार।

इस तरफ तपिश थी, बारिश थी, और उस तरफ बाग और बहार।

वो एकटक देखती रहती और समझती कि जा पाए उस पार

पर... करना क्या था??

शायद एक पत्थर मारकर तोड़नी थी वो दीवार, पर अपने मन की
जंजीरों ने उसे बांधा था कुछ यूँ कि पत्थर उठाना लगा मुश्किल और
लगा आसान सहना दर्द हजार।

सोचिए....

अगर आप उस जगह हों क्या करोगे आप? तोड़ देंगे वो दीवार??

या फिर जीवन में हम सभी कभी ना कभी उसी जगह होते हैं, तो
हम क्या करते हैं?

वो जो लगता हैं आसान और जिंदगी के ढर्रों पर चलते रहते हैं, या
हम हिम्मत करते हैं, करते हैं वो जो लगता है मुश्किल

करते हैं अपने सपनों को साकार

और तोड़ देते हैं वो शीशे की दीवार??

सांख्यिकी विज्ञान से कृषि विकास

प्रकाश कुमार
वैज्ञानिक

सांख्यिकी विज्ञान की है ये अनोखी राह,
संख्याओं का चलता जहां पर चाह।
आंकड़ों की जुबान, आंकड़ों का जाल,
सत्य की खोज इन्ही से होती, इसमें हमें न कोई सक और सवाल ।
माध्य और माध्यिका का होता खेल,
हर डेटा सेट में छुपा एक मेल।
मानक विचलन का होता बखान,
भटकावों का पाता जो मान।
प्रत्यावर्तन रेखा के संग,
बिखरे बिंदुओं का करती रंग।
संभावना की मधुर छवि,
भविष्य का पथ दिखा रही।
स्तंभ चित्र और पाई चार्ट,
आंकड़ों की बगिया में नया सार।
विश्वास अंतराल की शक्ति,
सच्चाई की पक्की धरती।
काई-स्क्वायर और एनोवा,
आंकड़ों के राज खोलता।
नमूनों की कला में है जादू,
नई राहों का मिल जाता है सुराग।

पी–मूल्य और टी–परीक्षण की गहराई,
अदृश्य बलों की मिलती सच्चाई।
सहसंबंध की मीठी बात,
रिश्तों की जड़ों का करती है आघात।
इस तथ्य और संख्याओं की दुनिया में,
मानव हृदय बसा हर आंकड़े में।
विज्ञान में है सत्य और कला,
हर संख्या में छुपा है जीवन का हाला।
तो आइए नृत्य करें डेटा की धुन पर,
हर संख्या में ढूंढ़ें अपना घर।
सांख्यिकी के इस सुंदर जहान में,
मानवता की दिशा खोजें हर आंकड़े के संग।

लहरें

मोहम्म्द हारून, वैज्ञानिक

भाग-1: नदी

एक लहर उठी एक नदी में,
आतुर सागर से जा मिलने को ।
एक बैचनी थी उसमें,
व्यग्र अपनी व्यथा सुनाने को ।।
शांत सतह से उसने खुद का यूं उदगार किया ।
तरुवर से लड़ते, पाहन से गुजराते,
अपने पथ पर प्रस्थान किया ।।
नहीं रुकी वो चट्टानों से टकराकर,
न रोक सके उसको जंगलों के पत्थर ।
होकर निडर जब हर डगर हर मुश्किल को पार किया,
हर कण-कण, हर प्राणी, हर पग ने उसका सम्मान किया ।।
सागर से मिलने को बेकल, एक बांध पर उसकी हार हुई ,
टूटे सपने टूटे अरमाँ और उस तरह वो निष्प्राण हुई।

भाग-2 नारी

एक लहर उठी मन में एक नारी के,
आतुर दुनिया को जा बदलने की ।
कुछ कहने की कुछ करने की,
कुछ जीने की कुछ लड़ने की ।।
नहीं रुकी वो हल-प्रपंचों से,
ना रोक सका उसको कोई लोभ ।
बनकर मशाल निकली लड़ने भ्रष्टाचारियों से
लेकर व्यधित चित, और हृदय में क्षोभ ।
मिल गए सारे रावण, फिर हमने तो ना एक चित्कार सुनी,
एक लहर उठी, एक लहर उठी, सबने बस ये बात सुनी,

