



Pannon Egyetem
Intézményi Tudományos Diákköri
Konferencia
2024



Növényi illóolajok szuperkritikus extrakciójának vizsgálata és fejlesztése kemometriai módszerekkel

Készítette: Juhász Csenge Judit
Témavezetők: Dr. Tarcsay Bálint Levente
Preiner Sára
Dr. Pethő Dóra

Mérnöki kar, Folyamatmérnöki Intézeti Tanszék
A pályamunka lezárásának dátuma: 2024. 11. 03.

Kivonat

Az illóolajok, mint különböző szerves anyagok keverékei, régóta fontos szerepet töltenek be a vegyipar finomkémiai ágazataiban (gyógyszer-, élelmiszer-, kozmetikai ipar). Gyártásuk számos összetevőjük miatt szintetikus úton nem gazdaságos, így a mai napig ipari szinten növényi alapanyagokból nyerik ki őket különböző fizikai-kémiai eljárásokkal. Ezen eljárások egyike az adott illóolajat tartalmazó növény részeinek szuperkritikus extrakciója. A szuperkritikus extrakciót az oldószer kiváló oldási képességei mellett számos további előnye (veszélytelen oldószer alkalmazása, könnyen változtatható oldási tulajdonságok, hőérzékeny komponensek kezelése, stb.) is kiemeli a hagyományos anyagelválasztási módszerek közül.

Munkám célja a közönséges levendulában (*Lavandula angustifolia*) található szuperkritikus extrakció révén kinyert illóolaj összetételének megismerése. Kísérleti eredmények felhasználásával vizsgálom az extrakció során kinyert illóolaj összetétele és az extrakció üzemeltetési paraméterei közti összefüggést regressziós modellek segítségével. Munkám során lineáris regressziós és LASSO regressziós modellt alkalmazok az illóolaj összetétel és kihozatal becslésére, keresztvalidáció segítségével pedig a felállított modellek predikciós képességét vizsgálom.

Az illesztett modellek paramétereit felhasználva rangsorolom az egyes extrakciós műveleti paraméterek és a növényt érő környezeti hatások (átlagos hőmérséklet, csapadékmennyiség a virágzás előtt) relatív súlyát a kinyert illóolaj tulajdonságaira gyakorolt befolyásuk szempontjából. A kapott összefüggéseket felhasználva optimális üzemviteli paramétereket lehet választani a művelethez elvárt mennyiségű és összetételű illóolaj kinyerésére.

Kulcsszavak: adatelemzés, regresszió, szuperkritikus extrakció, *Lavandula angustifolia*

Tartalomjegyzék

Ábrák jegyzéke	1
Rövidítésjegyzék	2
Jelölésjegyzék	4
1. Bevezetés	6
2. Illóolajok jellemzői és kinyerési módszereik	8
2.1. A levendulaolaj jellemzői	8
2.2. Illóolaj kinyerési módszerek	12
3. Illóolajok összetételének elemzése gépi tanulási módszerekkel	15
3.1. Az alkalmazott adatértékelési technikák elméleti áttekintése	17
3.1.1. Pearson korreláció analízis	18
3.1.2. Lineáris regresszió és LMG prediktor fontossági index	18
3.1.3. LASSO regresszió a prediktor változók fontosságának értékelésére	21
4. Kísérleti munka, eredmények feldolgozása	26
4.1. Eredmények elemzése	28
4.2. A kihozatalt és összetételt befolyásoló paraméterek rangsorolása	37
5. Összefoglalás	47

Ábrák jegyzéke

1.	Franciaország levendulaolaj termelése 2012-2018 [1].	11
2.	A modell hibájának változása a komplexitás függvényében [2].	22
3.	Két dimenziós paraméterter (β_1, β_2), az OLS becslése (fekete szaggatott vonal, β_{OLS}), az L_1 normából eredő korlátozás (piros négyzet) [3].	24
4.	Az adatkészlet ötszörös felbontása tanuló és tesztelő adathalmazokra. . .	25
5.	Az általunk használt szuperkritikus extrakciós berendezés sematikus ábrája.	27
6.	A kihozatali, illetve az egyes illóolaj alkotó komponensek eloszlása a nyomás (felül) és extrakciós hőmérséklet (alul) szerint.	29
7.	A kihozatali, illetve az egyes illóolaj alkotó komponensek eloszlása a virágzás alatti átlagos hőmérséklet (felül) és átlagos csapadék (alul) szerint.	31
8.	A kihozatali, illetve az egyes illóolaj alkotó komponensek eloszlása az extrahált növények származása (felül) és állási idő (alul) szerint.	33
9.	Az egyes változók korrelációs mátrixának szintértéke.	35
10.	Lineáris regressziós modell predikciós képessége a keresztvalidáció egyik becslő modelle esetében.	38
11.	A predikciós paraméterek LMG indexeinek szintértéke az egyes predikált paraméterek esetén.	39
12.	A predikciós paraméterek standardizált együtthatóinak változása a LASSO koefficiens változásának függvényében.	41
13.	LASSO regressziós modell predikciós képessége a keresztvalidáció egyik becslő modelle esetében a 7. táblázatban feltüntetett LASSO koefficiens értékekkel.	44
14.	LASSO regressziós modell által számított súlyozó tényezők abszolút értékének szintértéke az egyes predikált paraméterek esetén.	46

Rövidítésjegyzék

CNN	Convolutional Neural Network (Konvolúciós neurális hálózat)
CV	Cross-validation (keresztvalidáció)
EMA	European Agency for the Evaluation of Medicinal Products (Európai Gyógyszerügynökség)
EPA	Environmental Protection Agency (Környezetvédelmi Hivatal)
FANN	Fast Artificial Neural Network (Gyors mesterséges neurális hálózatok)
FDA	Food and Drug Administration (az USA Élelmiszerbiztonsági és Gyógyszerészeti Hivatala)
GC	Gas Chromatography (Gázkromatográfia)
GC-FID	Gas Chromatography Flame Ionization Detector (Gázkromatográfiás lángionizációs detektor)
GC-MS	Gas Chromatography-Mass Spectrometry (Gázkromatográfiás tömegspektrometria)
INCI	International Nomenclature of Cosmetic Ingredients (Kozmetikai alapanyagok nemzetközi elnevezése)
ISO	International Organization for Standardization (Nemzetközi Szabványügyi Szervezet)
LASSO	Least absolute shrinkage and selection operator (Legkisebb abszolút zsugorodás és szelekciós operátor)
LC	Liquid Chromatography (Folyadékkromatográfia)

LMG	Lindeman, Merenda and Gold módszer (más néven dominancia analízis (dominance analysis) vagy Sapeley-érték regresszió (Shapeley value regression))
LR	Linear Regression (Lineáris regresszió)
MLR	Multiple Linear Regression (Többszörös lineáris regresszió)
MNN	Multiple Neural Network (Többsztályos Neurális Hálózat)
NN	Neural Network (Neurális hálózat)
OLS	Ordinary least squares (Legkisebb négyzetek módszere)
QSRR	Quantitative structure retention relationship (Kvantitatív szerkezet retenciók összefüggések)
RF	Random Forest (Véletlen erdő)
SFE	Supercritical Fluid Extraction (Szuperkritikus fluid extrakció)

Jelölésjegyzék

β	Funkcionális operátor
β_{OLS}	OLS által számított funkcionális operátor
$\hat{\beta}$	LASSO regresszió által becsült koefficiensvektor
$\ \beta\ _1$	β koefficiensvektor L_1 normája
λ	LASSO koefficiens
μ	Átlag
ν	Adott struktúrában azon prediktorok halmaza, melyben nem szerepel a vizsgált prediktor (LMG vizsgálat)
σ	Szórás
π	Adott struktúrában az összes prediktor változó halmaza (LMG vizsgálat)
A_i	A adatsor i -edik eleme
d	Predikciós változók száma
LMG_j	j -edik LMG index
N	Minták száma
p	Nyomás [bar]
P^K	A K halmaz összes permutációja
R	Korreláció
R^2	Determinációs együttható

$R^2_{y(X_K)}$	X_K prediktor mátrix esetén becsült y predikált értékekhez tartozó R^2 érték a vizsgált tanuló adatsoron
$seqR^2_{y(X_j X_K)}$	A j prediktor kiegészítő R^2 értéke $\{j\} \in K$ vizsgált prediktorhalmaz esetén
T	Hőmérséklet [$^{\circ}\text{C}$]
u	Standardizált érték [-]
x_m	Tömegszázalék [m/m%]
x_V	Térfogatszázalék [V/V%]
X'	X mátrix transzponáltja
X^{-1}	X mátrix inverze
X_K	A K halmaz prediktor változóiból képzett prediktor mátrix
y	Predikált változó értékeinek vektora

1. Bevezetés

Szakmai munkám során közönséges levendula (*Lavandula angustifolia*) növényből szuperkritikus extrakcióval kinyert illóolajok olaj hozamát és az illóolaj összetételét vizsgálom kísérleti technikákkal és modellezési módszerekkel.

Az elmúlt években egyre több figyelmet szentelnek az emberek az egészséges életmódnak. A mindennapjaink részévé vált a különféle étrend-kiegészítők, gyógykészítmények megelőző célú használata. Fontos lett az ételmezésben, a gyógyászatban mind pedig a szépségápolásban a természetes anyagok használata, valamint a vegyszermentesség, ezért megnőtt a különböző iparágak természetes hatóanyagok iránti igénye [4].

Az illóolajok az illatos növények szöveteiben felhalmozódott hatóanyagok. A növény szinte bármely részében előfordulhatnak. Összetételüket, képződött mennyiségüket a környezeti tényezők (időjárás, talajminőség) befolyásolják. Fizikai tulajdonságokat tekintve légköri körülmények között általában folyékonyak, illékonyak, jellegzetes illatuk van. Apoláris oldószerekben jól oldódnak, vízzel nem elegyednek, (de előfordul, hogy bizonyos komponenseik – mint az oxigéntartalmúak – korlátlanul elegyednek vízzel). Apolárosságukból adódóan lehetséges a bőrön és légutakon keresztüli abszorpció. Sokszínű összetételükre jellemző 5-10 makro- és akár több száz minorkomponens, melyek elhanyagolhatatlanok a sajátos tulajdonságaik kialakításában. Az illóolajok a növény életében nagyon fontos védő szerepet töltenek be fertőzésekkel, penészgombákkal szemben. Régóta ismert, tartósító tulajdonságuk e jellemzővel magyarázható [5, 6].

Az illóolaj összetevőket azonban csak nagyon költségesen tudja előállítani a vegyipar. Egyrészt a komponensek sokasága, másrészt a szerves vegyületek számos sztereoizomere miatt. Gazdasági okokból a szintetikus előállítás helyett inkább a növényből való kinyerés mellett döntenek a vállalatok [4].

Ezen kinyerési módszerek közül is kiemelkedik a szén-dioxid alapú szuperkritikus extrakció számos előnyös tulajdonságának (oldószer újrahasznosíthatósága, egyszerű elválasztás, enyhe hőmérsékletviszonyok) köszönhetően.

A munka során célom olyan modellek alkotása mérési eredmények alapján, melyek-

kel előre lehet becsülni *Lavandula angustifolia*, azaz közönséges levendula növényből szuperkritikus extrakcióval kinyert illóolaj kihozatalát és az olaj összetételt tetszőleges extrakciós körülmények és más, a növény felhasználása előtti időtartamot jellemző paraméterek mellett. Dolgozatomban többek között azt is vizsgálom, hogy az olajösszetételt az egyes, kinyerés előtti vagy műveleti paraméterek milyen mértékben befolyásolják. Az eredményeket alkalmazva felhasználói igényeknek megfelelő mennyiségű és minőségű olaj előállítása lehetséges a műveleti paraméterek finomhangolásával.

Munkám során bemutatom az illóolajok általános tulajdonságait, illetve kinyerésük módját (2. fejezet). Utána ismertetem az illóolajok összetételének becslésére szolgáló matematikai módszereket (3. fejezet), majd az általam kidolgozott modelleket, valamint azok becslési pontosságát mutatom be (4. fejezet).

2. Illóolajok jellemzői és kinyerési módszereik

Általános esetben a tiszta illóolajokban több, mint 200 komponens található. Ezek a keverékek fenil-propán származékokat vagy terpéneket tartalmaznak. Két osztályba sorolhatóak: az illékony frakció a teljes olajtömeg 90-95%-át teszi ki. Tartalmaz mono-terpéneket, szeszkviterpéneket és azok oxigéntartalmú származékait, alifás alkoholokat, észtereket és aldehideket. A nem illékony maradék a tömeg 5-10%-a, mely zsírsavakból, szénhidrogénekből, szterolokból, viaszokból, flavonoidokból, illetve karotinoidokból állhat. Az illóolajokban található szénhidrogének közül a legjelentősebb az izoprén. A 1. táblázatban a jellemzően illóolajalkotó vegyületcsoportokat, azok élettani hatását, valamint példákat láthatunk az adott csoportból [5].

Vegyület-csoport	Élettani hatás	Példa
terpének	gyulladásgátló, baktericid, vírusölő	pinén, limonén, kamfén, piperin
alkoholok	antiszeptikus, vírusölő, baktericid	linalool (levendula, ilang-ilang), nerol (neroli), geraniol (rózsa, muskátli)
aldehidek	gyulladáscsökkentő, gombaellenes, fertőtlenítő, baktericid, vírusölő, nyugtató	citral (citrom), citronellal (citrusos eukaliptusz, citromfű)
savak	gyulladásgátló	benzoesav, fahéjsav (benzoin)
észterek	nyugtató, kiegyensúlyozó, antimikrobiális, gombaölő	linalil-acetát (levendula), bergamott és geranil-formiát (muskátli)
ketonok	gyógyító, hurutosodás ellen, köptető jelleg, elősegítik a sebgyógyulást, hegképződést, általában mérgezőek	legmérgezőbb: tujon (zsálya, üröm, tuja) nem mérgező: fenchone (édeskömény), jázmon (jázmin), menton (borsmenta), karvon (fodormenta)

1. táblázat. Az illóolajalkotó vegyületcsoportok, azok emberre gyakorolt hatásai, példák [5].

2.1. A levendulaolaj jellemzői

A levendula, mint gyógynövény, számos országban elterjedt, rengeteg faja és alfaja ismert. Összesen 32 *Lavandula* fajt sorolnak a *Lavandula* nemzetség tagjai közé, vala-

mint számos infraspecifikus taxont és hibridet. Őshonos a Kanári-szigeteken, Madeirán, a Földközi-tenger medencéjében (Franciaország, Spanyolország, Andorra, Olaszország), Észak-Afrikában, Délnyugat-Ázsiában, az Arab-félszigeten, valamint Indiában, e területekről terjedt el [7, 8].

Az összes szokásos kereskedelmi forgalomban kapható növényfajta két fő alcsaládba tartozik: *Stoechas* (*Lavandula stoechas*, *Lavandula dentata*, *Lavandula viridis* és *Lavandula pedunculata*) és *Spica* (*Lavandula angustifolia* (*Lavandula officinalis*), *Lavandula latifolia* és *Lavandula lanata*). A legtöbb valószínűleg a *Lavandula angustifolia* és a *Lavandula latifolia* hibridje [7].

A *Lavandula* nemzetség viszonylag nagy számban tartalmaz fenolos összetevőket. Tagjai 19 flavont és 8 antocianint tartalmaznak. A családra jellemzőek a hipolaetin és szkutellarein különböző glikozidjai. A levélflavonoidok főként flavon-glikozidok, melyek egyéni eloszlása a taxonok között jelentős taxonómiai jelentőséggel bír [7].

