Magyar Agrár- és Élettudományi Egyetem

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

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
Árukezelés, Kereskedelem, Ellátási Lánc és Érzékszervi Minősítési Tanszék

..... .....

Az iskolavezető jóváhagyása A témavezető(k) jóváhagyása

# 1. A MUNKA ELŐZMÉNYEI, CÉLKITŰZÉSEK

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 termékcsoportnál 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.

A kitűzött célokat a snackekre, zöld- és pörkölt kávéra vonatkozóan külön-külön fogalmaztam meg.

**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.

# 2. ANYAG ÉS MÓDSZER

## 2.1. Vizsgált minták

A kutatásaim során 155 kereskedelmi forgalomban kapható snack terméket vizsgáltam, melyek 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 termékek is. A minták gyártási technológia alapján a következő csoportokba sorolhatók: extrudált (50), olajban sütött burgonya chips (60), lúgosan főzött tortilla chips (31), puffasztott (4), szárított (3), egyéb (8). A legtöbb termék gyártása során burgonyát vagy kukoricát használtak alapanyagként, azonban bizonyos termékek búzát, rizst vagy szóját

A kávé esetében a kísérletekhez ötven zöld kávét szereztem be. Negyvennyolc arabica (egy koffeinmentes) és két robusta mintát vizsgáltam húsz különböző országból. A kávécseresznyék feldolgozása során, a legtöbb minta esetében (harminckilenc) a nedves eljárást alkalmazták, azonban száraz, félszáraz, félig mosott és „honey" eljárással készült kávékat is vizsgáltam. Az egyes mintákat, származási helyük alapján négy régióba soroltam be (Afrika, Ázsia, Dél-Amerika, Közép-Amerika). Az ötven eredeti mintát kisléptékben, egy előre meghatározott protokollt alkalmazva pörköltem meg. Három szintet állítottam be az első reccsenést figyelembe véve (pörkölés kezdete), melyek a világos, közepes és sötét pörkölés voltak.

## 2.2. Referencia és spektrális adatok meghatározása

A ***snackek*** minőségének vizsgálata során elengedhetetlen a makrokomponensek átlagos mennyiségének ismerete. 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. A sótartalom vizsgálatát direkt potenciometriás módszerrel hajtottam végre, amely az elektródpotenciálok mérésén alapuló elektroanalitikai eljárás.

A ***zöld- és pörkölt kávé***minták savasságát azok pH értékének meghatározásával adtam meg, tekintve, hogy a kávé érzékelt savassága annak

pH értékével erős korrelációt mutat. A koffeintartalom meghatározását az ISO 20481:2008 szabvány alapján, kisebb módosításokkal, HPLC-UV módszer alkalmazásával hajtottam végre. A kávébabban található klorogénsavak mennyiségi meghatározása HPLC-MS/MS módszerrel történt, gradiens elúcióval. A minták színét leíró paramétereket egy tristimulusos színmérő műszerrel rögzítettem a CIE *L*a*b** színtérben. 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 spektrális adatok rögzítéséhez egy Bruker MPA™ többfunkciós FT-NIR spektrométert (Bruker, Ettlingen, Németország) használtam, diffúz reflexiós mérési elrendezéssel. A snack termékekből 40-60 g darált mintát, míg a kávé mintákból kb. 50 g-ot eredeti, szemes formában vizsgáltam.

## 2.3. Kemometriai módszerek

A referencia és spektrális adatok kiértékelésére különböző kemometriai módszereket alkalmaztam. 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. A csoportok közötti különbségek vizsgálatára a Kruskal-Wallis próbát futtattam le ($\alpha$=0.05). Főkomponens elemzést (PCA) alkalmaztam az adatkészlet átfogó vizsgálata érdekében és a spektrálisan kieső minták meghatározására. A snackek esetén a spektrális adatok, míg a kávé vizsgálatánál mind a referencia, mind a spektrális adatok feltáró elemzésére alkalmaztam. Különböző adat-előkezelési módszereket futtattam le a spektrális adatokon a minták fizikai tulajdonságaiból eredő eltérések csökkentése érdekében, melyek a sor-normalizálás (SNV), átlagos szóráskorrekció (MSC), deriválás (első és második derivált, FD és SD), ill. az FD vagy SD és az SNV algoritmusok kombinációja voltak. A különböző adat-előkezelési módszereket kávé minták esetében az ortogonális jelkorrekció (OSC) előkezeléssel egészítettem ki.

### 2.3.1. Osztályozási módszerek

Az adatok elemzése során felügyelt tanítású algoritmusokat alkalmaztam a snack termékek és a kávé különböző minőségi jellemzőinek vizsgálatára. A snack termékek esetén a minták sokféleségéből adódóan, különböző lehetőségek álltak rendelkezésre azok osztályozására, amelyet az 1. táblázatban foglaltam össze. Az egyes szempontok alapján történő osztályozáshoz a véletlen erdő (RF), az RPropMLP algoritmusokat és a PLS-DA módszert alkalmaztam, melyeket minden esetben kereszt- és teszt-ellenőrzéssel is validáltam. 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.

