

TDK-dolgozat

Németh Csilla Márta

VPNet elemzése: Megmagyarázhatóság és adat szétválaszthatóság

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

INFORMATIKAI KAR

NUMERIKUS ANALÍZIS TANSZÉK



ELTE

EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

Szerző:

Németh Csilla Márta

programtervező informatikus MSc

Témavezető:

Dr. Kovács Péter

egyetemi docens

Budapest, 2024

Absztrakt

Mesterséges intelligencia (MI) modelleket a közelmúltban sikeresen alkalmaztak számos tudományos és gyakorlati probléma megoldására. Biztonságkritikus területeken, pl. autonóm rendszerek, egészségügy, ahol a teljesítmény mellett az döntéshozatal átláthatósága is fontos szempont, azonban továbbra is csak korlátozottan használhatók a legkorszerűbb MI algoritmusok. Ezt kiküszöbölendő jött létre a megmagyarázható mesterséges intelligencia (XAI) egy kitüntetett területe, a modellvezérelt neurális hálók kutatása, mely ötvözi a matematikai modell alapú megközelítés és az MI adatvezérelt szemléletének előnyeit. Ebbe a kategóriába sorolható a pár éve kifejlesztett ún. Variable Projection Network (VPNet) típusú neurális hálók, mely a paraméterezett ortogonális projekciókat kombinálja különféle MI technikákkal.

A VPNet első rétege egy tanítható és transzparens dimenziócsökkentő transzformációt valósít meg. A dolgozatban azt vizsgálom, hogy a betanított VP réteg által kinyert jellemzők, hogyan járulnak hozzá az utána kapcsolt háló kimenetéhez különböző osztályozási feladatok (EKG, VEP és kerékszenzoros jelek) estén. A VP jellemzők és a háló által becsült osztálycímkek közötti kapcsolatot a Case-Based Reasoning (CBR) módszer segítségével elemeztem, amely lehetővé teszi, hogy a neurális háló döntéseit korábbi, hasonló esetek alapján értsük meg.

A dolgozat másik fókusza a VP réteg hatása az adatok szeparálhatóságára. Ezt különböző dimenziócsökkentési módszerekkel (PCA, t-SNE, UMAP) és tanuló algoritmusokkal (KNN, SVM) vizsgáltam. Az eredmények alapján a VP réteg jelentősen javította az adatok nemlineáris szeparálhatóságát, amit kvantitatíve is alátámasztottak az elemzett szeparálhatósági mutatók, mint a Silhouette-metrika és a Davies-Bouldin index.

Tartalomjegyzék

1. Bevezetés	3
2. VPNet Áttekintése	5
2.1. A VPNet célja és felépítése	5
2.2. A VP Réteg Működése	6
2.3. A VPNet előnyei	7
2.4. Vizsgált adathalmazok	8
3. Megmagyarázható neurális hálózatok	11
3.1. A magyarázhatóság szükségessége	11
3.2. Post-hoc magyarázatok	12
3.3. Case-Based Reasoning (CBR)	13
3.3.1. A CBR alkalmazása a VPNet-ben	14
3.3.2. Eredmények	17
4. Dimenziócsökkentési módszerek	24
4.1. Dimenziócsökkentés szükségessége	24
4.2. Alkalmazott Dimenziócsökkentési Technikák	25
4.2.1. t-SNE (t-distributed Stochastic Neighbor Embedding)	25
4.2.2. UMAP (Uniform Manifold Approximation and Projection)	27
4.2.3. PCA (Principal Component Analysis)	28
4.3. Eredmények	28
4.3.1. EKG adathalmaz	28
4.3.2. EEG adathalmaz	34
4.3.3. Kerékszenzoros adathalmaz	39
5. Adatok szeparálhatósága	45
5.1. Szeparálhatósági vizsgálatok jelentősége	45
5.2. Lineáris és Nemlineáris Szeparálhatóság Vizsgálata	45

5.2.1. Lineáris szeparálhatóság	46
5.2.2. Nemlineáris szeparálhatóság	47
5.3. Eredmények	49
5.3.1. EEG adathalmaz	49
5.3.2. EKG adathalmaz	50
5.3.3. Kerékszenzoros adathalmaz	50
6. Összegzés és Kitekintés	52
Irodalomjegyzék	55

1. fejezet

Bevezetés

Az utóbbi évek technológiai fejlődésével a neurális hálózatok a mesterséges intelligencia és gépi tanulás területén kiemelkedő szerepet kaptak. Alkalmazásuk gyorsan terjed olyan komplex területeken, mint az orvosi diagnosztika, a pénzügyi előrejelzések vagy az önvezető rendszerek, amelyekben a döntések pontossága és megbízhatósága kritikus tényező (Mesterséges intelligencia az orvostudományban). Bár a neurális hálózatok képesek nagy mennyiségű adat feldolgozására és bonyolult mintázatok felismerésére, a belső működésük és döntési folyamataik gyakran átláthatatlanok maradnak az emberek számára. Ez a tulajdonság jelentős kihívásokat teremt olyan alkalmazásokban, ahol a modell döntéseinek megértése és indokolhatósága elengedhetetlen. Ezt kiküszöbölendő jött létre a megmagyarázható mesterséges intelligencia (Explainable Artificial Intelligence, XAI). Mostanra számos XAI módszer létezik, ezeket részletesen összegzi a „The Mythos of Model Interpretability” [1] című cikk. Az értelmezhetőségi megközelítések egyik gyakran alkalmazott kategóriája a post-hoc magyarázatok, amelyek a neurális hálózatok működésének utólagos megértését teszik lehetővé anélkül, hogy a modell belső struktúráját feltárnánk. A post-hoc módszerek közé tartoznak a szóbeli magyarázatok, a vizuális eszközök, például a saliency térképek, valamint a példákön alapuló magyarázatok (Case-Based Reasoning, CBR) is. A CBR alapelve, hogy egy adott döntést korábbi, hasonló esetek alapján magyarázzunk meg, ezzel érthetőbbé téve a neurális hálózat működését a felhasználók számára. Ez a megközelítés különösen jól alkalmazható az orvosi területen, ahol az egészségügyi szakemberek számára gyakran kulcsfontosságú, hogy a rendszer döntéseit a múltbeli esetekre hivatkozva értsék meg.

A dolgozat egyik kulcsfontosságú eleme a Case-Based Reasoning (CBR) módszer alkalmazása, amelyet a VPNet [2] magyarázhatóságának növelésére használtam. A CBR le-

hetővé teszi, hogy a neurális hálózat döntéseit korábbi, hasonló esetekre alapozva indokoljuk, így a felhasználók számára átláthatóbbá válik a modell működése. Az általam implementált CBR-alapú vizsgálat hozzájárult a VPNet megmagyarázhatóságához a *Weighted Hermite Variable Projection Networks for Classifying Visually Evoked Potentials*” című cikkben [3].

A dolgozat másik fókuszpontja a változóprojekciós (VP) réteg hatása az adatok szeparálhatóságára. Az adatok szeparálhatósága azt jelenti, hogy az egyes osztályok mennyire könnyen különíthetők el egymástól a térben, például lineáris (egy egyenes vagy hipersík) vagy nemlineáris (összetettebb, görbült döntési határvonalak) mentén. A VPNet osztályozórendszer sajátossága a változóprojekciós (VP) réteg, amely lehetővé teszi a paraméterek szétválasztását lineáris és nemlineáris változókra, ezáltal elősegítve az adatok hatékonyabb feldolgozását és osztályozását. A kutatásom során vizsgáltam, hogy a VP réteg miként befolyásolja az adatok lineáris és nemlineáris szeparálhatóságát. A kvalitatív vizsgálatok során különböző dimenziócsökkentési technikákat alkalmaztam, mint a PCA, t-SNE és UMAP, hogy vizuálisan elemezzem az osztályok elkülönülését. Ezzel szemben a kvantitatív vizsgálatokat KNN [4] és SVM [5] algoritmusokkal, valamint szeparálhatósági mutatókkal, mint a Silhouette-metrika és a Davies-Bouldin index végeztem, amelyek számszerűen értékelték a VP réteg hatékonyságát az adatok elkülöníthetőségére.

A dolgozat célja tehát a VPNet magyarázhatóságának növelése CBR alkalmazásával, valamint a VP réteg adatszeparálási hatékonyságának vizsgálata, különös tekintettel az orvosi alkalmazásokra, mint például az EKG-jelek osztályozása.

2. fejezet

VPNet Áttekintése

2.1. A VPNet célja és felépítése

A Variable Projection Network (VPNet) [2] célja az egy dimenziós nemlineáris modellezési problémák hatékony megoldása, amelyek során az adatokat paraméterezett függvényrendszerek segítségével közelítjük. A modell fő innovációja abban rejlik, hogy egyesíti a hagyományos, modellalapú jelfeldolgozás megközelítéseket a neurális hálózatok (NN) tanulási képességével, ezáltal ötvözi a VP (Variable Projection) reprezentációs képességeit a neurális hálózatokkal, kompozit modellt alkotva.

A VPNet architektúrájában (2.1 ábra) egy előre csatolt neurális hálózat (feedforward NN) található, amelynek első rétege egy VP operátor. A VP réteg működési módjai az alkalmazás céljától függően változhatnak, lehetőséget biztosítva mind regressziós és osztályozási feladatok elvégzésére. Az alkalmazás céljától függően a VP réteg vagy a bemenet alacsony dimenziós szűrt reprezentációját:

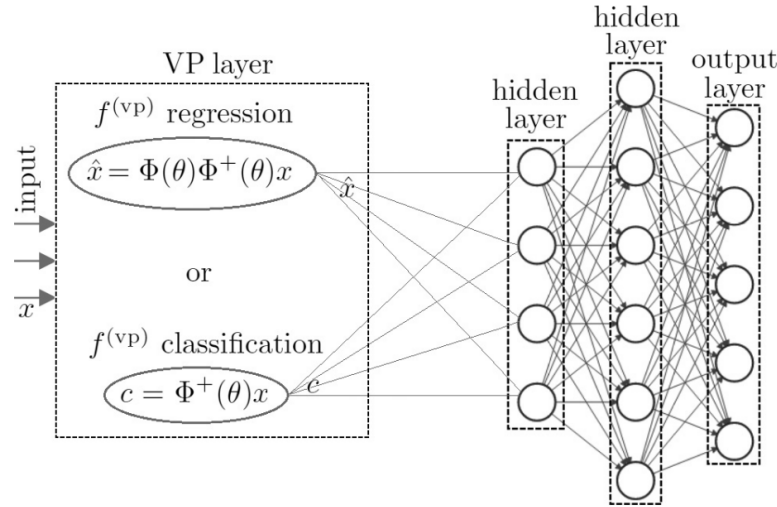
$$\hat{x} = \Phi(\theta)\Phi^+(\theta)x,$$

vagy a jellemzők extrakcióját végzi el:

$$c = \Phi^+(\theta)x$$

ahol $x \in \mathbb{R}^N$ a bemeneti jel, $\Phi(\theta) \in \mathbb{R}^{N \times n}$ pedig egy olyan (reguláris) mátrix, melynek oszlopai függnek a $\theta \in \mathbb{R}^m$ paraméter vektortól. Az első transzformáció az x vektort a $\Phi(\theta)$ oszlopterére vetíti, a második pedig megfelel ezen vetítés oszloprendszer-bázis szerinti koordinátáinak.

A $\Phi(\theta)$ a VP réteg által alkalmazott paraméterezett függvényrendszert reprezentáló mátrix, amely a bemeneti adatokat transzformálja. Ennek elemei lehetnek például Hermite-függvények. A $\Phi^+(\theta)$ a $\Phi(\theta)$ Moore–Penrose pszeudoinverze, amely a lineáris paraméterek (c) kiszámítására szolgál, biztosítva a legjobb illeszkedést az adatokhoz. A θ a tanítható paraméterek vektora a VP rétegnek a legtöbb alkalmazás esetében x egy $L_2(R)$ függvény mintavételezéséből adódik, Φ oszlopait pedig egy megfelelően kiválasztott $L_2(R)$ bázis mintavételezésének választjuk. Kvázi periodikus jelek esetén pl. több alkalmazáshoz használhatunk megfelelően paraméterezett Hermite-függvényeket.



2.1. ábra. VPNet architektúrája [2].

2.2. A VP Réteg Működése

A VP réteg két fő funkciót láthat el a VPNet architektúrájában: szűrést és egyedi jellemzők kinyerését. Az alkalmazott funkció függ az adott probléma típusától:

- **Szűrés:** Jellemzően regressziós feladatokhoz alkalmazható, ahol előrejelzéseket kell végezni. Ilyen esetek például a ház ár becslés vagy időjárás-előrejelzés, ahol a VP réteg releváns információkat szűr ki a bemenetekből, és így támogatja a hálózat pontosabb előrejelzéseit [6].
- **Egyedi jellemzők kinyerése az adatokból:** Ideálisak osztályozási feladatokhoz. Az ilyen feladatok közé tartozik például a kategorizálás, betegségdiagnosztika vagy képfelismerés, ahol a bemenetekből kinyert releváns jellemzők az adatok osztályozását segítik elő. Ezeket a jellemzőket a VPNet egy teljesen összekapcsolt, akár

mély hálózatnak továbbítja, amely a tanulási folyamat következő szakaszában dolgozza fel az információt.

Ugyan a VP réteg általános felépítése lehetővé teszi, hogy $\Phi(\theta)$ oszlopait tetszőleges lineárisan független rendszernek válasszuk, a dolgozatban olyan VP-Net realizációkra koncentrálnak, melyek ún. Hermite függvények segítségével valósítják meg az előző pontban ismertetett adaptív projekciókat. Ezek a függvények lehetővé teszik a csúcshullám jelalakok modellezését, így jól alkalmazhatók olyan orvosi biológiai jelminták feldolgozására, amelyek morfológiai jellemzői kiemelt fontossággal bírnak az osztályozási folyamat során. A Hermite-VP réteg lehetőséget nyújt ezen jellemzők hatékony és pontos kinyerésére.

2.3. A VPNet előnyei

A VPNet egyszerűsége és hatékonysága több szempontból is kiemelkedő előnyt nyújt a hagyományos neurális hálózatokkal, például teljesen összekapcsolt vagy konvolúciós neurális hálózatokkal (CNN) szemben:

- **Egyszerűség és hatékonyság:** A VP réteg dimenziócsökkentésével a rendszer összességében kevesebb számítási kapacitást és memóriát igényel, így a modell gyorsabban fut, különösen magas dimenziójú adathalmazok esetén
- **Paraméterszám csökkenése:** A VP réteg révén jelentősen csökken a tanítható paraméterek száma, mivel a VPNet az adatjellemzőket közvetlenül a bemenetekből állítja elő, nem pedig egymás után kapcsolt rétegek sorozatával. Ez csökkenti a hálózat túltanulásának kockázatát és a tanító halmaz méretét is.
- **Interpretálhatóság:** A VPNet VP rétegében alkalmazott nemlineáris paraméterek (pl. a Hermite-függvényeknél a transzláció és nyújtási paraméterek) közvetlen kapcsolatban állnak a feldolgozott jelek morfológiai jellemzőivel. Ez lehetővé teszi, hogy a tanult paraméterek alapján jobban megértsük az adatokat és a mögöttes mintázatokat.
- **Széles körű alkalmazhatóság:** A VPNet rugalmas architektúrája lehetővé teszi, hogy különböző területeken alkalmazzák, mint például az orvosi jelanalízis (pl. EKG és EEG feldolgozás), időbeli adatok osztályozása, valamint egyéb egydimenziós jel- és adatfeldolgozási problémák.

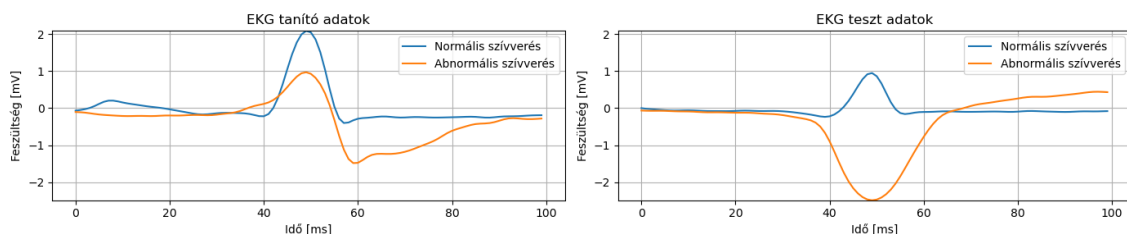
A dolgozatban a VPNetet osztályozási feladatokon keresztül vizsgáltam, ahol a VP réteg EKG (szív elektromos impulzusai) EEG (agyi aktivitás) és kerékszenzoros jelekből nyert ki egyedi jellemzőket. Az eredményeim azt mutatják, hogy a VP réteg alkalmazása javítja a jelek nemlineáris szeparálhatóságát, mivel a releváns jellemzők hatékony kinyerésével elősegíti az adatok elválasztását komplex térbeli struktúrákban. Ennek köszönhetően a VPNet pontosabb eredményeket produkált, mint a hagyományos, teljesen összekapcsolt vagy konvolúciós neurális hálózatok.

2.4. Vizsgált adathalmazok

A VPNet modell teljesítményét három különböző adathalmazon vizsgáltam, amelyek mindegyike tanító és teszt adatokra volt osztva. A tanító adathalmazt a modell betanítására, míg a teszt adatokat a modell általánosítási képességének értékelésére használják. Mindhárom adathalmaz közös jellemzője, hogy csúcshullámú, egydimenziós, bonyolult jeleket tartalmaznak. Ezeket a típusú jeleket jól lehet közelíteni Hermite-függvényekkel, emiatt a VPNet modell hatékonyan képes az adatok osztályozására.

Elektrokardiográfia (EKG)

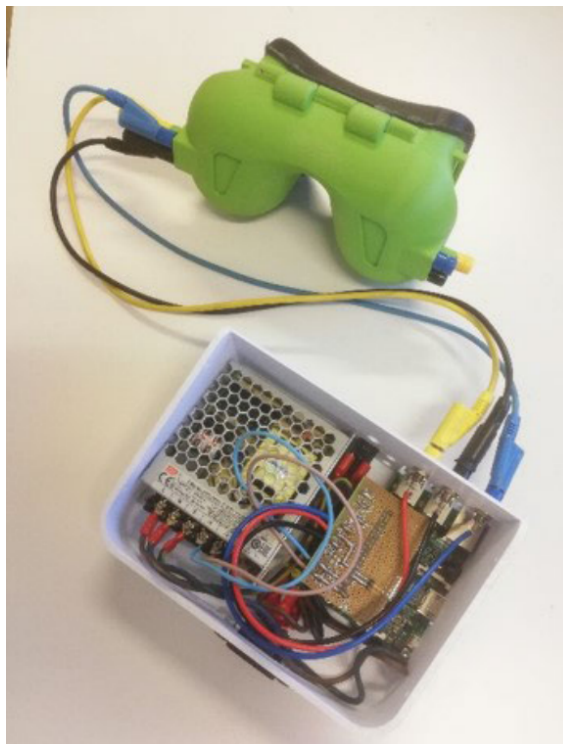
Az EKG egy diagnosztikai módszer, amely a szívizom összehúzódásai során keletkező elektromos aktivitás változásait rögzíti. A VPNet alkalmazását EKG jelek osztályozására a MIT-BIH Arrhythmia adatbázis alapján vizsgálták [2] és én ezen a betanított modellen végeztem a további vizsgálatokat a dolgozat során. Ez az adatbázis több mint 100 000 szívveréshez tartozó QRS komplexumot tartalmaz, amelyek két fő osztályba sorolhatók: normál szívverések és ritmuszavaros szívverések.



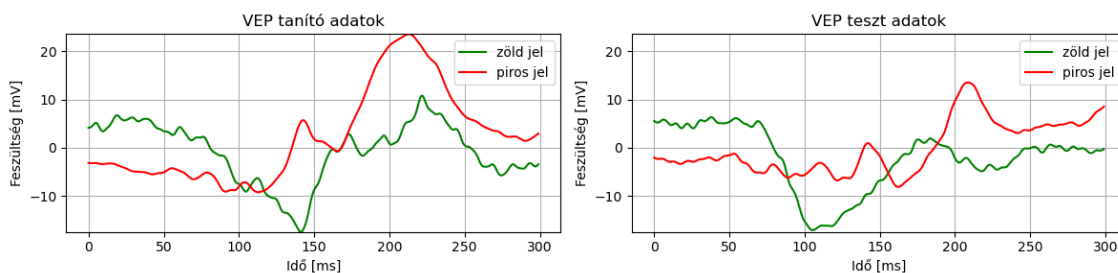
2.2. ábra. EKG jelek.