A kereskedelmi forgalomban kapható hibridek, a lavandinok változó koncentrációjú 1,8-cineolt és kámfort tartalmaznak. Ezek adják a keményebb illatjegyeket az illóolajokban. A „rhodinol-tartalom” (komponensei: citronellol, geraniol, nerol, neril-acetát és geranil-acetát) bár a teljes összetétel nagyon kis hányadát teszi ki, édes, rózsaszerű illatot kölcsönöz a levendulaolajoknak [7].

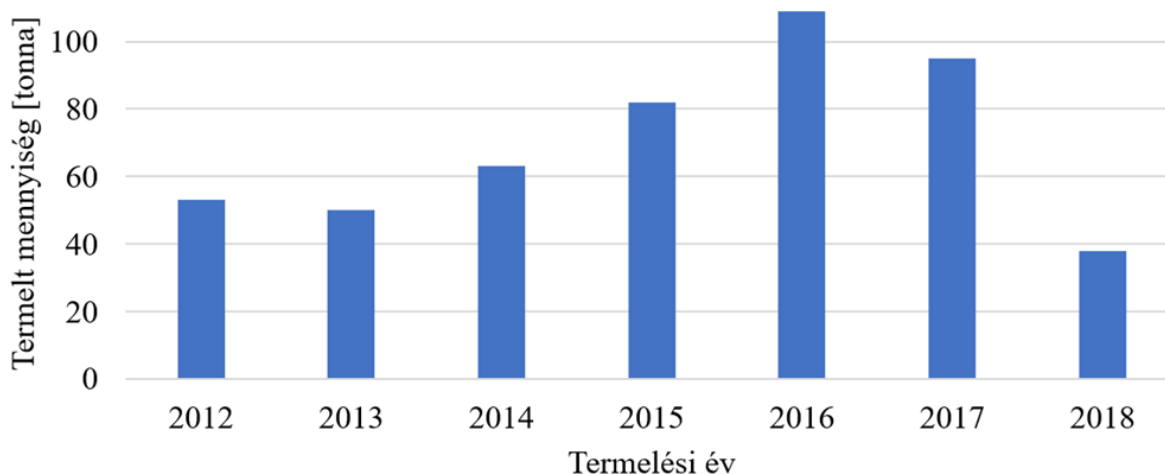
A kereskedelembe kapható *Lavandula angustifolia*, mely olájának fő komponensei a linalool (25-38%) és a linalil-acetát (25-45%), jelentős eltéréseket mutat alfajai között. A *Lavandula pyrenaica*, (vadon nő ÉK-Spanyolországban), illóolájának három lényeges komponense a linalool (20-66%), borneol (6-32%) és kámfor (2-14%). A linalool tartalma magas, azonban a kámfor tartalma elfogadhatatlanná teszi normál levendulaolajnak. A *Lavandula lanata*-nak nagyon magas kámforkoncentrációja (43-59%) és változó mennyiségű lavandulol (3-27%) mennyisége van. A *Lavandula dentata*-nak (vadon nő Spanyolorzágnak a Földközi-tenger partja menti részén), két kemotípusa van: 1,8-cineol/ β -pinén és β -pinén/ α -pinén. A *Lavandula stoechas* négy vadon élő populációját áttekintve, különböző százalékban tartalmaz α -pinént, 1,8-cineolt, fenchont, kámfort és mirtenil-acetátot. A *Lavandula lusieri* szintén két kemotípussal rendelkezik, amelyek fő összetevője egy azonosítatlan észter. A *Lavandula viridis*ben magas az 1,8-cineol, a kámfor és az α -pinén koncentrációja. A *Lavandula pinnata* (Madeirán őshonos) nagy

százalékban tartalmaz monoterpéneket (37-80%) és viszonylag kis arányban szeszkviterpéneket (13-22%) [7].

A levendulaolaj halványsárgás folyadék, amelynek illata friss, virágostól akár fás és aromás illatú is lehet. A levendulaolaj (levendula illóolaj) a levendula, a *Lavandula angustifolia* virágaiból nyert illóolaj. Az európai INCI (International Nomenclature of Cosmetic Ingredients – Kozmetikai alapanyagok nemzetközi elnevetése) neve *Lavandula angustifolia* olaj, az amerikai INCI neve pedig *Lavandula angustifolia* (levendula) olaj (CAS-számok 8000-28-0, 90063-37-9; EK-szám 289-995-2) [7].

A hagyományos gyógynövénygyógyászatban a levendulaolaj nyugtató, antidepresszáns, gombaellenes, altató, fájdalomcsillapító, görcsoldó, tonizáló, stresszoldó és afrodiziákum tulajdonságokkal rendelkezik [7].

A közelmúltban a levendulaolaj elsődleges forrásai: Franciaország, Bulgária, Kína és Spanyolország. A legtöbb levendulanövényt eredetileg a mediterrán Franciaország magasabb területein (600-1500 m) termesztették és desztillálták. Az olaj előállítására vonatkozó pontos adatokat nehéz megállapítani a hatalmas mennyiségű hamisítás, keverés, darabolás és szintetikus anyagok vagy szintetikus levendulaolaj hozzáadása miatt. 1984-ben a világ levendulaolaj-termelése 200 tonna volt. A fő termelőket tekintve Bulgária 100-129, Franciaország 55, Szovjetunió 35, Ausztrália 5 tonnát termelt, a maradék kisebb termelőktől érkezett. Később az USA importja évi 303 és 555 tonna között változott, exportja 52 és 121 tonna között mozgott. Több mint 30 különböző típusú levendulaolaj és keverék található meg a világpiacon [7].



1. ábra. Franciaország levendulaolaj termelése 2012-2018 [1].

Az 1. ábra Franciaország levendulaolaj termelését mutatja 2012 és 2018 között, évenként, tonnában mérve. 2016 után csökkenő tendenciát mutatott a termelés [1].

Az ISO (International Organization for Standardization) szabványokat ír elő, amelyeknek a kereskedelmi levendula olajoknak és azok előállítási módjainak meg kell felelniük. A fő komponensek minimális és maximális megengedett koncentrációja a 2. táblázatban látható (ISO 3515:2001). A legnagyobb megengedett koncentrációjú komponensek a linalil-acetát (45%), linalool (38%), β -cisz-ocimén (10,0%), β -transz-ocimén (6,0%) [7] [9].

Komponens	CAS	Minimum [(m/m)%]	Maximum [(m/m)%]
linalil-acetát	115-95-7	25,0	45,0
linalool	78-70-6-2	25,0	38,0
β -cisz-ocimén	3338-55-4	4,0	10,0
β -transz-ocimén	3779-61-1	2,0	6,0
3-oktanon	106-68-3	0,0	2,0
1,8-cineol	470-82-6	0,0	1,5
terpineol	98-55-5	0,0	1,0
β -fellandréen	555-10-2	0,0	1,0
kámfor	76-22-2	0,0	0,5
limonén	138-86-3	0,0	0,5
lavandulil-acetát	25905-14-0	2,0	-
terpinen-4-ol	562-74-3	2,0	-
lavandulol	498-16-8	0,3	-

2. táblázat. A levendulaolaj ISO szerinti paraméterei [9].

A növényi eredetű készítmények általánossá válásával párhuzamosan szigorodnak ezen termékek előállítását szabályozó minőségi követelmények. A hatósági szabályozás a környezetbarát technológiák bevezetésére, valamint az emberekre és környezetükre kevésbé káros termékek előállítására ösztönzi a gyártókat. Az FDA (Food and Drug Administration), az EMEA (European Agency for the Evaluation of Medicinal Products) és az EPA (Environmental Protection Agency) által meghatározott mértékadó előírások a szerves oldószert nem alkalmazó technológiák elterjedését támogatják [4].

Ezen szabályozások következtében nem mindegy, hogy ezeket az illó (és sokszor hőérzékeny) komponenseket milyen módon nyerjük ki a növényből. A következő, 2.2. alfejezetben a leggyakoribb illóolajkinyerési módszerek főbb jellemzőit, előnyeit és hátrányait tárgyalom.

2.2. Illóolaj kinyerési módszerek

Az illóolajokat különböző módszerek révén nyerhetjük ki. A hagyományos kinyerési módszerek pl. vízgőz-desztilláció, szerves oldószeres (pl. Soxhlet) extrakció, préselés mellett sajátos tulajdonságokkal bír a szuperkritikus extrakció a különféle növényi hatóanyagok kinyerésében [10].

Vízgőz-desztilláció során a vízgőzt meghatározott nyomáson és hőmérsékleten átvezetik a növényi részeken. A vízgőz az illékony komponensekkel együtt távozik, majd hűtés hatására a két fázis elkülönül. Hátránya a magas hőmérséklet, mely hatására a hőérzékeny vegyületek elbomolhatnak [5].

A vízdesztilláció folyamatában hasonló a vízgőz-desztillációhoz. A növényt vízben melegítjük (főzzük), majd az illóolajos párákat a vízgőz-desztillációhoz hasonlóan összegyűjtjük. Hátránya egyértelműen az előbb említett vízgőz-desztillációnál is felvetett, magas hőmérséklet miatti bomlékonyság [5].

Az előbb említett két desztillációhoz hasonló a vákuum-desztilláció, melyet a kényes növényi összetevők esetében alkalmaznak. A vákuum miatt az alkalmazott oldószer (víz) forráspontja lecsökken, így az alacsonyabb hőmérsékletű vízgőz alkalmazása révén nem

áll fenn a hőérzékeny molekulák bomlásának veszélye. A vákuum előállításának nagy energiaigénye, ezáltal magas költségei miatt kevés helyen használják [5].

Az előzőektől lényegesen eltérő folyamat az oldószeres extrakció. A növényi anyagot összekeverik a szerves oldószerrel, míg a komponensek oldódnak, majd a növény maradékát (melyből kioldódott már az értékes illóolaj) eltávolítják. Az oldott anyag ilyenkor az oldószer lepárlásának a maradéka. Hátránya, hogy az oldószer az illóolajon kívül más növényi összetevőket is képes oldani, ezért plusz komponenseket is tartalmaz a termék, valamint a lepárlás során az extraktum tartalmazhat oldószermaradékot [5].

Hőérzékeny vegyületeket tartalmazó növényekből jellemzően sajtolással állítanak elő illóolajat, melyet a növény héjából préseléssel és centrifugálással nyernek ki [5].

Enfleurage (pomádézás) során egy speciálisan kikészített és dezodorált zsírrétegre helyezik a növényi részeket (jellemzően virágokat), ahol az illékony vegyületek képesek beoldódni. Ezt a pomádénak nevezett telített zsírt alkohollal extrahálják, majd eltávolítják az oldószeret. A termék szintén tartalmazhat oldószermaradékot [5].

A szuperkritikus fluidum extrakció (Supercritical Fluid Extraction vagy SFE) egyre több helyen váltja fel a hagyományosnak tekinthető, szerves oldószeres extrakciót. A szuperkritikus extrakció során a legtöbb esetben oldószerként CO_2 -ot használnak, mivel alacsony a kritikus hőmérséklete (31°C) és nyomása (73 bar), szagtalan, nem mérgező, nem gyúlékony, olcsón, könnyen hozzáférhető [4, 10].

A CO_2 csak az apoláris komponenseket oldja nagymértékben, ezért gyakran alkalmaznak mellette keverék vagy kiegészítő oldószeret (koszolvenst), például etil-alkoholt, mely néhány százalék koncentrációban is jelentősen megnövelhetik a poláros komponensek oldékonyságát [11].

Egyik legjelentősebb előnye, hogy az oldószeret az extrakció végén maradéktalanul visszakapjuk, mivel standard körülmények között gázként távozik az elegyből, így az nyomokban sem tartalmaz szerves oldószeret, önmaga is lehet késztermék vagy alapanyag. Az SFE szintén kiemelten fontos előnye, hogy az alkalmazott hőmérséklet relatíve alacsony, a hőérzékeny hatóanyagok természetes előfordulásukban, szinte hőkárosodás nélkül kinyerhetőek, valamint olyan hatóanyagok is extrahálhatók, melyeket alapvetően hagyományos oldószeres extrakcióval nem, vagy csak kis mennyiségben lehetne kinyerni [4].

Az SFE során a működési paraméterek (pl. hőmérséklet és nyomás) megfelelő változtatásával, azaz optimalásával lehetőségünk van a szuperkritikus oldószer oldóképességének változtatására, tehát szelektív extrakció valósítható meg [12].

Ezen tulajdonság következtében a műveleti paraméterekkel hangolni lehet az extrakció során kinyert termék összetételét és hozamát. A műveleti paraméterek és a kinyert illóolaj közötti kapcsolatot matematikai modellek segítségével tárom fel. A témakörben gyakran adat alapú modelleket használnak ezen kapcsolatok elemzésére, melyeket a következő, 3. fejezetben mutatom be.

3. Illóolajok összetételének elemzése gépi tanulási módszerekkel

A mesterséges intelligencia az emberi képességeket utánzó gépek gyűjtőfogalma. A gépi tanulás ennek egy speciális területe, ahol a gépeket arra „tanítják meg”, hogy feltárják az adatokban a rejtett információkat. A gépi tanulás célja a hatékony adatgyűjtésen kívül az, hogy komolyabb emberi figyelem ráfordítása nélkül az egyre növekvő mennyiségű adatot felhasználja azok kezelésével és elemzésével [13].

Ezen módszereket számos helyen alkalmazzák, mint például a képfelismerés, beszéd-felismerés, kézírás-felismerés, orvosi diagnosztika, forgalom előrejelzés, termékajánlások, virtuális személyi asszisztensek, online csalásészlelés területeken [13].

Az illóolajok esetében egyrészt magából az illóolajösszetételből több dolog is becsülhető: több kutatás foglalkozik az illóolajtermelő növények biológiai aktivitásának és antimikrobiális hatásának a becslésével. Például Noha és társai tanulmányában a mély tanulási technikák közül a konvolúciós neurális hálózatokat (CNN) alkalmazták illóolajtermelő növények biológiai aktivitásának becslésére az illóolajuk kémiai összetételén keresztül. Többosztályos neurális hálózatot (MNN) alkalmaztak a CNN-modell teljesítményének értékelésére. Ezt az irányvonalat folytatva Jose és társai gyors mesterséges neurális hálózatok (FANN) segítségével vizsgálták az adott illóolaj antimikrobiális hatásának előrejelezhetőségét az illóolaj összetételéből. Vulic és társai *Helichrysum Italicum* illóolajának kihozatalát, antimikrobiális aktivitását és biológiai aktivitását vizsgálták és próbálták megbecsülni többek között ANN segítségével [14–16].

Más kutatások eredményei azt bizonyítják, hogy magát az illóolajösszetételt és az illóolajok kihozatalát is lehet becsülni. Pezo és társai mesterséges neurális hálózat (ANN) modell segítségével becsülték az ánizs estében a terméshozamot és az illóolajtartalmat az időjárási viszonyok függvényében. A tanulmány eredményei azt mutatták, hogy a trágyázás típusa a mezőgazdasági jellemzőkre (mint például 1000 mag tömege stb.), míg a termőhely az illóolaj-tartalomra volt nagyobb hatással [17].

Áćimović és társai a *Hyssopus Officinalis* kémiai összetételét vizsgálták. A vízgőz-desztillációval nyert illóolajak összetételét GC-FID és GC-MS módszerekkel elemezték. Az összetétel adatokat klaszteranalízis és korrelációs elemzés segítségével vizsgálták. Kvantitatív szerkezet-retenciós összefüggéseket (QSRR) kerestek az azonosított vegyületek visszatartási indexeinek becslésére [18].

Niazian és társai az ajowa illóolajtartalmát mesterséges neurális hálózattal (ANN) és a többszörös regressziós modellel (MLR) becsülték a könnyen mérhető növényi jellemzőkből (mint a virágok sugara, egy száron található virágok száma stb.) [19].