**1. 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 |
| **származási hely** (3. adatkészlet) | Belgium | 14 |
| | Cseh Köztársaság | 14 |
| | Lengyelország | 19 |
| | Egyesült Királyság | 12 |
| **gyártási technológia** (4. adatkészlet) | olajban sütött | 60 |
| | lúgosan főzött (tortilla) | 31 |
| | extrudált | 50 |

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 2. táblázatban foglaltam össze.

**2. 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.

### 2.3.2. Részleges legkisebb-négyzetek elve regresszió (PLSR)

A különböző paraméterek mennyiségi becslésére a részleges legkisebb négyzetek elve (PLS) regressziót alkalmaztam mind a két termékcsoportnál. A létrehozott modellek teljesítményének értékelésére különböző statisztikai mutatókat vettem figyelembe. A validáláshoz random, ötszegmenses keresztellenőrzést, ill. teszt-ellenőrzést is alkalmaztam. Utóbbi esetében az adatkészlet felosztása 70/30 arányban történt, Kennard-Stone algoritmus segítségével (KENNARD és STONE 1969).

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 különböző beállításokkal, melyek az intervallum részleges legkisebb négyetek elve regresszió (iPLSR), a genetikus algoritmus (GA) és a projekció során a változók fontosságán (VIP) alapuló módszer voltak.

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., Mansos, WA, USA), KNIME analytics platform 4.0.2 (KNIME AG, Zürich, Svájc).

# 3. EREDMÉNYEK

## 3.1. Snack termékek

### 3.1.1. 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 mért komponensekre és a számolt energia tartalomra vonatkozó legfontosabb mennyiségi adatokat a 3. táblázatban foglaltam össze.

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

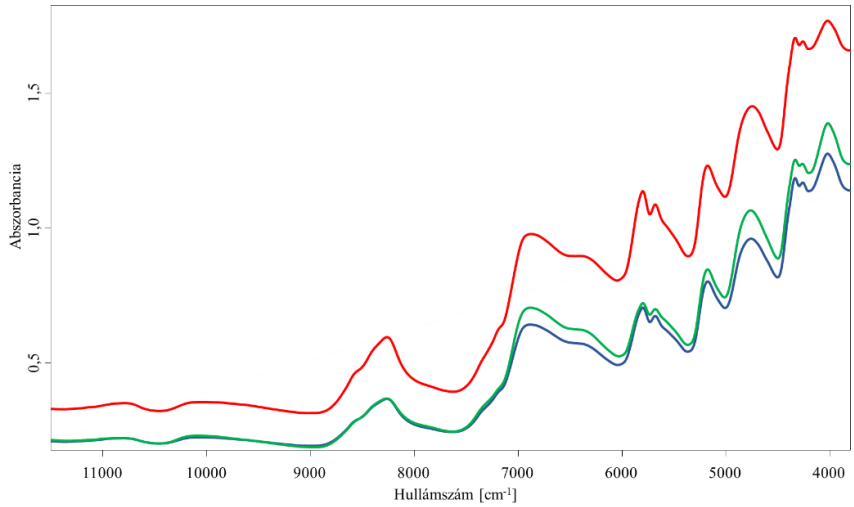
| | átlag±szórás | minimum | maximum | medián | mintaszám |
|---|---|---|---|---|---|
| **zsír [g/100 g]** | 26,68±9,57 | 2,24 | 56,13 | 27,48 | 156 |
| **szénhidrát [g/100 g]** | 56,07±10,21 | 20,51 | 84,02 | 55,91 | 90 |
| **cukor [g/100 g]** | 5,56±3,21 | 0,94 | 20,89 | 5,30 | 89 |
| **fehérje [g/100 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 összevetettem 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 azonosítottam a NIR spektrum jellegzetes abszorpciós sávjait, melyek kapcsolatba hozhatók a vizsgálni kívánt komponensekkel. Az átlagspektrumok két szakaszát (12800-11500 $cm^{-1}$ és 3800-3500 $cm^{-1}$) levágtam, mivel azok nem hordoztak szisztematikus információt. A snackek általában magas szénhidráttartalmú élelmiszerek, melyek végső formájának kialakítása olajban való sütéssel történik. Ebből következik, hogy a minták NIR spektrumai alapvető hasonlóságokat mutatnak (1. ábra). 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 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 felszínt hoznak létre. A fő különbségeket tehát a részecskeméret változása okozta, ami a spektrumok függőleges tengely mentén megfigyelhető eltolódásban nyilvánult meg (1. ábra).




**1. ábra:** Snack termékek átlag spektrumai gyártási technológia alapján: ● olajban sütött, ● tortilla, ● extrudált.

A PCA-t a spektrális kiesők feltárására használtam 95%-os konfidencia intervallum mellett, amelyhez az F-reziduális és a Hotelling-féle $T^2$ értékeket vettem figyelembe. Öt minta bizonyult valódi spektrális kiesőnek, ami a többitől eltérő mátrix tulajdonságokkal van összefüggésben. Ezek között volt sós ropi, puffasztott rizs, cékla és sült sertésbőr alapú termék is.

