

# A KLÍMAVÁLTOZÁS ÁLLAT- TENYÉSZTÉSI VONATKOZÁSAI

MATEMATIKAI MODELLEK, AI ÉS A DOLGOK INTERNETE A SZARVAS-  
MARHÁK METÁNTERMELÉSÉNEK MEGHATÁROZÁSÁBAN II.

**Szakértő  
munkatársunk írása**  
Állattenyésztési  
Teljesítményvizsgáló Kft.

A különféle modellek – ahogyan cikkünk első részében bemutattuk – fontos szerepet játszanak a szarvasmarhák enterális metán- ( $\text{CH}_4$ -) termelésének becslésében. Egyes típusaik azonban ennél szélesebb körű alkalmazást kínálnak: lehetővé teszik az állatok kibocsátási szintek szerint csoportosítását, a  $\text{CH}_4$ -csökkentést célzó – például takarmányozási – módszerek hatékonyságának értékelését, valamint a szokatlan kibocsátási mintázatok azonosítását, így mára értékes eszközökké váltak mind a kutatók, mind a gazdálkodók számára.

Az előző számunkban megkezdett áttekintést folytatva, írásunk második részében három modelltypussal foglalkozunk. Ezek különböző megközelítésekkel számszerűsítik a szarvasmarhák  $\text{CH}_4$ -kibocsátását. Elsőként a szakirodalomban közzétett empirikus modellek közül ismertetünk néhányat, majd röviden kitérünk a telepszintű modellekre, melyek átfogó képet nyújtanak az állattartó gazdaságok kibocsátásairól és azok forrásairól, elősegítve a rendszerszintű elemzést és optimalizálást. Végül az állattenyésztésben innovációnak számító mesterséges intelligencia

(artificial intelligence, AI) által alkotott modelleket vesszük górcső alá. Az AI-alapú megközelítések – különösen a gépi tanulási algoritmusok – áttörést jelentenek a  $\text{CH}_4$ -kibocsátás becslésében és mérséklésében, mivel képesek túllépni a matematikai modellek korlátain: gyorsan és pontosan dolgozzák fel a nagy adatállományokat, feltárják a bennük rejlő nemlineáris összefüggéseket, és precízebb előrejelzéseket adnak. Ráadásul az új adatok folyamatos integrálásával állandóan finomítják is a számításait, tovább növelve ezáltal a modellek hatékonyságát és alkalmazhatóságát.



## Empirikus megoldások a szarvasmarhák CH<sub>4</sub>-termelésének becslésére

Jelen írásban több empirikus modellt gyűjtöttünk egybe, amelyeket különféle összetételű és adagú takarmányok etetése mellett végzett CH<sub>4</sub>-mérések alapján fejlesztettek ki. (Lásd a táblázatot.) E modellek az egyszerű statisztikai megközelítéseknek köszönhetően könnyen alkalmazhatók a kidolgozásukhoz hasonló körülmények között. Ugyanakkor, mint más esetekben, ezeknél is szembesülnünk kell bizonyos korlátokkal. Jelentős hátrányuk például, hogy figyelmen kívül hagyják a takarmánybontási folyamatok mögött húzódó élettani mechanizmusokat, ami korlátozhatja a prognózisuk pontosságát. Továbbá, ha a modellépítéskor fennálló feltételekhez képest eltérő viszonyok között használjuk őket, több kritériumnak is teljesülnie kell a megbízható becslési eredmények érdekében. Ezek közé tartozik például a számításokhoz szükséges adatok teljes körű elérhetősége, az új adatok értéktartományának az eredeti adatállományéhoz való illeszkedése, valamint a korrelációs mintázatok hasonlósága a kiindulási és az új adathalmazokban.



Ahogy a táblázatban is látható, az itt bemutatott modellek inputváltozói, valamint az outputjuk (a szarvasmarhák CH<sub>4</sub>-termelésének) mértékegységei meglehetősen változatosak. A legtöbb kutatócsoport a szarvasmarhák nagy légzőkamrával mért, teljes enterális CH<sub>4</sub>-termelését (azaz a naponta kilélegzett,

felbőfögött és a végbélen át távozó CH<sub>4</sub> mennyiségét) vette alapul, ugyanakkor mások – például Ellis és mtsai. (2007), illetve Niu és mtsai. (2021) – alternatív mérési módszerekkel (a nagy légzőkamrák mellett/helyett fejkamrákkal, kén-hexafluoridos [SF<sub>6</sub>-os] nyomjelző technikával és/vagy GreenFeed-készülékkel) gyűjtött, „vegyes” adatokat használtak az egyenletek kidolgozásához. Bár az alternatív mérési eljárások többféle adatszerzési lehetőséget kínálnak, az eltérő módszertani háttérű adatok változó pontossága, nagyfokú heterogenitása és nehéz összehasonlíthatósága végső soron rontja a prognózisok megbízhatóságát. További kihívást jelent, hogy számos modell csupán korlátozott adatkészletre támaszkodik (például néhány állatra, egyetlen állományra vagy meghatározott étrendre vonatkozó mérési eredményekre; lásd a táblázat mintaelemszámot bemutató, utolsó oszlopát [n]), míg mások különböző tanulmányok metaanalízisére épülnek, ami tovább növelheti a becslések bizonytalanságát. Ráadásul a modellek túlnyomórészt egyszerű lineáris regressziós elemzésekben alapulnak, amelyek nem képesek megragadni az adatok közötti nemlineáris összefüggéseket. Ez torzítást vihet az előrejelzésekbe, különösen akkor, ha az inputadatok összetettebb kapcsolatokat jellemezhetők.





