



Magyar Agrár- és Élettudományi Egyetem

**EGYES SNACKEK, ILLETVE ZÖLD- ÉS PÖRKÖLT KÁVÉ RONCSOLÁSMENTES
ANALITIKAI, VALAMINT ÉRZÉKSZERVI VIZSGÁLATA**

DOI: 10.54598/003600

Benes Eszter Luca

Budapest

2023

A doktori iskola

megnevezése: Élelmiszertudományi Doktori Iskola

tudományága: Élelmiszertudományok

vezetője: Simonné Dr. Sarkadi Livia,
egyetemi tanár, DSc
Magyar Agrár- és Élettudományi Egyetem
Élelmiszertudományi és Technológiai Intézet
Táplálkozástudományi Tanszék

Témavezetők: Dr. Fodor Marietta,
habilitált egyetemi docens, PhD
Magyar Agrár- és Élettudományi Egyetem
Élelmiszertudományi és Technológiai Intézet
Élelmiszerkémia és Analitika Tanszék

Dr. Gere Attila,
habilitált egyetemi docens, PhD
Magyar Agrár és Élettudományi Egyetem
Élelmiszertudományi és Technológiai Intézet
Áruekezelés, Kereskedelem, Ellátási Lánc és Érzékszervi Minősítési
Tanszék

.....
Az iskolavezető jóváhagyása

.....
A témavezető(k) jóváhagyása

Tartalomjegyzék

Jelölések, rövidítések jegyzéke	6
1. Bevezetés	9
2. Célkitűzések.....	11
3. Irodalmi áttekintés.....	12
3.1. Snack termékek.....	12
3.1.1. Általános jellemzés	12
3.1.2. A snackgyártás különböző aspektusai	13
3.1.3. A minőségellenőrzés lehetőségei a snack termékekkel összefüggésben.....	14
3.2. Kávé.....	16
3.2.1. A zöldkávé általános jellemzése	16
3.2.2. A zöldkávé feldolgozása.....	18
3.2.3. A kávépörkölés	19
3.2.4. Aromavegyületek kialakulása.....	21
3.2.4.1. A kávé minőségellenőrzésének lehetőségei.....	22
3.3. Közeli infravörös spektroszkópia alapjai.....	27
3.3.1. A spektrális adatok rögzítése	29
3.3.2. Minőségbiztosítás során alkalmazható NIRS módszerek fejlesztése	31
3.3.2.1. A modellek teljesítményének javítása	36
4. Anyag és módszer	40
4.1. Vizsgált minták.....	40
4.1.1. Snack termékek.....	40
4.1.2. Kávé.....	40
4.1.2.1. A zöld kávé minták	40
4.1.2.2. Kávépörkölés	41
4.2. Vegyszerek és reagensek	41
4.3. Referencia módszerek.....	42
4.3.1. Snack termékek beltartalmi paramétereinek meghatározása	42
4.3.2. Zöld- és pörkölt kávéminták vizsgálata	43
4.3.2.1. Minta-előkészítés analitikai vizsgálatokhoz	43
4.3.2.2. Savasság meghatározása pH-mérés alapján.....	43
4.3.2.3. Koffeintartalom meghatározása	44
4.3.2.4. Klorogénsavak meghatározása	44
4.3.2.5. Tristimulusos színmérés	45
4.3.2.6. Érzékszervi bírálat	46

4.3.3.	FT-NIR spektroszkópia	46
4.3.4.	Kemometriai módszerek	47
4.3.4.1.	Kruskal-Wallis próba	47
4.3.4.2.	Főkomponens elemzés (<i>principal component analysis</i> , PCA)	47
4.3.4.3.	Adat-előkezelési módszerek	47
4.3.4.4.	Osztályozási módszerek	48
4.3.4.5.	Részleges legkisebb-négyzetek elve regresszió (PLSR)	54
5.	Eredmények és értékelésük.....	56
5.1.	Snack termékek.....	56
5.1.1.	Referencia adatok és a spektrális adatok áttekintése	56
5.1.2.	Kemometriai értékelés	60
5.1.2.1.	Főkomponens elemzés eredményei	60
5.1.2.2.	Snack termékek osztályozása a NIR spektrum alapján	61
5.1.2.3.	Részleges legkisebb négyzetek elve regresszió (PLSR) eredményei	65
5.1.2.4.	Az energiatartalom becslése iPLS regresszióval	67
5.2.	Kávé.....	70
5.2.1.	Referencia adatok kiértékelése	70
5.2.1.1.	Zöld kávé	70
5.2.1.2.	Pörkölt kávé	74
5.2.2.	Spektrális adatok minőségi vizsgálata	79
5.2.2.1.	Zöld kávé	79
5.2.2.2.	A pörkölés hatása a NIR spektrális adatokra	86
5.2.3.	Zöld kávé eredetének vizsgálata	89
5.2.4.	Részleges legkisebb négyzetek elve regresszió (PLSR) eredményei	95
5.2.4.1.	Zöldkávész	95
5.2.4.2.	Pörkölt kávé	99
6.	Következtetések és a javaslatok.....	106
7.	Új tudományos eredmények	109
8.	Összefoglalás	111
	Summary	115
	Mellékletek:.....	118
	M1. Irodalomjegyzék	118
	M2.: A vizsgált snackekre vonatkozó nominális adatok, egyéb információk.....	130
	M3.: A zöld kávék ismert jellemzői	136
	M4. Klorogénsavak meghatározása során alkalmazott tömegspektrometriás paraméterek...	137

M5.: A pörkölt kávé esetén meghatározott paraméterekre kapott korreláció elemzés eredménye	138
M6.: A pörkölt arabica kávéminták spektrumai második deriválás és sorstandardizálás adat- előkezelések után.....	139
M7.: Az afrikai és amerikai arabica kávék NIR spektrumaira futtatott főkomponens elemzés faktorsúly (loading) ábrája	139
M8.: Az ázsiai és afrikai arabica kávék NIR spektrumaira futtatott főkomponens elemzés korrelációs loading-jainak ábrája	140
M9.: A pörkölés hatása a kávé 5-CQA, 4-CQA, 3-CQA, 3,5-diCQA, 4,5-diCQA, 3-FQA koncentrációjára (mg/g)	141
Köszönetnyilvánítás	142

Jelölések, rövidítések jegyzéke

Kemometriai módszerek

AHC: *agglomerative hierarchical clustering*, agglomeratív hierarchikus klaszteranalízis

ANN: *artificial neural network*, mesterséges neurális háló

ANOVA: *analysis of variance*, varianciaanalízis

CA: *cluster analysis*, klaszteranalízis

CNN: *convolutional neural network*, konvolúciós neurális háló

DFA: *detrended fluctuation analysis*, detrendelt fluktuáció elemzés

FD: *first derivative*, első derivált

GA-OLS: *genetic algorithm-ordinary least squares*

HCA: *hierarchical cluster analysis*, hierarchikus klaszteranalízis

iPLSR: *interval partial least squares regression*, intervallum részleges legkisebb négyzetek elve regresszió

IPW-PLS: *iterative predictor weighting partial least squares regression*

ISE-PLS: *interactive stepwise elimination partial least squares regression*

k-NN: *k-nearest neighbors*, legközelebbi szomszéd módszer

KPLS: *kernel partial least squares regression*, kernel részleges legkisebb négyzetek elve regresszió

LDA: *linear discriminant analysis*, lineáris diszkriminancia elemzés

LS-SVM: *least squares-support vector machines*, legkisebb négyzetek elve-támogató vektorgépek;

MCA: *multiple correspondence analysis*, többszörös korrespondencia elemzés

MLR: *multiple linear regression*, többszörös lineáris regresszió

m-PLSR: *modified partial least squares regression*, módosított részleges legkisebb négyzetek elve regresszió

MSC: *multiplicative scattering correction*, átlagos szórás korrekció

OPS: *ordered predictors selection*

OSC: *orthogonal signal correction*, ortogonális jelkorrekció

PCA: *principal component analysis*, főkomponens elemzés

PLS-DA: *partial least squares- discriminant analysis*, részleges legkisebb négyzetek elve-diszkriminancia elemzés

PLSR: *partial least squares regression*, részleges legkisebb négyzetek elve regresszió;

RF: *random forest algorithm*, véletlen erdő algoritmus

RPropMLP: *neural network-based multi-layer feed-forward of resilient backpropagation network*, Neurális hálózat-alapú többrétegű előreccsatolt reziliens backpropagation hálózat

SD: *second derivative*, második derivált

SELECT-OLS: *stepwise orthogonalization of predictors-ordinary least squares*

SIMCA: *soft independent modelling by class analogy*

SNV: *standard normal variate*, sor-normalizálás

SOM: *self-organizing map*, önrendező térképek

SVM: *support vector machine*, támogató vektorgépek

t-SME: *t-distributed stochastic neighbour embedding*

UVE-PLS: *uninformative variable elimination*

YSI: *yield stability index*

Spektroszkópiai fogalmak

AOTF: *acousto-optic tunable filter*, akusztó-optikai hangolható szűrő

CCD: *charge coupled device*; töltéscsatolt érzékelő

FT: *Fourier-transform*; Fourier-transzformáció

HSI: *hyperspectral imaging*, hiperspektrális képfeldolgozás

MIR: *mid infrared*, középső infravörös

NIR: *near infrared*, közeli infravörös

Statisztikai metrikák

ROC: *receiver operating characteristic curve*, vevő működési karakterisztika görbe

AUC: *area under the curve*, görbe alatti terület

TPR: *true positive rate*, valódi pozitív arány, érzékenység

TNR: *true negative rate*, valódi negatív arány, specifikusság

ACC: *accuracy*, pontosság

PPV: *positive predictive value*; precizitás

RMSE: *root mean square error*, átlagos négyzetes hiba

RMSEC: *root mean square error of calibration*, átlagos négyzetes hiba a kalibrációs adatkészletre nézve

RMSECV: *root mean square error of cross validation*, átlagos négyzetes hiba a kereszt-ellenőrzés esetén

RMSEP: *root mean square error of prediction*, átlagos négyzetes hiba a teszt adatkészletre nézve

R^2 : *coefficient of determination using cross validation*, kalibrációs adatkészletre vonatkozó determinációs együttható kereszt-ellenőrzés esetén

R_t^2 : *coefficient of determination using random test set validation*, kalibrációs adatkészletre vonatkozó determinációs együttható teszt-ellenőrzés esetén

Q^2 : *goodness of validation using cross validation*, kereszt-ellenőrzésre vonatkozó determinációs együttható

Q_t^2 : *goodness of validation using random test set validation*, teszt-ellenőrzésre vonatkozó determinációs együttható

RMSECT: *root mean square error of calibration using test set validation*, átlagos négyzetes hiba a kalibrációs adatkészletre nézve teszt-ellenőrzésnél

Vegyületek, kémiai elnevezések

3,5-diCQA: *3,5-di-O-caffeoylquinic acid*; kinasav-3,5-O-dikávésav

3-CQA: *3-caffeoylquinic acid*, kinasav-3-O-kávésav; neoklorogénsav

3-FQA: *3-O-feruloylquinic acid*; kinasav-3-O-ferulasav

4,5-diCQA: *4,5-di-O-caffeoylquinic acid*; kinasav-4,5-O-dikávésav

4-CQA: *4-caffeoylquinic acid*; kinasav-4-O-kávésav; kriptoklorogénsav

5-CQA: *5-caffeoylquinic acid*, kinasav-5-O-kávésav; klorogénsav

CGA: *chlorogenic acid*, klorogénsavak

EC: *electrical conductivity*, elektromos vezetőképesség

TDS: *total dissolved solids*, összes oldott anyag

TPC: *total polyphenol content*, összes polifenol tartalom

1. Bevezetés

A napjainkra jellemző felgyorsult életmód egyre gyakrabban vezet fáradékonysághoz, energia- és időhiányhoz, ezáltal nagyobb tér jut az olyan élelmiszereknek, amelyek ezeket a „tüneteket” gyorsan képesek enyhíteni. A snack termékek és a kávé széles fogyasztói rétegekhez jutnak el a lokális és globális ellátólánc közreműködésével. Mind a snackipar, mind a kávétermesztésen és -feldolgozáson alapuló ipar egy gyorsan növekvő, változatos termékeket előállító, eltérő feldolgozástechnológiai módszereket alkalmazó piaci szereplőket magában foglaló szektor. A termékek előállítása és szállítása nagy volumenben történik, ezért olyan módszerekre van szükség, melyek hatékonyan és gyorsan képesek a termékminőség ellenőrzésére a termékpálya különböző szakaszaiban. A közeli infravörös spektroszkópia (NIRS), különböző kemometriai módszerekkel (ezen belül is a többváltozós adatelemzési módszerekkel) párosítva, eleget tesz ezeknek az elvárásoknak. Alkalmazása egyszerű és környezetbarát, mivel az élelmiszer vizsgálata becslési modellek alkalmazásán keresztül, noninvazív módon, vegyszer felhasználása nélkül történik, a spektrális és referencia adatokból létrehozott matematikai modellek segítségével. A NIRS nagymértékben támaszkodik a számítástechnikára és elektronikára nemcsak a műszer vezérlése, ill. az adatgyűjtés során, hanem az adatok elemzésénél is. Térnyerését elősegíti a napjainkban is tartó digitális forradalom, ami a számítógépek számítási kapacitásának növekedésével, az adatgyűjtés és -feldolgozás idejét is lerövidíti. Mivel a NIRS lehetővé teszi az élelmiszerek roncsolásmentes mérését is, a teljes élelmiszer-ellátási lánc során gyűjthetünk adatokat az adott élelmiszermátrix minőségi paramétereivel és mennyiségi összetételével összefüggésben, amibe beletartozik az alapanyagok minősítése is.

A doktori munkám során elsősorban a NIR spektroszkópiában rejlő lehetőségeket tártam fel különböző snack termékek, ill. zöld és pörkölt kávék vizsgálatával összefüggésben. Mindkét termékcsoporthoz más-más jellemzők (beltartalmi-, esetleg érzékszervi paraméterek) meghatározásának és elemzésének van relevanciája, melyek közül eddig nem, kevéssé vagy más szempontok alapján vizsgált paraméterek vizsgálatát tűztem ki célul. A snackek esetén a munka fókuszában olyan gyorsmódszer fejlesztése állt, amely alkalmas lehet a termékek átlagos összetételének mennyiségi becslésére a NIR spektrumok alapján. Emellett olyan statisztikai modelleket dolgoztam ki, melyek alkalmasak a snackek osztályozására a felhasznált alapanyag, sütőolaj, gyártási ország és gyártástechnológia alapján. A zöld kávémintáknál olyan vegyületeket (koffein, különböző klorogénsavak) vizsgáltam, melyek kapcsolatba hozhatók a pörkölt kávé érzékszervi tulajdonságaival, ezen kívül az ezek és a spektrális adatok közötti

összefüggéseket kerestem. A zöld kávé árát nagymértékben befolyásolja a származási hely, ezáltal a termék minősítésében szerepet játszik az eredetazonosítás, amelyre alkalmasak lehetnek a különböző spektroszkópiai módszerek, közöttük a NIRS is. A lehetőség vizsgálata érdekében a mintákat négy termesztési régióból, azon belül is huszonhárom országból szereztem be. Elvégeztem a kávék pörkölését és vizsgáltam az eltérő pörkölési idő hatását a fent említett vegyületekre, a kávébabok pH értékére, színére, az érzékelt savasságra és keserűségre, ill. a spektrális tulajdonságokra. Emellett becslési függvényeket fejlesztettem és optimalizáltam a mért jellemzők mennyiségének roncsolásmentes meghatározására.

2. Célkitűzések

Doktori értekezésem során a közeli infravörös spektroszkópia (NIRS) alkalmazhatóságát vizsgáltam snackekre, zöld- és pörkölt kávéra vonatkozóan, ezáltal a kitűzött célokat is eszerint csoportosítottam.

Snack termékek vizsgálata:

- spektrális adatbázis létrehozása;
- referencia adatbázis kiépítése a minták szénhidrát, cukor, zsír, nyers fehérje és sótartalmának meghatározása által;
- FT-NIR spektroszkópiai módszer fejlesztése a termékek átlagos tápértékének meghatározására;
- mintázatfelismerés és csoportosítás a felhasznált alapanyag és sütőolaj, a származási hely és a gyártási technológia alapján.

Zöld és pörkölt szemes kávé vizsgálata:

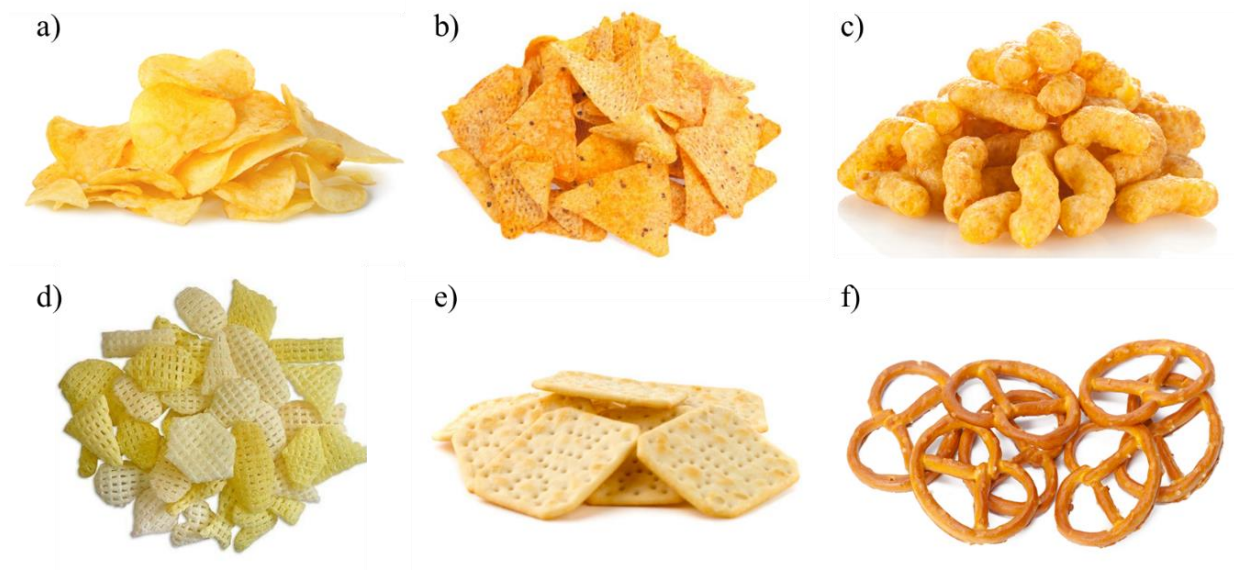
- spektrális adatbázis létrehozása;
- zöld kávé pH értékének, koffein tartalmának, klorogénsav összetételének és spektrális tulajdonságainak vizsgálata a származási hellyel összefüggésben;
- a pörkölés hatásának értékelése a vizsgált kémiai paraméterekre és a szemeskávé színére;
- a kávé képzett érzékszervi bírálók által meghatározott savasságának és keserűségének vizsgálata a kémiai és spektrális adatokkal összefüggésben;
- FT-NIR spektroszkópiai módszer fejlesztése és optimalizálása a vizsgált kémiai és érzékszervi paraméterekre vonatkozóan.

3. Irodalmi áttekintés

3.1. Snack termékek

3.1.1. Általános jellemzés

A snackek olyan élelmiszerek, melyeket alapvetően kis mennyiségben, főétkezések között fogyasztunk, beleértve a friss alapanyagokat, valamint a csomagolt és feldolgozott élelmiszereket. Ezek közé tartoznak pl. a diófélék és a magvak, a pattogatott kukorica, sült sertésbőr, burgonyachipsek (TUMULURU 2016). A legtöbb snack fogyasztásra kész termék magas olajtartalommal, sóval, vagy sóalapú fűszerrel ízesítve kerül forgalomba (1. ábra), azonban világszerte egyre nagyobb jelentőséggel bírnak az édes gyümölcs/zöldség alapú termékek is (SERNA-SALDIVAR 2022). Az 1333/2008 EK rendelet alapján a fogyasztásra kész fűszeres falatkák és snackek kategória a burgonya, gabona, liszt vagy keményítő alapú snackeket és dióféléket tartalmazza. A keményítő eredetét tekintve, származhat különböző gabonafélékből, gyökér, gumó, hüvelyes és hüvelytermés forrásból. A termékcsoportha általánosan jellemző a hosszú eltarthatóság, ami az alacsony nedvességtartalommal van összefüggésben.



1. ábra: Különböző fajta snackek. a) burgonyachips, b) tortilla chips, c) puffasztott, extrudált snack, d) pelletált, puffasztott snack, e) kréker, f) sós pereg (Internet 1-6.)

3.1.2. A snackgyártás különböző aspektusai

A snackgyártás során leggyakrabban használt alapanyag a burgonya, melyet a hagyományos burgonyachips előállításánál 0,9-1,8 mm vastagságúra szeletelnek egyenes, sima, hullámos vagy barázdált felülettel. Előfeldolgozásként a szeleteket vízzel mossák (a keményítő eltávolítása és az összetapadás megakadályozása érdekében) vagy blansírozzák (a sütés előtti barnulás elkerülése érdekében, a redukáló cukrok szintjének csökkentésével) (RIAZ 2016). Ezután a szeleteket általában olajban sütik, majd fűszerezik és csomagolják.

A burgonya nemcsak nyers, hanem szárított formában is felhasználásra kerül különböző burgonya alapú snackekben. A leggyakrabban használt alapanyagok ezek közül a burgonyakeményítő, a -pehely, a -granulátum és a -liszt. A termékcsoporthoz legismertebb képviselői közé tartoznak a PringlesTM és a Lay's Oven BakedTM termékek. A gyártás során alapvetően víz, illetve egyéb alap- és segédanyagok felhasználásával tésztát készítenek, amely azután különböző technológiai folyamatokon mehet keresztül, mint a nyújtás, vágás, extrudálás, sütés (sütőben, olajban) stb. (LUSAS és ROONEY 2001; PEDRESCHI et al. 2018; BHATTACHARYA 2023).

A tortilla chips szintén elterjedt a piacon, melynek alapanyaga a masa. Ennek összetevői a kukorica, a víz és a mész, melyeket lúgosan főznek a nixtamal előállításához, majd malomkövek segítségével dolgoznak fel. A tortilla chips durvára őrölt friss masa-ból vagy száraz masa lisztből készül, amelyet lapokra vágnak, formáznak és feldarabolnak a sütés, a kiegyenlítés, az olajban sütés és a sózás vagy fűszerezés további lépéseihez (ROONEY és SERNA-SALDIVAR 2015). A tortilla chipsek energiasűrűsége kisebb, mint pl. a burgonyachipseké, mert az olajban sütés során körülbelül 12%-kal kevesebb olajat szívnak magukba, ami annak köszönhető, hogy a megelőző sütési lépés során a termék nedvességtartalma lecsökken.

A snackgyártás során gyakran alkalmazott eljárás az extrudálás, ami kombinálja a keverés, főzés, dagasztás, nyírás, formázás és alakítás műveleteit. Alkalmazzák reggeliző pelyhek, kekszek, tészták stb. gyártásánál is (FELLOWS 2016; NIKINMAA et al. 2023). A gabonafélék típusának, a receptúrának és a folyamat körülményeinek változtatásával az extrudált snackek széles skálája állítható elő (LUSAS és ROONEY 2001).

A felhasznált sütőolajok általában a kereskedelemben nagy mennyiségben állnak rendelkezésre (SANDERS 2016). A sütőolaj kiválasztása függ a költségektől, az eltarthatósági időtől (stabilitástól), az íztől, illetve a kémiai összetételtől, melyet elsősorban a telített- és

transzzsírsvak mennyisége határoz meg (ROONEY és SERNA-SALDIVAR 2015). A két leggyakoribb típus a napraforgó- és a pálmaolaj, de léteznek egyéb alternatívák, mint pl. a szójabab, a repce, a kukorica, a gyapotmag, földimogyoró stb.

3.1.3. A minőségellenőrzés lehetőségei a snack termékekkel összefüggésben

A nassolás növekvő népszerűségére olyan tényezők vannak hatással, mint a felgyorsult életmód, az egyszemélyes háztartások, a dolgozó házastársak megnövekedett aránya; illetve a snack termékek egyszerű hozzáférhetősége pl. automatákban (RIAZ 2016). A magas zsír- és cukortartalom miatt az energiabevitel az ilyen típusú élelmiszerek által könnyen befolyásolhatja a napi energiamérleget és hozzájárulhat az elhízáshoz (PHILLIPS et al. 2004).

Az impulzusvásárlások nagy részét a különböző snackek teszik ki (DUARTE et al. 2013; STOCCHI et al. 2021; MACKENBACH et al. 2022). A fogyasztói trendek azt mutatják, hogy a fogyasztók egyre tudatosabbak az elfogyasztott élelmiszerek típusával, minőségével és az egészségre gyakorolt hatásaikkal összefüggésben (SHIROMA és RODRIGUEZ-SAONA 2009). A snack termékek tervezése, fejlesztése ezáltal nehéz, összetett feladat, ha meg akarunk felelni a fogyasztói elvárásoknak, pl. egészségesség, magas fehérjetartalom, egyedi megjelenés és íz. A legtöbb gyártó a már létező technológiák variációit alkalmazza, hogy új termékeket hozzon létre azáltal, hogy növeli az így kapott snack egészségtudatos imázsát, csökkentve a zsírok és a kalória mennyiségét, vagy hasznos tápanyagokat ad hozzá (RIAZ 2016). Mindezek, valamint a tápértékjelölés különböző követelményei miatt, a makrotápanyagok számszerűsítése rendkívül fontossá vált nemcsak a termelők számára, hogy megfeleljenek az előírásoknak, hanem a fogyasztók számára is, hogy hozzájussanak az általuk keresett információkhoz.

A számottevő érdeklődés miatt az ellenőrzés és minőségbiztosítás egyre fontosabb, ugyanakkor egyre összetettebb feladat. A globális ellátásilánc szélesedésével a snack termékek között nagyfokú változatosság figyelhető meg, mind a sütőolajok, nyersanyagok, származási helyek és a technológiai eljárások tekintetében, melyek kombinációi szinte végtelen számú terméket eredményeznek. A minőségellenőrzési folyamat kiterjed az alapanyagokra, félkész- és késztermékekre is, amelyre fokozott szükség van a nagy gyártási volumen miatt. A snackek vizsgálata során nagy hangsúlyt kapnak a textúrával összefüggésben lévő tulajdonságok, mint a nedvesség- és zsírtartalom, íz, szín, só és az aromaanyagok mennyisége (SERNA-SALDIVAR 2022). A snackek vizsgálata általában hagyományos analitikai módszerekkel, mint

a titrimetria, gravimetria, spektrofotometria vagy folyadékkromatográfia, ill. érzékszervi módszerekkel történik.

A hagyományos módszerek mellett, különböző molekulaszpektroszkópiai technikák alkalmazására van lehetőség, melyek biztosítják a gyártás során történő gyors, sok esetben roncsolásmentes analitikai mérést (PEDRESCHI et al. 2018; SERNA-SALDIVAR 2022). A NIR spektroszkópiával (*near infrared spectroscopy*, NIRS) és a snack termékek kemometriai elemzésével csak korlátozott számú publikáció foglalkozik, melyeket az 1. táblázat foglal össze. Mivel ezek főként egy adott terméktípust vizsgálnak, a mennyiségi becslésre alkalmas regressziós modellek elsősorban burgonyachipszek esetén állnak rendelkezésre. Ezekben a tanulmányokban a részleges legkisebb négyzetek elve regressziót (*partial least squares regression*, PLSR) alkalmazták sikeresen különböző kémiai és fizikai paraméterek előrejelzésére, azonban a szerzők nem végeztek komplex modell összehasonlítást, amelyben lineáris és nem lineáris modelleket is teszteltek volna.

1. táblázat: Molekula spektroszkópiai módszereket alkalmazó tudományos publikációk összefoglalása, snack termékek vizsgálatára vonatkozóan

Szerző(k)	Spektroszkópiai módszer	Mintaszám	Mátrix	Komponens, paraméter	Kemometriai módszer
C. Shiroma (SHIROMA és RODRIGUEZ-SAONA 2009)	MIR, NIR	15	burgonyachips	zsír, nedvesség	PLSR
Ni (NI et al. 2011)	NIR	80	burgonyachips	zsír, nedvesség, savszám, peroxidszám	PLSR, KPLSR, LS-SVM
T. Wang (WANG és RODRIGUEZ-SAONA 2012)	FT-NIR, FT-MIR	14	gabonaalapú és szacharózzal bevont minták	szacharóz, glükóz, fruktóz	PLSR
S. Mazurek (MAZUREK et al. 2016)	FT-Raman, MIR, NIR	43	burgonyachips	zsír	PLSR
S. Mazurek (MAZUREK et al. 2017)	FT-Raman, IR, NIR	18	burgonyachips	TPC, antioxidáns aktivitás, klorogénsav, szín L és ΔE , zsír	PLSR
J. M. Ramos-Diaz (RAMOS-DIAZ et al. 2019)	HSI-NIR	79	puffasztott kukoricaliszt és amaránt, quinoa, kañiwa keveréke	section expansion index (SEI), érzékszervi bíráló	PCA

MIR: középső infravörös; NIR: közeli infravörös, FT: Fourier-transzformáció; HSI: hiperspektrális képfeldolgozás; PLSR: részleges legkisebb négyzetek elve regresszió; KPLS: kernel részleges legkisebb négyzetek elve regresszió; LS-SVM: legkisebb négyzetek elve-támogató vektorgépek; PCA: főkomponenselemzés

3.2. Kávé

3.2.1. A zöldkávé általános jellemzése

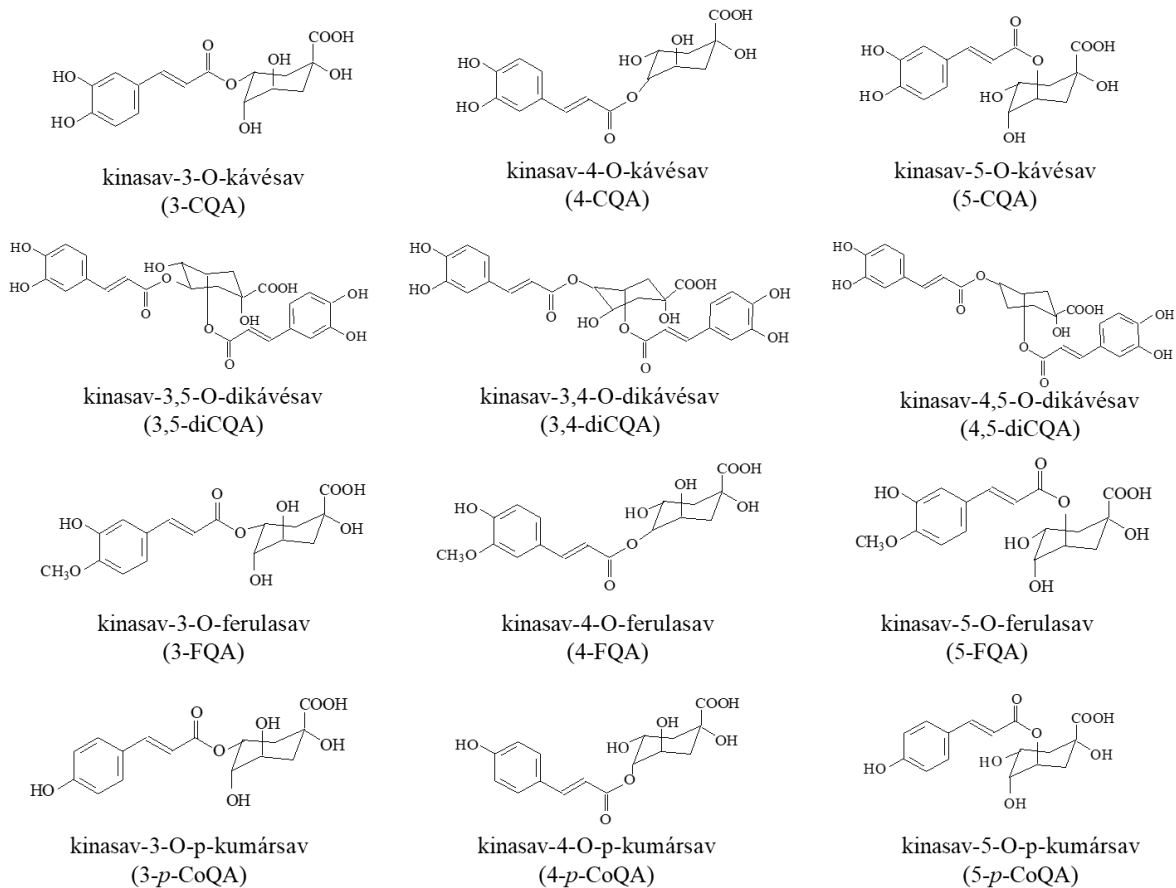
A kávé annak ellenére az egyik legnagyobb mennyiségben forgalmazott termék világszerte, hogy nem egy alapvető élelmiszer, hanem egy élvezeti cikk. Több mint 80 országban termesztik, főként Afrika, Amerika és Ázsia (szub)tropusai régióiban a kávétermő övben (HALL et al. 2022). A kávébabok fizikai tulajdonságainak és kémiai összetételének alakulásában fontos szerepet játszanak a földrajzi származáson belül a származási ország, a mikrorégiók, sőt még ugyanazon farmon belüli különböző ültetési helyek is jelentős eltéréseket okozhatnak az érzékszervi profilban (FARAH 2019). Az ital minősége emellett nagymértékben függ a pörkölési eljárástól. Jelenleg a *Coffea* nemzetségen belül több mint 100 faj van katalogizálva. E sokféleség ellenére a világpiacon valójában csak két fajnak van nagy jelentősége, melyek a *Coffea arabica* L. (arabica) és a *Coffea canephora* Pierre (robusta). Korábban az arabica kávé piaci részesedése 70%, míg a robustaé 30% volt (GEROMEL et al. 2006), azonban 2022 decemberében a százalékos megoszlás már 53,65% és 46,35% értékeket mutatott (USDA 2022). A két faj eltérő genetikai (tetraploid, diploid), morfológiai (a kávészem formája hosszúkás vagy kerekded) és kémiai tulajdonságokkal jellemezhető. A *C. arabica* fajon belül még a fajták sajátosságai is eltérnek (MALTA et al. 2020). A kávé minőségét és így az árát, a kávéfajták és -változatok, a földrajzi elhelyezkedés, a zöld kávébabok feldolgozásának módja és különösen a kávétermesztés során alkalmazott gondosság határozza meg (PREEDY 2015). A kávé minősége alapján különbséget tehetünk az átlagos (commodity) és a specialty kávé között, amelyre különböző osztályozási rendszereket használnak. Erre azért van szükség, mert a fogyasztók minőséggel, ízzel és egészséggel kapcsolatos növekvő tudatossága növeli a kiváló minőségű és különleges kávék iránti keresletet (CHENG et al. 2016).

Mivel a zöld kávébabokat többnyire eredeti formájukban értékesítik és szerzik be a kereskedelemben, a termelők és az ipar számára fontos a lehetséges minőségi markerek azonosítása, ill. azok az információk, amelyek összefüggésbe hozzák a zöld kávé összetételét a kávéital minőségével (BARBOSA et al. 2019). A genotípus kulcsfontosságú tényező, mivel nagymértékben meghatározza az olyan fontos tulajdonságokat, mint a bab mérete és alakja, valamint színe, kémiai összetétele és íze (WINTGENS 2004). A zöld kávé komplex kémiai összetételét poliszacharidok, monoszacharidok, lipidek, szterolok, zsírsavak, fenolsavak, polifenolok, alkaloidok, fehérjék, szabad aminosavak, vitaminok és ásványi anyagok együttese határozza meg. Emellett gazdag forrása az antioxidáns és gyökfogó hatással rendelkező vegyületeknek, mint pl. a klorogénsavaknak (CGA), hidroxifahéjsavaknak, koffeinnek és

kávésavnak (ŞEMEN et al. 2017). A poliszacharidok mint a mannán, galaktán, glükán és arabinán, a kávé száraz tömegének kb. 50%-át teszik ki. A lipidek, fehérjék, cukrok (főként szacharóz) és a különböző savak (főként a klorogénsavak) a *Coffea arabica* 16, 11, 8 és 7%-át teszik ki, míg a *Coffea canephora* esetén a 10, 11, 4 és 10%-át. A kávészemek koffein és trigonellin tartalma kb. 1-2% (FARAH 2019).

A koffein (1,3,7-trimetilxanitin) egy purinvázis alkaloid, amely számos növény levelében, magjában és gyümölcsében megtalálható, ilyen pl. a kávé, a kakaó, vagy a kóladió. Élénkítő hatása miatt a világon a legszélesebb körben fogyasztott központi idegrendszeri stimuláns, ún. pszichoaktív anyag (BUDIASTRA et al. 2018). Nagy hőstabilitása okán koncentrációja a pörkölés során alig változik. A koffein egy keserű vegyület, amely részt vesz a kávéital keserű ízének kialakításában (KY et al. 2001; FARAH et al. 2006), azonban mindössze kb. 30%-ban járul hozzá (GAO et al. 2023). Ezt alátámasztja az is, hogy a kávé keserősége a koffeinmentesítési eljárások hatására nem változik jelentősen. Ez azzal is magyarázható, hogy a kávéban található keserű vegyületek, mint a kahweol, a kafestol vagy a mozambiozid, kölcsönhatásba lépnek azokkal az ízérzékelésért felelős receptorokkal, amelyekkel a koffein is (LANG et al. 2020).

A klorogénsavak a kávéban és más növényekben is megtalálható, vízben oldódó, fenolos vegyületek, amelyek bizonyos transz-fahéjsavak, pl. a kávésav, a ferulasav és a kumársav kinasavval alkotott észterei (2. ábra). A legjelentősebbek a kávéban található klorogénsavak közül a kinasav-3-O-kávésav (3-CQA, neoklorogénsav), a kinasav-4-O-kávésav (4-CQA, kriptoklorogénsav), a kinasav-5-O-kávésav (5-CQA, klorogénsav), a kinasav-3,4-O-dikávésav észter (3,4-diCQA), a kinasav-3,5-O-dikávésav észter (3,5-diCQA), kinasav-4,5-O-dikávésav észter (4,5-diCQA). Ezek mellett kisebb mennyiségben vannak jelen a kinasav ferulasavval alkotott észterei (3-FQA, 4-FQA, 5-FQA), illetve a kinasav p-kumársavval alkotott észterei (3-p-CoQA, 4-p-CoQA, 5-p-CoQA) (LIANG és KITTS 2016). A klorogénsavakra jellemző a szerkezeti izoméria. A zöldkávé klorogénsav-profilját befolyásolja a kávénövény fajtája, fajtaváltozata, a származási hely, illetve a post-harvest eljárások is. Hozzájárulnak az ital végső savasságához, fanyarságához és keserűségéhez. A kávé pörkölése során a klorogénsavak mennyisége egyrészt csökken, másrészt számos kémiai átalakuláson mennek keresztül. A Maillard- és Strecker-reakciók eredményeként az ital keserű íze fokozódik a kávésav felszabadulása, valamint a laktonok és más, az ízért és az aromáért felelős fenolszármazékok képződése miatt (FARAH et al. 2005).

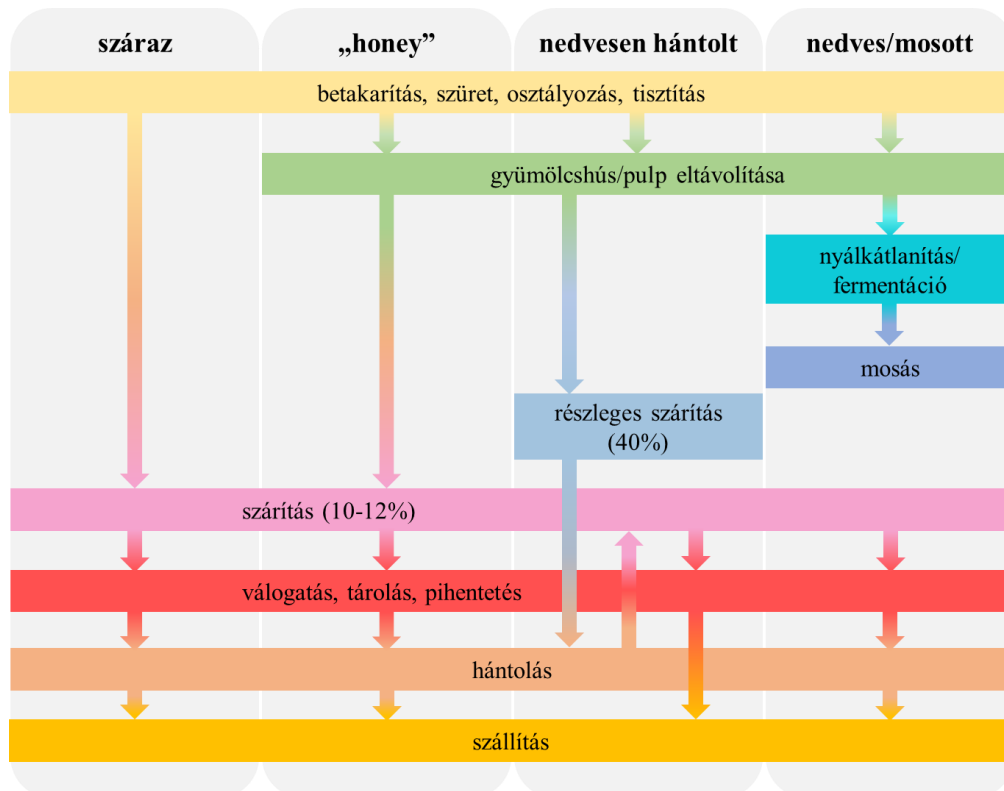


2. ábra: A kávéban található leggyakoribb klorogénsavak

3.2.2. A zöldkávészedés

A zöldkávészedés minőségét számos paraméter befolyásolja, azonban nagy általánosságban ezek két kategóriába sorolhatók: betakarítás előtti (pre-harvest) tényezők és betakarítás utáni (post-harvest) eljárások. A betakarítás előtti tényezők körülbelül 40%-ban határozzák meg a kávé íz- és illát, fizikai és kémiai tulajdonságait, a kávé minőségének fennmaradó 60%-át pedig a betakarítás utáni feldolgozás határozza meg (HAILE és KANG 2019). A betakarítás fontos lépés a jó minőségű kávé előállításához. Az ugyanazon a növényen lévő kávétermések általában nem egyszerre érnek be, ezért a betakarítás általában akkor kezdődik, amikor a termések többsége megérett. Ez a folyamat történhet kézzel vagy gépi úton. A kézi betakarítás történhet a cseresznyék egyenként történő leszedésével, vagy a gallyak lecsupaszításával és az összes gyümölcs összegyűjtésével, beleértve az érett, éretlen és túlérett gyümölcsöket is. A gépi betakarítás a fák megrázásával vagy az ágak lecsupaszításával történik (DAS 2021). A kézi szedésnek két típusa van: a gépekkel történő ún. szalagszedés (strip picking), amely növelheti a hatékonyságot, illetve a kézzel történő szelektív szedés (selective picking), amely során az

érett kávészemek leszedésével biztosítható az állandó minőség. A post-harvest feldolgozásra számos lehetőség áll rendelkezésre, mint pl. a száraz-, a nedves- és a „honey” eljárás, illetve a kávébabok nedvesen történő hántolása. Ezek főbb lépéseit a 3. ábra foglalja össze. A feldolgozás során az egyik legfontosabb szempont a kávé nedvességtartalmának kb. 40%-ról, 10-12%-ra való csökkentése, a káros mikrobiális tevékenység megelőzésére. A folyamat lényege általánosságban, hogy a zöldkávét előkészítsék a tároláshoz, szállításhoz és a pörköléshez megfelelő állapotra.



3. ábra: A zöldkávészemek feldolgozás leggyakoribb eljárásainak sematikus ábrája

3.2.3. A kávépörkölés

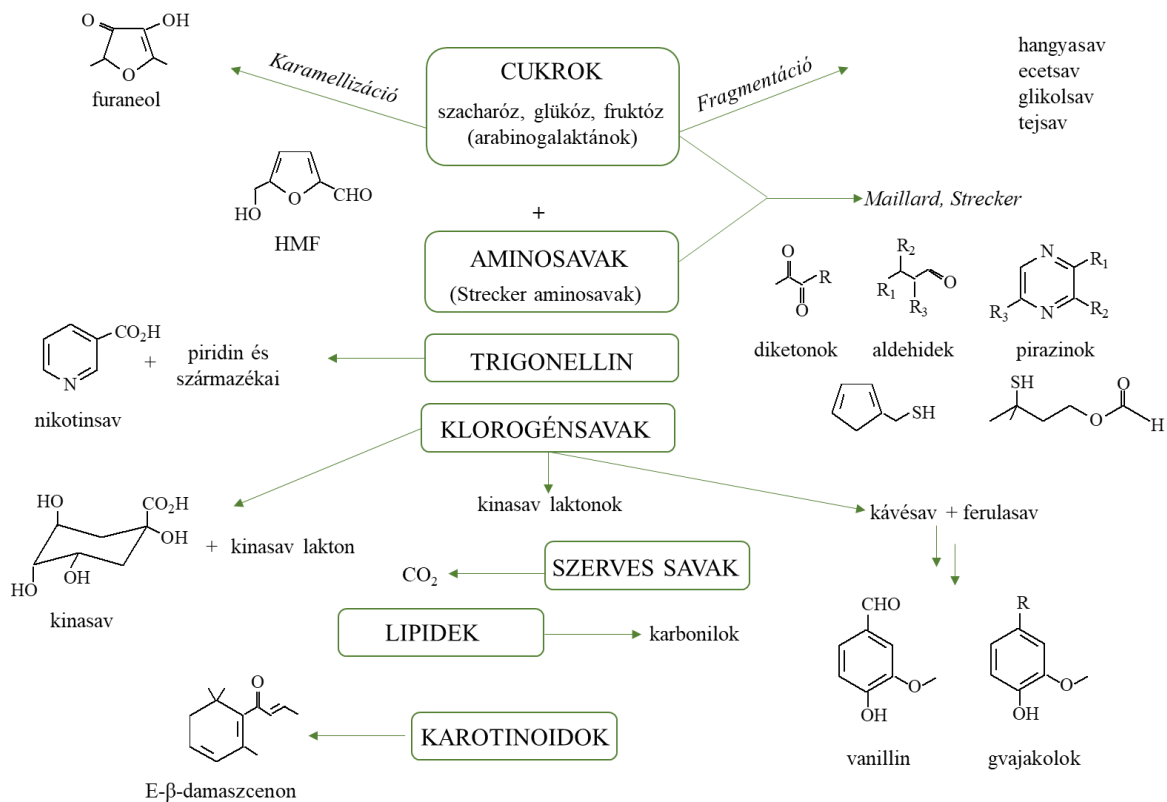
A pörkölés, melyet általános száraz hőkezelésként határoznak meg, az első olyan feldolgozási lépés, amely jellemzően nem az adott kávé termőterületén történik. A kávébabok forrólevegős pörkölése egy hagyományos hőkezelési eljárás, amelynek elsődleges célja a kívánt ízű pörkölt kávé előállítása, valamint a sötét szín és a törekeny, porózus, őrléshez és extrakcióhoz alkalmas textúra kialakítása. A pörkölés során a növekvő termékhőmérséklet kiterjedt kémiai reakciókat, dehidratációt és a mikroszerkezet mélyreható változásait idézi elő. Az élelmiszeripari alkalmazásokban (diófélék, kakaó stb.) alkalmazott más pörkölési eljárásokkal összehasonlítva a kávé pörkölése igényli a legmagasabb hőmérsékletet a kívánt termékjellemzők kialakításához.

Általában a kávébab hőmérsékletének egy bizonyos minimális időtartamig meg kell haladnia a 190°C-ot, hogy a pörkölés jellemző kémiai reakciói beinduljanak. A hagyományos kávépörkölés során a termék hőmérsékletének alakulását a végső maximumig tartó folyamatos emelkedés jellemzi, amelynél a folyamatot hirtelen, opcionális előhűtés (vízzel történő hűtés) és hűtés állítja le. A termék jellemző véghőmérséklete 200-250°C között lehet. A tipikus időtartam (pörkölési idő) 3 és 20 perc között változik (FOLMER 2017), amit befolyásol pl. a pörkölő típusa, a zöld kávé mennyisége stb. A pörkölés során, amikor a kávébab nedvességtartalma eléri a kb. 6%-ot, a hőmérséklet megközelíti a 170°C-ot. Pirolitikus reakciók játszódnak le, amelyek során nagy mennyiségű gáz (főként CO₂) keletkezik, amely a porózus bab mikroszerkezetében lévő zárt üregekben nagy nyomást eredményez. Ennek következtében a bab 50-100%-kal megduzzad, ami az első reccsenést eredményezi (YERETZIAN et al. 2002). Ahogy a kávé egyre sötétebb pörkölésű lesz, végbemegy a második reccsenés is, amely hatására a kávészemek belsejében lévő olaj a felszínre jut.

A pörkölt kávébabok kémiai összetevői határozzák meg a kávéital minőségét. A zöld kávébabok különböző kémiai vegyületek széles skáláját tartalmazzák, amelyek a kávépörkölés minden szakaszában reakcióba és kölcsönhatásba lépnek egymással és így nagyon változatos késztermék jön létre (BARBIN et al. 2014). Ennek eredményeképpen az egyes zöldkávék kémiai összetétele határozza meg az elérni kívánt pörköltiségi szintet, a kávészemek színét. A pörkölési folyamat lényege, hogy a kávészemek egyedi tulajdonságai, karaktere érvényesüljön. Bár a specialty kávéknál a világosabb pörkölés divatos, a túl világos pörkölés, az alul-pörkölés gyenge, füves ízt eredményez. A kávé érzékszervi minősítésére általában akkor kerül sor, amikor a kávébab éppen túl van ezen a fázison, de még mindig elég világos (BARBIN et al. 2014). A kávébab színe a pörkölési folyamat fontos mutatója, amely függ a pörkölési körülményektől és a zöldkávé jellemzőitől. Ezáltal a kávé színe a pörkölés egyfajta teljesítménymutatója (BERTONE et al. 2016). A pörkölést, így a folyamat ellenőrzését általában szakértők végzik, akik a kávészemek vizuális, érzékszervi és/vagy akusztikus tulajdonságainak kombinációján keresztül képesek kiválasztani a pörkölés optimális mértékét (CATELANI et al. 2018).

3.2.4. Aromavegyületek kialakulása

A kávéban több mint 1000 vegyület található, melyek között az illékony és nem illékony komponensek egyaránt befolyásolják az ital aromáját. Az aromaanyagok kialakulása egy nagyon komplex folyamat, ami a mai napig nem teljesen tisztázott. Habár a pörkölés folyamata egyszerűnek tűnhet, kémiai szempontból meglehetősen összetett, mivel egyszerre több száz kémiai reakció zajlik le. Tehát a fizikai változásokkal párhuzamosan a kémiai összetételben is mélyreható változások mennek végbe (4. ábra). A szacharóz a zöld kávé fő diszacharidja redukáló cukrokra, glükózra és fruktózra hasad. Ezek viszont a Maillard-reakcióban szabad aminosavakkal és a fehérjék szabad amino-csoportjaival reagálnak és a köztes glikozilamin-adduktok Amadori- vagy Heyns-átrendeződése után amino-ketonokat és/vagy amino-aldózokat képeznek. Az Amadori- és Heyns-termékek egy összetett reakciókaszkádban reagálnak tovább, amely számos illatanyagot és színes vegyületet eredményez. A pörkölési folyamat folytatása közben más reakciómechanizmusok is lejátszódnak, amelyek a szénhidrátok, nitrogéntartalmú anyagok, klorogénsavak és más szerves savak, valamint kisebb mértékben a lipidek, kölcsönhatás és pirolitikus hasadás útján történő lebontását foglalják magukban. Ez további víz, szén-dioxid és egyéb illékony komponensek képződéséhez vezet (YERETZIAN et al. 2002; FARAH et al. 2006). A kávépörkölés során a klorogénsavak termikus lebomlásához vezető reakciók közül sok, ha nem is mindegyik függ a pörkölés intenzitásától (pl. idő és hőmérséklet). A létrejövő változatos és összetett vegyületek kialakulásának folyamata a kinasav és fahéjsav egyszerű dekarboxileződésétől, a fenolos savak képződésén át, a klorogénsavakból képződő összetettebb laktonok képződéséig terjed. Az utóbbi esetében a magas hőmérséklet hatására intermolekuláris észterkötés alakul ki. Ezen kívül a peptidekből származó nukleofil amin- és tio-csoportok szintén példák a konjugált addíciókra, amelyek a klorogénsavakat érintik a hőkezelés során (LIANG és KITTS 2016).



4. ábra: A zöldkávét főbb alkotóiból pörkölés során kialakuló vegyületek főbb típusai (YERETZIAN et al. 2002)

3.2.4.1. A kávé minőségellenőrzésének lehetőségei

Mivel a kávé nagy gazdasági jelentőséggel rendelkező termék, minőségének vizsgálata elengedhetetlen a teljes termesztési és feldolgozási láncon keresztül. A különböző folyamatok azonban eltérő minőségi indikátorokat igényelnek. A kávé minősége meghatározható annak érzékszervi minőségén, fizikai megjelenésén, illetve a zöld- és pörkölt kávé kémiai összetételén keresztül, amely kritikus fontosságú a kávéipar számára. A kávé minősége magában foglalja a kávébabok illatát, aromáját, ízét, édességét, savasságát vagy általános ízét, amelyet a fogyasztó az ital fogyasztása után érez, valamint az olyan fizikai jellemzőket, mint a kávébabok hossza, szélessége, vastagsága vagy súlya, alakja és színe (GIRMA 2020). A kávé minősítését alapvetően szakértői érzékszervi bírálók értékelik, akik képesek az íz és aroma lehető legobjektívebb vizsgálatára, azonban személyes preferenciájuk ennek ellenére is érvényesülhet.

A Nemzetközi Szabványügyi Szervezet (ISO) hivatalosan úgy határozta meg a „minőség” fogalmát, mint „a belső tulajdonságok egy csoportjának a követelményeket kielégítő mértéke, ahol a követelmény szükségletet vagy elvárást jelent, amely lehet explicit, általában implicit vagy kötelező érvényű” (BAQUETA et al. 2020). Ennek ellenére a termelési- és ellátásilánc

során az egyes résztvevőknek, mint a kávétermesztők, a kereskedők, illetve a fogyasztók eltérő fogalmi és elvárásai vannak a minőséggel összefüggésben. Az objektívebb elemzés érdekében különböző szabványok, ajánlások pl. ISO (*International Organization for Standardization*, Nemzetközi Szabványosítási Testület), SCA (*Specialty Coffee Association*, Specialty Kávé Szövetség) állnak rendelkezésre, amelyek a kávé fizikai és kémiai paramétereinek műszeres mérésén alapulnak. Ezek értelmében a kávé minőségét leíró jellemzők közé tartozik zöld kávék esetén a nedvességtartalom, az idegen anyag jelenléte, a kávébabok mérete, formája, ill. koffeintartalma. A pörkölt kávéknál a nedvesség- és koffeintartalom mellett, a termék színe (Agtron skála) és a benne található akrilamid (CWIKOVÁ et al. 2022), ill. szénhidrát (KY et al. 2001; GEROMEL et al. 2006) is alkalmazható a minőség jellemzésére. A tudományos szakirodalom alapján a minőséget jelentősen meghatározó vegyületek lehetnek még a különböző klorogénsavak (BADMOS et al. 2019; AWWAD et al. 2021; BOBKOVÁ et al. 2021), illetve a trigonellin (KY et al. 2001; PERRONE et al. 2008; JESZKA-SKOWRON et al. 2020), amelyek mennyisége függ a pörkölés erősségétől.