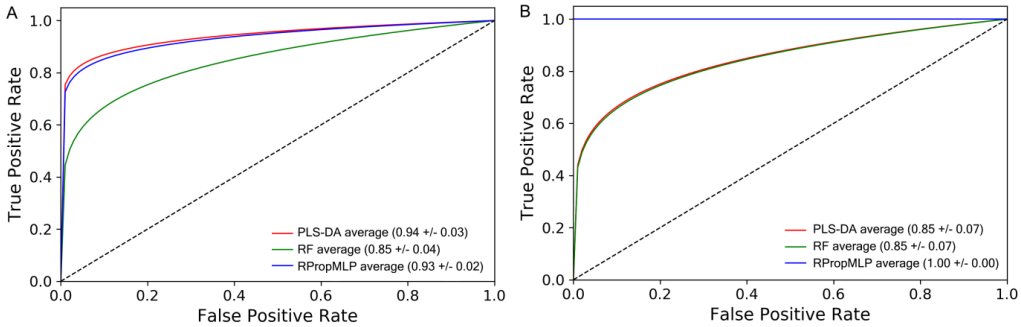
### 3.1.2. Snack termékek osztályozása a NIR spektrum alapján

Az elvégzett osztályozási eljárásokat a PCA során meghatározott spektrálisan kieső minták eltávolítása után hajtottam végre. A modellezéshez az első húsz főkomponenst használtam fel. Azokat a csoportokat, ahol a mintaszám ötnél

kisebb volt, kizártam a kiértékelésből, mert az osztályok mintaszámaiban jelentkező jelentős eltérések torzítják a modellt.

### *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 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 (ACC), a TPR és a valódi negatív minták aránya (TNR, specifikusság) értékek az AUC értékekkel együtt a 4. táblázatban vannak összefoglalva a kereszt- és a teszt-validálás esetében is.




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

**4. 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 |

### *Osztályozási modellek az alapanyag szempontjából*

A 2. adatkészlet (1. 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 5. 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 bírt, de a különbség nem jelentős. Következtetésképpen, a PLS-DA és az RPropMLP modellek ideálisak a snackminták alapanyag alapján történő osztályozására.

**5. 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 |

### *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. 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 (6. táblázat).

**6. 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 |

### *Osztályozási modellek gyártási technológia alapján*

Gyártási technológia alapján három nagy csoport hozható létre, amibe 141 mintát lehetett besorolni: olajban sütés, lúggal főzés és extrudálás. 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 7. táblázatban szereplő eredmények azt mutatják, hogy mindhárom modell használható a snacktermékek gyártása során alkalmazott technológiák meghatározására a NIR spektrumok alapján.

**7. táblázat:** A ötszegmenses 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 |

### *Részleges legkisebb négyzetek elve regresszió (PLSR) eredményei*

A snack termékek zsír, szénhidrát, cukor, fehérje és só tartalmának mennyiségi becslésére PLS regressziót alkalmaztam. Minden esetben random öt-elem-kihagyásos kereszt-ellenőrzést és teszt validálást használtam a faktorok optimális számának meghatározására, melyek megállapításához az RMSECV és a RMSEP 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 korábban felsorolt adat-előkezelési módszereket használtam.

A legjobbnak ítélt becslési függvényeket és azok legfontosabb statisztikai paramétereit a 8. táblázat tartalmazza.

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

| Paraméter | minták | 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 |

### *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 (teljes-spektrum modell), azonban az X-Y magyarázott variancia értékei ebben az esetben sem voltak megfelelőek, ezáltal a modell nem használható az energiatartalom becslésére. A különböző adat-előkezelési módszerekkel nem tudtam érdemben javítani az összefüggést. Az FD spektrum transzformáció a szóródási ábra és a modell nagy torzítása alapján nem ajánlott a teljes-spektrum modell esetében (9. táblázat).

A modell teljesítményének javítása érdekében intervallum PLS regressziót (iPLSR) futtattam. 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. Ennek

eredményeképpen azonban olyan abszorpciós sávokat kell kihagyni a kiértékelésből a meghatározott RMSECV értékek alapján, amelyek összefüggésben vannak a mintában található zsírokkal, olajokkal, fehérjékkel, a keményítővel és a cukrokkal is. 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 fentebb említett abszorpciós sávok 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 (9. táblázat).

**9. táblázat:** A fontosabb intervallum PLS (iPLS) modellek összehasonlítása a snackek energiatartalmára vonatkozóan. (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 – 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 |

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.

