

Land cover classification using machine-learning method and vegetation indices

Jargaldalai Enkhtuya^{1,*}, Amarsaikhan Damdinsuren¹, Bilguun Ulziibat¹,

Munkh-Erdene Altangerel²

¹*Division of Geographic Information System and Remote Sensing, Institute of Geography and Geoecology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia*

²*Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia*

*Corresponding author email: Jargaldalaie@mas.ac.mn

Received: 31 October 2022 /Accepted: 30 November 2022 /Published online: 29 December 2022

ABSTRACT

Machine-learning offers the potential for effective and efficient classification of remotely sensed imagery. The strength of the machine-learning include the capacity to handle data of high dimensionality and to map classes with very complex characteristics. This study aimed to apply the machine-learning method for an improved land cover classification. For this purpose, multispectral Sentinel-2 data along with 3 vegetation indices (NDVI - normalized difference vegetation index, TSAVI-transformed soil adjusted vegetation index, and MTVI-modified triangular vegetation index) were acquired in July 2021 around Zuunburen soum, Selenge province, and analysed to classify and compare land cover using random forest (RF) and support vector machine (SVM) techniques. For the actual classifications, three vegetation indices, NDVI, TSAVI, and MTVI, which were derived from the visible and infrared bands of Sentinel-2a, were used. As the result, the land cover was classified into 5 classes including forest, cropland, grass, soil, and water, and the overall accuracy of the machine-learning method was above 87%.

Keywords: *Land cover classification, Machine learning, RF, SVM*

Газрын бүрхэвчийг машин сургалттай ангиллын арга болон ургамлын индексүүд ашиглан ангилах нь

Жаргалдалай Энхтуяа^{1,*}, Амарсайхан Дамдинсүрэн¹, Билгүүн Өлзийбат¹,
Мөнх-Эрдэнэ Алтангэрэл²

¹Газарзүйн Мэдээллийн Систем, Зайнаас Тандан Судлалын Салбар, Газарзүй, Геоэкологийн Хүрээлэн,
Шинжлэх Ухааны Академи, Улаанбаатар, Монгол

²Математик, Тоон Технологийн Хүрээлэн, Шинжлэх Ухааны Академи, Улаанбаатар, Монгол

*Холбоо барих зохиогчийн цахим хаяг: jargaldalaie@mas.ac.mn

Хүлээн авсан: 2022 оны 10 сарын 31 өдөр /Зөвшөөрөгдсөн: 2022 оны 11 сарын 30 өдөр /Нийтлэгдсэн: 2022
оны 12 сарын 29 өдөр

ХУРААНГУЙ

Машин сургалттай ангиллын арга нь зайнаас тандсан мэдээг илүү үр дүнтэй, нарийвчлалтай сайтай ангилах боломжийг олгодог. Уг ангиллын давуу тал нь их хэмжээний өгөгдөлтэй ажиллах, маш нарийн төвөгтэй шинж чанар бүхий ангиудыг ангилахад оршино. Энэхүү судалгаа нь Сэлэнгэ аймгийн Зүүнбүрэн сумын нутгийн Sentinel-2 хиймэл дагуулын олон бүсчлэлийн мэдээнд санамсаргүй форест, тулах векторын зэрэг машин сургалттай ангиллын аргуудыг ашиглан, газрын бүрхэвчийн ангилал хийж, харьцуулсан дүгнэлт хийх зорилготой бөгөөд ангилалд NDVI, TSAVI, MTVI гэсэн 3 төрлийн ургамлын индексийг ашигласан болно. Дүн шинжилгээнд, газрын бүрхэвчийг ус, ой, тариалан, ногоон ургамал, хөрс гэсэн 5 ангид хуваан ангилсан ба эцсийн үр дүнгээс харахад машин сургалттай аргуудын ерөнхий нарийвчлал 87%-иас дээш байлаа.