Analitikai szempontból leggyakrabban hagyományos nedveskémiai módszereket alkalmaznak a minőséget leíró paraméterek meghatározására, mint a titrimetria vagy gravimetria, gyorsmódszereket, mint a refraktometria ($Brix^\circ$), színmérés, vezetőképesség (EC), illetve elválasztástechnikai módszereket, mint a gázkromatográfia (GC) és a folyadékkromatográfia (HPLC). A hagyományos módszerek mellett alternatíva lehet a NIR spektroszkópia (NIRS) alkalmazása is, amely egyre gyakoribb alacsony költsége, gyorsasága és a mintaelőkészítés rövidege miatt. Habár a NIR tartományban az egyes egyszerű molekulákhoz megfelelően hozzárendelhetők molekuláris rezgések, a biológiai minták pl. a kávé, komplex összetétele, ahol az egyszerű lineáris és elágazó szerkezetektől kezdve az összetett ciklikus és heterociklikus szerkezetekig több mint 1000 vegyület fordulhat elő, megnehezíti az elemzést. Ezáltal a spektrális adatok értelmezéséhez elengedhetetlen különböző kemometriai módszereket alkalmazni, melyek közül a leggyakrabban a főkomponens elemzést (*principal component analysis*, PCA) és a részleges legkisebb négyzetek elve regressziót (*partial least squares regression*, PLSR) használják. A NIR alkalmazhatóságát kávé esetén már számos esetben bizonyították tudományos publikációkban, amelyeket a 2. táblázat foglal össze. Zöld kávé esetében jellemzően eredetazonosításra, hamisítás feltárására és a fiziko-kémiai paraméterek becslésére, míg pörkölt kávé esetében ezeken túl a pörkölés hatásának és az érzékszervi tulajdonságok vizsgálatára alkalmazták a technikát.

2. táblázat: Közeli infravörös spektroszkópiát alkalmazó tudományos publikációk összefoglalása, zöld és pörkölt kávé vizsgálatára vonatkozóan

alkalmazás	mintatípus	mintaszám	paraméter, cél	spektroszkópiai módszer	statisztikai módszer	szerző, év
ZÖLD KÁVÉ						
klasszifikáció, eredetazonosítás, hamisítás	A, R	194	mosott A, száraz eljárással előállított A, R elkülönítése	FT-NIR	PCA, LDA	(BURATTI et al. 2015)
	A	90	genotípus, eredet	NIR	PLS-DA	(MARQUETTI et al. 2016)
	A	74	genotípus, eredet	FT-IR, NIR	SVM	(BONA et al. 2017)
	A, R	191	eredet	FT-NIR	PCA, PLS-DA	(GIRAUDO et al. 2019)
	A, R	7	eredet	NIR	PCA, SIMCA	(OKUBO és KURATA 2019)
	A	59	szín (L,a*,b*)	NIR	PCA, SOM	(WONGSAIPUN 2021)
	A, C	182	minőség (specialty, földrajzi eredetvédett, általános)	NIR	PCA, SIMCA, PLS-DA	(MUTZ et al. 2023)
fiziko-kémiai tulajdonságok becslése	A, R	83	koffein, teobromin, teofillin	NIR	PCA, PLSR	(HUCK et al. 2005)
	A	126	kafesztol, kaveol	NIR	PCA, PLSR	(SCHOLZ et al. 2014)
	A, R	108	nedvességtartalom, tömegveszteség, sűrűség	FT-NIR	MLR, PLSR	(ADNAN et al. 2017)
	A	250	pH, titrálható savtartalom	FT-NIR	PLSR	(ARAÚJO et al. 2020)
	A	74	koffein, 5-CQA	FT-NIR	LDA, PLS-DA	(ADNAN et al. 2020)
	A	250	elektromos vezetőképesség, kálium kioldódás	FT-NIR	PLSR	(DA SILVA ARAÚJO et al. 2020)
	A, R blendek	113	koffein, trigonellin, 5-CQA	NIR	OPS-PLSR	(RIBEIRO et al. 2021)
	A	50	lipidek, fehérje	FT-NIR	PCA, PLSR	(ZHU et al. 2021)
	A	250	nedvességtartalom, TDS, összes és redukáló cukor	FT-NIR	PLSR	(LEVATE MACEDO et al. 2021)
	A	250	TPC	FT-NIR	PLSR	(DA SILVA ARAÚJO et al. 2021)

alkalmazás	mintatípus	mintaszám	paraméter, cél	spektroszkópiai módszer	statisztikai módszer	szerző, év
PÖRKÖLT KÁVÉ						
klasszifikáció, eredetazonosítás, hamisítás	A, R blendek	99	A/R arány, hamisítás kukoricával, kávé törmelékekkel	microNIR	PCA, PLSR	(CORREIA et al. 2018)
	A, R	57	spektrális adatok	NIR	PCA, AHC, MCA, DFA, YSI	(BENES et al. 2020)
	A	30	hamisítás: cikória, kukorica, árpa	FT-NIR	PCA, t-SNE, PLS, iPLSR, CNN	(NALLAN CHAKRAVARTULA et al. 2022)
	A	4	hamisítás: R, kukorica, szója, rizs, árpa, kávé törmelék	NIR	PCA	(DE CARVALHO COUTO et al. 2022)
	A	18	specialty/nem specialty	NIR	PCA, HCA, SIMCA	(BAPTISTA MANUEL et al. 2022)
	A, R	97	földrajzi eredet, fajta	ATR-FTIR, NIR, NMR	PLS-DA	(MEDINA et al. 2017)
fiziko-kémiai tulajdonságok becslése	A, R	83	koffein, szín (L*, a*, b*)	NIR	PLSR, IPW-PLS, ISE-PLS, UVE-PLS, GA-OLS, SELECT-OLS	(PIZARRO et al. 2007)
	A, R	48	nedvességtartalom, tömegveszteség, sűrűség	FT-NIR	HCA, PLSR	(ALESSANDRINI et al. 2008)
	A	39	tömegveszteség	NIR	PLSR, MLR	(SHAN et al. 2015)
	A blendek	55	kávéhibák	FTIR, NIR	PLSR	(CRAIG et al. 2015)
	A, R blendek	130	szín (L*, a*, b*)	FT-NIR	PLRS	(BERTONE et al. 2016)
	A, R	168	titrálható savtartalom	NIR	PCA, PLSR	(SANTOS et al. 2016)
	A, R blendek	99	A/R arány, hamisítás kukoricával, ágakkal	microNIR	PCA, PLSR	(CORREIA et al. 2018)
	A, R blendek	78	nedvességtartalom, sűrűség, szemcseméret	NIR	PCA, PLSR	(TUGNOLO et al. 2019)
	spanyol márkák	36*4	koffein, 5-CQA, TPC, melanoidinek	NIR-HSI	ANOVA, PCA, m-PLSR	(NOGALES-BUENO et al. 2020)
	A, R	287	nedvességtartalom	NIR	PCA, PLSR	(TUGNOLO et al. 2021)

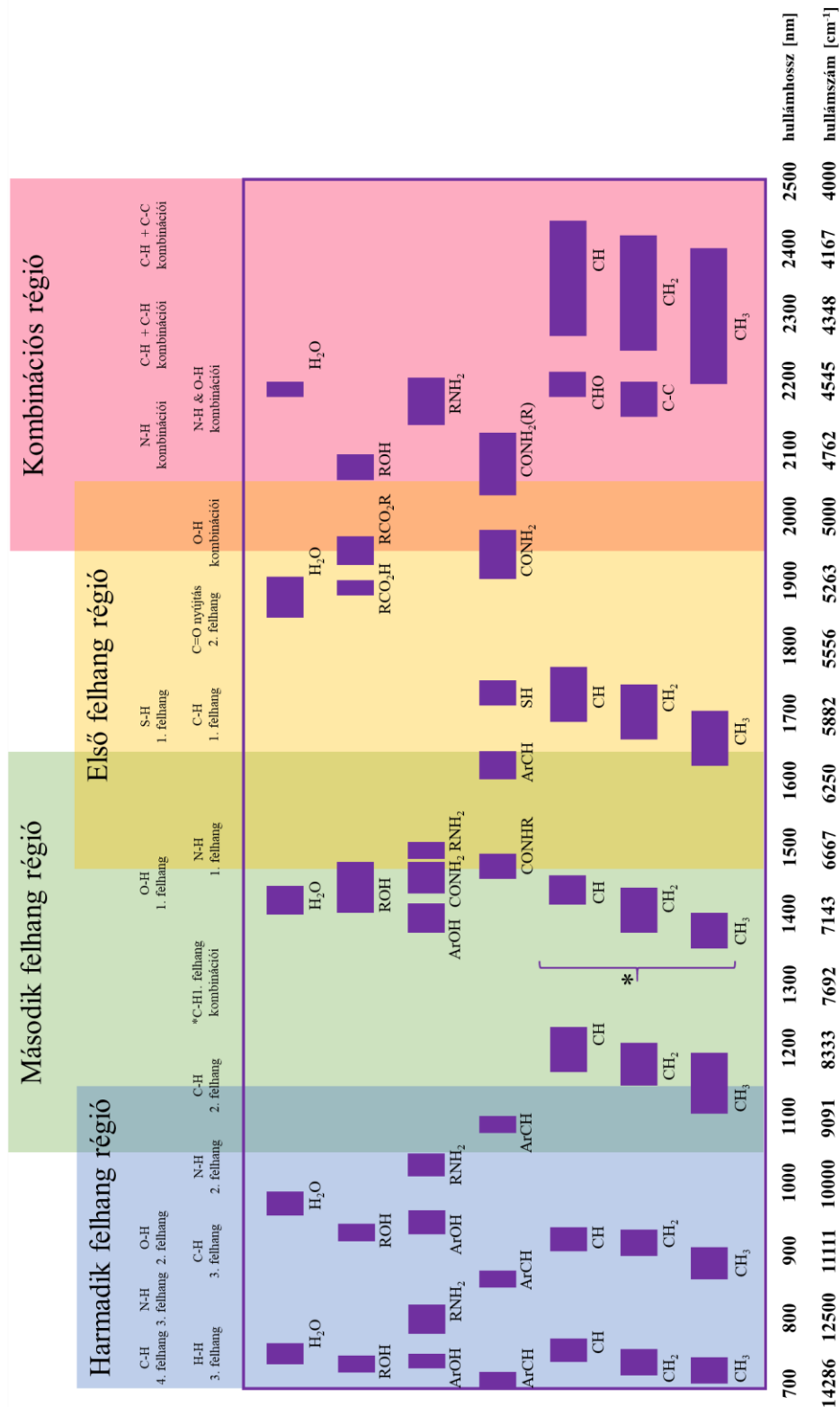
alkalmazás	mintatípus	mintaszám	paraméter, cél	spektroszkópiai módszer	statisztikai módszer	szerző, év
pörkölés hatásának vizsgálata	A	72 (8*3*3)	koffein	NIR	MCUVE-, CARS-, SCARS-PLSR	(ZHANG et al. 2013)
	A	65	5-CQA	HPLC-UV, NIR	PLSR	(SHAN et al. 2015)
	A, R	4(52)	a pörkölés valós idejű monitorozása szín, szacharóz tartalom	in-line NIR	PLSR	(SANTOS et al. 2016)
	A	86	5-CQA	HPLC-UV, NIR	PCA, GA-, SPA-PLSR	(SHAN et al. 2017)
	A	20	spektrális adatok	FT-NIR	PCA	(CATELANI et al. 2018)
	A	32	első és második recsenés kezdete, vége	in-situ FT-NIR	PLSR	(YERGENSON és ASTON 2020)
	A	1 (67)	pörkölés vizsgálata, Agtron skála	NIR	PLSR	(PIRES et al. 2021)
érzékszervi tulajdonságok becslése	A, R	35	érzékszervi bírálat	NIR	ANOVA, PCA, IPW-PLSR	(ESTEBAN-DÍEZ et al. 2004)
	A	51	érzékszervi bírálat	NIR	PLSR	(RIBEIRO et al. 2011)
	A	86	cup quality	FT-NIR	PLSR	(TOLESSA et al. 2016)
	A	50	érzékszervi bírálat	NIR	PLSR	(SAFRIZAL et al. 2018)
	A, R	217	érzékszervi bírálat	NIR	PLSR	(BAQUETA et al. 2019)
	A, R blendek	214	érzékszervi profil	microNIR	PCA, PLS-DA	(BAQUETA et al. 2021)
	kapszulás blendek	145	érzékszervi bírálat, eredetazonosítás, hagyományos és koffeinmentes kávé, A/R/blendek elkülönítése	NIR	PCA, PLSR	(SOUZA et al. 2022)

A: arabica; R: robusta; C: connilion; TDS: összes oldott anyag; TPC: összes polifenol tartalom; NIR: közeli infravörös; FTIR: Fourier transzformációs infravörös; ATR: esillapított teljes reflexió; PCA: főkomponenselemzés; PLSR: részleges legkisebb négyzetek elve regresszió; LDA: lineáris diszkriminancia elemzés; PLS-DA: részleges legkisebb négyzetek elve-diszkriminancia elemzés; SVM: támogató vektorgépek; SIMCA: soft independent modelling by class analogy, SOM: önrendező térképek, MLR: többszörös lineáris regresszió; OPS: ordered predictors selection, AHC: agglomeratív hierarchikus klaszteranalízis, MCA: többszörös korrespondencia elemzés, DFA: detrendelt fluktuáció elemzés; YSI: yield stability index; t-SME: t-distributed stochastic neighbour embedding; iPLSR: intervallum részleges legkisebb négyzetek elve regresszió; CNN: konvolúciós neurális háló; HCA: hierarchikus klaszteranalízis; SPA: success projection algorithm IPW-PLS: iterative predictor weighting; ISE-PLS: interactive stepwise elimination; UVE-PLS: uninformative variable elimination; GA-OLS: genetic algorithm-ordinary least squares; SELECT-OLS: stepwise orthogonalization of predictors- ordinary least squares; ANOVA: varianciaanalízis; m-PLSR: módosított részleges legkisebb négyzetek elve regresszió

3.3. Közeli infravörös spektroszkópia alapjai

Tekintettel arra, hogy az élelmiszerek vizsgálatára szolgáló módszereknek gyorsnak és könnyen alkalmazhatónak kell lenniük, az elmúlt 10 évben a spektroszkópai módszerek gyakoribb használatára irányuló tendencia volt megfigyelhető az élelmiszerek vizsgálata területén (DOWNEY 2016).

A közeli infravörös spektroszkópia (*near-infrared spectroscopy*, NIRS) egy széles körben alkalmazott analitikai módszer, különböző szerves anyagok minőségi és mennyiségi elemzése során. A NIR tartományban ($12500\text{-}4000\text{ cm}^{-1}$; $800\text{-}2500\text{ nm}$) a szerves molekulákat alkotó C, H, O, N és S atomok közötti, a középső infravörös (*mid-infrared*, MIR) tartományban gerjeszthető normálrezgések felhangjai és kombinációi jelennek meg. Egy adott minta bizonyos energiájú sugárzást képes elnyelni másokat nem, ezáltal kialakítva a spektrumokat. A NIR fotonok és az anyag kölcsönhatása során a molekulákban lévő rezgési és forgási átmenetek gerjesztődnek, ezáltal a létrejövő spektrumot széles, átlapoló abszorpciós sávok alkotják. Az abszorpció akkor lép fel, ha a sugárzás frekvenciája megegyezik a rezgés frekvenciájával és ha a molekula dipólus momentuma megváltozik. A folyamat során tulajdonképpen a sugárzás energiájának egy része mechanikai energiává alakul rezgések formájában. Egy abszorpciós sáv intenzitása összefügg az atomok rezgés közbeni elmozdulása során bekövetkező dipólusváltozás nagyságával és az anharmonicitás mértékével. A NIR tartományban a legtöbb sáv hidrogénatomot tartalmazó funkciók csoportoktól származik (pl. OH, CH, NH). Ez részben annak köszönhető, hogy az XH (X: C, O, N) kötés anharmonikus állandója nagy, részben pedig annak, hogy az XH vegyértékrezgések normálrezgései a magas frekvenciájú tartományban vannak ($3800\text{-}2800\text{ cm}^{-1}$) (OZAKI et al. 2021). A NIR tartományban jelentkező rezgéseket és azok jellemző elnyelési területeit az 5. ábra foglalja össze. Az $5263\text{-}4000\text{ cm}^{-1}$ közötti tartományt, amely a legkisebb energiájú, gyakran alkalmazzák komplex azonosítási célokra, tekintettel arra, hogy a tapasztalható kombinációs rezgések a MIR ujjlenyomat tartományával vannak összefüggésben. A $7692\text{-}5263\text{ cm}^{-1}$ közötti spektrális régió a vegyértékrezgések felhangjai, míg a nagyobb energiájú terület a magasabb rendű felhangok által dominált (ENGELSEN 2016; PASQUINI 2018). A sugárzás energiájának növelésével az abszorpciós sávok intenzitása csökken, ill. sávszélesedés is tapasztalható.



5. ábra: A közeli infravörös tartományban jelentkező molekuláris rezgések jellegzetes abszorpciós sávjai

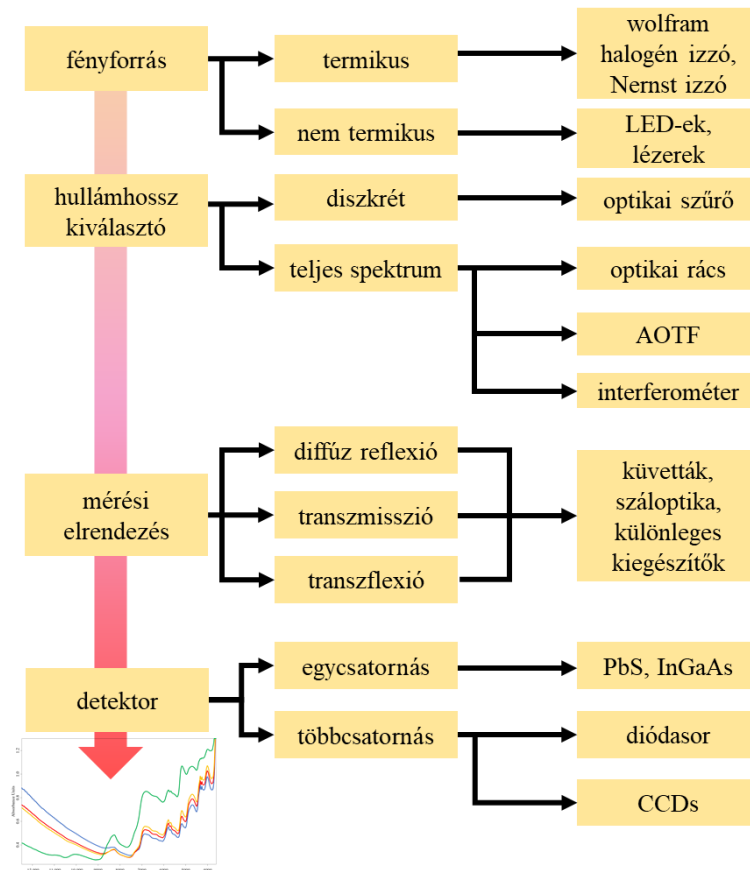
A molekuláris rezgések mellett, a spektrális adatokat a vizsgált anyag fizikai tulajdonságai is befolyásolják, mint pl. a részecskeméret, szín, tömörség. Amennyiben a részecskeméret egyenlő vagy nagyobb, mint a hullámhossz, a NIR spektrumok a részecskék felületén tapasztalható

fényszóródás által is befolyásoltak. Ez arányos a részecskeátmérő reciprokával ($1/d$). Ennek értelmében, ha az átlagos részecskeméret nő, a sugárzás mélyebbre hatol a mintába és az abszorbanca nő (Lorenz-Mie szórás) (ENGELSEN 2016). A fény többféle módon léphet kölcsönhatásba egy részecskével, melyek közé tartozik az abszorpció, a reflexió, a refrakció, a szóródás és a transzmisszió. A NIRS alkalmazása során, a fényvisszaverődést (reflexiót) és/vagy a fényáteresztést (transzmissziót) mérjük egy makroszkopikus mintán. Egy vizsgált mintán belül, vagy annak felszínén változás mehet végbe a fény törésmutatójában, amely hatására fényszóródás jön létre. Ennek a hatása leginkább a látható tartományhoz legközelebb eső $12500-10000\text{ cm}^{-1}$ közötti spektrális régióban jelentkezik, a mért abszorbanca növekedésében.

3.3.1. A spektrális adatok rögzítése

A NIRS alkalmazásának lényege abban rejlik, hogy amikor egy adott energiájú NIR sugárzás kölcsönhatásba lép az anyaggal, molekuláris rezgéseket idéz elő, amelyek a keletkező spektrumban megfigyelhető abszorpciós sávokat hoznak létre (NAGY et al. 2022). Tehát a mérés eredménye a minta spektruma, ami a hullámszám (hullámhossz) függvényében mért abszorbanca értékét adja meg. A NIR spektrométerek felépítése a céltól függően eltérő lehet (6. ábra), azonban a műszerek alapvetően tartalmaznak fényforrást, hullámhosszkiválasztó egységet, mintatartó egysége(ke)t, illetve detektor(oka)t. A minta jellemzőitől és a konkrét analitikai körülményektől, követelményektől függően a készülékek számos összeállításban elérhetőek, ami a technikát nagyon rugalmasá, sokoldalúvá teszi. A megfelelő technika kiválasztása általában a kívánt érzékenység, a megbízhatóság, a könnyű kezelhetőség, a kalibráció transzferálhatósága és a megvalósítási igény alapján történik.

A NIR spektrométereknek a hullámhossz kiválasztása tekintetében két típusa van, nevezetesen a diszkrét hullámhosszakkal és a teljes spektrummal dolgozó készülékek. Az előbbiek egyszerűbbek, mivel csak néhány hullámhosszú fotonnal sugározzák be a mintákat. Ennek eredményeképpen csak olyan alkalmazásokban használhatók, amelyekben a vizsgált komponensek meghatározott spektrumzónákban abszorbeálnak (OZAKI et al. 2021). A teljes spektrális tartománnyal dolgozó NIR műszerek tartalmazhatnak diffrakciós rácsot, azonban a modern készülékek leggyakrabban a Fourier-transzformáció (FT) elvét alkalmazzák, ezáltal interferométerrel vannak felszerelve. Ezek sokkal rugalmasabbak, mint a diszkrét hullámhosszú műszerek, így a legkülönbözőbb helyzetekben használhatók (6. ábra).

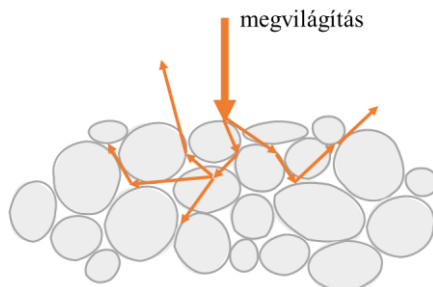


6. ábra: A közeli infravörös spektroszkópiában alkalmazott műszerek kialakításának lehetőségei
 AOTF: akusztó-optikai hangolható szűrő; CCDs: töltéscsatolt érzékelők
 (BLANCO és VILLARROYA 2002)

A NIRS különböző halmazállapotú anyagok vizsgálatára alkalmas, mint például folyadékok, kolloidok, szuszpenziók, szilárd anyagok és gázok. Mivel a minták vizsgálata általában eredeti formájukban történik, különböző optikai elrendezésekre és mintavizsgálati technikákra van szükség, melyek a diffúz reflexió, a transzmissziós és a transzflexiós.

Tekintettel arra, hogy a doktori kutatásom során a diffúz reflexiós mérési elrendezést használtam, csak ezt mutatom be részletesen. Alkalmazása szilárd halmazállapotú minták vizsgálata során a leggyakoribb. A mérés során a jelintenzitás változásához az abszorbancia mellett, a szilárd részecskék felületén létrejövő fényszóródás is hozzájárul (PASQUINI 2003). Ennek eredményeképpen a rögzített spektrum információtartalma megnő a transzmissziós és transzflexiós mérési elrendezésekhez képest. A vizsgált anyag fizikai, fiziko-kémiai tulajdonságai, mint pl. a szemcseméret, forma, rétegvastagság, szín, ill. a nedvességtartalom és az abból adódó eltérések nagyban befolyásolják a spektrális adatokat, ami az abszorbancia növekedésében, alapvonal-eltolódásban vagy csúcsetolódásban mutatkozik meg. A mérési elrendezés lényege, hogy a fényforrás felől érkező NIR sugárzás merőlegesen esik a mintára, majd részben elnyelődik, visszaverődik és véletlenszerű irányokba szóródik (7. ábra). Ennek a visszavert (reflektált)

sugárzásnak az intenzitását mérjük a detektor segítségével, amely a fényforrással megegyező oldalon helyezkedik el. Miután a fény véletlenszerűen szóródik, gyakorta alkalmaznak az összegyűjtésére arannyal bevont integráló gömböt, ami egy tökéletesen reflektáló referencia felület, mely javítja a diffúzan visszavert fény összegyűjtésének hatékonyságát.



7. ábra: Diffúz reflexió sematikus ábrája

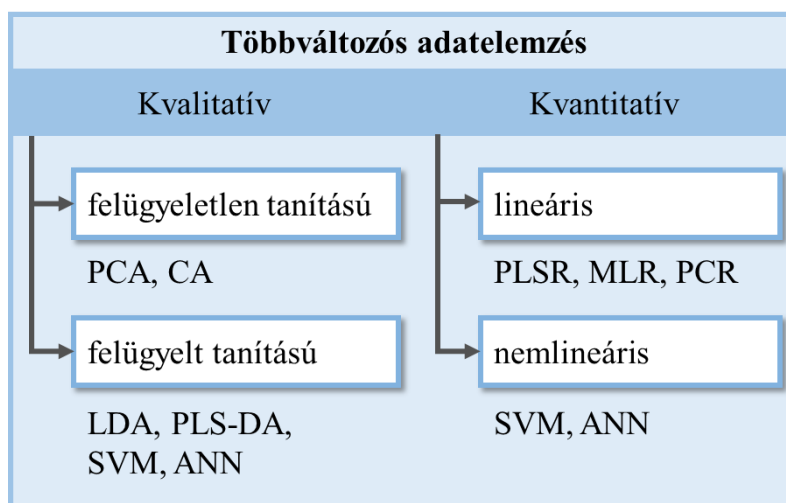
A NIR tartományban való mérés előnye, hogy a látható (*visible*, Vis) és a középső infravörös (MIR) tartománnyal ellentétben az abszorbanca értéke viszonylag alacsony, ami lehetővé teszi a nagy mintamennyiség egyszerre történő vizsgálatát. A mérések során célszerű állandó mintamennyiséggel (tömeg, térfogat) dolgozni, illetve, ha lehetőség van rá, a minták őrlésével viszonylag homogén részecskeméretet biztosítani (reprezentatív réteg). A mintahomogenitás további biztosítására nagyméretű küvetták és a minta forgatása is lehetőséget ad.

3.3.2. Minőségbiztosítás során alkalmazható NIRS módszerek fejlesztése

A mezőgazdasági termelés és az élelmiszeripari folyamatok során történő adatgyűjtés (pl. spektroszkópiai adatok, *in-situ* adatok) növekedése tükrözi az igényt a termékminőség folyamatos monitorozására. A legtöbb minőségellenőrzési rendszerben előnyt jelentő stratégia lehet a mintaáram (pl. élelmiszeripar, gyógyszeripar) NIR spektroszkópiával történő monitorozása, miután ez lehetővé teszi a gyors döntéshozatalt. A folyamatok megértése és valós idejű nyomon követése érdekében kemometriai módszereket alkalmazunk a különböző adatok kezelésére és a termékminőség becsléséhez szükséges releváns információk kinyerésére (ROGER et al. 2020). A kemometria az a tudományág, amely optimális matematikai, statisztikai és egyéb logikai alapú módszereket alkalmaz az adatok feldolgozására (ROUSSEL et al. 2014). A NIRS elterjedését elősegítette a számítástechnika és a többváltozós adatelemzés (*multivariate analysis*, MVA) területén elért fejlődés is. Miután a spektrumok rögzítése során egy adott hullámszám tartománnyal dolgozunk, a mérés eredményeképpen többváltozós adatkészlethez jutunk, amely kiértékelése összetett feladat. Attól függően, hogy minőségi vagy mennyiségi információt szeretnénk kinyerni, eltérő kemometriai módszerek alkalmazására van lehetőség, mint pl. a

főkomponens elemzés (*principal component analysis*, PCA), a polár minősítő rendszer (*polar qualification system*, PQS) (KAFFKA és SEREGÉLY 2002); klaszter analízis (*cluster analysis*, CA), részleges legkisebb négyzetek elve regresszió (*partial least squares regression*, PLSR). A fentieknek köszönhetően, az elmúlt évtizedekben a NIRS számos területen köztük a mezőgazdaságban és az élelmiszeriparban is fontos szerepet töltött be a gyors, roncsolásmentes módszerek között.

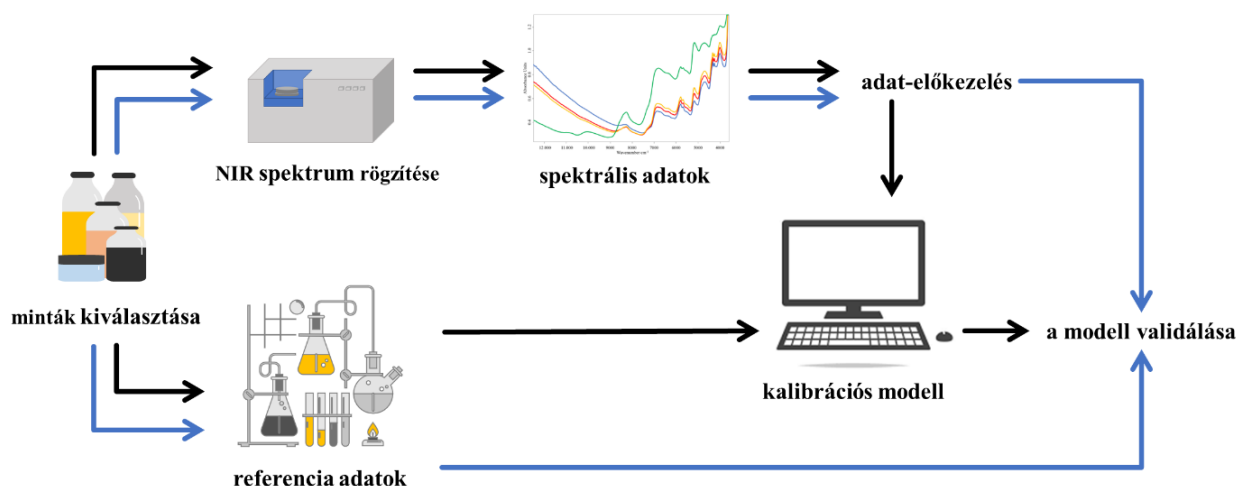
A NIR spektroszkópiát rutinszerűen használják különböző alapanyagok, élelmiszerek, félkész- és késztermékek összetételének, funkcionális és érzékszervi tulajdonságainak vizsgálatára. Mivel a NIRS egy közvetett analitikai módszer, a minőségi vagy mennyiségi paraméterek vizsgálatára statisztikai modelleket, becslési modelleket kell kidolgozni. Ehhez a NIR spektrumok mellett referencia adatokra van szükség a vizsgálni kívánt paraméterrel összefüggésben (pl. egy adott vegyület koncentrációja). A becslési függvények létrehozásához minimum 30-50 mintára van szükség (KADENKIN 2018), azonban ez erősen függ a vizsgált mintamátrixtól is. A minőségi jellemzők (pl. a származási hely, a termék típusa, a növény fajtája stb.) elemzése általában klasszifikációs módszerekkel történik, melyek segítségével lehetséges a vizsgált minták osztályokba való sorolására. A NIR spektroszkópiát azonban leggyakrabban a referencia adatokkal történő kalibrációra, majd kvantitatív becslésre használják. Erre a célra különböző lineáris és nemlineáris regressziós módszereket alkalmazhatunk (8. ábra).



8. ábra: NIR adatok elemzésére gyakran használt többváltozós statisztikai módszerek sematikus ábrája. PCA: főkomponens elemzés; CA: klaszteranalízis; LDA: lineáris diszkriminancia elemzés; PLS-DA: részleges legkisebb négyzetek elve-diszkriminancia elemzés; SVM: támogató vektorgépek; ANN: mesterséges neurális háló; MLR: többszörös lineáris regresszió

A NIR spektroszkópiában használt kemometriai módszerek alkalmazásával a nagy adathalmazok viszonylag könnyen interpretálhatók. Ezek segítségével feltárhatók az adott mintapopuláción belüli különbségek és hasonlóságok, melyek értelmezéséhez azonban a minták fizikai és kémiai

tulajdonságainak ismerete is szükséges. Mivel az egyes anyagok eltérő spektrális tulajdonságokkal rendelkeznek (ujjlenyomatszerű mintázat), a módszerfejlesztés adott mintamátrixokra történik, pl. búza, tej, pörkölt kávé stb. Az adatok elemzése sokféle módszer segítségével történhet, azonban a modellépítés főbb lépései megegyeznek (9. ábra): a minták kiválasztása, a spektrumok rögzítése, a referencia adatok meghatározása, az adatok előkezelése, a kalibráció és a modell validálása.



9. ábra: A NIR adatokon alapuló modellépítés főbb lépései.

- kalibráció ● validáció (AGELET, 2010)

A becslési modellek gyakorlati alkalmazhatóságában a **minták megfelelő kiválasztásának** nagy szerepe van. A folyamat során úgy kell eljárni, hogy a mintapopuláció a lehető legtöbb olyan eltéréssel (különbséggel) terhelt legyen, ami a későbbi alkalmazás során előfordulhat (pl. szemcseméret eltérése, szín stb.). Ennek azért van jelentősége, mert a minták vizsgálata elsősorban eredeti formájukban roncsolásmentesen, esetleg szárítás és őrlés után történik. Fontos kiemelni, hogy a különböző minta-előkészítési eljárások változásokat idézhetnek elő a mintamátrixban, ami mérési hibához vezethet (COZZOLINO 2020). Ajánlott a lehető legszélesebb koncentráció tartományt figyelembe venni a vizsgált paraméter, komponens szempontjából, esetleg extrém mintákat is bevonni a kiértékelésbe, hogy a szélsőértékek is a modell részét képezzék (COZZOLINO et al. 2011). Emellett azonban törekedni kell a referencia értékek egyenletes eloszlására a mérési tartományon belül, amellyel növelhető a becslés pontossága (SIESLER 2002; OZAKI 2006; GONZÁLEZ-CABALLERO et al. 2011), emiatt érdemes megvizsgálni és azonosítani a referencia adatok eloszlását.

A **NIR spektrumok (X-változók) rögzítése** tulajdonképpen egy spektrális adatbázis kiépítését jelenti, amely során a lehető legállandóbb mérési körülményeket kell biztosítani pl. a hőmérséklet vagy a vizsgált mintamennyiség tekintetében. A spektrumok felvétele előtt fontos a megfelelő mérési elrendezés kiválasztása, ami a szemcsés, őrlött vagy por állagú minták esetén általában a

diffúz-reflexió, kolloid állagú minták esetében a transzflexió, míg folyadékok esetén leggyakrabban a transzmissziós elrendezés (BEC et al. 2022; NAGY et al. 2022). A modell létrehozása előtt célszerű a spektrálisan kieső mintákat meghatározni főkomponens elemzés (PCA) alkalmazásával, mert ezek torzíthatják a becslést. Az eredmény részeként meghatározott mutatók, mint a Leverage vagy a Hotelling-féle T^2 statisztika, a modell reziduumainak ellenőrzésével kombinálva hatékony alternatívát jelent a kiugró értékek felderítésére (AGELET és HURBURGH, 2010).

A kalibrációhoz, alternatív mérési módszerekkel meghatározott **referencia adatok (Y-változó)** használhatók fel, melyek különböző analitikai és nedves kémiai módszerekkel határozhatók meg (gravimetria, titrimetria, HPLC stb.). Olyan referencia módszert kell választani, amely pontos és precíz értékeket szolgáltat, mivel ezek nagymértékben befolyásolják a fejlesztendő modell minőségét. Klasszifikációs modelleknél a referencia adatok valamilyen minőségi paraméterrel vannak összefüggésben, pl. az élelmiszer eredetével vagy a feldolgozás típusával. Azoknak a jellemzőknek a mennyiségi becslése, melyek kvantifikálhatók (pl. vegyületek koncentrációja, színadatok, érzékszervi adatok), regressziós modellekkel történik. A módszerfejlesztés során matematikai összefüggést keresünk pl. a vizsgált vegyület koncentrációja és a spektrális adatok között.

A berendezés, a környezet, illetve a minta fizikai és kémiai tulajdonságai egyaránt befolyásolhatják a rögzített spektrumot. Ennek megfelelően ezek a kívánt és nem kívánt hatások egyaránt meghatározzák a prediktív modell teljesítményét. A zavaró hatások általában a fizikai (pl. hőmérséklet) vagy kémiai (pl. páratartalom) paraméterek változásából adódnak. Ezeknek eltérő hatása van a mért NIR spektrumokra: a részecskeméret befolyásolja az alapvonalat; a hőmérséklet befolyásolja a csúcsok helyzetét és relatív amplitúdóját; a mérés geometriai elrendezése additív és multiplikatív hatásokat vált ki stb. (ROGER et al. 2020). Ezek csökkentése és a becslés szempontjából releváns adatok kinyerése érdekében különböző **adat-előkezelési lépések végrehajtására** van lehetőség, mint pl. a normalizálás, deriválás, szóráskorrekció stb.

A módszerfejlesztés során ki kell választani az adatok modellezésére leginkább alkalmas kemometriai módszert, majd optimalizálni a matematikai modellt. A spektrális adatokkal történő regresszió esetén, a lineáris algoritmusok alkalmazása a leggyakoribb, mint a PLSR, a PCR (*principal component regression*, főkomponens regresszió) és az MLR (*multiple linear regression*, többszörös lineáris regresszió). Ez azzal van összefüggésben, hogy általánosan véve a NIR spektroszkópiai mérések alapja a Lambert-Beer törvény, miszerint a meghatározott abszorbancia és az adott vegyület koncentrációja között egyenes arányosság áll fenn. A NIRS esetében a

többszörös kalibrációs módszerek célja egy összetevő mennyiségének becslése (vagy egy adott jellemző pl. szemcseméret) a rögzített spektrumok alapján. A folyamat során a spektrumok és a referencia adatok egy részét a modell tanítására használjuk (training set), míg egy másik részét a modell validálására (test set). A fentebb említett PLSR egy jól ismert és széles körben alkalmazott módszer a vizsgálni kívánt minták jellemzőinek mennyiségi becslése során. Különösen fontos a nagy dimenziójú adatok feldolgozásánál, főként akkor, ha a magyarázó változók száma meghaladja a minták számát (METZ et al. 2021), mivel ebben az esetben a hagyományos lineáris regressziós módszerek nem alkalmazhatók. Ezzel szemben, az osztályozási problémák megoldásánál gyakori a nem lineáris modellek alkalmazása (NI et al. 2011; BONA et al. 2017), mint pl. az mesterséges neurális háló (*artificial neural network*, ANN) vagy a támogató vektorgepek (*support vector machines*, SVM).

Tekintve, hogy a kapott kalibrációs modell megbízhatósága és alkalmazhatósága teljes mértékben a modell létrehozásához használt adatoktól függ, minden matematikai modell teljesítményét szükséges validálni. Ez kulcsfontosságú lépés annak megállapításához, hogy a modell alkalmas-e az új minták előrejelzésére, hiszen alapvetően ez a NIR kalibrációk kifejlesztésének célja (AGELET és HURBURGH, 2010). A leggyakrabban alkalmazott módszer a kereszt-ellenőrzés pl. alacsony mintaszámnál, azonban vitatható, hogy mennyire jól tükrözi a teljesítményt. Összességében jó kompromisszumként az öt- vagy tízszeres kereszt-ellenőrzés ajánlott (HÉBERGER és KOLLÁR-HUNEK 2019). Ideális esetben a modellek validálása független adatkészlet segítségével, teszt-ellenőrzéssel történik. Ilyenkor általában a teljes adatkészletet két részre osztjuk fel: kalibrációs- és validációs adatkészletre. A felosztás aránya leggyakrabban 60/40, 70/30, vagy 80/20. Ez történhet random felosztással, ill. olyan algoritmusok alkalmazásával, mint a Kennard-Stone (KENNARD és STONE 1969) vagy a Duplex algoritmus (XIAPING et al. 2011). Összességében a validáláshoz használt módszer az adatkészlet függvénye.

A modellek teljesítménye különböző statisztikai metrikák segítségével értékelhető. A determinációs együttható (R^2) azt méri, hogy a modell a válaszok varianciájának hány százalékát magyarázza. Biológiai adatok esetében a 70%-hoz közeli R^2 már nagynak számít (BROWN et al. 2020). Mivel a NIR adatokat erős multikollinearitás jellemzi, a modellezés minden esetben tartalmaz egy adatredukciós lépést, amely során egymástól független, ortogonális látens változók jönnek létre. Ezek az új változók (látens változók) az eredeti változók lineáris kombinációi és úgy vannak sorba rendezve, hogy elől állnak azok, amelyek az eredeti változó együttes varianciájának (tehát az X mátrix összes elemére számított szórásnégyzetnek) legnagyobb részét magyarázzák (HORVAI 2001). A látens változók száma fontos a modell komplexitása szempontjából, emellett a kalibráció potenciális hatékonyságának is a mérőszáma. Gyakorlatban a legtöbb sikeres modell

legfeljebb hat-nyolc látens változót használ, míg tizenkettő-tizenöt a számításban jelentkező bizonytalanságot jelezheti (WILLIAMS et al. 2017). Azonban számukat nagyban befolyásolja a vizsgált mátrix és a minták száma. A látens változók optimális számát a legkisebb átlagos négyzetes hiba (*root mean square error*, RMSE) érték alapján határozhatjuk meg (COZZOLINO et al. 2011), azonban a túlillesztés elkerülése érdekében figyelembe kell venni az X-Y magyarázott variancia értékét is. Utóbbi azt mutatja meg, hogy az egyes komponensek milyen mértékben magyarázzák a spektrális és referencia adatok varianciáját. A módszerfejlesztés során a cél az RMSE minimalizálása, amely értéke jó képet ad az abszolút hibáról, de nem mondja meg, hogy a hiba szisztematikus vagy véletlenszerű (SAEYS et al. 2019).

A modell teljesítőképessége javítható még különböző változókiválasztási módszerekkel, melyek alkalmazásával a modell komplexitása tovább csökkenthető azáltal, hogy a regresszió/klasszifikáció szempontjából lényeges hullámszámtartományokra korlátozódik a kiértékelés.

3.3.2.1. A modellek teljesítményének javítása

Adat-előkezelési módszerek

Az adat-előkezelés a nyers spektrális adatok matematikai feldolgozása, amely a modell létrehozásának vagy az algoritmus alkalmazásának előkészítéseként történik. Az adat-előkezelés oka alkalmazásonként eltérő lehet, de leginkább a nemkívánatos eltérések csökkentésére és a releváns különbségek kiemelésére használhatók. A releváns spektrális információk feltárására érdekében különböző technikák alkalmazhatók, mint pl. szórás korrekciós műveletek, deriváltak stb. Általában az adatokra alkalmazandó legjobb előkezelés vagy azok kombinációjának kiválasztása nem egyértelmű. Célszerű a különböző matematikai transzformációkat elvégezni, majd az eredmények fényében (statisztikai paraméterek) kiválasztani a céloknak leginkább megfelelőt.

A leggyakrabban alkalmazott szóráskorrekciós módszer az átlagos szórás korrekció (*multiplicative scatter correction*, MSC), amely eltávolítja mind az additív, mind a multiplikatív hatásokat a diffúz-reflexiós spektroszkópiában (SHIROMA és RODRIGUEZ-SAONA 2009). Az MSC egy modell-alapú módszer, amely során az adatkészletre vonatkozó átlagspektrummal korrigálunk minden spektrumot. Elsősorban olyan esetekben működik, amikor a spektrális eltérések elsődleges forrása a szóródás. A sor-normalizálás (*standard normal variate*, SNV) egy másik széles körben alkalmazott módszer a NIR adatok szóráskorrekciójára (RINNAN et al. 2009). Ez a módszer a spektrális adatokat soronként (mintánként) centrálja, ezáltal korrigálja az alapvonaleltolódást,

majd skálázza, amivel az optikai úthossz különbségeiből eredő eltéréseket csökkenti. Az alapvonalban jelentkező eltérések korrigálhatók a „straight line subtraction” (SLS) technikával is, amely során az algoritmus egy egyenest illeszt a spektrumra, majd ezeket az értékeket kivonja az eredeti spektrumból. Ezekon kívül, vagy az előzőkkel kombinálva, használhatók még különböző deriválási vagy simítási eljárások, mint a Savitzky-Golay algoritmus. A deriválási módszereket a NIR-spektrumok esetében a felbontás javítására és az alapvonal korrekciójára egyaránt használják. Az átlapoló abszorpciós sávok feloldásával a mennyiségi becslés pontossága is javítható. FT-NIR spektroszkópia esetében az első (*first derivative*, FD) és a második derivált (*second derivative*, SD) spektrumok alkalmazása a legelterjedtebb, de figyelembe kell venni, hogy a deriválás hatására a zaj mértéke is nő. Az egyes adat-előkezelési módszerek mellett, azok kombinációja is javíthatja a matematikai modellek teljesítményét pl. FD+SNV, SD+SNV.

Egy ezektől teljesen különböző eljárás az ortogonális jelkorrekció (*orthogonal signal correction*, OSC), amely a spektrális adatok esetében minimalizálja a vizsgált vegyület koncentrációjától független varianciát (BLANCO et al. 2001). Az OSC lényegében a PCA-t használja a spektrális adatmátrix maximális eltéréseit magyarázó, a koncentrációs mátrixra ortogonális faktorok eltávolítására (KUMAR 2020). Gyakran szolgál adat-előkezelési módszerként többváltozós regressziós modellek javítására. Az algoritmus bemenő adatai a spektrumadatok (X), ami egy $n \times p$ adatmátrix az n mintával és p spektrális adatponttal. Az n hosszúságú y vektor az X egyes spektrumainak megfelelő koncentrációkat (értékeket) jelenti. Az algoritmus háttérváltozókat hoz létre, melyek ortogonálisak a vegyület koncentrációjára (y). A háttérkomponensekhez kapcsolódó spektrális információk eltávolításával a kapott egyszerűbb spektrumok alkalmasabbak lehetnek mind a kalibrációs modell létrehozására, mind a jövőbeli mintákra vonatkozó előrejelzések elvégzésére (GESSELL és SMALL 2022).

Változókiválasztási módszerek

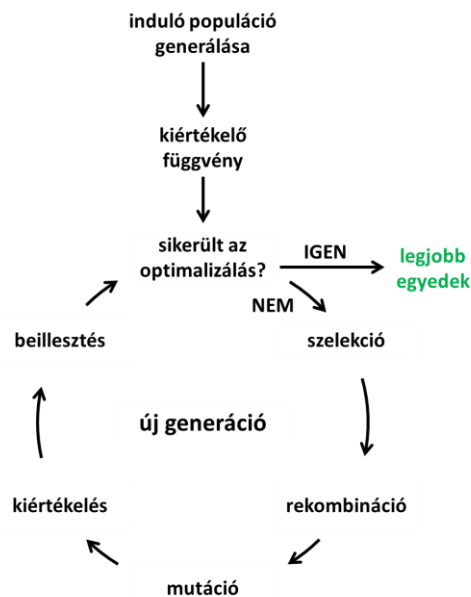
A változók kiválasztása fontos lépés a többváltozós elemzésben, mivel a nem informatív változók eltávolítása jobb becslési paramétereket és egyszerűbb modelleket eredményez (JIE et al. 2013). A spektrális adatok nagy száma gyakran bonyolulttá teszi a függő változók előrejelzését, ezért célszerű spektrális tartományokat vagy akár egy-egy hullámszámot kiválasztani a modellépítéshez. Erre azért van szükség, mert azok a tartományok, amelyek nem tartalmaznak releváns információt a vizsgált függő változóval kapcsolatban, esetleg túl zajosak, torzíthatják a regressziós modelleket. A változókiválasztási módszereknek az alap gondolata, hogy egy kisebb számú változó kiválasztásával az eredeti adatok értelmezése egyszerűbbé válik. Számos esetben alátámasztották (JIE et al. 2013; CARVALHO et al. 2021), hogy a különböző változókiválasztási

módszerek alkalmazásával a NIRS módszeren alapuló modellek előrejelző képessége növelhető, ill. csökkenthető a modellek komplexitása. Ezek közé a módszerek közé tartozik a változó-fontosság paramétere (*variable importance of projection*, VIP), az intervallum PLSR (iPLSR) és a genetikus algoritmus (GA) is. Természetesen fontos azt is figyelembe venni, hogy a vizsgált vegyület szerkezetéből adódóan, milyen abszorpciós sávokkal rendelkezik.

A **VIP** az egyes PLS-összetevők súlyozásával határozza meg az egyes változók fontosságát. A változók VIP-pontszámuk alapján csökkenő sorrendbe rendezhetők. A magasabb VIP-értékkel rendelkező változó nagyobb fontosságot képvisel. Az ilyen típusú változók esetében a PLS paraméteralapú módszerek gyakran adnak egy partíciós vonalat, hogy kizárják azokat a változókat, amelyek egy meghatározott küszöbértéknél kisebbek. Így gyorsak és egyszerűek (YUN et al. 2019). Az összes változó átlagos VIP-pontszáma 1, ezért a változók kiválasztására jellemzően egy "nagyobb, mint 1" kiválasztási szabályt alkalmaznak (SØRENSEN et al. 2021).

Az **iPLSR** egy olyan változó kiválasztáson alapuló módszer (NØRGAARD et al. 2000), amely átfogó képet ad a különböző spektrális tartományok releváns információtartalmáról. Az intervallumok kiválasztásán alapuló módszerek párhuzamba hozhatók a molekuláris kémiával is, mivel a molekuláris spektrumokat folytonos sávok, nem pedig diszkrét válaszjelek jellemzik (XIAOBO et al. 2010). Ez a spektrális adatkészlet erős korreláltságában is megmutatkozik. A módszer lényege, hogy a spektrális adatokat (tartományt) egyenlő szélességű intervallumokra osztjuk fel, amely során különböző számú szegmenst hozunk létre. A létrehozható szegmensek száma opcionális, azonban a leggyakrabban a 10, 20 és 40 szegmenses felosztásokat alkalmazzák (XIAOBO et al. 2010).

A **genetikus algoritmus (GA)** egy globális keresési technika, amelyet a természetes szelekció általi evolúció biológiai mechanizmusai ihlettek. Ebben az algoritmusban egy optimalizálási probléma minden lehetséges megoldását egy fejlődő populációhoz tartozó egyed képviseli. Az algoritmus a természetes szelekció folyamatát tükrözi, ahol a legmegfelelőbb egyedek kerülnek kiválasztásra a szaporodáshoz, hogy a következő generáció utódait hozzák létre. A GA egy véletlenszerű mintavételre alapuló módszer a változók kiválasztására (BROWN et al. 2020). A folyamat áttekintése érdekében, a különböző optimalizálási lépéseket a 10. ábra szemlélteti. Két szakasz különböztethető meg: a kezdeti és az evolúciós szakasz.



10. ábra: A genetikus algoritmus működésének sematikus ábrája (LAXALDE et al. 2011)

A kezdeti szakasz az egyedek (megoldások) első generációjának létrehozásából áll, amelyet a bitértékek véletlenszerű sorsolásával hoznak létre. A populáció mérete X egyed volt. Az evolúciós szakasz öt lépésből áll: a kiértékelés, a szelekció, a rekombináció, a mutáció és az újbóli beillesztés. Ezeket a lépéseket addig ismételjük, amíg az optimalizálási kritériumok valamelyike nem teljesül. Az evolúciós szakasz első lépése az egyedek kiértékelése egy rátermettségi függvény (*fitness function*) segítségével (LAXALDE et al. 2011). A PLS-modell optimalizálása esetén a cél a becsült hiba minimalizálása. A modell előrejelző képességét a kereszt-ellenőrzés átlagos négyzetes hibájának négyzetgyöke (RMSECV) alapján értékeljük.

4. Anyag és módszer

4.1. Vizsgált minták

4.1.1. Snack termékek

A mérések során 155 kereskedelmi forgalomban kapható snack terméket vizsgáltam. Ezek kiválasztásánál törekedtem a változatosságra, mind az alkalmazott gyártási technológia, mind a származási hely tekintetében. Huszonöt ország termékeit szereztem be, melyeket főként Európában állítottak elő, vannak azonban japán, kanadai és Afrikából származó; olajban sült, extrudált, puffasztott és szárított termékek is. A mintákra vonatkozó általános adatokat és a snackek fontosabb jellemzőit a 3. táblázat, valamint az M2. melléklet tartalmazza.

3. táblázat: A snack minták főbb jellemzői (zárójelben a mintaszám látható)

Származási hely	Amerikai Egyesült Államok (2), Ausztria (3), Belgium (14), Bulgária (2), Csehország (14), Egyesült Királyság (14), Egyiptom (3), Franciaország (2), Görögország (19), Írország (2), Japán (1), Kanada (4), Lengyelország (20), Líbia (2), Magyarország (8), Németország (13), Nigéria (5), Olaszország (4), Portugália (1), Románia (2), Spanyolország (5), Svédország (3), Szerbia (1), Szlovákia (4), Szlovénia (1)
Technológia	extrudált (50), olajban sült burgonya chips (60), lúgosan főzött tortilla chips (31), puffasztott (4), szárított (3), egyéb (8)
Alapanyag	burgonya, kukorica, búza, rizs és szója különböző formái
Sütőolaj	napraforgó (57), pálma (38), repce (16), olíva (3), kukorica (3), növényi olajok keverékei (27)
Egyéb snackek	banán, kókusz, cékla, paszternák chips, sült sertésbőr

4.1.2. Kávé

4.1.2.1. A zöld kávé minták

A kísérletekhez ötven zöld kávé szereztem be a Semiramis Kft-től (Budapest, Magyarország). Negyvennyolc arabica (egy koffeinmentes) és két robusta mintát vizsgáltam húsz különböző országból (4. táblázat). A kávécseresznye feldolgozása során, a legtöbb minta esetében (harminckilenc) a nedves eljárást alkalmazták, azonban száraz, félszáraz, félig mosott és „honey” eljárással készült kávékat is vizsgáltam. A minták általános jellemzőit az M3. mellékletben foglaltam össze. Az egyes mintákat, származási helyük alapján négy régióba soroltam be (Afrika, Ázsia, Dél-Amerika, Közép-Amerika).

4. táblázat: A kávéminták származási helyére vonatkozó általános adatok (zárójelben a mintaszámok láthatók)

Származási hely (ország)	Brazília (5), Costa Rica (1), Ecuador (1), Etiópia (4), Guatemala (4), Honduras (1), India (4), Indonézia (5), Jamaica (1), Kelet-Timor (1), Kenya (3), Kolumbia (5), Mexikó (2), Nepál (1), Nicaragua (1), Panama (2), Pápua-Új-Guinea (2), Tanzánia (1), Uganda (4), Zambia (2)
Származási hely (régió)	Afrika (14), Ázsia (13), Dél-Amerika (10), Közép-Amerika (13)

4.1.2.2. Kávépörkölés

Az ötven eredeti mintát kisléptékben, egy előre meghatározott protokollt alkalmazva pörkölttem meg. Három szintet állítottam be az első reccsenést figyelembe véve (pörkölés kezdete), melyek a világos, közepes és sötét pörkölés voltak. Ezáltal 150 mintát hoztam létre a további vizsgálatokhoz, melyek tárolása a vizsgálatok elvégzéséig -80°C -on történt. Mivel a különböző kávék eltérően viselkednek pörkölés során, az egyes szintek közötti határok elmosódhatnak. A világos pörkölésű kávék előállításához a folyamatot 40 másodperccel az első reccsenés után állítottam le, míg a másik két szint esetében az idő 60 és 90 másodpercre módosult. A kiindulási hőmérséklet kb. 150°C , a végső hőmérséklet 190°C volt. Az adott szintekhez minden esetben 300 g kávéval használtam fel, melyeket egy Probat minta pörkölővel (Leogap, Curitiba, PR, Brazil) pörkölttem meg. A protokoll beállításához, kávépörkölésben jártas szakemberek nyújtottak segítséget (Semiramis Kft., Budapest, Magyarország).