## 3.2. Zöld- és pörkölt kávék

### 3.2.1. Referencia adatok kiértékelése

***Zöld kávé***

Az ötven zöld kávé mintára vonatkozó kémiai mérések eredményeit és fontosabb statisztikai paramétereit a 10. táblázat foglalja össze. A mérési eredmények kiértékelés során a négy földrajzi régióból származó minták adatait egyszempontos varianciaanalízissel (ANOVA) hasonlítottam össze, azonban egyik változó tekintetében sem volt szignifikáns különbség tapasztalható.

**10. 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 |
| **∑CQA [mg/g]** | 76,07±9,76 | 61,18 | 98,07 | 74,19 | - |

RSD%: a referencia módszer átlagos relatív szórása (n=50)

A különböző régiókból származó minták kémiai összetételét meghatározzák az időjárási, termesztési körülmények és az alkalmazott feldolgozási módszerek. A vizsgált paraméterekre kapott mérési eredmények alapvetően megfelelnek korábbi tudományos kutatások megállapításainak. Annyit érdemes kiemelni, hogy az általam vizsgált minták különböző termőterületekről származtak, ezáltal a különböző komponensek mennyisége is széles tartományokban mozgott.

A mennyiségi adatok átfogó vizsgálata és vizualizálása érdekében PCA-t futtattam az adatkészleten. A koffeinmentes és a két robusta mintát kizártam a kiértékelésből, mivel kémiai összetételük jelentősen eltér a többi mintáétól. Az eredményeket figyelembe véve a minták közötti különbségeket elsősorban az 5-CQA, a 4-CQA, a 3-CQA, illetve a 3-FQA koncentrációjában jelentkező eltérések okozhatják a kapott faktorsúlyok alapján. Kisebb mértékben járulhat

hozzá a minták térben való szóródásához a 3,5- diCQA és a koffein mennyisége.

***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 (Agtron-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.

Három pörkölési szinten vizsgáltam a zöldkávé minták összetételének változását a különböző fizikai, kémiai és érzékszervi paraméterek mentén. A kapott eredményeket a 11. táblázat foglalja össze.

Általánosan a kávé savassága csökkenő tendenciát mutat a pörköltségi szint emelkedésével. A meghatározott pörkölési szintek között szignifikáns különbség volt tapasztalható ($\alpha$=0,05).

A koffein meghatározása során kapott eredmények összhangban vannak a korábbi kutatások eredményeivel. 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. Ezt alátámasztja a varianciaelemzés eredménye is, miszerint az egyes szintek között nincs szignifikáns különbség a koffein koncentrációja tekintetében (11. táblázat).

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

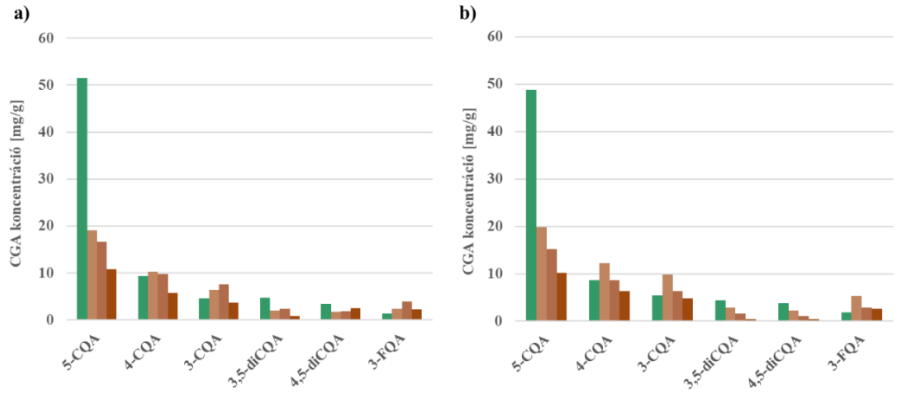
| | zöld | világos | közepes | sötét | RSD [%] |
|---|---|---|---|---|---|
| **pH** | 5,73±0,12 | 4,89±0,09[a] | 4,94±0,12[b] | 5,07±0,16[c] | 0,13 |
| **koffein [mg/g]** | 11,40±2,74 | 12,18±1,10[a] | 12,27±0,87[a] | 12,42±1,11[a] | 1,27 |
| **5-CQA [mg/g]** | 47,96±7,55 | 20,10±13,93[a] | 15,53±6,62[b] | 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 ($p<0,05$).
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. A 11. 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. Az egyes vizsgált klorogénsavak változására a következő főbb tendenciák jellemzők:

- 5-CQA – egyértelmű csökkenés, kivéve öt mintát; a sötét pörkölésnél nagyobb koncentráció, mint a közepesnél
- 4-CQA – nincs egyértelmű tendencia, sok esetben növekedés a zöld kávéhoz képest; legtöbb esetben a sötét pörkölésnél a legkisebb a mennyisége
- 3-CQA – két fő tendencia - egyértelmű csökkenés vagy növekedés és szisztematikus csökkenés
- 3,5-diCQA – kétféle tendencia - egyértelmű csökkenés vagy a zöldhöz képest csökken, majd nő és ismét csökken
- 4,5-diCQA és 3-FQA - legalább hatféle tendencia; a 4,5-diCQA-nál általában 50%-os csökkenés a pörkölés végére; a 3-FQA mennyisége a pörkölés végére kb. 70%-os növekedést mutatott (35db minta)

