



Cgrain Value -laitteen validointi vieraiden viljojen tunnistami- seen kauranäytteistä

Ville Alminojä

OPINNÄYTETYÖ
Kesäkuu 2020

Laboratoriotekniikka

TIIVISTELMÄ

Tampereen ammattikorkeakoulu
Laboratoriotekniikka

ALMINOJA, VILLE:

Cgrain Value -laitteen validointi vieraiden viljojen tunnistamiseen kauranäytteistä

Opinnäytetyö 26 sivua
Kesäkuu 2020

Opinnäytetyön tavoitteena oli validoida laboratorion vilja-analyysilaitte Cgrain Value vieraiden viljojen tunnistamiseen kauranäytteistä. Opinnäytetyön kokeellinen osuus tehtiin syksyn 2019 aikana Suomessa sijaitsevan myllyn laboratorion.

Validointia varten muodostettiin jyvänäytteitä toimeksiantajaa eniten kiinnostavista vierasviljalajeista. Jokaista näytettä analysoitiin 20 mittauskertaa. Mittauksista saadut lukuarvot kirjattiin sekaannusmatriisiin, ja mittauskertojen keskiarvot laskettiin matriisin luokille. Mittaustuloksista laskettiin laitteelle validoinnin tunnusluvut tunnistuskyky, täsmällisyys, tarkkuus, F1-arvo ja Matthewsian korrelaatiokerroin. Tunnuksista voitiin päätellä mm. kuinka hyvin laite luokitteli vierasviljalajin omaan luokkaansa, tai kuinka hyvin laite tunnisti viljan kaurasta poikkeavaksi vieraaksi viljalajiksi.

Tulosten perusteella Cgrain Value tunnisti hyvin vierasviljalajeja omiin luokkiinsa ja erityisesti kaurasta poikkeavaksi vieraaksi viljaksi. Ainoastaan yhden viljalajin näytteet laite tunnistasi huonosti sekä omaan luokkaan, että vieraaksi viljaksi, mikä tulee ottaa huomioon laitteen käyttösovellusta suunniteltaessa.

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Degree Programme in Laboratory Engineering

ALMINOJA, VILLE:

Validation of Cgrain Value for Recognition of Foreign Grain in Oat Samples

Bachelor's thesis 26 pages
June 2020

The purpose of this study was to validate The Cgrain Value laboratory device for recognising foreign grain in oat samples. The study was made for a Finnish mill in autumn 2019. The objective of the study was to determine the recall, precision and accuracy of the device.

To validate the device, different kernel samples were formed according to the interests of the client. Each sample was analysed 20 times and results were placed in a confusion matrix. The results were then used to calculate different validation characteristics and mean values for each matrix category.

The results of this study indicate that Cgrain Value can successfully recognise most of foreign grain types or at least identify them as foreign grain. The recognition was weak in one specific sample. This finding can be taken into account when utilization of the instrument is planned.

Key words: validation, grain, kernel

SISÄLLYS

1	JOHDANTO	6
2	VIERAIDEN VILJOJEN ANALYYSI	7
	2.1 Vieraiden viljojen analyysi viljanäytteistä.....	7
	2.2 Cgrain Value	7
3	MITTAUS- JA MÄÄRITYSMENETELMÄT	11
	3.1 Sekaannusmatriisi.....	11
	3.2 Validoinnin tunnusluvut	13
	3.2.1 Tunnistuskky.....	14
	3.2.2 Täsmällisyys.....	14
	3.2.3 Tarkkuus.....	14
	3.2.4 F1-arvo	15
	3.2.5 Matthewsian korrelaatiokerroin (MCC)	15
4	TYÖN SUORITUS	17
	4.1 Näytteet.....	17
	4.2 Näytteiden analysointi	18
5	TULOKSET	20
	5.1 Keskiarvo ja keskihajonta.....	20
	5.2 Tunnistuskky	21
	5.3 Täsmällisyys.....	22
	5.4 Tarkkuus, F1-arvo ja MCC	23
6	JOHTOPÄÄTÖKSET	24
	LÄHTEET	26

LYHENTEET JA TERMIT

hlp

hehtolitraino

1 JOHDANTO

Suomessa tuotettiin myllytuotantoon kelpaavaa, yli 52 hlp, kauraa 2010-luvulla keskimäärin n. 900 miljoonaa kilogrammaa vuodessa (SVT: Satotilasto). Kauran seassa on usein hieman vieraita viljalajeja, joiden määrän tunnistaminen on tärkeää laadun varmistamisen kannalta.

Tämän opinnäytetyön tavoitteena oli validoida Cgrain Value -laite vierasviljojen tunnistukseen kauranäytteistä. Työ suoritettiin Suomessa sijaitsevan myllyn laboratoriossa. Työssä tarkasteltiin laitteen tarkkuutta tunnistaa ohra, ruis ja vehnä, sekä työtä varten valmistetut luottamukselliset näytteet. Aiemmin vieraat viljalajit on tutkittu käsin, joten laitteen käyttöönotto näitä analyysejä varten vähentäisi työntekijöiden työtaakkaa ja vapauttaisi siten myös aikaa muihin työtehtäviin.