Түлхүүр үгс: Газрын бүрхэвчийн ангилал, Машин сургалт, Санамсаргүй форест, Тулах векторын арга

1. ОРШИЛ

Орчин үед, зайнаас тандах судлал (ЗТС) нь орон зайн төлөвлөлтийн хүчирхэг технологи хэмээн нэрлэгддэг газарзүйн мэдээллийн системийн мэдээллийн санг баяжуулах нэгэн чухал эх сурвалж болж байгаа ба өндөр хөгжилтэй орнууд агаар-сансрын мэдээг орон зайн мэдээллийн системийн сэдэвчилсэн давхаргуудыг шинэчлэн, баяжуулахад өргөнөөр ашиглаж байгаа юм. Үүнээс гадна, дэлхийн улс орнууд тандан судалгааны мэдээг төрөл бүрийн автомат тайллын болон өндөр түвшний аргуудыг ашиглан боловсруулж, гарган авсан сэдэвчилсэн үр дүнг тухайн зорилгоос хамаарч төрөл бүрийн салбарт ашиглаж байна [1].

XXI зууны эхэн үеэс, ЗТС-ын салбарт төрөл бүрийн сэдэвчилсэн мэдээлэл гарган авах шинэ аргазүй боловсруулах, олон янзын тайллын аргуудыг хөгжүүлэх, тэдгээрийг байгалийн ялгаатай бүс нутгуудад турших зэрэг чиглэлийн судалгаа эрчимтэй хөгжиж байна [2]. Зайнаас тандсан мэдээнд төрөл бүрийн өндөр түвшний аргаар тайлал хийж сэдэвчилсэн мэдээлэл гарган авах нь уламжлалт аргуудтай харьцуулахад цаг хугацаа, хүн хүч, хөрөнгө зардлыг хэмнэсэн, олон давуу талтай хэдий ч, уг аргууд нь

биесийн физик шинж чанар, магадлалын онол, хүчирхэг статистик загварт үндэслэдэг тул тодорхой бус байдал, хэт олон хэмжээс зэрэг бэрхшээлүүдтэй тулгардаг [3].

Сүүлийн үед эдгээр асуудлуудыг компьютерын тусламжтайгаар хүний сэтгэлгээний зарчимд тулгуурлан шийдвэрлэх төрөл бүрийн оролдлого ихээхэн хийгдэж байгаа юм [4] Өөрөөр хэлбэл, техник, технологийн хөгжилтэй уялдан олон хэмжээст, шугаман бус аргын хэрэглээ нэмэгдсэнээр ЗТС-ын салбарт томоохон ахиц дэвшил гарч, дараагийн шинэ чиглэл болох дүрс мэдээг боловсруулах хиймэл оюуны технологиуд, ялангуяа машин сургалтын аргуудын хэрэглээ эрчимтэй нэмэгдэж байна.

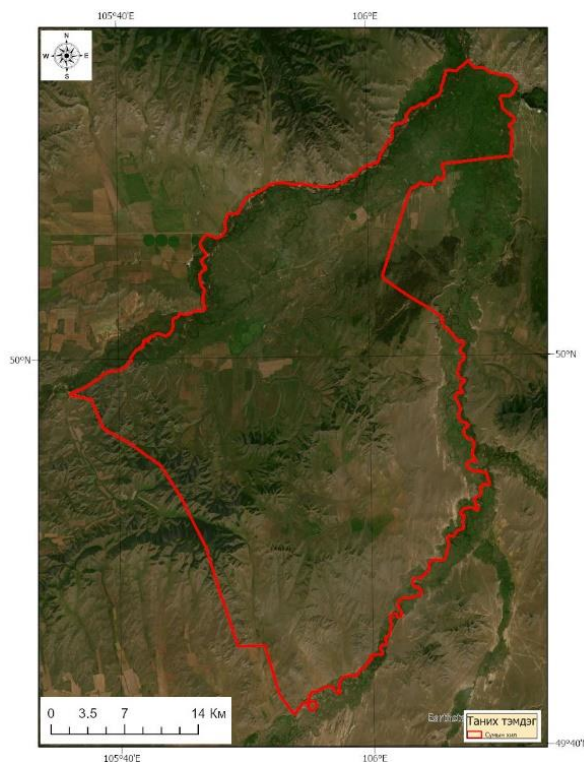
Хиймэл оюун нь голдуу тодорхой мэдлэг дээр суурилан, ухаалаг шийдвэр гаргах чадвартай байдаг бол машин сургалт нь хиймэл оюуны дэд хэсэг бөгөөд өгөгдсөн өгөгдлийг ашиглан тодорхой даалгаврыг математик, статистикийн аргаар гүйцэтгэх чадвартай байдаг [5]. Машин сургалтын загварыг өгөгдсөн даалгавраас нь хамааруулан сургалттай (хяналттай), сургалтгүй (хяналтгүй), хүч нэмэгдүүлсэн сургалт гэж ангилах [1] ба ямар аргыг ашиглах нь тухайн даалгаврын нөхцөл, байдлаас ихээхэн шалтгаална.