Táblázat: Néhány empirikus modell a szarvasmarhák CH<sub>4</sub>-kibocsátásának előrejelzésére

Forrás	Modell	R <sup>2</sup>	n
Moe és Tyrrell (1979) – holstein-fríz tehenek adatai alapján	CH <sub>4</sub> (teljes enterális; MJ/nap) = 0,341 + 2,65 × felvett cellulóz (kg/nap) + 1,74 × felvett hemicellulóz (kg/nap) + 0,51 × felvett NFC (kg/nap)	0,670	404
Mills és mtsai. (2003) – laktáló tejhasznú tehenek adatai alapján (fajta nem megnevezett)	CH <sub>4</sub> (teljes enterális; MJ/nap) = 5,93 + 0,92 × DMI (kg/nap);	0,600*	159
	CH <sub>4</sub> (teljes enterális; MJ/nap) = 8,25 + 0,07 × MEI (MJ/nap);	0,550	159
	CH <sub>4</sub> (teljes enterális; MJ/nap) = 7,30 + 13,13 × nitrogén (kg/nap) + 2,04 × ADF (kg/nap) + 0,33 × keményítő (kg/nap);	0,570	159
	CH <sub>4</sub> (teljes enterális; MJ/nap) = 1,06 + 10,27 × tömegtakarmány-arány (tömegtakarmány-DMI / takarmány-DMI) + 0,87 × DMI (kg/nap);	0,610	64
	Nem lineáris (Mitscherlich 3) modell: CH <sub>4</sub> (teljes enterális; MJ/nap) = 45,98 – 45,98e <sup>-c(ME)</sup> , ahol c = -0,0011 × $\frac{\text{keményítő} (\frac{\text{kg}}{\text{nap}})}{\text{ADF} (\frac{\text{kg}}{\text{nap}})} + 0,0045$	0,970	159
Hindrichsen és mtsai. (2005) – tejhasznú tehenek adatai alapján (fajta nem megnevezett)	CH <sub>4</sub> (teljes enterális; g/nap) = 17,1 × DMI (kg/nap) + 97,4;	0,842	35
	CH <sub>4</sub> (teljes enterális; g/nap) = 91 + 50 × lebontott cellulóz (kg/nap) + 40 × lebontott hemicellulóz (kg/nap) + 24 × lebontott keményítő (kg/nap) + 67 × lebontott cukrok (kg/nap);	0,843	
	Összes (teljes enterális + hígrágyából származó) CH <sub>4</sub> (g/nap) = 108 + 46 × lebontott NDF (kg/nap) + 30 × lebontott keményítő (kg/nap) + 85 × lebontott cukrok (kg/nap)	0,832	
Yan és mtsai. (2006) – holstein fríz tehenek adatai alapján	CH <sub>4</sub> (teljes enterális; l/nap) = 0,34 × testtömeg (kg) + 19,7 × DMI (kg/nap) + 12;	0,770	315
	CH <sub>4</sub> (teljes enterális; l/nap) = 47,8 × DMI (kg/nap) – 0,76 × DMI <sup>2</sup> – 4		
Ellis és mtsai. (2007) – metaanalízis, tej- és húshasznú szarvasmarhák adatai alapján (fajta nem megnevezett)	CH <sub>4</sub> (MJ/nap) = 3,23 + 0,809 × DMI (kg/nap)	0,650	89
	CH <sub>4</sub> (MJ/nap) = 2,16 + 0,493 × DMI (kg/nap) – 1,36 × ADF (kg/nap) + 1,97 × NDF (kg/nap)	0,630	
	CH <sub>4</sub> (MJ/nap) = 0,139 × takarmányadagon belüli tömegtakarmány-arány (%) + 8,56	0,560	
Jentsch és mtsai. (2007) – tehenek, üszők, fiatal bikák adatai alapján (fajta nem megnevezett)	CH <sub>4</sub> (teljes enterális; kJ/nap) = 1,62 × emészthető CP (g/nap) – 0,38 × emészthető CFA (g/nap) + 3,78 × emészthető CF (g/nap) + 1,49 × emészthető NFE (g/nap) + 1 142 (g/nap);	0,896	337
	CH <sub>4</sub> (teljes enterális; kJ/nap) = 1,28 × emészthető CP (g/nap) – 0,31 × emészthető CFA (g/nap) + 1,31 × emészthető keményítő (g/nap) + 1,16 × emészthető cukor (g/nap) + 2,40 × emészthető NFE (g/nap) + 1 835	0,889	
Nielsen és mtsai. (2013) – tejhasznú tehenek adatai alapján (fajta nem megnevezett)	CH <sub>4</sub> (MJ/nap) = 1,23 × DMI (kg/nap) – 0,145 × zsírsavtartalom (g/kg DM) + 0,012 × NDF (kg/nap)	0,750	211
Engelke és mtsai. (2018) – holstein-fríz tehenek adatai alapján	CH <sub>4</sub> (teljes enterális; l/nap) = 361,4 + 18,9 × DMI (kg/nap) + 28,5 × C18:0 (zsírsav%) – 23,6 × cisz-1-C18 (zsírsav%);	0,820	36
	CH <sub>4</sub> (teljes enterális; l/nap) = -1 364 + 9,58 × ECM (kg/nap) + 18,5 × SFA (tejzsír%) + 32,4 × C18:0 (zsírsav%)	0,790	
Niu és mtsai. (2021) – metaanalízis, tej- és húshasznú szarvasmarhák adatai alapján (fajta nem megnevezett)	CH <sub>4</sub> (MJ/nap) = 1,13 × DMI (kg/nap) – 0,114 × zsírsavtartalom (g/kg DM) + 0,017 × NDF (kg/nap)	0,694**	99