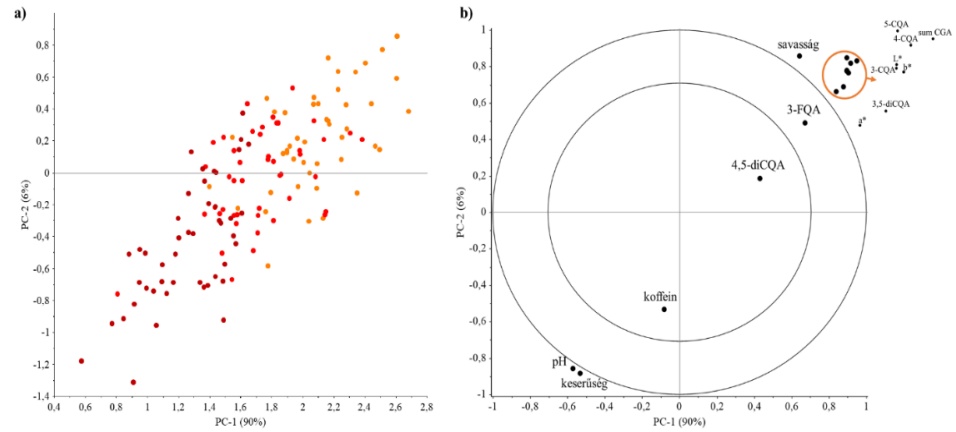
Megfigyelhető továbbá, hogy egy adott kávét vizsgálva, a 4-, és 3-CQA koncentrációja hasonló tendencia szerint változott a pörkölés során (42 mintánál). Ez azt jelenti, hogy ha két pörkölési szint között az egyik vegyület mennyisége nőtt/csökkent, akkor a másiké is.




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

A referencia adatokon korrelációelemzést és PCA-t is lefuttattam, melyek eredményei közül csak néhányat emelek ki. A PCA eredményeiből jól látszik (4. ábra), 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, sötéttől világos pörkölésig. A korrelációs loading ábra (4. ábra) megmutatja, hogy mely változóknak van a legjelentősebb hatása a minták pozíciójára. A

kávé színe mellett, a benne található klorogénsavak mennyisége is indikátora lehet a pörköltségi állapotnak. Az eredményekből 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 (4.b ábra).




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

Ha megfigyeljük a 4.b ábrát, a pH változó elhelyezkedése azzal magyarázható, hogy 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 érzékszeri 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).

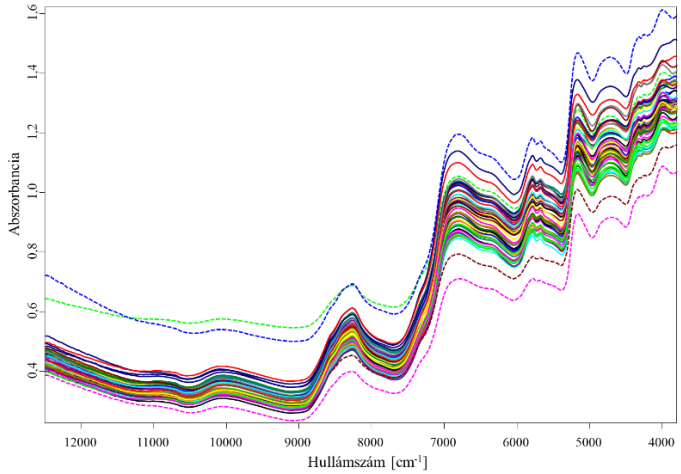
A minták pozícióját nagymértékben meghatározza a bennük található klorogénsavak mennyisége és összetétele is, ami a korrelációs loading-ok alapján egyértelműen látszik. Azok a minták, melyek nagyobb mennyiségű CGA-at tartalmaznak, savasabb karakterűnek bizonyultak az érzékszervi bírálat alapján is. Viszonylag 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 (α=0,01).

### 3.2.2. Spektrális adatok kiértékelése

***Zöld kávé***

A 5. ábra a zöld kávé minták 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, ugyanakkor szemmel látható eltérés tapasztalható a koffeinmentesített 1-es mintánál (●), a két robusta (● 25, ● 33) és a 49-es (●) arabica kávé esetében. Második deriválás spektrum transzformáció alkalmazása után meghatároztam a kávéra jellemző abszorpciós sávokat is.