Rashid és társai által írt tudományos cikk körültekintő áttekintést nyújt a gépi tanulási algoritmusok (többek között regressziós technikák és neurális hálók) használatáról a terméshozam előrejelzésére, különös tekintettel a pálmaolaj esetében. A pálmaolaj igaz nem illóolaj, de mivel ugyan úgy növény, ezért a közvetlen (például környezeti) hatások, mint bemeneti paraméterek jó közelítéssel párhuzamba hozhatóak az illóolaj-tartalmú növények esetével [20].

A hasonló témában megjelent tudományos publikációk között a leggyakrabban a predikciós adathalmaz részét nem a műveleti paraméterek, hanem az időjárási viszonyok képzik, az illóolaj kinyerése rögzített műveleti paraméterekkel történik.

Chen és társai szintén nem illóolajakat, hanem bioolajat vizsgáltak. A kutatásuk célja az volt, hogy különböző gépi tanulási módszerekkel becsüljék meg a bioolaj (H-bioolaj) hozamát és hidrogéntartalmát a biomassza, mint alapanyag összetétele és a pirolízis körülményei függvényében. A módszereket tekintve többszörös lineáris regressziót (MLR) és a véletlen erdőt (RF) alkalmazták [21].

A kutatási témám során Pezo és társai kutatásához hasonlóan figyelembe veszem a szüretelés előtti időjárási viszonyokat, illetve Chen és társai kutatásához hasonlóan a műveleti paraméterek változtatásával, azok függvényében vizsgálom az illóolajösszetételt és kihozataalt. Ezzel olyan körülményeket keresek, melyek mellett kinyerhető a kívánt összetételű (és mennyiségű) illóolaj.

Amint az előzőekben hivatkozott tudományos cikkekben látható, hogy a neurális hálók (NN) dominálják az illóolaj összetétel előrejelzésével kapcsolatos kutatásokat. Ezek a módszerek bonyolultak, aktivációs függvényeket, háló topológiát kell tervezni (neuronok/rétegek száma, stb.) hozzájuk. A kutatási témám során azt vizsgálom, hogy a

kevésbé összetett regressziós modellek mennyire megfelelően képesek becsülni az illóolaj-kihozatalt és összetételt.

A bemeneti paramétereket tekintve elsősorban a szuperkritikus extrakció során a műveleti paramétereket, tehát a nyomást és hőmérsékletet változtatva vizsgáltam a kihozatalt és összetételt. A műveleti paramétereken kívül meghatározó szerepet töltenek be a növény felhasználása előtti időtartamot jellemző paraméterek, mint a szüretelés és extrakció között eltelt idő, illetve az adott szezonban, a növényt ért környezeti hatások, mint a napsütéses órák száma, illetve a csapadékmennyiség. Ezeken kívül még a bemeneti paraméterek közé felvettem a növény származásának helyét. Azt, hogy ezen paraméterek hogyan befolyásolják a becsült változókat, korrelációanalízissel vizsgáltam. Az illóolajok összetételének becslésére regressziós modelleket állítottam fel, mely modellek becselő képességét megvizsgáltam, illetve a becselő változók fontosságának mértékét több módon vizsgáltam.

3.1. Az alkalmazott adatértékelési technikák elméleti áttekintése

A következőkben az illóolaj adatsor elemzésére alkalmazott statisztikai technikák elméleti alapját mutatom be. A munka során két céllal foglalkozom, az első regressziós modellek alkotása a levendula illóolaj összetételének és kihozatalának becsléséhez, az illóolaj kinyerés műveleti paraméterei, az állási idő, illetve a növény termesztési körülményeinek függvényében (származási hely, virágzás előtti átlag hőmérséklet és csapadékmennyiség). A második cél az egyes prediktor változók fontosságának rangsorolása és annak vizsgálata, hogy a kinyerés műveleti paraméterei, vagy a növény termesztésének és állásának körülményei határozzák-e meg jobban az illóolaj kihozatalt és összetételt. Az utóbbi cél eléréshez három technikát alkalmazok, az első az egyszerű korrelációanalízis, a második a lineáris regresszió és az LMG prediktor fontossági index alkalmazása, a harmadik a LASSO regresszió módszere.

3.1.1. Pearson korreláció analízis

A becsléshez alkalmazott regressziós módszerek alkalmazása előtt fontos az előzetes adatértékelés, melyhez a Pearson-korrelációt (R) használom mutatószámként. Ezzel az R korrelációs együttható értékével a változók közötti összefüggéseket vizsgálom, melyet az (1)-es összefüggés alapján számíthatunk [22].

$$R(A, B) = \frac{1}{(N-1)} \sum_{i=1}^N \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right) \quad (1)$$

Az $R(A, B)$ az A és B változó közötti korreláció, N a vizsgált minták darabszáma, A_i az i -edik mintát jellemző A változó értéke, μ az adott változó átlaga a vizsgált mintában, σ pedig az adott változó szórása az adatsorban. Amennyiben ez az érték 1, a két változó teljes mértékben lineáris kapcsolatban van, ugyanúgy változik az értékük, 0 és 1 között pozitívan korrelál egymással, azaz tendenciájában együtt nő és együtt csökken az értékük. 0 és -1 között negatívan korrelálnak egymással, ha az egyik változó nő, akkor a másik csökken. Ha 0, akkor a két változó függetlennek tekinthető. A korrelációs együttható ezáltal alkalmazható az adatsoron belüli összefüggések feltárásához. A korrelációs index előzetes képet adhat az egyes regressziós modellek felállítása során a fontos és kevésbé fontos prediktor változóról.

3.1.2. Lineáris regresszió és LMG prediktor fontossági index

Szakmai munkám során regressziós modelleket alkalmazok arra, hogy azt becsüljem, hogy az egyes bementi változók mekkora mértékben határozzák meg a kihozatali és összetétel értékeket, illetve hogy mennyire pontosan képesek becsülni azokat. A regresszió olyan statisztikai módszer, mely során az illesztett regressziós modell, képes előre jelezni, megjósolni bementi paraméterek alapján egy adott tulajdonság változását. Az ilyen modelleknél a bemeneti, becsülő változókat prediktor, a kimeneti, becsült változókat pedig predikált változóknak nevezzük [13]. A regressziós módszerek tanítása előtt szükségesnek tartottam az adatkészletet standardizálását (2), mivel egyes esetekben, akár két nagyságrend eltérés is mutatkozott két változó között. A standardizálás révén a

prediktor változók regressziós súlyai összehasonlíthatóvá válnak, ezek egymáshoz viszonyított súlya a regresszió során értékelhető, amennyiben a prediktor változók egymástól függetlenek.

A standardizált értékek jele: u [23].

$$u = \frac{A_i - \mu_A}{\sigma_A} \quad (2)$$

A regressziós modellek illeszkedését leggyakrabban az ún. R^2 együtthatóval szokás jellemezni. Az R^2 érték egyetlen számmal adja meg, hogy mennyivel teljesít jobban egy adott modell, mintha az átlaggal cserélnék ki a predikciót. A mérték a következőképpen számítható (3) [24]:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \mu_y)^2} \quad (3)$$

y a mért értékek, $\hat{y}$ a becsült érték, μ_y , az y értékek átlaga, N pedig a minták száma.

A munka során alkalmazott regressziós technikák közül első az egy kimenetű lineáris regresszió módszere. A modell mátrixalgebrai jelöléssel a következőképpen írható fel (4) [25]:

$$y = X\beta \quad (4)$$

$y \in R^{d \times 1}$ a becsült, predikált változó értéke, $X \in R^{N \times d}$ a predikciós változók mátrixa, $\beta \in R^{N \times 1}$ egy funkcionális vektor, N a minták száma, d a prediktor változók száma. A $\beta \in R^{d \times 1}$ funkcionális vektor a (5)-es összefüggés alapján határozhatjuk meg a legkisebb négyzetek módszerével (OLS - ordinary least squares) [26]:

$$\beta = (X'X)^{-1}X'y \quad (5)$$

β az a paraméter, amellyel megszorozva az egyes prediktor változókat, a legjobban illeszkednek a predikált adatpontok a mért értékekre (legközelebb helyezkednek el azokhoz) [27].

Mint korábban említésre került standardizált prediktor és predikált változók alkalmazásával az egyes prediktor változók súlyai tekinthetők relatív fontossági indexeknek,

azonban ez csak akkor lehet mérvadó, ha minden prediktor változó egymástól független. Amennyiben a prediktor változók lineárisan korreláltak, úgy a prediktor súlyok vizsgálata a relatív fontosság megállapításához félrevezető eredményeket adhat. A felállított lineáris regressziós modellben az egyes prediktor változók egymáshoz képesti relatív fontosságát Lindeman, Merenda and Gold (LMG) módszerrel értékelem, mivel a szakirodalomban is ezt alkalmazzák a leggyakrabban. Az LMG, más néven dominanciaanalízis (dominance analysis) vagy Shapley-érték regresszió (Shapley value regression) egy olyan módszer, mely átlagolja az ún. kiegészítő (additional) R^2 -összegeket a (6)-as egyenletben látható módon, mellyel megkaphatjuk az egyes LMG indexeket. Ezek az indexek megmutatják egy-egy prediktor változó relatív fontosságát a többi lehetséges prediktor változóhoz képest egy regressziós modell felállításakor. A kiegészítő R^2 az az érték, mely egy változó vizsgálata esetében megadja, hogy az adott regressziós modell illeszkedése mennyivel jobb a tanító adatokra, ha a vizsgált változó szerepel a prediktor változók között ahhoz képest, mintha nem szerepelne [28].

$$LMG_j = \frac{1}{d!} \sum_{\pi \in P^{\{1, \dots, d\}}} seqR_{y(X_j|X_\pi)}^2 \quad (6)$$

Ahol LMG_j a j -edik lehetséges prediktor változó LMG indexe, d a prediktor változók száma, $P^{\{1, \dots, d\}}$ az $\{1, \dots, d\}$ halmaz összes permutációja, $seqR_{y(X_j|X_\pi)}^2$ pedig a (7)-as összefüggés szerint számítható kiegészítő R^2 .

$$seqR_{y(X_j|X_\pi)}^2 = R_{y(X_\nu \cup \{j\})}^2 - R_{y(X_\nu)}^2 \quad (7)$$

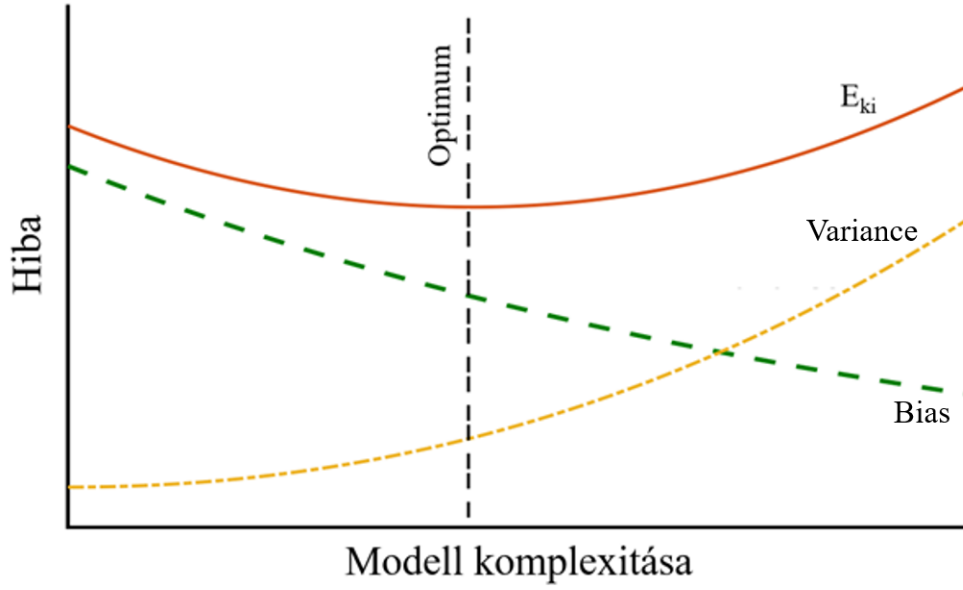
A kiegészítő R^2 értékének számításakor végig haladunk a permutációkon, melyek megmutatják az összes lehetséges regressziós modell struktúrát. A π -ik struktúrát ezután két halmazra bontjuk. Az egyik maga a $\pi = \nu \cup \{j\}$, melyben az összes prediktor változó benne van, melyet a π tartalmaz, a másik pedig $\nu = \pi \setminus \{j\}$, melyben nem szerepel a vizsgált j -edik prediktor változó. A prediktor adatok mátrixát, melyet ν -ból képzünk X_ν -vel jelöljük. Ezután definiáljuk az $R_{y(X_\nu)}^2$ változót, mely adott X_ν prediktor mátrix esetén becsült y predikált értékekhez tartozó R^2 érték a vizsgált tanuló adatsoron. A kiegészítő R^2 érték megadja, hogy mekkora a különbség $R_{y(X_\nu)}^2$ és az aktuálisan vizsgált prediktor változóval kiegészített modell $R_{y(X_\nu \cup j)}^2$ értéke közt. Az LMG értékeket az

összes lehetséges prediktor változóra kiszámítva, majd ezeket az LMG értékek összegével normalva megkapjuk az egyes prediktor változók relatív fontosságát.

3.1.3. LASSO regresszió a prediktor változók fontosságának értékelésére

Egy statisztikai modell teljesítményének vizsgálatakor figyelembe kell venni azt, hogy az (esetünkben a mért) adatoknak van egy természetes zaja. Egy statisztikai modell felállításának folyamata két nagy részre, a tanulásra és tesztelésre bontható. A tesztelés során a dolgozatomban a modell predikciós képességét mérem, melyet az R^2 együtthatóval jellemzek. Amikor fejlesztünk egy modellt, akkor fennáll annak a lehetősége, hogy a zajra is illesztünk, amely jelenség nem szerencsés, mivel így más, új adatokon (tesztelő adatokon) a felállított modell gyengén teljesít. Amennyiben növeljük a predikciós változók számát (és ezáltal a modell illeszkedését), a modell predikciós képessége csökkenhet pontosan a zajra illesztés jelensége miatt. A statisztikai modellek fejlesztésekor ezért fontos optimális számú prediktor változót kiválasztani, ez angolul a bias-variance trade-off problémaként ismert. Ha a predikciós változók száma nem elegendő úgy a modell alul illeszt, a modell illeszkedése a tanuló adatsorra rossz lesz (magas bias, alacsony variancia). Ellentétes esetben, ha túl sok prediktor változó van jelen a modell ugyan jól illeszkedik a tanuló adatsorra, viszont a teszt adatokra nagy hibával illeszt (magas variancia, alacsony bias), ez a túlillesztés jelensége [29].

Ezen jelenség a 2. ábrán szemléletesebben látható, ahol a modell teljes hibájának (E_{ki} , azaz a tanuló és tesztelő adatokra illeszkedés hibájának összessége) változása a modell komplexitásának függvényében lett ábrázolva. A modell komplexitása, azaz összetettsége alatt a bemeneti, prediktor változók számát értjük a lineáris regressziós modellek esetében [2, 30].



2. ábra. A modell hibájának változása a komplexitás függvényében [2].

A 2. ábrán látható, hogy a teljes hibának, a modell komplexitásának függvényében van egy minimuma, ez az optimum modell struktúrát mutatja meg. Az ezen pont előtti tartományban alulilleszkedést mutatnak azok a modellek, melyek túl kevés predikciós változót tartalmaznak, ami miatt nem képesek megfelelően leírni az adatokban fellelhető trendet [30].