Энэхүү судалгаанд, газрын бүрхэвчийг ангилахдаа Европын сансрын агентлагийн хөөргөсөн Sentinel-2 хиймэл дагуулын мэдээ ашиглан ургамлын индексүүдийг тооцоолон, улмаар гарсан үр дүнг машин сургалттай ангиллын аргуудаар ангилж, үр дүнгүүдийг харьцуулсан бөгөөд ангилалд санамсаргүй форестийн болон тулах векторын аргуудыг ашигласан болно.

2. СУДАЛГААНЫ ТАЛБАЙ, МАТЕРИАЛ

2.1 Судалгааны талбай

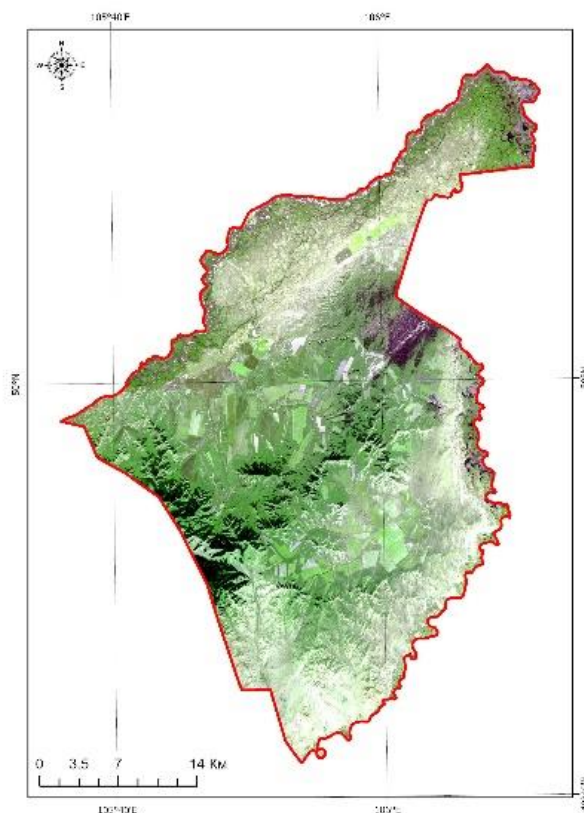
Судалгааны талбай болгон Монгол орны хойд хэсэгт Сэлэнгэ аймгийн Зүүнбүрэн сум орчмын газар нутгийг сонгон авсан болно. Уг нутаг нь далайн төвшнээс дээш 1000-1500 м өргөгдсөн бөгөөд физик, газарзүйн хувьд Хэнтийн захын уулсын мужид, геоморфологийн хувьд Орхон-Сэлэнгийн бэсрэг уулсын мужид багтдаг. Сумын нутаг дэвсгэр нь уул, гүвээ толгод бүхий тал хээр ба мал аж ахуй болон газар тариалан эрхлэхэд тохиромжтой. Зураг 1-д судалгааны талбайн байршилыг харуулав.



Зураг 1. Судалгааны талбайн байршил.

2.2 Сансрын мэдээ

Энэхүү судалгаанд 2021 оны 7-р сарын 12-ны өдрийн Sentinel-2A хиймэл дагуулын олон бүсчлэлийн мэдээг ашигласан болно. Тус мэдээ нь 13 сувагтай [6] бөгөөд тэдгээрийн үзүүлэлтийг Хүснэгт 1-д үзүүлэв. Судалгааны талбайн хиймэл дагуулын мэдээг Зураг 2-д үзүүлэв.



Зураг 2. Судалгааны талбайн Sentinel-2A хиймэл дагуулын мэдээ.

Хүснэгт 1. Sentinel-2A дагуулын мэдээний шинж чанар үзүүлэлтүүд.