2 VIERAIDEN VILJOJEN ANALYYSI

2.1 Vieraiden viljojen analyysi viljanäytteistä

Myllyn laboratoriossa työntekijät analysoivat viljanäytteitä manuaalisesti mm. tarkastamalla näytteen joukosta sinne kuulumattomat vieraiden viljojen jyvät. Analysointi voidaan toteuttaa esimerkiksi muodostamalla satunnaisotannalla 100-500 gramman viljanäytteitä, jotka työntekijä, esim. laborantti, analysoi visuaalisesti. Analyysin tarkoituksena on saada mahdollisimman hyvä tieto viljaerän laadusta, josta vieraiden viljojen määrä on yksi osa-alue.

Manuaalisen analyysin huonoja puolia on siihen kuluva työaika sekä se, että analyysin tarkkuuteen vaikuttavat työntekijän työkokemus ja vireystila. Jos tavoitteena on löytää näytteestä jokainen vieras jyvä, joudutaan näyte usein analysoimaan vähintään kaksi kertaa. Näytteitä manuaalisesti analysoitaessa tapahtuu myös inhimillisiä virheitä.

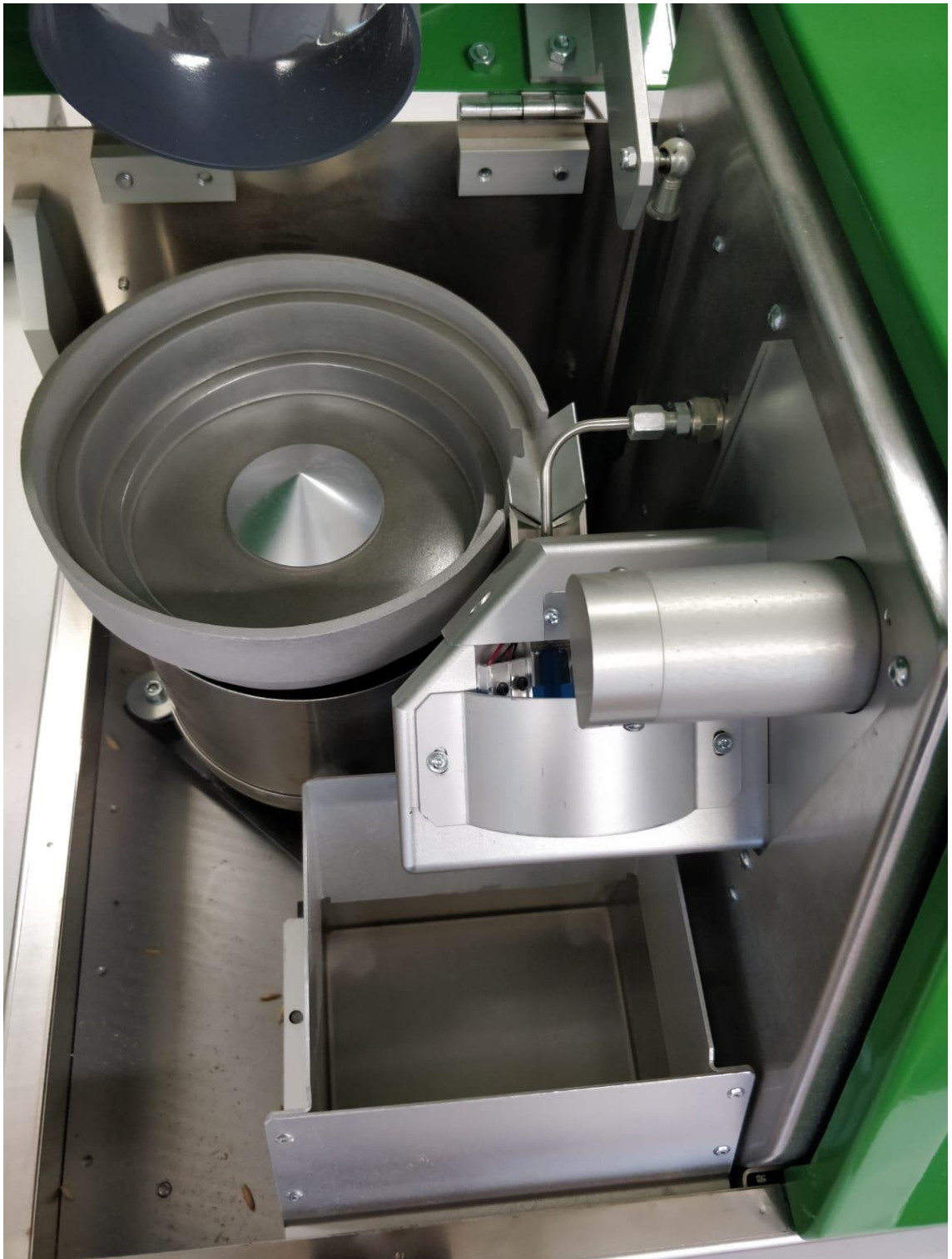
2.2 Cgrain Value

Cgrain Value on viljojen analysointiin erikoistunut laite, jolla pystytään mm. tunnistamaan kauranäytteen sisältämät vieraat viljalajit. Laite antaa myös näytteen sisältämien viljojen lukumäärät, sekä jyvien koot. Laitevalmistajan mukaan (Using Cgrain Value™ to ensure Gluten-Free Oats 2016) Cgrain Value -laitteen käyttöönotto vähentää vieraiden viljojen analysointiin kuluva työaika ja tuottaa tarkemman tuloksen kuin manuaalinen analyysi. Laite on esitetty kuvassa 1.



KUVA 1. Cgrain Value

Cgrain Value -laitteella voidaan analysoida korkeintaan 500 grammaa painavia roskista puhdistettuja jyvänäytteitä. Näytteet syötetään laitteessa olevalle alumiinikulholle. Kulho on säädetty värähtelemään tietyllä taajuudella, joka saa jyvät liikkumaan kulhon reunoja pitkin kulmikkaalle, led-valaistulle peilitasolle, jossa jyvät kuvataan yksitellen CMOS-kameralla. Kuvassa 2 on esitetty laite kansi avattuna sisältäpäin, jossa vasemmassa yläkulmassa näkyy alumiinikulho



KUVA 2. Cgrain Value kansi avattuna

Patentoidun peilitason avulla saadaan jyvän pinnasta kuvattua yli 90 %. Peilitaso on esitetty kuvassa 3. Kameran ottamat kuvat jokaisesta jyvästä tallentuvat laitteen muistiin. Laite analysoi kuvat hyödyntäen RGB-kuvantamista ja jyvien muodon tunnistusta. (Cgrain Technology 2013).