4.2. Vegyszerek és reagensek

A snack termékek általános összetételének meghatározása során alkalmazott vegyszerek és reagensek közül a bórsav, a cink-szulfát, az etil-alkohol, kálium-szulfát, a metilvörös- és brómkezelzöld indikátorok, a nátrium-klorid, nátrium-nitrát és a réz-szulfát a Reanal (Budapest, Magyarország) cégtől kerültek beszerzésre. Ezen kívül felhasználtam még izobutanolt, kálium-hexacianoferrátot, kálium-jodidot, kálium-nátrium-tartarátot, keményítőt, tömény kénsavat (VWR, Radnor, PA, USA), nátrium-hidroxidot (Lachner), nátrium-tioszulfátot, $40-60^{\circ}\text{C}$ forráspont tartományú petrolétert és tömény sósavat. Minden felhasznált vegyszer és reagens analitikai tisztaságú volt.

A kávéminták analitikai vizsgálata során, a folyadékkromatográfiás mérésekhez a következő referenciaanyagok kerültek beszerzésre: 1,3,7-trimetilxantin (koffein; 99%), kinasav-3-O-kávésav

(3-CQA; $\geq 98\%$), a kinasav-4-O-kávésav (4-CQA; $\geq 98\%$), a kinasav-5-O-kávésav (5-CQA; $\geq 98\%$), a kinasav-3,5-O-dikávésav (3,5-diCQA; $> 95\%$), a kinasav-4,5-O-dikávésav (4,5-diCQA; $> 95\%$) (Sigma-Aldrich, Schnellendorf, Németország) és a kinasav-3-O-ferulasav (3-FQA; 98%) (AOBIOUS INC, Gloucester, MA, USA). A kromatográfiai méréshez ezen kívül metanolt, magnézium-oxidot (Reanal, Budapest, Magyarország), hangyasavat (Honeywell Research Chemicals, Charlotte, NC, USA) és nagy tisztaságú ioncserélt vizet (Milli-Q-Plus ultrapure rendszer, Merck-Millipore, Milford, MA, USA) használtam fel.

4.3. Referencia módszerek

4.3.1. Snack termékek beltartalmi paramétereinek meghatározása

A snack termékek minőségének vizsgálatához elengedhetetlen a makrokomponensek átlagos mennyiségének ismerete. Erre a célra hagyományos analitikai módszereket alkalmaztam, melyeket kivitelezésük és elterjedtségük miatt nem mutatok be részletesen. A makrokomponensekre vonatkozó eredményeket 100 g friss termékre adtam meg. A mintákat az analitikai vizsgálatok elvégzése előtt konyhai daráló (SENCOR SCG 1050, Japán) segítségével aprítottam és a mérések elvégzéséig 4°C-on, sötétben tároltam.

A **zsírtartalom** meghatározása a 20501-1:2007 Magyar Szabvány (MSZ) referencia módszere alapján történt, petroléteres hideg extrakció alkalmazásával.

A minták **szénhidrát** és **cukor** tartalmát az MSZ 6369-12:1979 szabvány alapján határoztam meg, Schoorl-módszerrel.

A **nyersfehérje** tartalom meghatározása a Kjeldahl módszer alapján történt, az MSZ 20501-1:2007 szabványban leírtaknak megfelelően. Az eredmények számítása során, az alapvető fehérjeforrás figyelembevételével választottam ki a megfelelő szorzófaktorot.

A **sótartalom** vizsgálatát direkt potenciometriás módszerrel hajtottam végre, amely az elektródpotenciálok mérésén alapuló elektroanalitikai eljárás. A vizsgált komponens kémiai aktivitása és az elektród potenciálja (E) közötti kapcsolatot a Nernst-egyenlet írja le. A mérés során állandó ionerősséget biztosítottam, ezáltal az egyenlet egyszerűbb formában is felírható, 298,15 K hőmérsékleten:

$$E = E_0 - \frac{RT}{nF} \ln c \rightarrow E_{25^\circ\text{C}} = E_0 + \frac{0,059}{n} \lg c$$

ahol

n az elektródreakció során bekövetkező töltésszám változás,

R az egyetemes gázállandó (8,314 J/molK)

T az abszolút hőmérséklet (298,15 K)

F Faraday állandó (96 485 C/mol)

A snack termékek **energiatartalmának** kiszámításához a mért átlagos tápértékeket és a 1169/2011/EU rendelet XIV. mellékletében meghatározott átszámítási faktorokat használtam fel, amelyek megfelelnek 1 g tápanyag energiatartalmának. Az átszámítási faktorok a következők voltak a különböző makrotápanyagok esetén: szénhidrát 17 kJ/g, fehérje 17 kJ/g, zsír 37 kJ/g.

4.3.2. Zöld- és pörkölt kávéminták vizsgálata

4.3.2.1. Minta-előkészítés analitikai vizsgálatokhoz

A homogenitás biztosítása érdekében a kávémintákat ledaráltam. Mivel a zöldkávét meglehetősen kemény, először késes aprítóval (Sencor, Japán) csökkentettem a szemcseméretet, majd egy konyhai darálóval (Bosch TSM6A011W, Robert Bosch Hausgeräte GmbH, München, Németország) őröltem tovább. Ezután az őrleményt egy 510 µm pórusátmérőjű szitán (Microtrac Retsch GmbH, Haan, Németország) átszitáltam úgy, hogy annak legalább az 50%-a az átesett frakcióban legyen. A pörkölt kávéminták esetében az aprítási lépést kihagytam.

A mérési eredményeket minden esetben friss tömegre vonatkoztatva adtam meg.

4.3.2.2. Savasság meghatározása pH-mérés alapján

A minták savasságát azok pH értékének meghatározásával adtam meg, mivel a kávé érzékelt savassága annak pH értékével erős korrelációt mutat. A mérést BROLLO és mtsi. (2009) által kidolgozott módszer alapján végeztem el, melyhez 5 g őrölt mintát mértem be tárapontossággal főzőpohárba, majd 50 ml 95°C-os ioncserélt vizet adtam hozzá. Ezután egy percig állni hagytam a mintákat, majd mágneses keverő (Radelkis Type OP-951, Radelkis, Budapest, Magyarország) segítségével 3 percig kevertettem. Az extraktumot redős szűrőpapíron szűrtem Erlenmeyer lombikba, majd hűlés után a szűrlet pH értékét egy HANNA HI110 (Hanna Instruments, Rhode Island, USA) típusú pH mérő műszerrel határoztam meg.

4.3.2.3. Koffeintartalom meghatározása

A koffeintartalom meghatározását az ISO 20481:2008 szabvány alapján, kisebb módosításokkal, HPLC-UV módszer alkalmazásával hajtottam végre. Ehhez a szitált mintából kb. 0,2000 g-ot mértem be, 1 g magnézium-oxidot tartalmazó 50 ml-es centrifugacsőbe. Hozzáadtam 25 ml ioncserélt vizet és 40 percre 80°C-os rázófeltétes vízfürdőbe helyeztem a mintákat, majd folyóvíz alatt szobahőmérsékletűre hűtöttem. Az extraktumokat kiegészítettem 50 ml végtérfogatra, majd rövid ülepítés után a felülúszóból 1400 µl-t mikrocentrifuga csőbe pipettáztam. Szobahőmérsékleten 10 percig 10000 rpm-en centrifugáltam (Hettich Universal 22R, Andreas Hettich GmbH & Co., Tuttlingen, Németország) a mintákat, majd a felülúszóból 500 µl-t kromatográfias üvegbe pipettáztam.

Az elválasztás paraméterei a következők voltak:

- **készülék és szoftver:** Agilent 1200 HPLC UV detektorral, Openlab CDS (Agilent Technologies, Santa Carla, Kalifornia);
- **elúció:** izokratikus; 24/76 v/v% metanol-víz
- **kolonna:** C₁₈, Inertsil ODS-2, 5 µm töltetméret; 150x3 mm
- **hőmérséklet:** 40°C
- **áramlási sebesség:** 0,8 ml/perc
- **detektálási hullámhossz:** 272 nm

4.3.2.4. Klorogénsavak meghatározása

A klorogénsavak extrakciójához 1,0000 g szitált mintát mértem be egy 50 ml-es centrifugacsőbe, amihez 50 v/v% metanol-víz elegyet adtam. A mintákat 30 percre ultrahangos kádba helyeztem, majd 10 percig 6000 rpm-en centrifugáltam, végül a felülúszóból kb. 2 ml-t mikrocentrifuga csőbe szűrtem, 0,22 µm pórusátmérőjű, hidrofil PTFE fecskendőszűrővel. A mintákat ioncserélt vízzel százszorosára hígítottam.

A kávébabban megtalálható klorogénsavak mennyiségi meghatározása egy Shimadzu Nexera X2 szériájú HPLC-vel (Shimadzu, Kiotó, Japán) kapcsolt AB SCIEX 6500+ hármass kvadrupól tömegspektrométerrel (SCIEX, Framingham, MA, USA) és annak saját szoftverével az Analyst 1.7-tel történt. A komponensek detektálásához használt tömegspektrométeres paramétereket optimálással állítottam be. Mivel a mért vegyületekre jellemző a szerkezeti izoméria, izobár zavarás lép fel. Ezáltal a beállított ionátmenetek nem teljesen szelektívek az adott vegyületre (M4.), például a 353 > 191-es ionátmeneten mind a három CQA és az 515 > 353-as ionátmeneten

mindkét diCQA) detektálható. Ebből adódóan nélkülözhetetlen volt, hogy a kromatográfiás elválasztás során megvalósuljon az egymást zavaró komponensek alapvonalis elválasztása.

A komponensek elválasztáshoz ACE Excel 3 μm , C18-PFP kolonnát (100x3 mm) használtam. Mozgófázisként 0,1 v/v% hangyasavat tartalmazó vizet (A) és 0,1 v/v% hangyasavat tartalmazó metanolt (B) használtam. Az áramlási sebesség: 0,5 ml/perc, míg az injektált térfogat 1 μl . volt. A kromatográfiás elválasztás gradiens elúcióval történt, a következő módon: 0-1 perc, 10% B; 1-10 perc, 50% B; 10-13 perc, 100% B; 13-14 perc, 100% B; 14,01-17 perc, 10% B. Az eredmények kiértékeléséhez a MultiQuant 3.0.2 (SCIEX, Framingham, MA, USA) szoftvert használtam.

A komponensek detektálásához használt műszer- és komponensfüggő paramétereket az M4. melléklet tartalmazza.

4.3.2.5. Tristimulusos színmérés

A zöld- és pörkölt kávéminták színét leíró paramétereket Konica-Minolta CR410 (Konica Minolta Sensing Inc., Osaka, Japán) tristimulusos színmérő műszerrel rögzítettem a CIE $L^*a^*b^*$ színtérben. A mintákat eredeti formájukban kb. 15 g-os adagokban vizsgáltam, a mért felület átmérője pedig 50 mm volt. A megvilágítás szabványosított C fényforrással a mintára merőlegesen történt. A diffúz módon visszavert fény detektálása pedig 2° -os megfigyelési szögnél valósult meg. Az X, Y és Z értékek mérése után a készülék kiszámolja az L^* , a^* és b^* színekoordinátákat a következő egyenletek alapján (SCHANDA 2007):

$$L^* = 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16,$$
$$a^* = 500 \left(\left(\frac{X}{X_n} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \right),$$
$$b^* = 200 \left(\left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{3}} \right),$$

ahol X, Y és Z a vizsgált minta trikromatikus értékei, X_n , Y_n és Z_n az abszolút fehér felület mérőszámai ($X_n=98,07$, $Y_n=100,0$, $Z_n=118,22$).

4.3.2.6. Érzékszervi bírálat

A 150 pörkölt kávé minta érzékszervi bírálata több száz órás bírálói tapasztalattal rendelkező bírálópanel segítségével (hat fő) történt. A kávé érzékszervi tulajdonságai közül a savasság és a keserűség képezték a vizsgálat tárgyát. A kóstolás során a tulajdonságok értékelése egy tízpontos skálán történt (1: gyenge, 10: intenzív). A bírálat papír alapú bírálati lapokon történt, melyeket a bírálat után összesítettem. A szemes kávé őrléséhez egy Mahlkönig EK43 típusú darálót (MAHLKÖNIG GmbH & Co., Hamburg, Németország) használtam. Az őrlési fokozatot a kávékészítés módjához igazítva hatosra állítottam be, ami a filterkávékhoz ajánlott (kb. 400 µm-es szemcseméret maximum). A kávé-víz arány kb. 1:17 volt (60 g: 1000 ml). Az italkészítéshez Bunn Smart Wave filteres kávéfőzőt (Bunn-O-Matic Co., Illinois, USA) alkalmaztam. A bírálat során a mintákat számkódokkal láttam el, majd a bírálók blokkonként (6 minta/blokk) értékelték a mintákat. A blokkok között néhány perces szünetet tartottunk, közben a bírálók ásványvizet és natúr krékert fogyasztottak az ízsemlegesítéshez.

4.3.3. FT-NIR spektroszkópia

A spektrális adatok rögzítéséhez egy Bruker MPATM többfunkciós FT-NIR spektrométert (Bruker, Ettlingen, Németország) használtam. A spektrumok felvétele diffúz reflexiós mérési módban történt, melyhez a következő paramétereket állítottam be:

- optikai felbontás: 16 cm⁻¹
- alspektrumok száma: 32
- adatpontok száma: 1154
- tartomány: 12500-3600 cm⁻¹
- detektor: ólom-szulfid (PbS)
- adatfelvételi sebesség: 10 kHz
- szoftver: OPUS 7.2 (Bruker, Ettlingen, Németország)

A homogenitás növelése érdekében forgatható kvarc küvettát (Ø 85 mm) használtam. A referencia háttér felvétele arannyal borított integráló gömb alkalmazásával történt. A snack termékekből 40-60 g darált mintát helyeztem a küvettába, míg a kávé mintákból kb. 50 g-ot eredeti, szemes formában vizsgáltam. Az állandó mintamennyiséggel az optikai úthosszból adódó eltérések csökkenthetők. A snackek esetében három, a kávé mintáknál pedig hat párhuzamos spektrumfelvételt hajtottam végre. A kiértékeléshez az átlagspektrumok kerültek felhasználásra.

4.3.4. Kemometriai módszerek

4.3.4.1. Kruskal-Wallis próba

A kémiai és érzékszervi adatokon leíró statisztikát futtattam le, külön-külön a teljes adatkészletre és a négy termesztési régióból származó normál arabika kávékra vonatkozóan. Az utóbbi adatkészlet (negyvenhét minta) esetében azt is vizsgáltam, hogy az egyes régiók között van-e szignifikáns különbség a vizsgált paraméterekkel összefüggésben. Ehhez a referencia adatok normalitását a Shapiro-Wilk-próbával ellenőriztem ($\alpha=0.05$). A csoportok közötti különbségek vizsgálatára a Kruskal-Wallis próbát alkalmaztam ($\alpha=0.05$) (SAJTOS és MITEV 2007).

4.3.4.2. Főkomponens elemzés (*principal component analysis, PCA*)

A főkomponens elemzés (PCA) egy adatredukációs eljárás, ami az adatkészletben rejlő esetleges mintázatok feltárására és a kiugró értékek vizsgálatára egyaránt alkalmas (WOLD et al. 1987). Emellett vizuálisan is megjeleníti a minták és változók közötti kapcsolatokat. A módszer alap gondolata a "látens változók" (azaz főkomponensek) kiszámítása, amelyek az eredeti változók lineáris kombinációjaként jönnek létre. A látens változók (*latent variables, LV*) definíció szerint ortonormálisak egymáshoz képest. A nagyobb sajátértékkel rendelkező főkomponensek tartalmazzák az eredeti adathalmaz legtöbb információját.

A snackek esetén a PCA-t a spektrális adatok, míg a kávé vizsgálatánál mind a referencia, mind a spektrális adatok feltáró elemzésére alkalmaztam. A kiugró értékek meghatározásához a Hotelling-féle T^2 statisztika és az F-reziduális értékeit használtam. Az adatokat minden esetben az átlaggal centráltam. A modellek validálásához véletlenszerű, ötszögemes kereszt-ellenőrzést alkalmaztam.

4.3.4.3. Adat-előkezelési módszerek

A spektrális adatok transzformációja sor-normalizálással (*standard normal variate, SNV*), átlagos szóráskorrekcióval (MSC), deriválással (első és második derivált, *first and second derivative, FD* és SD), ill. az FD vagy SD és az SNV algoritmusok kombinációjának futtatásával történt. A különböző adat-előkezelési módszereket mindkét termékcsoporthoz alkalmaztam. Ezek mellett, a kávé esetén az ortogonális jelkorrekciót (*orthogonal signal correction, OSC*) is felhasználtam.

4.3.4.4. Osztályozási módszerek

Az adatok elemzése során felügyelt tanítású algoritmusokat alkalmaztam a snack termékek és a kávé különböző minőségi jellemzőinek vizsgálatára. A módszerek általános jellemzőit a 5. táblázatban foglaltam össze és a két termékcsoporthoz vonatkozóan külön-külön mutatom be.

5. táblázat: Az osztályozási módszerek rövid összefoglalása (KOTSIANTIS 2007; GERE et al. 2016)

	számítási idő	bemeneti adatok	főbb előnyök	főbb hátrányok	általános pontosság	magyarázó képesség, osztályozás átláthatósága	típus
DA	kicsi	folyamatos	maximális elkülönítés biztosítása	Gauss eloszlást feltételez, túllillesztés	magas	jó	statisztikai tanulás
DT	kicsi /közepes	folyamatos/ diszkrét	az osztályozás küszöbértékeken alapul, hiányzó adatok kezelése	túl nagy fák, túllillesztés	közepes	kiváló	logika-alapú
ensemble	nagy	folyamatos/ diszkrét	komplex, zajos adatok kezelése	hajlamosak az alul/túllillesztésre	közepes	átlagos	logika-alapú
k-NN	nagy	folyamatos	egyszerűség, könnyű megvalósítás	érzékeny a zajra, a változók és a megfigyelések számára	magas	jó	eset-alapú
MLP	nagy	folyamatos	hibatűrés	nehéz értelmezni, nincs garantált megoldás	magas	átlagos	perceptron-alapú
NB	közepes	folyamatos/ diszkrét	jól kezeli a nagy adatmennyiséget, tolerálja a zajos adatokat	nem rendelkezik minimum hibaarányal	közepes	kiváló	statisztikai tanulás
PLS-DA	kicsi	folyamatos/ bináris	robosztus a hiányzó adatok esetén, kezeli a multikollinearitással	az egyes megfigyelések több kategóriába is besorolhatók	magas	kiváló	statisztikai tanulás
RF	közepes	folyamatos/ diszkrét	nagy adatkészletek és hiányzó értékek kezelése	modellek komplexitása	magas	jó	logika-alapú
SVM	kicsi/ közepes	folyamatos/ bináris	hatékony nagy dimenziójú adatkészlet esetén, robusztus	érzékeny a zajos adatokra, nagy adatkészletekre	magas	átlagos	kernel-alapú

DA: diszkriminancia elemzés; DT: döntési fák; k-NN: legközelebbi szomszéd módszer; MLP: *multilayer perceptron neural network*, többrétegű perceptron neurális hálózat; NB: Naïve Bayes; PLS-DA: részleges legkisebb négyzetek elve-diszkriminancia elemzés; RF: véletlen erdő; SVM: támogató vektorgép

Snack termékek

A snack termékek esetén a minták sokféleségéből adódóan, különböző lehetőségek álltak rendelkezésre azok osztályozására. Az eltérő osztályozási rendszerekben a felhasznált minták száma nem volt azonos, ezért minden olyan csoportot kizártam, ami ötnél kevesebb terméket tartalmazott. A minták osztályozása a felhasznált alapanyag és sütőolaj, illetve az alkalmazott

gyártási technológia és a származási hely alapján történt. Az osztályozási szempontok jellemzőit az 6. táblázatban foglaltam össze.

6. táblázat: A snackminták összefoglalása a különböző osztályozási szempontok alapján

osztályozási szempont	Csoportok	minták száma
sütőolaj (1. adatkészlet)	Napraforgó	56
	Pálma	38
alapanyag (2. adatkészlet)	Burgonya	71
	Kukorica	55
	Belgium	14
származási hely (3. adatkészlet)	Cseh Köztársaság	14
	Lengyelország	19
	Egyesült Királyság	12
gyártási technológia (4. adatkészlet)	olajban sütött	60
	lúgosan főzött (tortilla)	31
	Extrudált	50

Véletlen erdő algoritmus (random forest, RF)

A véletlen erdő egy fa alapú módszer, ahol az eredeti adathalmaz alapján sok döntési fa kerül kiszámításra, melyek mindegyike egy adott osztályozást jelez előre (BREIMAN 2001). A döntési fák bináris felosztásokból állnak a legfontosabb változók adott értékeinél. A végső osztályozás szavazáson alapul, ugyanis a fák önmagukban gyenge előrejelzők, ugyanakkor együtt már egy olyan egységet alkotnak, amelynek általános teljesítménye nagymértékben javul. A véletlen erdő algoritmust az információnyereség arányával, mint osztási kritériummal alkalmaztam és a fák számát 100-ra állítottam be.

Neurális hálózat-alapú többrétegű előrecsatolt reziliens backpropagation hálózat (RPropMLP)

A backpropagation a legszélesebb körben használt algoritmus a többrétegű előrecsatolt hálózatok felügyelt tanítására. A backpropagation tanulási algoritmus alapgondolata a láncszabály ismételt alkalmazása, amely során egy tetszőleges hibafüggvénnyel számítja ki a hálózat egyes súlyainak hatását (RIEDMILLER és BRAUN 1993). Az osztályozáshoz minden esetben egy rejtett réteget és tíz rejtett neuront használtam.

Részleges legkisebb-négyzetek elve diszkriminancia analízis (PLS-DA)

A PLS-DA (*partial least squares-discriminant analysis*) az osztályozási és regressziós problémák megoldásának egyik klasszikus eszköze. A PLS-DA alapeleme a többszörös lineáris regresszió

robustusabb módon történő megvalósítása. Az új látens változókat (mint például a PCA esetében) az eredeti változók lineáris kombinációiként számítjuk ki úgy, hogy az Y (cél) változó független (kategorikus vagy folytonos) változó. Az új, látens változókat a modellezési fázisban alkalmazzuk. A PLS-DA-t az első húsz PCA score-vektor bemeneti mátrixként használva futtatam le minden egyes adatkészleten.

A teljesítmény értékelése és összehasonlítása

A snack termékek különböző szempont alapján történő osztályozásához négy adatkészletet hoztam létre. Ezeket minden esetben két részre osztottam 70:30 arányban (tanítási- és teszt adatkészlet), ahol a minták 70%-a véletlenszerűen, rétegzett mintavétellel (azaz a csoportok eredeti arányainak megőrzésével) lett kiválasztva. Ez a folyamat minden algoritmusnál hússzor ismétlődött (az eredmények összehasonlíthatósága érdekében minden esetben ugyanazt a véletlen magot használtam). A száz minta alatti adathalmazoknál háromszegmenses-, míg az ennél magasabb mintaszámnál ötszegmenses kereszt-ellenőrzést alkalmaztam. Mind a kereszt-ellenőrzésnél, mind a teszt ellenőrzésnél randomizációs eljárást használtam. A kidolgozott és a random modellek pontossága közötti különbségek szignifikanciáját kétmintás t -próbák alapján állapítottam meg. A modellek teljesítményének meghatározására és összehasonlítására az egyes osztályozási adathalmazok és algoritmusok esetében a pontosság (*accuracy*, ACC), a valódi pozitív arány (*true positive rate*, TPR), a valódi negatív arány (*true negative rate*, TNR) és a vevő működési karakterisztika görbe alatti terület (*receiver operating characteristic curve area under the curve*, ROC AUC) értékei lettek felhasználva. Röviden, a ROC-görbék az érzékenység (TPR) és a specifikusság (hamis pozitív arány, FPR) 2D-s ábrái minden lehetséges valószínűségi küszöbértékre. A ROC-görbék a [0; 0] ponttól [1; 1] pontig futnak, az átlós vonal pedig a véletlen besorolásnak felel meg. Azok a modellek, amelyek ROC-görbéje e vonal felett/alatt van, jobbak/rosszabbak a véletlen besorolásnál. A görbe alatti terület (AUC) értéke 0 és 1 között van, ahol 1, 0,5 és 0 megfelel a tökéletes, a véletlenszerű és a teljesen hibás osztályozásnak. Mivel a ROC-görbék egyszerű teljesítménymetriák a kétosztályos klasszifikációs feladatokhoz, több osztály esetén ezeket általánosítani kell. Az "egy az összes ellen" általánosítási sémát alkalmaztam, RÁCZ és munkatársai (2016) korábbi publikációja alapján. A modellek összehasonlításához a húsz futtatásból származó egyéb teljesítményparaméterek átlagértékeit is felhasználtam.

Kávé

A zöld kávé minőségét, ezáltal az árát jelentősen befolyásolja a származási hely, így az eredetazonosításnak fontos szerepe lehet a minőségellenőrzésben. A minták csoportosítását négy adatkészleten végeztem el, amit a 7. táblázatban foglaltam össze. A 2. és 3. adatkészletnél az Amerika alcsoport a Dél- és Közép-Amerikából származó mintákat tartalmazza. A csoportosításhoz a spektrálisan kiugró mintákat nem használtam fel, ezáltal a két ázsiai robusta, ill. két dél-amerikai minta (koffeinmentes, szennyezett felületű) került kizárásra a kiértékelésből.

7. táblázat: A kávéminták összefoglalása a különböző osztályozási szempontok alapján

osztályozási rendszer	alcsoportok	minták száma
négy régió (1. adatkészlet)	Afrika	14
	Ázsia	11
	Dél-Amerika	8
	Közép-Amerika	13
három régió (2. adatkészlet)	Afrika	14
	Amerika	20
	Ázsia	11
két régió (3. adatkészlet)	Afrika	14
	Amerika	20
két régió (4. adatkészlet)	Afrika	14
	Ázsia	11

A zöld kávéminták osztályozását a MATLAB szoftver Classification Learner alkalmazása segítségével hajtottam végre, mellyel számos modell létrehozására van lehetőség, beleértve a döntési fákat (DT), diszkriminancia elemzést (DA), a támogató vektor gépeket (SVM), a legközelebbi szomszéd (k -NN)- és az ensemble (együttes) módszereket, ill. a neurális hálózat alapú osztályozást.

Döntési fák (decision trees, DT)

A döntési fák (DT) algoritmust gyakran alkalmazzák többváltozós adatok elemzése során, mert egyszerűek, könnyen értelmezhetők, ill. alkalmazhatók minőségi, mennyiségi, diszkrét és folytonos változók vizsgálata során is (HAJJEJ et al. 2022). A döntési fa tulajdonképpen egy folyamatábraszerű struktúra, amelyben minden belső csomópont egy jellemzőre vonatkozó tesztet képvisel, az ágak pedig a jellemzők olyan kapcsolatait, amelyek az adott osztályhoz vezetnek. A gyökértől a levélig vezető utak különböző osztályozási szabályokat jelentenek.

Diszkriminancia elemzés (discriminant analysis, DA)

A diszkriminancia elemzés (DA) azon a feltételezésen alapul, hogy a különböző osztályokba tartozó adatok eltérő Gauss-eloszlásoknak felelnek meg. A fő folyamat először egy osztályozót képez ki egy olyan függvény illesztésére, amely képes megbecsülni az egyes osztályok eloszlásának paramétereit, majd az osztályozót az új minták előrejelzésére használja. A legszélesebb körben használt változata a lineáris diszkriminancia analízis (LDA). Az LDA-ban a legfontosabb lépés egy k -dimenziós térben lévő vetítési hipersík megtalálása, majd a minták különböző osztályainak a hipersíkra való vetítésével az osztályok közötti távolságok maximalizálása és az osztályon belüli távolságok minimalizálása. Egy másik tipikus diszkriminancia elemzés a kvadratikus diszkriminancia-elemzés (*quadratic discriminant analysis* QDA). A QDA és az LDA közötti fő különbség az, hogy a QDA egy kvadratikus döntési határral rendelkező osztályozó (HAN et al. 2019).

Támogató vektor gép (SVM)

A támogató vektor gép (SVM) a kernel módszerek közé tartozik, tehát a megoldásokat kernel-függvények súlyozott összegeként állítja elő (ALTRICHTER et al. 2006) ezáltal lehetővé téve akár a nem lineáris modellezést. Osztályozás esetén az SVM egyszerre minimalizálja az empirikus osztályozási hibát és maximalizálja az osztályok közötti geometriai margót, ami egyedi megoldáshoz vezet. Viszonylag kis adathalmazok esetén is jó általános teljesítmény jellemzi. Az SVM lényege, hogy megkeresi azt a hipersíkot, amely helyesen választja szét az adatokat miközben maximalizálja a legrövidebb távolságot ettől a hipersíktól az egyes osztályok legközelebbi mintáihoz. Ez a távolság határozza meg az adott hipersíkhhoz tartozó margót. SVM esetében az osztályozási szabály tehát az, hogy minden objektumnak a margó határainak jó oldalán kell lennie (DEVOS et al. 2009).

Legközelebbi szomszéd (k -nearest neighbors, k -NN) módszer

A legközelebbi szomszéd (k -NN) módszer alkalmazása során, nagy mennyiségű tanulási adatot használunk. Elméletileg minden egyes pontot egy sokdimenziós térben ábrázolunk, ahol a tér minden egyes tengelye egy-egy változónak felel meg. Ha van egy új (teszt) adatpontunk, akkor meg akarjuk találni a hozzá legközelebbi (azaz a leginkább "hasonló") k legközelebbi szomszédot. A k számot jellemzően az n négyzetgyökeként választjuk, ami a képzési adathalmazban lévő pontok teljes száma. (Tehát ha $n = 400$, akkor $k = 20$). A k -NN alapvetően egyszerű és előnye,

hogy nem parametrikus. Ennek értelmében a módszer akkor is használható, ha a változók kategorikusak, illetve nem követik a normál eloszlást (NADKARNI 2016).

Ensemble módszerek

Az ensemble módszerek a klasszifikációs modell létrehozásához több gyenge gépi tanító algoritmust használnak egyszerre és ezek előrejelzéseit kombinálják. Ezeket a módszereket jellemzően akkor alkalmazzák, amikor a pontosság javítása fontosabb az elemzés szempontjából, mint a modell értelmezésének biztosítása. Az ensemble-módszertan intuitív magyarázata az emberi természetből és abból a tendenciából ered, hogy különböző véleményeket gyűjtünk össze és ezeket mérlegelve és kombinálva összetett döntéseket hozunk. A fő gondolat az, hogy több egyéni vélemény mérlegelése és összevonása jobb eredményt ad, mintha egyetlen egyén véleményét választanánk (SAGI és ROKACH 2018).

Naïve Bayes (NB)

A Naïve Bayes (NB) egy egyszerű tanulási algoritmus, amely a Bayes-tételt használja, azzal az erős feltételezéssel együtt, hogy az attribútumok (változók) feltételezetten függetlenek (WEBB 2010). A Bayes-tétel egy feltételes valószínűség és a fordítottja között állít fel kapcsolatot. Azért nevezik naivnak, mert figyelmen kívül hagyja a paraméterek előzetes eloszlását és feltételezi az összes jellemző és az összes sor függetlenségét (YETURU 2020). Valójában a NB módszer hatékonysága ebből a feltételezésből ered, amely jelentősen leegyszerűsíti a feltételes valószínűségábrázolását és becslését (OUYANG et al. 2017; LI et al. 2020). A Bayes-féle osztályozási sémát használva a NB mérsékelheti a feltételes valószínűség becslésével kapcsolatos komplexitást a feltételes függetlenség feltételezésével, amely csökkenti a becslendő változók számát (PHAM et al. 2021).

Az osztályozáshoz a zöld kávéminták eredeti NIR spektrumait és különböző spektrum transzformációs módszerek alkalmazása után kapott spektrális adatait használtam fel. Az alkalmazott adat-előkezelési módszerek az SNV, az MSC, ill. az FD és az SNV kombinációja voltak. Az algoritmusokat minden esetben a főkomponens elemzés (PCA) által kapott score értékeken futtattam le. Az osztályozáshoz felhasznált főkomponensek számát a magyarázott variancia (hegyomlás ábra) alapján határoztam meg. Minden esetben random hét-szegmenses kereszt-ellenőrzéssel történt a modellek validálása. A modellek optimalálásához a MATLAB

Classification Learner alkalmazását (Mathworks Inc., Massachusetts, USA) használtam. Az eredmények értékelése a snack termékek esetén leírtakkal hasonló módon történt.

4.3.4.5. Részleges legkisebb-négyzetek elve regresszió (PLSR)

A különböző paraméterek mennyiségi becslésére a részleges legkisebb négyzetek elve (PLS) regressziót alkalmaztam mind a két termékcsoporthoz. A létrehozott modellek teljesítményének értékelésére különböző statisztikai mutatókat vettem figyelembe. A validáláshoz random, ötszegmenses kereszt-ellenőrzést, ill. teszt-ellenőrzést is alkalmaztam. Utóbbi esetben az adatkészlet felosztása 70/30 arányban történt, Kennard-Stone algoritmus segítségével.

A determinációs együttható (R^2) megadja, hogy az eredeti változók varianciájának hány százaléka jelenik meg a becslésben. Értéke minden esetben 0 (nincs kapcsolat) és 1 (tökéletes összefüggés) között változik, de megadható százalékban is. A validált modell esetén a kereszt-ellenőrzésnél Q^2 -tel, a teszt-validálásnál pedig Q_t^2 -tel jelöltem. A kalibrációs összefüggés teljes mérési tartományra vonatkozó hibáját, a kalibráció átlagos négyzetes hibájának négyzetgyöke (RMSEC) adja meg. Mértékegysége megegyezik a vizsgált paraméterével. A kereszt-ellenőrzés átlagos négyzetes hibájának négyzetgyöke (RMSECV) annak a pontosságát adja meg, amellyel a minták előrejelzése a validálás során történik (KUMAR 2013). A becslés átlagos négyzetes hibájának négyzetgyöke (RMSEP) szintén a modell pontosságát határozza meg, teszt-ellenőrzés alkalmazása esetében. A torzítás az előre jelzett értékeknek a valós értéktől való szisztematikus eltérése, amely egy adott mérőmódszer alkalmazásának eredménye. A „Rank” megadja a PLS-faktorok (látens változók) számát, amelyeket a becslési modellhez használtak. Az R^2 , Q^2 , RMSEC, RMSEP és RMSECV értékei a jól ismert matematikai összefüggések segítségével lettek meghatározva (KUMAR 2013; NÆS et al. 2017).

Változókiválasztási eljárások

A becslési modellek teljesítményének javítására különböző változókiválasztási módszerek állnak rendelkezésre. Alkalmazásukkal meghatározhatók azok a spektrális tartományok (változók), amelyek a regresszió során vizsgált paraméterrel leginkább összefüggésben lehetnek. Ezáltal csökkenthető a modellek komplexitása és a szükséges számítási kapacitást is. A doktori munkámban háromféle változókiválasztási eljárást alkalmaztam, melyek az intervallum részleges legkisebb négyzetek elve regresszió (iPLSR), a genetikus algoritmus (GA) és a projekció során a változók fontosságán (*variable importance in projection*, VIP) alapuló módszer voltak.

A snack termékeknek csak az iPLSR került alkalmazásra 10, 20 és 40 szegmens kialakítása mellett. A kiértékeléshez a 11500-3800 cm^{-1} közötti spektrális tartományt vettem alapul, ami 1000 változót

tartalmazott. A PLS regressziót minden szegmensre egyenként lefuttattam és a kapott RMSECV értékeket összehasonlítottam, a teljes spektrális tartományra vonatkozó RMSECV értékkel (avRMSECV). Azokat a szegmenseket, amelyek RMSECV értéke közel van az avRMSECV értékhez, kiválasztottam az intervallum PLSR-hez. Emellett, különböző adat-előkezelési technikákat is alkalmaztam a többváltozós regressziós modellek paramétereinek optimalizálása érdekében: FD, MSC, SNV, SLS, FD+MSC, FD+SNV.

A kávé esetében mindhárom változókiválasztási eljárást alkalmaztam különböző beállításokkal. Az iPLSR esetén 10, 20 és 40 szegmens, a GA-nál pedig 30-as és 50-es ablakszélesség (window width) beállításával futtattam le az algoritmusokat. A VIP score értékek alapján történő változókiválasztás során minden olyan hullámszám kiválasztásra került, melynek értéke nagyobb, mint 1 (az összes változó átlagos VIP score értéke 1). A vizsgált hullámszámtartomány a zöldkávénál $10000\text{-}3800\text{ cm}^{-1}$, míg a pörkölt kávéénál $9000\text{-}3800\text{ cm}^{-1}$ volt, melyek 800, illetve 680 változót tartalmaztak. A $12500\text{-}9000\text{ cm}^{-1}$ tartományban a mért abszorbanciát nagymértékben befolyásolja a minta színe, ami alapvonal emelkedést okoz. Ezáltal az esetleges kémiai információt elfedi.

Az adatok értelmezéséhez különböző statisztikai szoftvereket használtam, mint az SPSS Statistics 23 (IBM, New York, USA), The Unscrambler X 10.4 (CAMO Software, Oslo, Norvégia), MATLAB 2019b (Mathworks Inc., Massachusetts, USA) és PLS Toolbox 9.0 (Eigenvector Research Inc., Manson, WA, USA), KNIME analytics platform 4.0.2 (KNIME AG, Zürich, Svájc).

5. Eredmények és értékelésük

5.1. Snack termékek

5.1.1. Referencia adatok és a spektrális adatok áttekintése

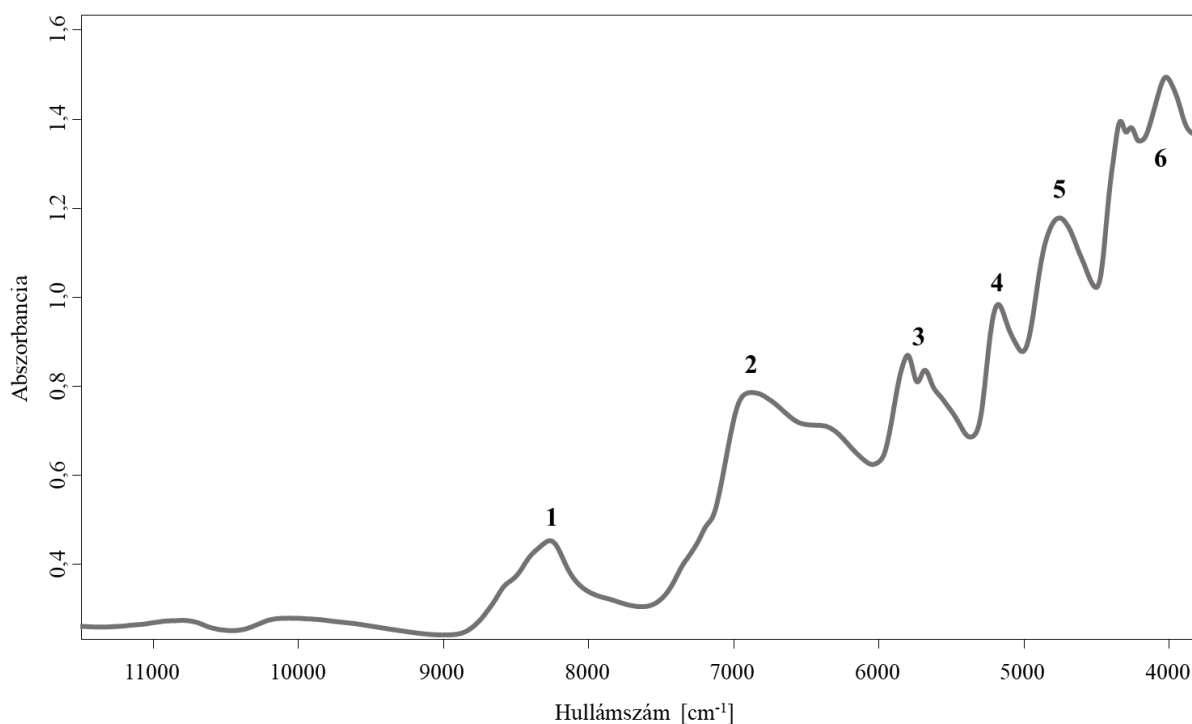
Első lépésben minden esetben a termékek NIR spektrumát rögzítettem, majd elvégeztem a makrokomponensek mennyiségi meghatározását. A minták feldolgozása kétéves periódusban zajlott. Nem minden minta esetében került meghatározásra az összes komponens, ezért a rendelkezésre álló mérési adatok száma eltérő. A snackekre vonatkozó részletes, nominális adatokat a M2. melléklet tartalmazza, azonban a mért komponensekre és a számolt energia tartalomra vonatkozó legfontosabb mennyiségi adatokat a 8. táblázatban foglaltam össze.

8. táblázat: A snackek esetén mért paraméterek fontosabb leíró statisztikai jellemzői

	átlag±szórás	minimum	maximum	medián	mintaszám
zsír [g/100 g]	26,68±9,57	2,24	56,13	27,48	156
szénhidrát [g/100 g]	56,07±10,21	20,51	84,02	55,91	90
cukor [g/100 g]	5,56±3,21	0,94	20,89	5,30	89
fehérje [g/100 g]	9,23±6,33	3,01	40,06	7,51	91
só [g/100 g]	1,56±0,77	0,17	4,44	1,47	91
energia [kJ/100 g]	2029±272	1264	2520	2058	90

A mért értékeket összevettem a termékek csomagolásán feltüntetett adatokkal, amelyek főként a szénhidrát- és fehérjetartalom tekintetében jelentős eltéréseket mutattak figyelembe véve a 1169/2011/EU rendeletben megadott tűréshatárokat. Ez valószínűleg abból adódhat, hogy a tápérték adatok megadása gyakran számításokon és adatbázisokon alapszik, nem analitikai vizsgálatokon. Ezen kívül a termék alapanyagainak nem állandó minősége is okozhatja az eltérést.

Az adatok statisztikai kiértékelése előtt érdemes a NIR spektrum jellegzetes abszorpciós sávjait azonosítani, melyek kapcsolatba hozhatók a vizsgálni kívánt komponensekkel (11. ábra). Az átlagspektrumok két szakaszát ($12800-11500\text{ cm}^{-1}$ és $3800-3500\text{ cm}^{-1}$) levágtam, mivel azok nem hordoztak szisztematikus információt. A 11. ábra az összes minta átlagolt spektrumát mutatja. A jellegzetes rezgési területeket sorszámokkal jelöltem, melyek a snack termékekben található makrokomponensekkel (szénhidrátok, zsírok, fehérjék) vannak összefüggésben (9. táblázat).



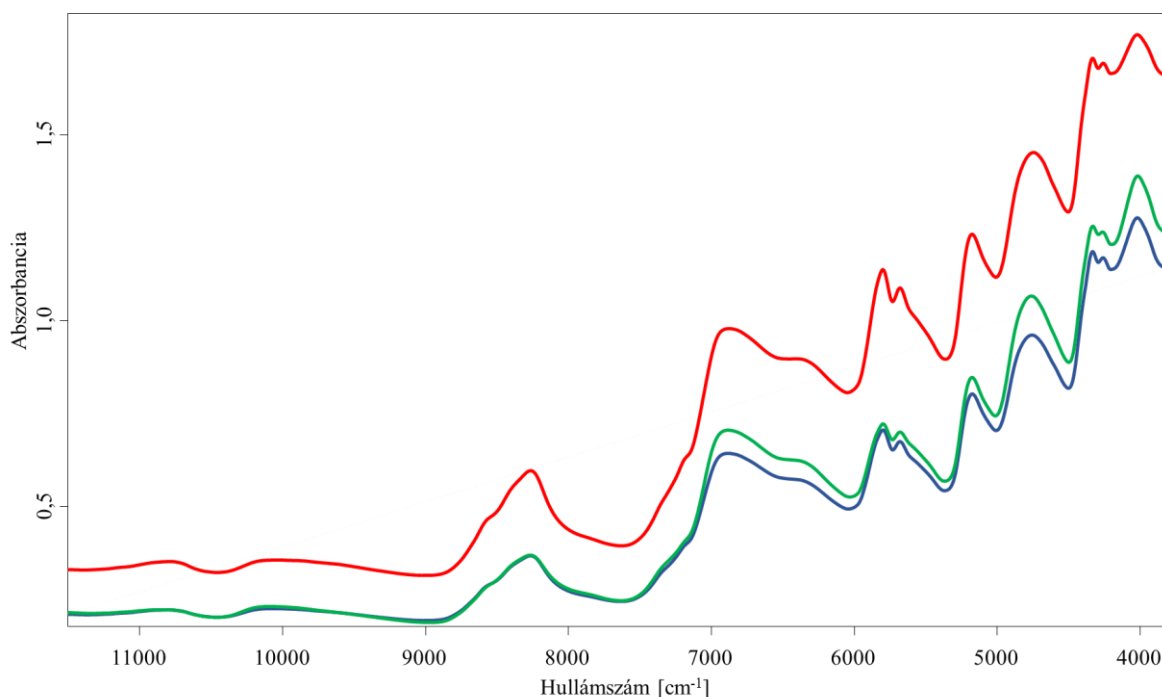
11.ábra: A snackek átlagspektruma. A számozott csúcsok a különböző makrokomponensek abszorpciós sávjaival hozhatók összefüggésbe (9. táblázat).

A snackek általában magas szénhidrát-tartalmú élelmiszerek, melyek végső formájának kialakítása olajban való sütéssel történik. Ebből következik, hogy a minták NIR spektrumai alapvető hasonlóságokat mutatnak. Ennek ellenére számottevő eltérés mutatkozott az egyes minták spektrális adatai között, főként a 6000-5400 és 4900-4000 cm⁻¹ hullámszám tartományban. Ez valószínűleg a gyártástechnológiából adódó különbségekkel, ill. a felhasznált alap- és segédanyagok változatosságával magyarázható. A kisebb spektrális eltérések feltárására a második derivált spektrum transzformációt alkalmaztam, aminek eredményeképpen a 90_18 jelzésű mintát a további vizsgálatokból kizártam. Spektrális tulajdonságai jelentős eltérést mutattak a többi mintához képest, ami azzal magyarázható, hogy ez a termék kókuszából készült, melynek kémiai összetétele nagymértékben eltér a magas keményítőtartalmú alapanyagokétól (pl. burgonya, kukorica).

9. táblázat: A snackekre jellemző NIR abszorpciós sávok (WORKMAN 2000; JERRY és WEYER 2012)

Zsírok, olajok		Hullámszám [cm⁻¹]
6	C-H vegyértékrezgés/C-C vegyértékrezgés kombinációja	4202
6	C-H deformációs rezgés 2. felhangja	4330
5	C-H vegyértékrezgés /C=O vegyértékrezgés/ C-H deformációs rezgés kombinációja	4673
5	O-H vegyértékrezgés /HOH deformációs rezgés kombinációja, O-H deformációs rezgés 2. felhangja	4831
2	C-H vegyérték- és C-H deformációs rezgés kombinációja	7355-7071
1	C-H vegyértékrezgés 2. felhangja	8584-8243
Keményítő és cukrok		
6	C-H /C-C és C-O-C vegyértékrezgések kombinációja	4000
6	C-H vegyértékrezgés/CH ₂ deformációs rezgés kombinációja	4386-4292
5	O-H/C-C vegyértékrezgések kombinációja	4405-4380
5	O-H vegyértékrezgés/HOH deformációs rezgés kombinációja	4762
2	O-H vegyértékrezgés 1. felhangja	6897
Fehérjék		
5	N-H deformációs rezgés 2. felharmonikus kombinációja a C-H /C=O vegyértékrezgés kombinációjával, a C=O vegyértékrezgés/N-H deformációs rezgés/C-N vegyértékrezgés kombinációjával	4650-4545
4	Primer amin C=O karbonil vegyértékrezgésének 2. felhangja (amid kötés)	4930-4810
4	N-H aszimmetrikus vegyértékrezgés és N-H deformációs rezgés kombinációja	5055
3	Aromás C-H vegyértékrezgés 1. felhangja	6173-5880
2	N-H vegyértékrezgés 1. felhangja	6760-6450

A legtöbb termék a gyártás során alkalmazott technológia alapján, három fő csoportba sorolható. Ezek az olajban sült burgonyachipsek, a lúgosan főzött tortilla chipsek és az extrudált termékek. Az adott csoportokba tartozó minták spektrális karakterisztikája alapvetően igen hasonló (12. ábra), ez a magas keményítőtartalommal és azzal van összefüggésben, hogy a snackek végső formájukat leggyakrabban olajban való sütéssel érik el. Emellett kisebb különbségek is megfigyelhetők, melyek a gyártás során felhasznált alap- és segédanyagokkal, illetve az alkalmazott gyártástechnológiával állhatnak összefüggésben.



12. ábra: Snack termékek átlag spektrumai gyártási technológia alapján:

● olajban sültött, ● tortilla, ● extrudált.

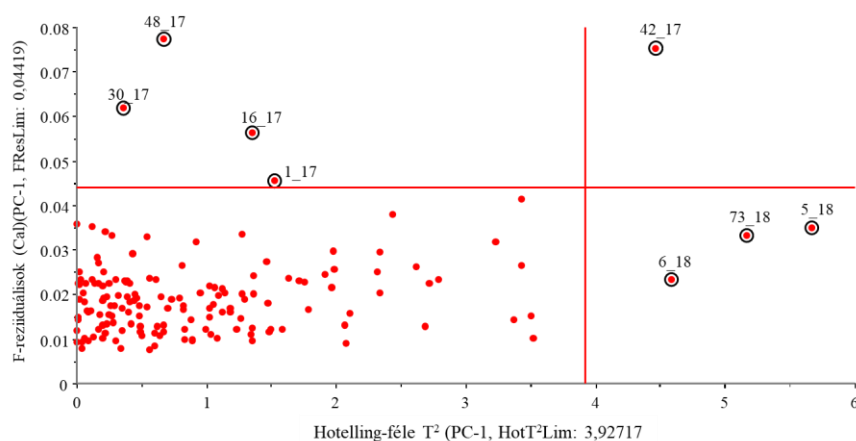
A minták NIR spektroszkópiával történő vizsgálata esetén a legnagyobb nehézséget a termékek összetételében és megjelenésében jelentkező különbségek okozták. Meg kell említeni, hogy a darálás hatására a magasabb zsírtartalmú termékek részecskéi jobban egymáshoz tapadnak, ezáltal a többi mintához képest jobban reflektáló felszín hoznak létre. A fő különbségeket tehát a részecskeméret változása okozta, ami a spektrumok függőleges tengely mentén megfigyelhető eltolódásban nyilvánult meg.

5.1.2. Kemometriai értékelés

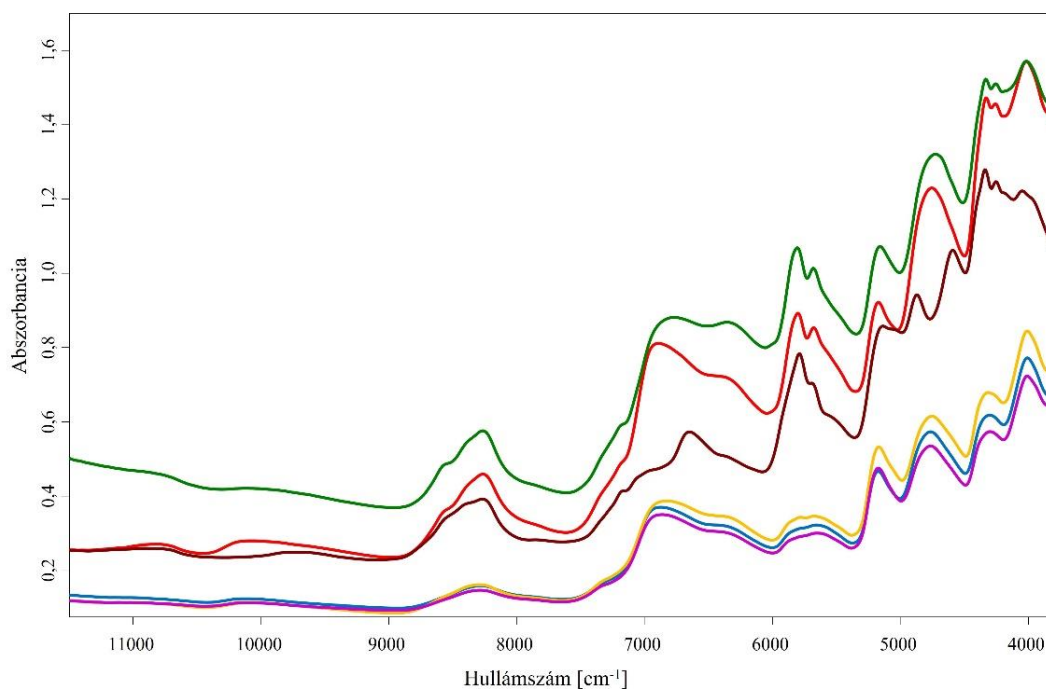
5.1.2.1. Főkomponens elemzés eredményei

A PCA-t a spektrális kiesők feltárására használtam 95%-os konfidencia intervallum mellett. A különböző makrokomponensek esetén eltérő számú adat állt rendelkezésemre (4.1.1. alfejezet) ezért minden egyes makrotápanyagra külön is elvégeztem a PCA-t. Ezáltal a spektrális kiesők sem voltak azonosak az egyes esetekben. A 13. ábra a 155 mintán futtatott PCA legfontosabb eredményeit mutatja be, amely a kiértékelés során megfigyelt szélsőséges minták mindegyikét tartalmazza. Az eredmények alapján, az adatkészlet varianciáját az első két főkomponens (PC) 99%-ban (PC1:98%, PC2:1%) magyarázza.

Az 1_17-es (sós burgonyachips), 16_17-es (paradicsomos burgonyachips) és 42_17-es (sós burgonyachips) mintákat nem tekintettem spektrálisan kiugró értéknek, mert sem az eredeti spektrum, sem annak deriváltja nem mutatott jelentős eltérést a többi mintához képest. A spektrálisan kiugró minták meghatározására az influence plot alkalmazható, amit a mintákra kapott F-reziduális és Hotelling-féle T^2 értékek határoznak meg (13. ábra). A bal alsó kvadránsba tartozó mintákat a modell jól jellemzi, míg a másik három kvadráns esetén valamilyen extrémításról beszélhetünk az adott minta vonatkozásában. Öt minta bizonyult valódi spektrális kiesőnek (14. ábra), ami a többitől eltérő mátrix tulajdonságokkal van összefüggésben. Az 5_18 és 73_18 kódú minták puffasztott rizsből készültek, ezáltal a többi mintához képest magasabb szénhidrát és alacsonyabb zsírtartalommal rendelkeznek. A 6_18-as minta búzalisztból készült sós ropi volt, aminek spektruma a puffasztott rizs mintákhoz hasonló profilt mutatott. A 48_17-es minta alapanyaga cékla volt, ezáltal összetétele és színe is jelentősen eltért a többitől. A spektrális különbségek a 30_17 kódú minta esetében is az alapanyagra vezethetők vissza, amely sertérbőr volt.