भाग-3 सार

लहरें उड़ती हैं नदियों में,
लहरें उड़ती हैं अन्तर्मन में ।
लहरें उठती है वाद से, लहरें उठती हैं समाज से,
हर लहरों का तुम सम्मान करो ।।
इनको आगे बढ़ने दो साहस से,
इनको जाकर मिलने दो सागर से ।
मत इनके पथ में तुम व्यवधान बनों,
मत इनके पथ में तुम व्यवधान बनो ।।
एक लहर उठी, एक लहर उठी,
जाकर सागर से मिलने को, आतुर दुनिया दो बदलने को ।

हिन्दी कार्यशाला 2024

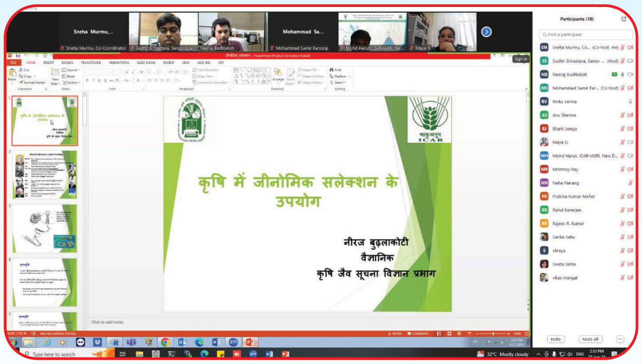
"कृषि शिक्षा में डिजिटल पहल"

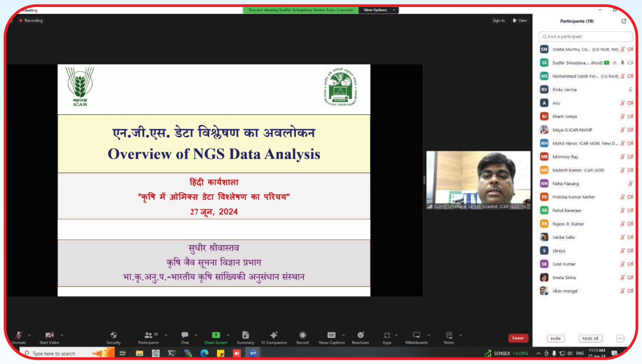
(समन्वयक डॉ. सपना निगम, डॉ. संचिता नाहा एवं डॉ. मधु)

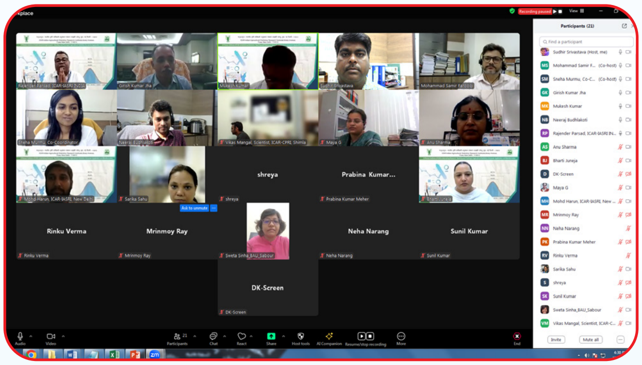
भाकृअनुप-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान, नई दिल्ली

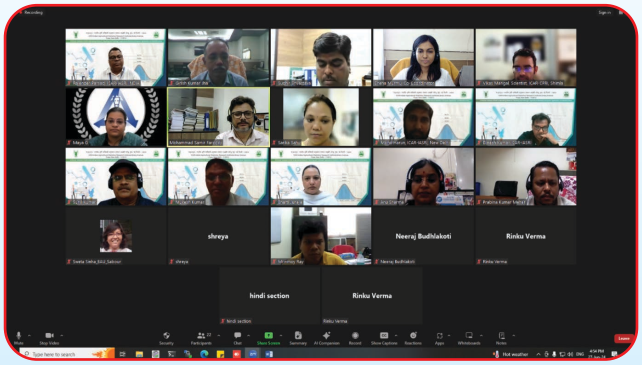
06-08 मार्च, 2024 (ऑफ-लाइन)

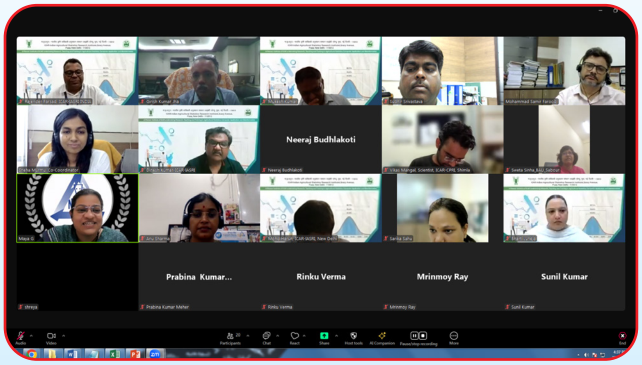
"कृषि में ओमिक्स डेटा विश्लेषण का परिचय"
(समन्वयक: डॉ. सुधीर श्रीवास्तव, डॉ. मोहम्मद समीर फारुकी एवं डॉ. स्नेहा मुर्मू)
27 जून 2024 (ऑन-लाइन) हिन्दी कार्यशाला