Elektroenkefalográfia (EEG)

Az EEG az agyi elektromos aktivitás mérésére szolgáló módszer, amely az agykéreg idegsejtjeinek elektromos kisüléseit rögzíti. A vizsgálathoz használt adathalmaz a [3] publikációban bemutatott kísérletekből származik, ahol a VPNet-et a látás által kiváltott potenciálok (VEP) osztályozására alkalmazták. Az adatok különböző vizuális ingerek: piros és zöld fényvillanások által kiváltott agyi válaszjelekből álltak. Ábra 2.3, amelyeket 88 résztvevőnél mértek. A jelek előfeldolgozása során a nyers EEG adatokból időben átlagolt mintákat hoztak létre, hogy csökkentsék a zajt és kiemeljék a jellegzetes mintázatokat. Az így kapott adathalmaz két osztályba sorolható: a piros és a zöld inger által kiváltott válaszjelek, amelyeket a VPNet modell nagy pontossággal osztályozott.



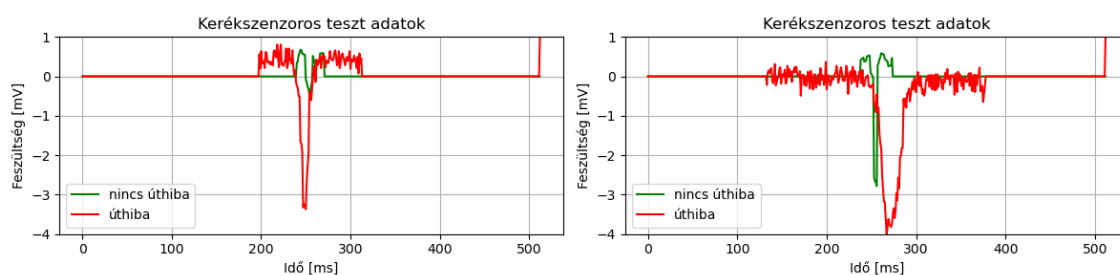
2.3. ábra. Az adatok gyűjtésére készült készülék [3] .



2.4. ábra. VEP jelek.

Kerékszenzoros jelek

A kerékszenzoros adathalmaz a VPNet alkalmazásának egy újabb példája [6], amely az "okos abroncsok" által gyűjtött adatok osztályozását célozza. A gumiabroncsokat erőmérő szenzorral szerelték fel, amely a mechanikai deformációkat elektromos ellenállás-változásokká alakítja, és ezek alapján becsüli a szenzorra ható mechanikai erők különböző irányú komponenseit. A szenzor által gyűjtött nyers adatok tüskeszerű jeleket eredményeznek, amelyek a kerék deformációit tükrözik. Az osztályozáshoz a jeleket a kerék teljes fordulatai alapján szegmentálták, hogy egy-egy fordulat jeleit külön szakaszokként lehessen elemezni. Az adathalmaz két fő osztályba sorolható: úthiba és „nincs úthiba”.



2.5. ábra. Kerékszenzoros jelek.

3. fejezet

Megmagyarázható neurális hálózatok

3.1. A magyarázhatóság szükségessége

A mesterséges intelligencia az utóbbi időben nagy népszerűségnek örvend számos területen. Ugyanakkor ezek a modellek gyakran nehezen alkalmazhatóak biztonságkritikus területeken, például autonóm rendszerek, pénzügyi döntéshozatal, egészségügyi elemzések, ahol a döntéshozatal átláthatósága kiemelten fontos.

A magyarázható mesterséges intelligencia (Explainable Artificial Intelligence, XAI) koncepciója éppen erre a problémára kínál megoldást. Az XAI célja, hogy olyan modelleket fejlesszenek ki, amelyek nemcsak pontosak, hanem döntéshozatali folyamataik is átláthatók és értelmezhetők az emberi felhasználók számára. Ezáltal növelhető az MI-rendszerek iránti bizalom, lehetővé válik a hibák azonosítása és javítása, valamint az etikus és igazságos döntéshozatal biztosítása.

Az Európai Unió is felismerte a mesterséges intelligencia alkalmazásának kockázatait, így erre vonatkozóan több átfogó szabályozási keretet alakítottak ki. A legfontosabb lépés ezen a területen a mesterséges intelligenciáról szóló rendelet (AI Act) elfogadása volt, amely 2024. augusztus 1-jén lépett hatályba.

Az AI Act célja, hogy biztosítsa az MI-rendszerek biztonságos, etikus és megbízható működését az Európai Unióban. A rendelet a mesterséges intelligencia előremutató fogalom meghatározásán és egy kockázatalapú megközelítésen alapul:

- **Elfogadhatatlan kockázat:** például azok az MI-rendszerek, amelyek lehetővé teszik a polgárok kormányok vagy vállalatok általi pontozását, egyértelmű fenyegetést jelentenek az emberek alapvető jogaira nézve. Az ilyen alkalmazások használata tilos.

- **Magas kockázat:** a magas kockázatú MI-rendszereknek, például a mesterséges intelligencián alapuló orvosi szoftvereknek vagy a munkaerő-felvételhez használt MI-rendszereknek szigorú követelményeknek kell megfelelniük, beleértve a kockázatsökkentő rendszereket, a magas színvonalú adatkészleteket, a felhasználók egyértelmű tájékoztatását, az emberi felügyeletet stb.
- **Az átláthatósággal kapcsolatos különleges körülmény:** az olyan rendszereknek, mint a csevegőrobotok, egyértelműen tájékoztatniuk kell a felhasználókat arról, hogy gépekkel lépnek kapcsolatba.
- **Minimális vagy semmilyen kockázat:** a legtöbb MI-rendszer, például a spamszűrők és a mesterséges intelligencián alapuló videojátékok nem tartoznak a mesterséges intelligenciáról szóló jogszabály szerinti kötelezettség hatálya alá, de a vállalatok önkéntesen további magatartási kódexeket fogadhatnak el.

A rendelet különös figyelmet fordít a magyarázhatóságra, főképp a magas kockázatú MI-rendszerek esetében. Ezeknek a rendszereknek biztosítaniuk kell, hogy működésük és döntéshozatali folyamataik átláthatóak és érthetőek legyenek a felhasználók számára.

3.2. Post-hoc magyarázatok

Az értelmezhetőség biztosításának egyik megközelítése a *post-hoc* magyarázatok alkalmazása. A post-hoc magyarázatok olyan megközelítéseket jelentenek, ahol a már betanított modellekből igyekszünk kinyerni információkat. Ezek a magyarázatok nem feltétlenül segítenek megérteni hogyan működik az adott modell, hanem a döntés helyességét hivatottak alátámasztani. A post-hoc módszer előnye, hogy a már betanított modell döntéseinek utólagos értelmezését teszik lehetővé anélkül, hogy módosítanánk a modell belső szerkezetét vagy működését. Ez különösen előnyös, mivel nem szükséges újra tanítani a modellt, és alkalmazható komplex, több millió paraméterrel rendelkező modellekre is.

Főbb Post-hoc technikák:

- **Szöveges magyarázatok:** Az MI-modellek működését természetes nyelvű magyarázatok segítségével próbálják megérthetővé tenni. Például egy modell egy megerősítéses tanulási algoritmus döntéseit képes verbálisan magyarázni egy különálló nyelvi modell segítségével. Ezt alkalmazzák például ajánlórendszerekben is, ahol a döntésekhez kapcsolódó felhasználói vélemények alapján készülnek magyarázatok [7].

- **Vizualizációk:** A tanult reprezentációk vizuális ábrázolása, például hő térképek használatával. Népszerű módszer a t-SNE algoritmus, amely magas dimenziójú adatokat két vagy háromdimenziós térben jelenít meg, hogy láthatóvá tegye az adatok közötti kapcsolatrendszer [8]. Ezt a módszert a VPNet elemzéséhez is használtam, a Alkalmazott Dimenziócsökkentési Technikák fejezetben térek ki a részletes ismertetésre.
- **Lokális magyarázatok:** A modell döntéseit egy adott pont környezetében elemzik, például saliency térképek vagy sparse lineáris modellek alkalmazásával. A saliency térképek azt mutatják meg, hogy a bemeneti adatok mely részei befolyásolják leginkább az adott döntést.
- **Példákon alapuló magyarázatok:** Case Base Reasoning (CBR). Az adott modell által hozott döntéseket a vizsgált bemenethez hasonló adatok alapján indokolják. Például egy daganatot tartalmazó kép esetében a modell olyan korábbi képekre hivatkozhat, amelyek szintén hasonló tulajdonságokat mutattak. A dolgozat további részében ezt a módszert alkalmazom a VPNet-re [9].

Ezek a módszerek lehetővé teszik a felhasználók számára, hogy betekintést nyerjenek a modell működésébe, ezáltal növelve a rendszer iránti bizalmat.

3.3. Case-Based Reasoning (CBR)

A *Case-Based Reasoning* (CBR) egy példákon alapuló megközelítés, amely a *post-hoc* magyarázatok egyik hatékony eszköze. A CBR alapelve, hogy egy új probléma megoldása hasonló a korábbi, már megoldott esetekhez, így a döntéseket ezek alapján lehet magyarázni. A mély neurális hálózatok tanításakor a rendszer nemcsak az előrejelzéseket, hanem az ezekhez kapcsolódó tanult reprezentációkat is képes előállítani. A rejtett rétegek aktivációit felhasználva azonosíthatók azok a konstansok, amelyek a modell által tanult térben a legközelebb helyezkednek el a bemenethez (például a legközelebbi k szomszéd meghatározásával). Ezek a példák segíthetnek az előrejelzések alátámasztásában és a modell működésének intuitív megértésében. Ez a módszer hasonlít az emberek által alkalmazott érvelési technikákhoz. Például betegségek kezelésekor az orvosok gyakran hivatkoznak korábbi esettanulmányokra, hogy alátámasszák a diagnózis vagy a kezelés indokoltságát. Ez a gondolkodási minta analóg módon alkalmazható a gépi tanulásban, ahol a modell döntéseit hasonló esetekkel lehet megmagyarázni.

A CBR előnyei:

- **Érthetőség:** A döntések hasonló esetekkel való magyarázata intuitív és könnyen érthető a felhasználók számára.
- **Átláthatóság:** Ez a módszer különösen akkor hasznos, ha a modell "fekete doboz" jellegű, és a döntéshozatali folyamat bonyolultsága miatt nehezen érthető.

Korlátok:

- **Példák kiválasztása:** A hasonló példák megtalálása a tanult térben a modell tanításának és optimalizációjának függvénye, ami nem mindig pontosan tükrözi a döntéshozatali logikát.
- **Relevancia:** Az azonosított példák nem mindig relevánsak a konkrét döntési kontextusban, különösen ha a modell hibás mintázatokat tanult meg.

3.3.1. A CBR alkalmazása a VPNet-ben

A VPNet felhasználási területei között szerepel az orvosi diagnosztika, ahol – ahogy korábban említettem – kiemelkedően fontos a megmagyarázhatóság, hiszen ez egy magas kockázatú terület. A való életben az orvostudományban gyakran indokolják az orvosok a döntéseiket korábbi, hasonló esetek alapján. Ezt a megközelítést követve a példaalapú magyarázatot választottuk a VPNet-hez. A VPNet-et párosítottam egy CBR rendszerrel, amely alapját Eoin M. Kenny és Mark T. Keane által kidolgozott iker rendszer [9] adta. Ebben a rendszerben egy Artificial Neural Network (ANN)-CBR párost használtak az ANN által hozott döntések megmagyarázására.

Az általam használt VPNet egy ún. FCVPNN volt, amelyet EKG jelek osztályozására alkalmaztak. Az FCVPNN azt jelenti, hogy a hálózat NN része egy hagyományos, teljesen összekapcsolt (fully connected) neurális hálózat (FCNN). Az első rétegben ReLU aktivációs függvény, míg a kimeneti rétegben Softmax aktiváció kapott helyet.

ReLU aktivációs függvény: A ReLU egy egyszerű, de hatékony aktivációs függvény, amely pozitív értékeket tart meg, míg negatív értékeket nullára állít. Formálisan a képlete:

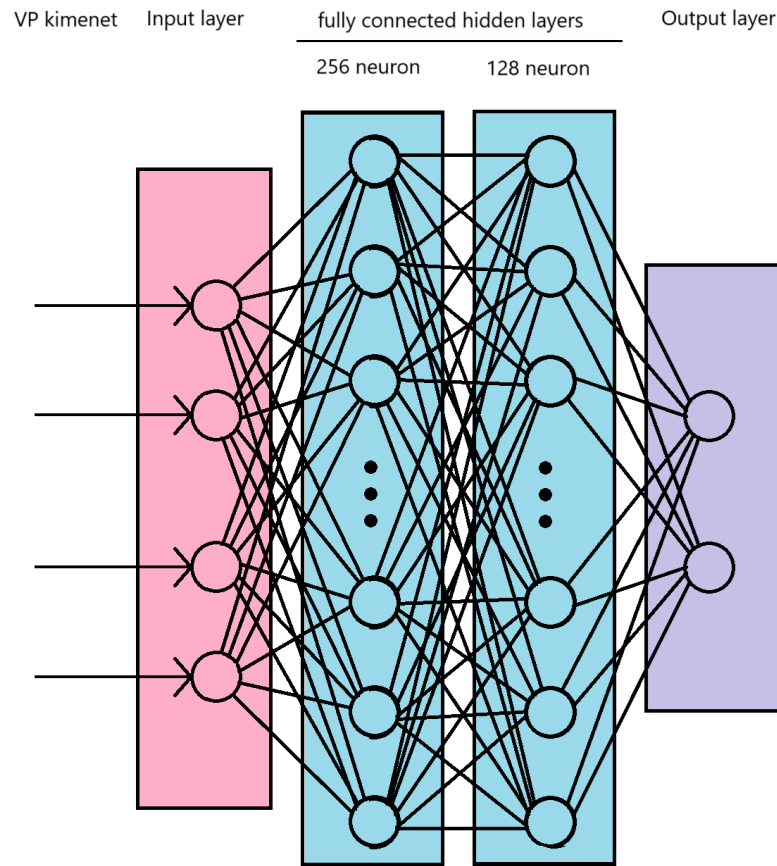
$$f(x) = \max(0, x),$$

ahol x az adott neuron bemeneti értéke. A ReLU előnye, hogy minimalizálja a gradiens eltűnésének problémáját, amely gyakran jelentkezik más aktivációs függvényeknél.

Softmax aktivációs függvény: A Softmax aktivációs függvény a kimeneti rétegben osztályozási problémák esetén használatos, mivel valószínűség-eloszlást generál az osztályok között. A Softmax képlete az i -edik kimeneti neuron esetén:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}},$$

ahol z_i az i -edik kimeneti neuron bemenete, N pedig az osztályok száma. Az egyes kimenetek valószínűségi értékei összege 1, ami intuitívvá teszi az osztályozási döntéseket.



3.1. ábra. Az FCVPNN NN részének felépítése.

Az implementáció során a VPNet betanítása után elmentettem a VP és az NN rétegek súlyait, majd szétválasztottam őket. Az adatok feldolgozására egy egyedi adatbetöltő rendszert (*data loader*) hoztam létre, amely egyenként továbbítja az adatokat a VP rétegen keresztül. A feldolgozott adatokat a NN részhez továbbítottam, ahol két különböző módszert alkalmaztam a legközelebbi szomszéd keresésére:

- **Második FC réteg aktivációi alapján:** A második teljesen összekapcsolt (FC) réteg által generált jellemzők aktivációját használtam konstansként a legközelebbi

szomszéd kereséséhez.

- **Softmax utáni valószínűségek alapján:** Egyszerűsített módszerként a Softmax réteg által generált kimeneti valószínűségeket használtam konstansként.

Mindkét módszerrel hasonló eredményeket kaptam az EKG adatok esetében, ezért az EEG jelek vizsgálatánál kizárólag a második, egyszerűbb módszert alkalmaztam, mivel ez gyorsabb implementációt és ugyanakkor pontos eredményeket biztosított.

A *k-Nearest Neighbors* (kNN) algoritmus egy egyszerű, mégis hatékony gépi tanulási módszer, amelyet alapvetően osztályozási és regressziós problémák megoldására használnak. A kNN alapelve, hogy egy adott adatponthoz a tanulási térben található legközelebbi k szomszédos adatpont alapján hoz döntést. A dolgozatomban az algoritmust a `sklearn.neighbors` modulból valósítottam meg, és arra használtam, hogy a tanító adathalmazból hasonló eseteket adhassak vissza, így megvalósítva a CBR-rendszert. Az alábbi lépésekben foglalható össze a kNN működése a megvalósítás során:

1. **Adattávolság mérése:** Az algoritmus minden új bemenethez kiszámolja a tanuló adathalmazban szereplő összes adatpont távolságát. A távolságméréshez a leggyakrabban alkalmazott metrika az euklideszi távolság, amely a két adatpont közötti geometriai távolságot méri:

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}.$$

ahol $x_1, x_2 \in \mathbb{R}^N$ a két adatpont koordinátáit és $\mathbf{x}_{1,i}$ a bemeneti i -edik koordinátáját jelöli.

2. **Legközelebbi K szomszéd azonosítása:** A távolságértékek alapján az algoritmus kiválasztja a bemenethez legközelebb álló k adatpontokat.

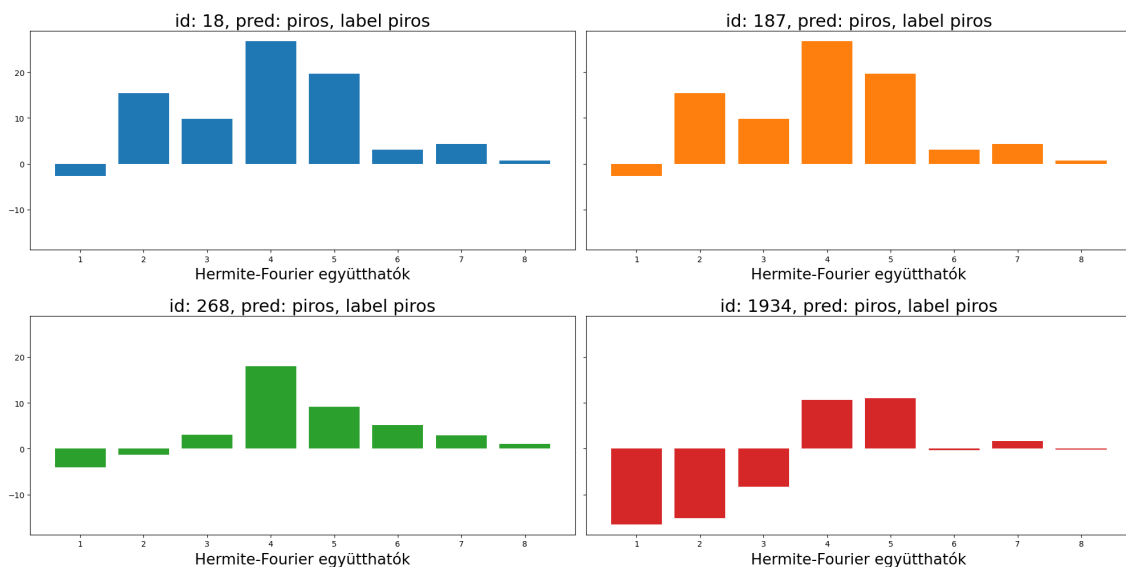
$$\hat{y} = \arg \max_{c \in \text{classes}} \sum_{i \in \mathcal{N}} \mathbb{I}(y_i = c),$$

ahol $\mathcal{N}$ a k legközelebbi szomszédok halmaza, $\mathbb{I}$ az indikátorfüggvény, amely 1-et ad, ha a feltétel teljesül, és 0-át, ha nem. y_i az i -edik szomszéd osztálycímkéje.