\* Korrigált R<sup>2</sup>-érték; \*\* konkordanciakorrelációs együttható, amely az R<sup>2</sup>-hez hasonlóan azt mutatja, hogy mennyire egyeznek meg a modell által becsült CH<sub>4</sub>-kibocsátási értékek a tényleges mérési eredményekkel, de emellett figyelembe veszi a közöttük levő eltérések mértékét és az adatok szórását is.

Megjegyzés: Ahol a „teljes enterális” kifejezés nincs feltüntetve, az egyenletek vegyes adatokon alapulnak (tehát azokat nem [kizárólag] nagy légzőkamrával,

hanem fejkamrával, SF<sub>6</sub>-os nyomjelző technikával és/vagy GreenFeed-készülékkel [is] mérték). NFC: nem rostjellelű szénhidrátok; DMI: szárazanyag-felvétel; MEI: metabolizálhatóenergia-felvétel; ADF: savdetergens rost; ME: metabolizálható energia; NDF: neutrális detergens rost; CP: nyersfehérje; CFA: nyerszsír; CF: nyersrost; NFE: nitrogénmentes kivonat; DM: szárazanyag; ECM: energiakorrigált tej; SFA: telített zsírsavak.



Ahogy a rovat korábbi cikkeiben bemutattuk, az egyes szénhidráttípusok eltérő mértékben járulnak hozzá a  $\text{CH}_4$ -kibocsátáshoz: a cellulóz és a hemicellulóz lebontása során több  $\text{CH}_4$  keletkezik, mint a könnyebben emészthető keményítő vagy cukrok esetében. Hindrichsen és mtsai.-nak (2005) táblázatban szereplő egyenletei azonban látszólag ellentmondanak ennek a megállapításnak, hiszen a cukrokhoz tartozó együtthatók (67 és 85) nagyobbak, mint a cellulóz és a hemicellulóz (2. egyenlet), valamint az NDF (3. egyenlet) együtthatói. E magas értékek a cukrok rostemésztést serkentő hatására utalnak, rávilágítva arra, hogy a  $\text{CH}_4$ -termelést nemcsak az egyes tápanyagfrakciók mennyisége, hanem a közöttük fennálló kölcsönhatások is jelentősen befolyásolják. Éppen ezért különösen fontos, hogy a kibocsátásra vonatkozó becslések a teljes takarmányadagot vegyék alapul, és ne csupán az egyes összetevőkre korlátozódjanak.

A nemlineáris modellek egyik fő előnye, hogy szélesebb takarmányfelvételi tartomány és változatos étrendek esetén is jobban teljesítenek a lineáris társaikhoz képest, ugyanakkor lényegesen nagyobb számítási kapacitást igényelnek. Példaként említhetjük ezekre a táblázatban szereplő Mitscherlich 3-modellt, amelyet Mills és mtsai. (2003) dolgoztak ki Egyesült Királyságból

származó adatok alapján, más modellváltozatokkal (például Mitscherlich 1, 2) együtt. A Mitscherlich 3 különösen figyelemre méltó, mivel fejlett matematikai struktúrájának köszönhetően képes elkerülni azokat a torzításokat, amelyek más (lineáris és nemlineáris) modelleknél gyakran előfordulnak, különösen akkor, ha az állatok keményítő-, ADF- és ME-felvétele túllépi a modellfejlesztés során használt értéktartományt. Ennek ellenére e modell előrejelző képessége nem feltétlenül bizonyul jobbnak, mint a többszörös regressziós modelleké (Mills és mtsai., 2001, 2003), amelyek egyszerűségük és robusztusságuk révén továbbra is versenyképes alternatívát jelentenek, különösen a regressziós elemzések feltételeihez jól illeszkedő, homogén adatkörnyezetben.