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

PCA alkalmaztam, az eredeti és a sor-normalizált (SNV) adatok felhasználásával a 12500-3800 $cm^{-1}$ tartományra. A PCA eredményei a minták származási helyével összefüggésben is vizsgálhatók az adatkészletben rejlő mintázatok elemzésén keresztül. Az eredmények alapján elmondható, hogy a négy régióra vonatkozóan a minták nem mutatnak egyértelmű elkülönülést, azonban kisebb klaszterek kialakultak. Hasonló eredményre jutottak amerikai és ázsiai minták vizsgálata során korábbi kutatások. A minták pozícióját a kémiai összetétel mellett, erősen befolyásolják a zöldkávé fizikai tulajdonságai

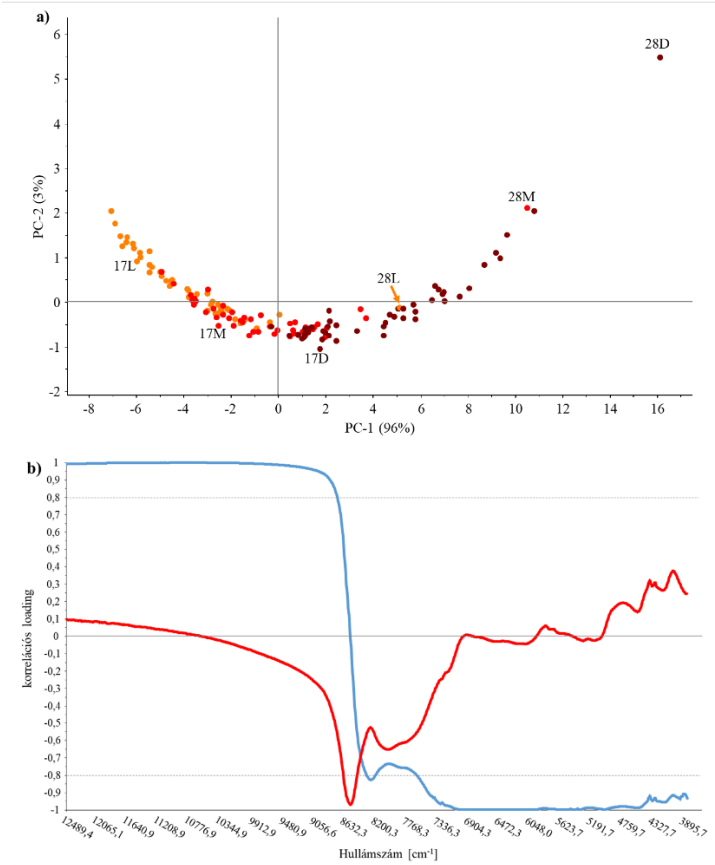
is, így a kávébab megjelenése, formája és színe. Ezek a jellemzők a 12500-9000 $cm^{-1}$ közötti spektrális tartományt befolyásolják leginkább. A meghatározott korrelációs loading-ok értelmében, erős pozitív korreláció tapasztalható a minták elhelyezkedése és azok fizikai tulajdonságai között.

***A pörkölés hatása a NIR spektrális adatokra***

A zöld- és a pörkölt kávé spektrális jellemzői a fő abszorpciós sávok tekintetében hasonlóak. A 8500 és 3800 $cm^{-1}$ közötti spektrális régióban mért abszorbancia értékek csökkenő tendenciát mutattak a pörkölés előrehaladtával (ahogy a nedvességtartalom csökken), ami egyetértésben van korábbi tanulmányok eredményeivel. Ez a jellegzetesség a 12500-8500 $cm^{-1}$ tartományban (látható tartományhoz közel) megfordul, a mért abszorbancia 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.

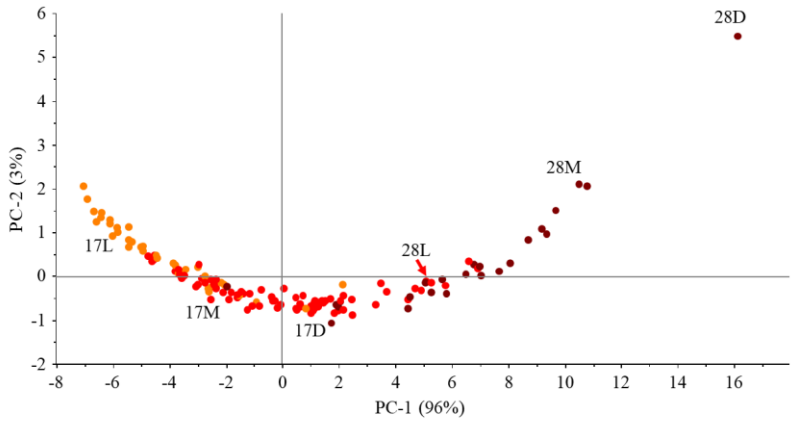
A pörkölt kávéminták NIR spektrumain is lefuttattam a PCA-t SNV adat-előkezelés után, ezeket az eredményeket a 6. ábra szemlélteti. A kiértékeléshez a 9000-3800 $cm^{-1}$ spektrális tartomány adatait használtam fel. Általánosságban elmondható, hogy a PC1 mentén a világos pörkölésű kávék inkább negatív értékeket (kivéve a 28L és a 44L mintát) vettek fel (6.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 „pörköltségi sor” alakult ki. A minták elhelyezkedése összefüggésben áll a pörkölési szint és a pörkölés során bekövetkező klorogénsavak koncentrációjának változásával.

