

Doktori (PhD) értekezés tézisei

**Integrált módszertanok az adatvezérelt szoftver szenzorok  
fejlesztésére**

Hanzelik Pál Péter

Pannon Egyetem

Vegyésmérnöki és Anyagtudományok

Doktori Iskola

Témavezetők:

Prof. Dr. Abonyi János

Dr. Kummer Alex



Folyamatmérnöki Intézeti Tanszék

Veszprém

2026



# 1. Bevezetés és a munka célja

A mai ipari környezetet egyre inkább a digitalizáció és az Ipar 4.0 megoldások integrációja határozza meg. Ezek a technológiák már nem csupán opcionális fejlesztések, hanem a piaci versenyképesség fenntartásának és a környezeti fenntarthatóság elérésének elengedhetetlen előfeltételei. Az intelligens gyárak hálózatba kapcsolt gépeket és Internet of Things (IoT) érzékelőket használnak az adatalapú döntéshozatal megkönnyítésére, ami fokozott működési rugalmasságot, csökkentett energiafogyasztást és minimalizált egészségügyi, biztonsági és környezeti (HSE) kockázatokat eredményez.

A vegyipar speciális kontextusában az elsődleges kihívás a nyersanyagok és termékek valós idejű minősítése. A hagyományos laboratóriumi minőségbiztosítás gyakran jelentős késleltetést okoz, ami megakadályozza a folyamatok ingadozásaira való gyors reagálást. Következésképpen a valós idejű adatelemzésre és az automatizált folyamatirányításra való átállás elengedhetetlen a termelékenység optimalizálása és a váratlan leállások megelőzése érdekében.

## Soft Szenzorok és Prediktív Modellezés

A kutatásnak egyik alapköve a soft szenzorok (lágy érzékelők) fejlesztése és alkalmazása. Ezek a virtuális műszerek adatvezérelt matematikai modellekként működnek, amelyek olyan kritikus folyamatparamétereket becsülnek meg – mint például a termék összetétele, viszkozitása vagy minősége –, amelyek fizikai hardverrel történő mérése egyébként nehéz, veszélyes vagy költséges lenne. Könnyen mérhető változók (pl. nyomás, hőmérséklet vagy spektroszkópiai adatok) felhasználásával a soft szenzorok nagyfrekvenciás becsléseket nyújtanak, amelyek szigorúbb folyamatirányítást és optimalizálást tesznek lehetővé.

Ezeknek a modelleknek a teljesítménye szorosan összefügg a gépi tanulás (ML) algoritmusok hatékonyságával. Míg az ML forradalmasítja az ipari előrejelzéseket, hosszú távú ipari hasznosíthatósága a robusztus adatkezelésen és az életciklus-monitorozáson múlik. Ez a kutatás a gépi tanulás műveleteinek (MLOps) módszertanát hangsúlyozza, amely magában foglalja az algoritmus teljes életciklusát – a fejlesztéstől és az automatizált telepítéstől a folyamatos teljesítmény nyomon követéséig és skálázásáig.

## **Kutatási Célok és Módszertan**

A disszertáció átfogó célja a soft szenzorok megbízhatóságának és robusztusságának javítása integrált Ipar 4.0 megoldások segítségével. A kutatás az ipari adattudomány használatával vizsgálja a hierarchikus adatok konzisztenciáját, az adatok fúzióját, a szintetikus adatok generálását, továbbá egy ipari keretrendszert javasol. Ennek elérése érdekében a munka a következő fő célok köré épül:

- *Hierarchikus idősor-mérleghiba kiegyenlítés:* A több szinten végzett rendszer modellezés gyakran magában hordozza a hibák és az inkonzisztenciák kialakulásának kockázatát. Kutatásom az optimális összehangolási technikákat vizsgálja – konkrétan összehasonlítva a független modellezést a hibákon alapuló finomhangolással –, hogy biztosítsa, hogy a gépi tanulásos előrejelzések megfeleljenek a fizikai és kémiai aggregációs korlátoknak.
- *Fejlett adatfúzió:* A modell rugalmasságának növelése érdekében egy komplex szintű adatfúziós (CLF) módszer fejlesztettem. A spektroszkópiai adatok (MIR és Raman) kétrétegű rétegező algoritmuson keresztüli integrálásával a kutatás célja, hogy túllépje a hagyományos alacsony és közepes szintű fúziós pontosságot a minőség-ellenőrzési alkalmazásokban.
- *Szintetikus adatok generálása:* A korlátozott ipari adathalmazok jelentette kihívásra reagálva megoldást javasoltam mesterséges infravörös spektrumok generálására a főkomponens-elemzés

(PCA) és a neurális hálózatok segítségével, biztosítva ezzel a robusztus modellek építését adathiányos helyzetekben.

- *CRISP-ML & Modellek életciklusa:* A kutatásomban javaslatot tettem a sikeres adat vezérelt projektek módszertanához. Az elsődleges cél egy átfogó keretrendszer létrehozása a gépi tanuláson alapuló soft szenzorok teljes életciklusának kezeléséhez. Ez magában foglalja a verziókezeléssel, a modellromlással és a komplex kémiai környezetekben történő valós idejű monitorozással kapcsolatos problémák megoldását (MLOps).

### **Disszertáció Felépítése**

A disszertáció a következőképpen épül fel: A 2. fejezet a soft szenzorok trendjeinek szisztematikus irodalmi áttekintését tartalmazza. A 3. fejezet a hierarchikus idősorokban végzett adatösszehangolást részletezi. A 4. fejezet a komplex szintű adatfúzió módszertanát mutatja be. Az 5. fejezet a spektroszkópos modellekhez szükséges mesterséges adatgenerálást tárgyalja. A 6. fejezet bemutatja az ipari soft szenzorokhoz használt MLOps keretrendszert, majd a 7. és 8. fejezetben található az eredmények átfogó összefoglalása és a jövőbeli kutatási irányok.

## 2. Tézisek

**Tézis #1.** Igazoltam, hogy a gépi tanulási modellek megfelelő fejlesztése integrálható egy hierarchikus modellkeretrendszerbe hierarchikus adategyeztetés segítségével, ami nagy predikciós teljesítményt, praktikusságot és a mérési hibákkal szembeni robusztusságot eredményez.

- Kifejlesztettem egy adategyeztetési technikát a hierarchikus idősorokhoz, amely javítja a gépi tanulási modellek használhatóságát.
- Három különböző adatkészletre alkalmaztam: egy ipari adatkészletre, egy referencia-adatkészletre és egy átfogó hulladékgazdálkodási adatkészletre. Az eredményekkel igazoltam, hogy a technika segít javítani a gépi tanulási modellek használhatóságát.
- Három esettanulmányon keresztül bemutattam a technika hasznosságát és annak fontosságát a komplex rendszerekben való alkalmazás szempontjából.

### **Kapcsolódó publikációk: 1, 6**

**Tézis #2.** Igazoltam, hogy a gyenge gépi tanulási algoritmusok teljesítménye jelentősen javítható a javasolt komplex szintű adatfúziós technikák alkalmazásával kiegészítő információt hordozó adathalmazok esetén.

- Kifejlesztettem egy adatfúziós technikát, amely képes különböző mérési spektrumokból származó adatok fúziójával nagyobb teljesítményű gépi tanulási modelleket létrehozni.
- A bemutatott két adatkészlet alapján a kifejlesztett komplex szintű ensemble modell más adatfúziós technikákhoz képest kiváló teljesítményt nyújtott.
- MIR- és Raman-spektrumokon alapuló adatfúziós technikákat teszteltem olyan adatkészletekben, ahol a kizárólag MIR- vagy Raman-adatokon alapuló alapmodellek pontossága gyenge volt (kevesebb mint 90%).

### **Kapcsolódó publikációk: 2, 7**

**Tézis #3.** Igazoltam, hogy mesterséges infravörös spektrumok célzottan előállíthatók összetett mátrixú minták esetén. A kifejlesztett módszertan független valós spektrumok validációjával történt, ami alkalmazható hiányos adatkészletek gyarapítására.