## Telepszintű matematikai modellek

Külön figyelmet érdemelnek a tejtermelő gazdaságok anyag-, energia- és üvegházhatásúgáz- (ÜHG-) mérlegének meghatározására használt komplex matematikai modellek. Ezek a tömegmérlegmódszeren alapulnak, amelynek célja a telep minden inputjának (például takarmány, víz, energia) és outputjának (állati termékek, trágya, elfolyások, kimosódások, gázkibocsátás stb.) pontos számszerűsítése. Az ilyen típusú modellek kidolgozása jelentős idő-, munka- és költségbefektetéssel jár, mivel a rendszer összes elemének és folyamatának nyomon követése rendkívül összetett feladat. A tömegmérlegmódszer a termelékenység vizsgálatán túl a környezeti hatások átfogó értékelését is lehetővé teszi. Segítségével pontos becslések készíthetők például a telepek szénlábnyomáról, ami megalapozza az emissziócsökkentést célzó döntéshozatalt. A környezetbe jutó gázok mennyisége az outputokból a szilárd és a folyékony komponensek levonásával számítható ki.

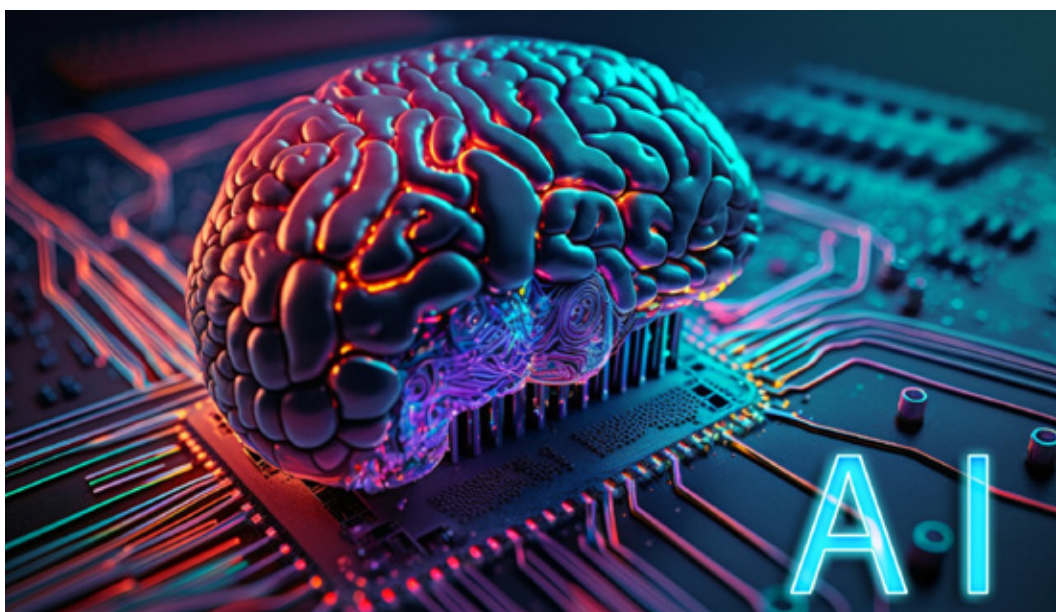
A tehenészetekben jelenleg alkalmazott modellek (például az Egyesült Államok Agrárminisztériumának Mezőgazdasági Kutatószolgálatá által kifejlesztett DairyGEM, a több ízben kibővített Integrated Farm System Model [Integrált Farmrendszermodell, Rotz és mtsai., 2022], az észak-floridai DyNoFlo [Cabrera és mtsai., 2006], az egyesült királysági SIMSDAIRY [Del Prado és mtsai., 2011] stb.) közül azonban több is különféle strukturális vagy alkalmazhatósági korlátokkal küzd. Egyesek például csak meghatározott földrajzi régiókban használhatók hatékonyan, míg mások kódtára nem támogatja új modulok integrálását, ami megnehezíti a precíziós mezőgazdasági szenzoradatok feldolgozását. E kihívások leküzdésére olyan új generációs modellek fejlesztésére van szükség, amelyek rugalmasan alkalmazkodnak a változatos környezeti feltételekhez, valamint képesek befogadni és feldolgozni a különböző típusú szenzoradatokat.