KUVA 3. Cgrain Valuen peilitaso

Cgrain Value pystyy luokittelemaan viljanäytteen jyvät muun muassa seuraaviin luokkiin: kaura, ohra, ruis, vehnä, tritikali, vieraat rikkoutuneet viljat (muu kuin kaura), rikkakasvin siemenet (Using Cgrain Value™ to ensure Gluten-Free Oats 2016). Lisäksi laitteelle voidaan asettaa luokittelemattomat-luokka, johon laite luokittelee jyvät ja partikkelit, joita se ei tunnista, mm. roskat. Tarvittaessa laitteella voidaan analyysin jälkeen käydä laitteen ottamat kuvat läpi, ja luokitella manuaalisesti jyvät oikeisiin luokkiin kuvien perusteella.

3 MITTAUS- JA MÄÄRITYSMENETELMÄT

3.1 Sekaannusmatriisi

Sekaannusmatriisi (confusion matrix) on taulukko, jonka avulla voidaan tarkastella, kuinka hyvin luokitellut asiat on luokiteltu oikeisiin luokkiin. Matriisin pystysarakkeilla on esitetty ennustetut luokat ja vaakariveillä todelliset luokat. Matriisin diagonaali kuvaa tapauksia, joissa ennustettu luokka on sama kuin todellinen luokka, ja vastaavasti ei-diagonaali tapauksia, joissa ennustettu luokka eroaa todellisesta luokasta. (Lantz 2019, 318).

Sekaannusmatriisin avulla voidaan tutkia Cgrain Value -laitteen onnistumista viljan luokittelussa. Sekaannusmatriisiin syötetyistä mittaustuloksista voidaan määrittää, kuinka hyvin laitteen tekemät luokittelut ovat osuneet oikeaan.

Laitteen validointia varten muodostettiin taulukon 1. mukainen sekaannusmatriisi. Sekaannusmatriisin pystysarakkeet täytettiin laitteen antamilla luvuilla ja vaakarivit oikeilla lukumäärillä kuvien tarkistuksen jälkeen.

TAULUKKO 1. Laitteen validointiin käytetty sekaannusmatriisi, täytetty esimerkkinä ohralle

	laitteen mukaan kaura	laitteen mukaan ohra	laitteen mukaan ruis	laitteen mukaan vehnä	laitteen mukaan tritikali	laitt. muk. vier. rikk.	laitt. muk. muut
oikeasti kaura	TN	FP	TN	TN	TN	TN	TN
oikeasti ohra	FN	TP	FN	FN	FN	FN	FN
oikeasti ruis	TN	FP	TN	TN	TN	TN	TN
oikeasti vehnä	TN	FP	TN	TN	TN	TN	TN
oikeasti tritikali	TN	FP	TN	TN	TN	TN	TN
oikeasti vieraat rikkou- tuneet	TN	FP	TN	TN	TN	TN	TN
oikeasti muut	TN	FP	TN	TN	TN	TN	TN

Taulukkoon täydennetään saadut jyvien lukumäärät seuraavasti (Lantz 2019, 319):

