

Óbudai Egyetem
Kandó Kálmán Villamosmérnöki Kar
Elektrofizika Intézet, Természettudományi Tanszék



TUDOMÁNYOS DIÁKKÖRI DOLGOZAT

Titán-dioxid és cink-oxid alapú festékérzékenyített napelem villamos paramétereinek előrejelzése statisztikai és gépi tanulási technikák segítségével

Szerző: **Varga Zoltán**
Alkalmazott Matematika MSc. szak,
I. évfolyam

Konzulens: **Dr. habil. Rácz Ervin PhD**
egyetemi docens

Budapest, 2024.

Tartalom

1	Bevezetés.....	2
2	Szakmai bevezetés.....	4
2.1	Festékérzékenyített napelem cella.....	5
2.2	DSSC hatásfokát meghatározó és befolyásoló fő tényezők kísérleti úton való vizsgálata.....	8
2.2.1	Pilot kutatásaim eredményei	9
2.2.2	Pilot kutatásaim eredményei elválasztása más tudományos munkáktól	10
2.3	Festékérzékenyített napelemekkel kapcsolatos predikciós modellek	11
3	Alkalmazott kutatás módszertan	15
3.1	Kutatásmódszertan	15
3.2	Tevékenység során elvárt eredmények	15
3.3	A tudományos munka gazdasági és társadalmi haszna.....	16
4	Elemzési módszerek és szimulációs eljárások	17
4.1	Kísérletek megtervezése és elvégzése.....	17
4.2	Nyers adatpontok elemzése	18
4.3	Alkalmazott statisztikai eljárás.....	19
4.4	Gépi tanulási modellek.....	22
4.5	Munka folyamatábrája.....	27
5	Eredmények.....	31
5.1	Statisztikai vizsgálat eredményei	31
5.1.1	Leíró statisztika	31
5.1.2	Homogenitás vizsgálat	33
5.1.3	ANOVA	33
5.2	Gépi tanulási algoritmusok eredményei.....	34
5.2.1	Hatásfok	36
5.2.2	Rövidzárási áramsűrűség.....	37

5.2.3	Üresjárási feszültség.....	38
5.2.4	Kitöltési tényező.....	39
5.3	Tudományos munkám elválasztása más tudományos eredményektől	40
6	Kitekintés	41
7	Összefoglaló.....	42
8	Köszönetnyilvánítás	45
9	Irodalomjegyzék.....	46
10	Ábrajegyzék	52
11	Táblázatjegyzék.....	54

Absztrakt

A megújuló energiaforrások iránti növekvő globális igény napjainkban különös hangsúlyt helyez az energiahatékony és környezetbarát villamosenergia-termelési módszerek és technológiák fejlesztésére és innovációjára. Ebben a kontextusban a festékszennyezett napelemek (angolul: Dye-Sensitized Solar Cell, rövidítve: DSSC) jelentős potenciállal bírnak, mivel alacsony gyártási költségeikkel és egyedülálló felépítésükkel egy vonzó alternatívát kínálhatnak a hagyományos szilícium alapú technológiákkal szemben.

Tekintettel arra, hogy nem csak az elektromágneses hőhatás befolyásolja a napelemek hatásfokát, hanem maga a napelem gyártási folyamata is, ideértve a megmunkálási hőmérsékletet és annak idejét. A megfelelő hőmérsékletkezelés kulcsfontosságú a DSSC napelemek optimális teljesítményének és tartósságának biztosításához. Mindezek mellett, a napelem kutatás az utóbbi években a mesterséges intelligencia felé mozdult el. Egyre több rangos folyóiratban (pl.: Nature, Solar Energy, MDPI) jelennek meg olyan tudományos publikációk, amelyek a gépi tanulási algoritmusok (rövidítve ML) alkalmazásán alapszanak. A rangos folyóiratokban megjelent publikációk rámutatnak az ML alkalmazásának fontosságára, mert a sok kísérleti próbálkozás helyett, a módszerek alkalmazásával fontos következtetések vonhatók le a cella készítése előtt.