Az optimum megtalálása (és ezáltal változó szelekció) érdekében a következő vizsgált regressziós technika a LASSO (least absolute shrinkage and selection operator) regresszió volt, amely torzított becsléssel képes szimultán módon becsülni adott paramétert és változószelekciót végrehajtani [31, 32].

A kutatási munkám során ezen tulajdonságai a módszernek azért használhatóak nagy hatékonysággal, mivel a változó szelekció képes egyfajta módon rangsorolni a predikciós változókat. A regresszió a becslés során minimalizálja a maradék négyzetek összegét, mely művelet a (8)-es összefüggéssel definiálható [33]:

$$\hat{\beta} = \operatorname{argmin}_{\beta} (y - X\beta)^T (y - X\beta) + \lambda \|\beta\|_1 \quad (8)$$

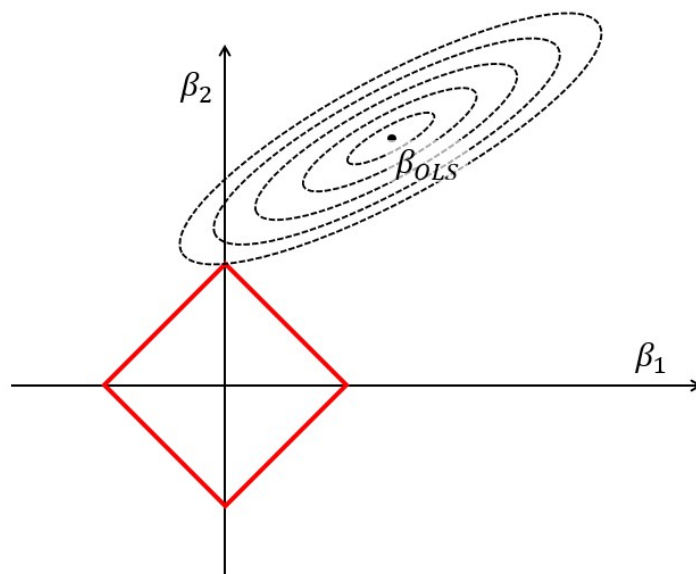
Ahol $\hat{\beta}$ a LASSO regresszió által becsült együttható vektor, y az aktuális predikált változó mért értékeinek vektora, X a predikciós adathalmaz, β a koefficiensvektor (a súlyozó tényezőket tartalmazza), $\lambda \geq 0$ a büntetés mértékét határozza meg, $\|\beta\|_1$ pedig

a β koefficiensvektor L_1 normája. Az L_1 norma a vizsgált vektor elemeinek abszolútértékének összege (9) [34]:

$$||\beta||_1 = \sum_{j=1}^d |\beta_j| \quad (9)$$

A hagyományos lineáris regresszió (OLS) módszerével számított β koefficiensvektor számítási módját a (5)-ös összefüggésben már láttuk. Összehasonlítva a LASSO regressziónál ismertetettel, az látható, hogy a LASSO regresszió esetében a (8)-as összefüggés szerint $\lambda=0$ esetén $\hat{\beta}$ megegyezik az OLS által számított β együtthatóvektorral. Illetve az is látható, hogy elegendően nagy λ érték a $\hat{\beta}$ értékét 0-ra zsugorítja. Így, a modell paraméterek becslésébe beépített L_1 típusú büntetőfüggvény révén a LASSO regresszió képes változó szelekciót végrehajtani [33].

Ahhoz, hogy egyszerűen szemléltessük, hogy hogyan történik a változószelekció, a 3. ábra segít. Két prediktor változó esetében a prediktor változók becsült súlyait (β_1, β_2) paraméterterben az L_1 normából eredő korlátozás négyzet formában jelenik meg, piros vonallal jelölve. A korlátok határait a LASSO regresszió paramétere λ határozza meg, a keresett predikciós súlyok értékeinek a korlátolt területen belül kell tartózkodniuk. Ehhez képest az OLS által becsült prediktor súlyok fekete, szaggatott vonallal láthatók [3].



3. ábra. Két dimenziós paraméterter (β_1, β_2), az OLS becslése (fekete szaggatott vonal, β_{OLS}), az L_1 normából eredő korlátozás (piros négyzet) [3].

Gyakorlatban a szelekció során ez annyit jelent, hogy amennyiben a veszteségkontúr érinti a (magasabb dimenziók esetén) hiper-kockát valamelyik sarokban vagy a sarkok közötti élen, akkor néhány paraméter pontosan nulla. Így, ha az L_1 normakényszer aktív, akkor az azt eredményezi, hogy néhány regressziós együttható pontosan nulla lesz, mely által megtörténik a változószelekció [3].

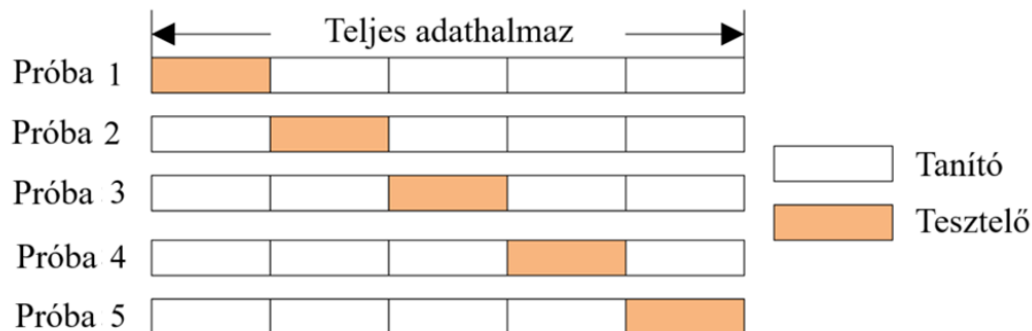
A LASSO regresszió során becsült β együtthatóvektort a LASSO koefficiens (λ) függvényében vizsgálom, mely során az látható, hogy az egyes predikciós változók súlyai mekkora mértékűek és hogyan változnak a λ függvényében. Ebből egy teljesen más megközelítésből lehet következtetni az egyes változók fontosságára, mint az LMG módszerénél látottaknál. Itt az fog számítani, (melyet a későbbiekben, a tényleges vizsgálat során prezentálok), hogy az egyes predikciós változók súlyozó együtthatóinak milyen gyors a lecsengése (mikor veszi fel a 0 értéket) a LASSO koefficiens függvényében.

A LASSO regresszió vizsgálata után, megfelelő tapasztalati λ értékkel az illesztett modell predikciós képességét keresztvalidáció segítségével becsülöm. Ez az egyik legszélesebb körben használt adat-újrámintavételezési módszer, melynek célja az adott prediktív modell valódi előrejelzési hibájának becslése, illetve a túlillesztés elkerülése [35].

A modell illesztése a változó fontosság szempontjából nem feltétlenül fontos, azonban

egy visszacsatolást ad a LASSO regresszió eredményeinek helyességéről, amennyiben az megfelelően képes predikcióra, tehát a dolgozatomban úgy használok, mint egy validálási modellt.

A módszer egy modell tanítására és értékelésére rendelkezésre álló adathalmazt véletlenszerű mintavételi módszerekkel felbontja tanítási és tesztelő adathalmazokra a 4. ábrán látható módon, melyen ötszörösen lett felbontva az adathalmaz [35].



4. ábra. Az adatkészlet ötszörös felbontása tanuló és tesztelő adathalmazokra.

A tanulási folyamat során az algoritmust az adatok tanuló adathalmazai tanítják és az így kapott modelleket azon a tesztadathalmazon tesztelik, amely nem szerepelt a tanulási adathalmazban. A betanított modellek teljesítményének átlaga a végső modell teljesítményének becslése [35].

4. Kísérleti munka, eredmények feldolgozása

Szakmai munkám során szuperkritikus extrakció során extrahált illóolajakat alkotó komponensek extrakciós és egyéb feldolgozási paraméterei közötti összefüggéseket vizsgálok. Célom, hogy optimális paramétereket találjak a kiértékelt kísérleti adatok alapján elvárt összetételű illóolaj kinyeréséhez. Ehhez a folyamathoz először azt vizsgálom, hogy az egyes bemeneti (becslő) változók milyen mértékben határozzák meg a kimeneti (becsült) változók értékeit. Regressziós modellek alkalmazásával, a végső célom, hogy a kinyert illóolaj összetételét és kihozatalát optimalizáljam, mely során a központi kérdés, hogy az extrakciós paraméterek finomhangolását érdemes-e elvégezni, vagy a növények termesztési körülményein lenne inkább érdemes módosítani.

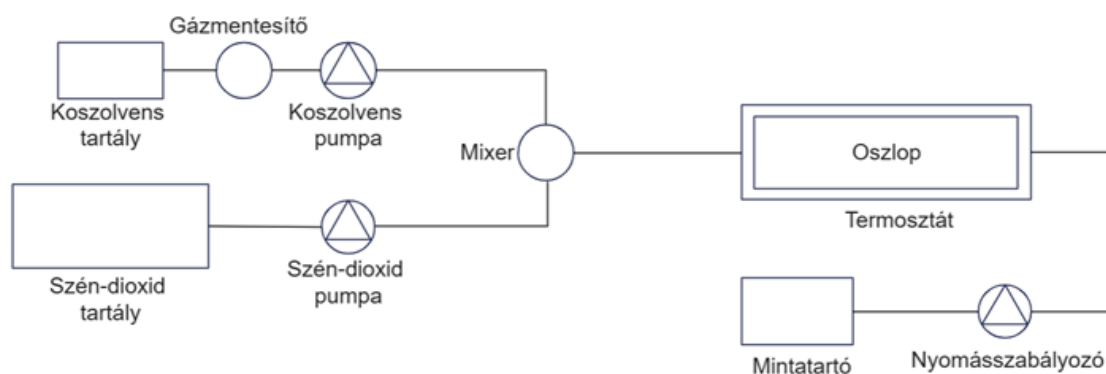
Ehhez egy mérési adathalmazt állítottam össze, melyet felhasználtam a korábban bemutatott regressziós modellek tanításához.

A mérések során az illóolaj szuperkritikus extrakcióval *Lavandula angustifolia* virágaiból lett kinyerve. Az extrakciós paraméterek 40-60°C hőmérséklet (T) és 100-200 bar nyomás (p) közt változtak, ezen tartományban vizsgáltam az extrakciós paraméterek hatását a kinyert illóolaj összetételére és kihozatalára. A kísérleti mátrix a 3. táblázatban látható, összesen 36 mérést végeztünk el.

$T[^\circ\text{C}]$ $P[\text{bar}]$	40°C	45°C	50°C	55°C	60°C
100	7M, 28M, 31M, 34M	10M, 16M	1M, 22M	13M, 19M	4M, 25M
150	8M, 29M, 32M, 35M	12M, 17M	2M, 23M	14M, 20M	5M, 26M
200	9M, 30M, 33M, 36M	11M, 18M	3M, 24M	15M, 21M	6M, 27M

3. táblázat. A szuperkritikus extrakció vizsgálatához alkalmazott kísérleti mátrix.

A vizsgált minták között Tihanyból és Tamásiból származó virágok voltak. A levendulák szüreteléstől számított állási ideje 1 és 20 hónap közt változott az extrakció időpontjában (a továbbiakban állási idő). Az általam vizsgált illóolajok kinyeréséhez alkalmazott extrakciós berendezést a 5. ábra mutatja.



5. ábra. Az általunk használt szupercritikus extrakciós berendezés sematikus ábrája.

Az egész folyamatot a szén-dioxid pumpán beállított program vezérelte. A szupercritikus állapotú oldószer etanoltól és szén-dioxidból állt. Minden mérés során az extrahálószer térfogatárama $6 \frac{\text{ml}}{\text{perc}}$ volt, ezen kívül a mérések során koszolvensnek 2 V/V% abszolút etanolt alkalmaztunk. A szén-dioxidot a szén-dioxid palackból a szén-dioxid pumpa, az etanolt pedig (gázmentesítés után) a (szén-dioxid pumpa által vezérelt) koszolvens szivattyú adagolja a keverőbe. Az extrahálószer az extraktorként működő oszlopba érkezett, amiben az általunk betömörített aprított levendula virágok voltak. Az oszlop egy termosztátban helyezkedett el, amely az extrakcióhoz szükséges hőmérsékletet biztosította. A szupercritikus nyomást az oszlophoz tartozó nyomásszabályozó segítségével állítottuk be. Az oszlopban történt meg az előbb említett paraméterek mellett az extrakció. A használt berendezések típusai a 4. táblázatban láthatóak.

Berendezés neve	Gyártó	Típuszám	Gyártó szerinti megnevezés
Gázmentesítő	Jasco	DG-2080-54	4-Line Degasser
Koszolvens pumpa	Jasco	PU-2080 Plus	Intelligent HPLC Pump
Szén-dioxid pumpa	Jasco	PU-1580-CO ₂	CO ₂ Delivery Pump
Keverő	Jasco	MX-2080-32	Dynamic Mixer
Termosztát	Jasco	CO-2060 Plus	Intelligent Column Thermostat
Nyomásszabályozó	Jasco	BP-1580-81	Back Pressure Regulator

4. táblázat. Az extrakciós berendezés részeinek típusai.

Az extrahált minták összetétele egy gázkromatográf (Agilent 5975C típusú) összekapcsolt tömegspektrométer (GC-MS) segítségével lettek kiértékelve. A gázkromatográfias elválasztás HP5 típusú oszlopon történt (30.0 m × 0.25 mm; 0.25 μm). Az injektor

és a detektor hőmérséklete 250°C volt. Az oszlop 10°C/perc fűtési sebességgel lett programozva 50°C-tól 250°C-ig. A kapott kromatogramok alapján 13 fő komponens lett beazonosítva, melyeket figyelembe vettem a modellezés során.

4.1. Eredmények elemzése

Az adatok elemzésének szemléletesebbé tételét különböző vizualizációs módszerekkel hajtom végre. A kapott eredmények eloszlását szórásdiagramok segítségével szemléltem, az egyes műveleti (6. ábra), időjárási (7. ábra), valamint feldolgozási (8. ábra) paraméterek függvényében, majd később az egyes változók közötti összefüggéseket korrelációanalízissel (9. ábra) vizsgálom.