- Oikea positiivinen, True positive (TP): Vilja luokiteltu oikein luokkaansa
- Oikea negatiivinen, True negative (TN): Vilja luokiteltu oikein ei kuuluvaksi luokkaan
- Väärä positiivinen, False positive (FP): Vilja luokiteltu väärin kuuluvaksi luokkaan
- Väärä negatiivinen, False negative (FN) Vilja luokiteltu väärin ei kuuluvaksi luokkaan

Niille näytteille, jotka sisältävät ainoastaan yhtä tutkittavaa viljalajia, ei voitu täydentää sekaannusmatriisin kaikkia soluja. Yhden viljan näytteessä jyvä voi kuulua ainoastaan oikea positiivinen (TP) -luokkaan tai väärä negatiivinen (FN) -luokkaan. Näin ollen näille näytteille ei voida laskea validoinnin tunnuslukuja, joihin tarvitaan väärä positiivinen (FP) -luokan tai oikea negatiivinen (TN) -luokan lukuarvoja.

Lisäksi muodostettiin toinen kaksiluokkainen sekaannusmatriisi, jonka avulla voitiin tarkastella laitteen luokittelukykyä karkeammalla luokituksella kauraan ja eikauraan. Tässä omana luokkana oli kaura ja toisena luokkana vieraat, joka sisälsi kaikki muut viljaluokat paitsi kauran. Tällaisen binäärisen sekaannusmatriisin avulla voitiin laskea tunnuslukuja silloin, kun kiinnostuksen kohteena oli, osaako laite ylipäätään tunnistaa vieraan viljan vieraaksi, eikä vieraan viljan omalla luokalla ollut tässä merkitystä. Binäärinen sekaannusmatriisi on esitetty taulukossa 2.

TAULUKKO 2. Muodostettu binäärinen sekaannusmatriisi.

	laitteen mukaan kaura	laitteen mukaan vieras
oikeasti kaura		
oikeasti vieras		

3.2 Validoinnin tunnusluvut

Näytteiden analysointia varten valittiin tunnusluvuiksi tunnistuskyky, täsmällisyys, tarkkuus, F1-arvo ja Matthewsian korrelaatiokerroin. Nämä tunnusluvut, lukuun ottamatta Matthewsian korrelaatiokerrointa, saavat arvoja välillä $[0,1]$, jotka voidaan myös esittää prosenttilukuina. Huonoin mahdollinen arvo tunnusluvulle on 0, ja vastaavasti paras mahdollinen on 1. Matthewsian korrelaatiokertoimen arvojoukosta on kerrottu kohdassa 3.2.5.

3.2.1 Tunnistuskyky

Tunnistuskyky (recall) ilmaisee, kuinka tarkasti luokiteltu asia löytyy sen omasta luokasta. Tunnistuskyky ei ota kantaa siihen, kuinka paljon luokkaan on luokiteltu ylimääräisiä, ”vääriä”, asioita. Se voidaan laskea sekaannusmatriisista seuraavasti (Lantz 2019, 330):

$$\text{tunnistuskyky} = \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} . \quad (1)$$

Tässä työssä tunnistuskyky ilmaisee, kuinka hyvin laite tunnistaa jyvän omaan viljaluokkaansa, sekä millä todennäköisyydellä laite luokittelee viljan muuhun luokkaan kuin kaura. Esimerkiksi ohran tapauksessa tunnistuskyky on sitä parempi, mitä suurempi osa ohrista on luokiteltu ohriksi. Sillä taas ei ole merkitystä, kuinka paljon vääriä jyviä luokiteltiin ohriksi.

3.2.2 Täsmällisyys

Täsmällisyys (precision) ilmaisee, kuinka suuri osa positiivisesti luokitelluista asioista luokiteltiin oikein (Lee 2019, 168). Täsmällisyys lasketaan sekaannusmatriisin luvuilla kaavalla (Lantz 2019, 329)

$$\text{täsmällisyys} = \text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} . \quad (2)$$

Täsmällisyys kertoo, kuinka tarkasti laite luokittelee pelkästään tutkittavaa viljaa omaan luokkaansa. Esimerkiksi ohran tapauksessa täsmällisyys on sitä parempi, mitä suurempi osuus ohraksi luokitelluista jyvistä on oikeasti ohria.

3.2.3 Tarkkuus

Tarkkuus (accuracy) ilmaisee, kuinka suuri osa kaikista luokitelluista asioista on luokiteltu oikein. Tarkkuus lasketaan sekaannusmatriisin avulla kaavalla (Lantz 2019, 320):

$$\text{tarkkuus} = AC = \frac{TP + TN}{TP + TN + FP + FN} . \quad (3)$$

Esimerkiksi näytteessä, joka sisältää ohraa ja kauraa, tarkkuus on ohralle sitä parempi, mitä enemmän laite on osannut luokitella ohria ohriksi ja kauraa ei-ohriksi.