13. ábra: Spektrális kiesők, torzító hatású minták minden snack NIR spektrumainak felhasználásával – influence plot



14. ábra: Egy „átlagos” snack (●) és a spektrális kiesők spektrumai: 5_18 (●), 73_18 (●), 6_18 (●), 48_17 (●), 30_17 (●)

5.1.2.2. Snack termékek osztályozása a NIR spektrum alapján

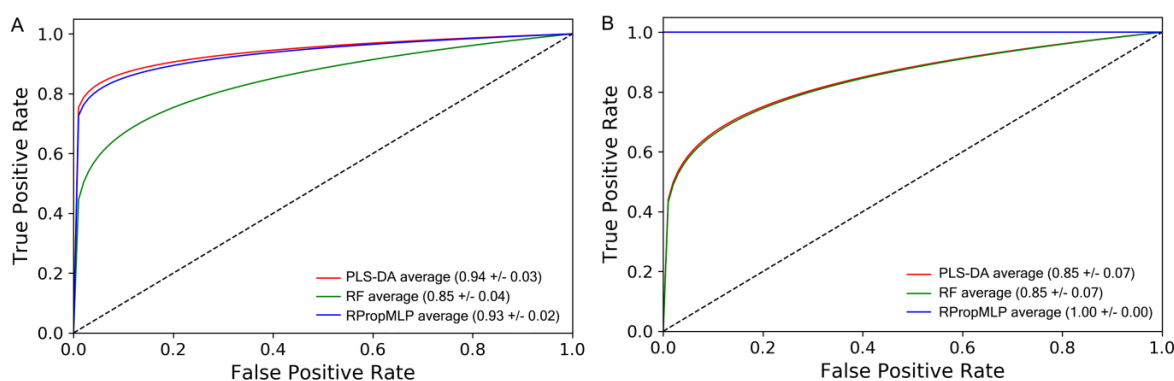
Az elvégzett osztályozási eljárásokat a PCA során meghatározott spektrálisan kieső minták eltávolítása után hajtottam végre. A modellezéshez az első húsz főkomponenst használtam fel. A tanítási/tesztelési ciklus után a részleges legkisebb négyzetek elve-diszkriminancia elemzést (PLS-DA), a véletlen erdő (RF) és a neurális hálózat-alapú többrétegű előre-csatolt reziliens backpropagation hálózat (RPropMLP) algoritmusokat alkalmaztam az osztályozáshoz.

Osztályozási modellek a felhasznált sütőolaj alapján

A leggyakrabban alkalmazott sütőolajok a vizsgált termékek esetében a napraforgó- és a pálmaolaj voltak, ami alapján két csoportot hoztam létre. A modellezéshez 94 mintát tudtam felhasználni, mivel a másfajta olajjal készült termékek esetében a mintaszám ötnél kisebb volt. Erre azért van általánosan szükség, mert az osztályok mintaszámaiban jelentkező jelentős eltérések torzítják a modellt. A modellek belső validálásaként háromszegmenses kereszt-ellenőrzést alkalmaztam, az osztályok alapján rétegzett mintavétellel. A modellekhez kiszámítottam a ROC-görbét (vevő működési karakterisztika görbe) és a megfelelő AUC-értékeket (*area under the curve*, görbe alatti terület). A ROC-görbe úgy jön létre, hogy a valódi pozitív minták arányát (TPR) a hamis pozitívok arányával (FPR) szemben ábrázoljuk. Lehetővé teszik az osztályozási módszerek gyors vizuális összehasonlítását (15. ábra). Teljesítményük azonban pontosabban számszerűsíthető a ROC-görbe alatti terület értékével (AUC). Definíció szerint az AUC-érték egy 0 és 1 közötti szám (százalékban

is megadható) és annak a valószínűségnek felel meg, hogy egy véletlenszerűen kiválasztott pozitív minta magasabbra kerül, mint egy véletlenszerűen kiválasztott negatív (RÁCZ et al. 2016). A modelleket minden esetben random csoportosítási besorolással is ellenőriztem, ami alapján a fejlesztett modellek szignifikánsan eltértek a random modellektől.

A PLS-DA és az RPropMLP módszerek a kereszt-ellenőrzés esetében hatékonyabbnak bizonyultak, mint az RF módszer, míg a PLS-DA teljesítménye rosszabb volt a teszt-validálás esetében. Utóbbi eredményei alapján az RPropMLP kedvezőbb statisztikai paraméterekkel rendelkezik, mint a másik két algoritmus. Mindent összevetve, az RPropMLP adta a legjobb eredményt a ROC-görbék és az AUC-értékek alapján. Ezáltal, a snack termékek gyártása során felhasznált sütőolaj meghatározására az RPropMLP modell a legalkalmasabb. A pontosság (*accuracy*, ACC), a TPR és a valódi negatív minták aránya (*true negative rate*, TNR, specifikusság) értékek az AUC értékekkel együtt a 10. táblázatban vannak összefoglalva a kereszt- és a teszt-validálás esetében is.



15. ábra: A három modell átlagos ROC-görbéinek összefoglalása a sütőolaj osztályozása esetén
A) kereszt-ellenőrzés és B) külső validálás esetén. PLS-DA (●), RF (●), RPropMLP (●).

10. táblázat: A háromszegmenses kereszt-ellenőrzéssel és a teszt-validálással kapott modellek teljesítményparaméterei a felhasznált sütőolajok alapján történő osztályozás során

Módszer	Validálás	Csoportok	AUC	TPR	TNR	ACC
PLS-DA	CV	2	0,94	0,93	0,85	0,89
PLS-DA	Teszt	2	0,85	0,56	0,83	0,68
RF	CV	2	0,85	0,91	0,53	0,76
RF	Teszt	2	0,85	0,76	0,75	0,76
RPropMLP	CV	2	0,93	0,79	0,89	0,83
RPropMLP	Teszt	2	1,00	0,65	1,00	0,79

PLS-DA: részleges legkisebb négyzetek elve-diszkriminancia elemzés; RPropMLP: Neurális hálózat-alapú többrétegű előreecsatolt reziliens backpropagation hálózat; RF: véletlen erdő algoritmus; CV: kereszt-ellenőrzés; AUC: görbe alatti terület; TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság

Osztályozási modellek az alapanyag szempontjából

A 2. adatkészlet (6. táblázat) vizsgálata során a csoportosítás alapja a két fő alapanyag volt: a) burgonya, b) kukorica. Mivel ez a két leggyakrabban felhasznált alapanyag, összesen 126 minta spektrális adataival tudtam dolgozni.

A 11. táblázat alapján elmondható, hogy a három modell minden teljesítménymutató alapján kiváló eredményeket produkált, mind a kereszt-ellenőrzés, mind a teszt-ellenőrzés során. Az RF modell pontossága valamivel kisebb, mint a másik két modellé. Az RF rosszabb statisztikai jellemzőkkel bír, de a különbség nem jelentős. A validálási protokoll részeként alkalmazott randomizációs tesztek eredményei is megerősítették, hogy minden létrehozott modell jobb statisztikai jellemzőkkel bír, mint a random modellek. Következésképpen, a PLS-DA és az RPropMLP modellek ideálisak a snackminták alapanyag alapján történő osztályozására.

11. táblázat: A háromszegmenses kereszt-ellenőrzéssel és a teszt-ellenőrzéssel kapott modellek teljesítményparaméterei a felhasznált alapanyagok alapján történő osztályozás során

Módszer	Validálás	Csoportok	AUC	TPR	TNR	ACC
PLS-DA	CV	2	1,00	0,98	1,00	0,99
PLS-DA	Teszt	2	1,00	0,95	1,00	0,97
RF	CV	2	0,98	0,96	0,91	0,94
RF	Teszt	2	0,98	0,97	0,91	0,94
RPropMLP	CV	2	1,00	0,97	1,00	0,98
RPropMLP	Teszt	2	1,00	1,00	1,00	1,00

PLS-DA: részleges legkisebb négyzetek elve-diszkriminancia elemzés; RPropMLP: Neurális hálózat-alapú többrétegű előrecsatolt reziliens backpropagation hálózat; RF: véletlen erdő algoritmus; CV: kereszt-ellenőrzés; AUC: görbe alatti terület; TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság

Osztályozási modellek gyártási ország alapján

Az adatkészlet származási hely szerinti osztályozásánál négy nagyobb csoportot lehetett létrehozni, ami több mint öt mintát tartalmazott: Belgium, Cseh Köztársaság, Lengyelország és Egyesült Királyság. Összesen ötvenkilenc mintát használtam fel az osztályozáshoz.

A modellek validálására (a korábbi esetekhez hasonlóan) háromszoros kereszt-validálást alkalmaztam (rétegzett mintavétel, randomizált módon) a szokásos 70/30 arányú tréning/teszt felosztási ciklusokkal. A modellek minőségét a 12. táblázatban szereplő teljesítményparaméterek, valamint a ROC-görbék és az AUC-értékek alapján lehet meghatározni. Ebben az esetben a neurális háló alapú RPropMLP algoritmus kedvezőbb, mint a másik kettő. A ROC görbék mind a

CV, mind a teszt-validációs esetekben távolabb vannak a véletlenszerű osztályozástól, azonban a modellek közötti különbségek nem jelentősek. Az AUC-értékek alapján mindhárom modell képes a minták osztályozására. A PLS-DA pontossága sokkal alacsonyabb az RPropMLP-hez képest a CV és a teszt-validációban is, így megállapíthatjuk, hogy az RPropMLP a javasolt megoldás a snack termékek származási hely szerinti osztályozására.

12. táblázat: A háromszegmenses kereszt-ellenőrzéssel és a teszt-ellenőrzéssel kapott modellek teljesítményparaméterei a gyártási ország alapján történő osztályozás során

Módszer	Validálás	Csoportok	AUC	TPR	TNR	ACC
PLS-DA	CV	4	0,93	0,81	0,91	0,81
PLS-DA	Teszt	4	0,92	0,71	0,91	0,71
RF	CV	4	0,93	0,78	0,90	0,78
RF	Teszt	4	0,92	0,78	0,90	0,78
RpropMLP	CV	4	0,97	0,90	0,96	0,90
RpropMLP	Teszt	4	0,97	0,83	0,94	0,83

PLS-DA: részleges legkisebb négyzetek elve-diszkriminancia elemzés; RPropMLP: Neurális hálózat-alapú többrétegű előrecsatolt reziliens backpropagation hálózat; RF: véletlen erdő algoritmus; CV: kereszt-ellenőrzés; AUC: görbe alatti terület; TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság

Osztályozási modellek gyártási technológia alapján

Gyártási technológia alapján három nagy csoport hozható létre, amibe 141 mintát lehetett besorolni: olajban sütés, lúggal főzés és extrudálás. Ez volt a legnagyobb adathalmaz, mivel a legtöbb mért mintát be lehetett sorolni valamelyik csoportba. A NIR spektrumokat az előző esetekhez hasonlóan húsz PC-re redukáltam és ezeket használtam fel az osztályozáshoz. A standard protokoll szerint, ötszegmenses kereszt-ellenőrzést hajtottam végre és elvégeztem az adatkészlet kalibrációra és validációra való felosztását is.

A kereszt-ellenőrzés eredményeit tekintve a három modell szinte egyformán teljesített. A teszt-validálásnál nagyobb, de statisztikailag még mindig nem mutatkoztak nagy különbségek. A ROC-görbék alapján mindhárom modell elfogadható volt. A 13. táblázatban szereplő eredmények azt mutatják, hogy a PLS-DA modell a pontosságok alapján valamivel jobb volt a többi modellhez képest és hogy az RPropMLP egy kicsit gyengébb teljesítményt nyújtott a teszt-validálásánál. Mindazonáltal, mindhárom modell használható a snacktermékek gyártása során alkalmazott technológiák meghatározására a NIR spektrumok alapján.

13. táblázat: A ötszermenses kereszt-ellenőrzéssel és a teszt-ellenőrzéssel kapott modellek teljesítményparaméterei a gyártási technológia alapján történő osztályozás során

Módszer	Validálás	Csoportok	AUC	TPR	TNR	ACC
PLS-DA	CV	3	0,96	0,92	0,96	0,92
PLS-DA	Teszt	3	0,97	0,88	0,95	0,88
RF	CV	3	0,97	0,89	0,93	0,89
RF	Teszt	3	0,95	0,91	0,94	0,91
RPropMLP	CV	3	0,96	0,91	0,96	0,91
RPropMLP	Teszt	3	0,91	0,84	0,92	0,84

PLS-DA: részleges legkisebb négyzetek elve-diszkriminancia elemzés; RPropMLP: Neurális hálózat-alapú többrétegű előrecsatolt reziliens backpropagation hálózat; RF: véletlen erdő algoritmus; CV: kereszt-ellenőrzés; AUC: görbe alatti terület; TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság

5.1.2.3. Részleges legkisebb négyzetek elve regresszió (PLSR) eredményei

A snack termékek zsír, szénhidrát, cukor, fehérje és só tartalmának mennyiségi becslésére részleges legkisebb négyzetek elve (PLS) regressziót alkalmaztam. Minden esetben random öt-elem-kihagyásos kereszt-ellenőrzést és teszt validálást használtam a faktorok optimális számának meghatározására, melyek megállapításához az RMSECV (átlagos négyzetes hiba a kereszt-validálási adatkészletre nézve) és a RMSEP (átlagos négyzetes hiba a teszt adatkészletre nézve) értékeket vettem figyelembe. A spektrálisan kieső minták eltávolítása után, a PLS regressziót a nyers adatokon futtattam le, amely alkalmas volt a változók közötti kapcsolatok tanulmányozására és a kezdeti statisztikai paraméterek rögzítésére. A szóródási ábrák (scores plot) minden tápanyag esetében azt mutatták, hogy az X-Y magyarázott variancia értékek nem megfelelőek, ezáltal a modellek nem alkalmazhatók mennyiségi becslésre. Ennek kiküszöbölésére a 4.3.4.3. alfejezetben felsorolt adat-előkezelési módszereket használtam. Megjegyzendő, hogy a második derivált alkalmazásával a jel/zaj viszony romlik, ezáltal a modellek statisztikai paraméterei gyengültek, emiatt ezen eredmények bemutatásától eltekintek.

A legjobbnak ítélt becslési függvényeket és azok legfontosabb statisztikai paramétereit a 14. táblázat tartalmazza, míg az adott komponensekre vonatkozó becslés relatív hibáját a teljes mérési tartományra a 16. ábra mutatja be.

14. táblázat: A NIR becslési függvények legfontosabb statisztikai jellemzői a snack termékek makrotápanyagaira vonatkozóan kalibráció, ötszeggmenses kereszt-ellenőrzés és teszt-validálás esetében.

Paraméter	Mintaszám	Adat-előkezelés	Kalibráció				Kereszt-ellenőrzés		Teszt-validálás		Mérési tartomány [g/100g]
			R ²	R _t ²	RMSEC [g/100g]	RMSEC _t [g/100g]	Q ²	RMSECV [g/100g]	Q _t ²	RMSEP [g/100g]	
zsír	135	SNV	0,99	0,99	1,05	1,11	0,98	1,17	0,98	1,11	2,2-45,1
szénhidrát	89	FD+SNV	0,95	0,94	1,53	1,53	0,92	1,77	0,92	1,90	45,1-69,7
cukor	146	FD	0,94	0,94	0,49	0,51	0,92	0,54	0,93	0,47	1,7-8,6
fehérje	84	FD+MSC	0,99	0,99	0,65	0,65	0,98	0,78	0,98	0,93	3,0-40,1
só	63	FD+SNV	0,96	0,98	0,11	0,11	0,93	0,14	0,91	0,16	0,7-2,5

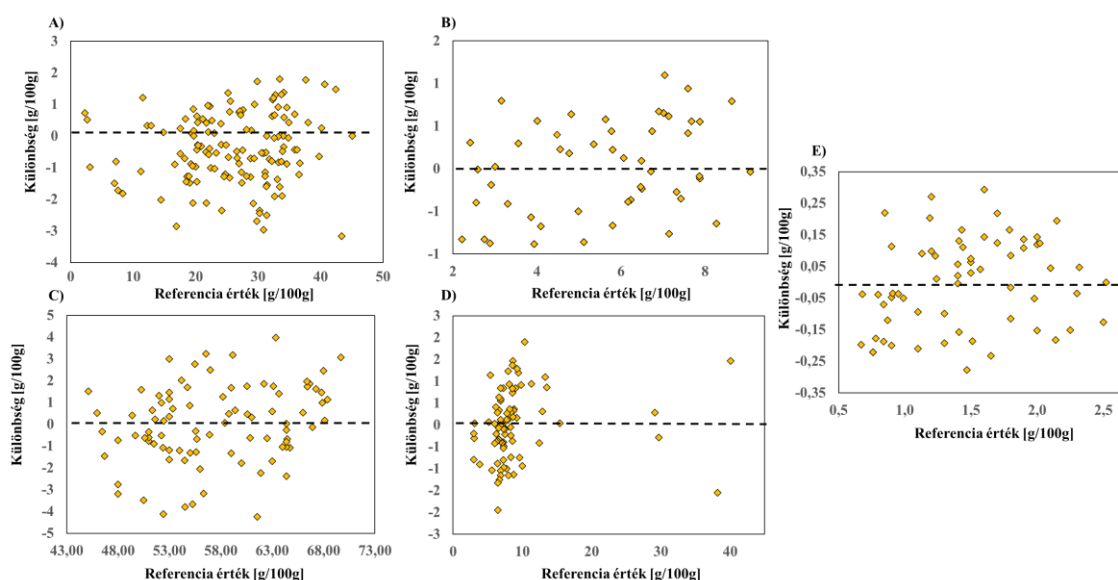
RMSEC: átlagos négyzetes hiba a kalibrációs adatkészletre nézve; RMSECV: átlagos négyzetes hiba a kereszt-ellenőrzés estén; RMSEP: átlagos négyzetes hiba a teszt adatkészletre nézve; R²: kalibrációs adatkészletre vonatkozó determinációs együttható kereszt-ellenőrzés esetén; R_t²: kalibrációs adatkészletre vonatkozó determinációs együttható teszt-ellenőrzés esetén; Q²: kereszt-ellenőrzésre vonatkozó determinációs együttható; Q_t²: teszt-ellenőrzésre vonatkozó determinációs együttható; RMSECV: négyzetes hiba a kalibrációs adatkészletre nézve teszt-ellenőrzésnél

A **zsírok és olajok** nagyon fontos szerepet játszanak a snack-gyártásban, mint sütőközeg és emellett befolyásolják a termékek textúráját és ízét. A becslési függvény mérési tartománya 2,24-45,10 g/100g, melynek az alsó tartományába eső minták (2,24-19,5 g/100 g) főként alacsony zsírtartalmú, extrudálással készült snackek. Az összefüggést kilenc PLS faktor felhasználásával, SNV adat-előkezelés után, a 9403-7498, 6102-5546 és 4605-4242 cm⁻¹ spektrális régiókra állítottam fel, melyek a zsírokra jellemző karakterisztikus elnyelési területekkel vannak összefüggésben.

A snack termékeket magas **szénhidrát-tartalom** jellemzi, amelynek becslésére létrehozott összefüggés a 45,12 és a 62,4 g/100g közötti tartományt fedi le. A mennyiségi becsléshez, SNV és FD spektrumtranszformációk után, a 9403-6094 cm⁻¹ és 5454-4242 cm⁻¹ közötti területeket és nyolc faktort használtam fel. Az Egyesült Államok Élelmiszer- és Gyógyszerügyi Hatósága (FDA), ill. az Európai Unió (EU) által kiadott meghatározás szerint a **cukrok** az összes szabad mono- és diszacharid összegét jelentik, ami magában foglalja a glükózt, fruktózt, galaktózt, laktózt, szacharózt és maltózt és nem tartalmazza a polioloikat (ERICKSON és SLAVIN 2015). A cukrokra vonatkozó mérési tartomány 1,72 és 8,58 g/100 g között volt, ami a referencia adatokat figyelembe véve viszonylag szűk. Ez abból ered, hogy a 8,58 g/100 g fölötti cukortartalommal rendelkező mintákat kizártam a kiértékelésből. Ezek között volt a két banánchips (magas cukortartalom) és egy cukros búzasnack is.

Napjainkban a fogyasztók jobban odafigyelnek a tudatos táplálkozásra, amihez a snackipar is igyekszik alkalmazkodni. Ennek eredményeképpen a piacon számos magas **fehérjetartalommal** rendelkező (20-40%) termék van. Mivel a kísérletekbe ilyen termékeket is bevontam, a PLS regresszió segítségével létrehozott összefüggés a 3,01-40,06 g/100 g nyers fehérjetartalommal rendelkező snack termékek vizsgálatára alkalmas. Az modellépítéshez az FD és MSC kombinált műveleteket és kilenc látens változót (faktort) alkalmaztam. A mennyiségi becslés során relevánsnak bizonyult spektrális régiók 9403-7489 cm^{-1} és 4605-4420 cm^{-1} között találhatók.

Az élelmiszerekben található oldott só (NaCl) a NIR tartományban nem nyel el, azonban mennyisége becsülhető a spektrális adatok alapján. Ennek az oka, hogy a só jelenlétében a víz abszorpciós sávjai a hullámszám mentén eltolódnak. A sáveltolódás nagysága összefüggésben van a sókoncentrációval (LAUB-EKGREEN et al. 2018; OZAKI et al. 2021). A modellépítéshez a 9403-5446 cm^{-1} és a 4605-4420 cm^{-1} régiókat használtam, kilenc faktor mellett. A mérési tartomány 0,67 és 2,52 g/100 g termék közé esik.



16. ábra: A becslés relatív hibája részleges legkisebb négyzetek elve regresszió (PLSR) alkalmazása mellett. (A) zsír, (B) cukor, (C) szénhidrát, (D) fehérje, (E) só.

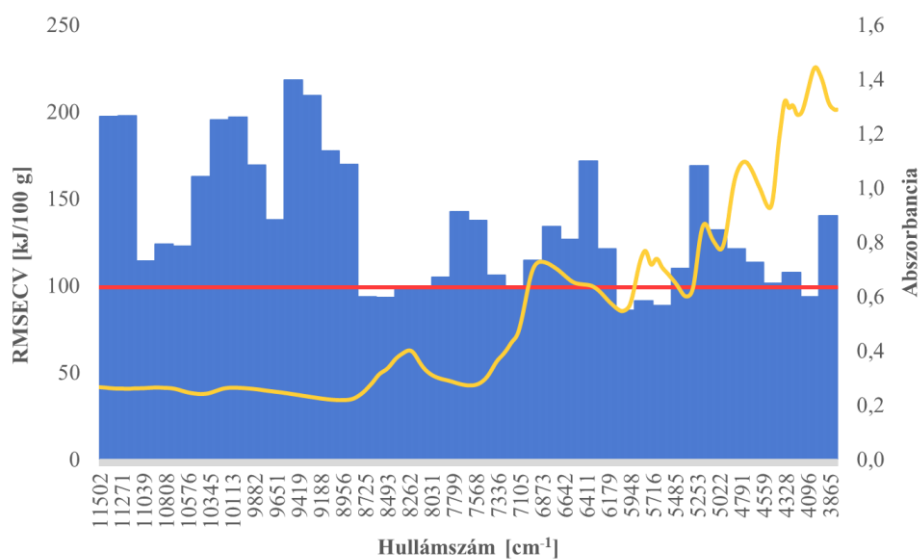
5.1.2.4. Az energiatartalom becslése iPLS regresszióval

Azáltal, hogy az energiatartalmat az élelmiszerben található makrotápanyagok mennyiségéből számoljuk ki, a teljes spektrum hordozhat lényeges információkat ezzel összefüggésben.

Először minden változó felhasználásával, adat-előkezelés nélkül futtattam le a PLS regressziót, aminek eredményeképpen a modell RMSECV értéke 94,67 kJ/100 g-nak adódott, míg a korreláció (Q^2) 0,87 volt (teljes-spektrum modell). A minták energiatartalma 1264,3 és 2520,2 kJ/100 g között mozgott. Az X-Y magyarázott variancia értékei ebben az esetben sem megfelelőek, ezáltal

a modell nem használható az energiatartalom becslésére. Az MSC és az SNV módszerek alkalmazásával a torzítás csökkent, azonban a becslést leíró főbb statisztikai paraméterek értékei romlottak (15.táblázat). FD spektrum transzformációt is alkalmaztam, azonban a szóródási ábra és a modell nagy torzítása alapján a technika alkalmazása nem ajánlott a teljes-spektrum modell esetében.

A modell teljesítményének javítása érdekében intervallum PLS regressziót (iPLSR) alkalmaztam, az energiatartalommal összefüggésben lévő spektrális tartományok meghatározására. Változókiválasztással azonosíthatók és eltávolíthatók azok a változók, amelyek redundánsak, zajosak vagy véletlenszerűen korrelálnak (XIAOBO et al. 2010). A spektrális tartományt 10, 20 és 40 egyenlő szélességű szegmensre osztottam fel, majd a regressziót a nyers és előkezelt adatokon is lefuttattam. Az egyes szegmensekre vonatkozó RMSECV értékeket összehasonlítottam a teljes spektrális tartományra vonatkozó átlagos hiba értékével (avRMSECV). A modellépítéshez azokat a szegmenseket használtam fel, amelyeknél a kapott RMSECV kisebb volt, mint az avRMSECV értéke (17. ábra).



17. ábra: A legjobb iPLS módszer grafikus ábrázolása 20 szegmens és SNV előfeldolgozás felhasználásával. Az egyes szegmensek RMSECV értékeit a hullámszámok függvényében ábrázoljuk. A spektrumot narancssárga színnel, az összes adathalmazra vonatkozó átlagos RMSECV értéket pedig piros vonallal jelöltem.

A statisztikai paraméterek alapján legjobbnak mondható modelleket 20-szegmens esetén, MSC vagy SNV előkezelés után lehetett létrehozni. A kiválasztott hullámszámtartományok mindkét esetben a 8802-8038 cm⁻¹ és 4559-3795 cm⁻¹ közötti változókat jelentették. Az iPLSR eredményeképpen a 2. (C-H vegyérték- és C-H deformációs rezgés kombinációja) és 4. csúcs (primer amin C=O karbonil vegyértékrezgésének 2. felhangja (amid kötés); N-H aszimmetrikus

vegyértékmezés és N-H deformációs rezgés kombinációja) abszorpciós sávjait (11. ábra) ki kell hagyni a kiértékelésből, a meghatározott RMSECV értékek alapján (17. ábra). Ezek a tartományok azonban összefüggésben vannak a mintában található zsírokkal, olajokkal, fehérjékkel, a keményítővel és a cukrokkal is, ezáltal a modellből való kizárásuk problémás lehet. A tapasztalt nagyobb átlagos hiba valószínűleg azzal állhat kapcsolatban, hogy a termékek változatos összetétele miatt nem egyértelműek a spektrális változások és az energiatartalom közötti összefüggések. Az eredmények alapján jól látszik, hogy az átlagos hiba értékének csökkentésére nem volt alkalmas az iPLSR, azonban az X-Y magyarázott variancia értékek javítására igen. A 2. és 4. csúcs abszorpciós sávjainak modellbe való bevonásával az átlagos hiba értéke nőtt, azonban a modell torzítása csökkent. További PLS faktorok alkalmazásával a statisztikai paraméterek tovább javulnának, azonban ezek erősen növelik a torzítás mértékét (túlillesztés). A legjobb modellt SNV adat-előkezelés után, 20-szegmenses iPLSR esetén értem el (15. táblázat).

15. táblázat: Intervallum PLS (iPLS) modellek összehasonlítása 10, 20 és 40 szegmenses a snackek energiatartalmára vonatkozóan. A modell teljesítménye, előfeldolgozás és a látens változók száma (mérési tartomány: 1264,3-2520,2 kJ/kg).

Módszer	Változók száma	adat-előkezelés	faktor	R ²	Q ²	RMSEC [kJ/100g]	RMSECV [kJ/100g]
Teljes-spektrum modell	1000	no	5	0,89	0,87	86,25	94,67
	1000	MSC	4	0,87	0,86	90,51	97,76
	1000	SNV	4	0,87	0,86	90,4	97,64
iPLS – 10 intervallum	100	no	7	0,91	0,88	78,58	87,96
	300	MSC	4	0,88	0,86	87,90	97,05
	400	SNV	4	0,89	0,87	86,33	94,82
iPLS – 20 intervallum	100	no	6	0,90	0,88	80,68	89,07
	300	MSC	4	0,89	0,87	86,49	94,66
	550	SNV	4	0,89	0,87	86,38	94,66
+ a 2. és 4. csúcs abszorpciós sávjai	300	MSC	4	0,88	0,86	90,03	96,26
+ a 2. és 4. csúcs abszorpciós sávjai	400	SNV	4	0,87	0,86	90,32	97,03
iPLS – 40 intervallum	225	MSC	6	0,90	0,87	81,58	93,31
	200	SNV	6	0,90	0,88	79,62	89,31

A legjobb modellek félkövér és dőlt betűvel jelölve.

RMSEC: átlagos négyzetes hiba a kalibrációs adatkészletre nézve; RMSECV: átlagos négyzetes hiba a kereszt-ellenőrzés esetén; RMSEP: átlagos négyzetes hiba a teszt adatkészletre nézve; R²: kalibrációs adatkészletre vonatkozó determinációs együttható kereszt-ellenőrzés esetén; R_t²: kalibrációs adatkészletre vonatkozó determinációs együttható teszt-ellenőrzés esetén; Q²: kereszt-ellenőrzésre vonatkozó determinációs együttható; Q_t²: teszt-ellenőrzésre vonatkozó determinációs együttható; RMSECV: négyzetes hiba a kalibrációs adatkészletre nézve teszt-ellenőrzésnél

Mivel a snackipart erősen jellemzi a termékek diverzifikációja, az energiatartalom pontosabb becslése érdekében további minták vizsgálata és egyéb, akár nem-lineáris kemometriai módszerek alkalmazása javasolt.

5.2. Kávé

5.2.1. Referencia adatok kiértékelése

5.2.1.1. Zöld kávé

Az ötven zöld kávé mintára vonatkozó kémiai mérések eredményeit és fontosabb statisztikai paramétereit a 16. táblázat foglalja össze. A mérési eredmények kiértékelés során a négy földrajzi régióból származó minták adatait egyszempontos varianciaanalízissel (*analysis of variance*, ANOVA) hasonlítottam össze, azonban egyik változó tekintetében sem volt szignifikáns különbség tapasztalható (17. táblázat).

16. táblázat: A zöld kávék kémiai összetételére vonatkozó eredmények összefoglalása

	átlag ±szórás	minimum	maximum	medián	RSD [%]
pH	5,73±0,12	5,42	6,00	5,71	0,32
koffein [mg/g]	11,40±2,74	0,33	22,70	11,36	1,93
5-CQA [mg/g]	47,96±7,55	18,93	61,61	47,43	4,99
4-CQA [mg/g]	11,15±2,89	5,52	19,06	10,49	5,14
3-CQA [mg/g]	7,59±2,89	3,75	18,37	6,80	4,97
3,5-diCQA [mg/g]	4,39±1,18	2,34	8,09	4,45	5,16
4,5-diCQA [mg/g]	3,08±1,08	1,85	8,00	2,89	4,72
3- FQA [mg/g]	2,35±1,07	0,98	5,37	1,93	5,62
ΣCQA [mg/g]	76,07±9,76	61,18	98,07	74,19	-

5-CQA: kinasav-5-O-kávésav; 4-CQA: kinasav-4-O-kávésav; 3-CQA: kinasav-3-O-kávésav; 3,5-diCQA: kinasav-3,5-O-dikávésav; 4,5-diCQA: kinasav-4,5-O-dikávésav; 3-FQA: kinasav-3-O-ferulasav; CGA: klorogénsavak; RSD%: a referencia módszer átlagos relatív szórása (n=50)

A minták pH értékét minden, az extraktumban jelenlévő savas karakterű komponens befolyásolja. Ez nemcsak az alifás savakat, hanem a szabad és fehérjéhez kötött fenol- és aminosavakat is magában foglalja (FARAH 2019). A zöld kávéknál meghatározott pH értékek az 5,42-6,00 tartományban mozogtak, ezáltal részben összhangban vannak a DA SILVA ARAÚJO és munkatársai (2020) által meghatározott eredményekkel (5,7-5,84). A szélesebb tartomány valószínűleg azzal függ össze, hogy míg az említett publikációban vizsgált kávék Brazília

különböző területeiről származtak, az általam vizsgált minták földrajzi eredete több országhoz köthető. A különböző régiókból származó minták kémiai összetételét, ezáltal a savas karakterű komponensek mennyiségét meghatározzák az időjárási, termesztési körülmények és az alkalmazott feldolgozási módszerek.

17. táblázat: A zöld kávék kémiai összetételére vonatkozó eredmények összefoglalása a négy régióra vonatkozóan

	Afrika (n=14)	Ázsia (n=11)	Közép-Amerika (n=13)	Dél-Amerika (n=9)
pH	5,71±0,08	5,72±0,14	5,80±0,15	5,71±0,06
koffein [mg/g]	10,86±0,81	11,01±1,03	11,82±0,93	11,11±0,63
5-CQA [mg/g]	51,22±6,17	46,95±4,95	47,39±6,91	46,87±6,61
4-CQA [mg/g]	10,55±2,49	11,95±3,01	10,43±2,47	10,96±3,26
3-CQA [mg/g]	6,58±1,98	8,74±2,98	7,10±2,37	7,34±2,51
3,5-diCQA [mg/g]	4,44±0,87	4,05±1,11	4,63±1,12	3,82±0,70
4,5-diCQA [mg/g]	3,02±0,70	2,85±0,43	3,10±1,08	2,73±0,74
3-FQA [mg/g]	1,88±0,62	2,72±1,23	2,10±0,91	2,28±0,85

* egyik változó tekintetében sem volt szignifikáns különbség a származási hellyel összefüggésben

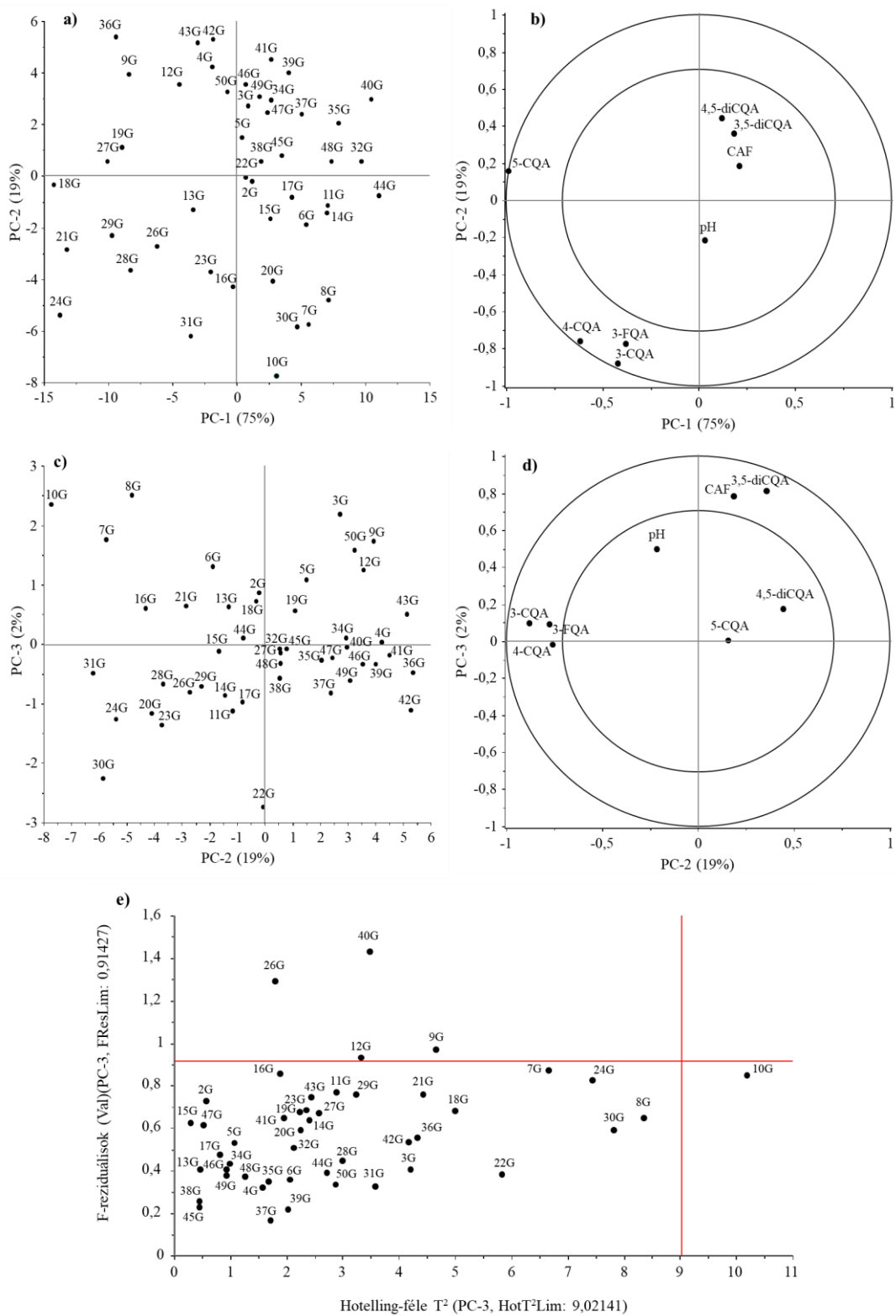
5-CQA: kinasav-5-O-kávésav; 4-CQA: kinasav-4-O-kávésav; 3-CQA: kinasav-3-O-kávésav; 3,5-diCQA: kinasav-3,5-O-dikávésav; 4,5-diCQA: kinasav-4,5-O-dikávésav; 3-FQA: kinasav-3-O-ferulasav; CGA: klorogénsavak

Az arabica minták koffeintartalma 8,73 és 13,54 mg/g között mozgott, kivéve a koffeinmentesített 1G kávé, amely 0,32 mg/g-ot tartalmazott. A vártnak megfelelően, a robusta minták (25G, 33G) esetében határozhatók meg a legmagasabb koncentráció értékek, melyek 20,08 és 22,47 mg/g-nak adódtak. A mért értékek egybeesnek korábbi tudományos publikációk megállapításaival (9,1-24,3 mg/g) (FUJIOKA és SHIBAMOTO 2008; MACHEINER et al. 2019; STEFANELLO et al. 2019). A zöldkávét klorogénsav tartalma különböző mértékben változhat függően a növény genetikájától (faj, fajta), a kávécsereznyé érettségi állapotától, illetve a mezőgazdasági gyakorlattól, a klímától és a talaj minőségétől is (FARAH és DONANGELO 2006). Mivel a minták különböző termőterületekről származtak, a komponensek mennyisége széles tartományokban mozgott. Korábbi kutatások megállapították (MACHEINER et al. 2019; JAISWAL et al., 2010; MOON et al. 2009), hogy a zöldkávétban legjelentősebb klorogénsavak (CGA) a kinasav-5-O-kávésav (5-CQA), melyet szerkezeti izomerjei követnek, a kinasav-4-O-kávésav (4-CQA) és a kinasav-3-O-kávésav (3-CQA), amit az általam vizsgált minták eredményei is alátámasztanak.

Fontos kiemelni, hogy az 1G minta esetében a klorogénsavak aránya valószínűleg a „Swiss Water” koffeinmentesítési eljárás hatására eltérést mutatott, a kezeletlen mintákhoz képest. A koffein

eltávolítása során a zöld kávébab számos kémiai változáson megy keresztül, amelyek nagymértékben függenek a módszer és az oldószer kiválasztásának termodinamikai feltételeitől (FARAH 2019; JESZKA-SKOWRON et al. 2020). Egy „átlagos” zöldkávét (a medián értékeket tekintve) 47,96 mg/g 5-CQA-t; 11,15 mg/g 4-CQA-t és 7,59 mg/g 3-CQA-t tartalmaz. Ehhez képest a koffeinmentes mintában meghatározott értékek kiegyenlítettebb képet mutatnak, ahol az 5-CQA, az 4-CQA és 3-CQA mennyisége 18,93; 18,50 és 18,37 mg/g-nak adódott (FARAH, DE PAULIS et al. 2006).

A mennyiségi adatok átfogó vizsgálata és vizualizálása érdekében főkomponens elemzést (PCA) futtattam az adatkészleten (18. ábra). A koffeinmentes és a két robusta mintát kizártam a kiértékelésből, mivel kémiai összetételük jelentősen eltér a többi mintáétól. Az eredmények alapján az első három főkomponens (*principal component*, PC) 96%-ban magyarázza az adatkészlet varianciáját (PC1:75%; PC2:19%; PC3:2%). Az első két PC-t figyelembe véve a minták közötti különbségeket elsősorban az 5-CQA, a 4-CQA, a 3-CQA, illetve a 3-FQA (kinasav-3-O-ferulasav) koncentrációjában jelentkező eltérések okozhatják a kapott faktorsúlyok alapján. Kisebb mértékben járulhat hozzá a minták térben való szóródásához a kinasav-3,5-O-dikávésav (3,5-diCQA) és a koffein mennyisége (PC3). Az F-reziduális és a Hotelling-féle T^2 értékek alapján a 10G (Szumátra, nedvesen hántolt), a 26G (Galápagos-szigetek, mosott, Bourbon változat) és a 40G (Guatemala, mosott) mintáknak van jelentős hatása a modellre (18.e ábra).



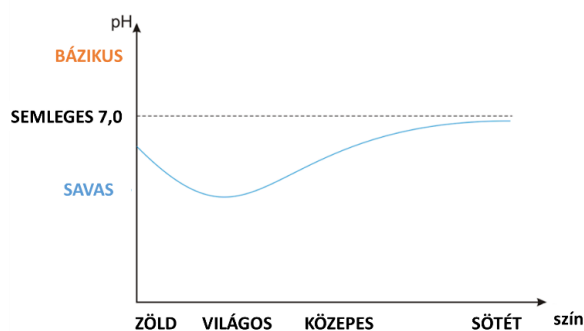
18. ábra: A zöld kávé mennyiségi adataira futtatott főkomponenselemzés eredményei. Scores plot-ok: a) PC1-PC2; c) PC2-PC3; loading plot-ok: b) PC1-PC2; d) PC2-PC3; e) influence plot

5.2.1.2. Pörkölt kávé

A pörkölés összetett folyamat, amely során a zöldkávé fiziko-kémiai tulajdonságai jelentősen megváltoznak köszönhetően a magas hőmérsékletnek (230-280°C). A pörkölt kávé kémiai összetétele nagyban függ az alkalmazott pörkölési hőmérséklettől, időtől, illetve a zöld kávé sajátos tulajdonságaitól. Az iparban bevett gyakorlat, hogy a kávé pörkölése adott színskálák alapján történik (Agron-skála). Mivel a kávé színét befolyásolja annak összetétele és a post-harvest technológia is, az optimális pörkölési szintet minden kávé esetén külön be kell állítani a kívánt érzékszervi tulajdonságok elérése érdekében.

A doktori munkám során három pörkölési szinten vizsgáltam a zöldkávé minták összetételének változását a különböző fizikai, kémiai és érzékszervi paraméterek mentén. A kapott eredményeket a 18. táblázat foglalja össze.

Általánosan a kávé savassága csökkenő tendenciát mutat a pörköltémi szint emelkedésével. A világos pörkölésű kávék minden esetben savasabb karakterűnek bizonyultak, mint a zöld kávék. Ez valószínűleg összefüggésben van a pörkölés korai szakaszában keletkező szerves savakkal, mint pl. az ecetsav és hangyasav, melyek a Maillard-reakció és a karamellizáció termékei. A pörkölés végső szakaszában azonban magasabb hőmérsékleten elbomlanak vagy elpárolognak (FOLMER 2017). Ennek eredménye, hogy a kávé savassága csökken, így a mért pH értékek nagyobbak (19. ábra). A közepes és a sötét pörkölésű minták esetében magasabb pH érték volt tapasztalható. A világos, közepes és sötét kávékra mért átlagos értékek 4,89, 4,94 és 5,07 voltak. A meghatározott pörkölési szintek között szignifikáns különbség volt tapasztalható ($\alpha=0,05$).



19. ábra: A kávé pH értékének általános változása a pörkölés függvényében (FARAH 2019)

A koffein meghatározása során kapott eredmények összhangban vannak a korábbi kutatások eredményeivel (FUJIOKA és SHIBAMOTO 2008; DE LUCA et al. 2018; ZANIN et al. 2020; ZHANG et al. 2013). Az arabica minták koffeintartalma 9,9-15,43 mg/g között mozgott. Mivel a koffein hőstabil vegyület, koncentrációja lényegesen nem változik a pörkölés során, esetlegesen enyhe emelkedés tapasztalható. A százalékos összetétel tekintetében a koffeintartalom növekedése figyelhető meg a hőérzékeny vegyületek vesztesége miatt (CHU 2012). Ezt alátámasztja a

varianciaelemzés eredménye is, miszerint az egyes szintek között nincs szignifikáns különbség a koffein koncentrációja tekintetében (18. táblázat).

18. táblázat: A kávémintákra mért referencia adatok átlagos értékei pörkölési szintenként

	zöld	világos	közepes	sötét	RSD [%]
pH	5,73±0,12	4,89±0,09 ^a	4,94±0,12 ^b	5,07±0,16 ^c	0,13
koffein [mg/g]	11,40±2,74	12,18±1,10 ^a	12,27±0,87 ^a	12,42±1,11 ^a	1,27
5-CQA [mg/g]	47,96±7,55	20,10±13,93 ^a	15,53±6,62 ^b	12,25±3,67 ^c	4,87
4-CQA [mg/g]	11,15±2,89	11,10±2,16 ^a	9,05±1,67 ^b	7,04±1,97 ^c	4,97
3-CQA [mg/g]	7,59±2,89	8,70±1,91 ^a	6,65±1,42 ^b	5,05±1,97 ^c	5,18
3,5-diCQA [mg/g]	4,39±1,18	2,36±0,70 ^a	1,94±0,50 ^b	0,92±0,30 ^c	4,61
4,5-diCQA [mg/g]	3,08±1,08	1,96±0,76 ^a	1,40±0,52 ^b	1,52±0,92 ^b	4,56
3-FQA [mg/g]	2,35±1,07	3,58±1,14 ^a	3,65±0,95 ^a	2,81±0,48 ^b	5,14
sum CGA [mg/g]	76,07±9,76	47,81±8,17 ^a	38,23±7,06 ^b	29,60±7,64 ^c	-
érzékszervi savasság	-	5,21±1,15 ^a	4,26±1,31 ^b	2,98±1,24 ^c	-
érzékszervi keserűség	-	3,63±1,53 ^a	4,33±1,45 ^b	5,55±1,45 ^c	-
L*	61,73±2,36	43,37±1,33 ^a	41,82±1,35 ^b	39,40±1,49 ^c	
a*	2,08±0,95	6,91±0,39 ^a	6,43±0,41 ^b	5,63±0,48 ^c	
b*	16,18±1,96	10,04±1,00 ^a	8,56±1,17 ^b	6,22±1,34 ^c	

*A sorokban szereplő változók alapján Tukey HSD post hoc teszt alapján meghatározott homogén alcsoportokat felsőindexben található betűkkel jelöltem (p<0,05).

5-CQA: kinasav-5-O-kávésav; 4-CQA: kinasav-4-O-kávésav; 3-CQA: kinasav-3-O-kávésav; 3,5-diCQA: kinasav-3,5-O-dikávésav; 4,5-diCQA: kinasav-4,5-O-dikávésav; 3-FQA: kinasav-3-O-ferulasav; CGA: klorogénsavak
RSD%: a referencia módszer átlagos relatív szórása pörkölt kávé esetén (n=150)

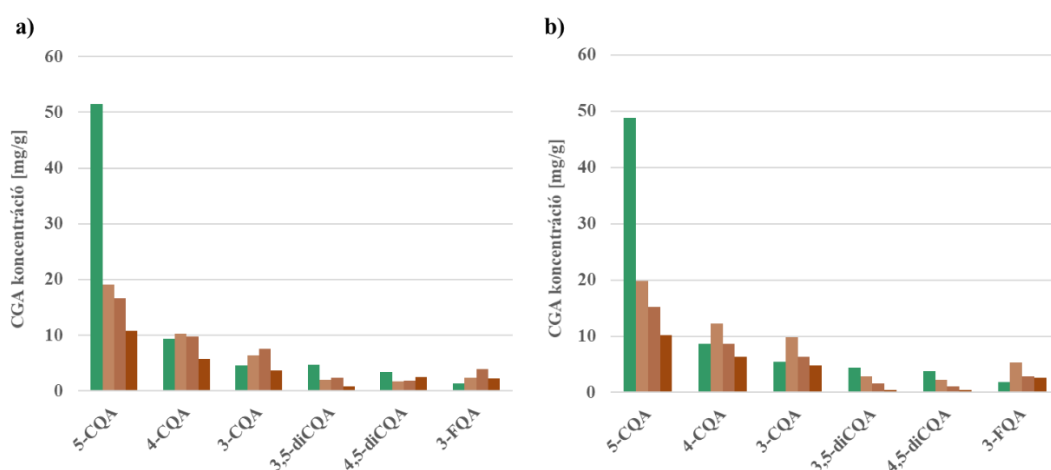
A pörkölésnek ezzel szemben jelentős hatása van a kávéban található klorogénsavak mennyiségére és összetételére. A feldolgozás típusától függően mennyiségük általában 0,5 és 6% között változhat. Ha a pörkölési folyamat intenzívebb, mennyiségük akár a szárazanyag kevesebb, mint 1%-a lehet (STEFANELLO et al. 2019). A 18. táblázat eredményei alapján is látszik, hogy az egyes pörkölési szintek között szignifikáns különbség tapasztalható. Az átlagértékeket figyelembe véve, az 5-, 4- és 3-CQA, illetve a 3,5-diCQA koncentrációja csökkenő tendenciát mutatott a pörkölés során. A minták összes CGA tartalma a zöldkávékhoz képest átlagosan 35,9%-kal csökkent a világos pörkölésű, 49,2%-kal a közepes és 60,5%-kal a sötét pörkölésű kávék esetén. Általánosságban valóban egy csökkenő tendenciát figyelhetünk meg a klorogénsavak mennyiségében (összes CGA), azonban az egyes zöldkávék eltérő genetikájából, termesztési és feldolgozási körülményeiből adódóan számos eltérés mutatkozik. Fontos észben tartani, hogy a klorogénsavak a pörkölés során degradálódnak, azonban emellett izomerizáció és egyéb

különböző kémiai átalakulások is megfigyelhetők. A változások részletesebb bemutatásával rávilágíthatunk a kávé minőségellenőrzésében jelentkező nehézségekre (M9.).

A legdrasztikusabb változás minden esetben az 5-CQA mennyiségében tapasztalható, ami egyértelműen van a tudományos szakirodalommal. A legtöbb minta esetén egyértelmű csökkenés volt tapasztalható, azonban öt kávénál (24, 30, 32, 39, 40) a sötét pörkölésnél nagyobb volt a vegyület koncentrációja, mint a közepesnél.

A 4-CQA-ra mért eredmények már nem mutatnak ennyire egyértelmű képet. Csökkenő tendencia csak tizenkettő mintánál volt megfigyelhető. A zöld kávéhoz képest nagyobb koncentrációt mértem a világos pörkölésű mintáknál huszonhárom esetben (20. ábra). Ez valószínűleg az 5-CQA izomerizációjával van összefüggésben. Bizonyos minták esetén a világos pörkölésű mintáknál nőtt, majd szisztematikusan csökkent a 4-CQA mennyisége. Egyes mintáknál a közepes és sötét szint esetén mért értékek között alig volt tapasztalható eltérés. Elmondható azonban az is, hogy a legtöbb esetben a sötét pörkölés esetén határozottan meg a legkisebb koncentrációt.

A 3-CQA eredményei alapján két fő tendenciát figyelhetünk meg; vagy egyértelmű csökkenést, vagy a zöldkávéhoz képest tapasztalt növekedést (huszonkilenc db), majd szisztematikus csökkenést (20. ábra). Bizonyos mintáknál a közepes és a sötét pörkölésre mért értékek között alig tapasztalható eltérés. A koncentráció a legtöbb mintánál a sötét pörkölésnél volt a legkisebb. Az adatokból megfigyelhető, hogy egy adott kávé vizsgálva a 4- és 3-CQA koncentrációja hasonló tendencia szerint változott a pörkölés során. Ez azt jelenti, hogy ha két pörkölési szint között az egyik vegyület mennyisége nőtt/csökkent, akkor a másiké is. Ez a jelenség az általam vizsgált ötven minta közül negyvenkettőnél volt megfigyelhető (20. ábra).



20. ábra: Az egyes klorogénsavak koncentrációjának változása a pörkölés során. a) 4. minta (Etiópia) b) 46. minta (Uganda). ● zöld kávé ● világos ● közepes ● sötét pörkölés.

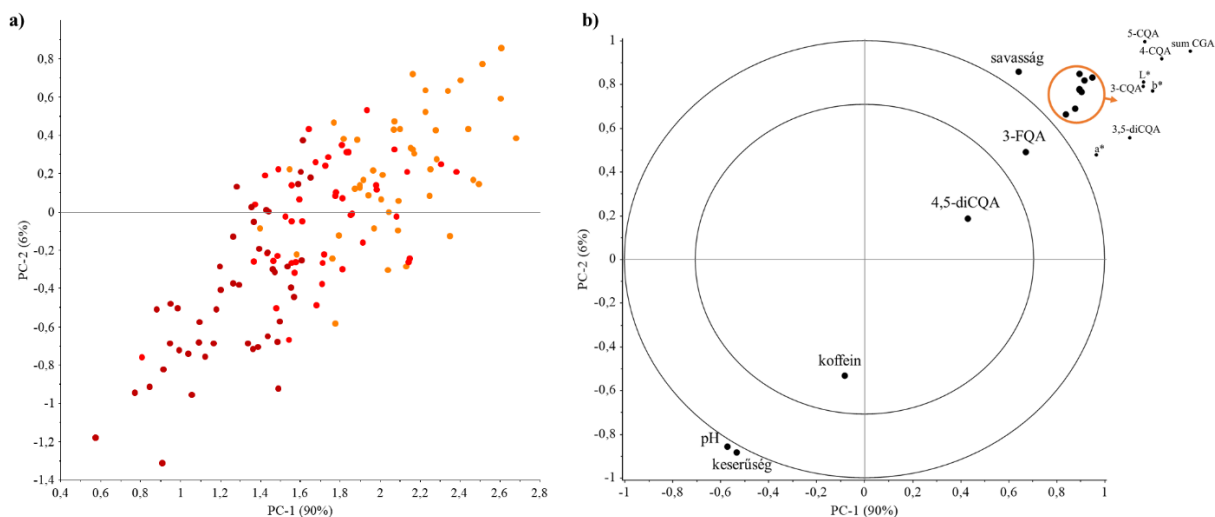
5-CQA: kinasav-5-O-kávésav, 4-CQA: kinasav-4-O-kávésav, 3-CQA: kinasav-3-O-kávésav, 3,5-diCQA: kinasav-3,5-O-dikávésav, 4,5-diCQA: kinasav-4,5-O-dikávésav, 3-FQA: kinasav-3-O-ferulasav

A 3,5-diCQA mennyiségének változását kétféle tendenciával írhatjuk le a vizsgált mintáknál (20. ábra): egyértelmű csökkenés, ill. a világos pörkölésű kávénál csökkenés tapasztalható a zöldkávéhoz képest, majd növekedés és ismét csökkenés.

A 4,5-diCQA és a 3-FQA koncentrációjának változását legalább hatféle tendencia jellemezte. Míg a 4,5-diCQA mennyiségében a pörkölés végére általában csökkenés lépett fel a kiindulási koncentrációhoz képest (kb. 50%-os), addig a 3-FQA mennyisége a minták nagyobb részénél (harmincöt db) növekedést mutatott. Ez átlagosan meghaladhatta akár a 70%-ot is.

Korrelációelemzést futattam le a különböző változók közötti kapcsolatok vizsgálata céljából. Ennek eredményei (M5.) alapján a mono-kávésavak mennyisége között pozitív korreláció figyelhető meg, ami a pörkölés során történő izomerizációval van összefüggésben. Az 5-CQA mennyisége nagyon erős, pozitív korrelációt mutatott a 4-CQA ($r(136)=0,94$) és a 3-CQA ($r(136)=0,90$) mennyiségével is ($\alpha=0,01$). Ugyanilyen irányú, de még erősebb korreláció volt megfigyelhető a 4-CQA és 3-CQA között ($r(136)=0,97$). A 3,5-diCQA esetén mért értékek is erősen korreláltak az 5-, a 4- és a 3-CQA koncentrációjával ($r(136)=0,71$; $r(136)=0,76$, $r(136)=0,74$). A 4,5-diCQA és a 3,5-diCQA mennyisége között közepes korreláció tapasztalható ($r(136)=0,53$), ami a pörkölés során fellépő folyamatokkal lehet összefüggésben. Fontos kiemelni a 3-FQA és az egyéb klorogénsavak koncentrációjára vonatkozó összefüggéseket is, mivel az általam feldolgozott tudományos szakirodalom nem tartalmazott erre vonatkozó adatokat. Minden esetben pozitív korrelációról beszélhetünk. Míg az 5-CQA-nál közepes ($r(136)=0,50$), addig a 4- és 3-CQA, ill. 3,5-diCQA esetén erős ($r(136)=0,67$; $r(136)=0,67$, $r(136)=0,72$) kapcsolat áll fenn a mintákban található 3-FQA koncentrációjával.