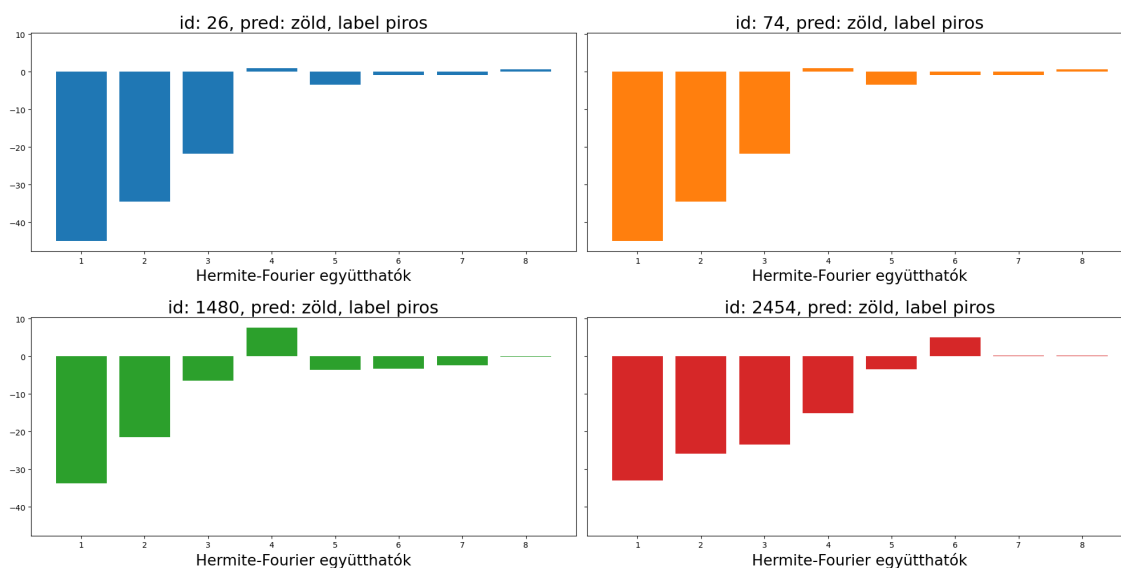
3.3.2. Eredmények

EEG adathalmaz

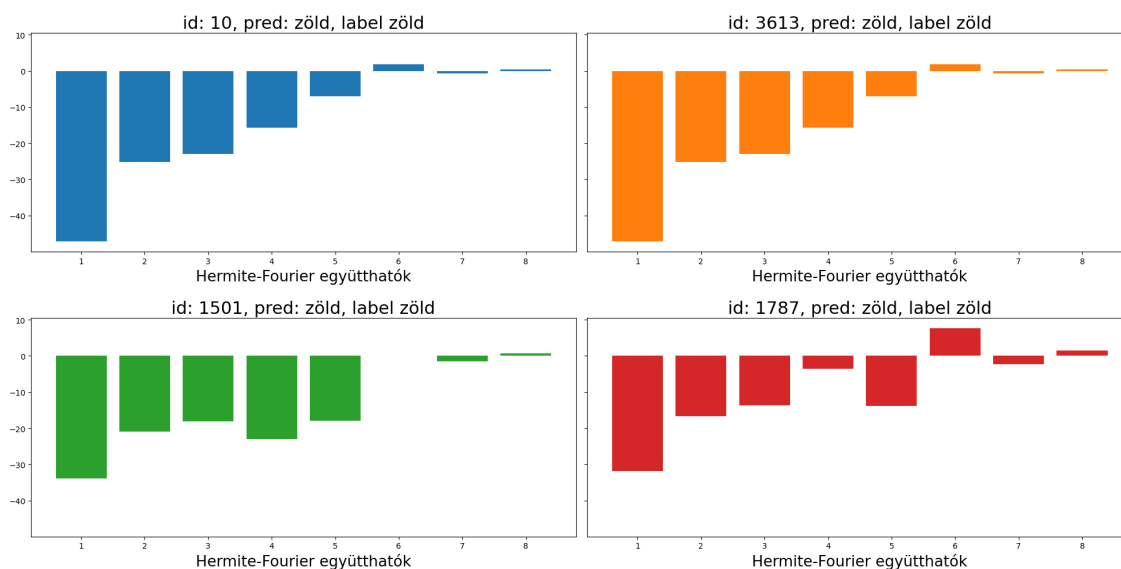
Az EEG adathalmaz feldolgozása során a VPNet modell betanításakor 8 Hermite-függvényt alkalmaztak a bemeneti jelek reprezentálására [3] a VP réteg segítségével. Az általam alkalmazott CBR (Case-Based Reasoning) módszerben minden bemenethez legközelebbi szomszédokat keresek a tanító adathalmazból. Ezeknél a szomszédoknál a következőket jelenítem meg: a bemenet Hermite-függvényekre vett Hermite-Fourier együtthatóinak sorozatát (azaz a VP réteg által létrehozott reprezentációt), a VPNet által adott besorolást, a valós osztályt, valamint a nyers jeleket egymásra rajzolva. A CBR-tesztelés során a tesztadatokhoz kerestem meg a legközelebbi három adatpontot a tanító halmazból. Ebben az esetben a bemeneti jel tényleges osztályát és a VPNet által jósolt osztályát is ismerjük, ami lehetővé teszi az algoritmus hatékonyságának értékelését. Természetesen éles alkalmazás során a valós osztály nem ismert, de a legközelebbi szomszédok osztályozási mintázatai alapján következtethetünk a bemenet helyes vagy helytelen besorolására. Ahogy a 3.2 ábrán is látszik, az implementáció által megtalált hasonló jelek VP utáni reprezentációja szinte azonos, ami az algoritmus robusztusságát és konzisztens viselkedését tükrözi. A VP utáni adatok kirajzolása lehetővé teszi az EEG jelek elemzésében kevésbé jártas megfigyelők számára is, hogy megértsék az osztályozási döntéseket. Ugyanakkor a nyers adatok kirajzolása 3.3 inkább a szakértők számára lehet hasznos, akik ezek alapján további elemzéseket végezhetnek.



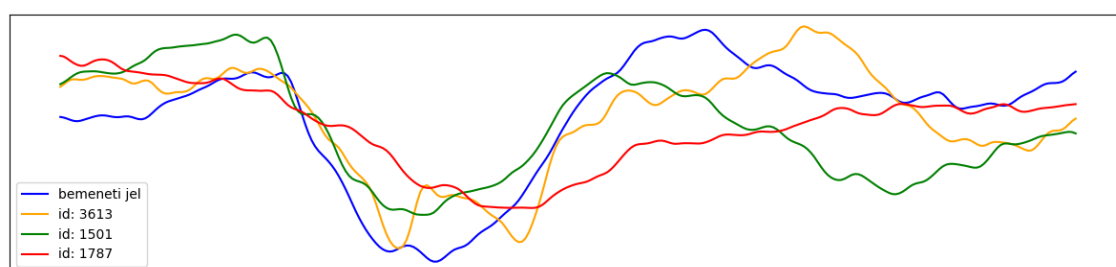
3.2. ábra. Példaalapú magyarázat a VP réteg utáni adatokkal, helyesen besorolt piros jel esetén.



3.3. ábra. Példaalapú magyarázat a bemeneti jelekkel, helyesen besorolt piros jel esetén.

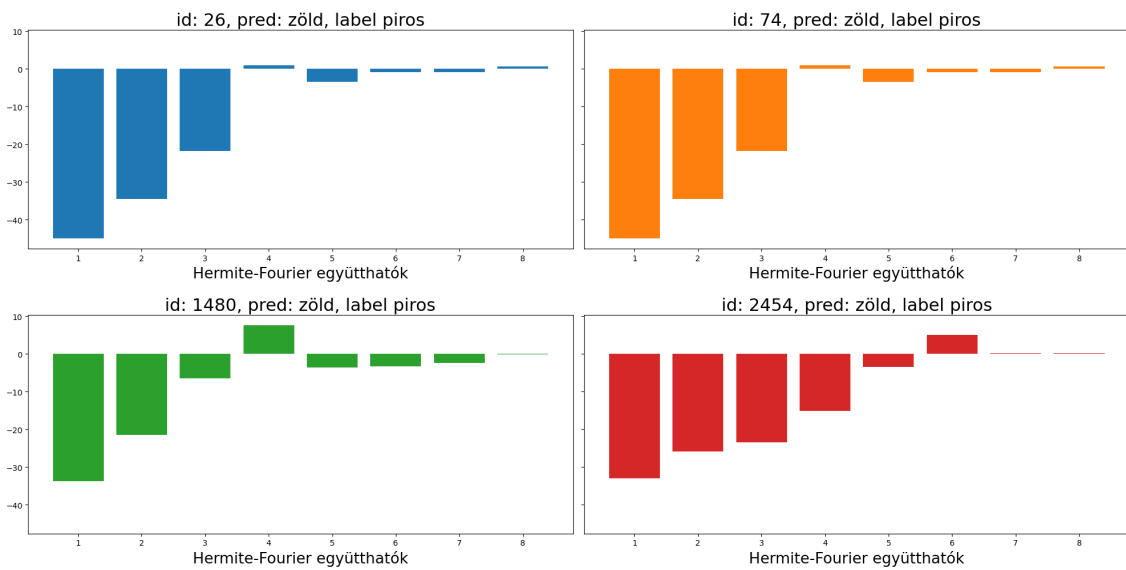


3.4. ábra. Példaalapú magyarázat a VP réteg utáni adatokkal helyesen besorolt zöld jel esetén.

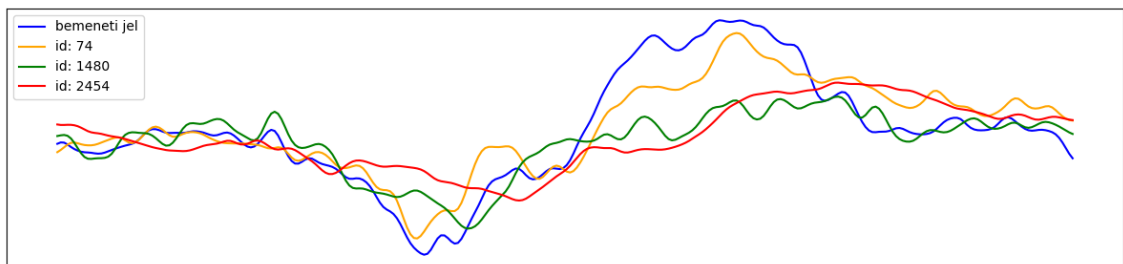


3.5. ábra. Példaalapú magyarázat a bemeneti jelekkel, helyesen besorolt zöld jel esetén.

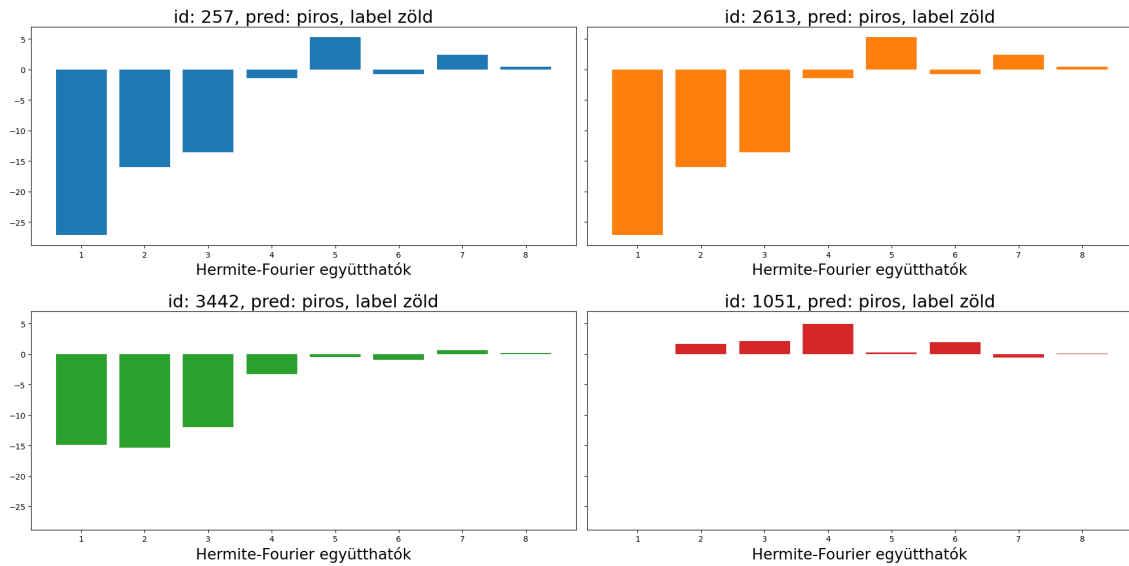
A 3.6 és a 3.8 ábrák olyan példákat mutatnak, ahol a bemeneti jelhez legközelebbi szomszédok mindegyike hibásan lett osztályozva, és maga a bemeneti jel is rosszul lett besorolva. Az általam végzett vizsgálatok során azt tapasztaltam, hogy amikor a CBR bemenetének olyan példákat választottam, ahol a VPNet hibázott, akkor a legközelebbi szomszédok többségénél is hibás volt az osztályozás. Ez azt jelenti, hogy a VPNet éles környezetben való alkalmazása során, ha egy bemeneti jelhez hasonló példák osztályozásai is hibásak, akkor nagy valószínűséggel a bemeneti jelet is helytelenül sorolta be a modell. Ez a megfigyelés kiemeli a CBR hasznosságát nemcsak az osztályozási döntések megértésében, hanem a potenciális hibák felismerésében is.



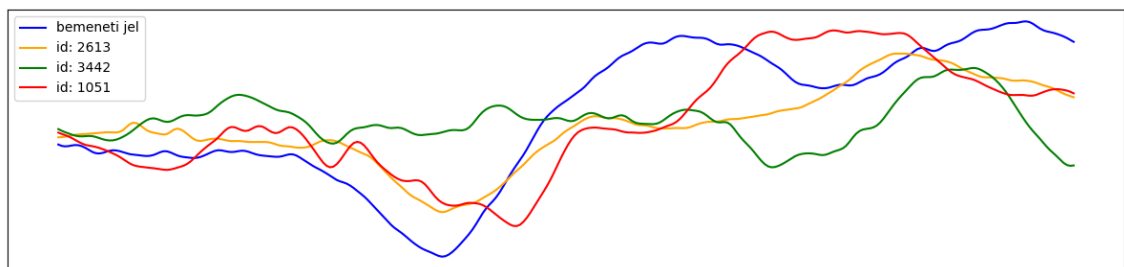
3.6. ábra. Példaalapú magyarázat a VP réteg utáni adatokkal helytelenül besorolt piros jel esetén.



3.7. ábra. Példaalapú magyarázat a bemeneti jelekkel, helytelenül besorolt piros jel esetén.



3.8. ábra. Példaalapú magyarázat a VP réteg utáni adatokkal helytelenül besorolt zöld jel esetén.



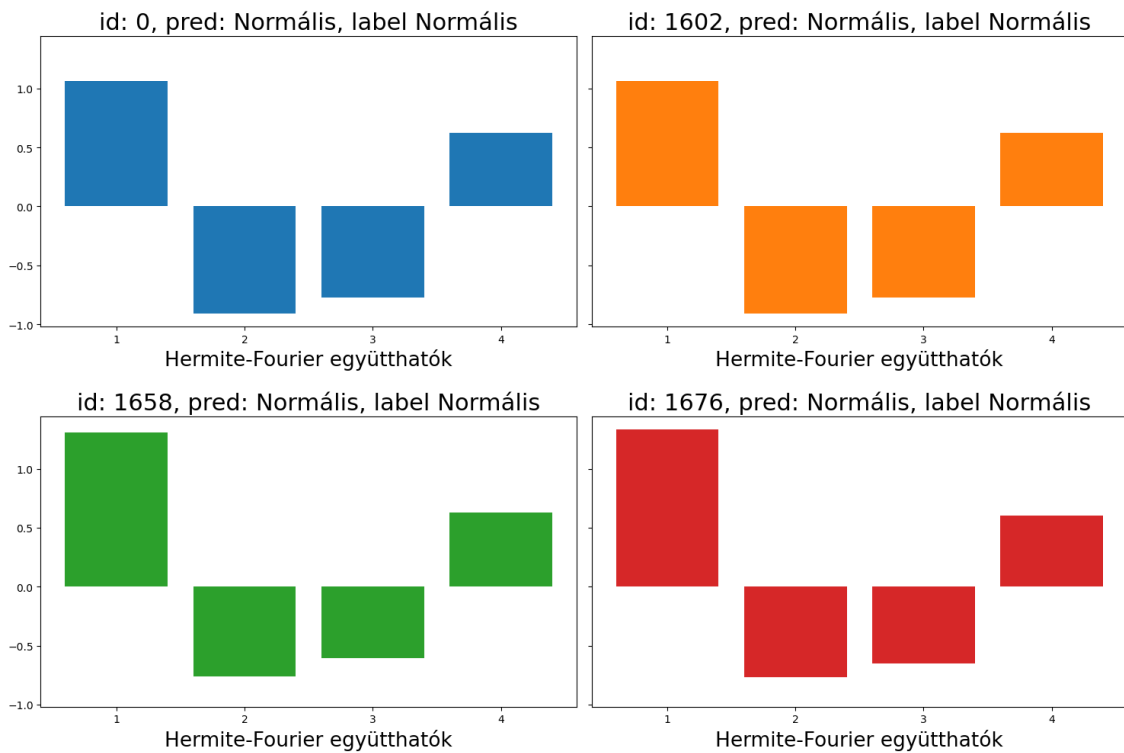
3.9. ábra. Példaalapú magyarázat a bemeneti jelekkel, helytelenül besorolt zöld jel esetén.

Összegzésül az EEG jelek esetében a CBR módszer sikeresen találta meg a hasonló adatokat, ezzel segítve az osztályozási döntések megértését és a lehetséges hibák felismerését. A dolgozat során megfigyeltem, hogy az adathalmazban a VP réteg után létrejött reprezentációkban az utolsó három Hermite-komponens értéke szinte mindig nulla volt. Ez arra utal, hogy a modell kevesebb Hermite függvénnyel is elég jól közelíti a jeleket. Ezáltal felmerül a kérdés, hogy ezen Hermite-függvények elhagyása esetén javulna-e a VPNet hatékonysága? További vizsgálatok szükségesek annak meghatározására, hogy a felesleges komponensek kizárása hogyan befolyásolná a modell teljesítményét.

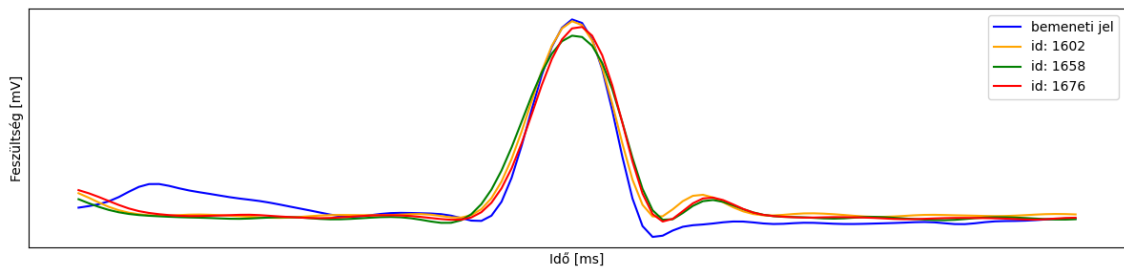
EKG adathalmaz

Az EKG adathalmaz feldolgozása során a VPNet modell a bemeneti jelek reprezentálására 4 Hermite-függvényt alkalmaztak a VP rétegben [2] amit a vizsgált QRS komple-

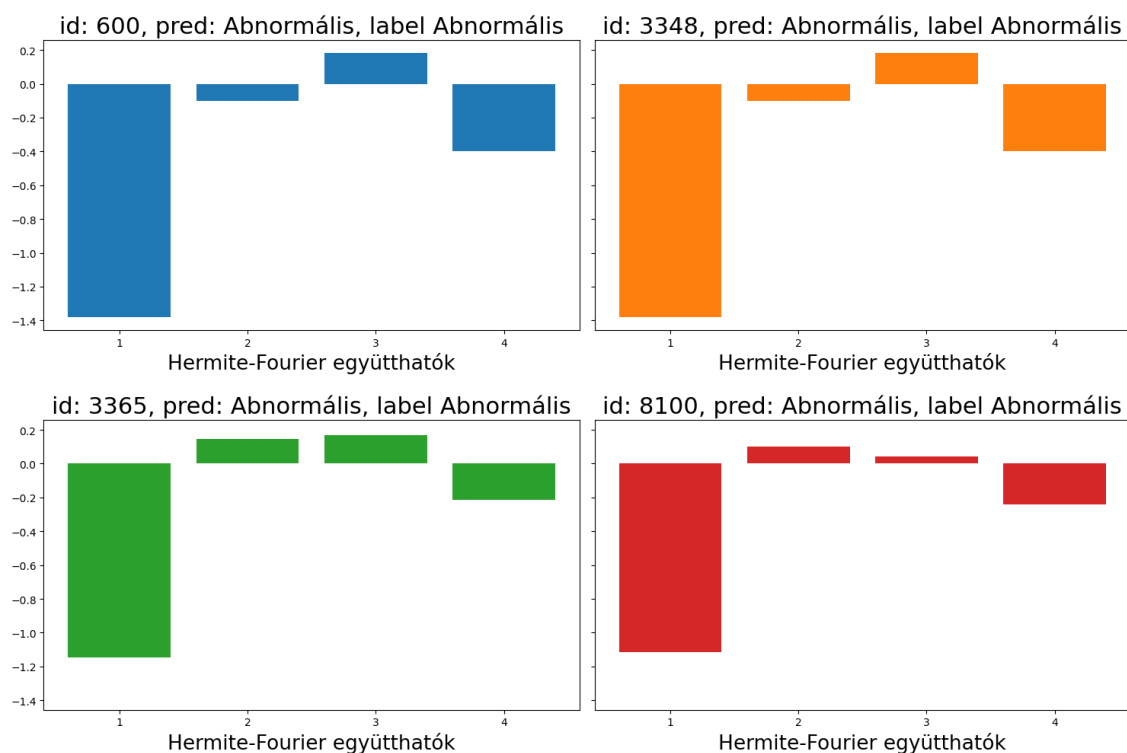
xumok EEG jelekhez viszonyított egyszerűbb morfológiája tett lehetővé. Az általam implementált CBR (Case-Based Reasoning) módszer itt is a három legközelebbi szomszédot keresi meg a tanító adathalmazból minden bemeneti jelhez. A szomszédoknál a következő elemeket jelenítem meg: a Hermite-függvények komponenseit (azaz a VP réteg által létrehozott reprezentációt), a VPNet által adott osztályozási eredményt, a valós osztályt, valamint a nyers jeleket egymásra rajzolva. A megfigyeléseim alapján az implementáció által megtalált hasonló jelek VP utáni reprezentációi itt is szoros egyezést mutattak. Itt is megfigyelhető, hogy amikor a VPNet ront, a például hozott adatpontok többségénél is rontott, így itt is igaz, hogy a hibák felismerésében is segít a rendszer.