"परिषद में ई-गवर्नेंस का अनुप्रयोग"

(समन्वयक: डॉ. एस.बी. लाल, डॉ. मुकेश कुमार एवं श्री संजीव कुमार)

11 सितम्बर, 2024

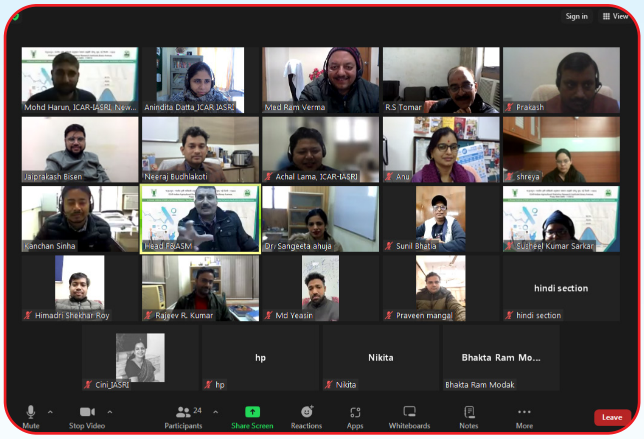
(ऑफ-लाइन) हिन्दी कार्यशाला

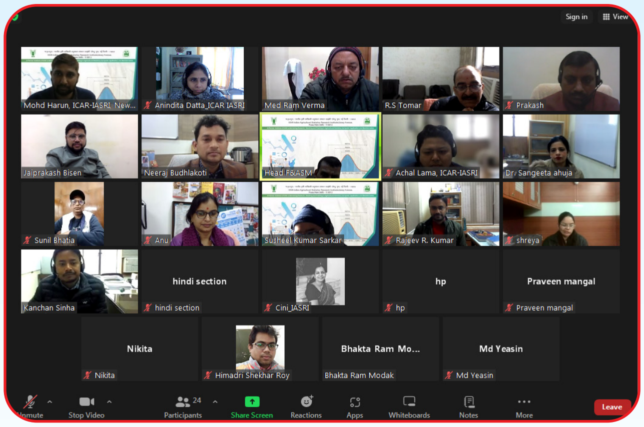
“परीक्षण अभिकल्पनाएं एवं विश्लेषण”

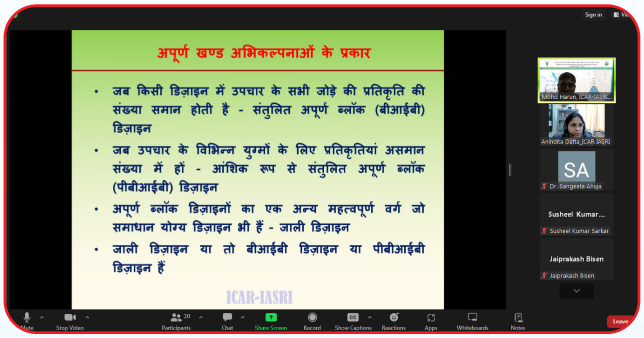
(समन्वयक: डॉ. मोहम्मद हारून एवं डॉ. अनिंदीता दत्ता)