Mivel az adatok áttekintése mennyiségük miatt nehéz feladat, PCA-t futtattam le a fenálló kapcsolatok szemléltetésére (21. ábra). Tekintve, hogy az adatok mértékegysége és mérési tartománya nem egyezett, a kiértékelés előtt a minimummal való centrálást és a tartománnyal való skálázást is elvégeztem.



21. ábra: PCA pörkölt kávé mennyiségi adatain futtatott főkomponenselemzés eredményei:
 a) scores plot; b) loading plot. ● sötét ● közepes ● világos

A 21. ábra az eredményül kapott score és faktorsúly (loading) ábrákat mutatja. Az első két PC az adatok variációjának 96%-át magyarázza. A 21.a ábrán jól látható, hogy a minták a pörkölési szinteknek megfelelően szóródnak a főkomponensek mentén. A kisebb PC1 és PC2, ill. a nagyobb PC1 és PC2 értékekhez tartozó változók között keresztirányú tendencia rajzolódik ki, sötétől világos pörkölésig. A korrelációs loading ábra (21.b) megmutatja, hogy mely változóknak van a legjelentősebb hatása a minták pozíciójára. Azok a változók, amelyek a középponthoz közel helyezkednek el, nem rendelkeznek nagy hatással. Azok a minták, amelyek a pörkölés során a legsötétebbnek bizonyultak kis PC1 és negatív PC2 értékekkel rendelkeznek. Az egyes változók fontossága a korrelációs loading ábrán jobban látható, mint a standard loading ábrán. A külső kör a 100%-os magyarázott variációt jelzi, míg a belső kör a magyarázott variancia 50%-át. A kávé színe mellett, a benne található klorogénsavak mennyisége is indikátora lehet a pörkölési állapotnak (MACHEINER et al. 2021; ALAMRI et al. 2022). Az M5. melléklet eredményeiből is jól látható, hogy a minták CGA tartalma és a meghatározott szín paraméterek között pozitív korreláció mutatkozik. Az összefüggés olyan módon áll fenn, hogy a pörkölés előrehaladtával a kávé színe egyre sötétebb lesz, a klorogénsavak koncentrációja pedig csökken. Az egyes vegyületekkel eltérő kapcsolat áll fenn, azonban, ha a vizsgált klorogénsavak összegét nézzük, nagyon erős korrelációról beszélhetünk az L^* ($r(136)=0,80$) és a b^* ($r(136)=0,81$) értékeivel. Emellett erős összefüggés áll fenn a mért a^* értékekkel is ($r(136)=0,73$). Ezek az összefüggések egyértelműen megmutatkoznak a PCA korrelációs loading-ján (21.b ábra).

A fent említett mintáknál határoztuk meg a legmagasabb értékeket az érzékszervi vizsgálatnál, a keserűséggel összefüggésben. Ha megfigyeljük a 21.b ábrát, a pH változó elhelyezkedése elsősorban furcsa lehet, azonban a magasabb pH érték kevésbé savas karakterű kávéra utal. A keserűség és a

pH között erős pozitív korreláció áll fenn ($r(136)=0,73$). Ugyanilyen mértékű, de ellentétes előjelű összefüggés van az érzékszervi savasság és a mért pH között ($r(136)=-0,73$). A negatív korreláció azért lehetséges, mert a kisebb pH-értékhez (savasabb) általában magasabb pontszám tartozott az érzékszervi bírálat során (ellentétes skála).

Sem a mintákban található koffeinnek, sem a 4,5-diCQA-nak nincs számottevő hatása az egyes minták elhelyezkedésére. Ez valószínűleg azzal van összefüggésben, hogy ezeknek a vegyületeknek a mennyisége csak kisebb mértékben változik a pörkölés során.

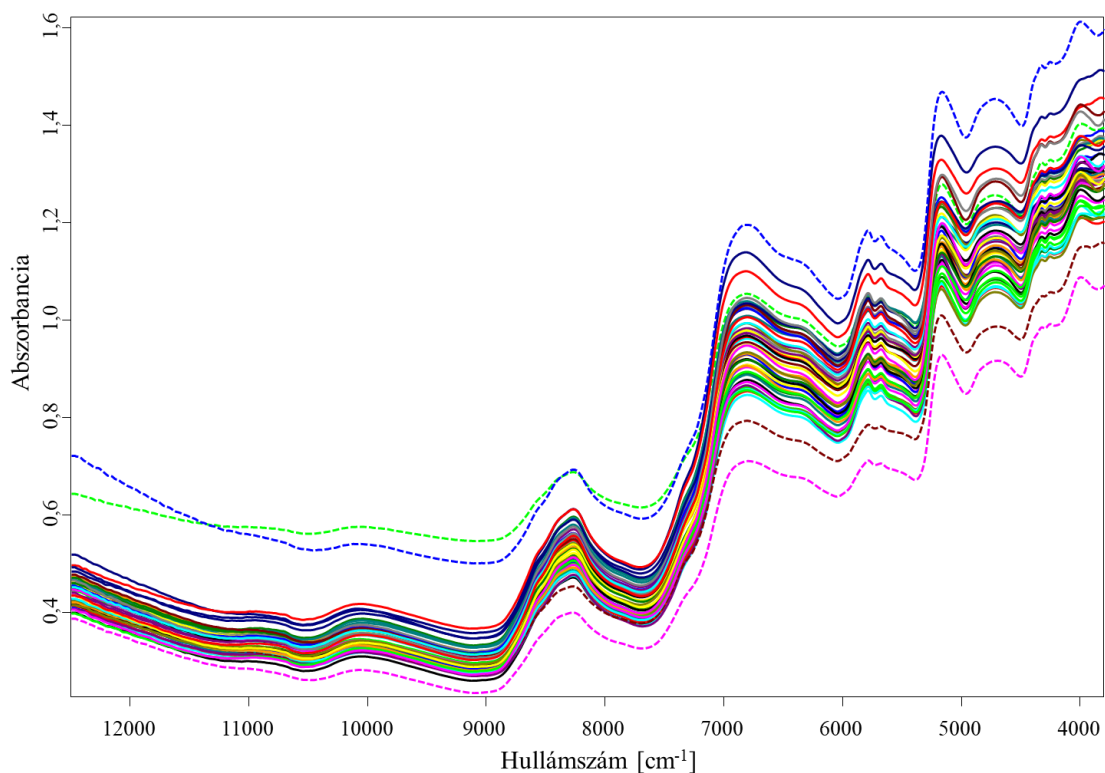
A minták pozícióját nagymértékben meghatározza a bennük található klorogénsavak mennyisége és összetétele is, ami a korrelációs loading-ok alapján egyértelműen látszik. Azok a minták, melyek nagyobb mennyiségű CGA-at tartalmaznak, savasabb karakterűnek bizonyultak az érzékszervi bírálat alapján is. Viszonylag erős korreláció is megfigyelhető a kávé CGA tartalma és érzékszervi savassága között ($r(136)=0,64$), azonban közepes, de negatív kapcsolat ($r(136)=-0,58$) az érzékelt keserűség között ($\alpha=0,01$).

5.2.2. Spektrális adatok minőségi vizsgálata

5.2.2.1. Zöld kávé

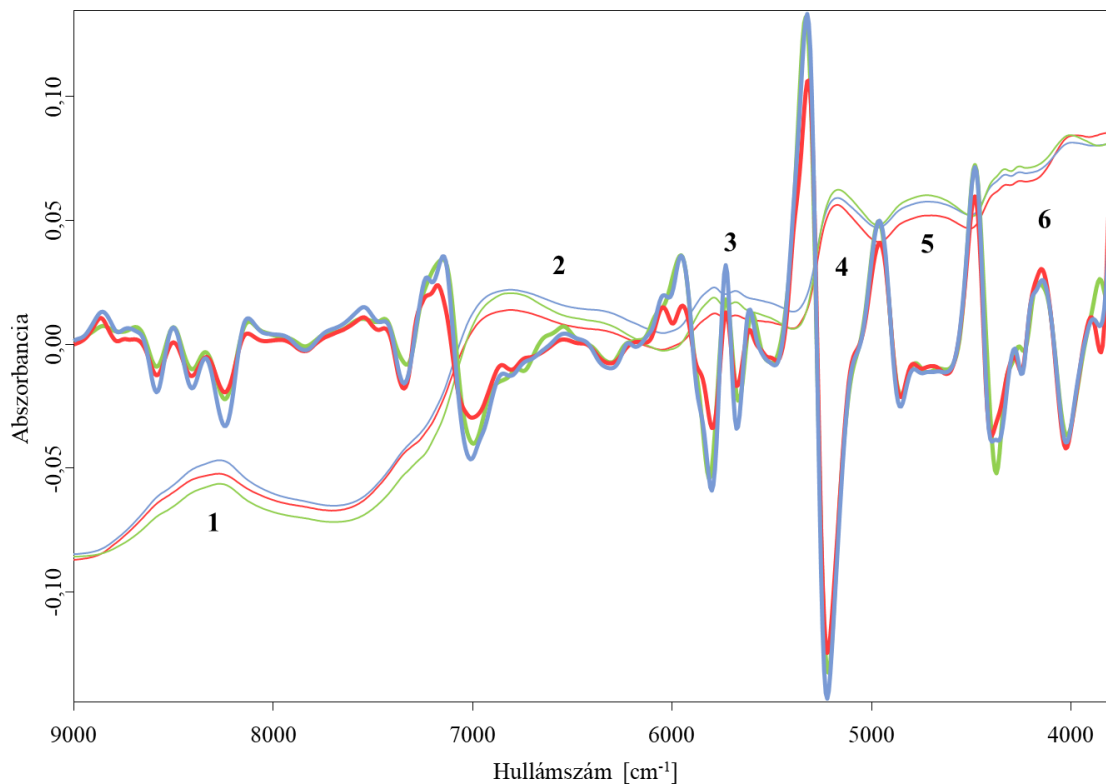
A további statisztikai elemzés előtt, ebben az esetben is célszerű a minták eredeti spektrumának és azok első vagy második deriváltjának minőségi kiértékelését elvégezni. Habár az első derivált transzformációval az additív alapvonal-eltolódás kiküszöbölhető, a spektrumok értelmezése meglehetősen nehézkes, mivel az eredeti spektrum lokális maximumai a deriválás után inflexiós pontként jelennek meg. A NIR spektrumokat vizsgálva, gyakorta lép fel lineáris alapvonal-eltolódás is, ami a második derivált spektrum transzformációval eltávolítható. Ebben az esetben negatív csúcsok jönnek létre, melyeknek lokális minimuma megegyezik az eredeti spektrumban észlelt abszorpciós sávokhoz tartozó maximum értékekkel. Emiatt gyakran alkalmazott módszer a spektrális adatok minőségi elemzése során (DAVIES 2007).

A 22. ábra a zöldkávéminták diffúz reflexiós NIR spektrumát szemlélteti, a 12500-3800 cm^{-1} tartományban. Mivel a spektrumok a minták fizikai tulajdonságairól is hordoznak információt, a kávészemek eltérő színe, formája, az esetleg a felületén maradt ezüsthéj is befolyásolja a spektrális adatokat. Jól látható, hogy a spektrumok karakterisztikája hasonló, azonban szemmel látható eltérés tapasztalható a koffeinmentesített 1-es mintánál (●), a két robusta (● 25, ● 33) és a 49-es (●) arabica kávé esetében.



22. ábra: Zöld kávé minták diffúz reflexiós NIR spektruma. Szaggatott vonallal jelölt minták:
 ● koffeinmentes 1G, ● 25G robusta, ● 33G robusta, ● 49G

Az eredeti spektrumokat célszerű a második deriválással transzformált spektrális adatokkal (23. ábra) együtt vizsgálni, hogy részletesebb képet kaphassunk az esetleges különbségekről. Habár az egyszerű vegyületek, komponensek rezgései a NIR tartományban is viszonylag jól értelmezhetők, a komplex minták minőségi kiértékelése nehéz feladat. Már a zöld kávé is több mint 800 vegyületet tartalmazhat, az egyszerű lineáris és elágazó szerkezetektől kezdve, az összetett ciklikus és heterociklusos szerkezetig (BARBIN et al. 2014). A nagy mintaszám miatt, a jellegzetes rezgési területeket az egyes mintatípusok eredeti, illetve második derivált és SNV módszerekkel transzformált spektrumán (koffeinmentes) és átlagspektrumán (arabica, robusta) mutatom be (23. ábra).



23. ábra: Minta típusok eredeti és második derivált spektrumai

● *Coffea arabica*; ● *Coffea canephora*; ● koffeinmentes *Coffea arabica*

A koffeinmentes minta a *Coffea arabica* fajba tartozik, ezáltal második derivált spektruma számos egyezést mutat az arabica minták átlagspektrumával. A Swiss Water eljárás hatása leginkább a 7500-6300 cm^{-1} (polifenolok, fehérjék, cukrok) és a 4500-4200 cm^{-1} (lipidek, cukrok, klorogénsavak) (19. táblázat) tartományban mutatkozik meg, a spektrum karakterisztikájának megváltozásában, illetve sávtozódásokban. Az arabica és robusta minták spektruma között is egyértelmű eltérések figyelhetők meg, szinte a teljes spektrumon keresztül, ami lehetővé teszi a minták elkülönítését a NIR spektroszkópiai adatok alapján. Ennek részleteire nem fogok kitérni, mivel a dolgozat fókuszában nem a fajok megkülönböztetése áll és számos tudományos publikáció foglalkozik a kérdéssel (VIGNOLI et al. 2014; BURATTI et al. 2015; CAPORASO et al. 2018). A különbségek ellenére a jellegzetes abszorpciós sávok (19. táblázat) együtt is vizsgálhatók, mivel a minták általános összetétele hasonló.

19. táblázat: A kávé jellegzetes elnyelési területei a közeli infravörös tartományban (BURATTI et al. 2015; GIRAUDO et al. 2019; DA SILVA ARAÚJO et al. 2021; RIBEIRO et al. 2021)

Hullámszám [cm ⁻¹]		Rezgés, csoportrezgés	Lehetséges komponensek
8586		C-H vegyértékrezgés 2. felhangja, HC=CH	lipidek
8409	1	C-H vegyértékrezgés 2. felhangja, CH ₂	alifás szénhidrogének, lipidek, koffein
8243		C-H vegyértékrezgés 2. felhangja, CH ₃	szacharóz
7343		C-H vegyérték és C-H deformációs rezgés kombinációja	koffein, cukrok
7020-7006	2	fenolos vegyületek O-H csoportjainak 1. felhangja; N-H szimmetrikus vegyértékrezgés 1. felhangja, (R-C=O-NH ₂)	polifenolok, CGA
6808		O-H vegyértékrezgés 1. felhangja	víz, CGA
6369-6256		N-H vegyértékrezgés 1. felhangja; O-H vegyértékrezgés 1. felhangja	fehérjék, cukrok
5789		C-H szimmetrikus vegyértékrezgés 1. felhangja	koffein, CGA
5675	3	C-H szimmetrikus vegyértékrezgés 1. felhangja	lipidek, cellulóz
5557-5417		O-H kombinációs rezgés; O-H vegyérték és C-H vegyértékrezgés kombinációjának 2. felhangja	víz, cellulóz
5200-5100	4	O-H vegyértékrezgés és HOH deformációs rezgés kombinációja	víz, CGA
5070-5060		N-H aszimmetrikus vegyértékrezgés és N-H deformációs rezgés kombinációja	fehérjék, aromás aminok
4866-4850		N-H vegyérték és C=O vegyértékrezgés kombinációja; N-H aszimmetrikus vegyértékrezgés és N-H deformációs rezgés kombinációja	fehérjék, trigonellin
4743	5	C=O-O vegyértékrezgés 3. felhangja; O-H deformációs és C-O vegyértékrezgés kombinációja	cukrok, trigonellin
4628		C-H vegyértékrezgés/C=O vegyértékrezgés/ C-H deformációs rezgés kombinációja	lipidek, koffein
4400		O-H és C-O vegyértékrezgés kombinációja	cellulóz, glükóz
4360-4329	6	C-H deformációs rezgés 2. felhangja	fehérjék, zsírok, CGA
4246-4232		C-H deformációs rezgés 2. felhangja; C-H és C-C vegyértékrezgés kombinációja	lipidek, CGA
4020	7	C-H deformációs rezgés 2. felhangja, C-H és C-C vegyértékrezgés kombinációja	aromás vegyületek, cellulóz

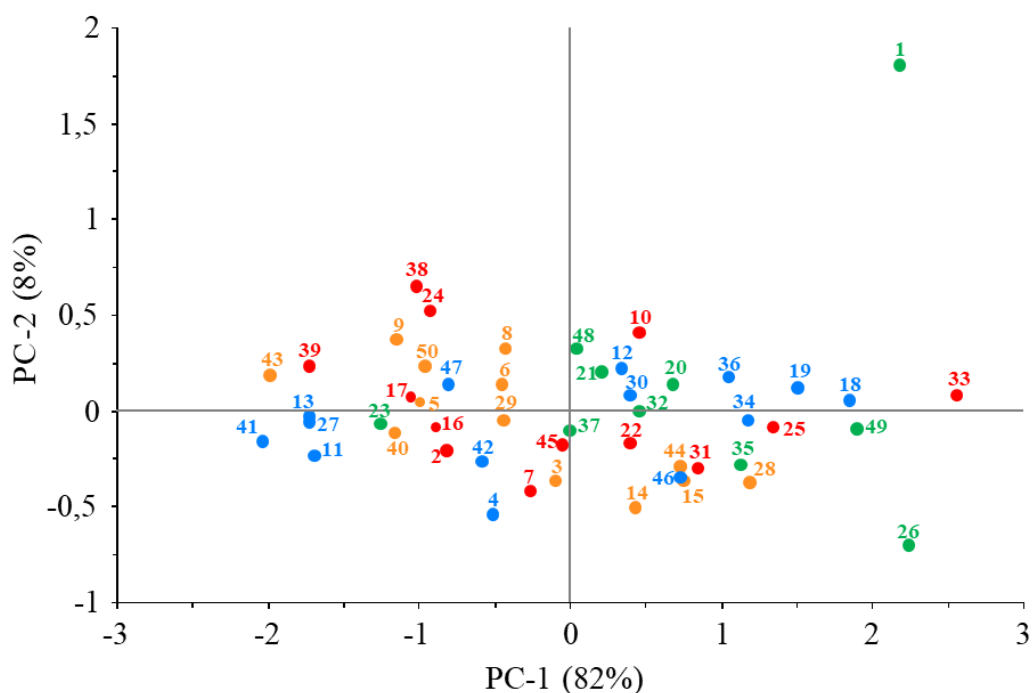
A spektrális adatok feltáró elemzésére főkomponens elemzést (PCA) alkalmaztam, az eredeti és a sor-normalizált (SNV) adatok felhasználásával a 12500-3800 cm⁻¹ tartományra. A modellek validálása random ötszegmenses kereszt-ellenőrzéssel történt. Habár a nyers adatokra létrehozott modellek esetében a magyarázott variancia az első három PC-re vonatkozóan 99%, csak a PC1-nek van hatása a mintázatra. Az SNV-vel előkezelt spektrumok eredményei alapján az első három PC az adatok varianciájának 95%-át (PC1: 82%, PC2: 8%, PC3: 5%) írja le, azonban a mintázat kialakításában mindhárom PC-nek van szerepe (loading plot). A hibatagok alapján spektrális

kiesőnek bizonyultak a koffeinmentes (1) és a két robusta (25, 33) minta, emellett a 49-es mosott, kolumbiai arabica minta is. Ennek a felületén fekete elszíneződés volt megfigyelhető, amely a minta mikroszkópos vizsgálata után egyértelműen látszott. A 24. ábra bal oldalán egy egészséges, míg jobb oldalán a szennyezett kávészem felülete látható.



24. ábra: Egészséges (ábra bal oldalán) és szennyezett (ábra jobb oldalán) kávészem felülete

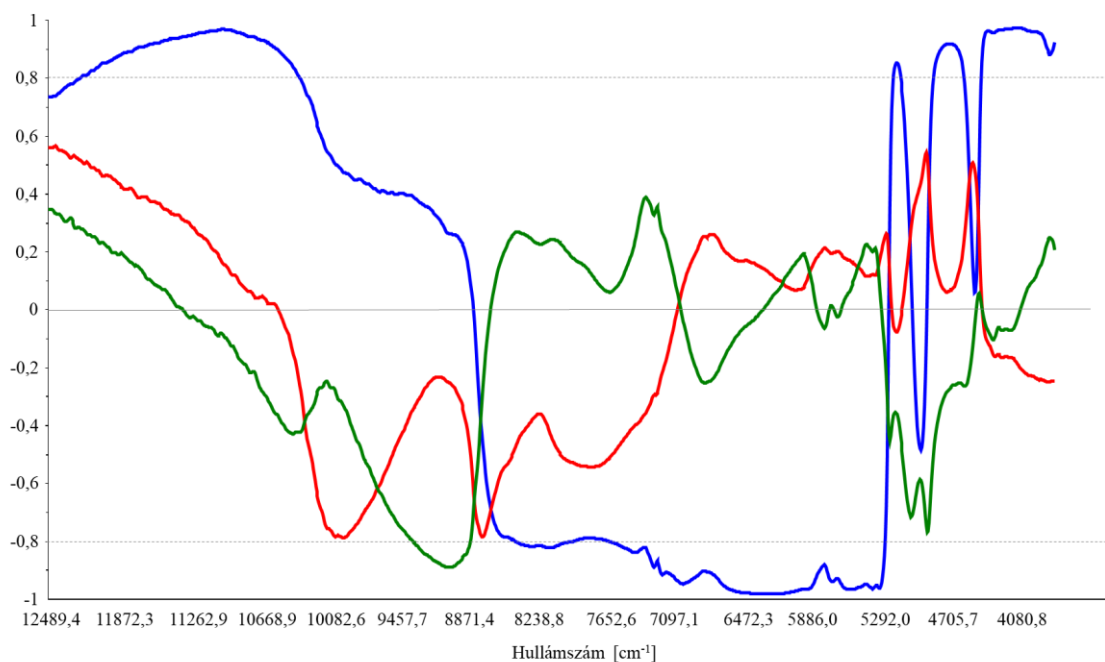
Nemcsak a borászatban, hanem a kávétermesztésben is fontos szerepet tulajdonítanak a termőterületnek az ital érzékszervi jellemzőivel összefüggésben. A termőhelyhez gyakran minőségbeli különbségek párosulnak (termesztett fajtaváltozatok, termesztési körülmények, feldolgozási technológia), melyek meghatározzák a zöld kávé árát. A PCA eredményei a minták származási helyével összefüggésben is vizsgálhatók az adatkészletben rejlő mintázatok elemzésén keresztül. A scores plot alapján elmondható (25. ábra), hogy a négy régióra vonatkozóan a minták nem mutatnak egyértelmű elkülönülést, azonban kisebb klaszterek kialakultak. Hasonló eredményre jutottak GIRAUDO és mtsi. (2019) amerikai és ázsiai minták vizsgálata során. A négy Brazíliából származó minta (20, 21, 32, 48) pl. a scores plot középső részén helyezkedik el (PC1-2 pozitív tartománya), míg a három kenyai kávé (13, 27, 41) a 25. ábra bal szélén (PC1-2 negatív tartománya).



25. ábra: A sor-normalizált zöldkávészektrumokra futtatott PCA scores plot-ja
 ● Afrika ● Ázsia ● Dél-Amerika ● Közép-Amerika

A PC1 mentén a minta pozícióját a kémiai összetétel mellett, erősen befolyásolják a zöldkávész fizikai tulajdonságai is, mint a kávébab megjelenése, formája és színe. Ezek a jellemzők a 12500-9000 cm^{-1} közötti spektrális tartományt befolyásolják leginkább. A meghatározott korrelációs loading-ok (26. ábra) értelmében, erős pozitív korreláció tapasztalható a minták elhelyezkedése és azok fizikai tulajdonságai között. A PC1-nél a legnagyobb score értékek olyan mintákhoz tartoztak, melyek színe barnás, ami a mért színadatokkal is alátámasztható. Viszonylag magas a^* érték volt tapasztalható (>3), ami a piros színezettel áll összefüggésben. Ezek közé a minták közé tartozik, pl. a 28-as, ami honey eljárással készült; a 18-as és 19-es minták, melyek szárítása napon történt, ezáltal alakult ki a sárgás, barnás színezet. Ebből jól látszik, hogy a kávészem színét, ezzel összefüggésben a felületen maradó ezüsthéj mennyiségét alapvetően a feldolgozás módja határozza meg (27. ábra).

A minták pozíciójával pozitív korrelációt mutat még az 5200-5000 cm^{-1} , a 4867-4600 cm^{-1} és a 4400-3800 cm^{-1} hullámszámtartomány is (26. ábra). Ezek a régiók számos vegyület molekuláris rezgéseivel összefüggésben lehetnek, mint pl. a víz, fehérjék, trigonellin, koffein, lipidek, klorogénsavak stb. A PC1 pozitív tartományában elhelyezkedő mintáknál, a fent említett spektrális régiókban mért abszorbancia értéke nagyobb, emellett az ezekhez tartozó csúcsok alakja is megváltozik. Negatív korrelációról beszélhetünk a PC1 esetén, 8600-5300 cm^{-1} között. Ezt a spektrális tartományt főként a kávéban található szénhidrátok, lipidek, fehérjék, klorogénsavak és egyéb polifenolok, illetve a víz molekuláris rezgései befolyásolják (19. táblázat).



26. ábra: A sor-normalizált zöldkávéspektrumokra futtatott PCA korrelációs loading-jai

● PC1 (82%) ● PC2 (8%) ● PC3 (5%)

A PC2-höz és a PC3-hoz tartozó loading értékek alapján erős negatív korreláció tapasztalható a 10100 cm^{-1} és 8700 cm^{-1} hullámszámtartományban. Itt figyelhetők meg a cellulóz (10000 cm^{-1} O-H vegyértékrezgések második felhangja), a különböző alifás szénhidrogének ($8600\text{--}8400\text{ cm}^{-1}$ C-H kötésrezgések) és a szacharóz (8243 cm^{-1} C-H vegyértékrezgés második felhangja) abszorpciós sávjai is.

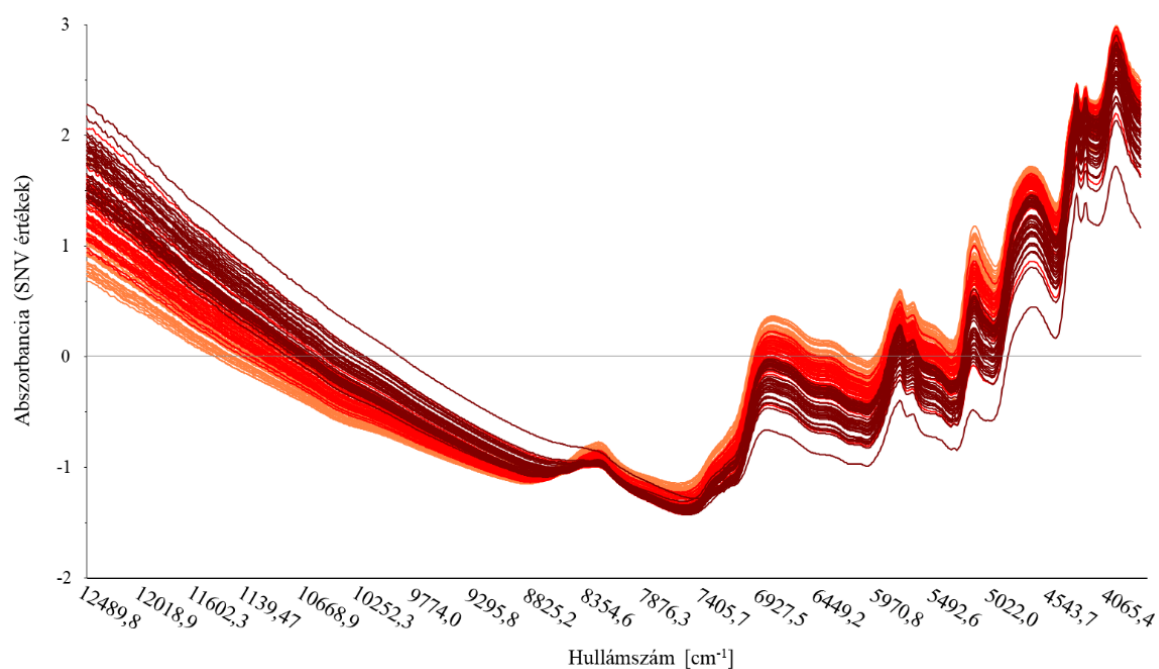


27. ábra: Különböző eljárással készült arabica kávék és egy robusta. a. mosott eljárás (8.); b. száraz eljárás (48.); c. száraz eljárás, robusta (33.); d. mosott, napon szárított (18.); e. Swiss Water koffeinmentesítés (1.); f. honey eljárás (28.)

5.2.2.2. A pörkölés hatása a NIR spektrális adatokra

A pörkölt kávéminták NIR spektrumát a 28. ábra szemlélteti, melyen a különböző pörkölési szintek eltérő színnel vannak jelölve. A zöld- és a pörkölt kávé spektrális jellemzői a fő abszorpciós sávok tekintetében hasonlóak. A 8500 és 3800 cm^{-1} közötti spektrális régióban mért abszorbanca értékek csökkenő tendenciát mutattak a pörkölés előrehaladtával (ahogy a nedvességtartalom csökken), ami egyetértésben van korábbi tanulmányok eredményeivel (ESTEBAN-DÍEZ et al. 2004; ALESSANDRINI et al. 2008; BENES et al. 2020).

Ez a jellegzetesség a 12500-8500 cm^{-1} tartományban (látható tartományhoz közel) megfordul, a mért abszorbanca értékekben növekedés tapasztalható. A pörkölés során a kávészemek egyre sötétebbek lesznek, ami az eredeti zöldes szín megváltozását eredményezi a sárgás színen keresztül, a barnán át, egészen a szinte feketéig. Ez a változás a Maillard-reakció, a pirolízis és a karamellizáció eredménye.

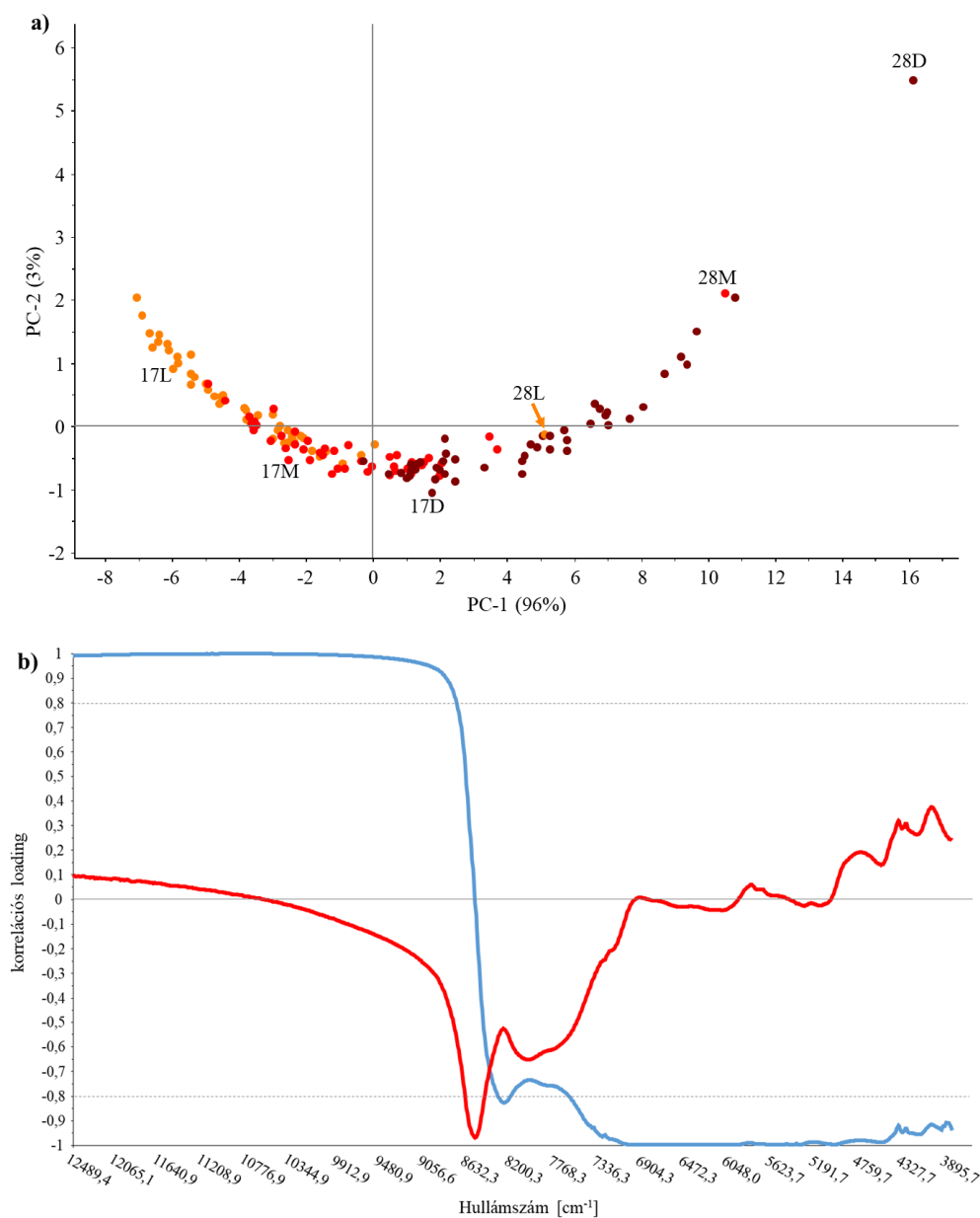


28. ábra: A pörkölt kávéminták spektrumai pörkölési szintenként. ● világos; ● közepes; ● sötét

A különbségek részletesebb vizsgálata érdekében érdemes ebben az esetben is elvégezni a spektrumok deriválását. A második derivált spektrumok (M6.) alapján a pörkölésnek a következő spektrális régiókra van számottevő hatása: 9000-8100 cm^{-1} , 7800-7200 cm^{-1} , 7000 cm^{-1} körül, 6100-5900 cm^{-1} , 5200 cm^{-1} körül, 4700-4200 cm^{-1} . Az ezekkel összefüggésbe hozható abszorpciós sávokat a 19. táblázat foglalja össze.

A pörkölt kávéminták NIR spektrumain is lefuttattam a PCA-t SNV adat-előkezelés után, aminek az eredményeit a 29. ábra szemlélteti. A kiértékeléshez a 9000-3800 cm^{-1} spektrális tartomány adatait használtam fel. A PC1 96%-ban, míg a PC2 3%-ban magyarázza az adatok varianciáját. Általánosságban elmondható, hogy a PC1 mentén a világos pörkölésű kávék inkább negatív értékeket (kivéve a 28L és a 44L mintát) vettek fel (29.a ábra). A közepes pörkölésű kávékhoz negatív és pozitív értékek egyaránt tartoztak, 54 és 46%-ban. Ez alapján valószínűsíthető, hogy ezek a minták illeszkednek legkevésbé az előre meghatározott pörkölési szintekhez. A sötét pörkölésű kávék jellemzően a PC1 tengely mentén a pozitív tartományban helyezkednek el (a 43D mintát kivéve). Mivel a pörkölés körülményei minden kávé esetében azonosak voltak, a mintákból tulajdonképpen egy „pörköltégségi sor” alakult ki. Ezt jól szemlélteti a 28. minta, ami már az első pörkölési szint (világos) esetén olyan spektrális tulajdonságokat mutat, mint sok másik mintánál a harmadik pörkölési szint. Ez összefüggésbe hozható a mért CGA-ak koncentrációjának változásával is, mivel ennél a mintánál volt a legnagyobb arányú a csökkenés (65,5%) a zöld kávé és a világos pörkölésű között. A csökkenés mértéke átlagosan 35,9%-nak adódott. Ezáltal egyfajta csúszás figyelhető meg a pörköltégségi állapot tekintetében az előre definiált pörkölési szintekhez képest. Ezzel szemben, a 17-es jelzésű kávénál a különböző pörkölésű minták (világos (L), közepes (M), sötét (D)) az előre definiált csoportoknak megfelelően helyezkednek el. Ennél a kávénál a zöld kávé és a világos pörkölésű minta között a CGA-ak csökkenésének mértéke 39,5%, ami közel van az átlagértékhez.

A NIR spektrumok alapján az egyes szintek között elmosódnak a határok, ami a zöld kávé minőségben fennálló különbségeiből ered. A korrelációs loading-ok alapján a legjelentősebb spektrális tartományok átfedést mutatnak a zöld kávénál kapott eredményekkel. A PC1 esetén pozitív korrelációról beszélhetünk a 9000-8650 cm^{-1} és negatív korrelációról a 8200-3800 cm^{-1} tartományban (29.b ábra). Legjelentősebb hatása a 7000-4500 cm^{-1} közötti abszorpciós sávoknak van, melyek számos vegyülettel összefüggésbe hozhatók, mint pl. a klorogénsavak, fehérjék, szénhidrátok, koffein, trigonellin és a víz (SANTOS et al. 2016; CATELANI et al. 2018). Tehát a kávé fizikai- és kémiai tulajdonságai is nagymértékben befolyásolják a minták pozícióját.

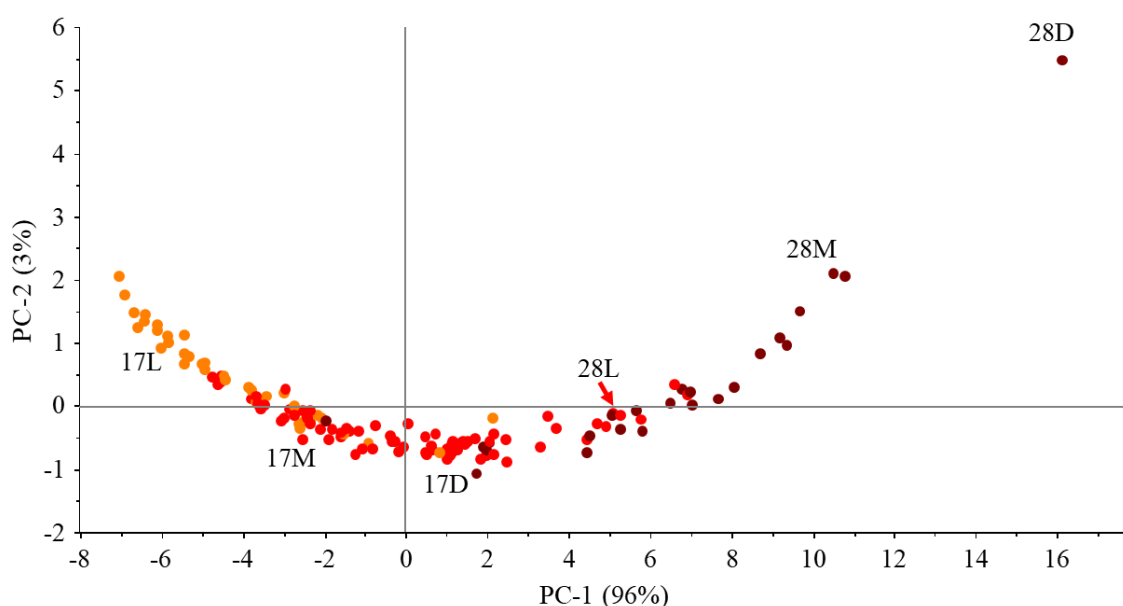


29. ábra: Pörkölt kávé sor-normalizálás (SNV) után kapott NIR spektrumain futtatott főkomponens elemzés eredményei.

- a) scores plot: a minták pörkölési szintje alapján jelölve: ●világos (L); ●közepes (M); ●sötét (D);
 b) loading plot: ● PC1, ● PC2.

A PC2-nek sokkal kisebb hatása van a mintázatra. Erős pozitív korreláció tapasztalható a 8650-8380 cm^{-1} közötti hullámszámtartományban, ami a lipidekre, különböző szénhidrátokra és egyéb alifás szénhidrogénekre jellemző C-H vegyértékrezgések 2. felhangjával hozható összefüggésbe (29.b ábra). Ezen kívül annak a hatását sem szabad elfelejteni, hogy az egyre sötétebb szín kialakulása erősen befolyásolja ezt a spektrális tartományt (28. ábra), ezáltal a minták pozícióját a PC2 mentén.

Az 5.2.1.2. alfejezetben bemutatott mérési eredmények alapján, a CGA-ak koncentrációja között szignifikáns különbség van az egyes pörkölési szinteket figyelembe véve. A pörkölés során a legegységesebb változások az 5-CQA mennyiségében mentek végbe, amelynél az esetek többségében csökkenés volt tapasztalható. Ez alapján, a vizsgált vegyületek közül az 5-CQA lehet a legalkalmasabb a kávé pörköltségi állapotának vizsgálata során. Ha a mintákat a scores plot-on (30. ábra) a meghatározott koncentráció alapján jelöljük, hasonló tendencia tapasztalható, mint a pörkölési szinteknél. Ez alapján jól látható, hogy a 28L minta a spektrális adatok és az 5-CQA mennyisége alapján is a közepes pörkölésű minták közé tartozik, míg a 28M és 28D minták a sötét pörkölésűekhez.



30. ábra: Pörkölt kávé sor-normalizálás (SNV) után kapott NIR spektrumain futtatott főkomponens elemzés scores plot-ja az 5-CQA koncentrációja alapján jelölve:

- 19,05-26,74 mg/g; ● 11,36-19,05 mg/g; ● 3,67-11,36 mg/g

Az eredmények alapján a pörkölési folyamat vizsgálható NIR spektroszkópiával, viszont a folyamatok megértéséhez átfogó adatbázissal kell rendelkezünk a minták fizikai és kémiai jellemzőiről is.

5.2.3. Zöld kávé eredetének vizsgálata

A kiértékeléshez negyvenhat minta, teljes spektrális tartományra (12500-3800 cm^{-1}) vonatkozó adatait használtam fel. A klasszifikációt elvégeztem a nyers adatokon és az SNV, MSC és FD+SNV adat-előkezelések után kapott spektrumokon is. Mivel a spektrális adatok esetében

általános jelenség a multikollinearitás, az algoritmusokat a PCA után kapott score értékeken futtattam le. Az egyes adat-előkezelések esetén az optimális főkomponensek (PC) számát a hegyomlás ábra alapján határoztam meg, így a nyers adatoknál négy, az SNV, illetve MSC után öt, míg az FD+SNV transzformációk alkalmazása után hat főkomponens score értékeit használtam fel. A modellek validálása random, hét-szegmenses kereszt-ellenőrzéssel és 70/30 felosztású teszt-ellenőrzéssel történt. A klasszifikációs folyamatot négy adatkészleten hajtottam végre, négy, három és két-két származási régióval, mint csoportosítási szempont (4. táblázat). A zöld kávé eredetének vizsgálatára korábbi publikációkban főként PCA-t, PLS-DA-t vagy SIMCA-t (*soft independent modeling of class analogy*) alkalmaztak (MARQUETTI et al. 2016; GIRAUDO et al. 2019; OKUBO és KURATA 2019). Ezáltal érdemes lehet további statisztikai módszereket is tesztelni. A Matlab 2019b szoftver „Classification Learner” alkalmazásával számos felügyelt tanítású algoritmus párhuzamosan lefuttatható és optimalizálható. Ezután történik a legjobb gépi tanulási algoritmus azonosítása a teljesítménymetriák alapján.

A kutatás során vizsgált zöld kávék a világ számos pontjáról származtak, ezáltal fizikai tulajdonságaik és kémiai összetételük is jelentős eltéréseket mutat. A minták az adott régión belül is eltérő országokból, termőhelyekről származnak, emellett a rendelkezésre álló adatok alapján nincs olyan paraméter, ami egyértelmű összefüggést mutat a származási régióval. Ezen kívül fontos kiemelni, hogy a növény genetikája is jelentősen befolyásolja a kémiai összetételt, ami egy ilyen diverz adatkészlet esetén nem elhanyagolható szempont.

A négy származási régió vizsgálata során, melyek Afrika (n=14), Ázsia (n=11), Dél- és Közép-Amerika (n=8 és n=13) voltak, minden nem spektrális kieső mintát bevontam a kiértékelésbe. Az eredeti spektrumok PCA score értékein futtatott algoritmusok esetén, a modellek pontossága 30,4-54,3% között változott, ahol az ensemble subspace *k*-NN (*k-nearest neighbors*, legközelebbi szomszéd) algoritmus alkalmazásával érhető el a maximális érték (20. táblázat). Az MSC és az SNV alkalmazásával a modellek teljesítménye romlott (maximális pontosság (*accuracy*, ACC) 50% és 47,8%), ezáltal alkalmazásuk nem ajánlott. FD és SNV adat-előkezelések kombinációjával javítható a modellek pontossága. A statisztikai metrikák alapján, szintén egy ensemble módszer (*bagged tree*) bizonyult a legmegfelelőbbnek. A modell általános pontossága 60,9%-ra nőtt. A helytelenül osztályozott minták száma tizenhét, amik közül a legtöbb a közép-amerikai régióból származott, de a modell hibásan az afrikai régióba sorolta. Az érzékenység (*true positive rate*, TPR) megadja egy adott csoporton belül a helyesen osztályozott minták arányát, míg a specifikusság (*true negative rate*, TNR) az mutatja meg, hogy mennyire képes a modell az adott csoport mintáit a többitől elkülöníteni. Ezek alapján az ázsiai kávéminták azonosítása a legbiztosabb az adatkészlet alapján. A dél- és közép-amerikai régiók esetén csak a minták 50 és

46%-át osztályozta helyesen a modell. A precizításra kapott értékek (0,56-0,67) alapján elmondható, hogy minden egyes régió esetén viszonylag magas a hamis pozitív minták aránya.

20. táblázat: A legjobbnak ítélt osztályozási modellek négy termesztési régió esetén, eltérő adat-előkezelési módszereknél

	Ensemble subspace k -NN		Gaussian Naïve Bayes		LDA		Ensemble bagged trees	
validálás	CV	teszt	CV	teszt	CV	teszt	CV	teszt
adat-előkezelés	nincs		MSC		SNV		FD+SNV	
PC-k	4		5		5		10	
pontosság [%]	54,3	54,5	50	38,5	47,8	46,2	60,9	53,8
érzékenység (TPR)								
<i>Afrika</i>	0,79	1,00	0,71	0,40	0,64	0,75	0,64	0,25
<i>Ázsia</i>	0,45	0,00	0,45	0,33	0,45	0,33	0,82	0,75
<i>Dél-Amerika</i>	0,63	1,00	0,00	0,00	0,13	0,00	0,50	0,00
<i>Közép-Amerika</i>	0,31	0,33	0,62	0,67	0,54	0,67	0,46	1,00
specifikusság (TNR)								
<i>Afrika</i>	0,84	0,88	0,72	0,63	0,72	0,89	0,78	1,00
<i>Ázsia</i>	0,83	1,00	0,91	0,90	0,86	0,90	0,86	1,00
<i>Dél-Amerika</i>	0,89	0,89	0,92	1,00	0,95	1,00	0,95	1,00
<i>Közép-Amerika</i>	0,82	0,63	0,76	0,60	0,76	0,70	0,88	0,40
pontosság (ACC)								
<i>Afrika</i>	0,83	0,91	0,72	0,64	0,70	1,00	0,74	0,91
<i>Ázsia</i>	0,74	0,73	0,80	0,91	0,76	0,73	0,85	1,09
<i>Dél-Amerika</i>	0,85	0,91	0,76	1,00	0,80	0,91	0,87	1,00
<i>Közép-Amerika</i>	0,67	0,55	0,72	0,73	0,70	0,82	0,76	0,64
precizitás (PPV)								
<i>Afrika</i>	0,69	0,75	0,53	0,40	0,50	0,75	0,56	1,00
<i>Ázsia</i>	0,45	0,00	0,63	0,50	0,50	0,25	0,64	1,00
<i>Dél-Amerika</i>	0,56	0,67	0,00	0,00	0,33	0,00	0,67	0,00
<i>Közép-Amerika</i>	0,40	0,25	0,50	0,33	0,47	0,40	0,60	0,33

TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság; PPV: precizitás

Az osztályozási folyamatot a dél- és közép-amerikai régiók összevonásával is elvégeztem. Ezáltal a vizsgált csoportok Afrika (n=14), Amerika (n=21) és Ázsia (n=11) voltak. Az eredményeket a 21. táblázat foglalja össze. Míg a spektrum-transzformáció nélkül futtatott modell esetén a Gaussian Naïve Bayes, addig a többinél az SVM (*support vector machine*, támogató vektorgépek) módszer (medium Gaussian, linear) alkalmazásával volt elérhető a legnagyobb modell pontosság. MSC és SNV hatására a modellek érzékenysége 0,09, illetve 0, tehát az Ázsia csoporton belül helyesen osztályozott minták száma egy és nulla volt. A teljesítménymetriák az FD+SNV

transzformációk után kapott modell bizonyult optimálisnak, amelynél a pontosság 69,6% volt. Míg a négy régió vizsgálata során a modellek érzékenysége az ázsiai mintákra volt a legnagyobb (0,82), addig három régiónál a legkisebb (0,45). Specifikusság tekintetében azonban jól teljesít a csoport, amely értelmében alacsony a hamis pozitív minták aránya. Az amerikai mintáknál határozta meg a legkisebb specifikusságot, mivel a tizennégy félreosztályozott mintából kilenc ebbe a csoportba került, melyből öt eredetileg az ázsiai csoportba tartozott. Az eredmények azt mutatják, hogy a modellek teljesítőképessége nem megfelelő az amerikai és ázsiai kávéminták osztályba sorolása során. A kialakuló klaszterek valószínűleg olyan tényezőkkel magyarázhatók, melyek számomra ismeretlenek, pl. a növény genotípusa.

21. táblázat: A legjobbnak ítélt osztályozási modellek három termesztési régió esetén, eltérő adat-előkezelési módszereknél

	Gaussian Naïve Bayes		medium Gaussian SVM		medium Gaussian SVM		linear SVM	
validálás	CV	teszt	CV	teszt	CV	teszt	CV	teszt
adat-előkezelés	nincs		MSC		SNV		FD+SNV	
PC-k	4		5		5		10	
pontosság [%]	67,4	61,5	63	61,5	60,9	53,8	69,6	69,2
érzékenység (TPR)								
<i>Afrika</i>	0,79	0,50	0,64	0,50	0,57	0,50	0,71	0,75
<i>Amerika</i>	0,76	0,83	0,90	1,00	0,95	0,83	0,81	0,83
<i>Ázsia</i>	0,36	0,33	0,09	0,00	0,00	0,00	0,45	0,33
specifikusság (TNR)								
<i>Afrika</i>	0,91	1,00	0,88	1,00	0,97	0,89	0,91	1,00
<i>Amerika</i>	0,64	0,43	0,48	0,29	0,32	0,29	0,64	0,71
<i>Ázsia</i>	0,91	0,90	1,00	1,00	1,00	1,00	0,94	0,80
pontosság (ACC)								
<i>Afrika</i>	0,87	1,00	0,80	1,00	0,85	0,91	0,85	1,09
<i>Amerika</i>	0,70	0,73	0,67	0,73	0,61	0,64	0,72	0,91
<i>Ázsia</i>	0,78	0,91	0,78	0,91	0,76	0,91	0,83	0,82
precizitás (PPV)								
<i>Afrika</i>	0,79	1,00	0,69	1,00	0,57	0,67	0,77	1,00
<i>Amerika</i>	0,64	0,56	0,59	0,55	0,54	0,50	0,65	0,71
<i>Ázsia</i>	0,57	0,50	1,00	0,00	0,00	0,00	0,71	0,00

TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság; PPV: precizitás

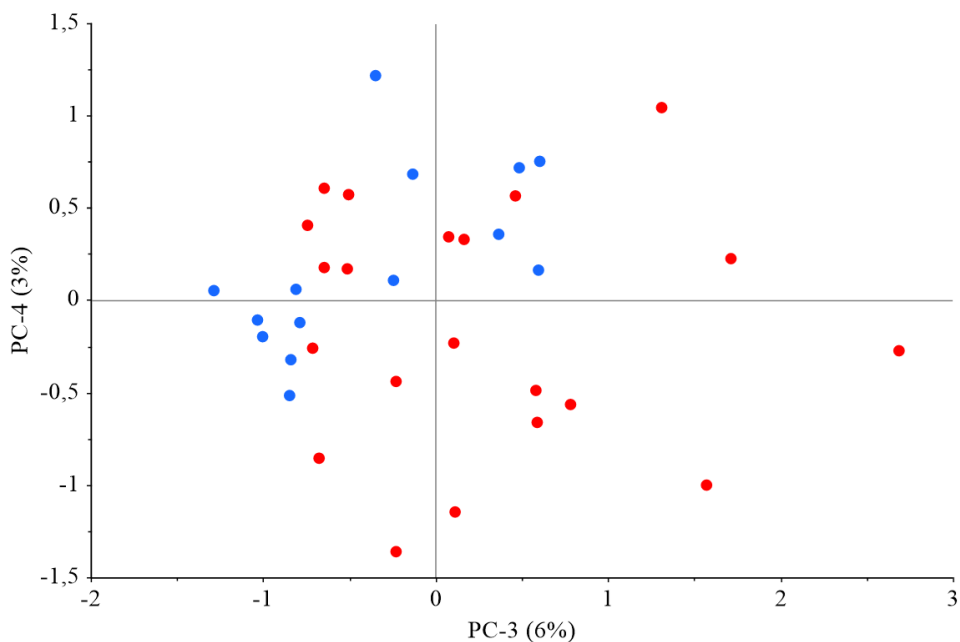
Az eredmények fényében érdemes megvizsgálni az egyes régiókat bináris rendszerekben is. Az afrikai és amerikai minták osztályozása során, mindhárom adat-előkezelési eljárás után a kvadratikus SVM alkalmazásával érhető el a legnagyobb pontosság (80-88,6%) (22. táblázat).

22. táblázat: A legjobbnak ítélt osztályozási modellek eltérő adat-előkezelési módszereknél Afrika-Amerika termesztési régió tekintetében

	weighted KNN		quadratic SVM		quadratic SVM		quadratic SVM	
validálás	CV	teszt	CV	teszt	CV	teszt	CV	teszt
adat-előkezelés	nincs		MSC		SNV		FD+SNV	
PC-k	4	4	5	5	5	5	10	10
pontosság [%]	80	80	82,9	80	80	80	88,6	80
érzékenység (TPR)								
<i>Afrika</i>	0,71	0,50	0,79	0,75	0,79	0,50	0,86	1,00
<i>Amerika</i>	0,86	1,00	0,86	0,83	0,81	1,00	0,90	0,67
specifikusság (TNR)								
<i>Afrika</i>	0,86	1,00	0,86	0,83	0,81	1,00	0,90	0,67
<i>Amerika</i>	0,77	1,00	0,79	0,75	0,79	0,50	0,86	1,00
precizitás (PPV)								
<i>Afrika</i>	0,77	1,00	0,79	0,75	0,73	1,00	0,86	0,67
<i>Amerika</i>	0,82	0,75	0,86	0,83	0,85	0,75	0,90	1,00

TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság; PPV: precizitás

Az optimális modellhez FD+SNV transzformációval jutottam, melynél minden teljesítménymetrika értéke legalább 0,86 volt. Mindkét csoportból két-két mintát osztályozott félre a modell. Az első két főkomponensnek (PC1: 74%; PC2:13%) nincs egyértelmű hatása a minták pozíciójára a származási hellyel összefüggésben. Annak ellenére, hogy az adatkészlet varianciáját csak kisebb mértékben magyarázza (6%), a két régióba tartozó minták elkülönítésében nagyobb szerepe van a PC3-nak, de a két csoport között még így is részleges átfedés tapasztalható, amit a 31. ábra szemléltet. A PC3-ra kapott faktorsúlyok (loadingok) alapján (M7.) az 5646 cm⁻¹ (C-H vegyérték és C-H deformációs rezgések kombinációja,) és az 5315 cm⁻¹ (C=O kötésrezgés 1. felhangja) körüli abszorpciós sávoknak jelentős hatása van a minták pozitív tartományban való elhelyezkedésére. Ezzel szemben negatív kapcsolat van az egyes minták pozíciója és az 5878 cm⁻¹ (C-H vegyértékrezgés 1. felhangja), ill. 4721 cm⁻¹ (N-H vegyértékrezgés és az amid III kombinációja) körül mért abszorbancia között. Ezek a területek számos vegyület, köztük a klorogénsavak, koffein, fehérjék, illetve a víz molekuláris rezgéseivel is kapcsolatba hozhatók.