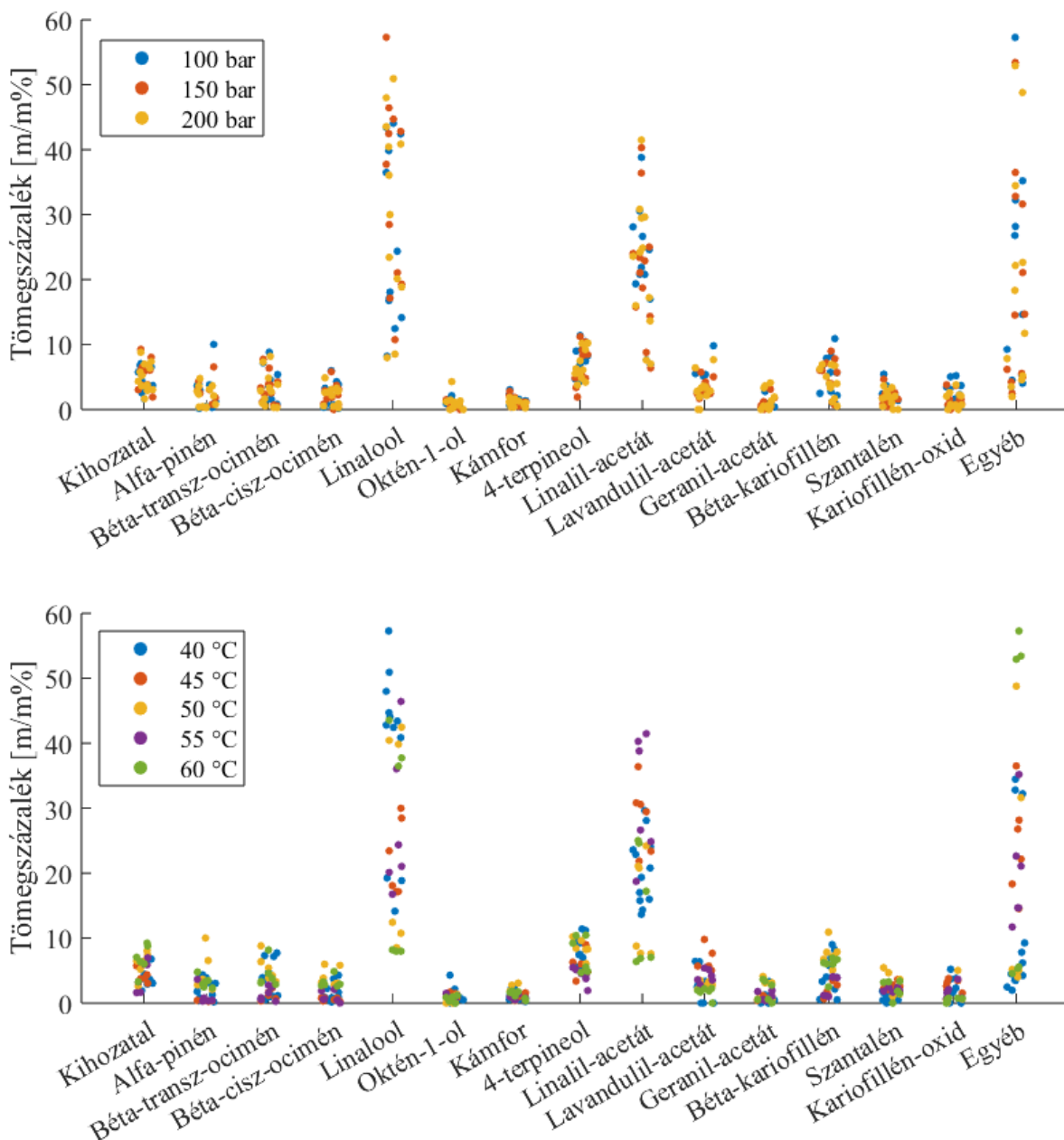
Az egyes kategorikus változókat (ebben az esetben a növények származási helyét), a vizsgálat során címkekódolással (label encoding) számosítom, a kódolt csoportokat és azok kódját a 5. táblázatban tüntettem fel.

Kódolt mennyiségek	0	1
Származás helye	Tihany	Tamási

5. táblázat. Kategorikus adatok címkekódolása.

A 6.-8. ábrák szórásdiagramokat ábrázolnak, melyeken az egyes komponensek tömegszázalék szerint ábrázolt kihozatali és összetétel értékei láthatók a mért mintákban. Az ábrázolt pontok színezése az egyes, későbbiekben jelölt műveleti, időjárási, illetve feldolgozási változók értékei szerint történt.

A 6. ábrán két olyan szórásdiagram látható, melyen az egyes komponensek tömegszázalék szerint ábrázolt összetételi és kihozatali értékei az extrakciós paraméterek, azaz hőmérséklet és nyomás szerint lettek színezve.

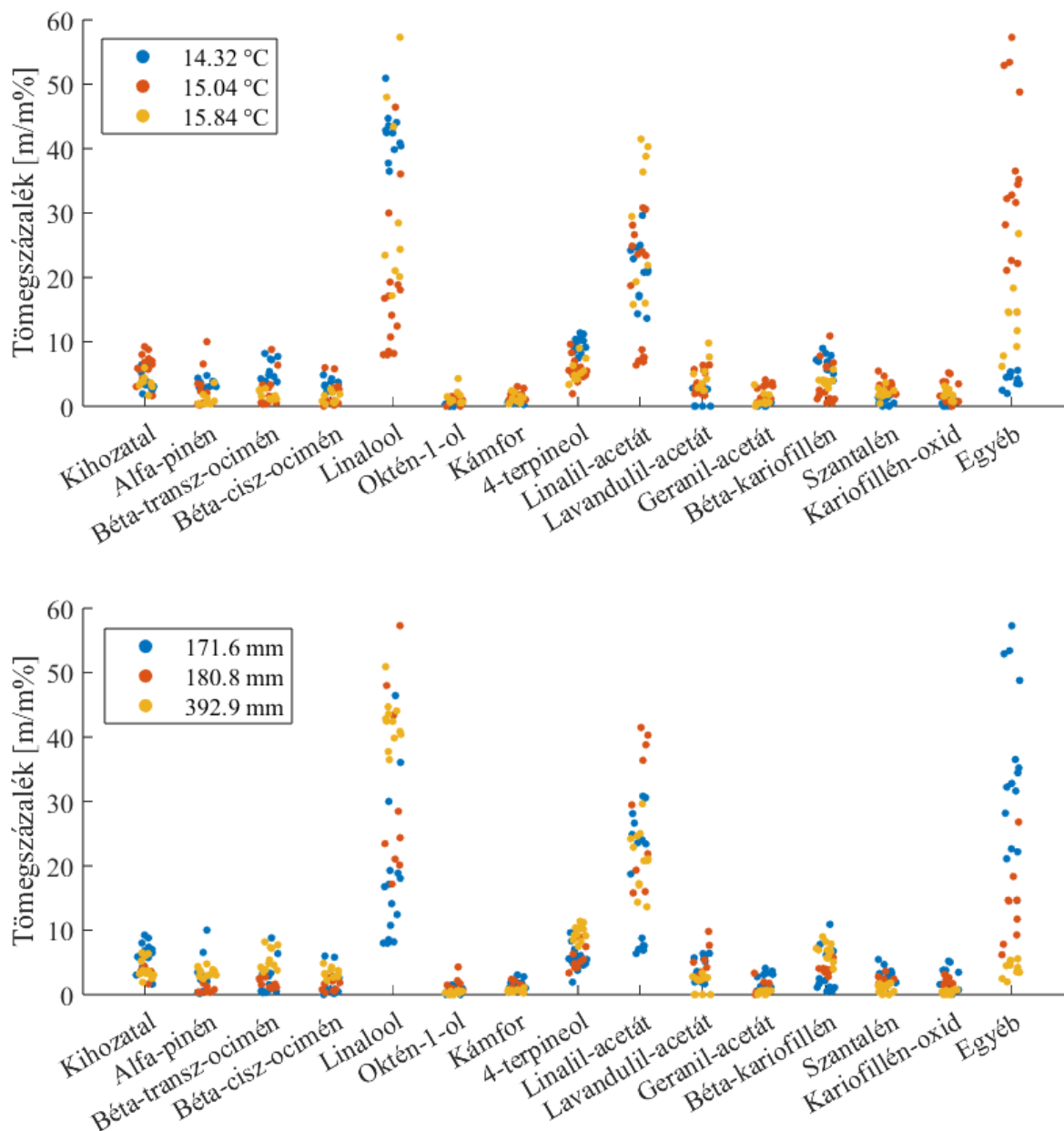


6. ábra. A kihozatali, illetve az egyes illóolaj alkotó komponensek eloszlása a nyomás (felül) és extrakciós hőmérséklet (alul) szerint.

A 6. ábráról kiemelendő, hogy több komponensnél, például a béta-transz-ocimén, béta-cisz-ocimén, linalool, lavandulil-acetát és kariofillén-oxid esetében is általánosan

elmondható, hogy a hőmérséklet növekedésével az említett komponensek mennyisége csökken, míg ezzel párhuzamosan a kihozatal és az egyéb illóolajalkotó komponensek mennyisége nő. A nyomás esetében a komponensek értékei saját nagyságrendjükhöz képest nagymértékben szórnak, ezáltal a diagramokon nem látható egyértelmű trend, azonban az extrakciós paraméterek együttes hatásáról elmondható, hogy az olaj minőségét meghatározó komponensek (bővebben a 2. táblázat), mint a linalool és kámfor esetében a viszonylag alacsony hőmérséklet és magas nyomás a kedvező (a linaloolnak ekkor nő a mennyisége, a kámfornak pedig nem túl magas).

A 7. ábrán szintén két szórásdiagram látható, melyen az egyes komponensek tömegszázalék szerint ábrázolt összetételi és kihozatali értékei láthatók, azonban a színezés az időjárási paraméterek, azaz a virágzás alatti átlagos hőmérséklet és csapadékmennyiség szerint történt.



7. ábra. A kihozatali, illetve az egyes illóolaj alkotó komponensek eloszlása a virágzás alatti átlagos hőmérséklet (felül) és átlagos csapadék (alul) szerint.

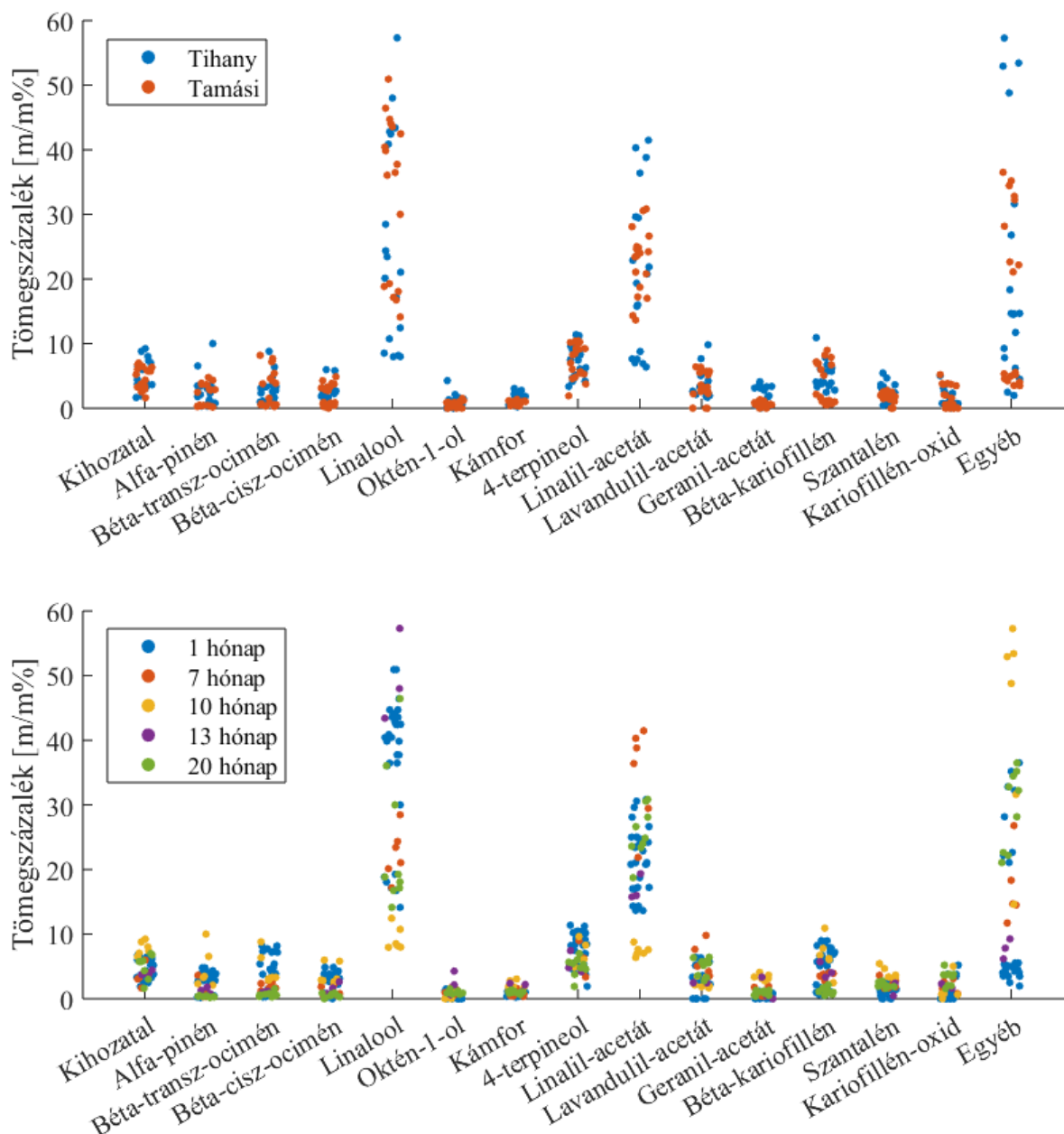
A 7. ábrán az látható, hogy a vizsgált időjárási paraméterek közül a középső átlaghőmérséklet és kis csapadékmennyiség mellett mutatkoztak a legmagasabb kihozatali értékek. Ugyanilyen trendet mutatnak az alfa-pinén, béta-transz-ocimén, és béta-cisz-ocimén komponensek.

Magasabb csapadékmennyiség és alacsonyabb hőmérsékletértékek a linalool, a 4-terpineol és béta-kariofillén mennyiségének kedvez, míg az egyéb komponensek mennyi-

sége ezen paraméterek mellett csökkenést mutat, ugyan úgy, mint a kámfor, lavandulil-acetát, geranil-acetát, szantalén és kariofillén oxid-mennyisége. Az illóolaj minősége szempontjából fontos linalil-actát mennyiségének általánosan elmondható, hogy a magas hőmérséklet és közepes csapadékmennyiség kedvezett.

Kiemelném, hogy mind az extrakciós hőmérséklet és nyomás, mind pedig az átlaghőmérséklet növekedésével az illóolajban található vizsgált (és minőségét meghatározó) komponensek aránya több esetben csökkent. Ennek több oka is lehet. Egyrészt a szuperkritikus oldószer oldási képessége az extrakciós paraméterek függvényében nagy mértékben változhat, így előfordulhat, hogy az eltérő paraméterek mellett az más komponenseket is nagyobb mértékben old, így a vizsgált komponensek hányada csökken. Másrészt levendula illóolajában (ezáltal növényében) található komponensek közül a főkomponensek többsége nemcsak illékony, de hőbomlékony is, így a hőmérséklet növekedése okozhatja a komponensek bomlását [36,37].

A 8. ábra szintén két szórásdiagramot mutat, melyen az egyes komponensek tömegszázalék szerint ábrázolt összetételi és kihozatali értékei láthatók. A színezés ebben az esetben a feldolgozási paraméterek függvényében történt, azaz a származás helye, illetve az állási idő szerint.



8. ábra. A kihozatali, illetve az egyes illóolaj alkotó komponensek eloszlása az extrahált növények származása (felül) és állási idő (alul) szerint.