Tarkkuuden ongelma on siinä, ettei se ota huomioon luokkien välisiä kokoeroja. Tämän vuoksi valittiin tarkkuuden lisäksi F1-arvo sekä Matthewsian korrelaatiokerroin (MCC) kuvaamaan binäärisen sekaannusmatriisin luokittelukykyä.

3.2.4 F1-arvo

F1-arvo (F1 Score) on tunnistuskyvyn ja täsmällisyyden harmoninen keskiarvo (Solanki ym. 2020, 39). F1-arvo lasketaan kaavalla (Lantz 2019, 331)

$$F1 - \text{arvo} = 2 \times \frac{\text{täsmällisyys} \times \text{tunnistuskyky}}{\text{täsmällisyys} + \text{tunnistuskyky}} . \quad (4)$$

3.2.5 Matthewsian korrelaatiokerroin (MCC)

Matthewsian korrelaatiokerroin (Matthews correlation coefficient, MCC) on yksi tapa tarkastella binäärisen sekaannusmatriisin luokittelukykyä. Tässä työssä MCC:tä käytetään tasoittamaan suuria luokkien välisiä kokoeroja 100 gramman kauranäytteissä, joihin oli lisättyä 30 vierasjyvää. MCC lasketaan jyvänäytteelle sekaannusmatriisista seuraavasti

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} . \quad (5)$$

MCC on vielä enemmän immuuni datan epäbalanssille kuin F1-arvo ja antaa johdonmukaisia tuloksia. MCC antaa tuloksia välillä $[-1, 1]$, jossa -1 tarkoittaa tässä asiayhteydessä huonointa mahdollista tulosta ja 1 parasta mahdollista. Tulos 0 näyttää, että luokittelu ei korreloinut havaintojen kanssa. Tällaiseen tulokseen

päästään, jos yksi TP, TN, FP tai FN on 0. Lisäksi Matthewsian korrelaatiokerrointa ei voida määrittää, mikäli kaksi edellisistä on 0. (Boughrobel, Jarray & El-Anbari 2017).

4 TYÖN SUORITUS

4.1 Näytteet

Opinnäytetyön jyvänäytteet valittiin toimeksiantajaa eniten kiinnostavista viljala-jeista. Osa näytteistä oli yhden vierasviljalajin näytteitä, ja osa sisälsi 30 kappa-letta vierasviljajyviä sekoitettuna 100 grammaan kauraa. Taulukossa 3 on esitetty muodostetut näytteet.

TAULUKKO 3. Valitut jyvänäytteet

Näyte	Sisältö	Satovuosi
1	ohra 100g	2018
2	ruis 100g	2018
3	vehnä 100g	2018
4	X2 100 kpl	2018
5	X4 104 kpl	2018
6	X1 100 kpl	2018
7	X3 100 kpl	2018
8	kaura 100g + ohra 30 kpl	2018
9	kaura 100g + ruis 30 kpl	2018
10	kaura 100g + vehnä 30 kpl	2018
11	kaura 100g + X1 30 kpl	2018
12	kaura 100g + X2 30 kpl	2018
13	kaura 100g + X3 30 kpl	2018
14	kaura 100g + X4 30 kpl	2018
15	ohra 100g	2019
16	ruis 100g	2019
17	vehnä 100g	2019
18	kaura 100g + X1 30 kpl	2019
19	kaura 100g + X2 30 kpl	2019

Näytteet 1-3, 8-10 ja 15-17 muodostettiin siten, että näytteet edustaisivat kutakin satovuotta riittävän tarkasti. Vertailukelpoisia keskenään näistä näytteistä ovat ohranäytteet 1, 8 ja 15, ruisnäytteet 2, 9 ja 16, sekä vehnänäytteet 3, 10 ja 17. Lisäksi muodostettiin salassa pidettävät näytteet X1, X2, X3 ja X4.

Näytteet 7-14 ja 18-19 ovat 100 gramman kauranäytteitä, joihin sekoitettiin tutkittavan viljan jyviä 30 kappaletta. Tällä haluttiin tutkia, vaikuttaako kauran mukanaolo jyvien luokittelutarkkuuteen.

Tutkittavana viljana, jonka mukaan tulokset laskettiin, on aina vieras vilja. Kauran ollessa mukana se toimii tavallaan taustamatriisina näytteille.

4.2 Näytteiden analysointi

Jokaista näytettä analysoitiin Cgrain Value -laitteella 20 kertaa. Laitteelle oli asetettu luokiksi kaura, ohra, ruis, vehnä, tritikali, vieraat rikkoutuneet, rikkakasvin siemenet ja luokittelemattomat. Lisäksi muodostettiin erillinen vieraat-luokka, johon kuului ohra, ruis, vehnä, tritikali, vieraat rikkoutuneet, rikkakasvin siemenet ja luokittelemattomat.

Yhtä viljalajia sisältävien näytteiden tuloksista kirjattiin laitteen antamat lukumäärät eri luokkiin lajitelluille jyville. Koska näytteen tiedettiin sisältävän vain yhtä viljalajia, kiinnostuksen kohteena oli, kuinka monta jyvää luokiteltiin väärään luokkaan, ja mikä tämä väärä luokka mahdollisesti oli. Laitteen ottamia kuvia yksittäisistä jyvistä ei tarvinnut erikseen tarkastella.