31. ábra: A kvadratikus támogató vektorgépek (SVM) módszerrel kapott osztályozás score ábrája a harmadik és negyedik főkomponens esetében. ● Amerika; ● Afrika

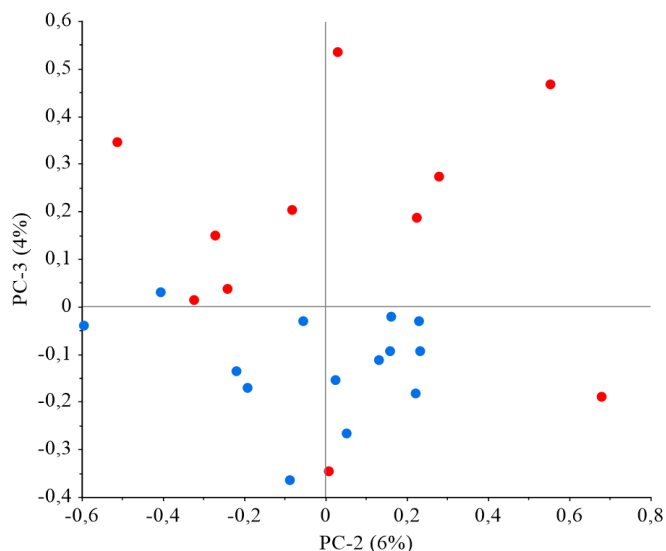
Az afrikai és ázsiai minták vonatkozásában az egyes adat-előkezelésekhez tartozó modellek pontossága 80-92% között változott. Az optimalizálási folyamat során minden esetben a kernel Naïve Bayes algoritmus alkalmazása bizonyult a legjobbnak. A modellek fontosabb statisztikai paramétereit a 23. táblázat foglalja össze. SNV alkalmazásával a modell pontossága 92% volt, emellett az osztályokra kapott érzékenység (TPR) és specifikusság (TNR) értékek is közel voltak 1,00-hez.

23. táblázat: A legjobbnak ítélt osztályozási modellek eltérő adat-előkezelési módszereknél Afrika-Ázsia termesztési régió tekintetében

	kernel Naïve Bayes	kernel Naïve Bayes	kernel Naïve Bayes	kernel Naïve Bayes
validálás		10-fold CV		
adat-előkezelés	nincs	MSC	SNV	FD+SNV
PC-k	4	5	5	10
pontosság [%]	88	88	92	80
érzékenység (TPR)				
<i>Afrika</i>	0,86	0,86	0,93	0,71
<i>Ázsia</i>	0,91	0,91	0,91	0,91
specifikusság (TNR)				
<i>Afrika</i>	0,91	0,91	0,91	0,91
<i>Ázsia</i>	0,92	0,86	0,93	0,71
precizitás (PPV)				
<i>Afrika</i>	0,92	0,92	0,93	0,91
<i>Ázsia</i>	0,83	0,83	0,91	0,71

TPR: valódi pozitív arány, érzékenység; TNR: valódi negatív arány, specifikusság; ACC: pontosság; PPV: precizitás

A minták régió alapján történő elkülönítésére a scores plot-ok alapján leginkább a PC3-nak (32. ábra) van hatása. A korrelációs loading-jai alapján (M8.) az ezt leginkább befolyásoló spektrális tartomány az $5230\text{-}4880\text{ cm}^{-1}$, amelynek korábbi tanulmányok is tulajdonítottak szerepet (GIRAUDO et al. 2019; OKUBO és KURATA 2019).



32. ábra: A kernel Naïve Bayes módszerrel kapott osztályozás score ábrája a második és harmadik főkomponens esetében. ● Ázsia; ● Afrika

5.2.4. Részleges legkisebb négyzetek elve regresszió (PLSR) eredményei

5.2.4.1. Zöldkáv

A viszonylag alacsony mintaszám és azok minőségében tapasztalat eltérések miatt, a regressziós modellek létrehozásához minden minta esetén három spektrumot használtam fel ($n=138$). A kiértékelés során a $10000\text{ és }3800\text{ cm}^{-1}$ közötti spektrális tartományt vizsgáltam, ami 800 változót jelentett. Becslési függvényeket hoztam létre a zöldkáv pH-értékének, koffein, 5-, 4-, 3-CQA, 3,5- és 4,5-diCQA, illetve a 3-FQA tartalmának kvantitatív vizsgálatára. Az egyes mátrixok és vizsgált paraméterek esetében a becslési függvények optimalása eltérő stratégiák mentén, empirikus úton történik. Ezáltal a regressziós összefüggések statisztikai paramétereinek javítására spektrum transzformációs módszereket és különböző változókiválasztási eljárásokat, illetve ezek kombinációit használtam. Először a nyers spektrális adatok és a referencia adatok felhasználásával hoztam létre a modelleket, majd külön-külön SNV (sor-normalizálás), MSC (átlagos szórás korrekció), FD+SNV (első deriválás és sor-normalizálás kombinációja), SD+SNV (második deriválás és sor-normalizálás kombinációja) és OSC (ortogonális jelkorrekció) adat-előkezelés után is. Ezután minden modellnél lefuttattam a VIP, az iPLS (20- és 40-szegmens) és a GA (ablakszélesség 30 és 50) változókiválasztási módszereket. Az összefüggések validálását random,

ötszeggmenses kereszt-ellenőrzéssel és teszt-ellenőrzéssel (70/30) is elvégeztem. A kapott eredményeket a teljes tartományra létrehozott PLS modellhez viszonyítva értékeltem. A kiindulási és a legjobbnak ítélt PLS modellek statisztikai paramétereit a 24. táblázat foglalja össze. A modellek összehasonlításához elsősorban a determinációs együtthatókat (R^2 ; Q^2) és az átlagos négyzetes hiba értékeket vettem figyelembe. A determinációs együtthatónál minél nagyobb értékre, míg az átlagos négyzetes hiba négyzetgyökénél (RMSE) minél kisebb értékre kell törekedni. Az optimális modell kiválasztásánál ezen kívül a látens változók (LV-k) száma is fontos szempont, az alul- vagy túlillesztés elkerülése miatt.

A zöld kávé pH-jának becslésénél a legjobb statisztikai paraméterekkel rendelkező modelleket az FD és a SNV spektrum transzformációk kombinációjával, 40-szeggmenses iPLSR esetén, ill. OSC alkalmazása után értem el. Korábbi tanulmányok alapján a kávé savasságával a 8278-8132 cm^{-1} (CH és CH_2 csoportok vegyértékrezgéseinek 2. felhangja), illetve a 9407-7473 cm^{-1} közötti spektrális területek hozhatók összefüggésbe (RIBEIRO et al. 2011; SANTOS et al. 2016; ARAÚJO et al. 2020). Ezek az általam létrehozott modellek esetében is jelentősnek bizonyultak. Az iPLSR-nél a 10000-9395 cm^{-1} (C-H rezgések 2. felhangja, O-H vegyértékrezgés 2. felhangja), a 7848-7545 cm^{-1} (CH_2 csoportok vegyértékrezgéseinek 1. felhangja), 7228-6928 cm^{-1} (C-H kombinációs rezgések, O-H vegyértékrezgés 1. felhangja), 5686-5076 cm^{-1} (C-H rezgések 1. felhangja, O-H kombinációs rezgések, C=O rezgés 2. felhangja) közötti spektrális tartományok kerültek kiválasztásra az RMSECV értékek alapján. Az RMSE értéke mindkét modellnél 0,05-nek adódott, a kereszt-ellenőrzés és a teszt-ellenőrzés esetén is. A R^2 alapján az iPLS modell jobb teljesítményt mutat, mint az OSC modell. A látens változók száma az előbbinél 7, az utóbbinál pedig 3 volt, ami jól mutatja az OSC dimenzióredukciós hatását. A modellek mérési tartománya az iPLS-nél pH=5,61-6, OSC esetén pedig pH=5,55-6 volt. A zöldkávé pH-jának becslésére ARAÚJO és munkatársai (2020) is kidolgoztak PLSR modelleket. Ezek statisztikai paramétereit hasonlóak az általam létrehozott becslési függvényekéhez, azonban kisebb RMSE értéket tudtak elérni. Ez valószínűleg azzal van összefüggésben, hogy a kutatás során vizsgált minták mind Brazíliából származtak, ezáltal a spektrális tulajdonságaik között nincsenek olyan nagy mértékű eltérések, mint az általam vizsgált, földrajzilag sokkal nagyobb variációt mutató minták között. Ezen kívül az általuk vizsgált mérési tartomány csak pH=5,7-5,84 közé esett, ami szintén a származási hellyel lehet összefüggésben.

A koffein becslésére létrehozott kiindulási modellnél az R^2 értéke viszonylag kicsi (0,75), azonban a legnagyobb problémát a látens változók száma jelenti (túlillesztés). Emiatt elengedhetetlen a spektrális adatok transzformációja. Az adat-előkezelés és a változókiválasztás tekintetében szintén az FD+SNV kombinációja és az iPLS40; illetve az OSC mutatkozott a legjobb stratégiának. A

statisztikai paraméterek alapján az OSC adat-előkezeléssel létrehozott modell rendelkezik a legjobb teljesítőképességgel a mennyiségi becslés során. A koffeinre jellemző és a regressziós összefüggések létrehozásához alkalmazott NIR abszorpciós sávok a következők voltak: 9542- 9241, 9079-8933, 8462-8316, 7999-7853, 7074-6927, 5994-5693, 5376-5230 és 4759-4304 cm^{-1} . Ezek a területek elsősorban, C-H kötések 1., 2. és 3. felhangjaival, ill. kombinációs rezgéseivel, N-H és O-H kötésrezgések kombinációival, ezen kívül C-H és C-C kötések kombinációs rezgéseivel állnak kapcsolatban. A vonatkozó molekuláris rezgéseket RIBEIRO és mtsi. (2021) kutatásukban részletesen tárgyalják.

24. táblázat: A PLS regresszióval zöld kávéra kapott modellek fontosabb statisztikai paraméterei. A koffein, 5-CQA, 4-CQA, 3-CQA, 3,5-diCQA, 4,5-diCQA, 3-FQA eredményeinek mértékegysége: mg/g.

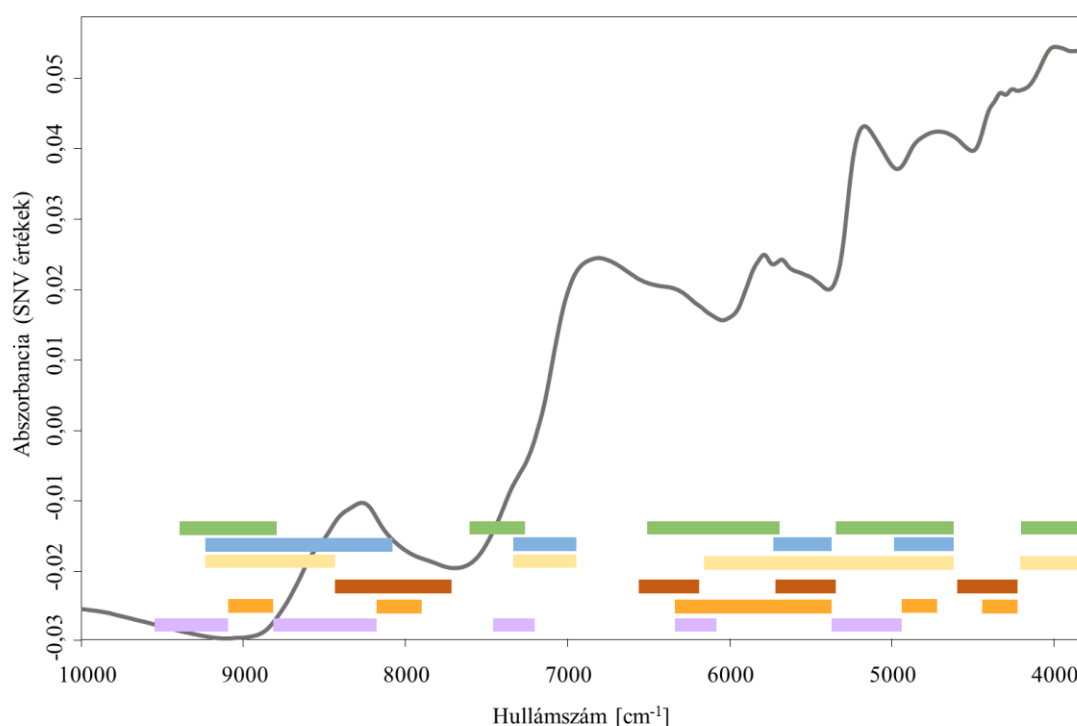
Paraméter	Adat-előkezelés	Változó-kiválasztás	Mintaszám	LV	Kereszt-ellenőrzés				Teszt-validálás				Mérési tartomány
					R ²	Q ²	RMSEC	RMSECV	R _t ²	Q _t ²	RMSEC	RMSEP	
pH	-	-	115	5	0,72	0,67	0,06	0,06	0,75	0,72	0,05	0,06	5,61-6
	FD+SNV	iPLS40	118	7	0,85	0,80	0,04	0,05	0,84	0,85	0,04	0,05	5,58-6
	OSC	-	125	3	0,81	0,78	0,05	0,05	0,82	0,77	0,05	0,05	5,55-6
koffein	-	-	129	10	0,75	0,65	0,43	0,53	0,75	0,63	0,47	0,43	8,87-13,49
	FD+SNV	iPLS40	124	8	0,86	0,80	0,35	0,42	0,85	0,84	0,34	0,45	8,32-13,6
	OSC	-	138	2	0,94	0,89	0,23	0,32	0,95	0,81	0,22	0,30	8,32-13,6
5-CQA	-	-	109	7	0,73	0,68	3,22	3,58	0,75	0,63	3,12	3,71	30,81-61,98
	FD+SNV	GA30	109	7	0,79	0,69	2,77	3,50	0,74	0,67	3,06	3,60	37,19-63,93
	OSC	-	107	2	0,99	0,98	0,52	0,86	0,99	0,98	0,60	0,70	33,54-61,98
4-CQA	-	-	130	7	0,83	0,80	1,16	1,28	0,83	0,80	1,18	1,22	5,52-16,92
	MSC	GA50	127	8	0,88	0,84	1,00	1,16	0,87	0,86	1,03	1,07	5,52-16,92
	OSC	-	135	3	0,97	0,90	0,48	0,91	0,98	0,86	0,43	0,91	5,52-16,92
3-CQA	-	-	130	6	0,81	0,79	1,18	1,26	0,82	0,79	1,13	1,34	3,27-13,80
	MSC	GA50	125	8	0,88	0,84	0,86	1,01	0,88	0,84	0,83	1,07	3,27-13,80
	OSC	-	130	2	0,90	0,86	0,75	0,91	0,91	0,87	0,73	0,84	3,27-13,80
3,5-diCQA	-	-	123	6	0,72	0,68	0,51	0,55	0,72	0,70	0,52	0,50	2,24-6,68
	FD+SNV	GA50	121	5	0,76	0,72	0,47	0,52	0,78	0,74	0,46	0,48	2,24-6,68
	OSC	-	131	3	0,99	0,99	0,11	0,15	0,99	0,99	0,08	0,09	2,24-6,68
4,5-diCQA	-	-	125	7	0,70	0,65	0,33	0,36	0,71	0,63	0,34	0,36	1,71-4,34
	MSC	GA30	121	6	0,77	0,74	0,28	0,30	0,79	0,76	0,27	0,29	1,71-4,34
	OSC	-	126	4	0,99	0,99	0,06	0,08	0,99	0,99	0,06	0,05	1,71-4,34
3-FQA	-	-	128	6	0,72	0,66	0,45	0,50	0,73	0,67	0,43	0,50	0,95-5,13
	SNV	GA30	114	8	0,77	0,70	0,27	0,31	0,77	0,72	0,27	0,31	0,95-3,36
	OSC	-	133	3	0,99	0,99	0,10	0,13	0,99	0,96	0,09	0,12	0,95-4,86

*A legjobb modellek félkövérrel vannak jelölve.

5-CQA: kinasav-5-O-kávésav; 4-CQA: kinasav-4-O-kávésav; 3-CQA: kinasav-3-O-kávésav; 3,5-diCQA: kinasav-3,5-O-dikávésav; 4,5-diCQA: kinasav-4,5-O-dikávésav; 3-FQA: kinasav-3-O-ferulasav; CGA: klorogénsavak

RMSEC: átlagos négyzetes hiba a kalibrációs adatkészletre nézve; RMSECV: átlagos négyzetes hiba a kereszt-ellenőrzés esetén; RMSEP: átlagos négyzetes hiba a teszt adatkészletre nézve; R²: kalibrációs adatkészletre vonatkozó determinációs együttható kereszt-ellenőrzés esetén; R_t²: kalibrációs adatkészletre vonatkozó determinációs együttható teszt-ellenőrzés esetén; Q²: kereszt-ellenőrzésre vonatkozó determinációs együttható; Q_t²: teszt-ellenőrzésre vonatkozó determinációs együttható; RMSEC_t: átlagos négyzetes hiba a kalibrációs adatkészletre nézve teszt-ellenőrzésnél

A klorogénsavak esetén a kiindulási PLS modellek, (amelyeknél se adat-előkezelés -csak átlaggal való centrálás-, se változókiválasztás nem történt), R^2 értéke 0,63-0,83 között változott. Míg az 5-CQA és a 3,5-diCQA mennyiségi becslése során FD+SNV adat-előkezelés, addig a többi vegyület esetén az MSC vagy az SNV szóráskorrekció, illetve a GA változókiválasztás kombinációja adta a legjobb eredményeket. Általánosan elmondható azonban, hogy a leggyakrabban alkalmazott adat-előkezelési módszerek és a különböző változókiválasztási eljárások nem javították nagymértékben a modellek predikációs képességét. A GA által kiválasztott, az egyes klorogénsavak becsléséhez lényegesnek ítélt spektrális tartományokat a 33. ábra szemlélteti.



33. ábra: A változókiválasztási eljárások által lényegesnek ítélt NIR elnyelési területek a klorogénsavak vizsgálatánál ● 5-CQA ● 4-CQA ● 3-CQA ● 3,5-diCQA ● 4,5-diCQA ● 3-FQA

A vizsgált vegyületek molekuláris szerkezete hasonlóságokat mutat, emiatt természetesen jelentkeznek átfedések az egyes modellek abszorpciós sávjai között, különösen a három monokávésav és a két di-kávésav esetén, melyek egymás szerkezeti izomerjei. Ezek közül egyedül az 5-CQA-nál áll rendelkezésre a vegyület NIR spektruma (RIBEIRO et al. 2021), ami alapján a kiválasztott területek releváns adatokat tartalmaznak a molekuláris rezgésekkel összefüggésben. Az előbb hivatkozott kutatás során sikeresen hoztak létre PLS modellt különböző természetes (arabica, robusta) és mesterséges kávématrixok, bináris keverékek és 5-CQA-val adalékolt minták NIR spektrumai alapján (mintaszám 54; $R^2=0,96$; RMSEP=0,27%; LV=5). Az általam létrehozott OSC modell, 2 látens változó használatával 33,54-61,98 mg/g mérési tartományban 0,99/0,98/0,98 $R^2/Q^2/Q_t^2$; és 0,52/0,86/0,70 mg/g RMSEC/RMSECV/RMSEP értékekkel jellemezhető.

Összességében elmondható, hogy a klorogénsavak esetén a statisztikai paraméterek tekintetében optimális modelleket OSC adat-előkezelés után kaptam. A determinációs együttható értéke 0,86 és 0,99 között mozgott, ezen kívül az RMSE értéke minden vegyület esetén kisebb volt, mint 1 mg/g, a kereszt- és teszt-ellenőrzés eredményeit tekintve. Az OSC hatására a látens változók száma 6-8-ról, 2-4-re változott, ezáltal a modellek komplexitása csökkent. A zöldkávémintákkal kapcsolatos modellek jellemző adatait az 24. táblázatban foglaltam össze.

5.2.4.2. Pörkölt kávé

A PLS modellek fejlesztéséhez 138 minta spektrális és referencia adatait használtam fel. Minden esetben az átlagértékekkel dolgoztam. A kiértékelés során a 9000-3800 cm^{-1} spektrális tartományt (680 változó) használtam, mivel a 12500-9000 cm^{-1} közötti szakaszt erősen befolyásolja a pörkölés során kialakult szín. A sötétebb szín a mért abszorbancia növekedését eredményezi, ami a molekuláris rezgésekkel kapcsolatos információkat elfedi. A végrehajtott spektrumtranszformációk és változókiválasztási technikák megegyeznek azokkal, amiket a zöldkávénál alkalmaztam. A modellek validálásához ebben az esetben is random, ötszempenses kereszt-ellenőrzést és teszt-ellenőrzést használtam. A kiindulási és a legjobbnak ítélt modellek legfontosabb statisztikai paramétereit a 25. táblázat foglalja össze.

Az eredmények alapján általánosságban elmondható, hogy a pörkölt kávé spektrumainak kiértékelése során az iPLS változókiválasztás segítségével felállított modellek statisztikai jellemzői nem javultak a teljes spektrum felhasználásával kialakított modellekhez képest. A legeredményesebb változókiválasztási eljárásnak minden esetben a GA bizonyult, amely az optimalizáláshoz minden változót és azok lehetséges kombinációit felhasználja. Az eljárás eredményessége magyarázható a spektrális adatok komplexitásával, ami a kávé molekuláris rezgéseinek összetettségéből ered.

A kávéital ízét jelentősen meghatározza a benne található savas karakterű vegyületek mennyisége, amire jó közelítés lehet a szemeskávéminták pH-jának vizsgálata, mivel az erős korrelációt mutat az érzékelt savassággal (5.2.1.2. alfejezet) (BROLLO et al. 2009). A pörkölt kávé pH-értékének becslésére kidolgozott kiindulási modell 0,75/0,71 R^2/Q^2 értékekkel és pH=0,07/0,08 RMSEC/RMSECV értékekkel rendelkezik. A statisztikai paramétereken önmagában a változókiválasztás nem tudott javítani, ezáltal mindenképpen szükséges valamilyen spektrumtranszformáció alkalmazása. Az FD+SNV és a GA50 futtatása után az R^2/Q^2 értéke 0,98/0,85-re emelkedett, míg a négyzetes hiba értéke pH=0,05/0,06-ra csökkent. A modellhez 350 változó felhasználására és 8 látens változóra volt szükség. Az optimális modellt OSC alkalmazása után

értem el, ahol az összefüggés ($R^2/Q^2=0,93/0,89$) javítása mellett az RMSEC/RMSECV értéke is $pH=0,04/0,05$ -re csökkent.

A kávé alapvetően a benne található koffein stimuláns hatása miatt fogyasztjuk, ezáltal mennyisége fontos minőségi paraméter. A kiindulási PLS regresszióval 10 látens változó használata mellett $0,74/0,66$ volt az R^2/Q^2 , míg az RMSEC=0,50 mg/g és az RMSECV=0,60 mg/g adódott (9,90-15,43 mg/g mérési tartomány). A látens változók nagy száma volt tapasztalható ezen kívül az SNV és az MSC adat-előkezelés után is (minimum 8 látens változó), amely hasonlóságot mutat a PIZARRO és mtsi. (2007) által publikált eredményekkel (7-17 látens változó, 191 kávé minta, arabica, robusta és blendek). Az összefüggés létrehozásához harminc mintát ki kellett zárni az optimalizálás során, amik a függvény segítségével nem modellezhetők jól, mert torzítják a becslést és rontják a pontosságot is. Ennek oka valószínűleg, hogy míg a spektrális adatokat a pörkölés jelentősen befolyásolja, addig a koffein koncentrációjára nincs számottevő hatással. Ahhoz, hogy az Y magyarázott varianciája is nagyobb legyen mint 70%, a látens változók relatíve magas száma szükséges. Az SNV, MSC, FD+SNV és SD+SNV spektrum transzformációkkal a modellek komplexitása csökken, illetve az RMSE értéke is. A kávé NIR spektrumaival történő modellezés során, a deriválás hatékonyságát több publikáció is igazolja (RIBEIRO et al. 2011; ZHANG et al. 2013; BUDIASTRA et al. 2018). Az adat-előkezelés által a minták fizikai tulajdonságaiból eredő eltérések csökkenthetők és a jel/zaj viszony is javítható, így csökkent azoknak a mintáknak a száma, amelyek torzítják a becslést. A változókiválasztással koffein esetében meghatározott, regresszió szempontjából jelentős elnyelési területek 7444-7220, 6981-6526, 6287-5600 cm^{-1} és 4200-3826 cm^{-1} között voltak. Az optimális modellt ebben az esetben is OSC adat-előkezelés után értem el, melynek eredményeképpen az adatkészlet 3 látens változó alkalmazásával modellezhető. A PLS modell mérési tartománya 9,9-15,43 mg/g, melyen belül az RMSEC/RMSECV értéke 0,30/0,41 mg/g.

25. táblázat: A PLS regresszióval, pörkölt kávék esetén kapott modellek fontosabb statisztikai paramétereit. A koffein, 5-CQA, 4-CQA, 3-CQA, 3,5-diCQA, 4,5-diCQA, 3-FQA eredményeinek mértékegysége: mg/g.

Paraméter	Adat-előkezelés	Változó-kiválasztás	Mintaszám	LV	Kereszt-ellenőrzés				Teszt-validálás				Mérési tartomány
					R ²	Q ²	RMSEC	RMSECV	R _t ²	Q _t ²	RMSEC _t	RMSEP	
pH	-	-	124	7	0,75	0,71	0,07	0,08	0,76	0,68	0,07	0,07	4,72-5,37
	FD+SNV	GA50	137	8	0,89	0,85	0,05	0,06	0,89	0,86	0,05	0,05	4,68-5,39
	OSC	-	137	5	0,93	0,89	0,04	0,05	0,93	0,86	0,04	0,05	4,68-5,39
koffein	-	-	125	10	0,74	0,66	0,50	0,60	0,74	0,65	0,39	0,38	9,90-14,88
	SD+SNV	GA30	125	7	0,82	0,75	0,40	0,48	0,84	0,73	0,36	0,52	9,90-14,6
	OSC	-	135	3	0,90	0,82	0,30	0,41	0,90	0,83	0,31	0,38	9,90-15,43
5-CQA	-	-	132	3	0,81	0,80	1,97	2,04	0,82	0,75	2,01	2,08	5,37-26,74
	SD+SNV	GA50	131	8	0,91	0,88	1,35	1,60	0,92	0,84	1,39	1,47	3,67-26,74
	OSC	-	130	2	0,90	0,89	1,46	1,52	0,90	0,90	1,51	1,29	3,67-24,96
4-CQA	-	-	127	5	0,79	0,77	1,11	1,19	0,80	0,78	1,11	1,14	3,13-14,70
	SD+SNV	GA50	129	8	0,86	0,81	0,89	1,06	0,86	0,81	0,88	1,11	3,43-14,71
	OSC	-	127	4	0,88	0,84	0,80	0,95	0,89	0,83	0,77	0,97	3,43-14,71
3-CQA	-	-	129	5	0,73	0,69	1,02	1,09	0,74	0,67	1,02	1,12	2,40-11,58
	SD+SNV	GA50	127	6	0,80	0,76	0,88	0,99	0,81	0,75	0,87	1,02	2,40-11,30
	OSC	-	127	4	0,87	0,81	0,72	0,89	0,87	0,86	0,70	0,83	2,40-11,58
3,5-diCQA	-	-	125	3	0,72	0,70	0,37	0,38	0,75	0,64	0,36	0,38	0,33-3,45
	SD+SNV	GA30	123	2	0,76	0,74	0,35	0,36	0,77	0,71	0,38	0,32	0,33-3,45
	OSC	-	125	5	0,89	0,81	0,25	0,34	0,90	0,81	0,23	0,39	0,33-3,45
savasság	-	-	121	5	0,71	0,66	0,78	0,85	0,71	0,66	0,80	0,79	1,10-6,80
	MSC	GA30	126	6	0,79	0,76	0,70	0,76	0,79	0,75	0,71	0,77	1,10-7,33
	OSC	-	125	2	0,85	0,81	0,57	0,66	0,87	0,78	0,54	0,70	1,17-7,00
keserűség	-	-	117	3	0,62	0,59	0,95	0,99	0,64	0,58	0,96	0,99	1,3-8,0
	FD+SNV	GA30	124	7	0,81	0,76	0,69	0,79	0,84	0,73	0,66	0,79	1,3-9,3
	OSC	-	130	3	0,89	0,83	0,53	0,68	0,91	0,82	0,48	0,69	1,3-9,3

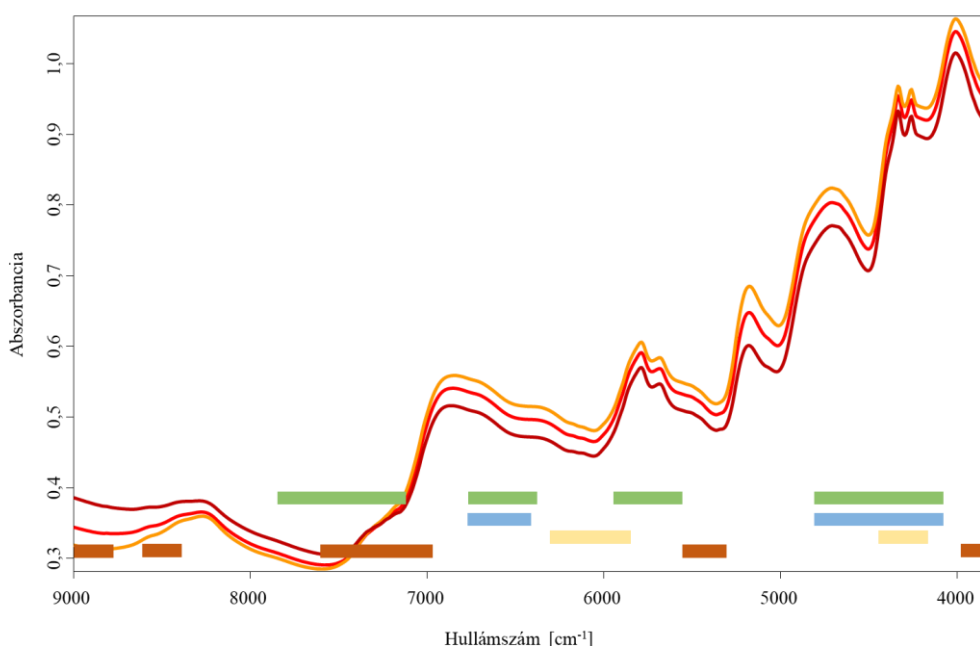
*A legjobb modellek félkövérrel vannak jelölve.

5-CQA: kinasav-5-O-kávészav; 4-CQA: kinasav-4-O-kávészav; 3-CQA: kinasav-3-O-kávészav; 3,5-diCQA: kinasav-3,5-O-dikávészav; 4,5-diCQA: kinasav-4,5-O-dikávészav; 3-FQA: kinasav-3-O-ferulasav; CGA: klorogénsavak

RMSEC: átlagos négyzetes hiba a kalibrációs adatkészletre nézve; RMSECV: átlagos négyzetes hiba a kereszt-ellenőrzés esetén; RMSEP: átlagos négyzetes hiba a teszt adatkészletre nézve; R²: kalibrációs adatkészletre vonatkozó determinációs együttható kereszt-ellenőrzés esetén; R_t²: kalibrációs adatkészletre vonatkozó determinációs együttható teszt-ellenőrzés esetén; Q²: kereszt-ellenőrzésre vonatkozó determinációs együttható; Q_t²: teszt-ellenőrzésre vonatkozó determinációs együttható; RMSEC_t: átlagos négyzetes hiba a kalibrációs adatkészletre nézve teszt-ellenőrzésnél

A különböző klorogénsavak mennyisége fontos paraméter a kávéital érzékszervi minősége, illetve az egészségre gyakorolt hatása miatt is. Ennek ellenére, korábbi tanulmányok nem foglalkoztak a NIRS alkalmazhatóságának vizsgálatával klorogénsavak mennyiségi becslésére pörkölt kávé esetén. Kvalitatív és kvantitatív meghatározásuk alapvetően folyadékkromatográfiai módszerekkel történik. Mivel a pörkölt kávé bizonyos klorogénsavakból legalább 0,1%-ot tartalmaz (tömegre vonatkoztatva), mennyiségük nagy valószínűséggel becsülhető NIR spektroszkópiával. Ezt erősíti meg továbbá a Fábrián és mtsi. (1994) által közölt tudományos publikáció, melyben sikeresen fejlesztettek becslési modellt instant kávék és különböző kávé

helyettesítők keverékeiben található klorogénsav mennyiségi meghatározására. Általánosan elmondható, hogy a vizsgált adat-előkezelések mind alkalmasak a PLSR modellek statisztikai paramétereinek javítására, azonban minden esetben az SD és SNV transzformációk kombinációja vezetett a legkisebb RMSE és legnagyobb R^2 értékekhez. A második derivált spektrum eredményes alkalmazása valószínűleg abban rejlik, hogy részletesebb képet ad a minták molekuláris rezgéseiről. A futtatott változókiválasztási eljárások közül az iPLS alkalmazása nem ajánlott, mivel nem volt hatása a modellek teljesítőképességére. Az egyes vegyületek esetében, a regresszió szempontjából lényeges, a változókiválasztási módszerekkel meghatározott spektrális tartományokat a 34. ábra szemlélteti.



34. ábra: A változókiválasztási eljárások által lényegesnek ítélt NIR elnyelési területek a klorogénsavak vizsgálatánál pörkölt kávé esetén. ● 5-CQA ● 4-CQA ● 3-CQA ● 3,5-diCQA

Az eredmények alapján megállapítható, hogy a NIRS különböző pörkölésű kávék 5-, 4- és 3-CQA, ill. a 3,5-diCQA mennyiségi becslésére alkalmazható. A kávéban található 5-CQA mennyisége és a pörköltégi állapot között fennálló kapcsolat miatt, a vegyület koncentrációja ígéretes eredménnyel becsülhető NIRS technikával. A létrehozott becslési függvény különböző pörkölésű kávék analitikai vizsgálatára alkalmas, ezáltal a mérési tartomány is igen széles 5,37-24,96 mg/g. A modell-építés során, az R^2/Q^2 értéke nem volt kisebb, mint 0,80/0,77.

A 4- és 3-CQA esetén a modellezéshez kiválasztott hullámszám tartományok $6742-6357\text{ cm}^{-1}$ és $4814-4050\text{ cm}^{-1}$, illetve $6279-5824\text{ cm}^{-1}$ és $4428-4204\text{ cm}^{-1}$ között találhatók. Ezek az elnyelési területek az 5-CQA modellezése során is fontos szerepet töltenek be, ami nem meglepő, hiszen

ezek a vegyületek egymás izomerjei, ezáltal molekuláris rezgéseik is hasonlóságokat mutatnak. A kiindulási modelleknél is erős kapcsolat van a spektrális és referencia adatok között, amit a R^2/Q^2 értéke ($>0,64$) is alátámaszt. Az SD+SNV transzformáció és a GA alkalmazásával a modellek teljesítőképessége javítható, azonban viszonylag nagyszámú látens változó felhasználását igényli. A 4-CQA esetén ezek száma 8, míg a 3-CQA esetén 6 volt. Az OSC transzformáció által az R^2 értéke nagymértékben nőtt ($>0,81$), az RMSE értéke pedig mindkét vegyületnél csökkent. Utóbbi értéke a 4-CQA-nál nem haladja meg a 0,97 mg/g-ot, ill. a 0,89 mg/g-ot a 3-CQA esetén (25. táblázat).

Habár a 4,5-diCQA és a 3-FQA koncentrációja is nagyobb mint 0,1%, a pörkölés során bekövetkező koncentráció-változásnak nincs egyértelmű iránya. Ahogy arra már korábban utaltam (5.2.1.2. alfejezet) a változás számos különböző tendenciával írható le. Tehát a vegyületek mennyisége és a pörkölési szint között nincs egyértelmű kapcsolat. Ezt alátámasztja a varianciaelemzés eredménye is, amely alapján a 4,5-diCQA esetén a közepes és sötét, míg a 3-FQA esetén a világos és közepes pörkölésű mintákra mért koncentráció között nincs szignifikáns különbség. Az eredmények alapján feltételezhető, hogy ezen vegyületek esetén a kapcsolat nem vagy csak nagyon nehezen modellezhető a NIR spektrális adatokkal. Ez azzal is alátámasztható, hogy a 4,5-diCQA becslésére létrehozható megfelelő statisztikai jellemzőkkel bíró PLS modell, azonban ehhez számos világos és sötét pörkölésű mintát ki kellett zárni a kiértékelésből. A világos pörkölésnél jellemzően azokról a mintákról van szó, melyeknél a mért koncentráció nagyobb, mint 2,33 mg/g, azaz a mérési tartomány felső kvartilisében található. Ezen kívül azokat a sötét pörkölésű mintákat kezelte a PLSR kiugró értéként, melyeknél a közepes pörköléshez képest, a 4,5-diCQA koncentrációja növekedést mutatott. Az adatkészlet alapján ezeknek a mintáknak a koncentrációját a függvény alulbecsülte. A modellezés azonban csak FD+SNV és SD+SNV adat-előkezelés után volt sikeres. Az optimális statisztikai paraméterekhez SD+SNV alkalmazása után, VIP változókiválasztási módszerrel jutottam, ahol az R^2/Q^2 értéke 0,74/0,69 volt. A függvény mérési tartománya 0,24-2,35 mg/g, míg az RMSE 0,26/0,29 mg/g, amiből jól látszik, hogy az alacsony koncentráció-tartomány vizsgálata problémás lehet. Az eredmények alapján az feltételezhető, hogy hasonló pörköltségű kávéminták esetén a mennyiségi becslés lehetséges. Ezen kívül, érdemes lehet megfontolni a nem-lineáris regressziós módszerek alkalmazását a modellezés során.

A kávé minőségének értékelése során az érzékszervi minősítésnek (cupping) kiemelkedő szerepe van, azonban ez egy viszonylag idő és munkaigényes feladat. Ezáltal igény mutatkozik gyorsabb módszerek alkalmazására, melyek megbízható és reprodukálható eredményt szolgáltatnak. Mivel az érzékszervi tulajdonságok a kávé kémiai összetételétől függenek, a NIRS alkalmas lehet

ezeknek a tulajdonságoknak a becslésére. Kutatásom során ezek közül az érzékelt savasságot és keserőséget vizsgáltam. A kávéban található szerves savak és klorogénsavak fontos szerepet játszanak az érzékelt savasság kialakításában. Ezáltal az ezekkel összefüggésben lévő abszorpciós sávoknak jelentős szerepe van PLS modellezés során, amit az adatok is alátámasztanak. Az érzékelt savasság becslése során jelentős abszorpciós sávok $7676-7444\text{ cm}^{-1}$ (C-H kombinációs rezgések 1. felhangja, CH_3), $5817-5593\text{ cm}^{-1}$ (C-H vegyértékrezgés 1. felhangja, O-H kombinációs rezgések) és az $5122-4899\text{ cm}^{-1}$ (C=O és C-H kombinációs rezgések 1. felhangja) között található. Ezek meghatározása a genetikus algoritmus alkalmazásával történt. A kiindulási modell teljesítménye mind a hagyományos adat-előkezelési és változókiválasztási módszerekkel, mind az OSC-vel javíthatók. Az érzékszervi tulajdonságok intenzitásának értékelése 1-10-ig terjedő skálán történt. Az érzékelt savasság és a NIR spektrális adatok között erős lineáris kapcsolat áll fenn, amit a determinációs együttható értéke is tükröz (0,66-0,85). Az optimális modellt OSC alkalmazásával értem el, melynél a kiindulási PLSR-hez képest az RMSECV értéke 0,79-ről, 0,70-re csökkent. A modell mérési tartománya pedig 1,17-7 között volt. Habár korábbi tanulmányok kisebb RMSECV értékről számolnak be arabica kávék vizsgálata során (RIBEIRO et al. 2011; DA SILVA ARAÚJO et al. 2021), mindkét esetben közepes pörkölésű mintákat vizsgáltak, ezáltal a mérési tartomány is szűkebb volt.

A kávé keserű íze nagymértékben függ a pörkölési foktól és összefüggésbe hozható a koffeinnel, egyes heterociklusos vegyületekkel, peptidekkel, illetve a klorogénsavak és cukrok bomlása során létrejövő termékekkel (RIBEIRO et al. 2011). A PLS modellezés során lényegesnek ítélt spektrális tartományok a $8362-7907\text{ cm}^{-1}$ (C-H vegyértékrezgés 2. felhangja), a $7668-7444\text{ cm}^{-1}$ (C-H vegyérték- és C-H deformációs rezgés kombinációja), a $6974-6750\text{ cm}^{-1}$ (O-H és N-H rezgések 1. felhangja), a $6048-5593\text{ cm}^{-1}$ (C-H és S-H rezgések 1. felhangja), az $5122-4667\text{ cm}^{-1}$ (C=O és C-H kombinációs rezgések 1. felhangja), illetve a $4428-3973\text{ cm}^{-1}$ (C-H+C-H kombinációs rezgések) voltak. A vizsgált minták a keserű íz tekintetében szélesebb skálán mozogtak, mint a savasságnál, ezáltal a PLSR mérési tartománya 1,3-9,3. Ennek az az oka, hogy míg a kávé savasságát befolyásoló vegyületek koncentrációja a pörkölés hatására csökken, addig a keserű ízérzettel kapcsolatban állóké növekszik. Általánosságban elmondható, hogy habár a hagyományos adat-előkezelési és változókiválasztási eljárások többnyire az R^2 növekedését és az RMSE csökkenését eredményezik, a látens változók száma is megnő. Az OSC alkalmazásával ez a probléma kiküszöbölhető. Az optimális modell 0,89/0,83 R^2/Q^2 értékkel rendelkezik, RMSEC/RMSECV=0,53/0,68 átlagos hiba mellett. Ez egyetértésben van a keserűség becslésére más kutatók által létrehozott PLS modellek eredményeivel (ESTEBAN-DÍEZ et al. 2004; RIBEIRO et al. 2011; BAQUETA et al. 2019).

Fontos kiemelni, hogy az általam feldolgozott tudományos szakirodalom alapján az érzékelt savasság és keserűség becslésére eddig nem állt rendelkezésre olyan összefüggés, amely alkalmas lehet eltérő országokból származó és különböző pörkölésű kávék vizsgálatára.

6. Következtetések és a javaslatok

Doktori munkám során egy eddig kevésbé vizsgált élelmiszer-mátrix (snackek) Fourier-transzformációs közeli infravörös spektroszkópiával (FT-NIRS) történő komplex elemzését végeztem el. Ezen kívül olyan makrokomponenseket és vegyületeket vontam be a kutatásba, melyekkel eddig kevesen, vagy egyáltalán nem foglalkoztak, ezáltal bővítve a tudományos szakirodalmat.

Törekedtem arra, hogy a spektrális adatok elemzése során eddig kevésbé használt kemometriai módszereket is alkalmazzak a modellezéshez, bemutatva az ezekben rejlő lehetőségeket is.

A fogyasztásra kész élelmiszerek, mint a snackek a mindennapi életünk szerves részét képezik. A termékminősítés szempontjából lényeges paraméter a snackek általános kémiai összetétele, melynek ismerete fontos a gyártó, ill. a hatóságok szempontjából is. A makrokomponensek (szénhidrát, cukor, zsír, fehérje, só) mennyiségi meghatározása adja a csomagoláson feltüntetett tápértéktáblázat alapját. A gyakorlatban alkalmazott hagyományos analitikai módszerek elvégzése több napot is igénybe vehet, ami megnehezíti a gyors döntéshozatalt. Az eredmények alapján elmondható, hogy az FT-NIR spektroszkópián alapuló becslési függvények alkalmasak a makrokomponensek mennyiségi becslésére, ami lehetővé teszi a snackek átlagos kémiai összetételének gyors és környezetbarát meghatározását.

A snackek energiatartalmának becslése során az intervallum PLS regresszió alkalmas volt a modell komplexitásának csökkentésére, azonban érdemes lehet további változó kiválasztási módszereket is lefuttatni a becslési modell teljesítményparamétereinek javítása érdekében.

A snackeket nagy változatosság jellemzi, a sütőolajok, nyersanyagok, származási helyek és technológiai eljárások kombinációi szintjén. A megfelelő kemometriai modellezéssel párosított NIRS azonban alkalmas a snacktermékek minőségi tulajdonságainak vizsgálatára is. A spektrális adatok lehetőséget adnak arra, hogy a vizsgált termékeket a fent említett jellemzők alapján csoportosítsuk különböző klasszifikációs módszerekkel. Mindhárom vizsgált algoritmus, a PLS-DA, az RF és a neurális hálózaton alapuló RPropMLP alkalmas volt a négy osztályozási rendszer alapján megfelelően osztályozni a snackmintákat. Emellett fontos megemlíteni, hogy jelen kutatás volt az első, amely a snackek NIR spektruma alapján osztályozási rendszerekben vizsgálta a termékeket. További osztályozási módszerek tesztelésével azonban még átfogóbb képet kaphatunk.

A zöld kávé minőségét, ezáltal az árát számos fiziko-kémiai, ezeken keresztül pedig számos érzékszervi tulajdonság, a növény genetikája és a származási hely is nagyban befolyásolja. Emellett a zöld kávé kémiai összetételének ismerete segítheti a profi pörkölőket a pörkölési

folyamat szabályozásában, ezáltal az általuk kívánt érzékszervi profil beállításában. A zöld kávé minősítése szempontjából fontos paraméterek, a savtartalom, a koffeintartalom és a klorogénsavak mennyisége is. Utóbbi kettő meghatározása alapvetően egy viszonylag hosszadalmas minta-előkészítés után, nagyműszeres analitikai módszerekkel történik, melyek költségesek és specifikus ismereteket igényelnek. Az általam létrehozott NIR spektroszkópián alapuló modellek azonban egyszerűen és gyorsan alkalmazhatók közvetlenül a szemeskávé roncsolásmentes vizsgálatára. Az optimális modellek eléréséhez mindenképpen szükséges valamilyen adat-előkezelés, emellett változókiválasztási eljárások alkalmazása is ajánlott a kávé komplex molekuláris rezgéseiből adódóan.

A spektrális adatokat gyakran használják eredetazonosítási célokra is, amellyel a zöld kávé esetében is számos tudományos publikáció foglalkozott. Tapasztalataim alapján, az osztályozás eredményes lehet egy diverz adatkészlet esetén is, azonban a különböző termesztési régiók és a spektrális tulajdonságok között nincs egyértelmű összefüggés, ami megnehezíti a feladatot. A modellek pontossága bináris rendszerek (két régió összehasonlítása) esetén a legmagasabb. A szokásosan használt osztályozási módszerek (LDA, PLS-DA stb.) mellett számos egyéb algoritmus alkalmazását (pl. döntési fák, ensemble módszerek stb.) érdemes lehet megfontolni.

Az általam vizsgált zöld kávék referencia adatai alapján (pH, koffeintartalom, 5-, 4-, 3-CQA, 3,5-diCQA, 4,5-diCQA, 3-FQA) nem találtam összefüggést vagy szignifikáns különbséget a négy termesztési régióra vonatkozóan. Ezek alapján elmondható, hogy a származási helyből adódó különbségek meghatározásához kiterjedtebb adatbázisra van szükség.

A pörkölés hatására a különböző klorogénsavak mennyiségében bekövetkező változásokat még napjainkban sem értjük teljesen, mivel összetett folyamatokról van szó. A vizsgált vegyületek közül az 5-CQA mennyiségében a legtöbb esetben csökkenés tapasztalható a pörkölés előrehaladtával. Ezek a változások párhuzamba állíthatók a NIR spektrális adatokban fellépő karakterisztikus változásokkal is. Kombinálva az információkat, egyértelműbb képet kaphatunk a kávé pörköltégi állapotáról. Ezzel szemben, a többi vizsgált klorogénsav esetén a pörkölés hatása változatos tendenciákon keresztül mutatkozik meg.

A pörkölt kávé koffeintartalmának NIR spektrumain alapuló mennyiségi becslését megnehezíti, hogy míg a koffein koncentrációja a pörkölés hatására nem változik jelentősen, a spektrális adatokban számottevő eltérés tapasztalható a koffeinnel összefüggésbe hozható abszorpciós tartományokban. A PLS regresszióval történő sikeres modellezéshez elengedhetetlen adat-előkezelés a deriválás, nem elegendő a hagyományos szórás-korrekciós műveletek alkalmazása. (MSC, SNV). A legjobb eredményt azonban az ortogonális jelkorrekcióval (OSC) értem el, ezért ennek alkalmazását javaslom.

Az eredmények alapján megállapítható, hogy a NIRS a pörkölt kávéban található 5-, 4- és 3-CQA, ill. a 3,5-diCQA mennyiségi becslésére alkalmazható. A 4,5-diCQA és a 3-FQA pörkölés során bekövetkező koncentráció-változásnak nincs egyértelmű iránya, a változás számos különböző tendenciával írható le. Tehát a vegyületek mennyisége és a pörkölési szint között nincs egyértelmű kapcsolat. Az eredmények alapján feltételezhető, hogy ezen vegyületek esetén a kapcsolat nem, vagy csak nagyon nehezen modellezhető a NIR spektrális adatokkal.

Általánosan elmondható, hogy a zöld és pörkölt kávé esetén a regressziós modell sikerességét nagymértékben meghatározza a kiválasztott adat-előkezelés és a változókiválasztási eljárásoknak is kiemelt szerepe van a cél szempontjából optimális modellek elérése érdekében. A legtöbb vizsgált célparaméter esetében az OSC alkalmazásával értem el a legjobb regressziós összefüggést, azonban ennek sikerességét nagyban befolyásolja a referencia adatok pontossága. Emellett elmondható, hogy az iPLSR változókiválasztási módszer alkalmazása a pörkölt kávé esetén, míg a VIP zöld kávék esetén sem javasolt.

7. Új tudományos eredmények

1. Elsőként foglalkoztam különböző gyártástechnológiával előállított snack termékek FT-NIR spektroszkópiával történő vizsgálatával, ami során feltérképeztem a termékcsoportha jellemző spektrális tulajdonságokat is. Megállapítottam, hogy a FT-NIR spektroszkópia kemometriai módszerekkel kombinálva lehetővé teszi az akár különböző gyártástechnológiával készített snackek azonos modellekkel való értékelését is. Gyorsmódszereket dolgoztam ki a snackek makrokomponenseinek mennyiségi becslésére, melyek alkalmasak a snackek szénhidrát-, cukor-, zsír-, fehérje-, sótartalmának előrejelzésére PLS regresszió alkalmazásával. Emellett elsőként alkalmaztam az iPLS regressziót snack mátrixra, az energiatartalom mennyiségi becslésére.

Eszter, Benes; Attila, Gere; Marietta, Fodor (2020). Predicting macronutrients and energy content of snack products using FT-NIR and analysis chemometric techniques. Journal of Food Engineering. IF érték: 5,354. SJR indikátor D1.

2. Megfelelő teljesítménymutatókkal rendelkező osztályozási modelleket dolgoztam ki a snackek különböző szempontok szerint történő osztályozására (felhasznált sütőolaj, alapanyag, gyártási ország, gyártási technológia), melyekkel elsőként alkalmaztam osztályozási módszereket snack termékek NIR spektrális adatainak vizsgálatára.

Eszter Benes; Dávid Bajusz; Attila Gere; Marietta Fodor; Anita Rác (2020). Comprehensive chemometric classification of snack products based on their near infrared spectra. LWT-FOOD SCIENCE AND TECHNOLOGY. IF: 4.952. SJR indikátor: D1

3. Igazoltam, hogy az FT-NIR spektroszkópia alkalmas mind zöld, mind pörkölt szemes kávé 5-, 4-, 3-CQA, 3,5-diCQA tartalmának mennyiségi becslésére, míg zöld kávé estén a 4,5-diCQA és a 3-FQA mennyiségi becslése is sikeres volt. Bizonyítottam, hogy a vizsgált arabica kávék eredményei alapján, nincs egyértelmű kapcsolat a 4,5-diCQA és a 3-FQA mennyisége, ill. a pörkölési szint között. Míg a 4,5-diCQA és a 3-FQA koncentrációjában a világos és a közepes pörkölésű minták között nem volt szignifikáns különbség, addig a többi vizsgált klorogénsav esetén ez megfigyelhető volt.

4. Bizonyítottam, hogy erős, szignifikáns pozitív korreláció figyelhető meg a kávé klorogénsav tartalma és érzékszervi savassága között, azonban közepes, de negatív kapcsolat az érzékelt keserűség között.

5. Regressziós modellt fejlesztettem olyan zöld arabica szemeskávék koffeintartalmának és pH értékének predikciójára, melyek származási hely tekintetében nagy eltéréseket mutatnak.
6. Bizonyítottam, hogy a PLS regresszió optimalizálása során az ortogonális jelkorrekció (OSC) alkalmazása nagymértékben javítja a becslési függvények teljesítményét. Bizonyítottam, hogy az eredmények értelmében az intervallum PLS regresszió és a VIP változókiválasztás alkalmazása nem befolyásolja jelentősen a modellek teljesítményét a vizsgált paraméterekkel összefüggésben szemeskávé estén.
7. A zöld kávé spektrális adatain eddig nem alkalmazott lineáris és nem lineáris kemometriai módszereket (pl. ensemble-, Naïve Bayes algoritmusok, döntési fák stb.) futtattam le. Megállapítottam, hogy a vizsgált adatkészlet felhasználásával a modellek teljesítőképesége bináris rendszerekben (Afrika-Amerika, Afrika-Ázsia) a legjobb.

8. Összefoglalás

Napjainkra jellemző a felgyorsult életmód, ami gyakran vezet fáradékonysághoz, energia- és időhiányhoz, ezért egyre több fogyasztó választ olyan élelmiszereket, amelyek ezeket a „tüneteket” gyorsan képesek enyhíteni. Ezek közé tartoznak a snackek és a kávé, melyek széles fogyasztói rétegekhez jutnak el a lokális és globális ellátóláncon keresztül. A termékek előállítása és szállítása nagy volumenben történik, ezért olyan módszerekre van szükség, melyek hatékonyan és gyorsan képesek a termékminőség ellenőrzésére a termékpálya különböző szakaszaiban. A közeli infravörös spektroszkópia (NIRS), különböző kemometriai módszerekkel (ezen belül is a többváltozós adatelemzési módszerekkel) párosítva lehetővé teszi az élelmiszerek gyors vizsgálatát, fizikai, kémiai és érzékszervi paraméterek szimultán meghatározásán keresztül, ezzel segítve a döntéshozatalt. Alkalmazása egyszerű, mivel a hagyományos analitikai eljárásokkal ellentétben, a modellek használata nem igényel semmilyen speciális ismeretet.

Doktori munkám fókuszában a NIRS alkalmazhatóságának vizsgálata állt, különböző snackek, ill. zöld és pörkölt kávé vonatkozásában. A két termékcsoporthoz más-más paraméterek (pl. beltartalmi, érzékszervi, fizikai) fontosak a termékminőség szempontjából, melyek közül eddig nem, kevéssé vagy más perspektívából vizsgált paraméterekkel összefüggésben végeztem kutatásaimat. A NIR spektroszkópiát minőségi és mennyiségi jellemzők meghatározására használtam a két termékcsoporthoz.