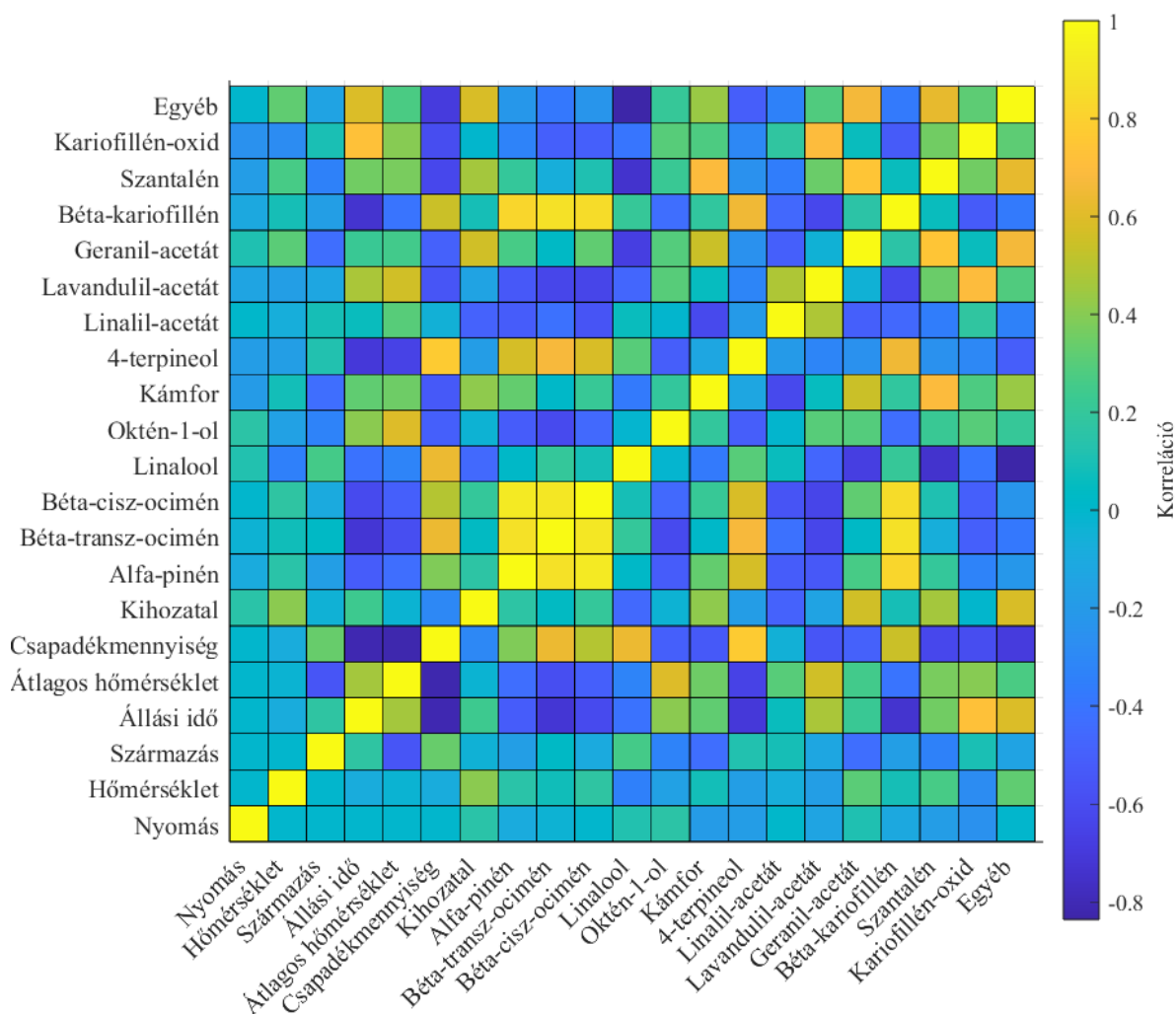
A 8. ábrán a származási hely és állási idő szerint is a diagramokon látható módon, viszonylag jól elkülöníthetőek az egyes csoportok. Átlagosan elmondható, hogy a Tihanyi levendulák esetében nagyobb szórás mutatkozik a kihozatali értékeknél, illetve az alfa pinén, linalool, kámfor, linalil-acetát, geranil-acetát, szantalén, illetve az egyéb illóolajalkotó komponensek tömegszázalékos értékénél is, míg a tamási minták esetében a kariofillén-oxid tömegszázalékos értékei mutatnak nagyobb szórást. A többi kompo-

nensnél nem figyelhető meg kifejezetten mennyiségbeli elkülönülés, azonban jól látszik, hogy a tihanyi levendulák több oktén-1-olt, kámfort és geranil-acetátot tartalmaznak.

Az állási időket tekintve az mondható, hogy a frissen extrahált (1 hónapot állt a növény az extrakció időpontja előtt) olajokban viszonylag magas az alfa-pinén, a béta-transz-ocimén, a béta-cisz-ocimén, a béta-kariofillén tartalom, illetve kimagasló a linalool és a 4-terpineol tartalom. Kiemelendő, hogy szinte az összes 1 hónapos állási idővel rendelkező mintánál az egyéb (beazonosítatlan) komponensek mennyisége a többi mintához képest csekély. Az alfa-pinén, béta-transz-ocimén és béta-cisz-ocimén komponensek értékei csökkenést mutatnak a hőmérséklet növekedésével.

A 10 hónapos állási időnél mért kihozatali, geranil-acetát, valamint szantalén értékek, a más állási idővel rendelkező mintákhoz képest kimagasló értékeket mutatnak, azonban az olajban található linalool és linalil-acetát mennyisége a többi értékhez képest nagyon kicsi. A 13 és 20 hónapos állási idő után mért minták esetében az látható, hogy több komponens esetében is ezen mintáknál tapasztalhatóak a legkisebb értékek. Ilyen összetevő az alfa-pinén, béta-transz-ocimén, béta-cisz-ocimén és béta-kariofillén. Fontos megjegyezni, hogy a szakirodalomban már többen vizsgálták azt, hogy az idő függvényében hogyan illannak el az egyes komponensek a levendula olajok esetében. Ezen kutatások eredményei, hogy az illékony komponensek mennyisége az időben (szárítási vagy állási idő alatt) folyamatosan csökken, melyből valószínűsíthető, hogy a 13 és 20 hónapos minták esetében is ez a folyamat játszódott le [38, 39].

A 9. ábra az általunk vizsgált adatmátrix Pearson-korrelációs együtthatóinak színtérképes ábrázolását mutatja, melyet a (1)-es összefüggés alapján számítottam [22].



9. ábra. Az egyes változók korrelációs mátrixának színtérképe.

A nyomás és az illóolaj összetevők és kihozatal korrelációja körülbelül -0,2 és 0,2 közé esik, tehát, míg a hőmérséklet nagyobb mértékben befolyásolja az illóolaj kihozatalát, valamint az egyes komponensek mennyiségét, a nyomás önmagában nem annyira befolyásoló tényező.

Kiemelném a hőmérséklet és a kihozatal, valamint az egyéb illóolaj alkotó komponensek közötti, körülbelül 0,4-es értékű korrelációt, mely szerint a hőmérséklet növekedésével nő ezen változók mennyisége.

A származás a linaloolal pozitív, míg a kámforral, illetve a geranil-acetáttal enyhe negatív korrelációt mutat. Ezen értékeket a 5. táblázatban feltüntetett kódolás alapján lehet értelmezni. Amennyiben a Tihanyból származó levendulák illóolajáról beszélünk, akkor azokban alacsonyabb a linalool és magasabb a kámfor, illetve geranil-acetát tartal-

lom, mint a Tamásiból származó levendulák esetében. Az utóbbi két eset összhangban van a 8. ábrán látottakkal, míg a linalool esetében a súlyozás nem volt egyértelmű a szórásdiagramon.

Az ábráról szintén kiolvashatjuk, hogy a szüretelés óta eltelt idő (állási idő) hogyan befolyásolja az összetételt. A később extrahált levendulák olajában magasabb az oktén-1-ol, a kámfor, a lavandulil-acetát, a szantalén, a kariofillén-oxid és az egyéb illóolaj alkotó komponensek aránya, míg a hamarabb extraháltakéban a béta-transz-ocimén, béta-cisz-ocimén, 4-terpineol és béta-kariofillén mennyisége a nagyobb. Ennek oka a már említett, állási idő függvényében tapasztalt, illékonyabb anyagok távozásának valószínűsíthető a növényből. Az eredmények összhangban vannak a 8. ábrán látottakkal.

A vizsgált időjárási paramétereket tekintve magasabb átlaghőmérséklet mellett nagyobb oktén-1-ol, kámfor, linalil-acetát, lavandulil-acetát, szantalén, kariofillén-oxid és egyéb illóolaj alkotó komponensértékek tapasztalhatók, míg alacsonyabb hőmérséklet mellett a béta-transz-ocimén, béta-cisz-ocimén és 4-terpineol értékei a magasabbak. Kis csapadékmennyiség mellett pedig a magasabb kihozatali, oktén-1-ol, kámfor, lavandulil-acetát, geranil-acetát, szantalén, kariofillén-oxid és egyéb komponensértékek mutatkoztak. Ezzel ellentétes trendet mutatnak az alfa-pinén, béta-transz-ocimén, és béta-cisz-ocimén, linalool, 4-terpineol és béta-kariofillén komponensek. Ezen eredmények is egyezést mutatnak a 7. ábráról leolvasottakkal.

A korrelációanalízis során szintén tapasztalható volt az a jelenség, hogy az egyes (extrakciós és átlagos) hőmérsékletek növekedésével az illóolajban található vizsgált (és minőségét meghatározó) komponensek aránya több esetben csökkent. A lehetséges okoknak tárgyalt hőbomlékonyságot, illetve az oldószer oldóképessége megváltozását az előzőekben részleteztem.

A komponensek egymás közötti összefüggések közül kiemelném, hogy a kihozattal nő az egyéb komponensek mennyisége, valamint az alfa-pinén, béta-transz-ocimén, béta-cisz-ocimén, 4-terpineol és béta-kariofillén mennyisége nagyrészt együtt változik, amit igazol a közöttük lévő, majdnem teljes mértékű, 1-hez közeli korreláció, valamint a linalool mértéke akkor számottevő, amikor az egyéb illóolajalkotó komponensek tömegszázalékos mennyisége alacsonyabb.

A 6.-9. ábrák sokszor jól elkülöníthető csoportokat, illetve egyszerű trendeket mutat-

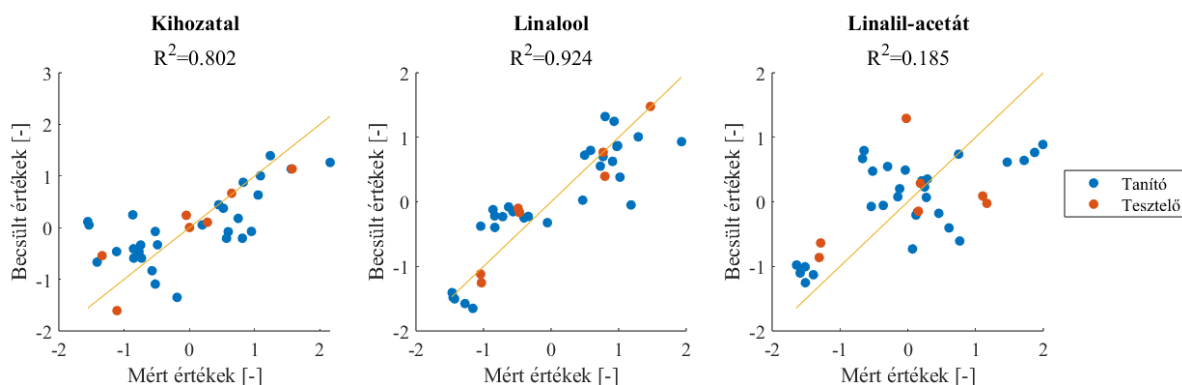
nak a vizsgált paraméterek szerint szétbontva. Szakmai munkám során fontosnak tartom ezen paraméterek együttes hatásának vizsgálatát. Nem minden paraméter ugyanolyan mértékben befolyásolja az extrahált illóolaj összetételét. A következőkben ezen paraméterek fontosságát vizsgálom.

4.2. A kihozatalt és összetételt befolyásoló paraméterek rangsorolása

Az illóolaj kihozatalt és összetételt nem minden predikciós változó egyforma mértékben befolyásolja. Ez látható volt már a 6-9. ábrákon is, melyeken szórásdiagramok, illetve az adathalmaz korrelációmátrixa volt látható. Az előbbin sokszor jól elkülöníthető csoportok voltak láthatók, utóbbi típusú ábra pedig azt mondta meg, hogy két paraméter változása hogyan hat egymásra. Ezen információk egy átfogó képet adnak az egyes becslő változók becsült változókra gyakorolt hatásáról, azonban olyan értelemben nem adnak számszerű értéket, hogy mégis milyen mértékben határozzák meg a vizsgált halmazt.

A következő alkalmazott módszer, amely eredményeit bemutatom, az LMG módszer, mely pontosan azt becsli, hogy egy-egy regresszor (becslő változó) mennyire fontos az adott tulajdonság becslésénél.

A módszer alapvetően a lineáris regresszió technikáján alapszik, emiatt először a lineáris regresszió predikciós képességét ellenőriztem ötszörös keresztvalidációt alkalmazva (4. ábra). A validálás során minden predikált változót külön összefüggéssel becsültem. Az egyes predikált változók nevei alatt feltüntettem az aktuális modellhez tartozó R^2 együtthatókat, melyek itt azt adják meg, hogy a modell mennyire helyesen tudja megbecsülni az vizsgált változó értékét. A 10. ábrán feltüntettem a keresztvalidáció egyik becslő modelle esetén a becsült értékeket a mért értékek függvényében, a kihozatal, illetve a minőséget tekintve, a két legfontosabb komponens esetében.



10. ábra. Lineáris regressziós modell predikációs képessége a keresztvalidáció egyik becslő modelle esetében.

A modell illesztése során kapott egyes modellek R^2 együttthatóinak átlaga adja meg a teljes keresztvalidáció során az átlagos R^2 értéket, mellyel a modellt egészében lehet jellemezni. Ezen átlagos R^2 értékek a 6. táblázatban láthatók az egyes predikált paraméterek esetében.

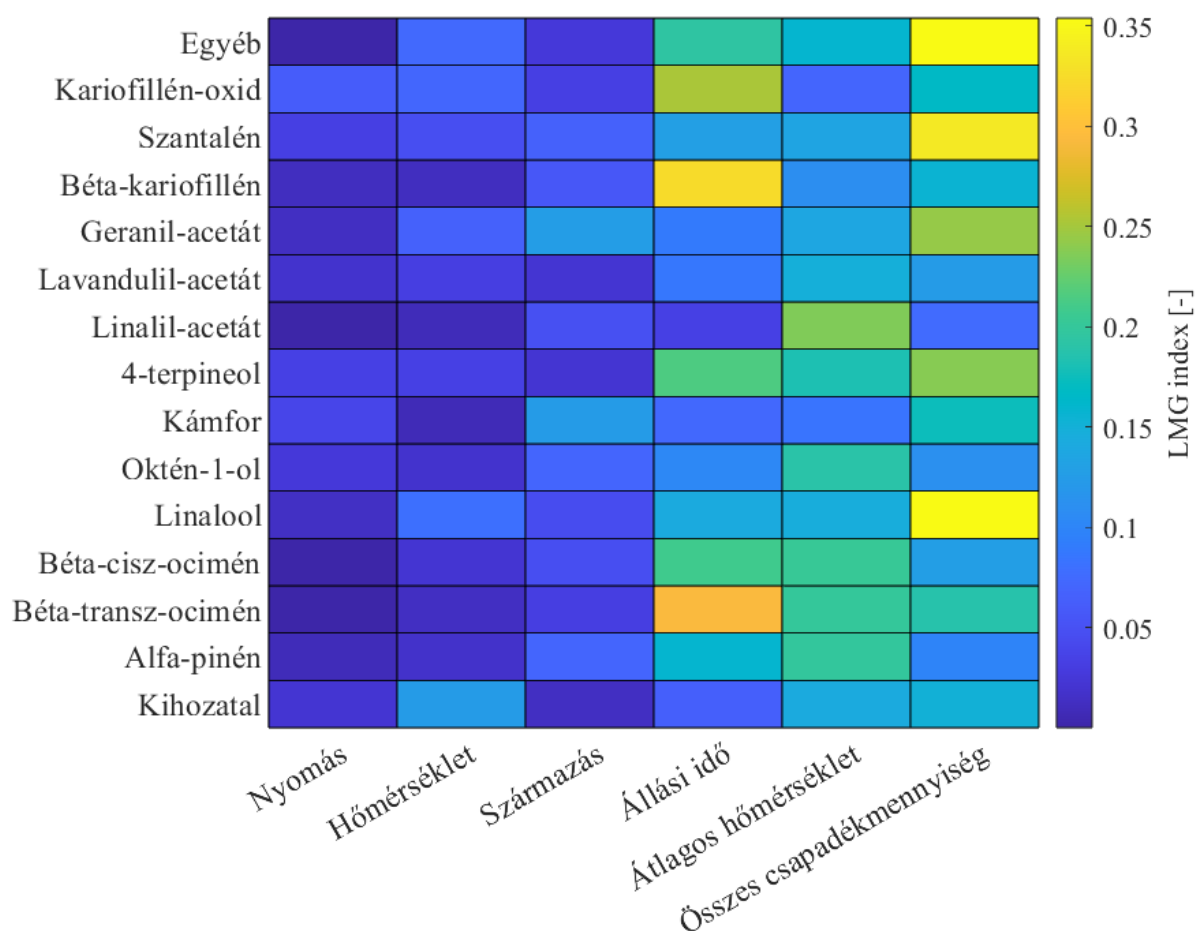
Predikált paraméter	Átlagos R^2 együttthatók [-]
Kihozatal	0,596
Alfa-pinén	0,409
Béta-transz-ocimén	0,623
Béta-cisz-ocimén	0,495
Linalool	0,636
Oktén-1-ol	0,385
Kámfor	0,387
4-terpineol	0,656
Linalil-acetát	0,528
Lavandulil-acetát	0,523
Geranil-acetát	0,464
Béta-kariofillén	0,532
Szantalén	0,467
Kariofillén-oxid	0,652
Egyéb	0,674

6. táblázat. Lineáris regressziós technika átlagos R^2 értékei az egyes predikált paraméterek esetében az ötszörös keresztvalidálás során.