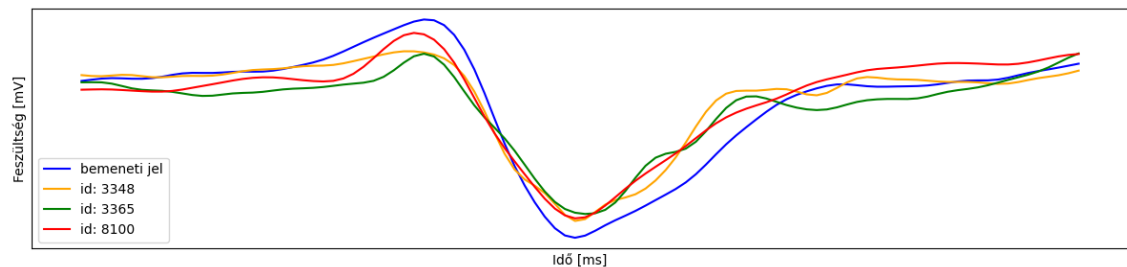
3.10. ábra. Példaalapú magyarázat a VP réteg utáni adatokkal helyesen besorolt normális szívverés esetén.



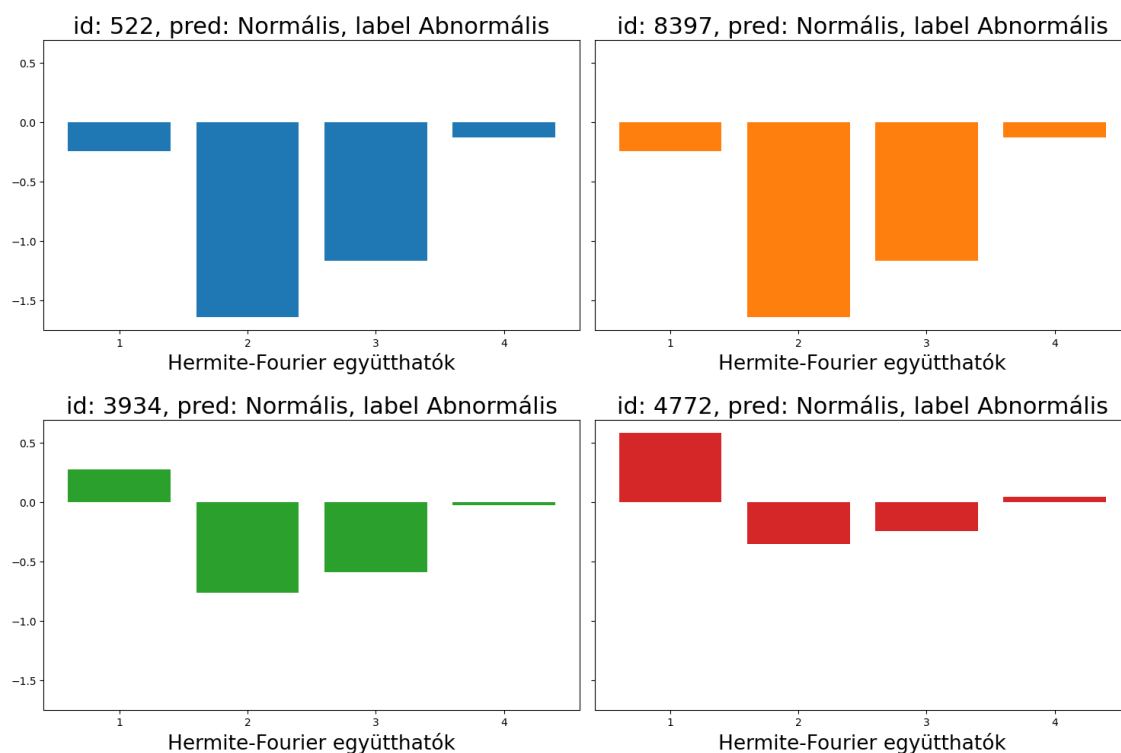
3.11. ábra. Példaalapú magyarázat a bemeneti jelekkel, helyesen besorolt normális szívverés esetén.



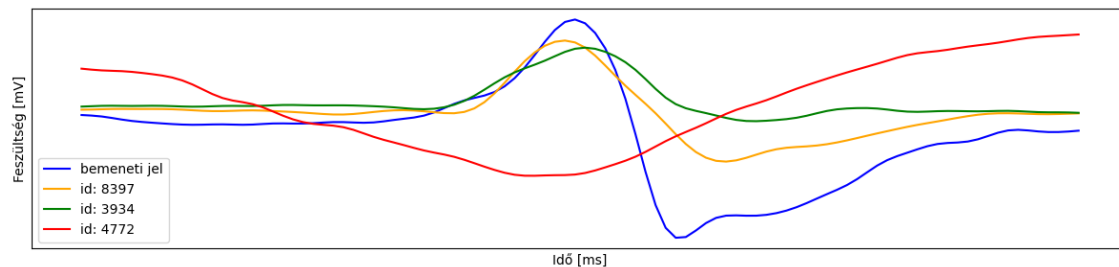
3.12. ábra. Példaalapú magyarázat a VP réteg utáni adatokkal helyesen besorolt abnormális szívverés esetén.



3.13. ábra. Példaalapú magyarázat a bemeneti jelekkel, helyesen besorolt abnormális szívverés esetén.



3.14. ábra. Példaalapú magyarázat a VP réteg utáni adatokkal helytelenül besorolt abnormális szívverés esetén.



3.15. ábra. Példaalapú magyarázat a bemeneti jelekkel, helytelenül besorolt abnormális szívverés esetén.

4. fejezet

Dimenziócsökkentési módszerek

4.1. Dimenziócsökkentés szükségessége

A magas dimenzionalitás problémái

A gépi tanulás és a neurális hálózatok alkalmazása során az adatok dimenzionalitása kulcsfontosságú tényezővé válik. A modern rendszerek gyakran rendkívül nagy mennyiségű, magas dimenziójú adatot dolgoznak fel, például képi, hang- vagy jelfeldolgozás során. Ez azonban számos problémát vet fel, amelyek megnehezítik a hatékony modellépítést és a pontos osztályozást [10]. Sokdimenziós adatok osztályozásánál fennál a veszély, hogy a modell túltanul, azaz a teszt adatokat nagy hibával prediktálja. Ez azt jelenti, hogy a modell megtanul olyan zajokat vagy irreleváns mintázatokat, amelyek nem általánosíthatók a tesztalmazra. Ez csökkenti a modell pontosságát az új adatokkal szemben, mivel az általánosítási képesség gyengül. Emellett a sokdimenziós adatoknál a számítási költségek megnövekednek, azaz a feldolgozás jelentős erőforrást igényel, így csökken a rendszer skálázhatósága. Továbbá az adatok vizualizációja és a mintázatok azonosítása több dimenzióban rendkívül bonyolult. A vizualizációs technikák alkalmazása alacsonyabb dimenziójú térben segíthet megérteni az adatok szerkezetét és azok szeparálhatóságát.

Dimenziócsökkentés a VPNet-ben

A VPNet-ben a VP réteg betölt egy dimenziócsökkentési funkciót, hiszen a bemeneti adatokat adaptív Hermite-függvények segítségével reprezentálja, így az ezt követő NN rétegnek már nagyságrendekkel kevesebb dimenziót kell feldolgoznia. Ezáltal a rendszer

kevésbé lesz hajlamos az ebből adódó problémákra (túltanulás, számítási költségek megnövekedése)

4.2. Alkalmazott Dimenziócsökkentési Technikák

Az adatok szeparáltságának kvalitatív vizsgálatához dimenziócsökkentési technikák alkalmazása szükséges, hogy az adatpontokat 2D-ben vagy 3D-ben vizualizálva megérthessük azok szerkezetét és mintázatait. Ahhoz, hogy lássuk a VP réteg hatását, a nyers adatokat és a VP réteg alkalmazása utáni kimenetet 2- illetve 3 dimenzióra csökkentjük. Számos módszer létezik a dimenziók csökkentésére [10], amelyeket kettő fő csoportra bonthatunk: lineáris és nemlineáris dimenziócsökkentési módszerek. A lineáris dimenziócsökkentési módszerek feltételezik, hogy az adatok egy magas dimenziójú vektortérben helyezkednek el, de egy alacsonyabb dimenziójú altér elemei jó minőségben közelítik őket. Ezek a módszerek főként a variancia és a korreláció vizsgálatára támaszkodnak. A nemlineáris módszerek abból indulnak ki, hogy az adatok magas dimenziójú térben manifoldszerűen helyezkednek el, és ezeket az összetett mintázatokat próbálják megőrizni egy alacsonyabb dimenziós térben. A "manifoldszerű elrendezés" kifejezés azt jelenti, hogy az adatok egy bonyolultabb, görbült struktúrában helyezkednek el, amely lokálisan egy egyszerűbb, alacsonyabb dimenziós térben ábrázolható, de globálisan összetett mintázatot mutat. A lineáris módszerek általában gyorsabbak és egyszerűbbek, ugyanakkor korlátozottak, ha az adatok szerkezete nemlineáris. A nemlineáris módszerek jobb eredményeket nyújthatnak az ilyen esetekben, de gyakran nagyobb számítási költséggel járnak, és nehezebben skálázhatók nagy adathalmazokra.

4.2.1. t-SNE (t-distributed Stochastic Neighbor Embedding)

A *t-SNE*-t nemlineáris dimenziócsökkentő módszerek közé sorolhatjuk [11]. Az adatok közötti páronkénti távolságokat a magas dimenziós térben valószínűségekkel alakítja. Ezt a valószínűség-eloszlást megpróbálja a lehető legjobban megközelíteni egy alacsony dimenziós térben. Jól megőrzi az adatpontok közötti lokális struktúrát, így különösen hasznos az adatok vizualizációjához. Lehetővé teszi a komplex mintázatok és szeparálhatóságok vizsgálatát is.

A t-SNE működési alapelvei:

- **Valószínűség-eloszlás létrehozása:** A magas dimenziós térben az adatpontok közötti páronkénti távolságokat valószínűségekkel alakítja, amelyek azt fejezik ki, hogy két pont mennyire lehet egymás szomszédja.
- **Valószínűség-eloszlás létrehozása:** A magas dimenziós térben az adatpontok közötti páronkénti távolságokat valószínűségekkel alakítja, amelyek azt fejezik ki, hogy két pont mennyire lehet egymás szomszédja. Ehhez Gauss-eloszlást használ, amely a következő képlettel számítja ki a két adatpont közötti valószínűséget:

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)}$$

Ahol $\|x_i - x_j\|$ a két adatpont közötti Euklideszi távolság, σ_i pedig a Gauss-eloszlás szórása, amely az adott adatpontra van központosítva.

- **Alacsony dimenziós tér modellezése:** Egy hasonló valószínűség-eloszlást próbál létrehozni az alacsony dimenziós térben, amely megközelíti a magas dimenziós tér eloszlását.
- **Optimalizáció Kullback-Leibler divergenciával:** A két eloszlás közötti különbséget minimalizálja a Kullback-Leibler divergencia [12]. Ez a mérőszám azt méri, hogy mennyivel tér el egy eloszlás a másiktól. A t-SNE esetében a cél az, hogy az alacsony dimenziós eloszlás minél jobban közelítse a magas dimenziós eloszlást, így az alacsony dimenziós ábrázolás pontosabban tükrözi az adat szerkezetét.

A t-SNE legfontosabb hiperparamétere a *perplexity*, amely a helyi és globális struktúrák közötti egyensúlyt szabályozza. Általában 5 és 50 közötti értéket vesz fel. Ezen kívül a célként kitűzött dimenziószám szintén kulcsfontosságú: vizualizációkhoz jellemzően 2 vagy 3 dimenziót választanak. A t-SNE előnye, hogy rendkívül intuitív vizualizációt biztosít, mivel az adatokat csoportokba rendezi, amelyek jól mutatják a lokális hasonlóságokat. Hátránya, hogy a nagy adathalmazoknál jelentős számítási költsége lehet, és a globális struktúrák nem mindig maradnak meg. Az utóbbi miatt döntöttem úgy, hogy a szeparálhatósági vizsgálatokhoz más típusú dimenziócsökkentési módszereket is alkalmazok, hogy egy teljesebb képet kaphassak a VP réteg ezen hatásáról.

4.2.2. UMAP (Uniform Manifold Approximation and Projection)

Az UMAP egy korszerű nemlineáris dimenziócsökkentési módszer, amely az adatok topológiai struktúráját próbálja megőrizni az alacsony dimenziós térben [13]. Az UMAP célja, hogy egyszerre biztosítsa a lokális és a globális struktúrák megőrzését, miközben gyorsabb és skálázhatóbb, mint a t-SNE.

Az UMAP működési alapelvei:

- **Adatok topológiai reprezentációja:** A magas dimenziójú adatokat egy gráfként reprezentálja, amely a lokális kapcsolatokon alapul.
- **Gráf leképezése alacsony dimenziós térbe:** Az adatok közötti kapcsolatok megőrzésére törekszik egy optimális alacsony dimenziós térben.
- **Topológiai szerkezet megőrzése:** Az UMAP, ahogy a nevében is benne van, a manifold-szerkezeteket is figyelembe veszi, azaz a magas dimenziójú térben lévő adatok alapszerkezetét (a geometriai és topológiai összefüggéseket) próbálja megőrizni az alacsony dimenziós térben, így jobban reprezentálja az adatok globális mintázatait.

Paraméterei:

- **n_neighbors:** Az egyik legfontosabb hiperparaméter, amely meghatározza, hogy hány szomszédot vegyen figyelembe az algoritmus az egyes pontok helyi szerkezetének modellezéséhez. Kisebb érték esetén a leképezés inkább lokális, míg nagyobb érték esetén globális szerkezeti információk is figyelembe vehetők. Általában 5 és 50 közötti értékek használata ajánlott.
- **n_components:** A céldimenziók száma, amelyre az adatok dimenziócsökkentése történik. Vizualizációs célokra gyakran 2 vagy 3 dimenziót választanak.
- **metric:** Ez a paraméter határozza meg, hogy milyen módszert alkalmazzon az algoritmus az adatpontok közötti távolság kiszámításához. Leggyakrabban az Euklideszi távolságot használják.

Az UMAP előnyei közé tartozik a gyorsabb számítási sebesség és a skálázhatóság nagy adathalmazok esetében. Az alacsony dimenziós térben történő reprezentációja gyakran stabilabb és jobban megőrzi a globális szerkezeteket, mint a t-SNE.

4.2.3. PCA (Principal Component Analysis)

A PCA lineáris dimenziócsökkentési módszer, amely az adatok varianciáját maximalizáló ún. főkomponenseket azonosítja [14]. Az eredeti adatokat egy új koordinátarendszerbe transzformálja, amelyben a főkomponensek mentén történik az adatok leképezése. Ezáltal csökkenti a dimenziók számát, miközben arra törekszik, hogy az információ legnagyobb része (az adatok varianciája) megmaradjon az alacsonyabb dimenziós térben.

A PCA működési alapelvei:

- **Adatok középértékre normalizálása:** Az adathalmaz minden dimenziójának középértéke levonásra kerül, hogy az adatok az origó körül koncentrálódjanak.
- **Kovarianciamátrix létrehozása:** Az adatok közötti kapcsolatok és varianciák meghatározására kovarianciamátrixot számol. Ez egy olyan mátrix, amely az adatok dimenziói közötti összefüggéseket reprezentálja. A diagonális elemei az egyes dimenziók varianciáját mutatják, míg a nem diagonális elemek az adatok különböző dimenziói közötti kovarianciát írják le, azaz, hogy az egyik dimenzió értékei milyen mértékben változnak együtt a másikkal.
- **Sajátvektorok és sajátértékek kiszámítása:** A kovarianciamátrix elemzése során meghatározzuk a sajátvektorokat, amelyek megadják, hogy az adatok mely irányokban tartalmazzák a legtöbb információt. Az sajátértékek azt mutatják meg, hogy ezek az irányok mennyire fontosak, azaz az adatok varianciájának mekkora része esik az adott irányra. Ez alapján a sajátvektorok segítségével kiválaszthatók az adatok legfontosabb mintázatai.
- **Főkomponensek kiválasztása:** A sajátértékek alapján rangsorolja a komponenseket, majd azokat az irányokat választja ki, amelyek a legnagyobb varianciát magyarázzák.

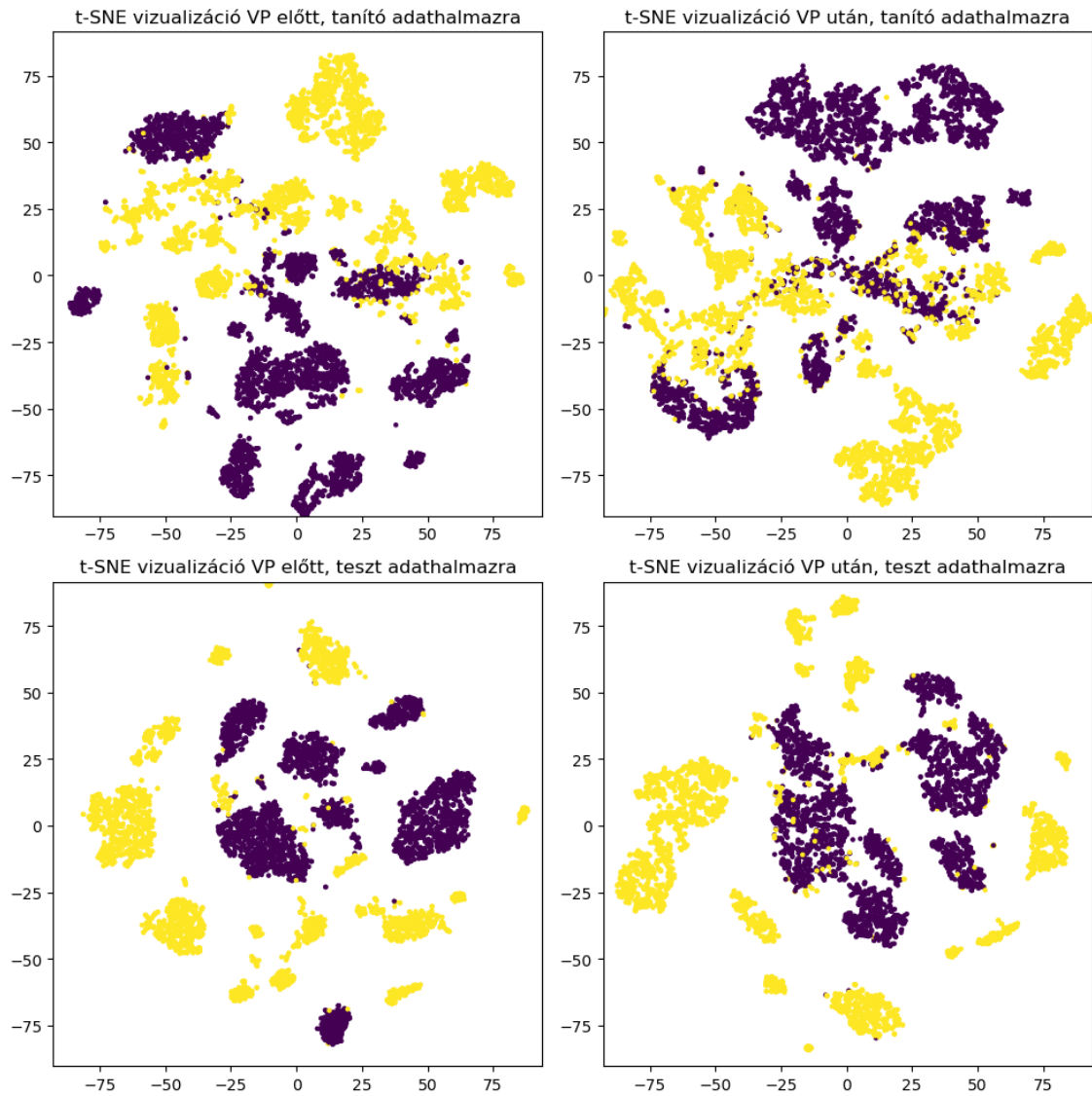
A PCA a t-SNE-hez képest alacsony számításigényű módszer, amelynek alkalmazása főleg lineárisan szétválasztható adatoknál lehet hasznos.