A kutatásom során elsődlegesen a kísérleti adatok alapján statisztikai elemzést végzek a hőmérséklet-változás és a villamos paraméterek közötti összefüggések tisztázására. A kutatás második szakaszában átfogó irodalomkutatást hajtok végre, melynek célja, hogy a releváns eredményeket szisztematikusan összegyűjtsem. Ezen adatgyűjtési folyamat eredményeképpen adatbázist fejleszték. Ezt követően, a létrehozott adatbázist felhasználva, gépi tanulási modellt implementálok, hogy predikciót végezzek a DSSC villamos paramétereinek meghatározására.

Kutatási munkámmal jelentős előrelépést kívánok elérni a DSSC napelem fejlesztésében a projekt innovatív és multidiszciplináris megközelítése révén. A kutatási eredményeim nem csupán a tudományos közösség számára nyújt új ismereteket, hanem a valós piaci igényeket is kielégíti a DSSC cella gyártástechnológiáját tekintve.

1 Bevezetés

A megújuló energiaforrások iránti növekvő globális igény napjainkban különös hangsúlyt helyez az energiahatékony és környezetbarát villamosenergia-termelési módszerek és technológiák fejlesztésére és innovációjára. Ebben a kontextusban a festékszennyezett napelemek (angolul: Dye-Sensitized Solar Cell, rövidítve: DSSC) jelentős potenciállal bírnak, mivel gyártási költségeikkel és egyedülálló felépítésükkel egy vonzó alternatívát kínálhatnak a hagyományos szilícium alapú technológiákkal szemben. A festékszennyezett napelem cella tudományterületen számos publikáció jelent meg, melyek rámutatnak, hogy a szilícium alapú napelemek dominanciáját a DSSC napelemek fogják forradalmasítani [1].

Korábbi tudományos munkáimban a DSSC villamos paramétereinek változását vizsgáltam az elektromágneses hőhatás függvényében és eredményeimet nemzetközi folyóirat publikációkban [2], [3] és konferencia publikációkban közöltem [4], [5]. Tekintettel arra, hogy nem csak az elektromágneses hőhatás befolyásolja a napelemek villamos paramétereit, hanem maga a napelem gyártási folyamata is, ideértve a szintézis hőmérsékletet és annak idejét. A megfelelő hőmérsékletkezelés kulcsfontosságú a DSSC napelemek optimális villamos paramétereinek és tartósságának biztosításához. A témában rejlő sok tudományok kérdésfeltevés még megválaszoltaltam, melyet a jelen kutatás keretein belül megvizsgáltok és kifejtetek. Mindezek mellett, a napelem kutatás az utóbbi években a mesterséges intelligencia felé mozdult el. Egyre több rangos folyóiratban (pl.: Nature, Solar Energy, MDPI) jelennek meg olyan tudományos publikációk, amelyek a gépi tanulási algoritmusok (rövidítve ML) alkalmazásán alapszanak [6], [7]. A rangos folyóiratokban megjelent publikációk rámutatnak az ML alkalmazásának fontosságára, mert a sok kísérleti próbálkozás helyett, a módszerek alkalmazásával fontos következtetések vonhatók le a cella készítése előtt [8].

Ezúton céloomul tűztem ki, hogy kísérleteket tervezek meg, kísérleteket végzek el a festékszennyezett napelem cella villamos paramétereire és az elektromágneses sugárzás interakcióját vizsgálva, adatokat gyűjtök be, dolgozok fel és értékelek, valamint az adatokon statisztikai tesztek futtatok. További céloom, hogy adatbázist fejlesszek, melyben a DSSC gyártástechnológiája során felmerülő releváns eredményeket szisztematikusan összegyűjtöm. Az adatgyűjtést követően a felépített adatbázis alapján gépi tanulási algoritmus segítségével predikciós modellt alakítok ki a DSSC villamos paramétereinek pontos előrejelzésére.

Kutatási munkámmal jelentős előrelépést kívánok elérni a festékszennyezett napelem fejlesztésében a projekt innovatív és multidiszciplináris megközelítése révén. A kutatási

eredményeim nem csupán a tudományos közösség számára nyújt új ismereteket, hanem a valós piaci igényeket is kielégíti a festékérzékenyített napelem cella gyártástechnológiáját tekintve.

Személyes célom az, hozzájáruljak és elősegítsem a fenntartható és megújuló elektromosenergia-termelést, valamint csökkentsem annak környezetre gyakorolt hatását. Kutatásommal törekszem, hogy nemzetközileg központi szerepet töltsön be a megújuló energiával kapcsolatos kutatásokban és innovációban. Hosszú távú célom, hogy hozzájáruljak a tiszta és fenntartható jövőhöz, harcolva a klímaváltozás ellen.