## Mesterséges intelligencia által alkotott modellek

Az AI-alapú technikák – különösen a gépi tanulási algoritmusok – napjainkban egyre meghatározóbb szerepet töltenek be a modellezésben. A modern számítógépes rendszerek hatékonyan tudják feldolgozni az „okosfarm-technológiák”, a dolgok internete (internet of things, IoT), valamint az új, fenotípusos teljesítményeket rögzítő platformok által generált hatalmas, heterogén adatállományokat,

és az elemzések során azonosítani tudják azokat a rejtett összefüggéseket, mintázatokat, trendeket, melyeket a hagyományos módszerekkel nehéz vagy lehetetlen kimutatni. Ennek eredményeként az AI-alapú előrejelzések nemcsak pontosabbak lehetnek a hagyományos matematikai modellek által nyújtottaknál, de új, eddig ismeretlen összefüggések feltárására is alkalmasak.



**A gépi tanulás** az AI egyik legsokoldalúbb alkalmazási területe, amely lehetővé teszi, hogy az algoritmusok önállóan felismerjék az adatállományokban rejlő mintázatokat és szabályszerűségeket, majd ezekből tanulva folyamatosan fejlődjenek. Az ilyen rendszerek emberi beavatkozás nélkül is képesek egyre pontosabb előrejelzéseket, osztályozásokat vagy döntéseket hozni, jelentősen növelve ezáltal a modellek megbízhatóságát és hatékonyságát.

Az AI-alapú technikák, bár még fejlesztés alatt állnak, már most át tudják hidalni a hagyományos matematikai modellek korlátait, mivel ötvözik az empirikus megközelítések sokoldalúságát a mechanisztikus modellek pontosságával. Az AI által alkotott modellek teljesítménye ugyanakkor nagyban függ attól, hogy a betanítási és a validálási fázisokban rendelkezésre állnak-e kellően nagy, konzisztens, a vizsgált folyamatokat jól reprezentáló adatkészletek. (A betanítás során az algoritmusokat olyan adathalmazokkal látják el, amelyek alapján megtanulják felismerni a releváns mintázatokat, míg a validálási fázis célja a modellek általánosíthatóságának és pontosságának tesztelése új, a betanításban nem szereplő adatokkal.)

Egyes AI-alapú modellek alkalmazása technikai kihívásokkal jár. Ezek közé tartozik például a „klasszikus”

gépi tanulási folyamatok sikeréhez elengedhetetlen adat-előkészítés és szabványosítás, mely műveletek a különféle forrásokból származó adatok egységesítését és integrálhatóságának biztosítását szolgálják. Emellett az AI-alapú modellek értelmezhetősége és átláthatósága is kiemelt figyelmet érdemel, mivel ezek közvetlen hatással vannak a gyakorlati alkalmazhatóságukra és a becslési eredményeik pontosságára.





Negussie és mtsai. (2022) tejelő szarvasmarhák  $CH_4$ -kibocsátásának proxyadatokon alapuló előrejelzése során a gépi tanuláson alapuló véletlen erdő (random forest, RF) modellek és a többszörös lineáris regressziós (empirikus) modellek hatékonyságát vizsgálták, továbbá felmérték, hogy miként befolyásolja a hiányzó adatok pótlása (imputálása) az előrejelzések pontosságát. A kutatás során 10 ország 20 szarvasmarha-állományának  $CH_4$ -kibocsátási és proxyadatait gyűjtötték össze, ami egy 43 519 rekordból álló, integrált adathalmazt eredményezett, 3 483 tehén adataival. Az elemzéshez három különböző méretű adatkészletet állítottak össze: a) egy 3 337 rekordból állót, ahol nem volt hiányzó adat; b) egy 21 215 rekordból állót, ahol a hiányzó DMI-adatokat a meglévő tej-, tejsír-, tejféhérjetermelési és testtömegadat alapján pótolták; valamint c) egy 40 532 rekordos készletet, ahol a hiányzó DMI-, tejsír- és tejféhérjetermelési adatokat egyaránt imputálták. Minden adatkészlet 11 különböző  $CH_4$ -kibocsátási proxyváltozót tartalmazott, beleértve a fajtát [5 kategória], az országot [10 kategória], az állományt [20 kategória], a laktációs napok számát, a laktációk számát [3 kategória], a  $CH_4$ -mérési technikát [5 kategória], a DMI-t, a testtömeget, valamint a tej-, tejsír- és tejféhérjetermelést. A modellezést különböző forgatókönyvek alapján végezték, például DMI-vel vagy anélkül, illetve imputált és nem imputált DMI-vel, tejsírral és tejféhérjével. Az eredmények szerint a DMI-adatok bevonása (akár imputációval is) javította az előrejelzések pontosságát. Az RF-modellek minden esetben felülmúlták az empirikus modelleket, mivel rugalmasabbak azoknál, és jobban kezelik az összetett, heterogén adatszerkezeteket. Az adatállományok méretének növelésével azonban nemcsak a becslések pontossága javult, de mérséklődött az RF-modellek előnye is. A tanulmány rámutat arra, hogy a fejlett imputációs technikák és a gépi tanulási algoritmusok alkalmazása ígéretes lehetőséget kínál robusztus  $CH_4$ -előrejelzési modellek kidolgozására a tejtermelő gazdaságok rutinszerűen mért adatainak felhasználásával.