4.3. Eredmények

4.3.1. EKG adathalmaz

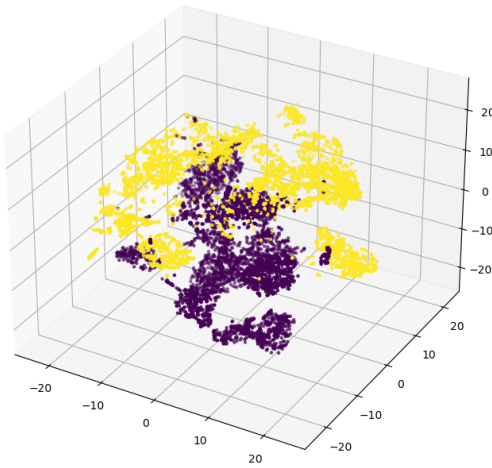
t-SNE vizualizációk

A t-SNE vizualizációk azt mutatják, hogy az EKG adatok már a VP réteg alkalmazása előtt is jól szeparálódnak, az egyes osztályok között vizuálisan egyértelmű klaszterek figyelhetők meg. Ez kihívást jelent annak megítélésében, hogy a VP réteg milyen mértékben járult hozzá az adatok további szeparációjához. Bár a VP réteg potenciális hatását nehéz vizuálisan kiértékelni, a szeparálhatóság finom javulása továbbra is feltételezhető, mivel látszólag a VP réteg utáni kisebb az átfedés a kettő csoport tulajdonságvektorai (feature vektorai) között. Fontos megjegyezni, hogy a két osztály között nem lehet egyértelmű, egyenes határt húzni, ami arra enged következtetni, hogy az adatok lineárisan nem szeparálhatóak. A t-SNE implementáció során a lehető legjobb vizuális hatás érdekében több perplexity értéket is kipróbáltam 5 és 50 között, amelyek közül a 30-at találtam a legoptimálisabbnak.

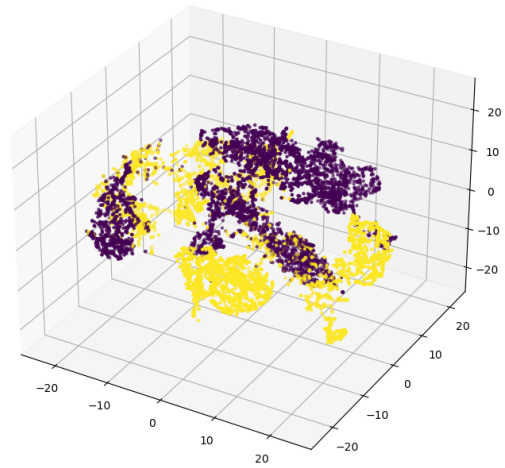


4.1. ábra. EKG adathalmazok t-SNE-vel 2D-re képezve.

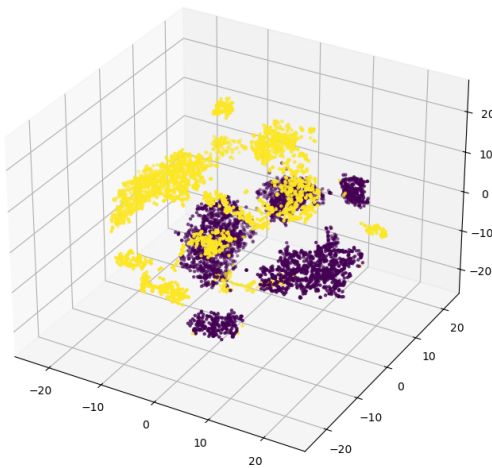
t-SNE vizualizáció VP előtt, tanító adathalmazra



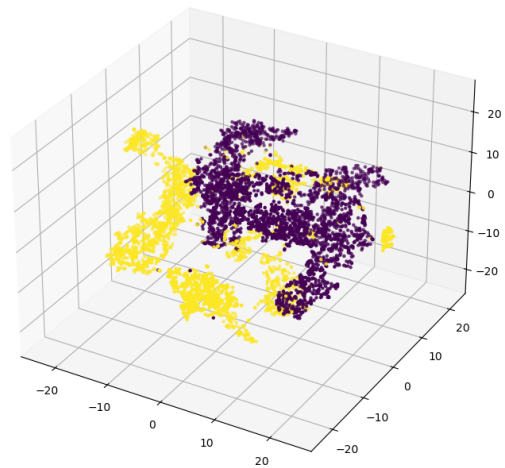
t-SNE vizualizáció VP után, tanító adathalmazra



t-SNE vizualizáció VP előtt, teszt adathalmazra



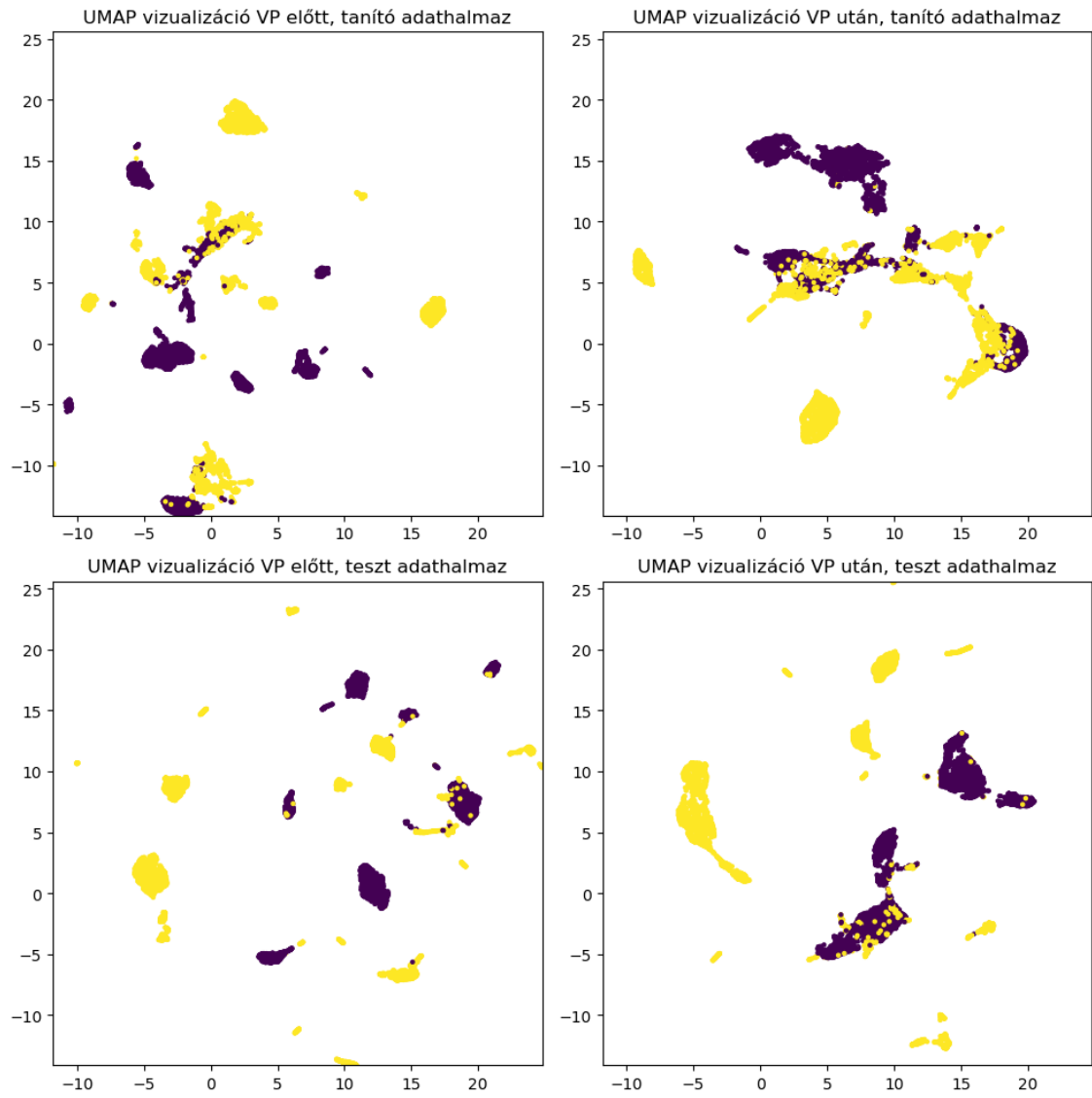
t-SNE vizualizáció VP után, teszt adathalmazra



4.2. ábra. EKG adathalmazok t-SNE-vel 3D-re képezve.

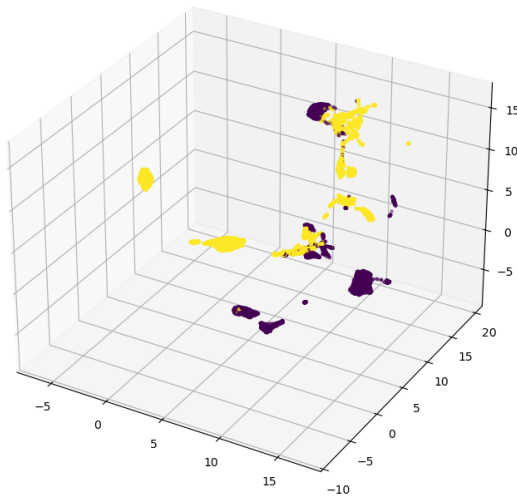
UMAP vizualizációk

A t-SNE-hez hasonlóan, az UMAP vizualizációknál is látszik, hogy az adathalmaz eredetileg is kompakt és jól szeparált részekből áll. Azonban a VP réteg alkalmazása után megfigyelhető, hogy a sok kisebb, széttöredezett halmaz helyett a modell inkább nagyobb, több elemű csoportokba rendezi az adatokat. Ez arra utal, hogy a VP réteg a lokális szerkezeteket globálisabbá alakítja, ezáltal az osztályok homogénebb, összefüggőbb struktúrákat alkotnak. A figyelembe vett szomszédok számát 15-re állítottam, így jobban őrzi meg a lokális struktúrákat. Azért erre esett a választás, mert így vált legjobban láthatóvá, hogy a szeparáltság mind a nyers, mind a VP réteg által transzformált adatokon jól érvényesül. Az adatok közötti távolság mérésére euklideszi távolságot használtam.

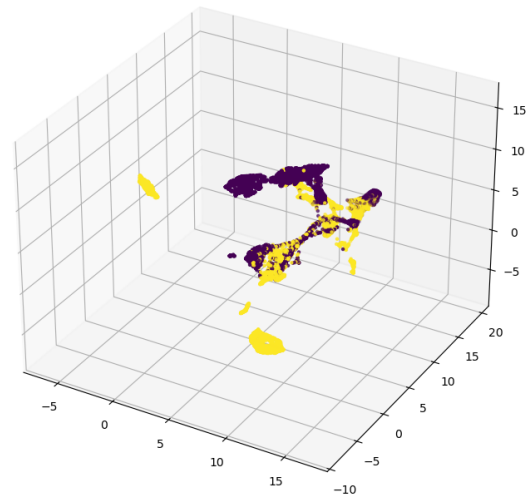


4.3. ábra. EKG adathalmazok UMAP-al 2D-re képezve.

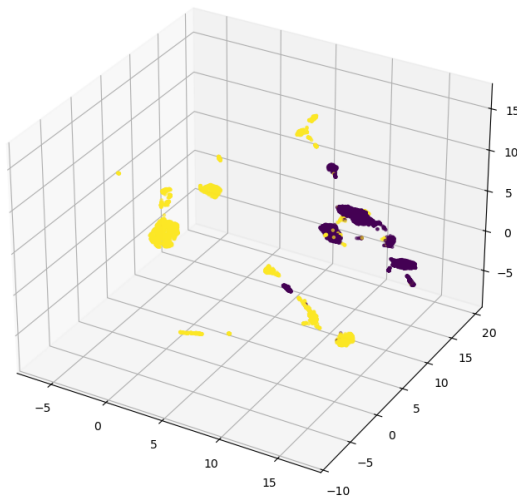
UMAP vizualizáció VP előtt, tanító adathalmazra



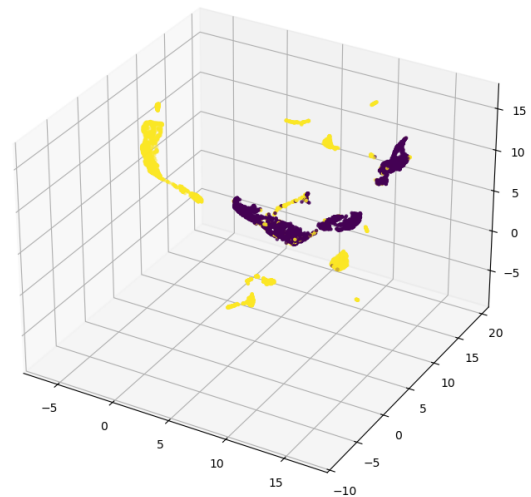
UMAP vizualizáció VP után, tanító adathalmazra



UMAP vizualizáció VP előtt, teszt adathalmazra



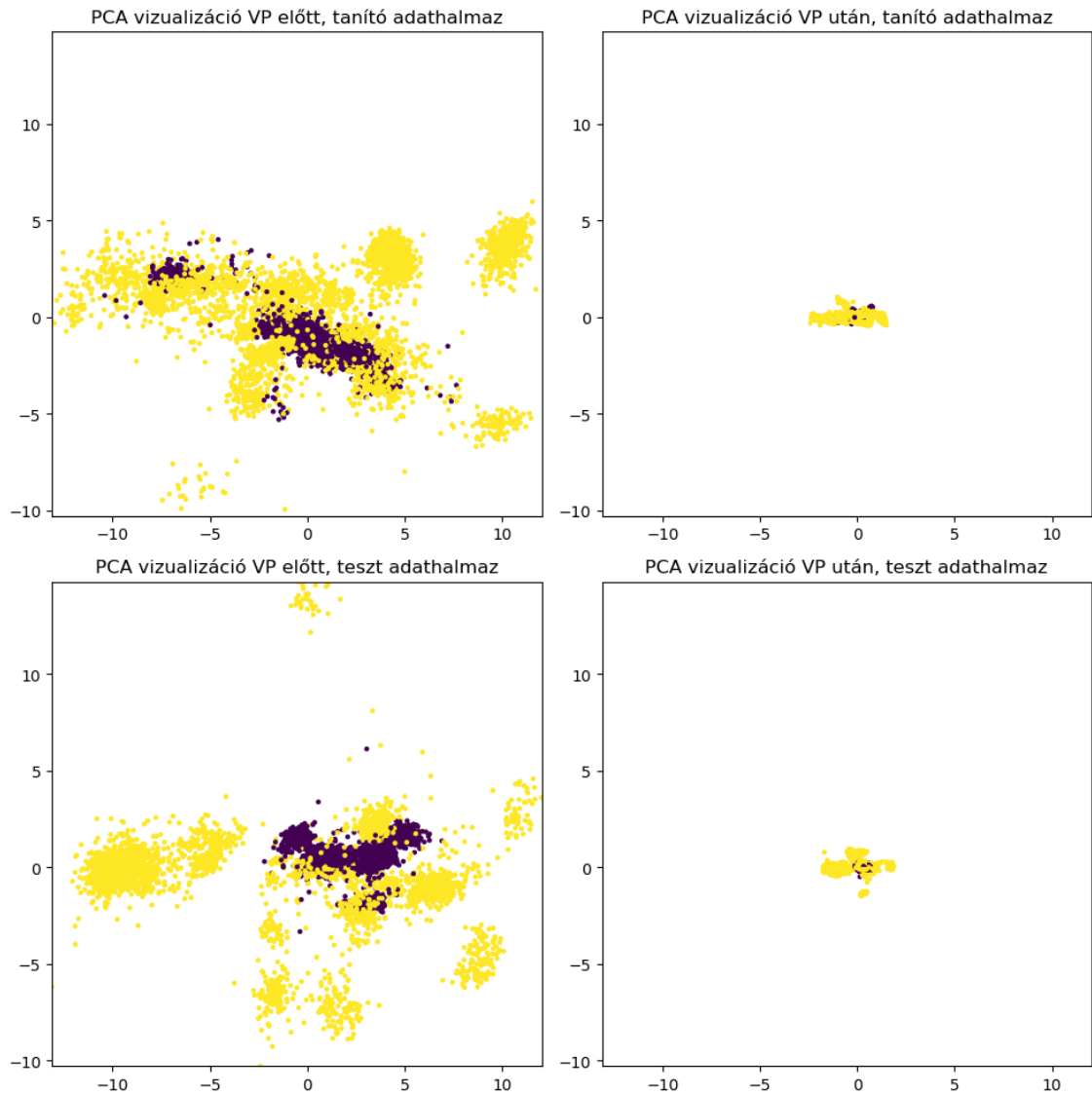
UMAP vizualizáció VP után, teszt adathalmazra



4.4. ábra. EKG adathalmazok UMAP-al 3D-re képezve.

PCA vizualizációk

Ahogy előre sejthető volt, a PCA által leképezett adatokon az osztályok nem különülnek el egymástól. Ez elsősorban abból adódik, hogy a vizsgált jelek nem lineárisan szeparálhatóak, így a PCA, mint lineáris dimenziócsökkentő módszer, nem képes megfelelően reprezentálni az adatok szerkezetét. Emiatt a lineáris elválasztási képessége korlátozott, és nem mutatja meg a jelek közötti rejtett nemlineáris összefüggéseket.