A 6. táblázatban az látható, hogy az átlagos R^2 értékek alsó hangon 0,385. A fontosabb komponensek esetében, mint a kihozatal, linalool és linalil-acetát, a modell minimum 0,5-ös R^2 értékkel rendelkezik. A 6. táblázatban feltüntetett értékek azt

mutatják, hogy a modell megbízható annyira, hogy egy átfogó képet tudjon adni a regresszorok relatív fontosságáról.

A 11. ábrán ezért az egyes predikációs paraméterek LMG indexei láthatóak az összes predikált paraméter esetében, szintértékpén ábrázolva. Az egyes LMG indexek értéke minden esetben 0 és 1 közé eshet. A vizsgálat során a legmagasabb LMG indexnek értéke körülbelül 0,35 körül adódott.



11. ábra. A predikációs paraméterek LMG indexeinek szintértékpén az egyes predikált paraméterek esetén.

A 11. ábrán az látható, hogy a kihozataalt az extrakciós hőmérséklet körülbelül ugyanannyira befolyásolja, mint az időjárást jellemző paraméterek (átlagos hőmérséklet és összes csapadékmennyiség), mely vegyipari szempontból fontos lehet. Az alfa-pinén, oktén-1-ol és lavandulil-acetát becslésében az időjárási paraméterek (leginkább az átlagos

hőmérséklet) játszanak szerepet, azonban nem elhanyagolható az állási időnek a figyelembevétele sem. A béta-cisz-ocimén és béta-transz-ocimén becslésében szintén az időjárási paraméterek, illetve (a transz forma esetében inkább) az állási idő játszik szerepet. Az olaj minősége szempontjából egyik legfontosabb komponens, a linalool mennyiségi becslésében a csapadékmennyiség körülbelül kétszer olyan fontos, mint az átlaghőmérséklet vagy állási idő és háromszor annyira fontos, mint az extrakciós hőmérséklet és származási hely. A kámfor mennyiségét az összes csapadékmennyiség és származás helye, illetve kevésbé, de nem elhanyagolható módon az állási idő és az átlagos hőmérséklet. A 4-terpineol, illetve az egyéb illóolajalkotó komponensek mennyiségi becslése az állási idő, illetve az időjárást jellemző paraméterekből szinte olyan eredményt ad az LMG módszere szerint, mintha minden eredeti becselő paramétert figyelembe vennénk. Az olaj minőségét meghatározó másik legfontosabb komponens, a linalil-acetát mennyiségét nagyrészt a származás helye és az időjárási paraméterek határozzák meg. A geranil-acetát és szantalén mennyisége főként az összes csapadékmennyiségtől függ, azonban nem elhanyagolható az extrakciós hőmérséklet, a feldolgozási paraméterek (származási hely és állási idő), illetve az átlagos hőmérséklet sem. A béta-kariofillén és kariofillén-oxid esetében a becslés során a feldolgozási paraméterek, azon belül is az állási idő játszik kiemelkedő szerepet. Mindkét komponens esetében bár kevésbé, de fontosak az időjárási paraméterek, illetve megjegyzendő, hogy a kariofillén-oxidnál az extrakciós paraméterek a többi komponenshez képest nagy súlyozó tényezőket kaptak az elemzés során. Az egyéb komponensek becslésénél az állási idő, illetve az időjárási paraméterek a meghatározóak (főként az összes csapadékmennyiség), azonban az extrakciós hőmérséklet sem elhanyagolható.

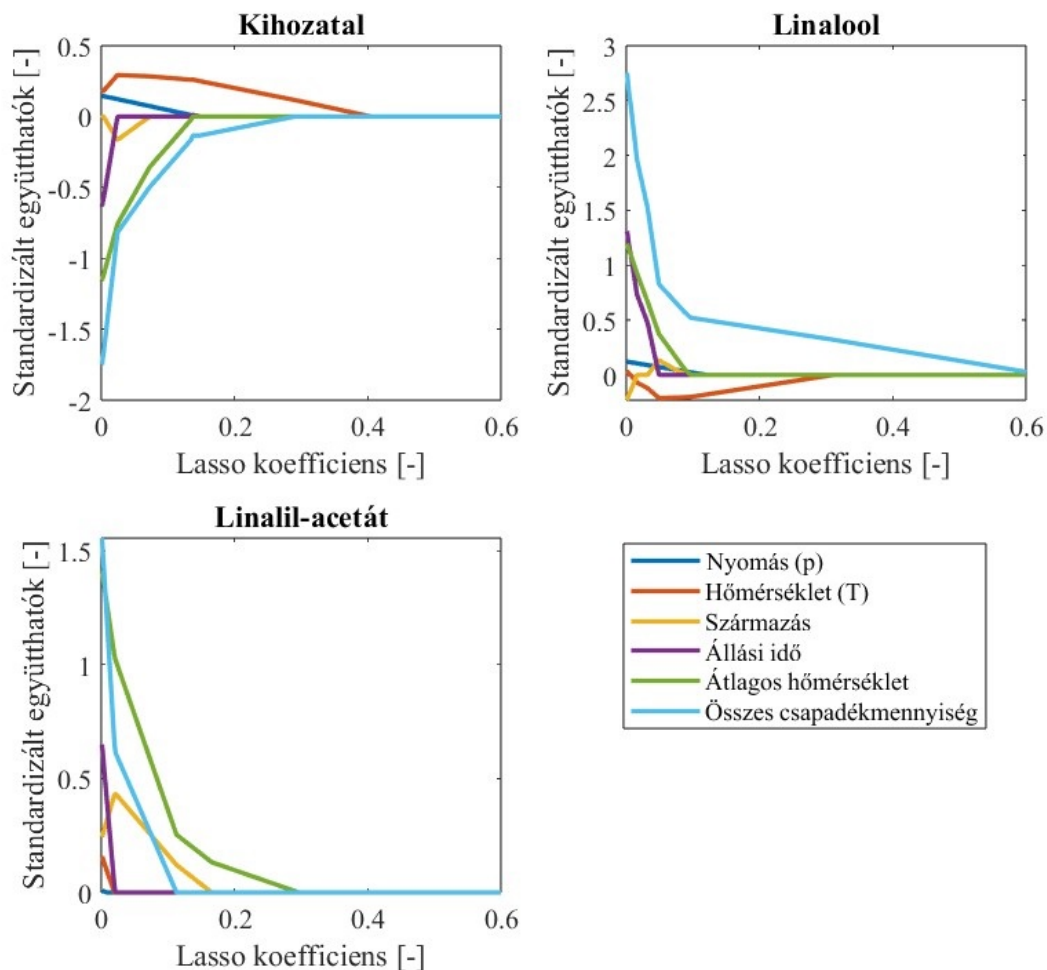
Összefoglalva az LMG módszer alkalmazása során kapott eredményeket azt kaptam, hogy mind a kihozatal, mind pedig az összetétel becslése során az időjárást jellemző adatok, illetve az állási idő kiemelkedő szerepet játszik, ezzel a többi paramétert kissé a háttérbe szorítva.

A változó fontosságot egy másik szemszögből a LASSO regresszió segítségével vizsgáltam, mely a változószelekcióra képes tulajdonsága révén azt képes többek között megmondani, hogy egyáltalán szükséges-e az adott változó a becslés során (bővebben: 3.1. fejezet).

A módszer alkalmazása során az egyes, a regresszió során kapott súlyozó együttha-

tókat vizsgálom a LASSO koeficiens függvényében, majd a nagyobb változások lejátszódása után az egyes predikált változóhoz kiválasztok egy optimálisnak talált LASSO paramétert. A predikációs képességet ezen koefficienst alkalmazva vizsgálom, ezzel ellenőrizve a módszer pontosságát. A pontosság alatt itt azt értem, hogy amennyiben a vizsgált LASSO paraméternél egy predikációs változó súlyát a módszer 0-ra csökkentette, akkor a modell képes-e megfelelő mértékben becsülni a vizsgált predikált változó értékét.

A 12. ábrán az egyes vegyipari, illetve az olaj minőségét leginkább meghatározó komponensek (azaz a kihozatal, linalool és linalil-acetát) esetében, a predikció során kapott súlyozó paramétereket ábrázoltam a LASSO koeficiens függvényében.



12. ábra. A predikációs paraméterek standardizált együtthatóinak változása a LASSO koeficiens változásának függvényében.

A 12. ábrán a kihozatal esetében az látható, hogy a legnagyobb súlyokkal az időjárás-

si paraméterek rendelkeznek, míg az extrakciós hőmérséklet súlyozó tényezője a LASSO koefficiens függvényében először nő, majd lassan elkezdi csökkenni. Ezen eredmények többnyire egyeznek az LMG által becsült súlyokkal (11. ábra). Ott az extrakciós hőmérséklet majdnem akkora LMG indexet kapott, mint az időjárást jellemző paraméterek, míg a LASSO regresszió esetében ez az a predikciós változó, mely a legtovább marad a becsülő változók között.

A linalool és linalil-acetát esetében szintén az látható, hogy az időjárási paraméterek a legfontosabbak (mely információk nagyrészt megegyeznek a megfelelő LMG indexeknél látottakkal), azonban ezen változók esetében a származás súlya, a linaloolnál pedig az extrakciós nyomás is számottevő.

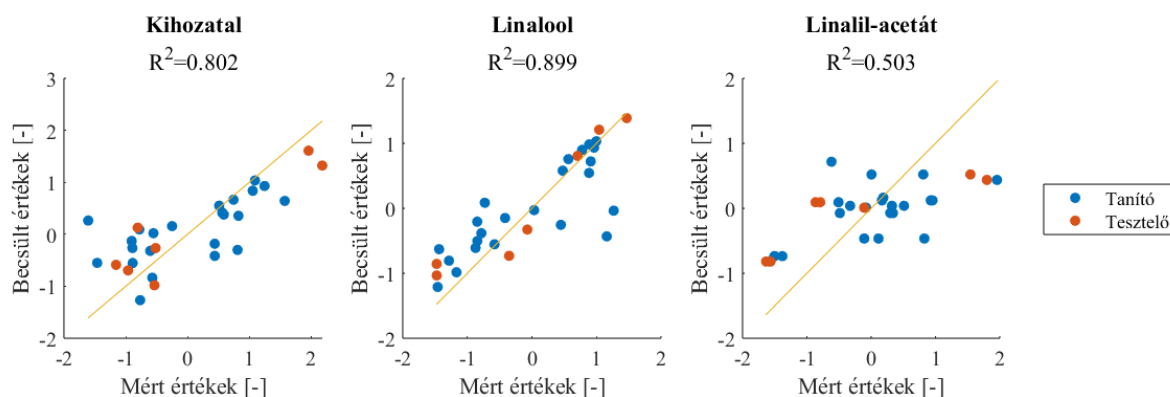
Általánosan elmondható, hogy a kihozatal és a két főkomponens mennyiségét elsősorban az időjárást jellemző predikciós változók értékei határozzák meg.

Amennyiben jobban megnézzük a 12. ábrát a kihozatal esetében, akkor az összes predikciós paraméterhez tartozó súlynak az intenzív változása körülbelül $\lambda = 0,025$ -nél lecseng. Ilyen módon a linalool esetében a λ értékére 0,03-as érték, a linalil-acetátnál pedig $\lambda = 0,025$ -ös érték határozható meg. Ezen értékeket választom az optimális LASSO koefficiensnek, melyet a későbbiekben behelyettesítek a LASSO regressziós modellbe és az így felállított modell predikciós képességét keresztvalidációval vizsgálom. A többi predikált paramétert is megvizsgáltam és a 7. táblázatban foglaltam össze.

Predikált paraméter	Választott LASSO koefficiens (λ) [-]
Kihozatal	0,025
Alfa-pinén	0,04
Béta-transz-ocimén	0,02
Béta-cisz-ocimén	0,04
Linalool	0,045
Oktén-1-ol	0,035
Kámfor	0,02
4-terpineol	0,01
Linalil-acetát	0,025
Lavandulil-acetát	0,05
Geranil-acetát	0,04
Béta-kariofillén	0,04
Szantalén	0,045
Kariofillén-oxid	0,02
Egyéb	0,025

7. táblázat. Az egyes predikált paramétereknél meghatározott LASSO koefficiensek értékei.

A 7. táblázatban feltüntetett értékeket behelyettesítettem a LASSO regressziós modellbe. Ezen modell predikációs képességét keresztvalidációval vizsgálom. A keresztvalidációs vizsgálat során ötszörös keresztvalidációt alkalmaztam (4. ábra), mely során minden predikált változót külön összefüggéssel becsültem. Az egyes predikált változók nevei alatt feltüntettem az aktuális modellhez tartozó R^2 együtthatókat, melyek esetünkben azt adják meg, hogy a modell mennyire helyesen tudja megbecsülni az adott predikált változó értékét. A 13. ábrán feltüntettem a keresztvalidáció egyik becselő modelle esetén a becsült értékeket a mért értékek függvényében, a kihozatal, illetve a két legfontosabb komponens esetében.



13. ábra. LASSO regressziós modell predikációs képessége a keresztvalidáció egyik becsülő modelle esetében a 7. táblázatban feltüntetett LASSO koefficiens értékekkel.

A modell illesztése során kapott egyes modellek R^2 együttthatóinak átlaga adja meg a teljes keresztvalidáció során az átlagos R^2 értéket, mellyel a modellt egészében lehet jellemezni. Ezen átlagos R^2 értékek a 8. táblázatban láthatók az egyes predikált paraméterekre.