2 Szakmai bevezetés

A villamosenergetikai iparágban a kutatók és a mérnökök folyamatosan keresik azokat az innovatív megoldásokat, melyek csökkentik a fosszilis energiahordozók égetését, ezáltal minimalizálva a villamosenergia termelés ökológiai lábnyomát [9], [10]. A fosszilis energiahordozók égetése során a légkörben lévő szén-dioxid kibocsátás megnövekszik, amely klimatikai problémákat okoz úgy, mint a szélsőséges időjárások kialakulása és a földfelszín átlagos hőmérsékletének emelkedése [11]. A félvezetőipar ugrásszerű növekedése lehetővé tette, hogy a megújuló energiaforrásokat átalakító eszközök, különös tekintettel a napelemre, előtérbe kerüljön a piaci alkalmazásokban [12]. Az egyik legígéretesebb fejlesztési irány a napenergia, melynek hasznosítása hőenergia és villamosenergia formájában lehetséges [13]. A Napból a Föld felszínre érkező elektromágneses sugárzást közvetlenül átalakító eszköz a napelem (rövidítve: PV), mely technológia jelentős mértékben hozzájárul a környezeti terhelés csökkentéséhez.

A napelemek számos típusát különböztetjük meg az alapanyag szerinti kategorizálás alapján. Ebben a rendszerben az elsődleges klasztert a kristályos szilícium alapú napelemek alkotják, melyeken belül két domináns altípust különböztetünk meg: a monokristályos és a polikristályos szilícium alapú napelemeket, amelyek a szilícium kristályrácsának rendezettségében mutatkozó különbségeik alapján különböztethetők meg egymástól. Ezen csoportra jellemző, hogy a hétköznapi felhasználás tekintetében az egyik legelterjedtebbek, valamint hosszú az élettartamuk, amely 20-25 évre is becsülhető [14]. A második nagy csoport a vékonyréteg napelemek, melyek közül az amorf szilícium alapú napelem a legelterjedtebb a vékonyréteg csoporton belül az egyes piaci elemzések alapján [13]. Mindezek mellett, a harmadik csoportba tartoznak a speciális alkalmazási területen, ideértve az űrkutatást, előforduló napelem típusok úgy, mint a gallium-arszenid és más vegyület félvezető napelemek, melyek gyártási költsége magasabb a korábbi típusokhoz képest, de ezzel szemben magasabb hatásfokkal rendelkeznek [15], [16]. Az utolsó megemlítendő csoport a kutatási-fejlesztési fázisban lévő napelemek, melyek közé sorolható a kutatásom középpontjában álló festékérzékenyített napelem cella is. Ezen csoportban lévő innovatív napelemek számos potenciállal bírnak a széles alkalmazási terület és gazdaságosság szempontjából [17].

A festékérzékenyített napelem cellák előnyei, hogy kiemelkedő teljesítményt nyújtanak alacsony fényviszonyok mellett és képesek diffúz megvilágítás esetén is hatékonyan működni. Mindezek mellett, a gyártásuk egyszerűbb és kevésbé energiaigényes, ami a gyártási költségek

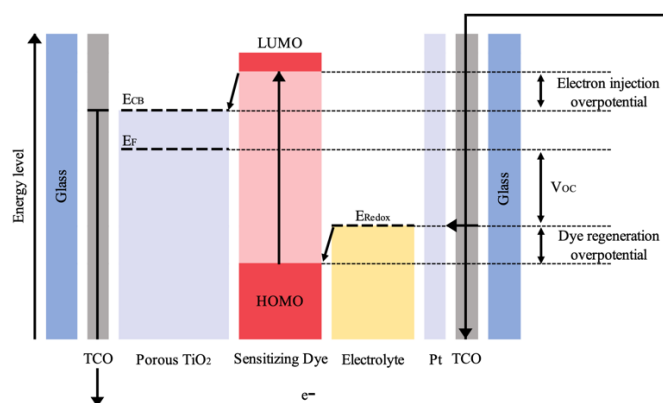
csökkenését eredményezi. Az áttanulmányozott irodalmak alapján a festékérzékenyített napelemek világpiaci szerepe csekély, de egyes prognosztizációk alapján a szilícium alapú napelemek dominanciáját a jövőben meg fogja törni [18].