- Kifejlesztettem egy technikát mesterséges spektrumok generálására, amely kifejezetten ritka, vagy hiányos adatsorokhoz képes hatékonyan, megfelelő minőségben adatokat létrehozni.
- A mesterséges infravörös spektrumokat a fejlesztett módszertanban, PCA és ANN technikák segítségével hoztam létre.
- A mesterséges spektrumok hitelességét részletesen teszteltem, és bemutattam validálásuk módszertanát is.

**Kapcsolódó publikációk: 3, 4**

**Tézis #4.** Igazoltam, hogy a CRISP-ML módszertanon alapuló architektúra hatékonyan támogatja a folyamatosan változó teljesítményű gépi tanulási modellek üzemeltetését.

- Kifejlesztettem egy ipari szintű keretrendszert, amely felhő- és edge-alapú számítástechnikát használ a laboratóriumi adatok szinte valós idejű előrejelzésére.
- Kifejlesztettem egy keretrendszert, amely magában foglalja a CRISP-ML módszertan összes elemét, lehetővé téve a modell teljesítményének folyamatos monitorozását és szükség esetén újratanítását.
- Kipróbáltam a keretrendszert egy ipari hálózaton, és pontosságát két, gépi tanulási modelleket használó esettanulmányon keresztül bizonyítottam.

**Kapcsolódó publikációk: 5, 8, 9**

## **Eredmények és jövőbeli célok**

### **A kutatási eredmények ipari hasznosítása**

A disszertációban bemutatott kutatás áthidalja a szakadékot az elméleti gépi tanulás és a vegyipar és a feldolgozóipar szigorú követelményei között. A kidolgozott módszertanok az Ipar 4.0 szélesebb ökoszisztémájának integrált elemeiként lettek kialakítva, és sokrétű előnyöket kínálnak a minőségbiztosítás és a működési hatékonyság terén.

### **Robusztus soft szenzorok megvalósítása**

Az eredmények elsődleges felhasználása a „soft szenzorok” virtuális műszerként való alkalmazásában rejlik. A tárgyalt hibrid modellezési megközelítés alkalmazásával az ipari üzemek összehangolhatják a magas szintű rendszerkorlátozásokat az alacsony szintű szenzoradatokkal. Ez biztosítja, hogy a becsült paraméterek, például a termék összetétele, fizikailag összhangban maradjanak a tömeg- és energiaegyensúlyokkal. A minőségbiztosítás érdekében ez valós idejű, megbízható adatokat biztosít, amelyek jelentősen csökkentik az időigényes, roncsolásos laboratóriumi mintavételek gyakoriságát.

### **Döntéshozatal támogatása adatfúzióval**

A komplex szintű adatfúziós (CLF) módszertan átvihető keretrendszert biztosít a spektroszkópos elemzéshez. Olyan környezetekben, ahol érzékelők (pl. Raman és MIR) állnak rendelkezésre, a CLF több spektrális információkat használ fel a predikációs hibák drasztikus csökkentése érdekében. Moduláris jellege lehetővé teszi új elemző eszközök integrálását a rendszer újratervezése nélkül, ezáltal biztosítva az ipari minőség-ellenőrzési infrastruktúrák hosszú távú használatát.

### **Adathiány célzott kezelése mesterséges adatgenerálással**

Az ipari mesterséges intelligencia egyik gyakori szűk keresztmetszete a ritka folyamatokhoz kapcsolódó átfogó modell tréning adatok hiánya. A mesterséges adatgeneráló keretrendszer ezt úgy oldja meg, hogy a paraméterteret fizikailag értelmes szintetikus mintákkal gazdagítja. Ez lehetővé teszi robusztus modellek fejlesztését még akkor is, ha nem állnak rendelkezésre kiterjedt régebből valós adatok,

lerövidítve az új termékek piacra kerülésének idejét és csökkentve az adatvezérelt rendszerekkel kapcsolatos kockázatokat.