Shadpour és mtsai.-nak (2022) kutatása két fő kérdést vizsgált: 1. A tehének  $CH_4$ -kibocsátásának előrejelzése: Mennyire használhatók a tehéntej összetételére vonatkozó középínfravörös spektroszkópiás (middle infrared spectroscopy, MIRS) adatok önállóan vagy a tejtermelési mutatókkal (tej-, tejsír- és tejféhérjetermeléssel) kombinálva? 2. Modellek összehasonlítása: Hogyan teljesítenek a lineáris és

nemlineáris mesterséges neurális hálózatok (artificial neural network, ANN) a  $CH_4$ -kibocsátás becslésében a részleges legkisebb négyzetek (partial least squares, PLS) elvére épülő, regressziós (matematikai) modellekhez képest?

A vizsgálatban 158 kanadai és 44 dán holstein-fríz tehén  $CH_4$ -kibocsátását mérték GreenFeed-készülékkel és snifferekkel (Kanada: 181 hét; Dánia: 217 hét időtartamban). Az így kapott heti átlagos kibocsátási értékeket tejösszetételre vonatkozó tejspektrumrekordokhoz, teszt nap (befejési) tej-, tejsír- és tejféhérjetermeléshez, illetve egyéb adatokhoz társították.

A kutatók a tehének napi átlagos  $CH_4$ -emissziójának becslésére először lineáris és nemlineáris ANN-eket alkalmaztak, nyolc változókészletet tesztelve: 1. – tejtermelés, ellési kor (age at calving, AC) és tejelő napok száma (days in milk, DIM); 2. – tejsírtermelés, AC és DIM; 3. – tejféhérjetermelés, AC és DIM; 4. – tej- és tejsírtermelés, AC és DIM; 5. – tej- és tejféhérjetermelés, AC és DIM; 6. – tej-, tejsír- és tejféhérjetermelés, AC és DIM; 7. – MIRS, AC és DIM; valamint 8. – tej-, tejsír- és tejféhérjetermelés, MIRS, AC és DIM. Mindegyik változókészletben szerepelt az ország, az ellési időszak és a laktációk száma is. Ezt követően Shadpour-ék az állatok  $CH_4$ -kibocsátását PLS-regresszióval is megbecsülték. Főbb eredményeik a következők:

- Az előrejelzés pontossága **kizárólag a tejtermelés** alapján alacsony ( $r = 0,245-0,457$ ), a RMSE pedig magas (87,28–99,39) volt. (A Pearson-féle korrelációs együttható [ $r$ ] nagysága [0–1] a vizsgált változók közötti kapcsolat erősségét fejezi ki, míg a RMSE [root mean square error, azaz az átlagos négyzetes hiba négyzetgyöke] a modell-előrejelzések és a tényleges értékek közötti eltérést számszerűsíti.)
- A **tejsírtermelés bevonása** némileg növelte a pontosságot ( $r = 0,288-0,491$ ; RMSE = 85,94–98,04), de a **tejféhérjetermelés hozzáadása** nem hozott javulást.
- A **tej- és a tejsírtermelés kombinációja** ( $r = 0,272-0,469$ ; RMSE = 87,21–100,76) csak minimális előnyt jelentett.
- A **MIRS-adatok adták**, akár önállóan, akár más adatokkal kombinálva, **a legjobb eredményeket** ( $r = 0,586-0,717$ ; RMSE = 69,09–96,20).



Amikor a változóképzletek nem tartalmaztak MIRS-adatokat, a nemlineáris ANN-modellek pontosabb előrejelzést nyújtottak, mint a lineáris ANN-ek és a PLS-regresszió. Ez arra utal, hogy a CH<sub>4</sub>-kibocsátás és a tejtermelési jellemzők között nemlineáris kapcsolat áll fenn, amely korlátozza a lineáris modellek hatékonyságát. Ugyanakkor a MIRS-adatok bevonása jelentősen javította a lineáris modellek teljesítményét, így azok pontossága megközelítette a nemlineáris ANN-modellekét.



A **mesterséges neurális hálózatok** (artificial neural networks, **ANN-ek**) a gépi tanulás egy speciális formáját képviselik, amely az emberi agy működését utánozza. Ezek a hálózatok rétegekbe rendezett, egymáshoz kapcsolódó csomópontokból, ún. neuronokból állnak. A rétegek közül a bemeneti réteg gyűjti össze az adatokat (például a CH<sub>4</sub>-termelést befolyásoló tényezőket), a rejtett réteg(ek) végzi(k) az adatfeldolgozást, a kimeneti réteg pedig előrejelzéseket ad (esetünkben a CH<sub>4</sub>-kibocsátásra vonatkozóan). Minden neuron csak a szomszédos rétegek neuronjaitól fogad inputokat, amelyeket matematikai függvények segítségével dolgoz fel, majd a kapott eredményeket továbbítja a következő réteg felé (kivéve az utolsó réteg neuronjait). Ezt a folyamatot – ahol a bemeneti adatok végighaladnak a hálózaton, és a kimeneti réteg megadja az előre jelzett értéket – előterjesztésnek (forward propagation) nevezzük.