A snackek esetén olyan gyorsmódszer fejlesztése volt a cél, amely alkalmas lehet a termékek átlagos összetételének mennyiségi becslésére a NIR spektrumok alapján. Ehhez meghatároztam a minták szénhidrát-, cukor-, zsír-, fehérje- és sótartalmát, ill. kiszámoltam a rendelkezésre álló adatok alapján az energiatartalmat is. Az adatkészletre jellemző diverzitás széles mérési tartományt biztosított az mennyiségi adatok tekintetében. A minták darálása után rögzítettem a spektrumokat, majd különböző adat-előkezelési technikák alkalmazása után predikciós modellt fejlesztettem az egyes makrokomponensek mennyiségi becslésére PLS regresszióval. A modellekre kapott determinációs együttható értéke minden esetben nagyobb volt, mint 0,90. A legjobb teljesítménymutatói a zsír- és fehérjetartalom becslésére létrehozott modelleknek voltak. Azáltal, hogy az energiatartalmat a termékben található makrotápanyagok mennyisége határozza meg, a teljes spektrum hordozhat lényeges információkat ezzel összefüggésben. Annak érdekében, hogy a releváns változókat meghatározzam, ezáltal csökkentsem a modell komplexitását, intervallum PLSR-t futtatam le az adatkészleten 10, 20 és 40 szegmens kialakítása mellett.

A modellek statisztikai paramétereinek javítása érdekében adat-előkezelési módszereket is alkalmaztam. A legjobb eredményt SNV vagy MSC adat-előkezelés után, a 20-szegmenses iPLSR futtatásával értem el. Ennek eredményeképpen csökkent a modell komplexitása és a torzítás is,

azonban a főbb statisztikai paraméterek nem javultak. Az átlagos hiba értéke nem csökkent az optimalálás során, ami valószínűleg azzal állhat kapcsolatban, hogy a termékek változatos összetétele miatt nem egyértelműek a spektrális változások és az energiatartalom közötti összefüggések.

Emellett különböző osztályozási modelleket is kidolgoztam a snackekkel összefüggésben, a következő szempontok alapján: felhasznált alapanyag (burgonya-kukorica), sütőolaj (napraforgó-pálma), gyártási ország (Belgium - Cseh Köztársaság - Lengyelország - Egyesült Királyság) és gyártástechnológia (olajban sült-lúgosan főzött-extrudált). Három osztályozási algoritmust futtattam le, melyek az RF, PLS-DA, RPropMLP voltak. A modellek teljesítményének meghatározására és összehasonlítására az egyes osztályozási adathalmazok és algoritmusok esetében a pontosság (ACC), a valódi pozitív arány (TPR), a valódi negatív arány (TNR) és a vevő működési karakterisztika görbe alatti terület (ROC AUC) értékei lettek felhasználva. Mind a három vizsgált algoritmus alkalmas volt a négy osztályozási rendszer alapján megfelelően besorolni a snackmintákat. A legjobb modell pontossága a sütőolajok osztályozásánál 83%, az alapanyagoknál 98%, a származási országnál 90%, míg a gyártási technológiánál 91% volt.

Az ötven zöld kávé (48 db arabica, 2 db robusta) NIR spektrumát a termék eredeti állapotában, mintaelőkészítés nélkül rögzítettem. A minták kiválasztásánál arra törekedtem, hogy a származási hely tekintetében változatos legyen az adatkészlet. Meghatároztam a minták savtartalmát (pH-érték alapján), színét (CIE Lab színtérben), a koffein- és néhány klorogénsav (5-CQA, 4-CQA, 3-CQA, 3,5-diCQA, 4,5-diCQA, 3-FQA) koncentrációját. Utóbbiakhoz folyadékkromatográfiai módszereket alkalmaztam. Elvégeztem a referencia adatok kiértékelését, hogy átfogóbb képet kaphassak a vizsgált mintákról és összefüggéseket keressek az egyes paraméterek között. A termesztési régió alapján négy csoportot alakítottam ki (Afrika, Ázsia, Dél-Amerika, Közép-Amerika), amelyek között a különbségek vizsgálatára ANOVA-t alkalmaztam. Az eredmények alapján nem volt szignifikáns különbség az egyes paraméterek között a termesztési régiókkal összefüggésben. A spektrális és referencia adatok felhasználásával regressziós modelleket hoztam létre különböző adat-előkezelési, ill. változókiválasztási módszerek alkalmazása után, amelyek során az egyes vegyületek szempontjából optimális modell kidolgozása volt a cél. Ezekhez minden esetben ortogonális jelkorrekció (OSC) keresztül jutottam el. Az eredmények értelmében az FT-NIR spektroszkópia alkalmas a zöld arabica szemeskávéban található 5-, 4-, 3-CQA, 3,5- és 4,5-diCQA, ill. 3-FQA koncentrációjának becslésére. A PLS modellek az R^2/Q^2 értékét tekintve 0,86 és 0,99 közötti értékekkel jellemezhetők.

A zöld kávé árát nagymértékben befolyásolja a származási hely, ezáltal a termék minősítésében szerepet játszhat az eredetazonosítás, amelyre alkalmasak lehetnek a különböző spektroszkópiai módszerek. A NIRS alkalmazhatóságát már korábbi tanulmányok igazolták, azonban a további lehetőségek feltárása érdekében a mintákat négy termesztési régióból, azon belül is huszonhárom országból szereztem be. Különböző, jellemzően korábban nem alkalmazott osztályozási módszereket futtattam le az adatkészleteken, a termesztési régiók spektrális adatok alapján történő azonosítására. A legjobb eredményt bináris rendszereknél értem el, melyek esetén az afrikai és amerikai minták megkülönböztetése 88,6%-ban, míg az afrikai és ázsiai mintáké 92%-ban volt sikeres.

Elvégeztem a kávék pörkölését és vizsgáltam az eltérő pörkölési idő hatását a fent említett vegyületekre, a kávébabok pH értékére, színére, az érzékelt savasságra és keserűségre, ill. a spektrális tulajdonságokra vonatkozóan. Három pörkölési szintet alakítottam ki, melyek a világos, közepes és a sötét pörkölés voltak. Ezeket az első reccsenés után eltelt idő alapján definiáltam (40, 60, 90 másodperc). Az eredmények összehasonlítása érdekében leíró statisztikát futtattam a teljes adatkészletre és az egyes pörkölési szintekhez tartozó mintákra is. Utóbbiak esetében ANOVA-t is alkalmaztam az egyes pörkölési szintek közötti esetleges különbségek feltárására. Az adatok átfogó és egyszerűbb értelmezése érdekében PCA-t is használtam. Megállapítást nyert, hogy a koffein, a 4,5-diCQA és a 3-FQA mennyiségét kivéve, minden paraméter tekintetében szignifikáns különbség van a pörkölési szintek között. A koffeinnél ez a vegyület hőstabilitásával magyarázható, a klorogénsavaknál pedig azzal, hogy a pörkölés során bekövetkező változások nem egyértelműek. Korrelációelemzést futattam le a különböző változók közötti kapcsolatok vizsgálata céljából, amely alapján a mono-kávésavak mennyisége között nagyon erős pozitív korreláció figyelhető meg ($r(136)=0,90-0,97$), ami a pörkölés során történő izomerizációval van összefüggésben. Érdeemes kiemelni, hogy viszonylag erős korreláció volt megfigyelhető a kávé CGA tartalma és érzékszervi savassága között ($r(136)=0,64$), azonban közepes, de negatív kapcsolat ($r(136)=-0,58$) az érzékelt keserűség között ($\alpha=0,01$).

A pörkölés jelentősen befolyásolja a kávé fiziko-kémiai tulajdonságait, ezáltal a spektrális adatokat is. A NIR spektrumokon futtatott PCA eredményeképpen kvázi egy pörköltégi sor alakul ki a minták pozíciója alapján. Ezeket az eredményeket az 5-CQA mennyiségével együtt értelmezve hasonló jelenség volt megfigyelhető. A spektrális és referencia adatok felhasználásával becslési függvényeket fejlesztettem és optimalizáltam (adat-előkezelés, változókiválasztás) a mért jellemzők mennyiségének roncsolásmentes meghatározására, zöld kávéhoz hasonló módon. Az FT-NIR spektroszkópia alkalmas a pörkölt arabica szemeskávéban található 5-, 4-, 3-CQA és a

3,5-diCQA mennyiségének becslésére, azonban a vizsgált adatkészlet alapján nem alkalmazható a 4,5-diCQA, ill. a 3-FQA meghatározására. Az eredmények alapján feltételezhető, hogy ezen vegyületek esetén a kapcsolat nem, vagy csak nagyon nehezen modellezhető a NIR spektrális adatokkal. A PLS regresszió optimalizálása során, minden paraméternél az ortogonális jelkorrekció (OSC) alkalmazásával értem el a legjobb összefüggéseket.

Summary

Our fast-paced lifestyles often lead to fatigue, and lack of energy and time, so more and more consumers tend to choose foods that can quickly alleviate these "symptoms". These foods include snacks and coffee, which reach a wide range of consumers through local and global supply chains. Produced and transported in large volumes, these products require methods able to control product quality effectively and quickly at different stages of the product life cycle. Near-infrared spectroscopy (NIRS), coupled with various chemometric methods (including multivariate data analysis methods), allows rapid analysis of food products through the simultaneous determination of physical, chemical, and sensory parameters to aid decision-making. Unlike traditional analytical methods, NIRS is easy to apply and the used chemometric methods do not require any specific knowledge.

My doctoral thesis focuses on investigating the applicability of NIRS for different snacks and green and roasted coffee. For the two product groups, different parameters (*e.g.*, nutritional, sensory, physical) are considered as important for product quality. I completed my research on parameters that have not been studied, have been studied in a limited way, or from a different perspective.

For snacks, my aim was to develop a rapid method able to estimate the average composition of products from NIR spectra quantitatively. For this purpose, the carbohydrate, sugar, fat, protein, and salt contents of the samples were determined, while the energy content was calculated from the available nutritional data. The diversity of the dataset ensured a wide measurement range for quantitative data. The samples were ground, and the spectra were recorded. After using different data processing techniques, prediction models were developed for the quantitative estimation of each macro-component using PLS regression. In all cases, the coefficient of determination obtained for the models was above 0.90. The best performance metrics were achieved by models developed to predict fat and protein content. Since the energy content is determined by the amount of macronutrients in the product, the whole spectrum can carry relevant information in this relation. In order to identify the relevant variables and thus reduce the complexity of the model, interval PLSR was performed on the data set which was subdivided into 10, 20, and 40 intervals.

Data pre-processing methods were also used to improve the statistical parameters of the models. The best results were obtained after SNV or MSC data pre-treatment by running iPLSR with 20 intervals. As a result, the model's complexity and bias were reduced, but the main statistical parameters have not been improved. The value of root mean square error did not decrease during the model optimization, which may be related to the fact that the spectral changes are not clearly related to the energy content due to the different compositions of the products.

Besides, different classification models were developed for snacks based on the following aspects: raw material used (potato – corn); cooking oil (sunflower – palm); country of production (Belgium – Czech Republic – Poland – UK), and production technology (deep-fried – cooked – extruded). The following three classification algorithms were performed on the dataset: RF, PLS-DA, RPropMLP. Accuracy (ACC), true positive rate (TPR), true negative rate (TNR) and receiver operating characteristic area under the curve (ROC AUC) were used to determine and compare the performance of the models for each classification dataset and algorithm. All three algorithms tested were able to classify snack samples correctly based on the four classification schemes. The best model accuracy was 83% for the classification of frying oils, 98% for raw materials, 90% for country of origin, and 91% for production technology.

The NIR spectra of fifty green coffees (48 arabicas, 2 robustas) were recorded without sample preparation. The samples were selected to ensure a diverse data set in terms of origin. The acidity (based on pH-value), colour (CIE Lab colour space), the concentration of caffeine and some chlorogenic acids (5-CQA, 4-CQA, 3-CQA, 3,5-diCQA, 4,5-diCQA, 3-FQA) have also been defined. For the latter, liquid chromatography methods were used. The reference data were evaluated to obtain a more comprehensive picture of the samples tested and to look for correlations between the parameters. Based on origin, four groups have been created (Africa, Asia, South America, and Central America) and ANOVA was used to test the differences between them. There were no significant differences between the parameters in relation to the origin. Using spectral and reference data, regression models were generated after applying different data treatment and variable selection methods, with the aim of developing an optimal model for each compound. In all cases, these were obtained by using orthogonal signal correction (OSC). FT-NIR spectroscopy was found to be suitable for predicting the concentration of 5-, 4-, 3-CQA, 3,5-, and 4,5-diCQA and 3-FQA in green arabica bean coffee. The PLS models have R^2/Q^2 values between 0.86 and 0.99.

The price of green coffee is strongly influenced by its origin. Thus, authenticity testing can play a role in the assessment of the quality of the product, which can be done by various spectroscopic methods. The applicability of NIRS has already been demonstrated in previous studies, but to explore further possibilities, samples were obtained from four regions, including twenty-three countries. Different classification methods, typically not used before, were tested on the datasets to identify regions based on spectral data. The best results were achieved with binary systems, where the discrimination between African and American samples was 88.6% successful, while the discrimination between African and Asian samples was 92% successful.

Roasting of the coffee beans was performed and the effect of different roasting times on the above-mentioned compounds, pH, colour, perceived acidity and bitterness, and spectral properties of the beans were investigated. Three roasting stages were developed, which were light, medium, and dark roasting. These were defined according to the time elapsed after the first crack (40, 60, and 90 s). Descriptive statistics were run on the whole data set and also on the samples belonging to each roasting level. ANOVA was also used to explore possible differences between the individual roasting levels. PCA was also performed to provide a more comprehensive and simpler interpretation of the data. There was a significant difference between roasting levels for all parameters except caffeine, 4,5-diCQA, and 3-FQA. For caffeine, this can be explained by the thermal stability of the compound, and for chlorogenic acids by the fact that the changes during roasting are not clear. Correlation analysis was run to investigate the relationships between the different variables, showing a very strong positive correlation ($r(136)=0.90-0.97$) between the amount of mono-caffeic acids, which are related to isomerization during roasting. It is worth pointing out that a relatively strong correlation was observed between the CGA content of coffee and the sensory acidity ($r(136)=0.64$), but a moderate but negative correlation ($r(136)=-0.58$) was observed between the perceived bitterness ($\alpha=0.01$).

Roasting significantly affects the physicochemical properties of coffee, and thus the spectral data. The PCA performed on the NIR spectra resulted in a quasi-sequence of roasting based on the position of the samples. Interpreting these results together with the amount of 5-CQA, a similar phenomenon was observed. Using the spectral and reference data, predictive models were developed and optimized (data pre-processing, variable selection) for non-destructive quantification of the measured parameters in a similar way to green coffee. FT-NIR spectroscopy is suitable for predicting the concentration of 5-, 4-, 3-CQA, and 3,5-diCQA in roasted arabica bean coffee, but cannot be used to estimate the concentration of 4,5-diCQA and 3-FQA, respectively. Based on the obtained results, the relationship between these components and the NIR spectra cannot (or only hardly can) be modelled. In the optimization of the PLS regression, the best correlations were found using orthogonal signal correction (OSC) for all parameters.

Mellékletek:

M1. Irodalomjegyzék

- 1169/2011 (X. 25.) Európai Parlament és a Tanács rendelete a fogyasztók élelmiszerekkel kapcsolatos tájékoztatásáról. <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX%3A32011R1169>
- 1333/2008 (XII. 16.) Európai Parlament és a Tanács rendelete az élelmiszer-adalékanyagokról. <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=celex%3A32008R1333>
- ADNAN, A., HÖRSTEN, D. von, PAWELZIK, E., MÖRLEIN, A.D. (2017): Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy. *Foods*. 6 (5), p. 38. <https://doi.org/10.3390/foods6050038>
- ADNAN, A., NAUMANN, M., MÖRLEIN, D., PAWELZIK, E. (2020): Reliable Discrimination of Green Coffee Beans Species: A Comparison of UV-Vis-Based Determination of Caffeine and Chlorogenic Acid with Non-Targeted Near-Infrared Spectroscopy. *Foods*. 9 (6), p. 788. <https://doi.org/10.3390/foods9060788>
- AGELET, L.E., HURBURGH, C.R., (2010): A Tutorial on Near Infrared Spectroscopy and Its Calibration. *Critical Reviews in Analytical Chemistry*. 40 (4), pp. 246–260. <https://doi.org/10.1080/10408347.2010.515468>
- ALAMRI, E., ROZAN, M., BAYOMY, H. (2022): A study of chemical Composition, Antioxidants, and volatile compounds in roasted Arabic coffee. *Saudi Journal of Biological Sciences*. 29 (5), pp. 3133–3139. <https://doi.org/10.1016/j.sjbs.2022.03.025>
- ALESSANDRINI, L., ROMANI, S., PINNAVAIA, G., ROSA, M.D. (2008): Near infrared spectroscopy: An analytical tool to predict coffee roasting degree. *Analytica Chimica Acta*. 625 (1), pp. 95–102. <https://doi.org/10.1016/j.aca.2008.07.013>
- ALTRICHTER, M et al. (2006): *Neurális hálózatok*. Budapest: Panem Könyvkiadó Kft. Elérhető: <http://mialmanach.mit.bme.hu/neuralis/index> Lekérdezés időpontja: 2022. 12. 25.
- ARAÚJO, C. da S., MACEDO, L.L., VIMERCATI, W.C., FERREIRA, A., PREZOTTI, L.C., SARAIVA, S.H. (2020): Determination of pH and acidity in green coffee using near-infrared spectroscopy and multivariate regression. *Journal of the Science of Food and Agriculture*. 100 (6), pp. 2488–2493. <https://doi.org/10.1002/jsfa.10270>
- AWWAD, S., ISSA, R., ALNSOUR, L., ALBALS, D., AL-MOMANI, I. (2021): Quantification of Caffeine and Chlorogenic Acid in Green and Roasted Coffee Samples Using HPLC-DAD and Evaluation of the Effect of Degree of Roasting on Their Levels. *Molecules*. 26 (24), p. 7502. <https://doi.org/10.3390/molecules26247502>
- BADMOS, S., LEE, S.-H., KUHNERT, N. (2019): Comparison and quantification of chlorogenic acids for differentiation of green Robusta and Arabica coffee beans. *Food Research International*. 126, p. 108544. <https://doi.org/10.1016/j.foodres.2019.108544>
- BAPTISTA MANUEL, M.N., DA SILVA, A.C., LOPES, G.S., DIAS RIBEIRO, L.P. (2022): One-class classification of special agroforestry Brazilian coffee using NIR spectrometry and chemometric tools. *Food Chemistry*. 366, p. 130480. <https://doi.org/10.1016/j.foodchem.2021.130480>
- BAQUETA, M.R., CAPORASO, N., COQUEIRO, A., VALDERRAMA, P. (2020): A review of coffee quality assessment based on sensory evaluation and advanced analytical techniques. In: *Innovations in Coffee Quality*. Nova Science Publishers, Inc.
- BAQUETA, M.R., COQUEIRO, A., MARÇO, P.H., VALDERRAMA, P. (2021): Multivariate classification for the direct determination of cup profile in coffee blends via handheld near-infrared spectroscopy. *Talanta*. 222, p. 121526. <https://doi.org/10.1016/j.talanta.2020.121526>
- BAQUETA, M.R., COQUEIRO, A., VALDERRAMA, P. (2019): Brazilian Coffee Blends: A Simple and Fast Method by Near-Infrared Spectroscopy for the Determination of the Sensory Attributes Elicited in Professional Coffee Cupping. *Journal of Food Science*. 84 (6), pp. 1247–1255. <https://doi.org/10.1111/1750-3841.14617>

- BARBIN, D.F., FELICIO, A.L. de S.M., SUN, D.-W., NIXDORF, S.L., HIROOKA, E.Y. (2014): Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*. 61, pp. 23–32. <https://doi.org/10.1016/j.foodres.2014.01.005>
- BARBOSA, M. de S.G., SCHOLZ, M.B. dos S., KITZBERGER, C.S.G., BENASSI, M. de T. (2019): Correlation between the composition of green Arabica coffee beans and the sensory quality of coffee brews. *Food Chemistry*. 292, pp. 275–280. <https://doi.org/10.1016/j.foodchem.2019.04.072>
- BEĆ, K.B., GRABSKA, J., and HUCK, C.W. (2022): Miniaturized NIR Spectroscopy in Food Analysis and Quality Control: Promises, Challenges, and Perspectives. *Foods*. 11 (10), p. 1465. <https://doi.org/10.3390/foods11101465>
- BENES, E., FODOR, M., KOVÁCS, S., GERE, A. (2020): Application of Detrended Fluctuation Analysis and Yield Stability Index to Evaluate Near Infrared Spectra of Green and Roasted Coffee Samples. *Processes*. 8 (8), p. 913. <https://doi.org/10.3390/pr8080913>
- BERTONE, E., VENTURELLO, A., GIRAUDO, A., PELLEGRINO, G., GEOBALDO, F. (2016): Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground coffee. *Food Control*. 59, pp. 683–689. <https://doi.org/10.1016/j.foodcont.2015.06.055>
- BHATTACHARYA, S. (2023): *Snack foods Processing and Technology*. London, United Kingdom: Academic Press, an imprint of Elsevier.
- BLANCO, M., VILLARROYA, I. (2002): NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry*. 21 (4), pp. 240–250. [https://doi.org/10.1016/S0165-9936\(02\)00404-1](https://doi.org/10.1016/S0165-9936(02)00404-1)
- BOBKOVÁ, A., JAKABOVÁ, S., BELEJ, Ľ., JURČAGA, L., ČAPLA, J., BOBKO, M., DEMIANOVÁ, A. (2021): Analysis of caffeine and chlorogenic acids content regarding the preparation method of coffee beverage. *International Journal of Food Engineering*. 17 (5), pp. 403–410. <https://doi.org/10.1515/ijfe-2020-0143>
- BONA, E., MARQUETTI, I., LINK, J.V., MAKIMORI, G.Y.F., DA COSTA ARCA, V., GUIMARÃES LEMES, A.L., FERREIRA, J.M.G., DOS SANTOS SCHOLZ, M.B., VALDERRAMA, P., POPPI, R.J. (2017): Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. *LWT - Food Science and Technology*. 76, pp. 330–336. <https://doi.org/10.1016/j.lwt.2016.04.048>
- BREIMAN, L. (2001): Random Forests. *Machine Learning*. 45 (1), pp. 5–32.
- BROLLO, G., CAPPUCCI, R., and NAVARINI, L. (2009): Acidity in coffee: bridging the gap between chemistry and psychophysics. *22nd International Conference on Coffee Science, ASIC 2008, Campinas, SP, Brazil, 14-19 September, 2008*. pp. 270–280.
- BROWN, S. et al. (2020): *Comprehensive chemometrics. Chemical and biochemical data analysis*. Elsevier.
- BUDIASTRA, I.W., SUTRISNO, WIDYOTOMO, S., AYU, P.C. (2018): Prediction of Caffeine Content in Java Preanger Coffee Beans by NIR Spectroscopy Using PLS and MLR Method. *IOP Conference Series: Earth and Environmental Science*. 147 (1), p. 012004. <http://dx.doi.org/10.1088/1755-1315/147/1/012004>
- BURATTI, S., SINELLI, N., BERTONE, E., VENTURELLO, A., CASIRAGHI, E., GEOBALDO, F., (2015): Discrimination between washed Arabica, natural Arabica and Robusta coffees by using near infrared spectroscopy, electronic nose and electronic tongue analysis: NIR and e-devices for coffee classification. *Journal of the Science of Food and Agriculture*. 95 (11), pp. 2192–2200. <https://doi.org/10.1002/jsfa.6933>
- CAPORASO, N., WHITWORTH, M.B., CUI, C., and FISK, I.D. (2018): Variability of single bean coffee volatile compounds of Arabica and robusta roasted coffees analysed by SPME-GC-MS. *Food Research International*. 108, pp. 628–640. <https://doi.org/10.1016/j.foodres.2018.03.077>

- DE CARVALHO COUTO, C., FREITAS-SILVA, O., MORAIS OLIVEIRA, E.M., SOUSA, C., CASAL, S. (2022): Near-Infrared Spectroscopy Applied to the Detection of Multiple Adulterants in Roasted and Ground Arabica Coffee. *Foods*. 11 (1), p. 61. <https://doi.org/10.3390/foods11010061>
- CARVALHO, D.G., RANZAN, L., JACQUES, R.A., TRIERWEILER, L.F., TRIERWEILER, J.O. (2021): Analysis of total phenolic compounds and caffeine in teas using variable selection approach with two-dimensional fluorescence and infrared spectroscopy. *Microchemical Journal*. 169, p. 106570. <https://doi.org/10.1016/j.microc.2021.106570>
- CATELANI, T.A., SANTOS, J.R., PÁSCOA, R.N.M.J., PEZZA, L., PEZZA, H.R., LOPES, J.A. (2018): Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. *Talanta*. 179, pp. 292–299. <https://doi.org/10.1016/j.talanta.2017.11.010>
- CHENG, B., FURTADO, A., SMYTH, H.E., HENRY, R.J. (2016): Influence of genotype and environment on coffee quality. *Trends in Food Science & Technology*. 57, pp. 20–30. <https://doi.org/10.1016/j.tifs.2016.09.003>
- CHU, Y.-F. (2012): Coffee: emerging health effects and disease prevention. Ames, Iowa: Wiley-Blackwell/IFT Press.
- CORREIA, R.M., TOSATO, F., DOMINGOS, E., RODRIGUES, R.R.T., AQUINO, L.F.M., FILGUEIRAS, P.R., LACERDA, V., ROMÃO, W. (2018): Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta*. 176, pp. 59–68. <https://doi.org/10.1016/j.talanta.2017.08.009>
- COZZOLINO, D. (2020): The Sample, the Spectra and the Maths—The Critical Pillars in the Development of Robust and Sound Applications of Vibrational Spectroscopy. *Molecules*. 25 (16), p. 3674. <https://doi.org/10.3390/molecules25163674>
- COZZOLINO, D., CYNKAR, W.U., SHAH, N., SMITH, P. (2011): Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food Research International*. 44 (7), pp. 1888–1896. <https://doi.org/10.1016/j.foodres.2011.01.041>
- CRAIG, A.P., FRANCA, A.S., OLIVEIRA, L.S., IRUDAYARAJ, J., ILELEJI, K. (2015): Fourier transform infrared spectroscopy and near infrared spectroscopy for the quantification of defects in roasted coffees. *Talanta*. 134, pp. 379–386. <https://doi.org/10.1016/j.talanta.2014.11.038>
- CWIKOVÁ, O., KOMPRDA, T., ŠOTTNÍKOVÁ, V., SVOBODA, Z., SIMONOVÁ, J., SLOVÁČEK, J., JŮZL, M. (2022): Effects of Different Processing Methods of Coffee Arabica on Colour, Acrylamide, Caffeine, Chlorogenic Acid, and Polyphenol Content. *Foods*. 11 (20), p. 3295.
- DAS, S., 2021. Review Post-harvest processing of coffee: An overview. *Coffee Science - ISSN 1984-3909*. 16, pp. e161976–e161976. <https://doi.org/10.3390/foods11203295>
- DAVIES, A.M.C. (2007): Back to basics: spectral pre-treatments, derivatives. [online]. Hozzáférés: <https://www.spectroscopyeurope.com/td-column/back-basics-spectral-pre-treatments-derivatives>. Lekérdezés időpontja: 2023. 03. 20.
- DE LUCA, S., CIOTOLI, E., BIANCOLILLO, A., BUCCI, R., MAGRÌ, A.D., MARINI, F. (2018): Simultaneous quantification of caffeine and chlorogenic acid in coffee green beans and varietal classification of the samples by HPLC-DAD coupled with chemometrics. *Environmental Science and Pollution Research*. 25 (29), pp. 28748–28759. <https://doi.org/10.1007/s11356-018-1379-6>
- DEVOS, O., RUCKEBUSCH, C., DURAND, A., DUPONCHEL, L., HUVENNE, J.-P. (2009): Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation. *Chemometrics and Intelligent Laboratory Systems*. 96 (1), pp. 27–33. <https://doi.org/10.1016/j.chemolab.2008.11.005>
- DOWNEY, G. (2016): Advances in food authenticity testing. Woodhead Publishing.
- DUARTE, P., RAPOSO, M., FERRAZ, M. (2013): Drivers of snack foods impulse buying behaviour among young consumers. *British Food Journal*. 115 (9), pp. 1233–1254. <https://doi.org/10.1108/BFJ-10-2011-0272>

- ENGELSEN, S.B. (2016): Near Infrared Spectroscopy—A Unique Window of Opportunities. *NIR news*. 27 (5), pp. 14–17. <https://doi.org/10.1255/nirn.1620>
- ERICKSON, J. and SLAVIN, J. (2015): Total, Added, and Free Sugars: Are Restrictive Guidelines Science-Based or Achievable? *Nutrients*. 7 (4), pp. 2866–2878. <https://doi.org/10.3390/nu7042866>
- ESTEBAN-DÍEZ, I., GONZÁLEZ-SÁIZ, J.M., PIZARRO, C. (2004): Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. *Analytica Chimica Acta*. 525 (2), pp. 171–182. <https://doi.org/10.1016/j.aca.2004.08.057>
- FARAH, A. (2019): Coffee: Production, Quality and Chemistry. Royal Society of Chemistry.
- FARAH, A., DONANGELO, C.M. (2006): Phenolic compounds in coffee. *Brazilian Journal of Plant Physiology*. 18, pp. 23–36. <https://doi.org/10.1590/S1677-04202006000100003>
- FARAH, A., MONTEIRO, M.C., CALADO, V., FRANCA, A.S., TRUGO, L.C. (2006): Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chemistry*. 98 (2), pp. 373–380. <https://doi.org/10.1016/j.foodchem.2005.07.032>
- FARAH, A., DE PAULIS, T., MOREIRA, D.P., TRUGO, L.C., MARTIN, P.R. (2006): Chlorogenic Acids and Lactones in Regular and Water-Decaffeinated Arabica Coffees. *Journal of Agricultural and Food Chemistry*. 54 (2), pp. 374–381. <https://doi.org/10.1021/jf0518305>
- FARAH, A., DE PAULIS, T., TRUGO, L.C., MARTIN, P.R. (2005): Effect of Roasting on the Formation of Chlorogenic Acid Lactones in Coffee. *Journal of Agricultural and Food Chemistry*. 53 (5), pp. 1505–1513. <https://doi.org/10.1021/jf048701t>
- FÁBIÁN, Z., IZVEKOV, V., SALGÓ, A., és ÖRSI, F. (1994): Near-infrared reflectance and Fourier transform infrared analysis of instant coffee mixtures. *Anal. Proc.* 31 (9), pp. 261–263.
- FELLOWS, P.J. (2016): Extrusion cooking. *Food Processing Technology*. Woodhead Publishing. pp. 753–780.
- FOLMER, B. (2017): The craft and science of coffee. Academic Press.
- FUJIOKA, K., SHIBAMOTO, T. (2008): Chlorogenic acid and caffeine contents in various commercial brewed coffees. *Food Chemistry*. 106 (1), pp. 217–221. <https://doi.org/10.1016/j.foodchem.2007.05.091>
- GAO, C., TELLO, E., PETERSON, D.G. (2023): Identification of compounds that enhance bitterness of coffee brew. *Food Chemistry*. p. 135674. <https://doi.org/10.1016/j.foodchem.2023.135674>
- GERE, A., DANNER, L., DE ANTONI, N., KOVÁCS, S., DÜRRSCHMID, K., és SIPOS, L. (2016): Visual attention accompanying food decision process: An alternative approach to choose the best models. *Food Quality and Preference*. 51, pp. 1–7. <https://doi.org/10.1016/j.foodqual.2016.01.009>
- GEROMEL, C., FERREIRA, L.P., GUERREIRO, S.M.C., CAVALARI, A.A., POT, D., PEREIRA, L.F.P., LEROY, T., VIEIRA, L.G.E., MAZZAFERA, P., MARRACCINI, P. (2006): Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *Journal of Experimental Botany*. 57 (12), pp. 3243–3258. <https://doi.org/10.1093/jxb/erl084>
- GESSELL, A., SMALL, G.W. (2022): Longitudinal Study Comparing Orthogonal Signal Correction Algorithms Coupled with Partial Least-Squares for Quantitative Near-Infrared Spectroscopy. *Analytical Letters*. 55 (3), pp. 449–466. <https://doi.org/10.1080/00032719.2021.1939362>
- GIRAUDO, A., GRASSI, S., SAVORANI, F., GAVOCI, G., CASIRAGHI, E., GEOBALDO, F. (2019): Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control*. 99, pp. 137–145. <https://doi.org/10.1016/j.foodcont.2018.12.033>
- GIRMA, B. (2020): Review on coffee quality markers. *Academic Research Journal of Agricultural Science and Research*. 8 (4), pp. 378–389. <http://dx.doi.org/10.14662/ARJASR2020.019>
- GONZÁLEZ-CABALLERO, V., PÉREZ-MARÍN, D., LÓPEZ, M.-I., és SÁNCHEZ, M.-T. (2011). Optimization of NIR Spectral Data Management for Quality Control of Grape Bunches during On-Vine Ripening. *Sensors*. 11 (6), pp. 6109–6124.

- HAILE, M., KANG, W.H. (2019): The Harvest and Post-Harvest Management Practices' Impact on Coffee Quality. In: *Coffee - Production and Research*. IntechOpen.
<http://dx.doi.org/10.5772/intechopen.89224>
- HAJJEJ, F., ALOHALI, M.A., BADR, M., RAHMAN, M.A. (2022): A Comparison of Decision Tree Algorithms in the Assessment of Biomedical Data. *BioMed Research International*. 2022, pp. 1–9. <https://doi.org/10.1155/2022/9449497>
- HALL, R.D., TREVISAN, F., DE VOS, R.C.H. (2022): Coffee berry and green bean chemistry – Opportunities for improving cup quality and crop circularity. *Food Research International*. 151, p. 110825. <https://doi.org/10.1016/j.foodres.2021.110825>
- HAN, S., LI, M., REN, Q. (2019): Discriminating among tectonic settings of spinel based on multiple machine learning algorithms. *Big Earth Data*. 3 (1), pp. 67–82.
<https://doi.org/10.1080/20964471.2019.1586074>
- HÉBERGER, K., KOLLÁR-HUNEK, K. (2019): Comparison of validation variants by sum of ranking differences and ANOVA. *Journal of Chemometrics*. 33 (6), p. e3104.
<https://doi.org/10.1002/cem.3104>
- HORVAI GY. (Szerk.) (2001): Sokváltozós adatelemzés (kemometria). Budapest: Nemzeti Tankönyvkiadó, 356 p.
- HUCK, C.W., GUGGENBICHLER, W., BONN, G.K. (2005): Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. *Analytica Chimica Acta*. 538 (1–2), pp. 195–203. <https://doi.org/10.1016/j.aca.2005.01.064>
- JAISWAL, R., PATRAS, M.A., ERAVUCHIRA, P.J., KUHNERT, N. (2010): Profile and Characterization of the Chlorogenic Acids in Green Robusta Coffee Beans by LC-MSⁿ: Identification of Seven New Classes of Compounds. *Journal of Agricultural and Food Chemistry*. 58 (15), pp. 8722–8737.
<https://doi.org/10.1021/jf1014457>
- JERRY, W.Jr., WEYER, L. (2012): Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy, CRC Press, 326 p.
- JESZKA-SKOWRON, M., FRANKOWSKI, R., ZGOŁA-GRZEŚKOWIAK, A. (2020): Comparison of methylxantines, trigonelline, nicotinic acid and nicotinamide contents in brews of green and processed Arabica and Robusta coffee beans – Influence of steaming, decaffeination and roasting processes on coffee beans. *LWT*. 125, p. 109344. <https://doi.org/10.1016/j.lwt.2020.109344>
- JIE, D., XIE, L., FU, X., RAO, X., YING, Y. (2013): Variable selection for partial least squares analysis of soluble solids content in watermelon using near-infrared diffuse transmission technique. *Journal of Food Engineering*. 118 (4), pp. 387–392. <https://doi.org/10.1016/j.jfoodeng.2013.04.027>
- KADENKIN, A. (2018): *Near-Infrared Spectroscopy: Quantitative analysis according to ASTM E1655*. Metrohm AG.
 Hozzáférés: https://www.metrohm.com/hu_hu/applications/whitepaper/wp-029.html.
 Lekérdezés dátuma: 2023. 03. 20.
- KAFFKA, K.J. és SEREGÉLY, Z. (2005): PQS (polar qualification system) the new data reduction and product qualification method. *Acta Alimentaria*. 31 (1), pp. 3–20.
<https://doi.org/10.1556/aalim.31.2002.1.2>
- KENNARD, R.W., STONE, L.A. (1969): Computer Aided Design of Experiments. *Technometrics*. 11 (1), pp. 137–148. <https://doi.org/10.2307/1266770>
- KOTSIANTIS, S.B. (2007): Supervised Machine Learning: A Review of Classification Techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*. IOS Press. pp. 3–24.
- KUMAR, S. (2013): Fourier transform-near infrared reflectance spectroscopy calibration development for screening of oil content of intact safflower seeds. *International Food Research Journal*. 20, pp. 759–762.

- KY, C.-L., LOUARN, J., DUSSERT, S., GUYOT, B., HAMON, S., NOIROT, M. (2001): Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. *Food Chemistry*. 75 (2), pp. 223–230.
[https://doi.org/10.1016/S0308-8146\(01\)00204-7](https://doi.org/10.1016/S0308-8146(01)00204-7)
- LANG, T., LANG, R., DI PIZIO, A., MITTERMEIER, V.K., SCHLAGBAUER, V., HOFMANN, T., BEHRENS, M. (2020): Numerous Compounds Orchestrate Coffee's Bitterness. *Journal of Agricultural and Food Chemistry*. 68 (24), pp. 6692–6700.
<https://doi.org/10.1021/acs.jafc.0c01373>
- LAUB-EKGREEN, M.H., MARTINEZ-LOPEZ, B., JESSEN, F. SKOV, T. (2018): Non-destructive measurement of salt using NIR spectroscopy in the herring marinating process. *LWT*. 97, pp. 610–616. <https://doi.org/10.1016/j.lwt.2018.07.024>
- LAXALDE, J., RUCKEBUSCH, C., DEVOS, O., CAILLOL, N., WAHL, F., DUPONCHEL, L. (2011): Characterisation of heavy oils using near-infrared spectroscopy: Optimisation of pre-processing methods and variable selection. *Analytica Chimica Acta*. 705 (1), pp. 227–234.
<https://doi.org/10.1016/j.aca.2011.05.048>
- LEVATE MACEDO, L., DA SILVA ARAÚJO, C., COSTA VIMERCATI, W., GHERARDI HEIN, P.R., PIMENTA, C.J., HENRIQUES SARAIVA, S. (2021): Evaluation of chemical properties of intact green coffee beans using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*. 101 (8), pp. 3500–3507. <https://doi.org/10.1002/jsfa.10981>
- LI, C.-Q., FANG, Z., XU, Q.-S. (2020): A partition-based variable selection in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*. 198, p. 103935.
<https://doi.org/10.1016/j.chemolab.2020.103935>
- LIANG, N. and KITTS, D.D. (2016): Role of Chlorogenic Acids in Controlling Oxidative and Inflammatory Stress Conditions. *Nutrients*. 8 (1), p. 16. <https://doi.org/10.3390/nu8010016>
- LUSAS, E.W., ROONEY, L.W. (2001): *Snack foods processing*. Lancaster, Pa: Technomic Pub. Co.
- MACHEINER, L., SCHMIDT, A., MAYER, H.K. (2021): A novel basis for monitoring the coffee roasting process: Isomerization reactions of 3-caffeoylquinic and 4-caffeoylquinic acids. *LWT*. 152, p. 112343. <https://doi.org/10.1016/j.lwt.2021.112343>
- MACHEINER, L., SCHMIDT, A., SCHREINER, M., MAYER, H.K. (2019): Green coffee infusion as a source of caffeine and chlorogenic acid. *Journal of Food Composition and Analysis*. 84, p. 103307.
<https://doi.org/10.1016/j.jfca.2019.103307>
- MACKENBACH, J.D., HOBBS, M., PINHO, M.GM. (2022): Where do Dutch adults obtain their snack foods? Cross-sectional exploration of individuals' interactions with the food environment. *Health & Place*. 75, p. 102802. <https://doi.org/10.1016/j.healthplace.2022.102802>
- MALTA, M.R., FASSIO, L. de O., LISKA, G.R., CARVALHO, G.R., PEREIRA, A.A., BOTELHO, C.E., FERRAZ, V.P., SILVA, A.D., PEDROSA, A.W., ALVARO, L.N., PEREIRA, R.G.F.A. (2020): Discrimination of genotypes coffee by chemical composition of the beans: Potential markers in natural coffees. *Food Research International*. 134, p. 109219.
<https://doi.org/10.1016/j.foodres.2020.109219>
- MARQUETTI, I., LINK, J.V., LEMES, A.L.G., SCHOLZ, M.B. dos S., VALDERRAMA, P., BONA, E., (2016): Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Computers and Electronics in Agriculture*. 121, pp. 313–319. <https://doi.org/10.1016/j.compag.2015.12.018>
- MAZUREK, S., SZOSTAK, R., KITA, A. (2016): Application of infrared reflection and Raman spectroscopy for quantitative determination of fat in potato chips. *Journal of Molecular Structure*. 1126, pp. 213–218. <https://doi.org/10.1016/j.molstruc.2016.01.064>
- MAZUREK, S., SZOSTAK, R., KITA, A., KUCHARSKA, A.Z., SOKÓŁ-ŁĘTOWSKA, A., HAMOUZ, K. (2017): Determination of Antioxidant Activity and Polyphenols Content in Chips by Raman and IR Spectroscopy. *Food Analytical Methods*. 10 (12), pp. 3964–3971.
<https://doi.org/10.1007/s12161-017-0972-5>

- MEDINA, J., CARO RODRÍGUEZ, D., ARANA, V.A., BERNAL, A., ESSEIVA, P., WIST, J. (2017): Comparison of Attenuated Total Reflectance Mid-Infrared, Near Infrared, and ¹H-Nuclear Magnetic Resonance Spectroscopies for the Determination of Coffee's Geographical Origin. *International Journal of Analytical Chemistry*. 2017, p. e7210463. <https://doi.org/10.1155/2017/7210463>
- METZ, M., ABDELGHAFOR, F., ROGER, J.-M., LESNOFF, M. (2021): A novel robust PLS regression method inspired from boosting principles: RoBoost-PLSR. *Analytica Chimica Acta*. 1179, p. 338823. <https://doi.org/10.1016/j.aca.2021.338823>
- MOON, J.-K., YOO, H.S., SHIBAMOTO, T. (2009): Role of Roasting Conditions in the Level of Chlorogenic Acid Content in Coffee Beans: Correlation with Coffee Acidity. *Journal of Agricultural and Food Chemistry*. 57 (12), pp. 5365–5369. <https://doi.org/10.1021/jf900012b>
- MUTZ, Y.S., DO ROSARIO, D., GALVAN, D., SCHWAN, R.F., BERNARDES, P.C., CONTE-JUNIOR, C.A. (2023): Feasibility of NIR spectroscopy coupled with chemometrics for classification of Brazilian specialty coffee. *Food Control*. p. 109696. <https://doi.org/10.1016/j.foodcont.2023.109696>
- MSZ 20501-1:2007. Sütőipari termékek vizsgálati módszerei. 1. rész: Kémiai vizsgálatok.
- MSZ 6369-12:1979. Lisztvizsgálati módszerek. Nyersrost-, keményítő-, dextrin-, összes és oldható szénhidrát- és összes cukortartalom meghatározása.
- NADKARNI, P. (2016): Core Technologies: Data Mining and “Big Data”. In: *Clinical Research Computing*. Academic Press. <http://dx.doi.org/10.1016/B978-0-12-803130-8.00010-5>
- NÆS, T., ISAKSSON, T., FEARN, T., DAVIES, T. (2017): A user-friendly guide to Multivariate Calibration and Classification. IM Publications Open.
- NAGY, M.M., WANG, S., FARAG, M.A. (2022): Quality analysis and authentication of nutraceuticals using near IR (NIR) spectroscopy: A comprehensive review of novel trends and applications. *Trends in Food Science & Technology*. 123, pp. 290–309. <https://doi.org/10.1016/j.tifs.2022.03.005>
- NALLAN CHAKRAVARTULA, S.S., MOSCETTI, R., BEDINI, G., NARDELLA, M., MASSANTINI, R. (2022): Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. *Food Control*. 135, p. 108816. <https://doi.org/10.1016/j.foodcont.2022.108816>
- NI, Y., MEI, M., KOKOT, S. (2011): Analysis of complex, processed substances with the use of NIR spectroscopy and chemometrics: Classification and prediction of properties - The potato crisps example. *Chemometrics and Intelligent Laboratory Systems*. 105 (2), pp. 147–156. <https://doi.org/10.1016/j.chemolab.2010.12.003>
- NIKINMAA, M., MUSTONEN, S.A., HUITULA, L., LAAKSONEN, O., LINDERBORG, K.M., NORDLUND, E., SOZER, N. (2023): Wholegrain oat quality indicators for production of extruded snacks. *LWT*. 174, p. 114457. <https://doi.org/10.1016/j.lwt.2023.114457>
- NOGALES-BUENO, J., BACA-BOCANEGRA, B., ROMERO-MOLINA, L., MARTÍNEZ-LÓPEZ, A., RATO, A.E., HEREDIA, F.J., HERNÁNDEZ-HIERRO, J.M., ESCUDERO-GILETE, M.L., GONZÁLEZ-MIRET, M.L. (2020): Control of the extractable content of bioactive compounds in coffee beans by near infrared hyperspectral imaging. *LWT*. 134, p. 110201. <https://doi.org/10.1016/j.lwt.2020.110201>
- NØRGAARD, L., SAUDLAND, A., WAGNER, J., NIELSEN, J.P., MUNCK, L., ENGELSEN, S.B. (2000): Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy*. 54 (3), pp. 413–419. <https://doi.org/10.1366/0003702001949500>
- OKUBO, N., KURATA, Y. (2019): Nondestructive Classification Analysis of Green Coffee Beans by Using Near-Infrared Spectroscopy. *Foods*. 8 (2), p. 82. <https://doi.org/10.3390/foods8020082>
- OUYANG, Q., LIU, Y., CHEN, Q., ZHANG, Z., ZHAO, J., GUO, Z., GU, H. (2017): Intelligent evaluation of color sensory quality of black tea by visible-near infrared spectroscopy technology: A

- comparison of spectra and color data information. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 180, pp. 91–96. <https://doi.org/10.1016/j.saa.2017.03.009>
- OZAKI, Y. et al. (2021): Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications. Springer Singapore, 593 p.
- OZAKI, Y., MCCLURE, W.F., és CHRISTY, A.A. (2006): Near-Infrared Spectroscopy in Food Science and Technology: Ozaki/Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc, 408 p. <http://doi.wiley.com/10.1002/0470047704>
- PASQUINI, C. (2003): Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*. 14 (2), pp. 198–219. <http://dx.doi.org/10.1590/S0103-50532003000200006>
- PASQUINI, C. (2018): Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*. 1026, pp. 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>
- PEDRESCHI, F. et al. (2018): Potato Crisps and Snack Foods. In: *Reference Module in Food Science*. Elsevier. <http://dx.doi.org/10.1016/B978-0-08-100596-5.21137-2>
- PERRONE, D., DONANGELO, C.M., FARAH, A. (2008): Fast simultaneous analysis of caffeine, trigonelline, nicotinic acid and sucrose in coffee by liquid chromatography–mass spectrometry. *Food Chemistry*. 110 (4), pp. 1030–1035. <https://doi.org/10.1016/j.foodchem.2008.03.012>
- PHAM, B.T., JAAFARI, A., PHONG, T.V., MAFI-GHOLAMI, D., AMIRI, M., VAN TAO, N., DUONG, V.-H., PRAKASH, I. (2021): Naïve Bayes ensemble models for groundwater potential mapping. *Ecological Informatics*. 64, p. 101389. <https://doi.org/10.1016/j.ecoinf.2021.101389>
- PHILLIPS, S.M., BANDINI, L.G., NAUMOVA, E.N., CYR, H., COLCLOUGH, S., DIETZ, W.H., MUST, A. (2004): Energy-Dense Snack Food Intake in Adolescence: Longitudinal Relationship to Weight and Fatness. *Obesity Research*. 12 (3), pp. 461–472. <https://doi.org/10.1038/oby.2004.52>
- PIRES, F. de C., PEREIRA, R.G.F.A., BAQUETA, M.R., VALDERRAMA, P., ALVES DA ROCHA, R. (2021): Near-infrared spectroscopy and multivariate calibration as an alternative to the Agtron to predict roasting degrees in coffee beans and ground coffees. *Food Chemistry*. 365, p. 130471. <https://doi.org/10.1016/j.foodchem.2021.130471>
- PIZARRO, C., ESTEBAN-DÍEZ, I., GONZÁLEZ-SÁIZ, J.-M., FORINA, M. (2007): Use of Near-Infrared Spectroscopy and Feature Selection Techniques for Predicting the Caffeine Content and Roasting Color in Roasted Coffees. *Journal of Agricultural and Food Chemistry*. 55 (18), pp. 7477–7488. <https://doi.org/10.1021/jf071139x>
- PREEDY, V.R. (2015): Coffee in health and disease prevention. Amsterdam: Academic Press.
- RÁCZ, A., BAJUSZ, D., FODOR, M., HÉBERGER, K. (2016): Comparison of classification methods with “n-class” receiver operating characteristic curves: A case study of energy drinks. *Chemometrics and Intelligent Laboratory Systems*. 151, pp. 34–43. <https://doi.org/10.1016/j.chemolab.2015.11.009>
- RAMOS-DIAZ, J.M., RINNAN, Å., JOUPPIA, K. (2019): Application of NIR imaging to the study of expanded snacks containing amaranth, quinoa and kañiwa. *LWT*. 102, pp. 8–14. <https://doi.org/10.1016/j.lwt.2018.12.029>
- RIAZ, M.N. (2016): Snack Foods, Processing. In: *Reference Module in Food Science*. Elsevier. pp. 98–108.
- RIBEIRO, J.S., FERREIRA, M.M.C., SALVA, T.J.G. (2011): Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*. 83 (5), pp. 1352–1358. <https://doi.org/10.1016/j.talanta.2010.11.001>
- RIBEIRO, J.S., SALVA, T. de J.G., SILVAROLLA, M.B. (2021): Prediction of a wide range of compounds concentration in raw coffee beans using NIRS, PLS and variable selection. *Food Control*. 125, p. 107967. <https://doi.org/10.1016/j.foodcont.2021.107967>
- RIEDMILLER, M., BRAUN, H. (1993): A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *IEEE International Conference on Neural Networks*. Presented at the IEEE

- International Conference on Neural Networks. pp. 586–591 vol.1.
<https://doi.org/10.1109/ICNN.1993.298623>
- RINNAN, Å., BERG, F., van den, ENGELSEN, S.B. (2009): Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*. 28 (10), pp. 1201–1222.
<https://doi.org/10.1016/j.trac.2009.07.007>
- ROGER, J.-M., BIANCOLILLO, A., MARINI, F. (2020): Sequential preprocessing through ORTHogonalization (SPORT) and its application to near infrared spectroscopy. *Chemometrics and Intelligent Laboratory Systems*. 199, p. 103975. <https://doi.org/10.1016/j.chemolab.2020.103975>
- ROGER, J.-M., BOULET, J.-C., ZEAITER, M., RUTLEDGE, D.N. (2020): Pre-processing Methods. In: *Comprehensive Chemometrics*. Elsevier. (3) p. 1–75.
<https://doi.org/10.1016/B978-044452701-1.00074-0>
- ROONEY, L.W., SERNA-SALDIVAR, S.O. (2015): Tortillas: Wheat Flour and Corn Products. Elsevier.
- ROUSSEL, S., PREYS, S., CHAUCHARD, F., LALLEMAND, J. (2014): Multivariate Data Analysis (Chemometrics). In: *Process Analytical Technology for the Food Industry*. New York, NY: Springer. pp. 7–59. https://doi.org/10.1007/978-1-4939-0311-5_2
- SAEYS, W., NGUYEN DO TRONG, N., VAN BEERS, R., NICOLAÏ, B.M. (2019): Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. *Postharvest Biology and Technology*. 158, p. 110981. <http://dx.doi.org/10.1016/j.postharvbio.2019.110981>
- SAFRIZAL, SUTRISNO, LILIK, P.E.N., AHMAD, U., SAMSUDIN. (2018): Estimation of Sensory Analysis Cupping Test Arabica Coffee Using NIR Spectroscopy. *IOP Conference Series: Earth and Environmental Science*. 147, p. 012008. <http://dx.doi.org/10.1088/1755-1315/147/1/012008>
- SAGI, O., ROKACH, L. (2018): Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*. 8 (4), p. e1249. <https://doi.org/10.1007/s11704-019-8208-z>
- SAJTOS L., MITEV A. (2007): SPSS Kutatási és adatelemzési kézikönyv. Budapest: Alinea Kiadó.
- SANDERS, T., A.B. (2016): Functional Dietary Lipids: Food Formulation, Consumer Issues and Innovation for Health. Woodhead Publishing.
- SANTOS, J.R., LOPO, M., RANGEL, A.O.S.S., LOPES, J.A. (2016): Exploiting near infrared spectroscopy as an analytical tool for on-line monitoring of acidity during coffee roasting. *Food Control*. 60, pp. 408–415. <https://doi.org/10.1016/j.foodcont.2015.08.007>
- SCHANDA, J. (2007): Colorimetry: understanding the CIE system. Hoboken: John Wiley & Sons.
- SCHOLZ, M.B.S., PAGIATTO, N.F., KITZBERGER, C.S.G., PEREIRA, L.F.P., DAVRIEUX, F., CHARMETANT, P., LEROY, T. (2014): Validation of near-infrared spectroscopy for the quantification of cafestol and kahweol in green coffee. *Food Research International*. 61, pp. 176–182. <https://doi.org/10.1016/j.foodres.2013.12.008>
- ŞEMEN, S., MERCAN, S., YAYLA, M., AÇIKKOL, M. (2017): Elemental composition of green coffee and its contribution to dietary intake. *Food Chemistry*. 215, pp. 92–100.
<https://doi.org/10.1016/j.foodchem.2016.07.176>
- SERNA-SALDIVAR, S.O. (2022): Snack Foods: Processing, Innovation, and Nutritional Aspects. Boca Raton: CRC Press.
- SHAN, J., SUZUKI, T., OGAWA, Y., KONDO, N. (2015): Coffee roasting degrees prediction in terms of weight loss with selected wavebands based on near-infrared spectroscopy. *Engineering in Agriculture, Environment and Food*. 8 (4), pp. 195–199. <https://doi.org/10.1016/j.eaef.2015.10.003>
- SHAN, J., WANG, X., HAN, S., KONDO, N. (2017): Application of Curve Fitting and Wavelength Selection Methods for Determination of Chlorogenic Acid Concentration in Coffee Aqueous Solution by Vis/NIR Spectroscopy. *Food Analytical Methods*. 10 (4), pp. 999–1006.
<https://doi.org/10.1007/s12161-016-0650-z>
- SHIROMA, C., RODRIGUEZ-SAONA, L. (2009): Application of NIR and MIR spectroscopy in quality control of potato chips. *Journal of Food Composition and Analysis*. 22 (6), pp. 596–605.
<https://doi.org/10.1016/j.jfca.2008.09.003>