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 kávé fizikai- és kémiai tulajdonságai is nagymértékben befolyásolják a minták pozícióját.




**6. 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 bemutatott mérési eredmények alapján, a CGA-ak koncentrációja között szignifikáns különbség van az egyes pörkölési szinteket figyelembe véve. A pörkölés során a legegyértelműbb 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 (7. 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.




**7. 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 vizsgáltható NIR spektroszkópiával, viszont a folyamatok megértéséhez átfogó adatbázissal kell rendelkeznünk a minták fizikai és kémiai jellemzőiről is.

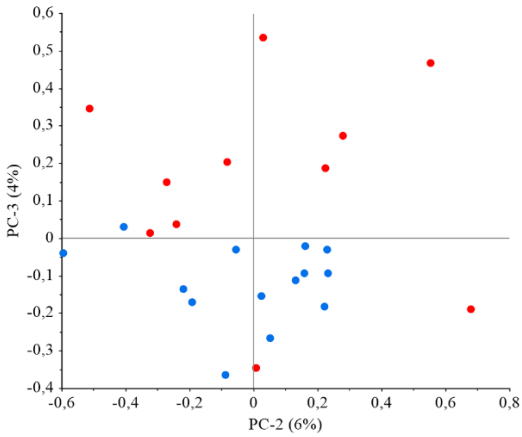
### 3.2.3. Zöld kávé eredetének vizsgálata

A kiértékeléshez negyvenhat minta, teljes spektrális tartományra (12500-3800 cm-1) vonatkozó adatait használtam fel. A klasszifikációs folyamatot négy adatkészleten hajtottam végre, négy, három és két-két származási régióval, mint csoportosítási szempont.

A kutatás során vizsgált zöld kávék a világ számos pontjáról származtak, ezáltal fizikai tulajdonságaik és kémiai összetételük is jelentős eltéréseket mutatnak. 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.

Négy és három termesztési régió vizsgálatánál az eredmények azt mutatják, hogy a modellek teljesítőképessége nem megfelelő a kávéminták osztályba sorolása során. A kialakuló klaszterek valószínűleg olyan tényezőkkel magyarázhatók, melyek számomra ismeretlenek, pl. a növény genotípusa. A modellek statisztikai paraméterei nagymértékben javultak bináris rendszerek

estén (Afrika-Amerika, Afrika-Ázsia). A legjobb eredményt az afrikai és ázsiai minták osztályozása során FD+SNV adat-előkezeléssel értem el (8. ábra), amelynél a modell pontossága 92%-nak adódott.




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

### 3.2.4. Részleges legkisebb négyzetek elve regresszió (PLSR) eredményei

***Zöld kávé***

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 és 3800 $cm^{-1}$ közötti spektrális tartományt vizsgáltam. Becslési függvényeket hoztam létre a zöldkávé 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 optimálása eltérő stratégiák mentén, empirikus úton történik. Ezáltal a regressziós összefüggések statisztikai paramétereinek javítására spektrum transzformációs módszereket és különböző változókiválasztási eljárásokat, illetve ezek kombinációit használtam. A legjobbnak ítélt PLS modellek statisztikai paramétereit a 12. táblázat foglalja össze. A modellek összehasonlításához elsősorban a determinációs együtthatókat ($R^2$; $Q^2$) és az átlagos négyzetes hiba értékeket vettem figyelembe.

**12. 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 | Minták | LV | Kereszt-ellenőrzés | | | | Teszt-validálás | | | | Mérési tartomány |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $R^2$ | $Q^2$ | RMSEC | RMSECV | $R^2_t$ | $Q^2_t$ | RMSEC | RMSEP | |
| **pH** | OSC | 125 | 3 | 0,81 | 0,78 | 0,05 | 0,05 | 0,82 | 0,77 | 0,05 | 0,05 | 5,55-6 |
| **koffein** | 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** | 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** | 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** | 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** | 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** | 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** | OSC | 133 | 3 | 0,99 | 0,99 | 0,10 | 0,13 | 0,99 | 0,96 | 0,09 | 0,12 | 0,95-4,86 |

Minden vizsgált paraméternél sikeres volt a modellezés, melyeknél a legjobb eredményt minden esetben OSC adat-előkezelés után értem el, amelynek jelentős hatása van a modellek komplexitására (látens változók számának csökkentése) és a statisztikai paraméterek értékeire is.

### *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 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.

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 optimáláshoz minden változót és azok lehetséges kombinációit felhasználja. Az eljárás eredményessége magyarázható a spektrális adatok komplexitásával, ami a kávé molekuláris rezgéseinek összetettségéből ered.

Az optimális eredményt a pörkölt kávék esetében is az OSC adat-előkezelést követően értem el. 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.