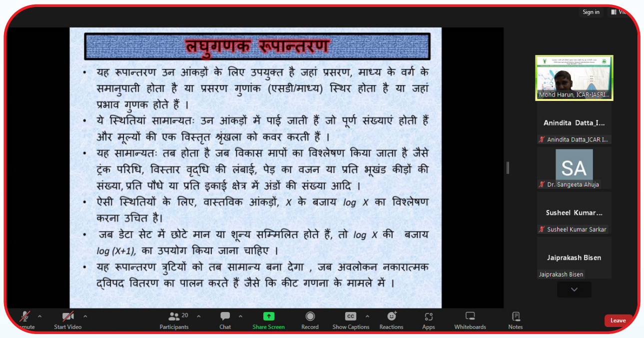
19 दिसम्बर, 2024

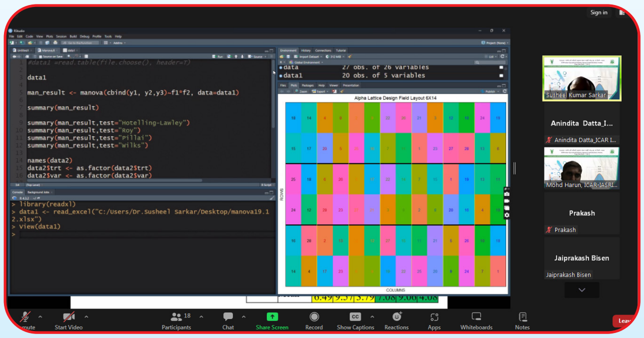
(ऑन-लाइन) हिन्दी कार्यशाला

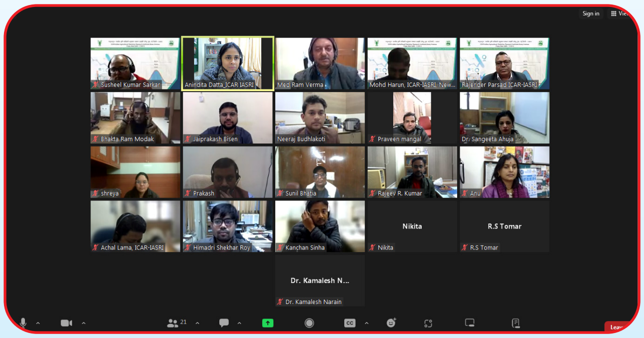












“भारतीय संविधान”

(समन्वयक: हिन्दी एकक द्वारा आयोजित की गई, इसमें व्याख्यान श्री राम दयाल शर्मा, संयुक्त निदेशक (राजभाषा) को व्याख्यान दाता के लिए आमंत्रित किया गया)

24 दिसम्बर, 2024

(ऑफ-लाइन) हिन्दी कार्यशाला

हिन्दी पखवाड़ा - 2024

संस्थान में हिन्दी पखवाड़े का आयोजन दिनांक 14-30 सितम्बर, 2024 के दौरान किया गया। हिन्दी पखवाड़े के अंतर्गत काव्यपाठ, अंताक्षरी, प्रश्नमंच, हिन्दी शोध-पत्र प्रस्तुतिकरण प्रतियोगिता, चल-शील्ड प्रतियोगिता का आयोजन किया गया। पखवाड़े के अंतर्गत डॉ. दरोगा सिंह स्मृति व्याख्यान श्री राकेश कुमार त्यागी, पूर्व महानिदेशक, राष्ट्रीय प्रतिदर्श सर्वेक्षण कार्यालय (एनएसएसओ) द्वारा किया गया। पखवाड़े से संबंधित कुछ छायाचित्र निम्नवत है:

कुछ यादगार लम्हें.....

ऑर्गेनोग्राम

अनुसंधान सलाहकार समिति ↔ निदेशक ↔ संस्थान प्रबंधन समिति

प्रभागाध्यक्ष
- परीक्षण अभिकल्पना
- सांख्यिकी आनुवंशिकी
- पूर्वानुमान एवं कृषि प्रणाली मॉडलिंग
- प्रतिदर्श सर्वेक्षण
- संगणक अनुप्रयोग
- कृषि जैव-सूचना
- आईटी एकक

अधिकारी प्रभारी
- पुस्तकालय एवं प्रलेखन
- उन्नत संकाय प्रशिक्षण केंद्र
- संस्थान प्रौद्योगिकी प्रबंधन एकक
- प्राथमिकीकरण, निगरानी एवं मूल्यांकन प्रकोष्ठ
- कृषिज्ञान प्रबंधन प्रकोष्ठ
- एनएएसएम

प्रोफेसर (कृषि सांख्यिकी) प्रोफेसर (संगणक अनप्रयोग) प्रोफेसर (जैवसूचना विज्ञान) वार्डन
- प्रशिक्षण प्रशासन प्रकोष्ठ

सतर्कता अधिकारी
- ओएमवी एकक

मुख्य प्रशासनिक अधिकारी
- प्रशासन I अनुभाग
- प्रशासन II अनुभाग
- रोकड़ अनुभाग
- केंद्रीय खरीद एवं उपकरण रखरखाव
- पावती एवं प्रेषण अनुभाग
- कार्य एवं रखरखाव अनुभाग
- रखरखाव अनुभाग
- भण्डार
- हिन्दी एकक

मुख्य वित्त एवं लेखाधिकारी
- लेखा परीक्षा अनुभाग
- लेखा अनुभाग

NOTES

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली - 110012