- SIESLER, H.W. (Szerk.) (2002): Near-infrared spectroscopy: principles, instruments, applications. Wiley-VCH, 348 p.
- da SILVA ARAÚJO, C., MACEDO, L.L., VIMERCATI, W.C., SARAIVA, S.H. (2021): Spectroscopy Technique Applied to Estimate Sensory Parameters and Quantification of Total Phenolic Compounds in Coffee. *Food Analytical Methods*. 14 (9), pp. 1943–1952. <https://doi.org/10.1007/s12161-021-02025-0>
- da SILVA ARAÚJO, C., VIMERCATI, W.C., MACEDO, L.L., FERREIRA, A., PREZOTTI, L.C., TEIXEIRA, L.J.Q., SARAIVA, S.H. (2020): Predicting the Electric Conductivity and Potassium Leaching of Coffee by NIR Spectroscopy Technique. *Food Analytical Methods*. 13 (12), pp. 2312–2320. <https://doi.org/10.1007/s12161-020-01843-y>
- SØRENSEN, K.M., VAN DEN BERG, F., ENGELSEN, S.B. (2021): NIR Data Exploration and Regression by Chemometrics—A Primer. In: Y. OZAKI, C. HUCK, S. TSUCHIKAWA and S.B. ENGELSEN, eds. *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*. Singapore: Springer. pp. 127–189. https://doi.org/10.1007/978-981-15-8648-4_7
- SOUZA, J.C., PASQUINI, C., HESPANHOL, M.C. (2022): Feasibility of compact near-infrared spectrophotometers and multivariate data analysis to assess roasted ground coffee traits. *Food Control*. 138, p. 109041. <https://doi.org/10.1016/j.foodcont.2022.109041>
- STEFANELLO, N., SPANEVELLO, R.M., PASSAMONTI, S., PORCIÚNCULA, L., BONAN, C.D., OLABIYI, A.A., TEIXEIRA DA ROCHA, J.B., ASSMANN, C.E., MORSCH, V.M., SCHETINGER, M.R.C. (2019). Coffee, caffeine, chlorogenic acid, and the purinergic system. *Food and Chemical Toxicology*. 123, pp. 298–313. <https://doi.org/10.1016/j.fct.2018.10.005>
- STOCCHI, D.L., KEMPS, E., ANESBURY, D.Z. (2021): The effect of mental availability on snack food choices. *Journal of Retailing and Consumer Services*. 60, p. 102471. <https://doi.org/10.1016/j.jretconser.2021.102471>
- TOLESSA, K., RADEMAKER, M., DE BAETS, B., BOECKX, P. (2016): Prediction of specialty coffee cup quality based on near infrared spectra of green coffee beans. *Talanta*. 150, pp. 367–374. <https://doi.org/10.1016/j.talanta.2015.12.039>
- TUGNOLO, A., BEGHI, R., GIOVENZANA, V., GUIDETTI, R. (2019): Characterization of green, roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices. *Journal of Near Infrared Spectroscopy*. 27 (1), pp. 93–104. <https://doi.org/10.1177/0967033519825665>
- TUGNOLO, A., GIOVENZANA, V., MALEGORI, C., OLIVERI, P., CASSON, A., CURATITOLI, M., GUIDETTI, R., BEGHI, R. (2021): A reliable tool based on near-infrared spectroscopy for the monitoring of moisture content in roasted and ground coffee: A comparative study with thermogravimetric analysis. *Food Control*. 130, p. 108312. <https://doi.org/10.1016/j.foodcont.2021.108312>
- TUMULURU, J.S. (2016): Snack Foods: Role in Diet. In: *Encyclopedia of Food and Health*. Elsevier. pp. 6–12. <http://dx.doi.org/10.1016/B978-0-12-384947-2.00632-2>
- USDA (2022): Coffee: World Markets and Trade. *USDA Foreign Agricultural Service* [online]. Hozzáférés: <https://www.fas.usda.gov/data/coffee-world-markets-and-trade> Lekérdezés dátuma: 2023. 03. 10.
- VIGNOLI, J.A., VIEGAS, M.C., BASSOLI, D.G., BENASSI, M. de T. (2014): Roasting process affects differently the bioactive compounds and the antioxidant activity of arabica and robusta coffees. *Food Research International*. 61, pp. 279–285. <https://doi.org/10.1016/j.foodres.2013.06.006>
- WANG, T., RODRIGUEZ-SAONA, L.E. (2012): Rapid Determination of Sugar Level in Snack Products Using Infrared Spectroscopy. *Journal of Food Science*. 77 (8), pp. 874–879. <https://doi.org/10.1111/j.1750-3841.2012.02824.x>
- WEBB, G.I. (2010): Naïve Bayes. *Encyclopedia of Machine Learning*. Boston, MA: Springer US. pp. 713–714. https://doi.org/10.1007/978-0-387-30164-8_576

- WILLIAMS, P., DARDENNE, P., FLINN, P. (2017): Tutorial: Items to be included in a report on a near infrared spectroscopy project. *Journal of Near Infrared Spectroscopy*. 25 (2), pp. 85–90. <http://dx.doi.org/10.1177/0967033517702395>
- WINTGENS, J.N. (2004): Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers. Wiley. Available from: <https://doi.org/10.1002/9783527619627>
- WOLD, S., ESBENSEN, K., GELADI, P. (1987): Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. 2 (1), pp. 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- WONGSAIPUN, S. (2021): Application of Artificial Neural Network for Tracing the Geographical Origins of Coffee Bean in Northern Areas of Thailand Using Near Infrared Spectroscopy. *Chiang Mai Journal of Science*. 48 (1), pp. 163–175.
- WORKMAN, J.J. (2000): Near-Infrared Spectroscopy of Polymers and Rubbers. *Encyclopedia of Analytical Chemistry*. Chichester, UK: John Wiley & Sons, Ltd. p. a2021. <https://doi.org/10.1002/9780470027318.a2021>
- XIAOBO, Z., JIEWEN, Z., POVEY, M.J.W., HOLMES, M., HANPIN, M. (2010): Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*. 667 (1–2), pp. 14–32. <https://doi.org/10.1016/j.aca.2010.03.048>
- XIAPING F., YIBIN Y., DANTING Y. (2011): A comparative study of representative subset selection for NIR model updating. *American Society of Agricultural and Biological Engineers*. <https://doi.org/10.13031/2013.37411>
- YERETZIAN, C., JORDAN, A., BADOUD, R., LINDINGER, W. (2002): From the green bean to the cup of coffee: investigating coffee roasting by on-line monitoring of volatiles. *European Food Research and Technology*. 214 (2), pp. 92–104. <https://doi.org/10.1007/s00217-001-0424-7>
- YERGENSON, N., ASTON, D.E. (2020): Monitoring coffee roasting cracks and predicting with in situ near-infrared spectroscopy. *Journal of Food Process Engineering*. 43 (2). e13305. <https://doi.org/10.1111/jfpe.13305>
- YETURU, K. (2020): Chapter 3 - Machine learning algorithms, applications, and practices in data science. *Handbook of Statistics*. Elsevier. pp. 81–206. <https://doi.org/10.1016/bs.host.2020.01.002>
- YUN, Y.-H., LI, H.-D., DENG, B.-C., CAO, D.-S. (2019): An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends in Analytical Chemistry*. 113, pp. 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>
- ZANIN, R.C., KITZBERGER, C.S.G., BENASSI, M. de T. (2020): Characterization of Roasted *Coffea arabica* Species by the Relationship Between Caffeine and Diterpenes Contents. *Brazilian Archives of Biology and Technology*. 63, p. e20180752. <https://doi.org/10.1590/1678-4324-2020180752>
- ZHANG, X., LI, W., YIN, B., CHEN, W., KELLY, D.P., WANG, X., ZHENG, K., DU, Y. (2013): Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted Arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 114, pp. 350–356. <https://doi.org/10.1016/j.saa.2013.05.053>
- ZHU, M., LONG, Y., MA, Y., CHEN, Y., YU, Q., XIE, J., LI, B., TIAN, J. (2021): Comparison of chemical and fatty acid composition of green coffee bean (*Coffea arabica* L.) from different geographical origins. *LWT*. 140, p. 110802. <https://doi.org/10.1016/j.lwt.2020.110802>

Internetes hivatkozások

Internet 1.: https://www.alimente.elconfidencial.com/consumo/2022-10-14/retiran-lay-s-al-horno-al-punto-de-sal_3506591/

- Internet 2.: <https://pt.dreamstime.com/foto-de-stock-royalty-free-microplaquetas-de-milho-da-tortilha-do-queijo-do-nacho-image28780325>
- Internet 3.: <https://www.ingredientsnetwork.com/47/product/125/24/17/erdnussflips.jpg>
- Internet 4.: <https://www.indiamart.com/proddetail/3d-papad-snacks-4412994562.html>
- Internet 5.: <https://www.istockphoto.com/hu/fot%C3%B3/keksz-feh%C3%A9r-alapon-as%C3%B3s-kekszek-feh%C3%A9r-alapon-z%C3%A1rnak-gm1173654045-326075932>
- Internet 6.: https://www.freepik.com/free-photo/glazed-salted-pretzels-isolated-white-background_21016552.htm#query=pretzel%20white%20background&position=3&from_view=search&track=ais

M2.: A vizsgált snackekre vonatkozó nominális adatok, egyéb információk

Mintakód	Energia [kJ]	Energia [kcal]	Zsír [g]	Amelyből telített zsírsavak [g]	Szénhidrát [g]	Amelyből cukrok [g]	Rost [g]	Fehérje [g]	Só [g]	Gyártási technológia	Származási hely	Sütőolaj	Ízesítés	Alapvető fehérjeforrás
1_17	2271	545	35	2,7	52	0,5	4,2	4,5	1,7	burgonya (hullámos)	Ausztria	napraforgó	sós	burgonya
2_17	2255	541	35	3,2	48	0,6	-	5,5	0,78	burgonya	Csehország	napraforgó	sós	burgonya
3_17	2267	544	35	3,2	48	2,8	-	6,3	0,85	burgonya	Csehország	napraforgó	sajtos	burgonya
4_17	2277	547	35	3,2	49	3,4	-	5,8	0,85	burgonya	Csehország	napraforgó	paprikás	burgonya
5_17	2192	526	32	11	52	0,6	4,4	5,9	1,4	burgonya	Lengyelország	növényi olajok	sós	burgonya
6_17	1979	472	20	9	64,4	1,5	4	6,6	1,6	tortilla	Belgium	pálma	sajtos	kukorica
7_17	2105	503	27,6	12,8	55,6	1,7	5,9	5,1	1,7	extrudált	EU	pálma	sós	burgonya
8_17	2123	508	24	3,6	59	0,7	6,9	11	1,84	extrudált	Ausztria	napraforgó	mogyorós	kukorica
9_17	2137	512	29	13	54	4	-	7	3,3	extrudált	Lengyelország	pálma	sajtos	kukorica
10_17	2039	487	25	2,7	60	1,7	-	4,1	2,25	extrudált		napraforgó	paprikás	burgonya
11_17	2245	539	35	16	47	1,9	-	5,9	1,5	burgonya	Csehország	pálma	bacon	burgonya
12_17	2264	543	35	16	48	2,2	-	6	1,8	burgonya (hullámos)		pálma	sajtos-tejfölös	burgonya
13_17	2271	545	35	3,3	48	2,3	-	6,2	0,84	burgonya	Csehország	napraforgó	jalapenos-sajtos	burgonya
14_17	2170	519	28	2,9	61	3,1	-	4,4	0,9	extrudált	Magyarország	napraforgó	sajtos	burgonya
15_17	2050	489	23	2	63	4,6	1,9	6,5	1,6	extrudált	Lengyelország	repce	pizzás	kukorica
16_17	2254	541	34	3	50	3,8	-	5,4	0,7	burgonya	Csehország	napraforgó	fűszeres paradicsomos	burgonya
17_17	2157	516	29	3	55	2,8	-	7	2,1	extrudált	Magyarország	napraforgó	sajtos	kukorica
18_17	2166	519	31	10	53	2,9	4,4	6,3	1,7	burgonya (hullámos)	Lengyelország	pálma, napraforgó	sajtos-hagymás	burgonya
19_17	2018	482	22	7	63	4		5,9	2	tortilla	Németország	pálma, napraforgó	sajtos	kukorica
20_17	2148	515	30	10	53	3,3	4,3	5,8	1,9	extrudált	Lengyelország	pálma, napraforgó	ketchupos	burgonya
21_17	2139	513	30	10	53	3,2	4,4	5,8	2	burgonya	Lengyelország	pálma, napraforgó	hagymás-tejfölös	burgonya
22_17	2163	518	30	10	53	2,2	4,3	6,2	1,8	burgonya	Lengyelország	pálma, napraforgó	sajtos	burgonya
23_17	2141	512	28	2,7	56	4,4	-	7,1	1,7	extrudált	Magyarország	napraforgó	pizzás	kukorica
24_17	1979	472	20	9	64,4	1,5	-	6,6	0,9	tortilla	Belgium	pálma	sós	kukorica

Mintakód	Energia [kJ]	Energia [kcal]	Zsír [g]	Amelyből telített zsírsavak [g]	Szénhidrát [g]	Amelyből cukrok [g]	Rost [g]	Fehérje [g]	Só [g]	Gyártási technológia	Származási hely	Sütőolaj	Ízesítés	Alapvető fehérjeforrás
25_17	1979	472	20	9	64,4	1,5	-	6,6	1,6	tortilla	Belgium	pálma	chilis	kukorica
26_17	2100	505	24	11	68	2,5	5	6,1	0,9	tortilla	Görögország	pálma	chilis	kukorica
27_17	2100	505	24	11	68	2,5	5	6,1	0,9	tortilla	Görögország	pálma	sajtos	kukorica
28_17	2100	505	24	11	68	2,5	5	6,1	0,9	tortilla	Görögország	pálma	BBQ	kukorica
29_17	2100	505	24	11	68	2,5	5	6,1	0,9	tortilla	Görögország	pálma	sós	kukorica
30_17										egyéb	USA	növényi olaj	sós	sertésbőr
31_17	1979	472	20	9	64,4	1,5	-	6,6	1,3	egyéb	USA	napraforgó, repce	sós	búza
32_17	1979	472	20	2	64,4	1,5	4	6,6	0,9	tortilla	Belgium	pálma	BBQ	kukorica
33_17	1979	472	20	9	64,4	1,5	-	6,6	0,9	tortilla		napraforgó	sós	kukorica
34_17	1979	472	20	9	64	1,5	-	6,6	1,6	tortilla	Belgium	pálma	sós	kukorica
35_17	1979	472	20	9	64,4	1,5	-	6,6	1,6	tortilla	Belgium	pálma	chilis	kukorica
36_17	1979	472	20	9	64,4	1,5	-	6,6	1,6	tortilla	Belgium	pálma	sajtos	kukorica
37_17	2113	506	27	7,5	49	2,1	-	14	1,6	tortilla	Belgium	pálma	chilis	kukorica
38_17	2341	562	38	18	48	1,1	3,2	5,5	2	extrudált	Lengyelország	pálma	mogyorós	kukorica
39_17	2092	500	27	2,5	56	0,8	-	8,3	1,3	burgonya	Lengyelország	pálma	vajas-sós	burgonya
40_17	2168	520	30	10	53	3,1	4,5	6,3	1,3	burgonya	Svédország	napraforgó	sós	burgonya
41_17	2260	541	33,8	2,4	52,4	0,2	-	7,2	0,9	burgonya (hullámos)	Lengyelország	pálma, napraforgó	kebabos	burgonya
42_17	2253	541	35	3,2	48	0,6	4	5,4	1,5	burgonya	Csehország	repce	sós	burgonya
43_17	2252	541	35	3,2	47	2	-	6,8	1,8	burgonya	Csehország	napraforgó	sós	burgonya
44_17	2252	541	35	3,2	47	2	-	6,8	1,8	burgonya	Szlovákia	napraforgó	paprikás	burgonya
45_17		525	26,3		62,5		5	7,5	-	tortilla	Kanada	növényi olaj	sajtos	kukorica
46_17		500	25,8		59		3	7,5	-	tortilla	Kanada	növényi olaj	chilis	burgonya
47_17	2209	530	33,8	4,8	48	1,3	4,4	6,3	1,8	burgonya	Görögország Egyesült	kukorica	oregánós	burgonya
48_17	2037	489	30,9	3,5	44,4	15,8	-	6	1,3	egyéb	Királyság	napraforgó	sós	cékla, édes burgonya, fehérrepa
49_17										extrudált	Görögország	pálma	sajtos	kukorica
50_17	2223	532	30	4,6	56	3,4	1,7	8,4	2,56	extrudált	Görögország	kukorica	sajtos-paradicsomos	kukorica
51_17	2201	527	31,8	12,7	50,6	1,3	4,4	7,3	1,5	burgonya	Görögország Egyesült	pálma	sós	burgonya
52_17	2094	502	28,4	3,3	53	1,3	6,1	5,4	1,9	burgonya	Királyság	napraforgó	sós-balzsametes	burgonya

Mintakód	Energia [kJ]	Energia [kcal]	Zsír [g]	Amelyből telített zsírsavak [g]	Szénhidrát [g]	Amelyből cukrok [g]	Rost [g]	Fehérje [g]	Só [g]	Gyártási technológia	Származási hely	Sütőolaj	Ízesítés	Alapvető fehérjeforrás
53_17	2169	519	28,8	2,4	58,6	1,9	2,5	5,1	1,1	burgonya	Egyesült Királyság	napraforgó	érett cheddar-vöröshagyma	burgonya
54_17	2139	531	30,1	3,5	51,5	0,5	6,5	5,8	0,9	burgonya	Egyesült Királyság	napraforgó	enyhén sózott	burgonya
55_17										burgonya				
56_17	2138	511	30	12	50,5	3,6	4,4	7,8	3,1	burgonya	Görögország	pálma	BBQ	burgonya
57_17	2252	540	33,8	14	51	<0,5	5,1	6,8	1,5	extrudált	Görögország	pálma	sós	burgonya
58_17	2168	520	30,1	12,9	51	0,5	5,5	7	1,5	burgonya	Görögország	pálma	sós	burgonya
59_17	2410	579	40	17,6	51	3,5	0,3	4,6	1,9	extrudált burgonya	Görögország	pálma	sajtos-sonkás	kukorica
60_17	2160	518	30	10	53	3,4	4,4	6,1	1,3	(hullámos)	Lengyelország	pálma, napraforgó	csípős csirkeszárny	burgonya
61_17	2319	556	35	3	57	0,6	-	2,4	1	extrudált	Magyarország	napraforgó	sós	burgonya
62_17	2382	550	28,8	2,1	66,4	3,2	2,5	8	2,15	extrudált	Egyesült Királyság	repce	hagymás-tejfölös	quinoa
63_17	1883	449	17	1,2	68,4	2,2	4,5	6,5	1,98	extrudált burgonya	Egyesült Királyság	repce	sós	csicseriborsó
64_17	2105	504	28	2,6	53	4	-	6,3	1,5	(hullámos)	Csehország	napraforgó	édes chilis	burgonya
1_18	-	560	36	4	52	0	2	6	-	burgonya	Kanada		sós	burgonya
2_18	1979	472	20	9	64,4	1,5	-	6,6	0,9	tortilla	Belgium	pálma	sós	kukorica
3_18	-	540	30	4	60	2	4	6	-	tortilla	Kanada	pálma, napraforgó	sajtos	kukorica
4_18	2192	526	32	11	52	0,6	4,4	5,9	1,4	burgonya	Lengyelország	pálma, napraforgó	sós	burgonya
5_18	1482	350	3,3	0,6	72	1,8	2,8	6,7	0	puffasztott	Szlovénia	-		rizs
6_18	1722	408	5,5	2,6	75,4	2,5	-	12,1	4,51	egyéb	Magyarország	repce, pálma, napraforgó	sós	búza
7_18	1979	472	20	2	64,4	1,5	4	6,6	1,6	tortilla	Belgium	napraforgó	sajtos	kukorica
8_18	2137	512	31	2,2	50	1,6	4,3	6,2	1,7	burgonya (hullámos)	Lengyelország	napraforgó	tejfölös	burgonya
9_18	2011	480	23	2	62	6,7	1,6	5,6	3,3	extrudált	Lengyelország	repce	sajtos	kukorica
10_18	2048	489	23	2	63	4,6	1,9	6,5	1,6	extrudált	Lengyelország	repce	pizzás	kukorica
11_18	2123	508	27	3,8	53	3,3	3,1	11	1,7	extrudált	Lengyelország	repce	mogyorós	kukorica
12_18	1861	443	14	2	71	7,6	4	5	1,47	burgonya (sütőben)	Lengyelország	repce	sós	burgonya
13_18	1999	477	20,9	4,9	56,8	1	3,2	13,8	0,9	extrudált	Magyarország	napraforgó	mogyorós	köles
14_18	1954	466	19,5	1,4	66	2,2	3,2	9,3	2,8	extrudált	Egyesült Királyság	repce	tejszínes kapros	lencseliszt

Mintakód	Energia [kJ]	Energia [kcal]	Zsír [g]	Amelyből telített zsírsavak [g]	Szénhidrát [g]	Amelyből cukrok [g]	Rost [g]	Fehérje [g]	Só [g]	Gyártási technológia	Származási hely	Sütőolaj	Ízesítés	Alapvető fehérjeforrás
15_18	1704	406	14,4	2,5	58,8	0,7	8	6,4	1,5	puffasztott	Bulgária	olívaolaj	pesztós	barnarizs
16_18	2299	549	33	5	52	6	6	11	0,56	egyéb	Egyesült Királyság	repce	csípős indiai	rizs, burgonya
17_18	1994	476	22	6,2	63	0,8	-	5,3	3,3	egyéb	Magyarország	napraforgó, pálma	burgonyás	búzáliszt, burgonya
18_18	1935	464	26,7	2,7	48,5	0,9	-	4,6	0,9	tortilla	Németország	napraforgó	sós	kukorica
19_18	1862	446	24,9	2,6	49,6	5,4	-	4,8	2	tortilla	Németország	napraforgó	sajtos	kukorica
20_18	1979	472	20	2	64,4	1,5	4	6,6	1,6	tortilla	Németország	napraforgó	sajtos	kukorica
21_18	1979	472	20	2	64,4	3	4	6,6	1,4	tortilla	Németország	napraforgó	paprikás	kukorica
22_18	2153	514	33	3,6	51	1,2	2,7	4	1,3	extrudált	Belgium	napraforgó, kukorica	sós	burgonya
23_18	2331	559	34	3	57	2,3	-	4,5	1,5	tortilla	Németország	napraforgó	sós	kukorica
24_18	2104	504	28	2,6	53	4,4	4,4	6,3	1,3	burgonya	Csehország	napraforgó	chili és paprika	burgonya
25_18	2229	535	32,8	4,1	49,6	3,7	3,9	12,3	1,4	extrudált	Szlovákia	napraforgó	mogyorós	kukorica
26_18	2170	519	28	2,5	62	5,2	-	3,3	0,93	extrudált	Magyarország	napraforgó	ketchupos	burgonya
27_18	2259	542	34	2,7	54	1,7	-	6,2	1,5	bugonya	Németország	napraforgó	paprikás	burgonya
28_18	2267	544	35	3,2	48	2,8	-	6,3	0,85	burgonya	Csehország	napraforgó	sajtos	burgonya
29_18	2144	514	30	10	52	2,5	4,3	6,2	1,9	burgonya (hullámos)	Lengyelország	pálma, napraforgó	salsa	burgonya
30_18	2165	519	31	15	53	2,6	4,3	6	1,5	burgonya	Lengyelország	pálma	hagymás-tejfölös	burgonya
31_18	2212	530	32	2,6	53	2,4	3,5	5,8	2	burgonya	Németország	napraforgó	rozmaryingos	burgonya
32_18	2111	506	30	2,1	51	3,5	4	6	1,7	burgonya (hullámos)	Szlovákia	napraforgó	újhagymás	burgonya
33_18	2008	479	22	7	62	3,7	-	5,9	2,1	tortilla	Németország	pálma, napraforgó	paprikás	kukorica
34_18	1979	472	20	9	64,4	1,5	-	6,6	1,6	tortilla	Németország	pálma	chilis	kukorica
35_18	2407	575,3	44,8	17,7	42,9	1,7	-	1,1	2,6	extrudált	Lengyelország	repce, pálma	roston sült burgonya	tejsavópor, kukorica
36_18	1757	418	11,3	0,8	52,9	2,2	8,4	22	2,2	extrudált	Svédország	napraforgó	BBQ/paprika	burgonya, szója
37_18	1677	399	7,6	0,5	54,2	0,6	8,6	23,9	1,7	extrudált	Svédország	napraforgó	sós	burgonya, szója
38_18	1496	354	2,5	0,2	54	0,8	13,5	36	1,5	extrudált	Írország	repce	sós-paprikás	tejfehérje
39_18	1516	379	2,8	0,2	54	1,6	13,6	36	0,9	extrudált	Írország	repce	BBQ	tejfehérje
40_18	1622	387	6,5	<1,9	29	4	5,3	50	2,5	extrudált	Egyesült Királyság	repce	BBQ	tejfehérje
41_18	1622	387	6,5	<1,9	29	4	5,3	50	2,5	extrudált	Egyesült Királyság	repce	hagymás-tejfölös	tejfehérje

Mintakód	Energia [kJ]	Energia [kcal]	Zsír [g]	Amelyből telített zsírsavak [g]	Szénhidrát [g]	Amelyből cukrok [g]	Rost [g]	Fehérje [g]	Só [g]	Gyártási technológia	Származási hely	Sütőolaj	Ízesítés	Alapvető fehérjeforrás
42_18	1979	472	20	2	64	0,8	-	6,6	0,9	tortilla	Belgium	napraforgó	sós	kukorica
43_18	2229	535	33	2,5	52	1,4	3,4	5,7	1,7	burgonya	Németország	napraforgó	paprikás	burgonya
44_18	2252	540	34	3	50	4	-	5,4	0,7	burgonya (hullámos)	Csehország	napraforgó	mustáros hús	burgonya
45_18	-	496,7	29	13	51	1	5,3	8	-	burgonya	Egyiptom	növényi olaj	chili-lemon	burgonya
46_18	2105	505	31	3,3	50	1,7	2,4	5,2	2,2	extrudált	Ausztria	napraforgó	fűszeres	kukorica
47_18	2082	497	23	10	65	1,4	-	6	1,4	tortilla	Szlovákia	pálma	sajtos	kukorica
48_18	2277	547	35	3,2	49	3,4	-	5,8	0,85	burgonya	Csehország	napraforgó	paprikás	burgonya
49_18	2278	547	35	3,2	49	2,1	-	5,8	0,91	burgonya	Csehország	napraforgó	bacon	burgonya
50_18	1979	472	20	2	64,4	1,5	4	6,6	1,6	tortilla	Belgium	napraforgó	erős chili	kukorica
51_18	-	496,7	28	12,7	52	3,3	5,3	7,7	-	burgonya	Egyiptom	növényi olaj	paradicsomos	burgonya
52_18	2060	490	29	-	51	0,8	5,3	8	-	burgonya	Líbia	növényi olaj	paprikás	burgonya
53_18	1955	467	21,4	-	64	<1	3	4,7	-	extrudált	Líbia	növényi olaj	paprikás	burgonya
54_18	2097	502	30	13	51	0,6	5,4	7,9	-	burgonya	Egyiptom	növényi olaj	sós	burgonya
55_18	2012	480	22,9	10,3	64,1	2	-	4,4	3	extrudált	Spanyolország	pálma	ketchupos	burgonya
56_18	2141	512	28	3,8	57,3	5,1	3,1	6,3	2,4	extrudált	Spanyolország	kukorica	sajtos-bacon	kukorica
57_18	2173	519	23,1	2,3	67,3	7,5	-	5,8	1,4	extrudált	Spanyolország	napraforgó	BBQ	kukorica
58_18	1963	468	20,9	2,6	64,7	8,2	-	5,3	2	extrudált	Spanyolország	napraforgó	ketchupos	kukorica
59_18	2196	526	29,8	3,6	60,7	0,9	4,6	6,1	1,2	extrudált	Spanyolország	napraforgó	BBQ	kukorica
60_18	2253	541	34,3	4,1	49,3	0,4	5,2	6,3	1,36	burgonya	Portugália	napraforgó	sós	burgonya
61_18	1862	446	24,9	2,6	49,6	5,4	-	4,8	2	tortilla	Németország	napraforgó, vajolaj	sajtos	kukorica
62_18	2269	544	34	2,7	52	1,2	-	5,8	1,6	burgonya	Németország	napraforgó	paprikás	burgonya
63_18	1979	472	20	2	64	1,5	-	6,6	1,6	tortilla	Belgium	pálma	sajtos	kukorica
64_18	2099	502	26	3,3	59	1,2	2,3	6,8	2,1	extrudált	Olaszország	napraforgó	sajtos	kukorica
65_18	2168	520	30,1	12,9	51	2,1	6,2	7,5	2,1	burgonya	Görögország	pálma	oregánós	burgonya
66_18	2095	501	26	12	60	3,3	4,3	4,6	1,2	extrudált	Románia		pizza-sajt-paprika	
67_18	2070	494	25	11	60	6,1	4,2	4,5	1,4	extrudált	Románia		sajt-bacon-sajtos-ketchupos	
68_18	2010	480	23	2	59	0,5	2,5	8	1	burgonya	Olaszország	napraforgó	rozmaryingos	burgonya
69_18	2116	507	28	2,6	54	2,1	-	6,3	1,3	burgonya		napraforgó	sós-rozmaryingos	burgonya

Mintakód	Energia [kJ]	Energia [kcal]	Zsír [g]	Amelyből telített zsírsavak [g]	Szénhidrát [g]	Amelyből cukrok [g]	Rost [g]	Fehérje [g]	Só [g]	Gyártási technológia	Származási hely	Sütőolaj	Ízesítés	Alapvető fehérjeforrás
70_18	1841	437	12	1,1	75	1,6	2,5	6	1,8	extrudált	Olaszország	napraforgó	sós	kukorica, burgonya, amaránt
71_18	2104	503	26,2	2,4	57,7	3,8	3,7	7,3	1,6	burgonya	Egyesült Királyság	repce	édes chilis	burgonya
72_18	2150	514	26	11,4	62	2,4	0,9	7,6	2,63	extrudált	Görögország	pálma	bacon	kukorica
73_18	1708	407	4,6	1	83	-	3,56	7,4	1,7	puffasztott	Szerbia		tökmaggal	barnarizs
74_18	1878	448	18	2,8	61	2,9	7,5	6,7	0,9	puffasztott	Bulgária	olívaolaj	erős chili	barnarizs
75_18	2168	520	30,1	12,9	51	0,5	5,5	7	1,5	burgonya	Görögország	pálma	sós	burgonya
76_18	2244	539	37	18	45	1	-	6	2,2	extrudált	Görögország	repce	sajtos-paradicsomos	kukorica
77_18	2081	499	28,7	2,5	52,2	0,2	-	5,8	2,5	burgonya (hullámos)	Egyesült Királyság	napraforgó	sós-ecetes	burgonya
78_18	2139	513	30,1	3,5	51,5	0,5	6,5	5,8	0,9	burgonya (hullámos)	Egyesült Királyság	napraforgó	enyhén sózott	burgonya
79_18	2367	568	36	17	52	0,5	-	6,9	2,1	burgonya (hullámos)	Görögország	pálma	oregánós	burgonya
80_18	2510	600	42	19	50	0,6	1,5	6	1,9	extrudált	Görögország	pálma	sajtos	kukorica
81_18	2425	583	42	20	45	0,8	-	6	1,9	extrudált	Görögország	pálma	sajtos	kukorica
82_18	2221	532	32	2,6	52	2,8	4,4	6,4	1,3	burgonya	Franciaország	napraforgó, repce	BBQ	burgonya
83_18	2151	516	31	4	53	1,2	4,1	4,1	1,4	burgonya	Olaszország	olívaolaj	sós	burgonya
84_18	1680	399	8,7	0,8	58,5	10,3	7	18,1	2,2	extrudált	Egyesült Királyság	napraforgó	paradicsomos-zöldfűszeres	szója, burgonya, rizsliszt
85_18	2221	533	32	2,7	52	2,2	4,2	6,4	1,5	burgonya	Franciaország	napraforgó, repce, vajolaj	sültcsirkés	burgonya
86_18		553	34,8	-	54,5	-		5,6	1,2	burgonya	Japán		wasabi	burgonya
87_18	-	510	23	12	63	16	4	9	-	extrudált	Nigéria	margarin		búza
88_18	934	223	5,2	-	36,2	-	-	7,4	-	extrudált	Nigéria	vaj		búza
89_18	-	-	-	-	-	-	-	-	-	szárított	Nigéria		fűszeres	éretlen banán
90_18	-	-	-	-	-	-	-	-	-	szárított	Nigéria			kókusz
91_18	-	-	-	-	-	-	-	-	-	szárított	Nigéria			éretlen banán

M3.: A zöld kávék ismert jellemzői

mintakód	kontinens, terület	származási hely	megjegyzés (származási hely, egyéb)	faj	post-harvest
1	Dél-Amerika	Kolumbia	koffeinmentes	<i>Coffea arabica</i>	mosott
2	Ázsia	Indonézia	Sulawesi; bio	<i>Coffea arabica</i>	félíg-mosott
3	Közép-Amerika	Costa Rica	San Marcos de Tarrazú, San José	<i>Coffea arabica</i>	mosott
4	Afrika	Etiópia	Yirgacheffe	<i>Coffea arabica</i>	szárász
5	Közép-Amerika	Guatemala	Huehuetenago, Amate	<i>Coffea arabica</i>	mosott
6	Közép-Amerika	Guatemala	Antigua	<i>Coffea arabica</i>	mosott
7	Ázsia	India	Kerala, Rathnagiri Estate	<i>Coffea arabica</i>	mosott
8	Közép-Amerika	Jamaica	Blue Mountain régió, Typica	<i>Coffea arabica</i>	mosott
9	Közép-Amerika	Panama	Boquete, Chiriqui	<i>Coffea arabica</i>	mosott
10	Ázsia	Indonézia	Szumátra	<i>Coffea arabica</i>	nedvesen hántolt
11	Afrika	Tanzánia	Machere Estate, Blue Mountain	<i>Coffea arabica</i>	mosott
12	Afrika	Uganda		<i>Coffea arabica</i>	mosott
13	Afrika	Kenya		<i>Coffea arabica</i>	mosott
14	Közép-Amerika	Nicaragua	Maragogype	<i>Coffea arabica</i>	mosott
15	Közép-Amerika	Mexikó	Chiapas state, Jaltenango	<i>Coffea arabica</i>	mosott
16	Ázsia	Indonézia	Jáva	<i>Coffea arabica</i>	félíg-mosott
17	Ázsia	Pápua-Új-Guinea	Sigri estate	<i>Coffea arabica</i>	mosott
18	Afrika	Etiópia	Sidama	<i>Coffea arabica</i>	mosott
19	Afrika	Etiópia	Jima (Djimmah)	<i>Coffea arabica</i>	mosott
20	Dél-Amerika	Brazília	Cerrado, Pantano Farm; Yellow Bourbon	<i>Coffea arabica</i>	szárász
21	Dél-Amerika	Brazília	Sul de Minas régió	<i>Coffea arabica</i>	szárász
22	Ázsia	Nepál	Himalaya fennsík, Mt. Everest Supreme	<i>Coffea arabica</i>	mosott
23	Dél-Amerika	Kolumbia		<i>Coffea arabica</i>	mosott
24	Ázsia	Indonézia	Kalosi	<i>Coffea arabica</i>	mosott
25	Ázsia	India	cherry robusta	<i>Coffea canephora</i>	szárász
26	Dél-Amerika	Ecuador	Galapagos; Bourbon	<i>Coffea arabica</i>	mosott
27	Afrika	Kenya		<i>Coffea arabica</i>	mosott
28	Közép-Amerika	Honduras		<i>Coffea arabica</i>	mosott, honey
29	Közép-Amerika	El Salvador		<i>Coffea arabica</i>	mosott
30	Afrika	Zambia		<i>Coffea arabica</i>	mosott
31	Ázsia	Kelet-Timor		<i>Coffea arabica</i>	mosott
32	Dél-Amerika	Brazília	Santos	<i>Coffea arabica</i>	szárász
33	Ázsia	India	cherry robusta	<i>Coffea canephora</i>	szárász
34	Afrika	Uganda		<i>Coffea arabica</i>	mosott
35	Dél-Amerika	Kolumbia		<i>Coffea arabica</i>	mosott
36	Afrika	Uganda		<i>Coffea arabica</i>	mosott
37	Dél-Amerika	Kolumbia		<i>Coffea arabica</i>	mosott
38	Ázsia	Indonézia	Szumátra	<i>Coffea arabica</i>	mosott
39	Ázsia	Pápua-Új-Guinea		<i>Coffea arabica</i>	mosott
40	Közép-Amerika	Guatemala		<i>Coffea arabica</i>	mosott
41	Afrika	Kenya		<i>Coffea arabica</i>	mosott
42	Afrika	Etiópia		<i>Coffea arabica</i>	mosott
43	Közép-Amerika	Panama		<i>Coffea arabica</i>	mosott
44	Közép-Amerika	Mexikó		<i>Coffea arabica</i>	mosott
45	Ázsia	India		<i>Coffea arabica</i>	mosott
46	Afrika	Uganda		<i>Coffea arabica</i>	mosott
47	Afrika	Zambia		<i>Coffea arabica</i>	mosott
48	Dél-Amerika	Brazília		<i>Coffea arabica</i>	szárász
49	Dél-Amerika	Kolumbia		<i>Coffea arabica</i>	mosott
50	Közép-Amerika	Guatemala	Antigua	<i>Coffea arabica</i>	mosott

M4. Klorogénsavak meghatározása során alkalmazott tömegspektrometriás paraméterek

Függőnygáz (CUR):	45,0 psi
Ionspré feszültség (IS):	-4500 V
Hőmérséklet (TEM):	450 °C
Porlasztógáz (GS1):	45,0 psi
Szárítógáz (GS2):	45,0 psi
Ütközőgáz (CAD):	Alacsony

MRM táblázat. A kvantitatív átmenetek félkövér betűtípussal vannak jelölve.

Precursor ion (m/z)	Product ion (m/z)	Dwell time (msec)	Name	DP (volt)	EP (volt)	CE (volt)	CXP (volt)
353,123	191	8	5-CQA.1	-20	-10	-20	-19
353,123	85	8	5-CQA.2	-20	-10	-52	-9
353,123	93,1	8	5-CQA.3	-20	-10	-56	-11
354,046	191,9	8	5-CQA.4	-40	-10	-22	-11
353,075	191	8	3-CQA.1	-20	-10	-22	-13
353,075	178,9	8	3-CQA.2	-20	-10	-24	-19
353,075	134,9	8	3-CQA.3	-20	-10	-40	-15
354,104	192	8	3-CQA.4	-30	-10	-24	-11
353,022	173	8	4-CQA.1	-60	-10	-20	-15
353,022	179	8	4-CQA.2	-60	-10	-20	-11
353,022	134,9	8	4-CQA.3	-60	-10	-34	-11
353,022	191	8	4-CQA.4	-60	-10	-26	-15
514,948	352,9	8	4,5 DCQA.1	-35	-10	-26	-55
514,948	173,2	8	4,5 DCQA.2	-35	-10	-34	-9
514,948	178,9	8	4,5 DCQA.3	-35	-10	-34	-13
514,948	135,2	8	4,5 DCQA.4	-35	-10	-60	-9
514,986	352,9	8	3,5 DCQA.1	-40	-10	-24	-29
514,986	190,9	8	3,5 DCQA.2	-40	-10	-46	-17
514,986	178,9	8	3,5 DCQA.3	-40	-10	-36	-11
514,986	135	8	3,5 DCQA.4	-40	-10	-68	-15
366,912	193	8	3-O-feruloylquinic acid.1	-40	-10	-24	-11
366,912	134,2	8	3-O-feruloylquinic acid.2	-40	-10	-40	-13
368,104	193,9	8	3-O-feruloylquinic acid.3	-55	-10	-22	-13
366,912	173,2	8	3-O-feruloylquinic acid.4	-40	-10	-22	-21

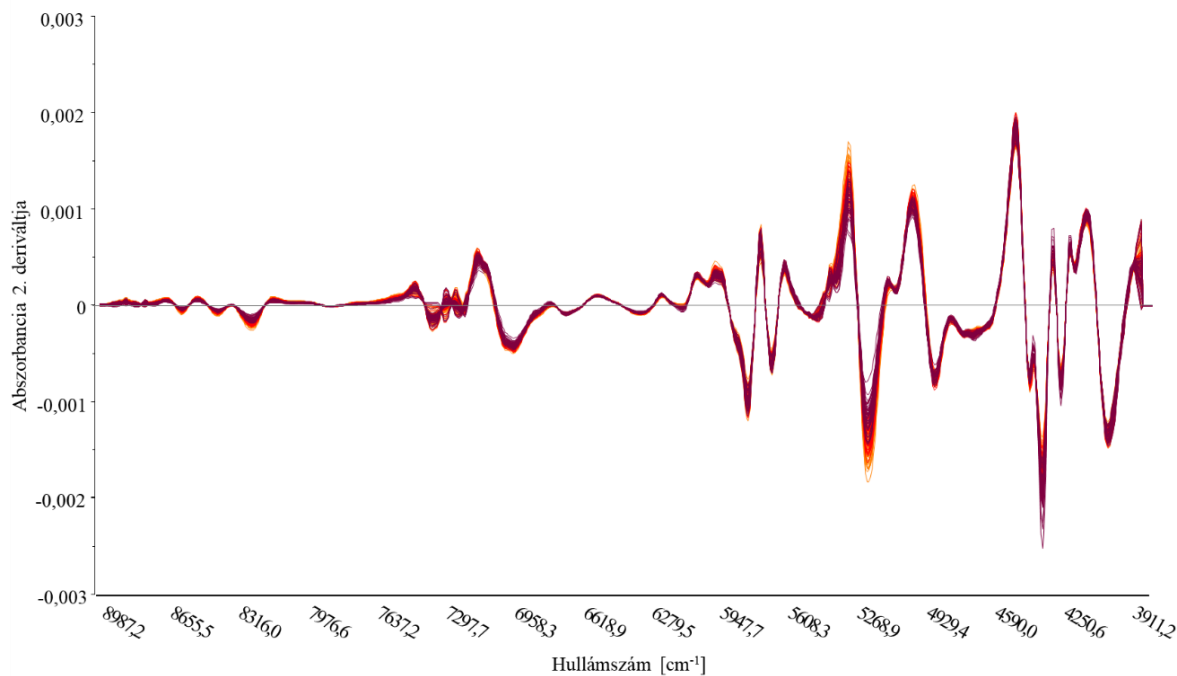
DP: declustering potential; EP: entrance potential; CE: collision energy, CXP: collision cell exit potential

M5.: A pörkölt kávé esetén meghatározott paraméterekre kapott korreláció elemzés eredménye

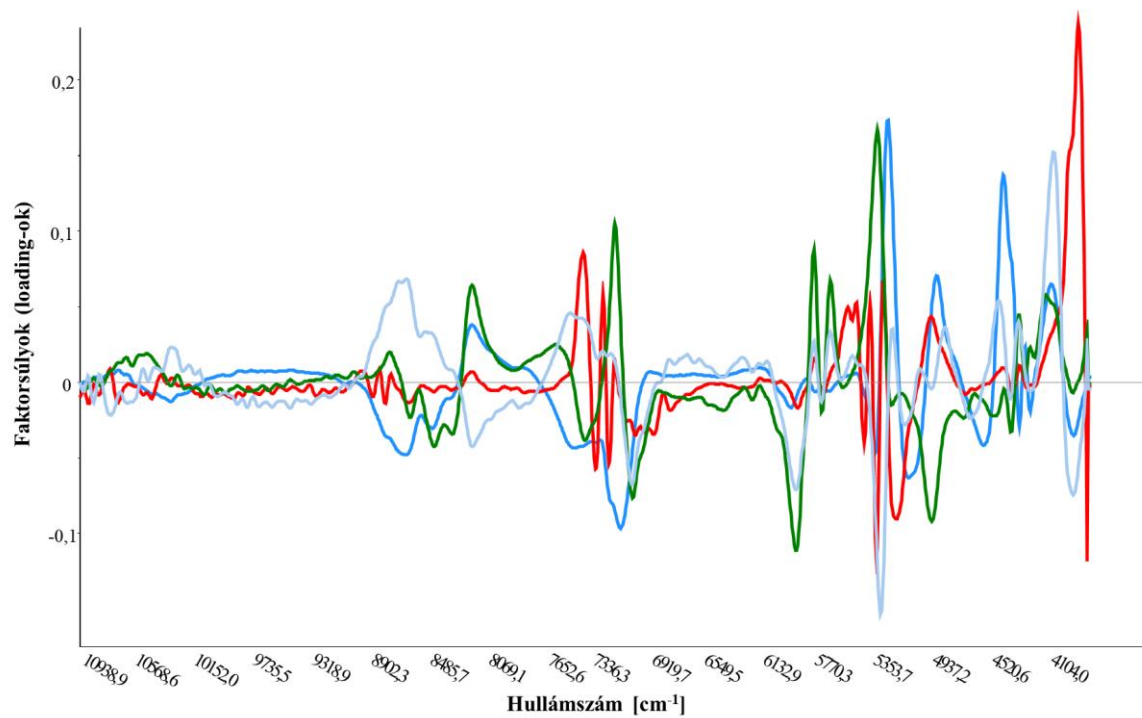
*p<0,05; ** p<0,01

	pH	koffein	5-CQA [mg/g]	4-CQA [mg/g]	3-CQA [mg/g]	3,5-dtCQA [mg/g]	4,5-dtCQA [mg/g]	3-FQA [mg/g]	sum CGA [mg/g]	észékszeri savasság	észékszeri keserűség	L*	a*	b*
pH	1													
koffein [mg/g]	0,33**	1												
5-CQA [mg/g]	-0,69**	-0,28**	1											
4-CQA [mg/g]	-0,59**	-0,28**	0,94**	1										
3-CQA [mg/g]	-0,54**	-0,24**	0,90**	0,97**	1									
3,5-dtCQA [mg/g]	-0,54**	-0,08	0,71**	0,76**	0,74**	1								
4,5-dtCQA [mg/g]	-0,08	0,11	0,24**	0,28**	0,25**	0,53**	1							
3-FQA [mg/g]	-0,33**	-0,06	0,50**	0,67**	0,67**	0,72**	0,33**	1						
sum CGA [mg/g]	-0,63**	-0,24**	0,96**	0,98**	0,96**	0,82**	0,37**	0,69**	1					
észékszeri savasság	-0,73**	-0,37**	0,69**	0,60**	0,56**	0,55**	0,21*	0,25**	0,64**	1				
észékszeri keserűség	0,73**	0,47**	-0,60**	-0,58**	-0,54**	-0,47**	-0,13	-0,29**	-0,58**	-0,81**	1			
L*	-0,66**	-0,17*	0,81**	0,77**	0,73**	0,75**	0,27**	0,45**	0,80**	0,65**	-0,60**	1		
a*	-0,49**	-0,11	0,73**	0,72**	0,68**	0,67**	0,22**	0,38**	0,73**	0,55**	-0,54**	0,86**	1	
b*	-0,62**	-0,18*	0,81**	0,78**	0,75**	0,75**	0,27**	0,45**	0,81**	0,65**	-0,58**	0,98**	0,91**	1

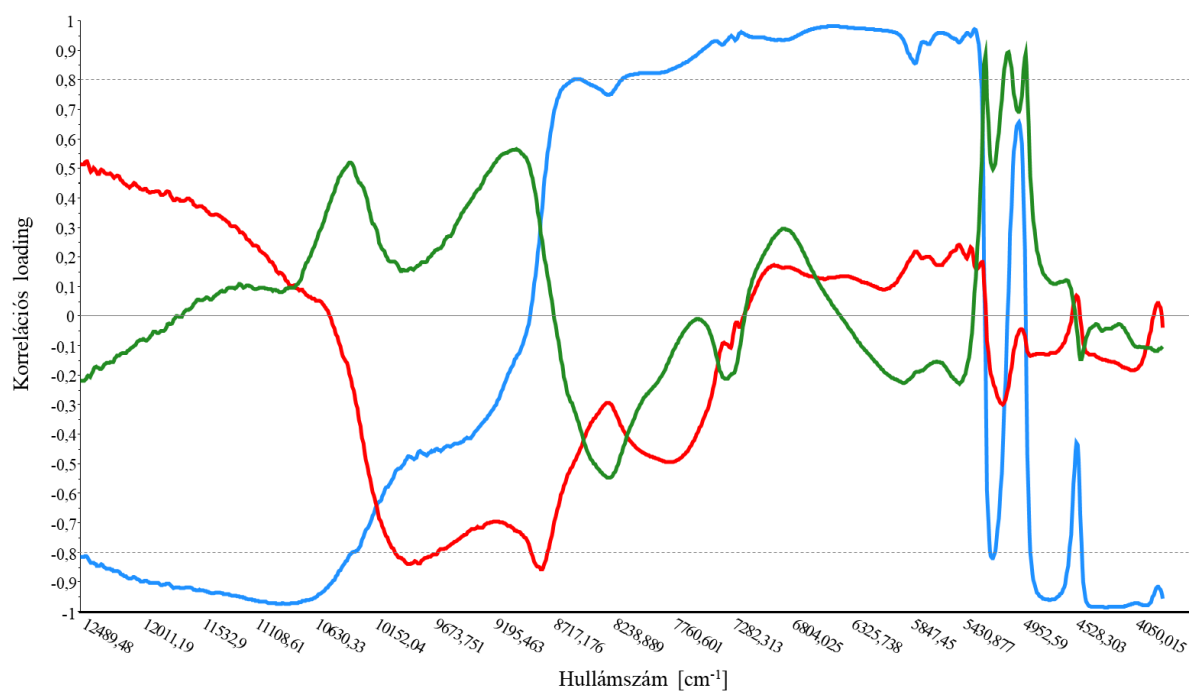
M6.: A pörkölt arabica kávéminták spektrumai második deriválás és sorstandardizálás adat-előkezelések után



M7.: Az afrikai és amerikai arabica kávék NIR spektrumaira futtatott főkomponens elemzés faktorsúly (loading) ábrája

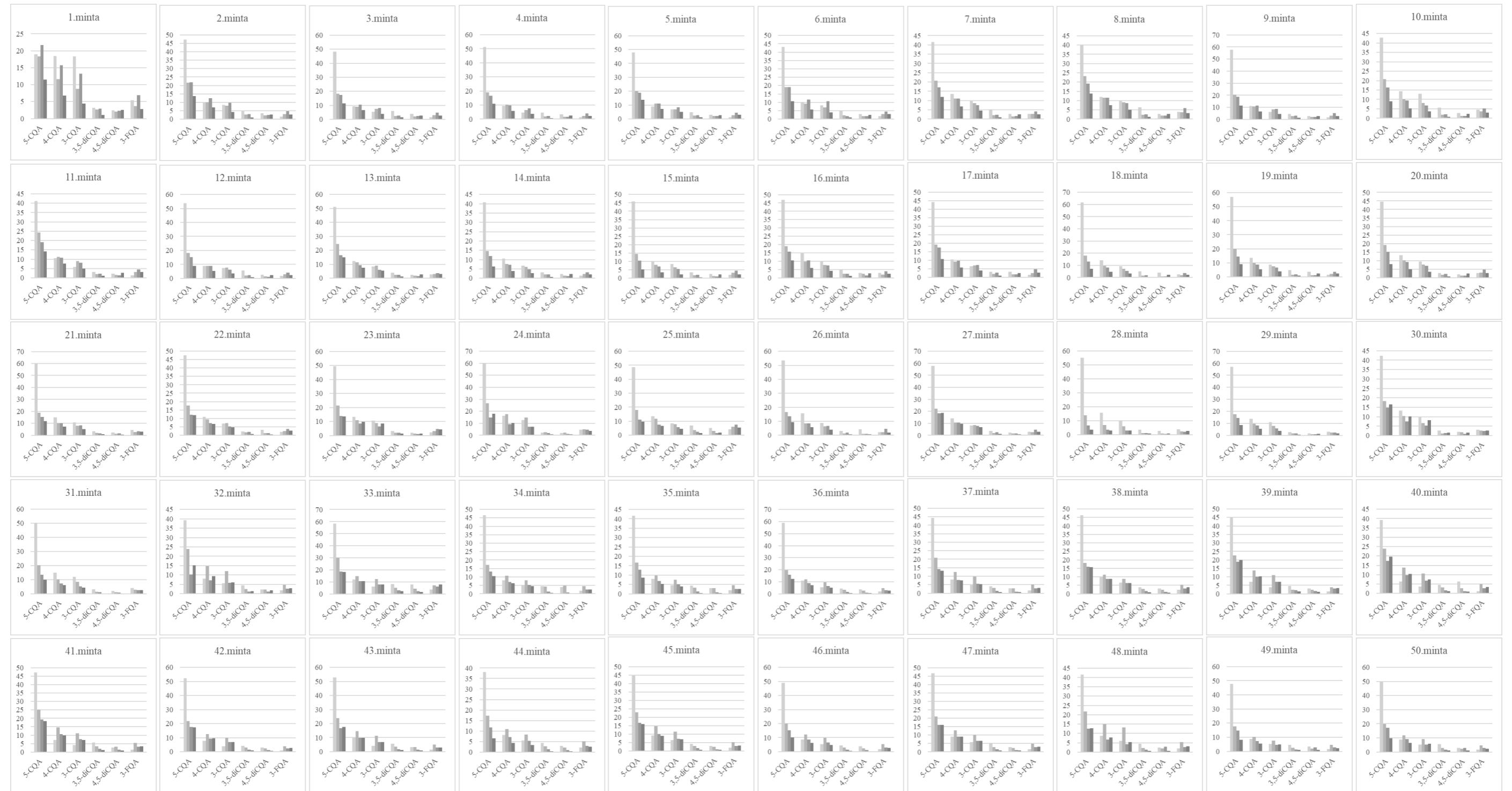


M8.: Az ázsiai és afrikai arabica kávék NIR spektrumaira futtatott főkomponens elemzés korrelációs loading-jainak ábrája



M9.: A pörkölés hatása a kávé 5-CQA, 4-CQA, 3-CQA, 3,5-diCQA, 4,5-diCQA, 3-FQA koncentrációjára (mg/g)

● zöld kávé ● világos pörkölés ● közepes pörkölés ● sötét pörkölés



*5-CQA: 5-caffeoylquinic acid, kinasav-5-O-kávészav; klorogénsav; 4-CQA: 4-caffeoylquinic acid; kinasav-4-O-kávészav; kriptoklorogénsav; 3-CQA: 3-caffeoylquinic acid, kinasav-3-O-kávészav; neoklorogénsav; 3,5-diCQA: 3,5-di-O-caffeoylquinic acid; kinasav-3,5-O-dikávészav; 4,5-diCQA: 4,5-di-O-caffeoylquinic acid; kinasav-4,5-O-dikávészav; 3-FQA: 3-O-feruloylquinic acid; kinasav-3-O-ferulasav

Köszönetnyilvánítás

Elsősorban konzulenseimnek, Dr. Fodor Mariettának és Dr. Gere Attilának szeretnék köszönetet mondani, hogy belém vetették a bizalmukat és megteremtették a kutatásom feltételeit. Doktoranduszi éveim alatt hozzáértésükkel végig segítettek a szakmai fejlődésemet az analitika, a statisztika és az érzékszervi minősítés területén is. Támogatásukat mind emberi, mind szakmai téren nagyra értékelem és hálás vagyok érte.

Köszönetet szeretnék mondani az Élelmiszerkémia és Analitika Tanszék minden doktoranduszának és munkatársának, akik a kutatásomat segítettek az évek alatt. Külön köszönet Firisz Zsuzsannának, Lippai Erzsébetnek és Nagy-Rédei Rékának a sok segítségért és támogatásért a hosszú laboratóriumban töltött napok során. Szeretném megköszönni Dr. Üveges Mártának és Dr. Abrankó Lászlónak, hogy segítettek a folyadékkromatográfia gyakorlati alkalmazásának elsajátításában. Nem utolsó sorban Dr. Jókainé Szatura Zsuzsannának, akihez mindig fordulhattam segítségért és Tormási Juditnak, akire közös doktoranduszi éveink során mindig számíthattam és hozzáállásával engem is motivált.

Köszönet a hallgatóimnak kitartó és lelkiismeretes munkájukért, akik részt vettek a mérések végrehajtásában.

A kávék beszerzése, pörkölése és érzékszervi minősítése során nyújtott szakmai segítségéért köszönettel tartozom Szongoth Jánosnak, Gál Miklósnak és Hajcsuk Tibornak.

Külön szeretném megköszönni páromnak Tóth Edgárnak, hogy az évek során mindenben támogatott, biztatott és szakmai tudásával végig segítette a munkámat. Hatalmas részt vállalt a klorogénsavak tömegspektrometriás vizsgálatában, amiért végtelenül hálás vagyok neki.

Végül szeretném megköszönni családomnak, akik mindig bíztak bennem és mindenben támogattak a doktoranduszi éveim alatt is.