2.1 Festékérzékenyített napelem cella

A festékérzékenyített napelem cellával kapcsolatos első publikációt M. Greatzel és O' Brian közzétették 1991-ben, melyben 7,1%-os hatásfokról számoltak be [18]. A munkájuk azóta is referencia pont a festékérzékenyített napelem kutatásban. A több, mint 30 éves távlatban a 7,1%-os hatásfokot megdöntötték és az áttanulmányozott irodalmak alapján az eddig elért világ rekord 14%, melyet Greatzel mutatott be a Hybrid and Organic Photovoltaic konferencián [19]. Habár a publikációkból megállapítható, hogy az időtényező változásának függvényében a DSSC hatásfokának gradiense jelentéktelen mértékű, de még sok kiaknázatlan kérdést foglal magába ezen tudományterület [20].

A hagyományos festékérzékenyített napelem cella öt fő alkotó elemet tartalmaz, amely (i) a üveghordozó réteg átlátszó vezetőképes oxiddal bevonva (angolul: Transparent Conductive Oxid, amely rövidítve TCO); (ii) vezetőképes üveghordozóra felvitt általában n-típusú félvezetőrétteg, leggyakrabban titándioxid (TiO_2) vagy cink-oxid (ZnO); (iii) félvezetőréttegre abszorbeált fényérzékeny, ruthenium-alapú festékanyag (angolul: Sensitizing Dye); (iv) elektrolit oldalt, jellemzően jodid-trijodid; (v) és az utolsó réteg az ellen elektróda, mely leggyakrabban platina [21], [22]. A festékérzékenyített napelem cella működésének energiadiagramját, valamint elektrontranszfer folyamatát az 1. ábra szemlélteti, ahol bemutatásra kerül az átlátszó vezető oxid (TCO), a pórúsos félvezető réteg, fényérzékeny festékanyag, elektrolit és platina réteg. Az energia szintek és azok relatív helyzetét úgy, mint LUMO, HOMO szintek, Fermi-szint és üresjárási feszültség az ábra szemlélteti. A beeső elektromágneses sugárzás hatására a festékérzékenyített napelem cella fényérzékeny molekuláinak belső energiája megnő, ami a HOMO (Highest Occupied Molecular Orbital) szinten lévő elektronok LUMO (Lowest Unoccupied Molecular Orbital) szintre való gerjesztését jelenti [23]. A gerjesztést követően a megnőtt belső energiájú elektronok a félvezető réteg vezetési sávjába injektálódnak, majd diffúzió hatására kilépnek az áramkörbe, ezáltal villamos áramot generálva. A leadott elektronokat követően a festékanyagban elektronhiány alakul ki, más néven elektronhiányos állapotba kerül, melyet elektrolit oldószerben lévő redox párok redukálnak és az elvesztett elektronokat pótolják. Az elektrolit

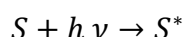
oldószer pedig a platina ellenelektrodán keresztül visszaállítja a redox párt [24]. A cella működéséből jól látszik, hogy technológiai oldalról jelentős különbség van a festékszennyezett napelem és a piacot domináló monokristályos szilícium alapú napelem között, hiszen a hagyományos p-n átmenetes félvezető napelem cella esetén a félvezető nyeli el a fotonokat és választja szét a töltéshordozókat, addig a festékszennyezett napelem cellánál ezen két funkció külön-külön jelenik meg. Pontosabban, a belépő fotonokat a fényérzékeny festékanyag nyeli el és a töltésszétválasztódás a fényérzékeny festékanyag és a félvezető réteg határán történik meg [25], [26].



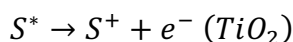
1. ábra. Energiadiagram a festékszennyezett napelem cella (DSSC) működésének bemutatására az elektrontranszfer folyamatot is szemléltetve. A sötét kékkel jelölt rész az üveghordozót jelenti (glass felirat az ábrán); a szürke rész az átlátszó vezetőképes oxid réteget szemlélteti (Transparent Conductive Oxid, TCO); a világos kék félvezető réteg, bal oldalon jellemzően titán-dioxid (jelölve Porous TiO₂), míg az ábra jobb oldalán pedig platina (Pt); piros színnel a fényérzékeny festékanyag látható (ábrán jelölve, mint Sensitizing Dye), ahol a HOMO: Highest Occupied Molecular Orbital, azaz legmagasabb molekula orbitál, és a LUMO: Lowest Unoccupied Molecular Orbital, azaz legalacsonyabb nem foglalt molekula orbitál; és a világossárga rész pedig az elektrolit oldal. (Az ábra saját konferencia publikációból került másolásra) [5].