100 gramman kauranäytteissä, joissa oli lisäksi tutkittavaa viljaa, pelkkien laitteen antamien lukumäärien tarkastelu ei välttämättä riittänyt. Jos jyvät eivät löytyneet omasta luokastaan, tarkastettiin lisäksi laitteen ottamia kuvia, kunnes löytyi luokat, joihin laite oli jyvät luokitellut. Näin saatiin myös selville, oliko tutkittava vilja luokiteltu vieraat-luokkaan vai virheellisesti kaura-luokkaan. Tulokset kirjattiin ylös sekaannusmatriisiin.

Sekaannusmatriisiin tallennetuista tuloksista laskettiin jyvänäytteille keskiarvo, keskihajonta, ja tunnistuskyky viljan omalle luokalle ja vieraiden luokalle. Lisäksi 100 gramman kauranäytteille, joihin oli sekoitettuna tutkittavaa viljalajia, laskettiin täsmällisyys, tarkkuus, F1-arvo sekä Matthewsinkin korrelaatiokerroin.

Yhden viljalajin jyvänäytteille ei ole mielekästä laskea kaikkia tunnuslukuja, sillä niissä ei ole vääriä positiivisia (FP) eikä oikeita negatiivisia (TN). Esimerkiksi näytteessä, joka sisältää pelkästään ohraa, ei muun viljalajin jyvä voi tulla luokitelluksi virheellisesti ohraksi. Näyte ei myöskään sisällä vierasta viljaa, joka tulisi oikein luokiteltua ei-ohraksi.

Yhden viljalajin tapauksessa täsmällisyys antaa tulokseksi aina 1, koska näytteessä ei ole vääriä positiivisia. Täsmällisyyden arvoa tarvitaan F1-arvon laskemiseen, joten se jätettiin näille näytteille myös laskematta. Kun näytteessä ei ole vääriä positiivisia eikä oikeita negatiivisia arvoja, tarkkuuden kaava supistuu samaksi kuin tunnustuskyvyn, eikä Matthewsian korrelaatiokerrointa ole määriteltä.

5 TULOKSET

5.1 Keskiarvo ja keskihajonta

Taulukossa 4 on esitetty keskiarvot ja keskihajonnat näytteille. Tulokset on esitetty viljan omalle luokalle ja vieraat-luokalle.

TAULUKKO 4. Jyvänäytteiden keskiarvot ja keskihajonnat

Näyte	Sisältö	Viljan omalle luokalle		Vieraat-luokalle	
		keskiarvo, kpl	keskihajonta	keskiarvo, kpl	keskihajonta
1	ohra 100g, 2021 kpl	1999,0	7,0	2020,1	3,9
2	ruis 100g, 2766 kpl	2744,7	6,5	2765,7	1,7
3	vehnä 100g, 2856 kpl	2826,2	26,6	2854,0	3,0
4	X2 100 kpl	65,8	3,3	93,1	1,7
5	X4 104 kpl	101,3	1,3	104,0	0,0
6	X1 100 kpl	76,8	3,5	93,7	2,4
7	X3 100 kpl	63,0	4,3	71,9	3,4
8	kaura 100g + ohra 30 kpl	29,9	0,7	30,1	0,3
9	kaura 100g + ruis 30 kpl	30,0	0,5	30,1	0,3
10	kaura 100g + vehnä 30 kpl	29,2	1,3	29,9	0,6
11	kaura 100g + X1 30 kpl	25,0	1,6	29,3	1,4
12	kaura 100g + X2 30 kpl	24,3	2,9	29,3	1,4
13	kaura 100g + X3 30 kpl	18,7	2,7	20,1	2,5
14	kaura 100g + X4 30 kpl	29,2	1,4	29,9	0,4
15	ohra 100g, 1971 kpl	1962,4	9,8	1968,9	6,5
16	ruis 100g, 2904 kpl	2885,4	11,2	2899,5	8,4
17	vehnä 100g, 2307 kpl	2293,8	3,6	2306,4	2,3
18	kaura 100g + X1 30 kpl	21,8	1,6	28,5	0,8
19	kaura 100g + X2 30 kpl	24,7	2,5	30,5	0,7

5.2 Tunnistuskyky

Näytteiden tunnistuskyvyn tulokset on esitetty taulukossa 5. Tulokset on esitetty viljan omalle luokalle ja vieraat-luokalle.