### **CRISP-ML keretrendszer és modell életciklus-kezelés**

Ennek a munkának a hosszú távú hasznossága a javasolt CRISP-ML életciklusban testesül meg. A vázolt szisztematikus monitoring protokollok alkalmazásával a szervezetek átállhatnak a statikus modell bevezetéséről egy dinamikus, önkorrekciós életciklusra. A keretrendszer edge- és cloud computing technológiákat használ, hogy a soft szenzorokat szinkronban tartsa a gyár aktuális állapotával, biztosítva ezzel a stabil termelést, minimalizálva a nem megfelelő termékeket és a vegyi hulladékok mennyiségét.

### **Jövőbeli célok és stratégiai irányok**

A cél az, hogy ezeket a módszereket integráljuk egy autonóm, folyamatorientált intelligens rendszerbe, ha nem is teljes mértékben, akkor legalábbis a konkrét üzleti problémának megfelelő módon.

- A rugalmasság növelése érdekében az adatok egyeztetésére vonatkozó szabályok – például a fizikai és kémiai korlátozások – egy felhasználóbarát felületen keresztül testreszabhatóvá válnak. Ez lehetővé teszi a kezelők számára, hogy szakértői programozási ismeretek nélkül is módosítsák a modell alapjait.
- Autonóm rendszerek feléptése. A rendszer ML modelljei automatikus újratartással minden nap a lehető legpontosabb adatokat prediktálják az üzem minőségbiztosításához.
- Adatgenerálás esetén a jövőbeli cél, hogy részletesen teszteljük és számszerűen bemutassuk a mesterséges adatok által használt modellek pontosságának javulását. A sikeres adatgenerálás révén a konvolúciós neurális hálózatok (CNN) alkalmazása megvalósíthatóvá válik.

### **Összefoglalás**

Konkluzióként elmondható, hogy ez a disszertáció több módszertant javasol egy következő generációs intelligens gyártáshoz. Az adatok összehangolásának, az érzékelők fűzőjének, a mesterséges adatgenerálásnak és az életciklus-kezelésnek a szinergiájával a kutatás iparágra módszereket javasol, amelyek pontosabb, rugalmasabb és fenntarthatóbb ipari döntéshozatalt tesznek lehetővé a jövőben.

### 3. Tézisekkel kapcsolatos publikációk

#### Cikkek nemzetközi folyóiratokban

1. Pál Péter Hanzelik, Alex Kummer, and János Abonyi. „Edge-computing and machine-learning-based framework for soft-ware sensor development” *Sensors*, 22(11):4268, 2022.  
doi: [10.3390/s22114268](https://doi.org/10.3390/s22114268)  
Scimago Journal Ranking: Q1-Q2, Impact factor: 4.532
2. Pál Péter Hanzelik, Alex Kummer, and János Abonyi. „Data reconciliation-based hierarchical fusion of machine learning models” *Machine Learning and Knowledge Extraction*, 6(4):2601–2617, 2024.  
doi: [10.3390/make6040125](https://doi.org/10.3390/make6040125)  
Scimago Journal Ranking: Q1, Impact factor: 8.980
3. Pál Péter Hanzelik, Szilveszter Gergely, János Abonyi, and Alex Kummer. „Data fusion of spectroscopic data for enhancing machine learning model performance” *Digital Chemical Engineering*, page 100271, 2025.  
doi: [10.1016/j.dche.2025.100271](https://doi.org/10.1016/j.dche.2025.100271)  
Scimago Journal Ranking: Q1, Impact factor: 4.775
4. László Győry, Szilveszter Gergely, and Pál Péter Hanzelik. „Generating realistic infrared spectra using artificial neural networks” *Journal of Chemometrics*, 38(9):e3573, 2024.  
doi: [10.1002/cem.3198](https://doi.org/10.1002/cem.3198)  
Scimago Journal Ranking: Q3, Impact factor: 2.189
5. Pál Péter Hanzelik, Szilveszter Gergely, Csaba Gáspár, and László Győry. „Machine learning methods to predict solubilities of rock samples” *Journal of Chemometrics*, 34(2):e3198, 2020.  
doi: [10.1002/cem.3198](https://doi.org/10.1002/cem.3198)  
Scimago Journal Ranking: Q3, Impact factor: 1.569