**13. táblázat:** A PLS regresszióval, pörkölt kávék esetén 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 | Minta | LV | Kereszt-ellenőrzés | | | | Teszt-validálás | | | | Mérési tartomány |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $R^2$ | $Q^2$ | RMSEC | RMSECV | $R^2_t$ | $Q^2_t$ | $RMSEC_t$ | RMSEP | |
| **pH** | OSC | 137 | 5 | 0,93 | 0,89 | 0,04 | 0,05 | 0,93 | 0,86 | 0,04 | 0,05 | 4,68-5,39 |
| **koffein** | 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** | 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** | 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** | 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** | OSC | 125 | 5 | 0,89 | 0,81 | 0,25 | 0,34 | 0,90 | 0,81 | 0,23 | 0,39 | 0,33-3,45 |
| **savasság** | OSC | 125 | 2 | 0,85 | 0,81 | 0,57 | 0,66 | 0,87 | 0,78 | 0,54 | 0,70 | 1,17-7,00 |
| **keserűség** | OSC | 130 | 3 | 0,89 | 0,83 | 0,53 | 0,68 | 0,91 | 0,82 | 0,48 | 0,69 | 1,30-9,30 |

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 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. 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.

# 4. KÖVETKEZTETÉSEK ÉS A JAVASLATOK

Doktori munkám során egy eddig kevéssé vizsgált élelmiszermátrix (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 kevéssé 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 képzik. 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 teszik 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 változókiválasztási 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 vizsgálata 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 pörköltségi 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 szóráskorrekciós 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.

# 5. ÚJ TUDOMÁNYOS EREDMÉNYEK

1. Elsőként foglalkoztam különböző gyártástechnológiával előállított snack termékek FT-NIR spektroszkópiával történő vizsgálatával, ami során feltérképeztem a termékcsoportra 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.

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.

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 teljesítőképessége bináris rendszerekben (Afrika-Amerika, Afrika-Ázsia) a legjobb.

# AZ ÉRTEKEZÉS TÉMAKÖRÉHEZ KAPCSOLÓDÓ PUBLIKÁCIÓI

**IF-es folyóiratcikkek:**

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

Eszter **Benes**; Dávid Bajusz; Attila Gere; Marietta Fodor; Anita Rácz (2020): Comprehensive chemometric classification of snack products based on their near infrared spectra. LWT-FOOD SCIENCE AND TECHNOLOGY. SJR indikátor: **D1**. IF-érték: 4,952

Eszter **Benes**; Marietta Fodor; Sándor Kovács; Attila Gere (2020): Application of Detrended Fluctuation Analysis and Yield Stability Index to Evaluate Near Infrared Spectra of Green and Roasted Coffee Samples. PROCESSES. SJR indikátor: SJR indikátor: **Q2**. IF-érték:2,847

John-Lewis Zinia Zaukuu; Eszter **Benes**; György Bázár; Zoltán Kovács; Marietta Fodor (2022): Agricultural Potentials of Molecular Spectroscopy and Advances for Food Authentication: An Overview. PROCESSES. SJR indikátor: **Q2.** IF-érték: 2,847

Brigitta Jesztl; Eszter **Benes**; Marietta Fodor (2019): FT-NIR origin identification of coffee samples. JOURNAL OF FOOD INVESTIGATION. SJR indikátor: **Q4** IF-érték:0,05

Zsóka Kárpáti; Eszter Benes; Marietta Fodor (2018): Nutritional analysis of coffee dregs for utilization purposes using classical, ICP-OES and FT-NIR tech Multimédiás tananyag - Élelmiszer- és biomérnöki analitika tárgy niques. JOURNAL OF FOOD INVESTIGATION. SJR indikátor: **Q4.** IF: 0,05

**Magyar nyelvű (teljes)**

Benes, Eszter Luca; Fodor, Marietta. Különböző snack termékek beltartalmi értékeinek mennyiségi meghatározása FT-NIR spektroszkópia alkalmazásával. Magyar Tudományos Akadémia, Kertészeti és Élelmiszertudományi Bizottság, Élelmiszertudományi Albizottság Workshop (2018. december 6.).

**Nemzetközi konferencia (teljes)**

Eszter, Benes; Marietta, Fodor. Non-destructive measurement of carbohydrate content using NIR-spectroscopy the snack example. Third International Conference on Food Science and Technology konferenciaközlemény. 2018.

**Nemzetközi konferencia (összefoglaló)**

Eszter, Benes; Marietta, Fodor. Development of rapid, non-destructive method for quantitative determination of fat in different chips. II. YRICCCE: Program and Book of Abstracts p. 58 2018.

Eszter, Benes; Marietta, Fodor; Attila, Gere. Applicability of detrended fluctuation analysis (DFA) to near-infrared spectroscopic data. SZIEntific Meeting for Young Researchers. 2019

Eszter, Benes; Marietta, Fodor; Attila Gere: Prediction of coffee acidity and bitterness using sensory, near infrared spectroscopy and color measurement data. EuroSense 2020 - European Conference on Sensory and Consumer Research (Dublin, Írország)