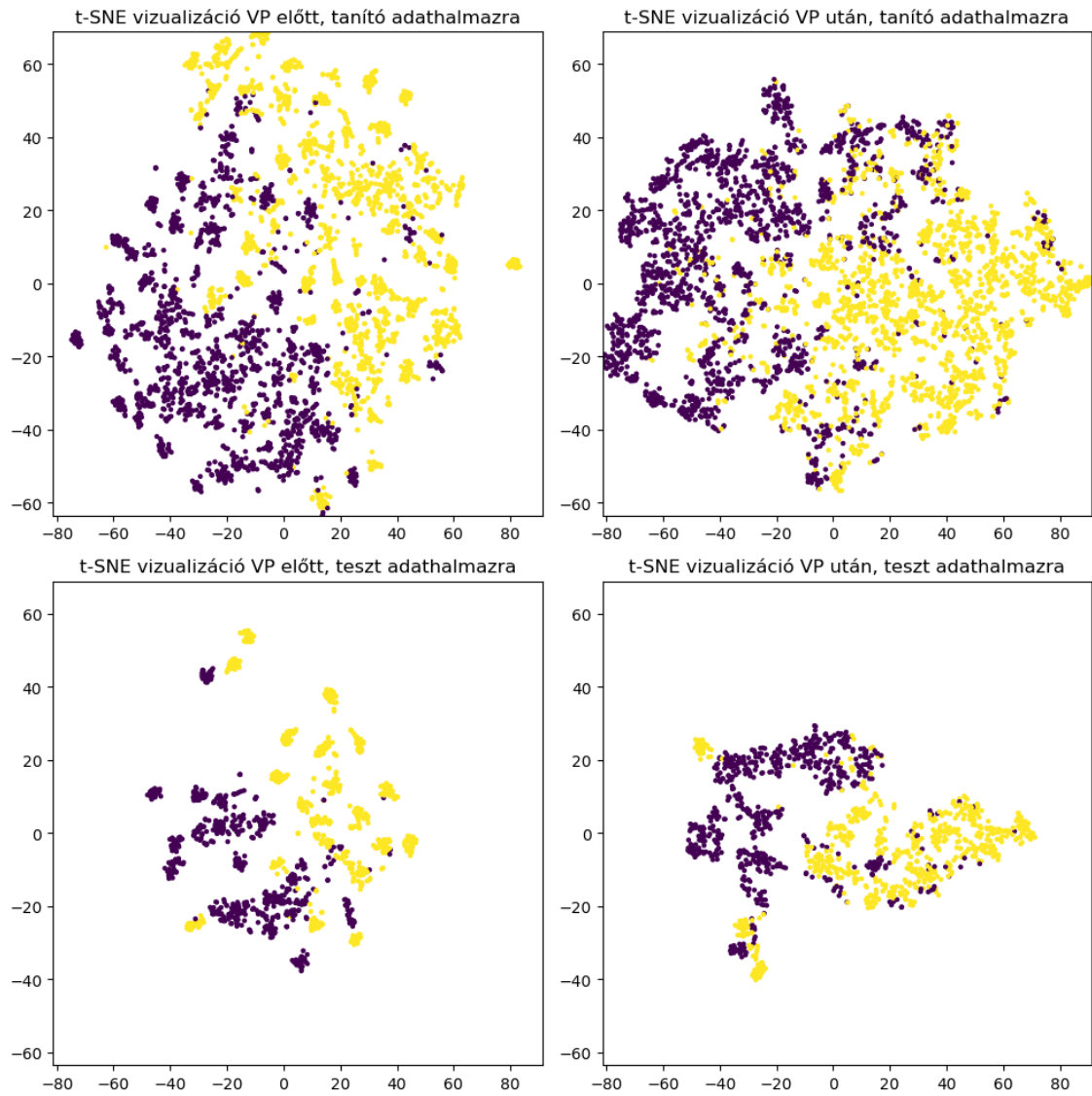


4.5. ábra. EKG jelek PCA-val 2D-re képezve.

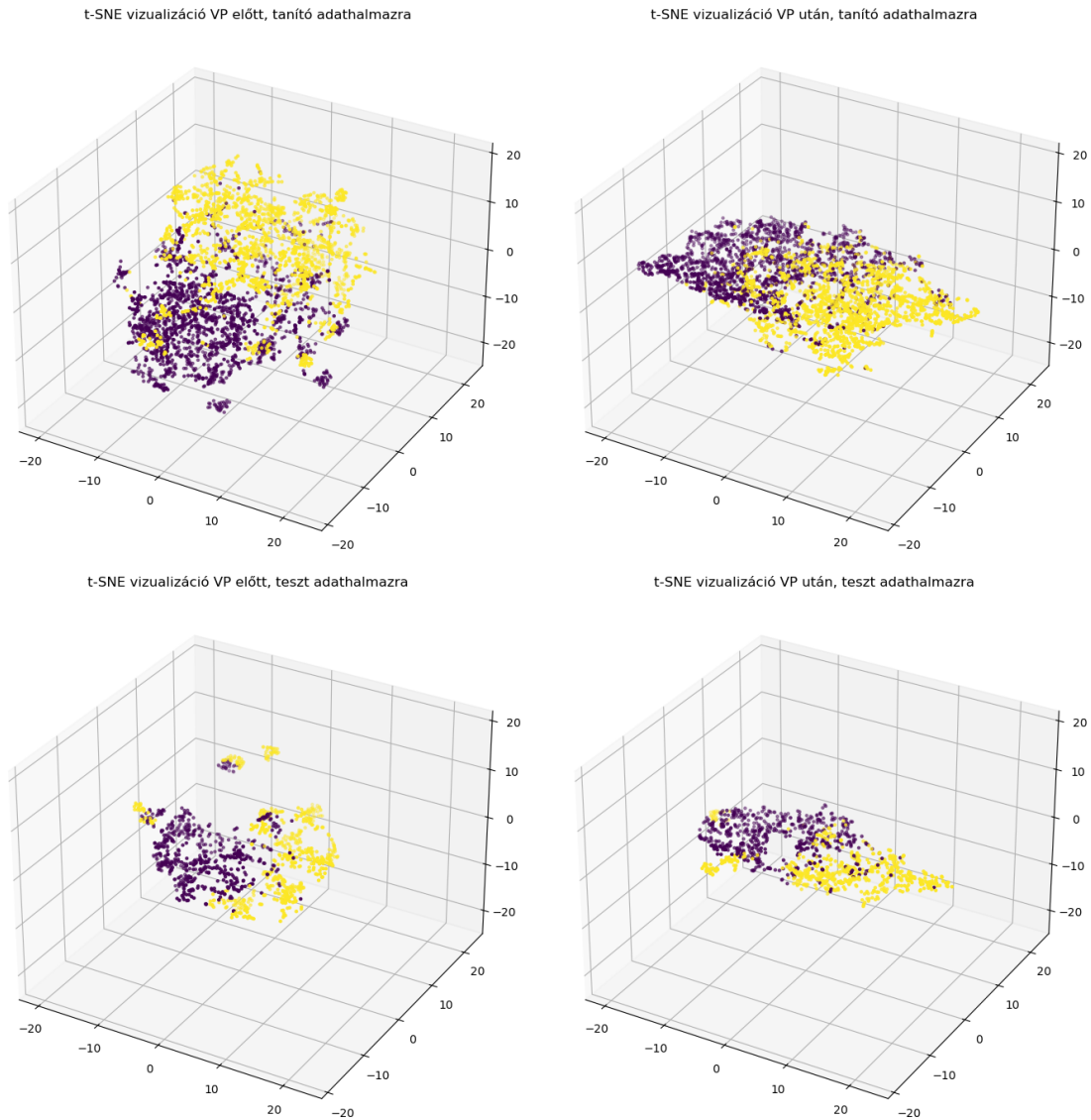
4.3.2. EEG adathalmaz

t-SNE vizualizációk

Hasonlóan az EKG jelekhez, itt is jól látszik, hogy a nyers adatok is szeparáltak bizonyos mértékig. Azonban az EKG adatokhoz képest lényeges különbség, hogy mind az adatok, mind a VP feature-ök lineárisan is elkülönülnek egymástól. A VP réteg alkalmazása után a lineáris szeparáltság megmarad, sőt megfigyelhető, hogy az osztályok közötti átfedés látszólag tovább csökkent. Ez azt jelzi, hogy a VP réteg elősegíti az osztályok közötti tisztább határok kialakulását. A perplexity értéket itt is 30-ra állítottam.



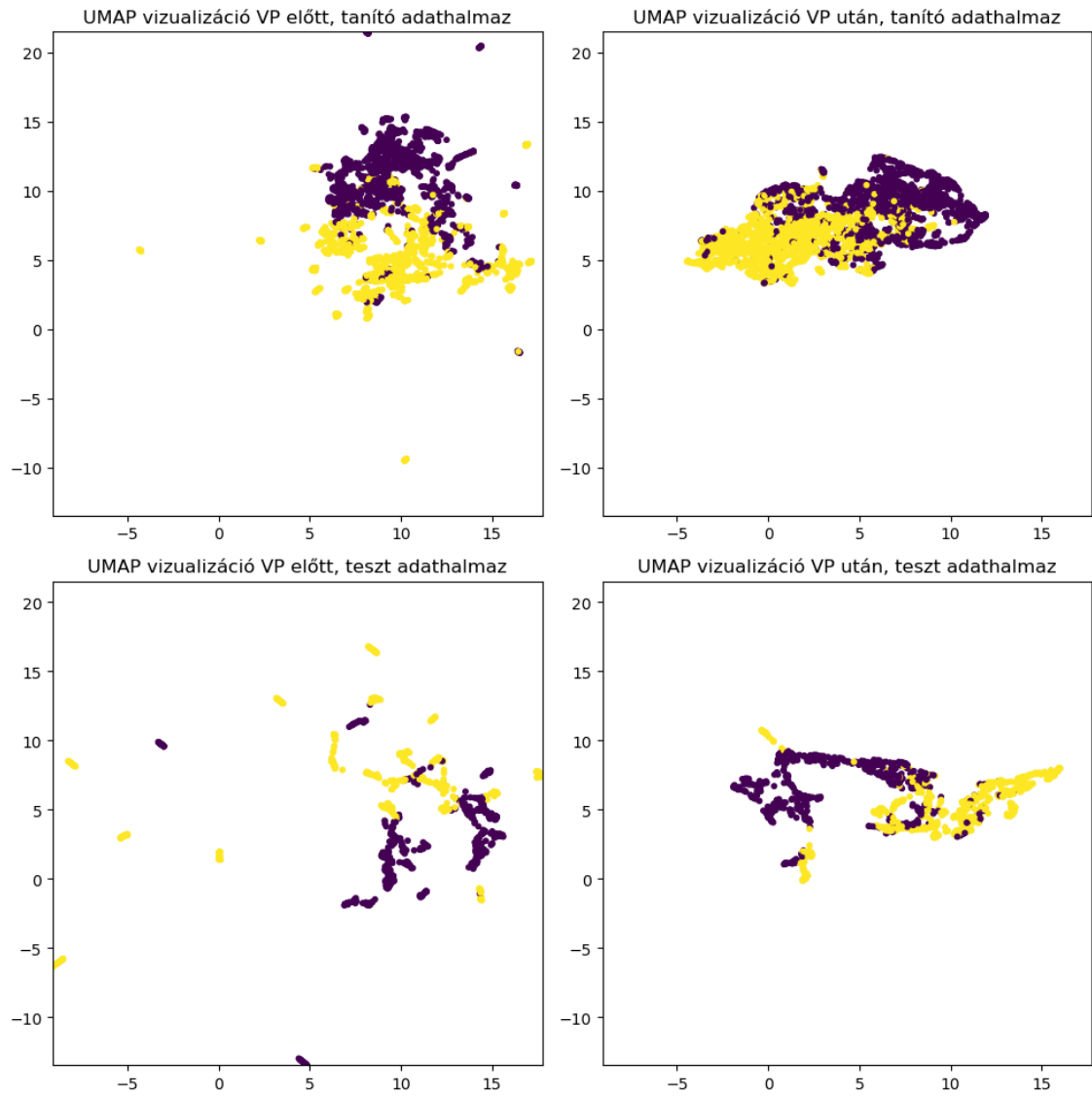
4.6. ábra. EEG adathalmazok t-SNE-vel 2D-ra képezve.



4.7. ábra. EEG adathalmazok t-SNE-vel 3D-re képezve.

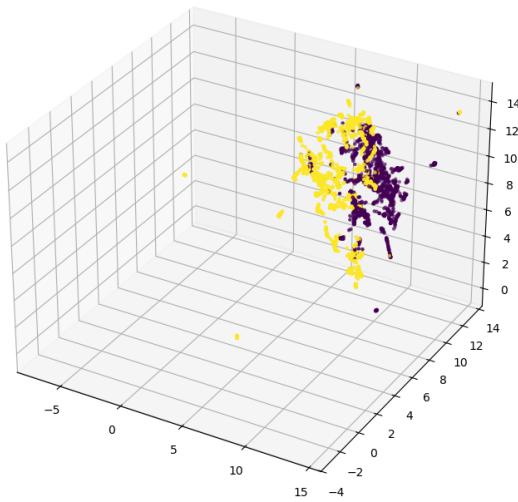
UMAP vizualizációk

A t-SNE-hez hasonlóan az UMAP esetében is megfigyelhető, hogy az adatok már a VP réteg alkalmazása előtt is jól szeparáltak. A nyers adatokon az osztályok különböző csoportokat alkotnak, ami a jól szeparálhatóság alapját képezi. A tesztadatokon pedig az EKG jelekhez hasonló mintázat figyelhető meg: míg a VP réteg előtt az adatok több kisebb klaszterre oszlanak, a VP réteg alkalmazása után ezek a kisebb klaszterek összeolvadnak, és az adatok két nagyobb, tisztábban elkülönülő csoportba rendeződnek át. Ez azt mutatja, hogy a VP réteg képes egységesebb szerkezetet kialakítani az osztályok között.

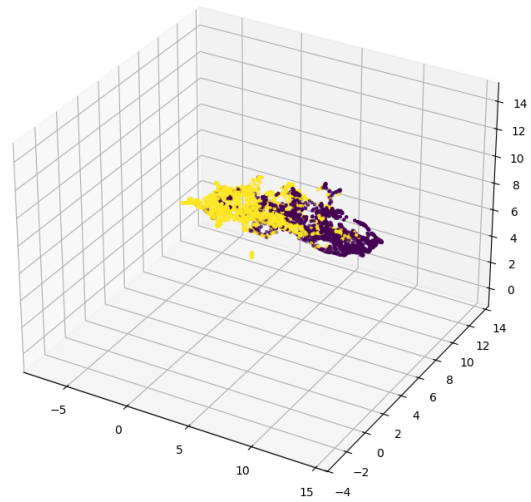


4.8. ábra. EEG adathalmazok UMAP-al 2D-re képezve.

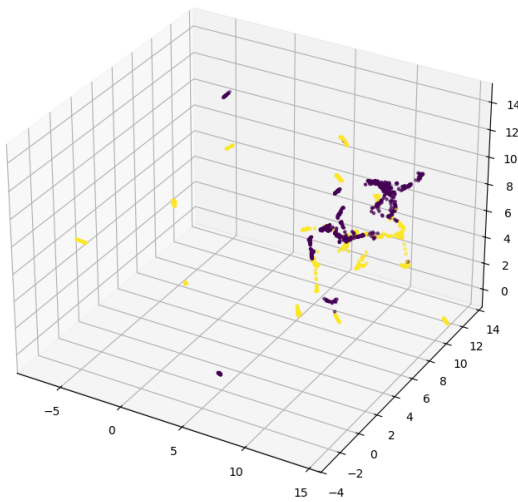
UMAP vizualizáció VP előtt, tanító adathalmazra



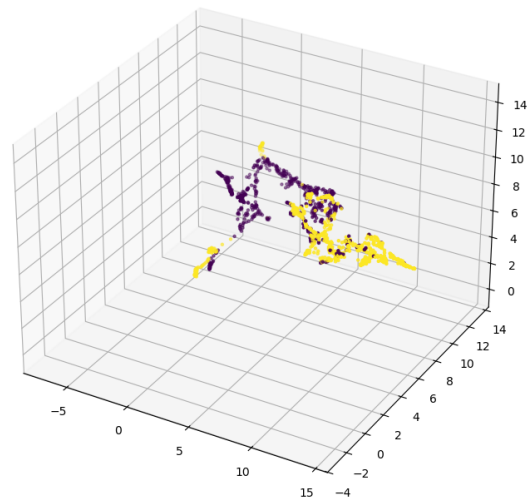
UMAP vizualizáció VP után, tanító adathalmazra



UMAP vizualizáció VP előtt, teszt adathalmazra



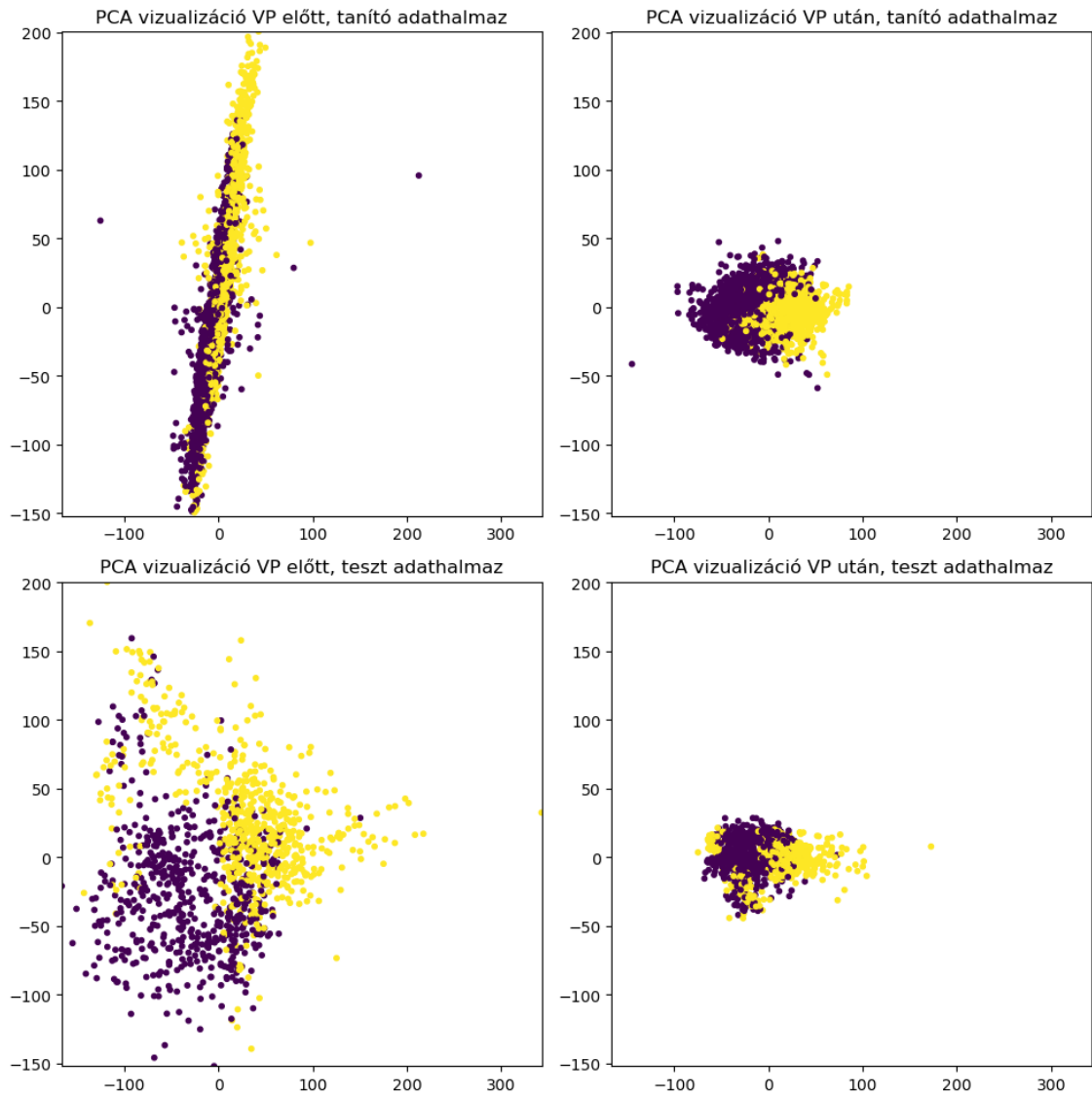
UMAP vizualizáció VP után, teszt adathalmazra



4.9. ábra. EEG adathalmazok UMAP-al 3D-re képezve.

PCA vizualizációk

Hasonlóan az EKG jelekhez, sajnos a PCA által leképezett adatokból itt sem vonhatunk le sokatmondó következtetéseket. Ennek ellenére itt jól látszik, hogy az adatok szeparáltak. Azt gondolom ennek az lehet az oka, hogy az EEG jelek az EKG-val ellentétben lineárisan is sokkal jobban elkülönülnek, és ezt figyelhetjük meg itt a PCA leképezésnél.