Сувгууд	Долгионы урт (μм)	Орон зайн шийд (м)	Сүвгийн өргөн(нм)
Суваг 1	0.443	60	27/45 (2A/2B)
Суваг 2	0.490	10	98
Суваг 3	0.560	10	45/46 (2A/2B)
Суваг 4	0.665	10	38/39 (2A/2B)
Суваг 5	0.705	20	19/20 (2A/2B)
Суваг 6	0.740	20	18
Суваг 7	0.783	20	28
Суваг 8	0.842	10	115
Суваг 8A	0.865	20	20
Суваг 9	0.945	60	20
Суваг 10	1.375	60	20
Суваг 11	1.610	20	90
Суваг 12	2.190	20	180

3. СУДАЛГААНЫ АРГАЗҮЙ

Дүн шинжилгээнд, санамсаргүй форест, тулах векторын ангиллын аргуудыг ашиглаж үр дүнгүүдийг харьцуулан шинжилсэн болно.

Санамсаргүй форестийн арга нь орчин үеийн өндөр түвшний машин сургалтын арга бөгөөд ангиллын үндсэн нэгж нь шийдвэрийн мод байдаг. Уг мод нь анхдагч өгөгдлийн шинж чанарууд буюу ижил төстэй байдал дээр тулгуурлан нийт мэдээний шаталсан хэлбэрийн бүтцийг үүсгэдэг. Энэхүү аргад мод тус бүрийн зангилаа цэг дээрх таамагласан үр дүнгүүдийн хамгийн их тохиолдлыг тодорхойлон тэдгээрийг нэгтгэх замаар эцсийн шийдвэрийг гаргана. Санамсаргүй форестийн арга нь нарийвчлал өндөртэй бөгөөд их хэмжээний мэдээтэй ажиллахад илүү тохиромжтой байдаг [7].

Тулах векторын арга нь төрөл бүрийн тоон мэдээг ашиглан ангилал хийж регрессийн дүн шинжилгээ хийдэг машин сургалтын нэг хэлбэр бөгөөд сургалттай ангиллын төрөлд хамаардаг. Тулах векторыг статистик сургалтын онолоос гаралтай ангилал гэж үздэг бөгөөд уг арга нь анги хоорондын зөрүүг хамгийн их байлгах гадаргууг ашиглан ангиудыг ялгадаг.

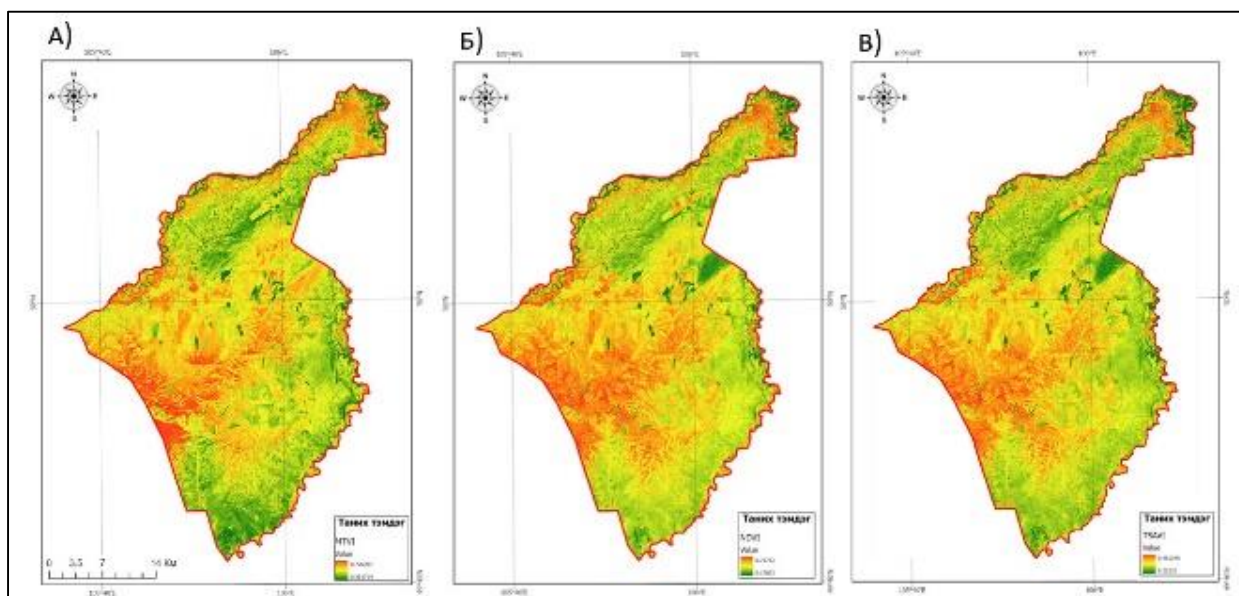
Гадаргууг “оптимал хайпер хавтгай” гэх бөгөөд түүнтэй ойролцоо өгөгдлийн цэгүүдийг тулах буюу туслах векторууд гэж нэрлэдэг. Эдгээр тулах векторууд нь ангиллын чухал элементүүд болдог бөгөөд ангилалд түгээмэл хэрэглэгддэг шугаман, полиномын, радиал суурь функцийн болон сигмоид хэлбэрийн гэсэн үндсэн 4 төрөл байдаг. Тандан судлалд тулах векторын аргыг анхандаа хайперспектрийн мэдээг ангилах, объектыг танихад түлхүү ашигладаг байсан бол одоо олон бүсчлэлийн мэдээг ангилахад ч өргөнөөр ашиглаж байна [8].