Predikált paraméter	Átlagos R^2 együttthatók [-]
Kihozatal	0,846
Alfa-pinén	0,690
Béta-transz-ocimén	0,801
Béta-cisz-ocimén	0,698
Linalool	0,887
Oktén-1-ol	0,553
Kámfor	0,553
4-terpineol	0,580
Linalil-acetát	0,612
Lavandulil-acetát	0,712
Geranil-acetát	0,579
Béta-kariofillén	0,608
Szantalén	0,572
Kariofillén-oxid	0,542
Egyéb	0,852

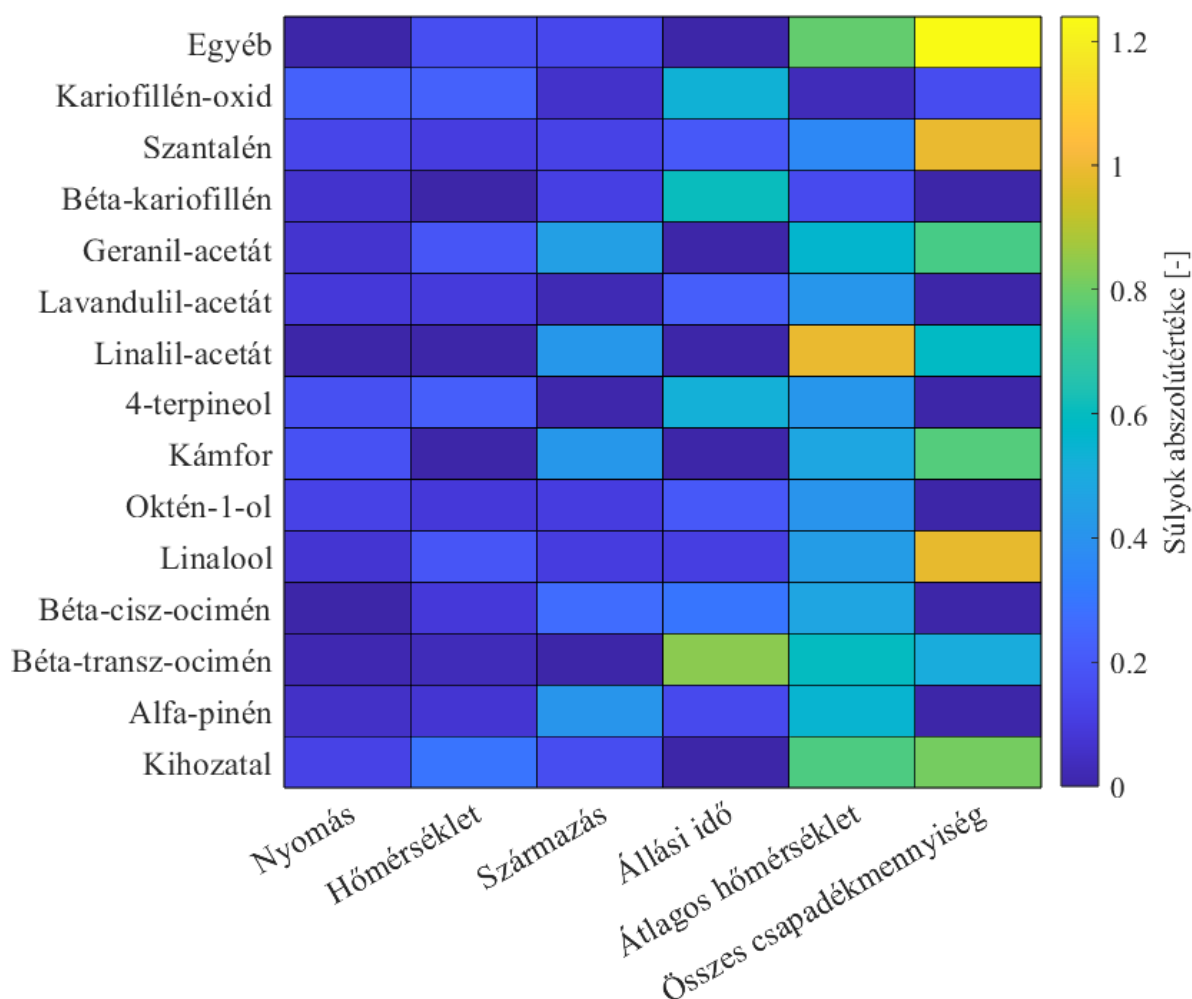
8. táblázat. LASSO regressziós technika átlagos R^2 értékei az egyes predikált paraméterek esetében az ötszörös keresztvalidálás során.

A 8. táblázatban az látható, hogy az átlagos R^2 értékek alsóhangon 0,542. A fontosabb komponensek esetében, mint a kihozatal és linalool, a modell minimum 0,8-as R^2 értékkel rendelkezik, azonban a harmadik legfontosabb komponensnél, a linalil-acetát

esetében ez az érték valamivel alacsonyabb.

A 6. és 8. táblázatban látható átlagos R^2 értékeket összehasonlítva az látható, hogy a LASSO regresszió esetében a predikciós képesség majdnem minden becsült változó esetében nőtt, tehát a lineáris regressziós modell (ami igazából megegyezik a $\lambda=0$ értékkel számított LASSO regresszió becslésével), nagy valószínűséggel túlillesztést mutat.

A 8. táblázatban feltüntetett értékek azt mutatják, hogy a modell megbízható annyira, hogy egy átfogó képet tudjon adni a regresszorok relatív fontosságáról. A 14. ábrán ezért az LMG módszerénél látott 11. ábrához hasonló szintériképet ábrázoltam a LASSO regresszió esetén is, melyen a színezés a LASSO regresszió által számított β koefficiensvektorok egyes elemeinek abszolútértéke szerint történt. Azért választottam az elemek abszolútértéket, mivel így az egymáshoz viszonyított mértékeket egyszerűbben észre lehet venni.



14. ábra. LASSO regressziós modell által számított súlyozó tényezők abszolút értékének szintértéke az egyes predikált paraméterek esetén.

A 14. ábrán szintén látható, hogy a nagyobb súlyokat az időjárást jellemző predikációs változók kapták, azonban a 11. ábrával ellentétben ez a módszer sok helyen kiszelektálta az összes csapadékmennyiséget, mint becsülő változót. Megjegyzendő, hogy a feldolgozási paraméterek több helyen ennél a módszernél is kiemelkedő szerepet töltenek be a becslés során, ilyen a béta-transz-ocimén és a béta-kariofillén esete például (ez a trend egyébként az LMG indexek mátrixába ugyan úgy látható). Az extrakciós paraméterek itt sem töltenek be annyira fontos szerepet, azonban az extrakciós hőmérséklet habár jóval kisebb mértékben, mint más becsülő paraméterek, de több komponens becsléséhez is nem elhanyagolható módon járul hozzá, mint például a kihozatal.

5. Összefoglalás

Munkám célja *Lavandula angustifolia* virágainak szuperkritikus extrakciója során extrahált illóolajok összetételének és kihozatalának vizsgálata volt. A vizsgálat során az extrakciós, feldolgozási és időjárási paraméterek és az illóolaj alkotó komponensek, valamint az illóolaj kihozatal közötti korrelációt elemeztem. Az extrakciós paraméterek kis, a feldolgozási paraméterek mérsékelt, az időjárási paraméterek pedig jelentős mértékben befolyásolják a predikált eredményeket.

A mérések során gyűjtött adatokból különböző regressziós technikákkal próbáltam becsülni, hogy a vizsgált minták predikációs paraméterei mekkora mértékben befolyásolják az illóolajok kihozatalát és összetételét, milyen fontosak azok predikációjában. Az LMG módszer alkalmazása során kapott eredmények igazolják, hogy az időjárás szerepe kiemelkedő helyen van az illóolajok összetételének becslésében, illetve a módszer rámutatott arra, hogy az állási idővel is befolyásolhatjuk egyes komponensek mennyiségét. A LASSO regresszió során hasonló eredményeket kaptam. Röviden, a két módszerből arra következtettem, hogy amennyiben egy adott összetételű illóolajat szeretnénk kinyerni, akkor az optimalizálás szempontjából a növény termesztési körülményeire szükséges inkább fókuszálni, mintsem a műveleti paraméterek finomhangolására.

A továbbiakban célom a kísérleti adatbázis bővítése különböző forrásból származó mintákkal, valamint más tanítási modellek illesztése, illetve azok predikációs képességének tesztelése. Ezen kívül szeretném feltárni, hogy a mintákból kinyert illóolaj tulajdonságait esetleg milyen paraméter határozhatja meg még meg. Ezen folyamat során szeretném vizsgálni még többek között azt, hogy az extrakciós töltet szemcsemérete, illetve a szuperkritikus oldószer térfogataránya hogyan befolyásolja az extrakció kihozatalát, a kinyert illóolaj összetételét. Amennyiben lesz lehetőségem, szeretném a növények származási helyén lévő táptalajt jellemző paraméterek függvényében is vizsgálni az adathalmazt.

Hivatkozások

- [1] MS Windows NT volume of the production of essential oil of lavender in france from 2012 to 2018. <https://www.statista.com/statistics/975529/essential-oil-lavender-production-volume-france>. Accessed: 2022-12-14.
- [2] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, 810:1–124, 2019.
- [3] Morten Arendt Rasmussen and Rasmus Bro. A tutorial on the lasso approach to sparse modeling. *Chemometrics and Intelligent Laboratory Systems*, 119:21–31, 2012.
- [4] Dániel Cossuta. Gyógyászati, kozmetikai és élelmiszeripari célra alkalmas növényi hatóanyagok kinyerése szuperkritikus extrakcióval. 2010.
- [5] Muhammad Asif Hanif, Shafaq Nisar, Ghufrana Samin Khan, Zahid Mushtaq, and Muhammad Zubair. Essential oils. *Essential Oil Research: Trends in Biosynthesis, Analytics, Industrial Applications and Biotechnological Production*, pages 3–17, 2019.
- [6] Ravi Kant and Anil Kumar. Review on essential oil extraction from aromatic and medicinal plants: Techniques, performance and economic analysis. *Sustainable Chemistry and Pharmacy*, 30:100829, 2022.
- [7] MT Lis-Balchin. Lavender. In *Handbook of herbs and spices*, pages 329–347. Elsevier, 2012.
- [8] Anton de Groot and Erich Schmidt. Essential oils, part v: peppermint oil, lavender oil, and lemongrass oil. *Dermatitis*, 27(6):325–332, 2016.
- [9] Maria Lis-Balchin. 11 lavender essential oil. *Medicinal and aromatic plants—industrial profiles*, page 117, 2002.
- [10] Janka Bobek and Dóra Rippelné Pethő. Levendula extrakciója nagy nyomású széndioxid-dal. *Műszaki Tudományos Közlemények*, (7):107–110, 2017.

- [11] E Cséfalvay, A Deák, T Farkas, L Hanák, LT Mika, P Mizsey, J Sawinsky, B Simándi, T Szánya, E Székely, et al. Vegyipari műveletek ii., anyagátadó műveletek és kémiai reaktorok, 2011.
- [12] Erika Csordásné Rónyai. Növényi hatóanyagok kinyerése szuperkritikus extrakcióval. 2002.
- [13] Attila Nagy. Klasszikus és kvantum gépi tanulás dióhéjban. *Biztonságtudományi Szemle*, 4(4.):105–113, 2022.
- [14] Milica Aćimović, Jovana Ljujić, Jelena Vulić, Valtcho D Zheljazkov, Lato Pezo, Ana Varga, and Vesna Tumbas Šaponjac. Helichrysum italicum (roth) g. don essential oil from serbia: Chemical composition, classification and biological activity—may it be a suitable new crop for serbia? *Agronomy*, 11(7):1282, 2021.
- [15] Noha E El-Attar, Mohamed K Hassan, Othman A Alghamdi, and Wael A Awad. Deep learning model for classification and bioactivity prediction of essential oil-producing plants from egypt. *Scientific Reports*, 10(1):21349, 2020.
- [16] Mathieu Daynac, Alvaro Cortes-Cabrera, Jose M Prieto, et al. Application of artificial intelligence to the prediction of the antimicrobial activity of essential oils. *Evidence-Based Complementary and Alternative Medicine*, 2015, 2015.
- [17] Lato Pezo, Biljana Lončar, Olja Šovljanski, Ana Tomić, Vanja Travičić, Milada Pezo, and Milica Aćimović. Agricultural parameters and essential oil content composition prediction of aniseed, based on growing year, locality and fertilization type—an artificial neural network approach. *Life*, 12(11):1722, 2022.
- [18] Milica Aćimović, Lato Pezo, Tijana Zeremski, Biljana Lončar, Ana Marjanović Jeromela, Jovana Stanković Jeremic, Mirjana Cvetković, Vladimir Sikora, and Maja Ignjatov. Weather conditions influence on hyssop essential oil quality. *Processes*, 9(7):1152, 2021.
- [19] Mohsen Niazian, Seyed Ahmad Sadat-Noori, and Moslem Abdipour. Artificial neural network and multiple regression analysis models to predict essential oil content of

- ajowan (*carum copticum* l.). *Journal of applied research on medicinal and aromatic plants*, 9:124–131, 2018.
- [20] Mamunur Rashid, Bifta Sama Bari, Yusri Yusup, Mohamad Anuar Kamaruddin, and Nuzhat Khan. A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE access*, 9:63406–63439, 2021.
- [21] Qinghui Tang, Yingquan Chen, Haiping Yang, Ming Liu, Haoyu Xiao, Ziyue Wu, Hanping Chen, and Salman Raza Naqvi. Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions. *Energy & Fuels*, 34(9):11050–11060, 2020.
- [22] Dominic Edelmann, Tamás F Móri, and Gábor J Székely. On relationships between the pearson and the distance correlation coefficients. *Statistics & probability letters*, 169:108960, 2021.
- [23] K Muralidharan. A note on transformation, standardization and normalization. *IUP J. Oper. Manag*, 9:116–122, 2010.
- [24] Julian Karch. Improving on adjusted r-squared. *Collabra: Psychology*, 6(1), 2020.
- [25] Jürgen Groß. *Linear regression*, volume 175. Springer Science & Business Media, 2003.
- [26] Kovács Péter. A multikollinearitás vizsgálata lineáris regressziós modellekben. *Statistikai szemle*, 86(1):38–67, 2008.
- [27] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.
- [28] Ulrike Grömping. Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61(2):139–147, 2007.
- [29] Geoffrey I Webb and Paul Conilione. Estimating bias and variance from data. *Pre-publication manuscript (<http://www.csse.monash.edu/webb/-Files/WebbConilione06.pdf>)*, 2005.

- [30] Xue Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019.
- [31] Son Ha Ngo. Komplex szénhidrogénelegyek modellezése folytonos termodinamikával. 2005.
- [32] Ramakrishnan Muthukrishnan and R Rohini. Lasso: A feature selection technique in predictive modeling for machine learning. In *2016 IEEE international conference on advances in computer applications (ICACA)*, pages 18–20. Ieee, 2016.
- [33] Chris Hans. Bayesian lasso regression. *Biometrika*, 96(4):835–845, 2009.
- [34] Youjuan Li and Ji Zhu. L 1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 17(1):163–185, 2008.
- [35] Daniel Berrar et al. Cross-validation., 2019.
- [36] Bhaskar Protim Mahanta, Pranjit Kumar Bora, Phirose Kemprai, Gitasree Borah, Mohan Lal, and Saikat Halder. Thermolabile essential oils, aromas and flavours: Degradation pathways, effect of thermal processing and alteration of sensory quality. *Food Research International*, 145:110404, 2021.
- [37] GN Sapkale, SM Patil, US Surwase, and PK Bhatbhage. Supercritical fluid extraction. *Int. J. Chem. Sci*, 8(2):729–743, 2010.
- [38] Valtcho D Zheljazkov, Tess Astatkie, and Alexander N Hristov. Lavender and hyssop productivity, oil content, and bioactivity as a function of harvest time and drying. *Industrial Crops and Products*, 36(1):222–228, 2012.
- [39] Ameneh Tarakemeh, Vahid Rowshan, and Sharareh Najafian. Essential oil content and composition of *lavandula angustifolia* mill. as affected by drying method and extraction time. *Analytical Chemistry Letters*, 2(4):244–249, 2012.