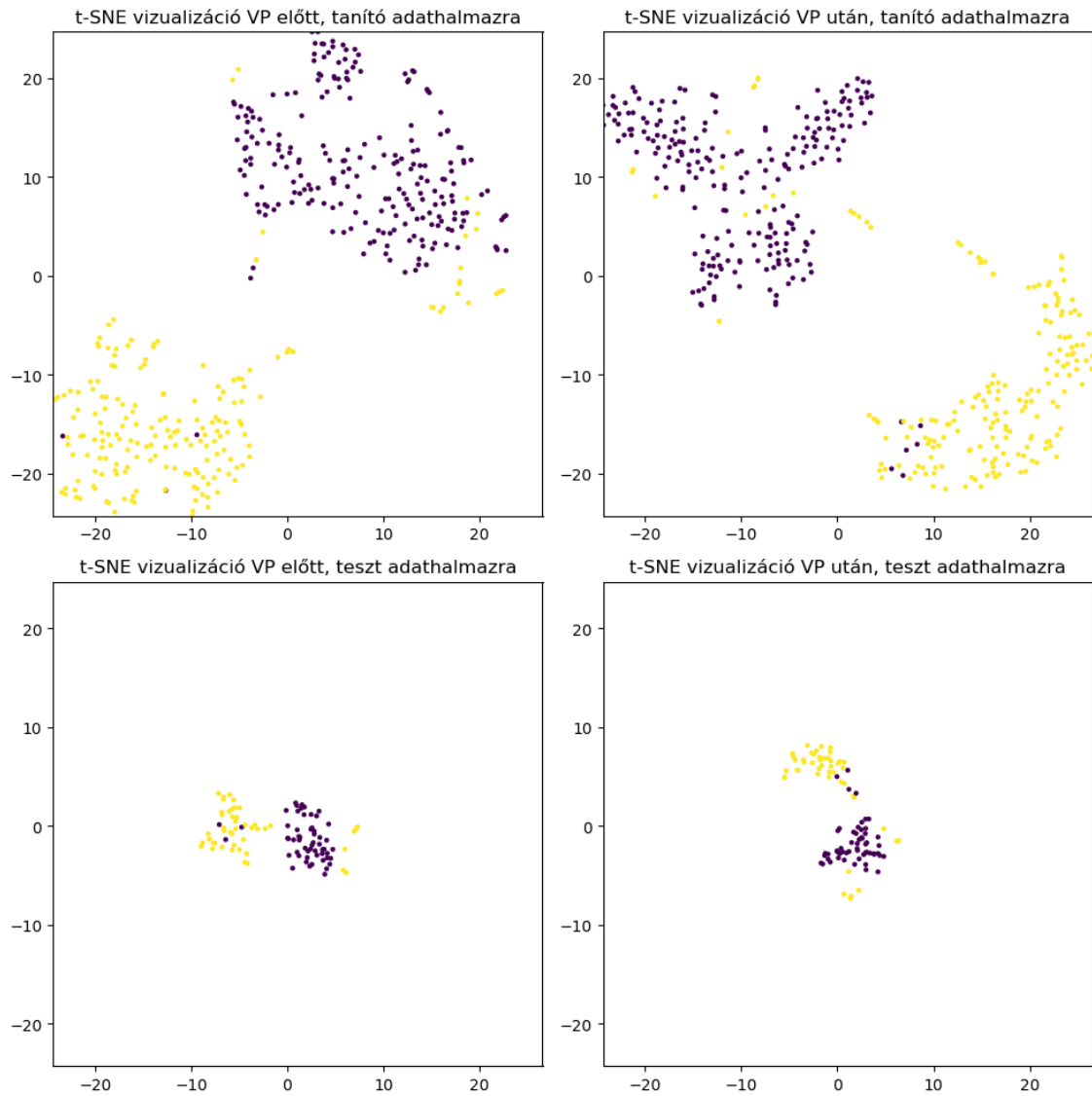


4.10. ábra. EEG jelek PCA-val 2D-ra képezve.

4.3.3. Kerékszenzoros adathalmaz

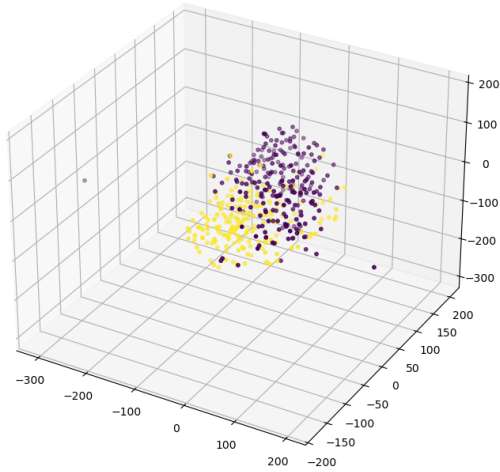
t-SNE vizualizációk

Az előző adathalmazokkal ellentétben, itt rögtön látszik, hogy jelentősen kevesebb adatponttal rendelkezünk. Így nehezebb a szeparáltságra általánosságban levonni biztos következtetéseket. Hasonlóan a korábbiakhoz, jól látszik, hogy a nyers adatok is szeparáltak. A korábbiakhoz hasonlóan továbbá az is jól megfigyelhető, hogy az adatok lineárisan is elkülönülnek egymástól. A VP réteg alkalmazása után a lineáris szeparáltság megmarad. A perplexity értéket itt is 30-ra állítottam.

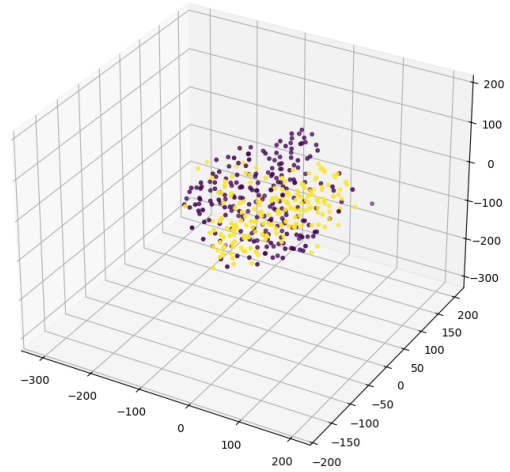


4.11. ábra. Kerékszenzoros jelek t-SNE-vel 2D-re képezve.

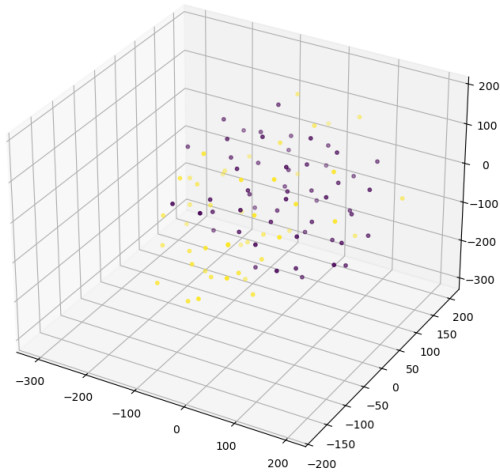
t-SNE vizualizáció VP előtt, tanító adathalmazra



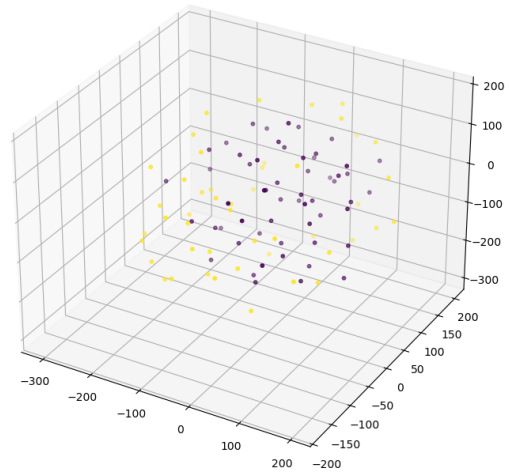
t-SNE vizualizáció VP után, tanító adathalmazra



t-SNE vizualizáció VP előtt, teszt adathalmazra



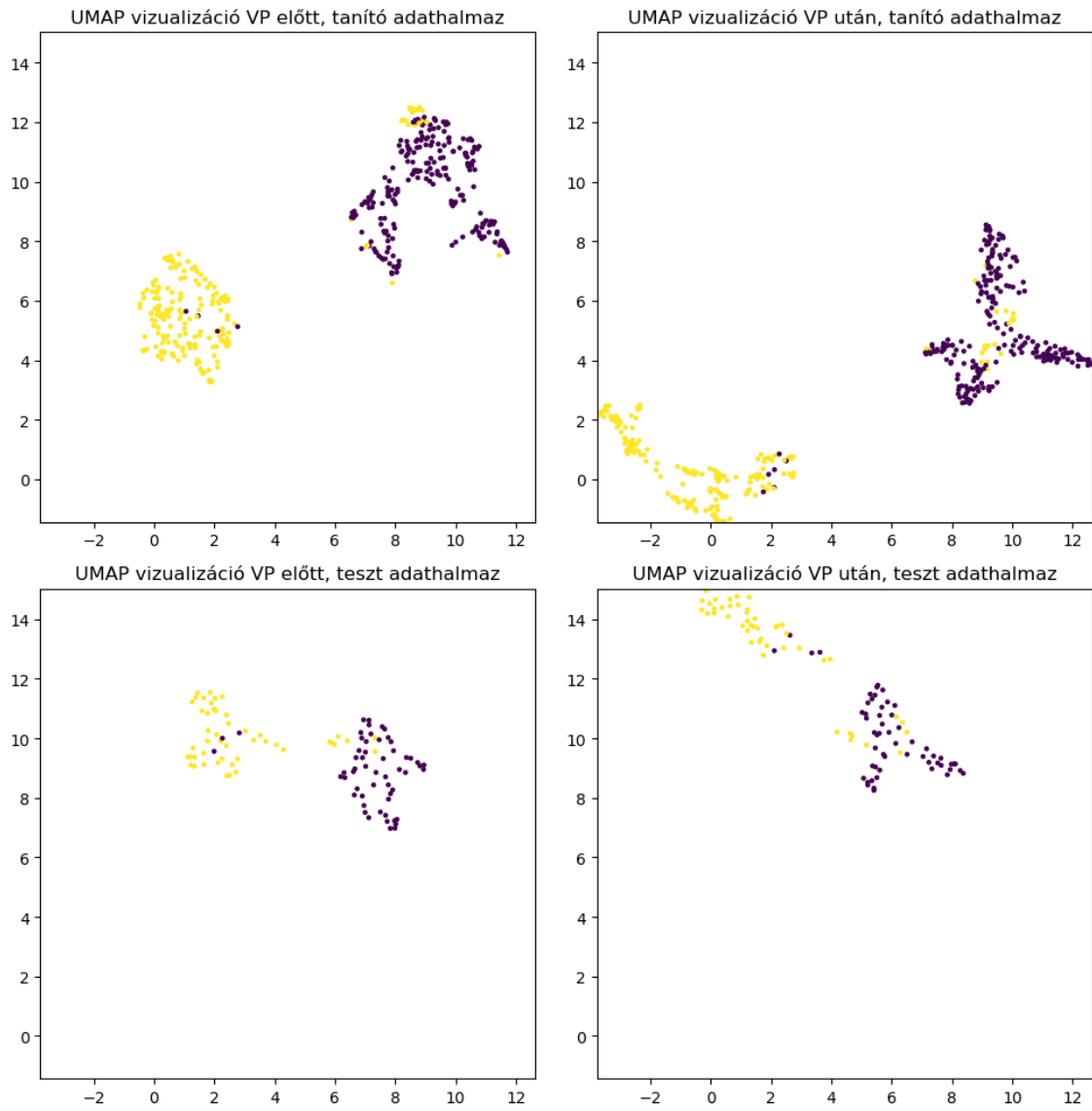
t-SNE vizualizáció VP után, teszt adathalmazra



4.12. ábra. Kerékszenzoros jelek t-SNE-vel 3D-re képezve.

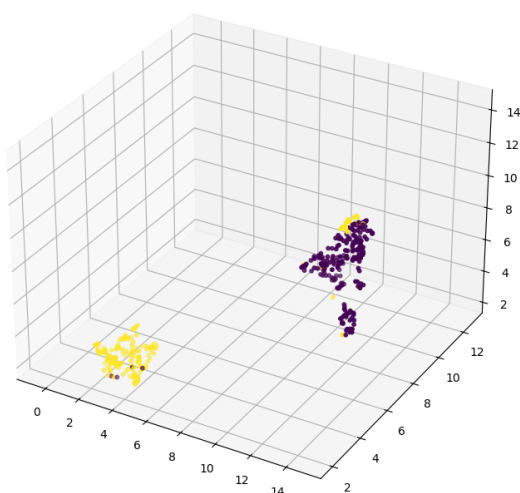
UMAP vizualizációk

Az UMAP vizualizáció alapján, hasonlóan a t-SNE-hez, az adatok a VP réteg előtt és után is jól szeparálódnak. Az osztályok közötti elkülönülés itt még kifejezettebbnek tűnik, sőt vizuálisan látszik, hogy az osztályok akár lineárisan is szeparálhatók. Ugyanakkor az alacsony adatpontoszám miatt az a néhány pont, amely a valós osztályától eltérő csoportban helyezkedik el, aránytalanul nagy hatást gyakorolhat a szeparálhatóság metrikák szerinti számszerűsítésére. Ez a néhány pont torzíthatja az értékelést, mivel súlyukhoz képest aránytalanul nagy befolyást gyakorolhatnak az eredményekre.

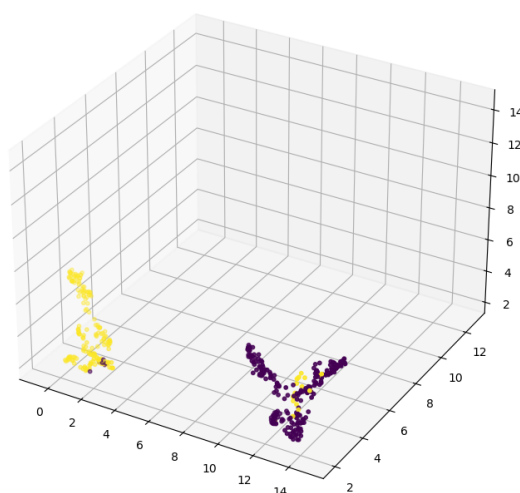


4.13. ábra. Kerékszenzoros jelek UMAP-al 2D-re képezve.

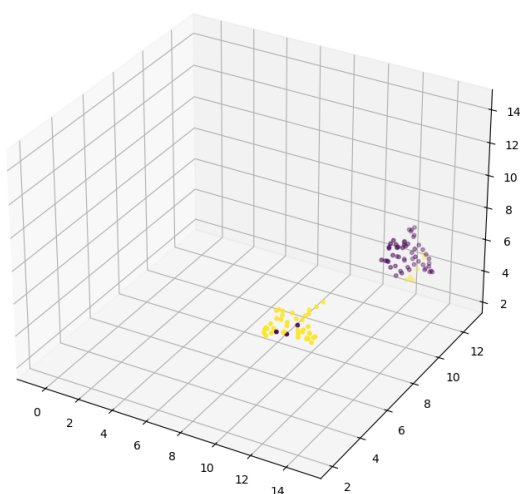
UMAP vizualizáció VP előtt, tanító adathalmazra



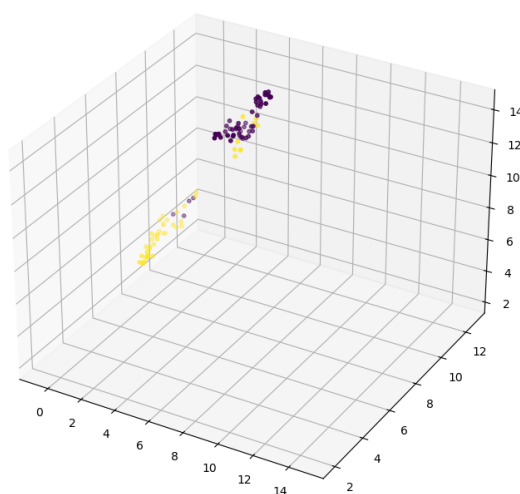
UMAP vizualizáció VP után, tanító adathalmazra



UMAP vizualizáció VP előtt, teszt adathalmazra



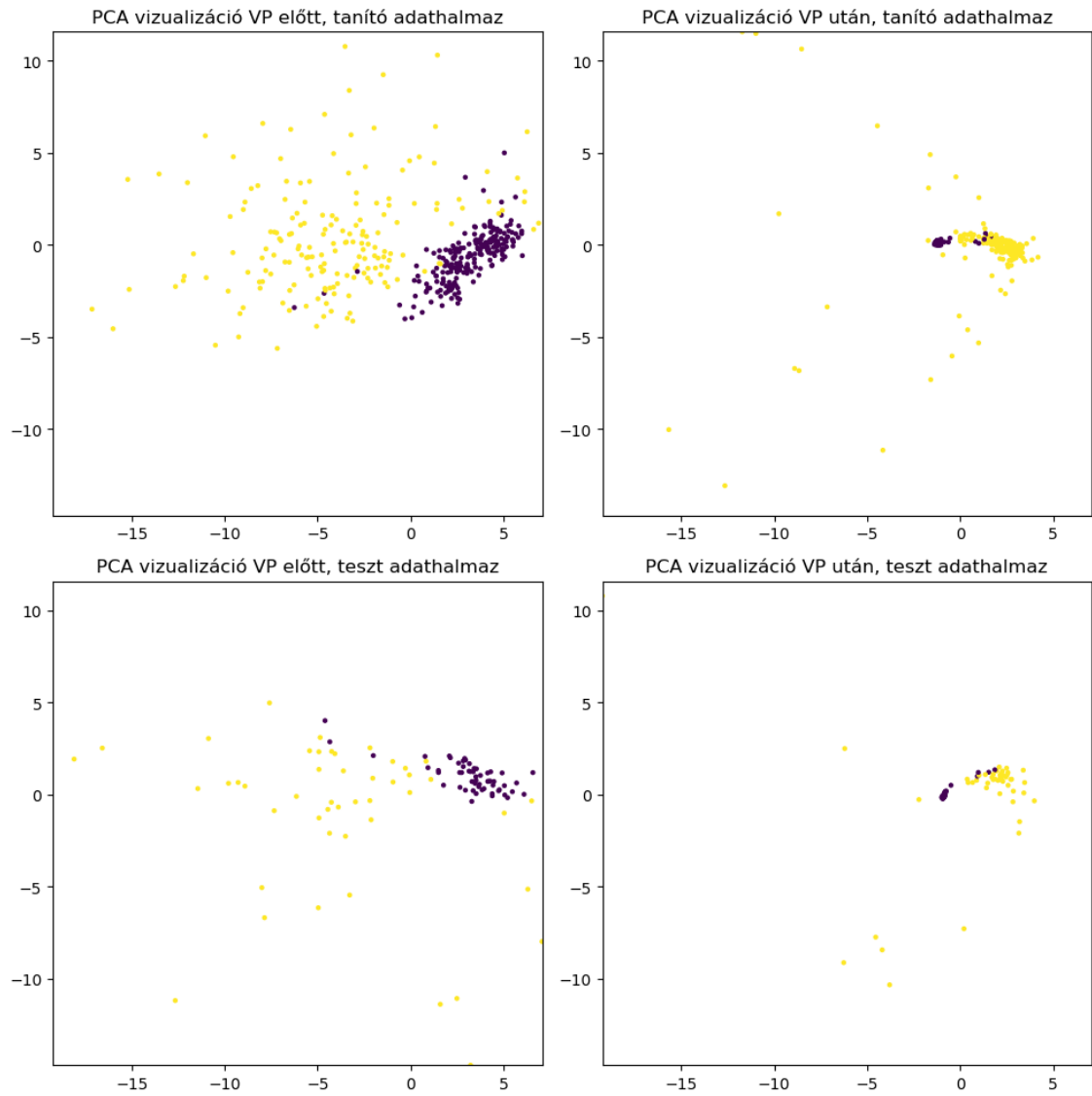
UMAP vizualizáció VP után, teszt adathalmazra



4.14. ábra. Kerékszenzoros jelek UMAP-al 3D-re képezve.

PCA vizualizációk

Hasonlóan a korábbi típusú adatokhoz, itt, a kerékszenzoros jelek esetében is megfigyelhető, hogy a PCA által leképezett adatokon az osztályok csak részben különülnek el egymástól. A PCA, mint lineáris dimenziócsökkentő módszer, nem képes megragadni a jelek komplex, nemlineáris szerkezetét.



4.15. ábra. Kerékszenzoros jelek PCA-val 2D-ra képezve.

5. fejezet

Adatok szeparálhatósága

5.1. Szeparálhatósági vizsgálatok jelentősége

Az adatok szeparálhatósága az osztályozási problémák egyik alapvető jellemzője, amely meghatározza, hogy az adatok különböző osztályai mennyire könnyen elkülöníthetők egymástól. A jól elkülönülő osztályok lehetővé teszik a gépi tanulási algoritmusok, például neurális hálózatok vagy más osztályozók számára, hogy pontosabb előrejelzéseket adjanak, mivel az osztályok közötti különbségek egyértelműen azonosíthatók.

A szeparálhatósági vizsgálatok célja kvantitatív módon elemezni az osztályok elkülöníthetőségét, ami azt jelenti, hogy számszerűsíthetően mérjük, az adatok különböző osztályai mennyire jól elkülöníthetők egymástól a térben, például távolsági metrikák vagy csoportosulási mutatók segítségével. Ez különösen fontos a modellek teljesítményének optimalizálása és a háló által jellemzők relevanciájának értékelése szempontjából.

A dolgozatban a VP réteg hatását vizsgálom az adatok szeparálhatóságára. A VPNet által elért pontosabb osztályozási teljesítmény alapján az a hipotézis, hogy a VP réteg növeli az adatok szeparálhatóságát.