4. СУДАЛГААНЫ ҮР ДҮН

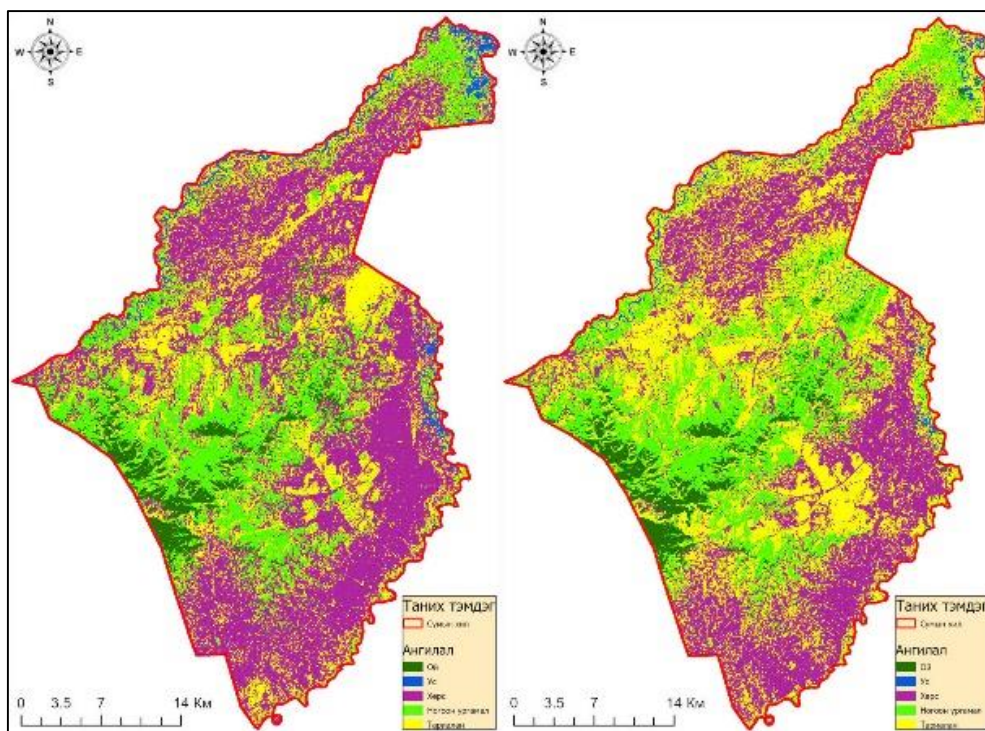
Судалгааны эхэнд, NDVI, TSAVI, MTVI гэсэн 3 индексийг тооцоолов. Тооцоолсон индексүүдийг ашиглан машин сургалттай ангилал хийхдээ 5 ангийг (ус, ой, тариалан, ногоон ургамал, хөрс) тодорхойлон төлөөллийн сигнатуруудыг контекстийн мэдлэгт тулгуурлан сонгож авна. Дараа нь санамсаргүй форест, тулах векторын аргуудыг ашиглан газрын бүрхэвчийг ангиллаа.

Ангиллын үр дүнг тохирлын матриц ашиглан ерөнхий нарийвчлалын аргаар үнэлсэн ба уг арга нь цэвэр пикселийг ангилагдсан пикселүүдтэй харьцуулах замаар нарийвчлалын хувийг илэрхийлдэг [9,10]. Санамсаргүй форестэд 135691, тулах векторын аргад 141918 цэвэр пикселийг тус тус сонгон авсан бөгөөд тохирлын матрицыг ангиллын дүн тус бүрд тооцож ерөнхий нарийвчлалыг тооцоолов. Ангиллын нарийвчлалын үнэлгээг Хүснэгт 2-д үзүүлэв.