TAULUKKO 5. Näytteiden tunnistuskyky

Näyte	Sisältö	Viljan omalle luokalle Tunnistuskyky	Vieraat-luokalle Tunnistuskyky
1	ohra 100g	0,989	1,000
2	ruis 100g	0,992	1,000
3	vehnä 100g	0,990	0,999
4	X2 100 kpl	0,660	0,933
5	X4 104 kpl	0,974	1,000
6	X1 100 kpl	0,769	0,938
7	X3 100 kpl	0,644	0,736
8	kaura 100g + ohra 30 kpl	0,993	1,000
9	kaura 100g + ruis 30 kpl	0,996	1,000
10	kaura 100g + vehnä 30 kpl	0,969	0,993
11	kaura 100g + X1 30 kpl	0,826	0,968
12	kaura 100g + X2 30 kpl	0,801	0,965
13	kaura 100g + X3 30 kpl	0,623	0,671
14	kaura 100g + X4 30 kpl	0,975	0,998
15	ohra 100g	0,973	0,999
16	ruis 100g	0,984	0,999
17	vehnä 100g	0,994	1,000
18	kaura 100g + X1 30 kpl	0,721	0,946
19	kaura 100g + X2 30 kpl	0,808	0,998

5.3 Täsmällisyys

Näytteiden täsmällisyys laskettiin ainoastaan viljan omalle luokalle. Sisäisen tarkkuuden tulokset on esitetty taulukossa 6.

TAULUKKO 6. Näytteiden täsmällisyys

Näyte	Sisältö	Viljan omalle luokalle Täsmällisyys
1	ohra 100g	..
2	ruis 100g	..
3	vehnä 100g	..
4	X2 100 kpl	..
5	X4 104 kpl	..
6	X1 100 kpl	..
7	X3 100 kpl	..
8	kaura 100g + ohra 30 kpl	0,832
9	kaura 100g + ruis 30 kpl	0,829
10	kaura 100g + vehnä 30 kpl	0,731
11	kaura 100g + X1 30 kpl	0,846
12	kaura 100g + X2 30 kpl	0,890
13	kaura 100g + X3 30 kpl	0,295
14	kaura 100g + X4 30 kpl	0,864
15	ohra 100g	..
16	ruis 100g	..
17	vehnä 100g	..
18	kaura 100g + X1 30 kpl	0,916
19	kaura 100g + X2 30 kpl	0,892

.. *Ei laskettavissa*

5.4 Tarkkuus, F1-arvo ja MCC

Näytteiden tarkkuus, F1-arvo ja Matthewsian korrealaatiokerroin (MCC) laskettiin ainoastaan viljan omalle luokalle. Tunnusluvut on esitetty taulukossa 7.

TAULUKKO 7. Näytteiden tarkkuus, F1-arvo ja MCC

Näyte	Sisältö	Tarkkuus	F1-arvo	MCC
1	ohra 100g
2	ruis 100g
3	vehnä 100g
4	X2 100 kpl
5	X4 104 kpl
6	X1 100 kpl
7	X3 100 kpl
8	kaura 100g + ohra 30 kpl	0,999	0,905	0,928
9	kaura 100g + ruis 30 kpl	0,998	0,905	0,935
10	kaura 100g + vehnä 30 kpl	0,997	0,833	0,864
11	kaura 100g + X1 30 kpl	0,992	0,836	0,855
12	kaura 100g + X2 30 kpl	0,992	0,844	0,863
13	kaura 100g + X3 30 kpl	0,993	0,400	0,433
14	kaura 100g + X4 30 kpl	0,993	0,916	0,942
15	ohra 100g
16	ruis 100g
17	vehnä 100g
18	kaura 100g + X1 30 kpl	0,991	0,807	0,823
19	kaura 100g + X2 30 kpl	0,991	0,848	0,867

.. Ei laskettavissa

6 JOHTOPÄÄTÖKSET

Tarkkuuden tulokset olivat epäluotettavia kaikille kahta viljaa sisältäville jyvänäytteille, sillä luokkien välillä oli suuret kokoerot. Tällöin laskettaessa tarkkuutta esimerkiksi näytteelle, joka sisälsi 30 kpl ohraa sekoitettuna 100 grammaan kauraa, kauran määrä dominoi laskukaavoissa paljon. Matthewsian korrelaatiokerroin (MCC) on tässä parempi ottaen huomioon suuret kokoerot luokkien välillä.

Ohralle (näytteet 1, 8 ja 15) ja rukiille (näytteet 2, 9 ja 16) tunnistuskyvyn tulokset olivat erittäin hyviä. 30 kappaleen näytteissä 100 grammassa kauraa täsmällisyydet olivat melko matalia, johtuen suuresta erosta jyvien lukumäärien välillä, mutta F1-arvo ja MCC olivat molemmat hyviä. Rukiin tulokset olivat samansuuntaisia ohran kanssa.