A modell pontossága nagymértékben függ a neuronok közötti kapcsolatok erősségét meghatározó súlyok és az eltolási értékek (bias-ok) megfelelő beállításától, amelyre a betanítási fázis kezdetén, véletlenszerűen kerül sor. (Az eltolási értékek hozzáadódnak az inputok és a súlyok szorzatának összegéhez, elősegítve ezzel az ANN-ek rugalmasabb és pontosabb tanulását.) Az előterjesztést követően a rendszer az ún. visszaterjesztés (backward propagation) módszerével folyamatosan finomítja ezeket a paramétereket, minimalizálva a tényleges és a becsült eredmények közötti eltéréseket. Ez az iteratív folyamat biztosítja, hogy a modell fokozatosan tanuljon, és egyre pontosabb előrejelzéseket készítsen. (Siki, 2024)

Az ANN-ek előnye, hogy hatékonyan kezelik a nagyméretű és komplex adathalmazokat, mivel képesek az adatmintázatok felismerésére és az összefüggések elsajátítására.

A **mélytanulás** az ANN-ek továbbfejlesztett változata, amely többrétegű hálózatokat használ, lehetővé téve az adatok még összetettebb és mélyrehatóbb elemzését. Ez a folyamat nagyrészt automatizált, így nem igényli a klasszikus gépi tanulásnál megszokott, időigényes előzetes adatfeldolgozást (strukturálást és címkézést). Ennek eredményeként az algoritmusok nyers, strukturálatlan adatok – például képek, szövegek és videók – azonnali feldolgozására is képesek, miközben önállóan azonosítják az előrejelzések szempontjából legfontosabb jellemzőket. Ez a rugalmasság különösen hasznos olyan területeken, mint a CH<sub>4</sub>-kibocsátás becslése, ahol komplex és heterogén adatokkal kell dolgozni.

A gépi tanulással fejlesztett modellek nemcsak a CH<sub>4</sub>-termelés előrejelzésére alkalmasak, hanem a tejelő szarvasmarhák CH<sub>4</sub>-kibocsátási szint szerinti csoportosítására is. Zheng és mtsai. (2016) egy olyan gépi tanulási módszert, a Bayes-hálót használták a célra, amely aciklikus gráfok segítségével modellezi az inputváltozók, ez esetben az energiakorrigált tejhozam, a takarmányozás és a szarvasmarhafajták közötti valószínűségi kapcsolatokat. A kutatók megállapítása szerint a Bayes-hálóra épülő modell 66%-os

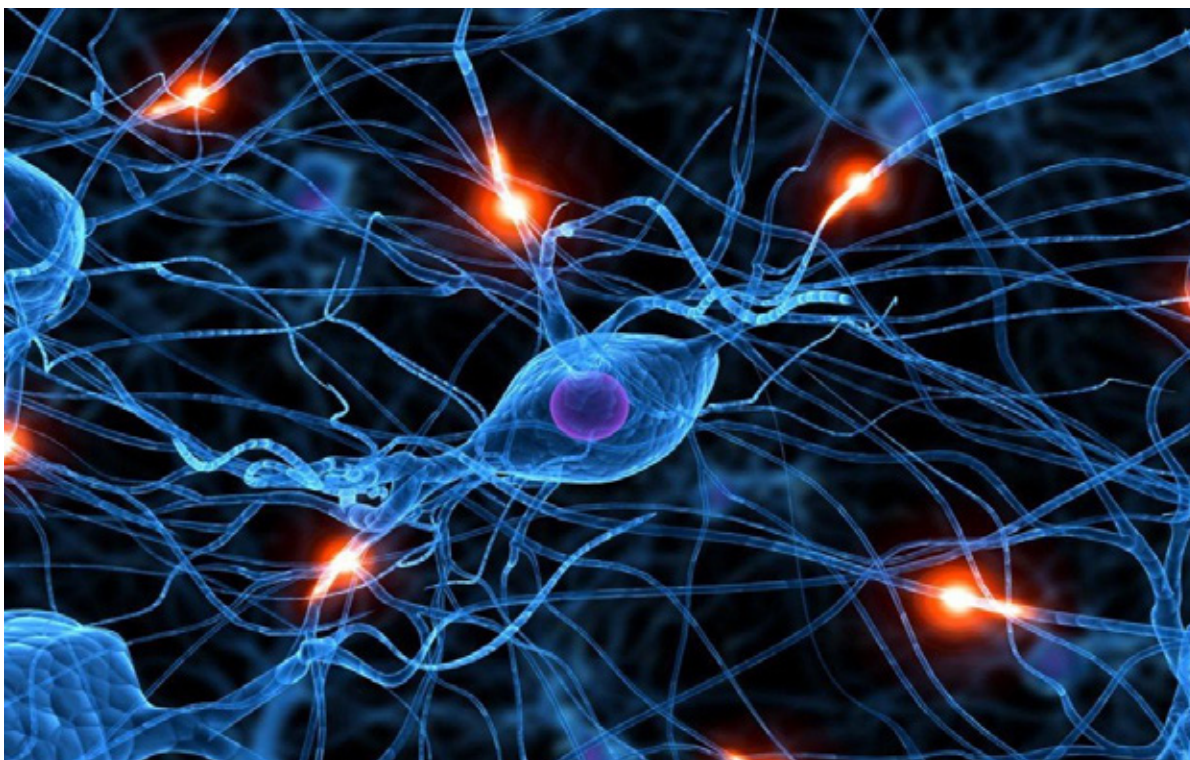
pontossággal sorolta be a teheneket a megfelelő CH<sub>4</sub>-kibocsátási kategóriákba, de kevésbé bizonyult hatékonynak az alacsony kibocsátású csoportba tartozó egyedek azonosításában, amelyek a vizsgált minta mindössze 5%-át tették ki. Ez rávilágít a gépi tanulási osztályozási modellek egyik hátrányára: ha egy kategória alulreprezentált, a modellek csak nehezen sajátítják el az annak megkülönböztetéséhez szükséges jellemzőket, ami az osztályozási pontosság csökkenéséhez vezet.