## Konferencia-kiadványokban megjelent cikkek

6. Pál Péter Hanzelik, Alex Kummer, Ádám Ipkovich, and János Abonyi. „Fusion and integrated correction of chemometrics and machine learning models based on data reconciliation” In M.C. Georgiadis A. Kokossis and S. Pistikopoulos, editors, *Computer Aided Chemical Engineering*, volume 52, pages 1379–1384. Elsevier, 2023. Preceeding and presented at the *33rd European Symposium on Computer-Aided Process Engineering (ESCAPE-33) conference*. Availability: [link](#)
7. Pál Péter Hanzelik, Alex Kummer, Márton Mócz, Szilveszter Gergely, Dorián L Galata, and János Abonyi. „Comparison of different data and information fusion methods to improve the performance of machine learning models” In F. Manenti and G.V.R. Reklaitis, editors, *Computer Aided Chemical Engineering*, volume 53, pages 3007–3012. Elsevier, 2024. Preceeding and presented at the *34th European Symposium on Computer Aided Process Engineering /15th International Symposium on Process Systems Engineering (ESCAPE-34/PSE2024) conference*. Availability: [link](#)

## Konferencia összefoglalók és előadások

8. Pál Péter Hanzelik, Alex Kummer, and János Abonyi. „Development of the edge- and cloud computing framework in the analytical chemistry” Conference presentation, August 2022. at the *International Congress of Chemical and Process Engineering (CHISA) conference*, Availability: [link](#)
9. Pál Péter Hanzelik, Alex Kummer, Ádám Ipkovich, and János Abonyi. „Performance monitoring of machine learning models” Conference presentation, September 2022. at the *Process Control Systems (PCS) conference*, Availability: [link](#)

## 4. Disszertációhoz nem kapcsolódó publikációk

### Cikkek nemzetközi folyóiratokban

- Lei Fu, Yanxiang Yu, Chicheng Xu, Michael Ashby, Andrew McDonald, Wen Pan, Tianqi Deng, István Szabó, Pál Péter Hanzelik, Csilla Kalmár, et al. „Well-log-based reservoir property estimation with machine learning: a contest summary” *Petrophysics*, 65(01):108–127, 2024.  
doi: [10.30632/PJV65N1-2024a6](https://doi.org/10.30632/PJV65N1-2024a6)  
Scimago Journal Ranking: Q3, Impact factor: 0.885
- Norbert Péter Szabó, Károly Nehéz, Oliver Hornyák, Imre Piller, Csaba Deák, Pál Péter Hanzelik, Csaba Kutasi, and Károly Ott. „Cluster analysis of core measurements using heterogeneous data sources: An application to complex miocene reservoirs” *Journal of Petroleum Science and Engineering*, 178:575–585, 2019.  
doi: [10.1016/j.petrol.2019.03.067](https://doi.org/10.1016/j.petrol.2019.03.067)  
Scimago Journal Ranking: Q1, Impact factor: 5.168

### Konferencia összefoglalók és előadások

- Márton Mócz, Pál Péter Hanzelik, János Slezsák, and Gergely Szilveszter. „Impact of different model transfer algorithms on dilution series and oil samples” Presentation, September 2023. Poster presented at the *Conferencia Chemometrica 2023 conference*.
- Puskas, S., Vago, A., Toro, M., Ordog, T., Kalman, G., Hanzelik, P. P., Bihari, Zs., Blaho, J., Tabajdi, R., Dekany, I., Dudas, J., Nagy, R., Bartha, L., and I. Lakatos. „Surfactant-Polymer EOR from Laboratory to the Pilot” *Paper presented at the SPE EOR Conference at Oil and Gas West Asia*, Muscat, Oman, March 2018. doi: [10.2118/190369-MS](https://doi.org/10.2118/190369-MS)