Vehnälle (näytteet 3, 10 ja 17) tunnistuskyvyn tulokset olivat hyvin samansuuntaisia kuin ohralle ja rukiille. Täsmällisyyden tuloksista voitiin päätellä, että laite luokittelee jonkin verran enemmän muita viljoja vehnän luokkaan kuin ohralla ja rukiilla. Toisaalta laite tunnistaa vehnän hyvin omaan ja vieraiden luokkaan, ja tärkeintä olikin, että mahdollisimman moni vieras viljalaji luokiteltaisiin muuhun luokkaan kuin kaura. Vehnän luokkaan laite sijoittikin yli 97% vehnänäytteistä, josta vuoden 2019 sadon lähes sataprosenttisesti. 2018 sadon vehnänäytteessä keskihajonta oli suurempi kuin ohralla tai rukiilla johtuen yhdestä mittauskerrasta, jossa laite luokitteli huomattavasti enemmän vehnäjyviä luokittelemattomat-luokkaan verrattuna muihin mittauskertoihin. Olisi voinutkin olla perusteltua jättää tämä mittauskerta huomioimatta, mutta se otettiin kuitenkin mukaan, sillä sen vaikutus validoinnin tunnuslukuihin oli lähes merkityksetön.

Näytteillä, jotka sisälsivät viljaa X1, tunnistuskyky oli matala viljan omalle luokalle, mutta toisaalta se oli 94-97 prosenttia vieraat-luokalle, mikä on hyvä tulos. Keskihajonta oli pientä, joten laite tunnistaa viljan tasaisesti ja mittauskertojen välinen vaihtelu oli pientä. Täsmällisyys, F1-arvo ja MCC olivat hyviä.

Viljaa X2 sisältävillä näytteillä tulokset olivat hyvin samansuuntaiset kuin viljan X1 kohdalla. Laite tunnisti viljan yli 93 prosenttisesti vieraaksi viljalajiksi. 2018 sadon

X2 (näyte 4) tunnistuskyky viljan omalle luokalle oli myös matala. Täsmällisyys, F1-arvo sekä MCC olivat hyviä.

Viljaa X4 sisältävät näytteet laite tunnisti erinomaisesti, ja lähes sataprosenttisesti vieraaksi viljaksi. F1-arvo ja MCC olivat myös hyviä.

Näytteiden, jotka sisälsivät viljaa X3, keskihajonta oli samansuuruista X1-viljanäytteisiin verrattuna, mutta tunnistus viljan omaan ja vieraiden luokkiin huomattavasti heikompa. Laite tunnisti jyvistä vain hieman yli 60% omaan luokkaan ja 67%-74% vieraiden viljojen luokkaan. Tämä tarkoittaa, että laite luokittelee enimmillään jopa kolmasosan X3-viljan jyvistä kauran luokkiin. Myöskin tämän viljalajin F1-arvo ja MCC olivat todella matalia.

Tutkielman yksi tärkeimpiä asioita oli saada todennettua laitteen luotettavuus siihen, ettei se luokittelisi kauranäytteiden vieraita viljoja kauran luokkiin. Tulosten perusteella laite luokitteli luotettavasti kaikki muut näytteet, paitsi X3-viljan näytteen. X3-viljanäytteen haastavuus on tärkeä tieto laitteen käyttöönottoa suunniteltaessa. Kun tämä huomioidaan tarpeellisin varmistuksin, laitetta voidaan pitää hyvinkin luotettavana vieraiden viljojen tunnistus- ja määrittämissä menetelmänä manuaalisen tarkastuksen vaihtoehtoksi.

Tulevaisuudessa voisi muodostaa usean viljalajin näytteitä, joissa kokoerot olisivat samaa suuruusluokkaa, esim. 10 grammaa / viljalaji. Näin voitaisiin luotettavammin tarkastella myös täsmällisyyttä ja tarkkuutta. Yleinen luokittelutarkkuus ei kuitenkaan ollut validointia suunniteltaessa tärkeintä, vaan kiinnostavinta oli, kuinka hyvin mahdollisesta 100 gramman kauranäytteestä laite tunnistaa vieraat viljat omiin luokkiinsa, tai ylipäätään vieraiksi viljoiksi.

LÄHTEET

Boughrobel, S., Jarray, F. & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. Plos One. 12 (6): e0177678. <https://doi.org/10.1371/journal.pone.0177678>

Cgrain Technology. 2013. Cgrain. Luettu 12.5.2020. <https://www.cgrain.se/technology>

Lantz, B. (2019). Machine Learning with R (3rd Edition) - 10.1.2 A Closer Look at Confusion Matrices. Packt Publishing.

Lee, W-M. (2019). Python® Machine Learning - 7.2.3.3 Computing Accuracy, Recall, Precision, and other Metrics.

Solanki, A., Kumar, S. & Nayyar, A. (2020). Handbook of Research on Emerging Trends and Applications of Machine Learning - 2.3.10 Evaluating Performance Metrics Using Confusion Matrix. (pp. 39). IGI Global.

Suomen virallinen tilasto (SVT): Satotilasto. Helsinki: Luonnonvarakeskus. Luettu 17.5.2020.

<http://statdb.luke.fi/PXWeb/sq/ec698220-35dd-49d2-a69a-6de203ea19e1>

Using Cgrain Value™ to ensure Gluten-Free Oats. 2016. Cgrain. Luettu 10.5.2020.

<https://www.cgrain.se/pdf/Gluten-Free%20Oats%20Poster%20m%20copywrite.pdf>