5.2. Lineáris és Nemlineáris Szeparálhatóság Vizsgálata

Az adatok szeparálhatóságának vizsgálatánál két alapvető megközelítést különböztetünk meg: a lineáris és a nemlineáris szeparálhatóságot. Ezek a vizsgálatok segítenek megérteni, hogy az osztályok közötti különbségek egyszerű lineáris határokkal leírhatók-e, vagy összetettebb, görbült határookra van szükség. A kvalitatív vizsgálattal szemben itt a

nyers adatokat vizsgáljuk dimenziócsökkentés nélkül, és csak a megértést segítő ábrákhoz használunk t-SNE dimenziócsökkentést.

5.2.1. Lineáris szeparálhatóság

A lineáris szeparálhatóság azt jelenti, hogy az adatok különböző osztályai egyetlen egyenes (vagy magasabb dimenzióban egy hipersík) segítségével elkülöníthetők egymástól. Az ilyen osztályozási feladatok egyszerűbb modellekkel is megoldhatók, például lineáris *Support Vector Machines* (SVM) vagy *k*-legközelebbi szomszéd (*K-Nearest Neighbors*, *KNN*) algoritmus alkalmazásával. Az adatok lineáris szeparálhatóságát leggyorsabban úgy határozhatjuk meg, hogy a fent említett modelleket betanítjuk a tanuló (train) adathalmazon, majd az így létrehozott modell teljesítményét teszteljük az ellenőrző (test) adathalmazon.

SVM:

A *Support Vector Machines* (SVM) [5] egy felügyelt tanulási algoritmus, amely célja, hogy megtalálja azt a lineáris határvonalat (hipersíkot), amely a legjobban elkülöníti az adatok osztályait. A modell a különböző osztályok közötti távolságot maximalizálja, ezáltal biztosítva a döntési határ stabilitását.

SVM működése:

- Az algoritmus megkeresi azt a hipersíkot, amely a legnagyobb távolságot (margin) biztosítja az osztályok között.
- Azokat az adatpontokat, amelyek legközelebb esnek a szeparáló hipersíkhhoz, *tartó vektoroknak* (support vectors) nevezzük. Ezek határozzák meg a hipersík pontos helyzetét.

SVM alkalmazása: Az SVM segítségével megállapítottam, hogy az adathalmazok lineárisan szeparálhatók-e. Egy modellt betanítottam a nyers adatokon dimenziócsökkentés után és egy másikat a VP utáni adatokon. Az így betanított modelleket megvizsgáltam a tesztadatokon, ha a pontosság növekedett, akkor a VP réteg pozitív hatással van az adatok lineáris szeparáltságára.

KNN:

A k -legközelebbi szomszéd (*K-Nearest Neighbors*, *KNN*) [4] algoritmust korábban már használtam a VPNet megmagyarázására, de hatékony módszert nyújt a szeparálhatóság mérésére is. A lineáris szeparálhatóság esetén a KNN általában pontos eredményt ad, mivel az osztályok közötti határ egyszerűen meghatározható.

KNN működése:

- Az algoritmus megméri az új adatpont távolságát a tanuló adathalmaz minden egyes pontjától (euklideszi távolság segítségével).
- Az adatpont osztályát a legközelebbi k szomszéd többségi osztálya határozza meg.

KNN alkalmazása: Az SVM-hez hasonlóan itt is betanítottam egy modellt a nyers adatokra és egy másik modellt a VP utáni adatokra. Amennyiben a második modell pontosabban osztályoz, az azt jelenti, hogy növekedett az adatok lineáris szeparáltsága.

5.2.2. Nemlineáris szeparálhatóság

A nemlineáris szeparálhatóság azt jelenti, hogy az osztályok nem különíthetők el egy egyszerű lineáris döntési határral, hanem bonyolultabb, nemlineáris határookra van szükség. Az ilyen típusú szeparálhatóságot általában nemlineáris modellek, például neurális hálózatok, kernel SVM-ek, vagy nemlineáris dimenziócsökkentési technikák, például t-SNE vagy UMAP segítségével vizsgálják.

A nemlineáris szeparálhatóság elemzéséhez a Silhouette-metrikát és a Davies-Bouldin indexet használtam:

Silhouette-metrika:

A Silhouette érték [15] ($S(i)$) megmutatja, hogy egy adott adatpont mennyire illeszkedik a saját klaszterébe, és mennyire van távol a többi klasztertől. Mivel a vizsgált adatok mindössze két osztályba sorolhatók, a legközelebbi klasztert nem kell külön kiválasztanunk, így az egyetlen másik osztály automatikusan adott. A távolságokat euklideszi távolságként számítjuk, amely a geometriai távolságot méri az adatpontok között. A metrika értéke 1 és -1 közötti tartományban kerül meghatározásra minden egyes adatra, majd ezt követően az átlagot veszem így megkapva egy számszerűsíthető értéket az adathalmaz szeparáltságára.

Számítás alapjai:

- $A(i)$: Az adatpont (i) átlagos távolsága az ugyanahhoz a klaszterhez tartozó többi adatponttól: minél kisebb, annál erősebben tartozik saját klaszteréhez/osztályához.
- $B(i)$: Az adatpont (i) átlagos távolsága a másik klaszter adatpontjaitól: minél nagyobb, annál erősebben tartozik saját klaszteréhez/osztályához.

Silhouette érték képlete:

$$S(i) = \frac{B(i) - A(i)}{\max(A(i), B(i))}$$

Az értéke -1 és 1 közötti érték, ahol:

- 1: Az adatpont jól illeszkedik a saját klaszterébe, és jól el van választva a többi klasztertől.
- 0: Az adatpont egyformán közel van a saját és a másik klaszterhez.
- -1: Az adatpont inkább a másik klaszterhez tartozna, nem a sajátjához

Az átlagos Silhouette érték alapján az adathalmaz általános szeparálhatósága értékelhető: minél közelebb van az 1-hez, annál jobb a klaszterezés.

Davies-Bouldin index:

A Davies-Bouldin index (DBI) [16] egy másik mérőszám a klaszterek minőségének értékelésére, amely az osztályok közötti euklideszi távolságot és az osztályokon belüli kohéziót (osztályon belüli euklideszi távolságot) méri. A DBI értékének csökkenése a klaszterek jobb szeparáltságát jelzi. A klaszterek közötti távolságot az osztályok középpontjainak távolsága határozza meg.

Davies-Bouldin index képlete:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d_{ij}} \right)$$

ahol:

- σ_i : Az i -edik klaszter kohéziója (átlagos távolság a klaszter középpontjától).
- d_{ij} : Az i és j klaszterek középpontjainak távolsága.
- k : A klaszterek száma.

Minél nagyobb az érték annál kevésbé szeparáltak az adatok általánosságban.

5.3. Eredmények

5.3.1. EEG adathalmaz

Metrika	VP előtt	VP után
Osztályozási Pontosság (%)	92.23	92.8
KNN Pontosság (%)	83.5	84.8
SVM Pontosság (%)	80.5	87.4
Silhouette-metrika	0.136	0.212
Davies-Bouldin index	2.347	1.582

5.1. táblázat. VP réteg hatása az osztályozási teljesítményre és az EEG adatok szeparálhatóságára.

A táblázat első sorában az adathalmazon mért pontossága olvasható a VPNet-nek egy egyszerű FCNN-hez képest. Abból adódóan, hogy a VP réteggel együtt tanított és alkalmazott rendszer jobb eredményt ér el az adathalmazon, arra gondolhatunk, hogy a VP réteg javíthatja az adatok szeparáltságát.

A nyers adatokon betanított SVM modell kisebb pontosságú, mint a VP réteg utáni adatokkal betanított. Ugyan ez megfigyelhető a KNN modelleknél. Ezek alapján arra következtethetünk, hogy a VP réteg javítja az adatok lineáris szeparáltságát.

A KNN modellek szomszédszámának növelése nyers adatok esetében csak minimális pontosságnövekedést eredményezett (csúcs: 83.5% 35 szomszédnál), míg a VP réteg utáni adatoknál a pontosság stabilan 85% körül maradt. Ez azt mutatja, hogy a nyers adatok globálisan jobban szeparálhatók, de a VP réteg után az adatok lokálisan is jobban elkülönülnek.

A Davies-Bouldin index (DBI) csökkenése és a Silhouette-metrika növekedése azt jelzi, hogy a VP réteg javítja a nemlineáris szeparáltságot. A DBI alacsonyabb értéke azt mutatja, hogy a klaszterek távolságai nőttek, míg a magasabb Silhouette-érték jobb belső kohéziót és elkülönülést jelez. Bár a Silhouette érték még nem éri el a tökéletes szeparáltságot (1), az eredmények alapján a VP réteg jelentősen javította az adatstruktúrát.

Az eredmények alapján a VP réteg növeli mind az adatok lineáris, mind a nemlineáris szeparáltságát az EEG adathalmaz esetében, ezáltal javítva az osztályozási teljesítményt.

5.3.2. EKG adathalmaz

Metrika	VP előtt	VP után
Osztályozási Pontosság (%)	97.01	99.96
KNN Pontosság (%)	94.5	93.8
SVM Pontosság (%)	90.6	93.5
Silhouette-metrika	0.273	0.291
Davies-Bouldin index	2.026	1.884

5.2. táblázat. VP réteg hatása az osztályozási teljesítményére és az EKG adatok szeparálhatóságára.

A táblázat első sorában az EKG adathalmazon a VPNet-tel mért pontosság olvasható egy egyszerű FCNN-hez képest. Mivel az általános pontosságon javított a VP réteg alkalmazása és az együttes betanítás, így ennél az adathalmaznál is sejthetjük, hogy a VP réteg pozitív hatással lehet az adatok szeparálhatóságára.

Lineáris szeparálhatóság szempontjából azt láthatjuk, hogy az SVM modell esetében a VP réteg alkalmazása után pontosabb modell tanítható, míg a KNN modellnél minimális pontosságcsökkenés figyelhető meg. Ez a vegyes eredmény arra utal, hogy a VP réteg nem növeli egyértelműen az adatok lineáris szeparáltságát ezen az adathalmazon.

Ezzel szemben a nemlineáris szeparálhatóságnál azt látjuk, hogy a Silhouette-metrika növekedik, ez azt jelzi, hogy a klaszterek szeparáltsága javult a VP réteg alkalmazása után. Emellett a Davies-Bouldin index (DBI) csökken, ez azt mutatja, hogy a klaszterek közötti távolságok növekedtek, és a klaszterek belső homogenitása javult. Ezek alapján kijelenthetjük, hogy a VP réteg javítja az EKG adatok nemlineáris szeparálhatóságát.

Összességében azt látjuk, hogy míg a VP réteg hatása a lineáris szeparálhatóságra vegyes eredményeket mutatott, míg a nemlineáris szeparálhatóság javulását egyértelműen eredményezte az EKG adathalmazon.

5.3.3. Kerékszenzoros adathalmaz

Metrika	VP előtt	VP után
Osztályozási Pontosság (%)	98.04	99.02
KNN Pontosság (%)	94.2	93.2
SVM Pontosság (%)	98.1	88.3
Silhouette-metrika	0.304	0.248
Davies-Bouldin index	1.55	2.982

5.3. táblázat. VP réteg hatása az osztályozási teljesítményére és a kerékszenzoros adatok szeparálhatóságára.

Lineáris szeparálhatóság szempontjából azt láthatjuk, hogy a KNN esetében a VP réteg alkalmazása után kevésbé pontosabb modell tanítható, hasonló eredményt kapunk az SVM modell esetében is. Ezek azt jelzik, hogy a VP réteg nem javította a lineáris szeparálhatóságot, sőt, bizonyos mértékben rontotta azt a kerékszenzoros adatok esetében.

A nemlineáris szeparálhatóság vizsgálatakor a Silhouette értékek és a Davies-Bouldin Index alapján szintén romlást tapasztaltam a VP réteg alkalmazása után. Mivel a Silhouette érték maximális értéke 1, ami a tökéletes szeparáltságot jelentené, a csökkenő értékek azt mutatják, hogy az adatok szeparáltsága romlott, de még így is inkább szeparáltak, mint nem. A magasabb Davies-Bouldin Index rosszabb klaszterezést jelez, ami arra utal, hogy a VP réteg alkalmazása után a klaszterek közötti elkülönülés csökkent, és a klasztereken belüli adatok homogenitása romlott.

Összességében a Kerékszenzoros adatok esetében a VP réteg alkalmazása nem javította, hanem rontotta az adatok szeparálhatóságát mind lineáris, mind nemlineáris szempontból. Ennek egyik lehetséges oka, hogy a kerékszenzoros adathalmaz mérete jelentősen kisebb a másik két vizsgált adathalmaznál (EKG és EEG). A kisebb minta miatt a VP réteg ezen hatása nem mutatkozott meg egyértelműen.

6. fejezet

Összegzés és Kitekintés

A dolgozat során számos technikát alkalmaztunk a VPNet megmagyarázására és az adatok szétválaszthatóságára. Az esetalapú érvelés (case-base-reasoning CBR) segítségével a hálózat kimeneti döntései könnyebben értelmezhetővé válnak és a hibák felismerése is egyszerűbbé vált. A dimenziócsökkentési technikákból, mint például a t-SNE, UMAP és PCA, készült vizualizációk alapján kvalitatív szempontból első betekintést kaptunk az adatok lineáris és nemlineáris szeparáltságába. Ezek a vizualizációk segítettek megérteni az adatok belső szerkezetét és azt, hogy a különböző osztályok mennyire különülnek el egymástól a térben. Az SVM és K-NN modellekkel meghatároztuk a lineáris szeparáltság mértékét mindhárom adathalmaz esetében. Megállapítottuk, hogy bár az adatok jól elkülönülnek lineárisan, csak a kerékszenzoros adathalmaz esetén ér el olyan magas értéket, amely alapján kijelenthetjük, hogy az adatok lineárisan szeparáltak. Az EKG adathalmazon kimutattuk, hogy a VP réteg alkalmazása növeli az adatok lineáris és nemlineáris szeparáltságát. A VP réteg előtti és utáni eredmények összehasonlítása révén láthatóvá vált, hogy a VP réteg hatékonyan javítja az osztályok közötti elkülönülést. A metrikákon keresztül, mint a Davies-Bouldin Index és a Silhouette érték, kvantitatív szempontból vizsgáltuk az adatokat, és ezek az értékek alátámasztották a vizualizációk alapján levont következtetéseket.

Összegzés

Összefoglalva a dolgozatban több fontos eddig nem ismert megállapítást sikerült tenünk a VPNet-re vonatkozóan:

- Az elvégzett elemzésekkel sikerült azonosítani azokat a VP komponenseket, melyek vélhetően nem járulnak hozzá a modell kimenetéhez. Így ezek törlésével a

modell redukálható.

- A CBR vizsgálatok rámutattak, hogy ha egy bemeneti jelhez hasonló példák osztályozása is hibás, akkor nagy valószínűséggel a bemeneti jelet is helytelenül sorolta be a modell. Ez a tulajdonság éles környezetben segíthet eldönteni egy megbízunk-e a háló egy adott kimenetében vagy sem. Ez különösen biztonságkritikus alkalmazások esetén fontos, ahol egy-egy hibás döntésnek komoly következményei is lehetnek.
- A szeparáltság kvalitatív vizsgálatánál a dimenztócsökkentő módszerek segítségével készült vizualizációk segítenek megérteni az adatok belső szerkezetét.
- A szeparálhatóság kvantitatív vizsgálata során megállapítottuk, hogy a VP réteg alkalmazása növeli az adatok lineáris és nemlineáris szeparáltságát.
- A kvantitatív vizsgálat során továbbá sikerült azonosítani, hogy nyers adatok globális szinten jobban szeparálhatók, míg a VP réteg alkalmazása után az adatok nemcsak globálisan, hanem lokálisan is jobban elkülönülnek egymástól.

Kitekintés

A jövőben több irányban is érdemes lenne folytatni a kutatást:

- **A VP réteg adatszeperációs hatásának összehasonlítása más technikákkal:** Jelen dolgozat csak a VP réteg szeparáló hatását vizsgálja kvalitatív és kvantitatív elemző módszerekkel. Ezt érdemes lenne összevetni más reprezentáció tanuló algoritmusokkal (pl. autoenkóder). Ez lehetővé tenné annak megállapítását, hogy a VP réteg mennyire hatékony az adatok szeparáltságának növelésében más módszerekhez képest.
- **A nem használt Hermite-függvények eltávolítása:** Az esetalapú magyarázatnál a 3.2 ábrán megfigyelt nem releváns Hermite komponensek elhagyásával tovább csökkenthető a VP réget által megtanult tulajdonságvektorok dimenziója. Megfontolandó, hogy ez a módosítás javítja-e a modell teljesítményét és hatékonyságát.
- **A VPNet vizsgálata más adathalmazokon:** A VPNet csak egy keretrendszer, melynek számos eleme cserélhető, ilyen például az alkalmazott függvényrendszer

is. Érdeemes lenne más, nem feltétlenül tüskeszerű jelalakokhoz igazított függvényrendszerek vizsgálata új adathalmazokon, ami hozzájárulna a modell általánosítóképességének vizsgálatához.

Ezek a jövőbeli irányok hozzájárulhatnak a VPNet továbbfejlesztéséhez és a módszer alkalmazhatóságának kiterjesztéséhez. Az alternatív technikákkal való összehasonlítás mélyebb megértést nyújthat a VP réteg működéséről és előnyeiről. A Hermite-függvények optimalizálása pedig a modell egyszerűsítését és hatékonyabbá tételét eredményezheti. Új adathalmazok bevonása pedig lehetővé teszi a VPNet képességeinek szélesebb körű tesztelését és az eredmények általánosíthatóságának növelését.

Irodalomjegyzék

- [1] Zachary Chase Lipton. „The Mythos of Model Interpretability”. *CoRR* abs/1606.03490 (2016). arXiv: 1606.03490. URL: <http://arxiv.org/abs/1606.03490>.
- [2] Péter Kovács és tsai. „VPNet: Variable projection networks”. *International Journal of Neural Systems* 32.01 (2022), 2150054. old.
- [3] Tamás Dózsa és tsai. „Weighted Hermite Variable Projection Networks for Classifying Visually Evoked Potentials”. *IEEE Transactions on Neural Networks and Learning Systems* (2024).
- [4] Gongde Guo és tsai. „KNN model-based approach in classification”. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*. Springer. 2003, 986–996. old.
- [5] MO Stitson és tsai. „Theory of support vector machines”. *University of London* 117.827 (1996), 188–191. old.
- [6] Tamás Dózsa és tsai. „Road abnormality detection using piezoresistive force sensors and adaptive signal models”. *IEEE Transactions on Instrumentation and Measurement* 71 (2022), 1–11. old.
- [7] Julian McAuley és Jure Leskovec. „Hidden factors and hidden topics: understanding rating dimensions with review text”. *Proceedings of the 7th ACM conference on Recommender systems*. 2013, 165–172. old.
- [8] Alexander Mordvintsev, Christopher Olah és Mike Tyka. „Inceptionism: Going deeper into neural networks”. *Google research blog* 20.14 (2015), 5. old.

- [9] Eoin M Kenny és Mark T Keane. „Twin-systems to explain artificial neural networks using case-based reasoning: Comparative tests of feature-weighting methods in ANN-CBR twins for XAI”. *Kraus, S.(ed.). Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. 2019.
- [10] Laurens Van Der Maaten, Eric O Postma, H Jaap Van Den Herik és tsai. „Dimensionality reduction: A comparative review”. *Journal of machine learning research* 10.66-71 (2009), 13. old.
- [11] Laurens van der Maaten és Geoffrey Hinton. „Visualizing Data using t-SNE”. *Journal of Machine Learning Research* 9 (2008), 2579–2605. old. URL: <http://www.jmlr.org/papers/v9/vandermaaten08a.html>.
- [12] Solomon Kullback és Richard A Leibler. „On information and sufficiency”. *The annals of mathematical statistics* 22.1 (1951), 79–86. old.
- [13] Leland McInnes, John Healy és James Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. 2020. arXiv: 1802.03426 [stat.ML]. URL: <https://arxiv.org/abs/1802.03426>.
- [14] Andrzej Maćkiewicz és Waldemar Ratajczak. „Principal components analysis (PCA)”. *Computers & Geosciences* 19.3 (1993), 303–342. old.
- [15] Peter J Rousseeuw. „Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. *Journal of computational and applied mathematics* 20 (1987), 53–65. old.
- [16] David L Davies és Donald W Bouldin. „A cluster separation measure”. *IEEE transactions on pattern analysis and machine intelligence* 2 (1979), 224–227. old.