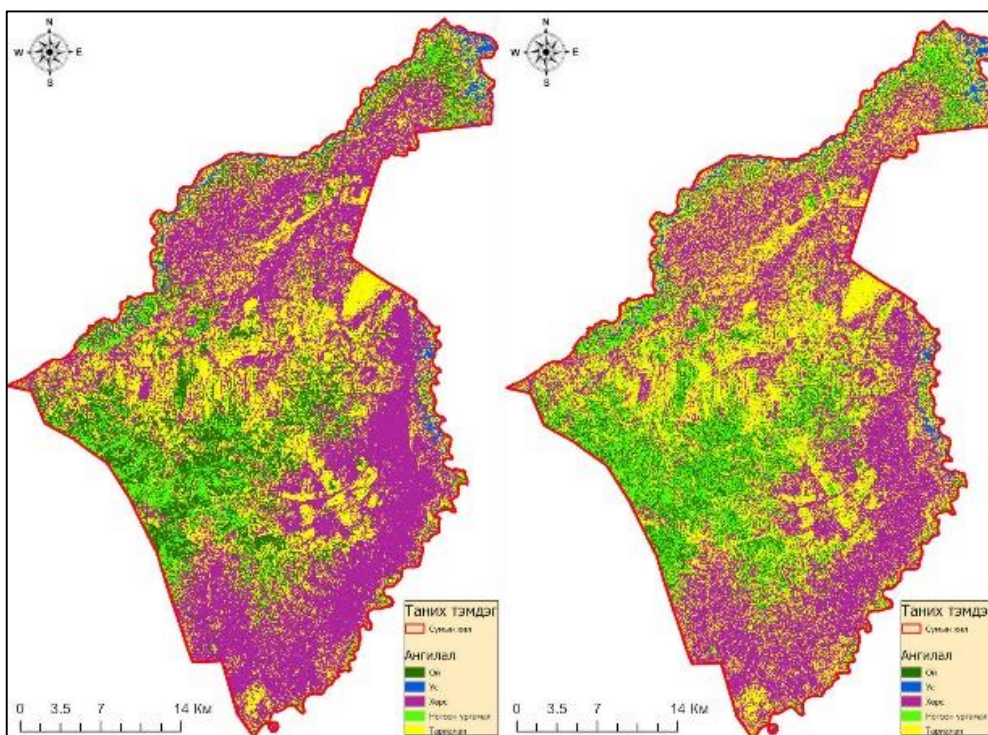
Ангиллын үр дүнгүүд болон нарийвчлалын үнэлгээнээс харахад, машин сургалттай аргын үр дүн уламжлалт ангиллын аргын үр дүнгээс илүү гэдэг нь харагдаж байна. Ангиллын үр дүнгүүдийг Зураг 4,5,6,7-д үзүүлэв.



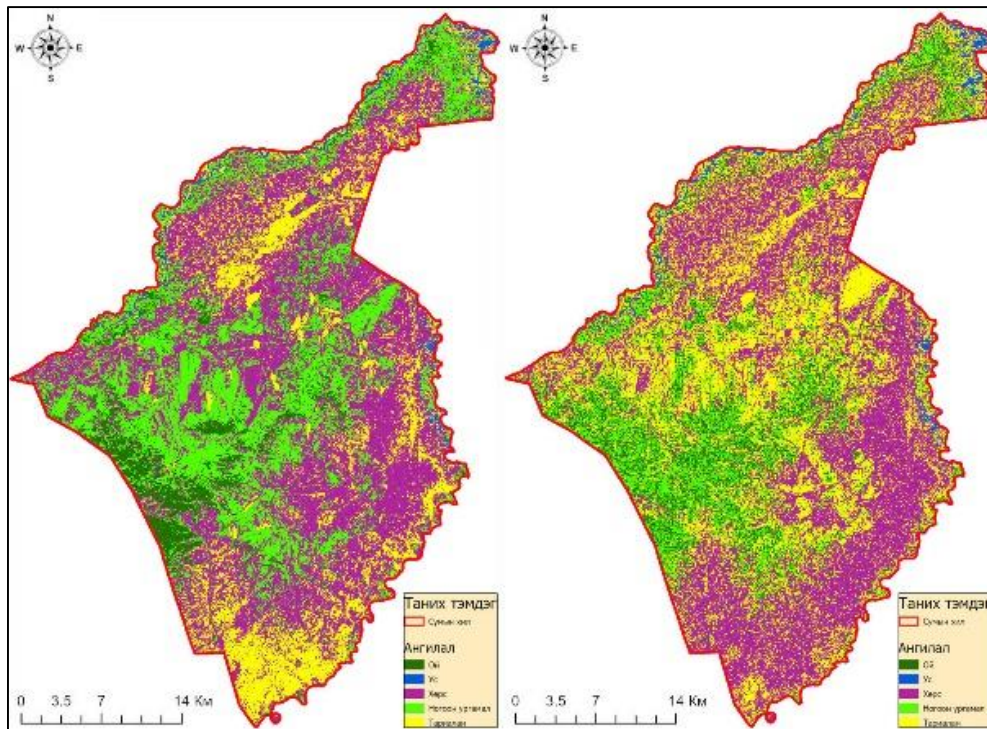
Зураг 3. Тооцоолсон ургамлын индексүүд (MTVI-(a), NDVI-(б), TSAVI-(в)).