Jeong és mtsai. (2022) tejtermelő szarvasmarhák mélytanulási modellel becsült  $\text{CH}_4$ -kibocsátását két, Egyesült Államokban használt referenciamodell, a CALGEM és a Vista-CA adataival hasonlították össze a kaliforniai San Joaquin-völgyben. A CALGEM-et és a Vista-CA-t Kalifornia Állam Légszennyezés-ellenőrzésért Felelős Hivatala fejlesztette ki az ÜHG-, különösen a  $\text{CH}_4$ -kibocsátások részletes tér- és időbeli elemzésére, az emissziós források nyomon követésére és a környezetvédelmi döntéshozatal támogatására. A két modell alkalmazási területe ugyanakkor eltérő: míg a CALGEM a légszennyezés átfogó monitorozására szolgál, addig a VISTA-CA elsősorban a közlekedésből származó emissziók értékelésére és a közlekedési infrastruktúra fenntarthatóságának elemzésére irányul. A kutatók mélytanulási modellje által becsült kibocsátások erős korrelációt mutattak a CALGEM ( $r = 0,80$ ) és a Vista-CA adataival ( $r = 0,96$ ). Az AI-modell azonban besorolási hibákat követett el: a Nemzeti Mezőgazdasági Képfeldolgozó Program keretében készült légi felvételek „kis emissziójú” pixeleit némelykor túlszámlálta, a „nagy emissziójú”

pixeleket pedig alulbecsülte. Emellett hibásan azonosította egyes nem szarvasmarhatartó területek emisszióit is, így némely „nem mezőgazdasági” pixeleket tévesen „mezőgazdasági” pixelként kezelte. Ezek a pontatlanságok feltehetően a légi felvételek alacsony száma és esetenkénti gyengébb minősége (homályos, kis felbontású vagy torzított képek), valamint a betanításhoz rendelkezésre álló korlátozott adatmennyiség miatt jelentkeztek.

Az ANN-modellek – különösen a mélytanulási típusok – nemcsak az előbb említett hátránnyal küzdenek, de hajlamosak a túlillesztésre is, vagyis a betanítási adatkészlet mintáinak „túlzott” elsajátítására. Ilyenkor a véletlenszerű ingadozásokat is megtanulják, ami ronthatja a  $\text{CH}_4$ -termelésre vonatkozó előrejelzések pontosságát új, ismeretlen adathalmazokon. Szerencsére több módszer is létezik e probléma kiküszöbölésére; ezek közé tartozik például az újramintavételezés, a validáló adatkészletek használata, a súlyok validálása, illetve a rejtett rétegekben levő neuronok számának optimalizálása.



Gépi tanulási modellek esetén a már említett imputáció – tehát a hiányzó adatok proxy- vagy becsült értékekkel való helyettesítése – is kihívásokat okozhat. Ez a gyakran használt technika lehetővé teszi, hogy a modellek kibővített, részletesebb adathalmazokkal dolgozzanak, és így valóságosabb szimulációkat, illetve pontosabb előrejelzéseket nyújtsanak. A helytelenül végrehajtott imputáció azonban torzíthatja az

adatokat, ami téves következtetésekhez vezethet. Ezért különös gondot kell fordítani a helyettesítő adatok és az imputációs módszerek alapos validálására azok ténylegesen alkalmazása előtt.

A felhasznált források listáját a cikk terjedelmi korlátai miatt nem közöljük, az a szerkesztőségben érhető el.