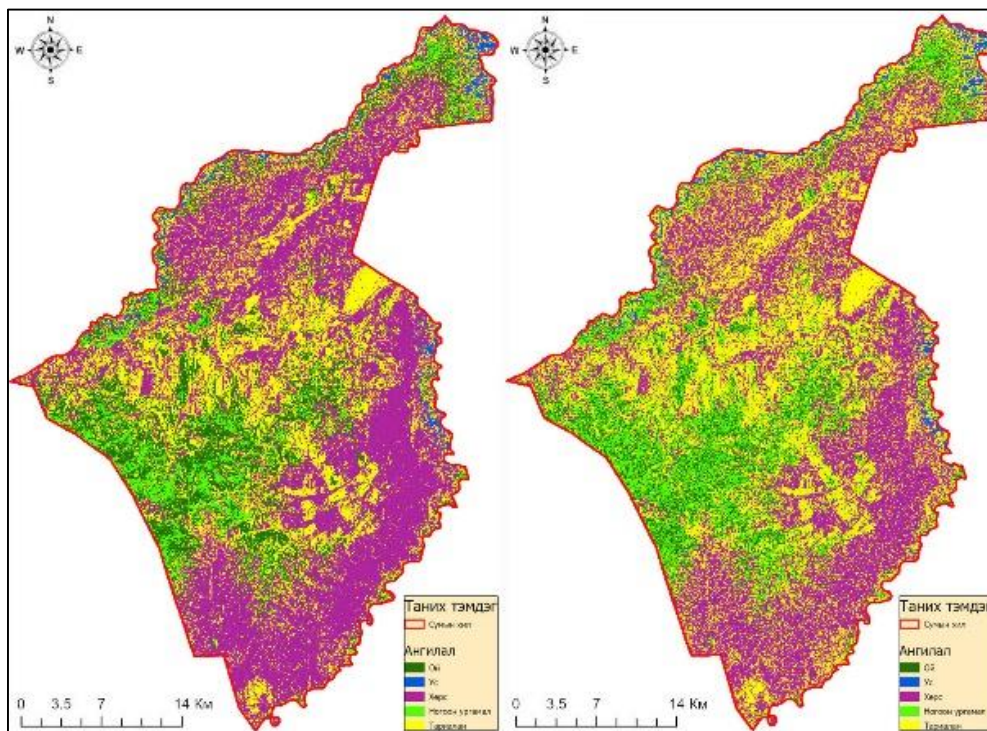
Зураг 4. Тулах векторын арга (а) болон Санамсаргүй форестын(б) ангиллын аргаар Sentinel-1a хиймэл дагуулын УНХ (RGB) мэдээнд ангилал хийсэн дүн.



Зураг 5. Тулах векторын арга (а) болон санамсаргүй форестын(б) ангиллын аргаар NDVI мэдээнд ангилал хийсэн дүн.



Зураг 6. Тулах векторын арга (а) болон санамсаргүй форестын (б) ангиллын аргаар MTVI-ийн мэдээнд ангилал хийсэн дүн.



Зураг 7. Тулах векторын арга (а) болон санамсаргүй форестын (б) ангиллын аргаар TSAVI мэдээнд ангилал хийсэн дүн.

Хүснэгт 2. Ангиллын нарийвчлалын үнэлгээ.

№	Хөрс, ургамлын индексүүд	Тулах векторын арга	Санамсаргүй форестын
1	УНХ	0.942	0.915
2	NDVI	0.925	0.871
3	MTVI	0.958	0.898
4	TSAVI	0.947	0.914

5. ДҮГНЭЛТ

Энэхүү судалгаанд Sentinel-2A хиймэл дагуулын олон сувгийн мэдээг ашиглан машин сургалттай (санамсаргүй форест, тулах вектор) ангиллын аргуудаар газрын бүрхэвчийг ангилан, үр дүнгүүдийг харьцуулж үзсэн болно.

Судалгаанаас харахад, төрөл бүрийн индексүүдийг тооцоолж ангилал хийхэд спектрийн хувьд хоорондоо төстэй ангиуд болох хөрс, тариалан, ногоон ургамал гэсэн ангиудын зааг сайн гарч ангилалын нарийвчлал нэмэгдэж байгаа нь харагдаж байна.

Ангиллын нарийвчлалыг үнэлэхэд, тулах векторын ангиллын аргаар ангилсан хуурмаг дүрс болон индексүүдийн ерөнхий нарийвчлал 0.942, 0.925%, 0.958%, 0.947% байсан бол санамсаргүй форестын аргаар ангилсан хуурмаг дүрс болон 3 индексийн ангиллын нарийвчлал 0.915, 0.871, 0.898, 0.914% байлаа.

Судалгааны үр дүнгээс харахад, орчин үеийн зайнаас тандсан дүрс мэдээг боловсруулах машин сургалттай аргууд нь газрын бүрхэвч, газар ашиглалтын үнэн бодит сэдэвчилсэн мэдээлэл гарган авахад үр дүнтэй гэдэг нь харагдаж байна.

АШИГЛАСАН НОМ, ХЭВЛЭЛ

- [1] А. Мөнх-Эрдэнэ, О. Мөнхдулам, “Зайнаас тандсан дүрс мэдээг боловсруулах хиймэл оюуны аргууд ба тэдгээрийн хэрэглээ”, “Орон зайн мэдээ” сэтгүүл, ГЗБГЗЗГ . 2021.
- [2] Д. Амарсайхан, “Орчин үеийн Газарзүйн мэдээллийн систем, тандан судлалын зарчмууд”, нэгэн сэдэвт бүтээл, хх32,8, ШУА Монгол улс, 2019.
- [3] G. Camps-Valls, Machine learning in remote sensing data processing. IEEE international 463 workshop on machine learning for signal processing, pp 1–6, September. 2009. Available: <https://doi.org/10.1109/MLSP.2009.5306233>
- [4] N. Sisodiya, N. Dube, & P. Thakkar, Next-generation artificial intelligence techniques for satellite data processing. In Artificial Intelligence Techniques for Satellite Image Analysis (pp. 235-254). Springer, Cham. 2020. Available: https://doi.org/10.1007/978-3-030-24178-0_11
- [5] Д. Батсүрэн, Даатгуулагчийн эрсдэлийг машин сургалтаар үнэлэх. Хүмүүнлэгийн Ухааны Их Сургуулийн Бизнесийн Сургууль, Санхүү Эдийн засгийн тэнхим. 2020.
- [6] Sentinel-2 User Handbook, ESA Standard Document, ESA, pp.64. 2015.
- [7] S. Chakraborty, Bayesian Additive Regression Tree for Seemingly Unrelated Regression with Automatic Tree Selection, Handbook of Statistics, Volume 35, Pages 229-251. 2016. Available: <https://doi.org/10.1016/bs.host.2016.07.007>
- [8] А. Мөнх-Эрдэнэ, Сансрын олон эх сурвалжийн мэдээгээр ойн мэдээллийн санг баяжуулах аргазүйн асуудалд. докторын зэрэг горилсон бүтээл, 54 хуудас. 2019.
- [9] P. M. Mather, M. Koh, Computer Processing of Remotely-Sensed Images: an Introduction. Fourth edition (Wiley-Blackwell). 2011. Available: <https://doi.org/10.1002/9780470666517>
- [10] G. Li, D. Lu, E. Moran, and S. Hetrick, Land-cover classification in a moist tropical region of Brazil with Landsat TM imagery, International Journal of Remote Sensing, 32(23), pp.8207–8230. 2011. Available: <https://doi.org/10.1080/01431161.2010.532831>