A korábban részletesen leírt jelenség az alábbi folyamatokkal írható le, ezzel hozzásegítve az 1. ábra könnyebb megértését:

1. A beérkező elektromágneses sugárzást a fényérzékeny festékanyag elnyeri és a kiindulási állapotból (S) gerjesztett állapotba (S^*) kerül:

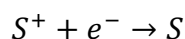
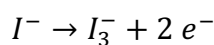


2. A gerjesztett elektronok a félvezető réteg (TiO₂) vezetési sávjába (E_{CB}) kerülnek

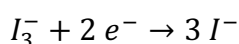


3. A félvezető réteg vezetési sávjából az elektronok ezt követően az elektródán át az áramkörbe jutnak.

4. A fényérzékeny festékanyag elvesztett elektronjai a jodid/trijodid oldatból visszanyeri az alábbiak szerint:



5. Végül a trijodid a katódhoz diffundál és redukálódva visszalakul jodiddá:



Számos kutató adaptálta Ferbernek a DSSC-re készített elektromos modelljét, mely integrálja a töltéstranszportot a határrétegen keresztül [27], [28]. Megállapításra került, hogy a diffúziós folyamat uralja a töltéstranszportot a félvezető rétegben és az anyagi jellemzők úgy, mint a részecske átmérője, porozitása jelentős mértékben befolyásolja az elektronok diffúziós képességét, együttthatóját [1]. Mindezek mellett, az áttanulmányozott irodalmakból kiderül, hogy a leggyakrabban alkalmazott félvezető rétegek a titán-dioxid (TiO₂) és a cink-oxid (ZnO) [22]. A két anyagról elmondható, hogy tiltott-sávszélességük megegyezik (3,2 eV) [29]. Érdeemes megjegyezni, hogy a fentebb említett félvezető rétegek mellett más, hasonló tiltott-sávszélességű anyagok használatot is tesztelték úgy, mint kéndioxid (SnO₂) [30] és nióbium-pentoxid (Nb₂O₅) [31]. Az anyagok mellett a különböző struktúrák kialakításával is igyekeztek a DSSC villamos paramétereit javítani. Ilyen típusú struktúrák a nano csövek [32], nano rudak [33], nano vezetékek [34], nano erdők [35], nano gombák [36]. A publikációk rámutatnak, hogy a hidrotermális szintézis az egyik legmeghatározóbb módszer a félvezető réteg kialakítása során, mely eljárás magas nyomáson és hőmérsékleten történő kristály növekedést eredményez. A szintézis hőmérséklete és időtartama ugyancsak kulcsfontosságú paraméterek, amelyek befolyásolják a félvezető réteg morfológiáját és a kristály növekedésének dinamikáját, ezáltal meghatározva a festékanyag adszorpciók képességét [37].

A festékkérszenyített napelem cella másik alkotóeleme a fényérzékeny festékréteg, mellyel szemben az alábbi elvárásokat támasztották a kutatók: (i) megfelelő stabilitás működés közben és degradáció elleni védelem; (ii) széles fényabszorpciók képesség, a látható tartomány minél nagyobb hasznosítása [22]. A fényérzékeny festékanyagokat két típusra oszthatjuk fel, melyek a fémes komplex festékkérszenyítő és a fémentes organikus festékkérszenyítő, attól függően, hogy tartalmaznak-e fémes elemeket. A fémes komplex fényérzékeny festékanyagok főleg ruténium poliporodol komplexben, fém-porfinokban, ftalocianinekben koncentrálnak [38]. A fémentes organikus fényérzékeny festékanyagok közé tartoznak a természetes festékek és

a szerves szintetikus festékek [23]. A leggyakrabban alkalmazott fényérzékeny festékréteg a ruténium-komplexek, melyek közül a legismertebb az N3 és N719 festékek [39]. Az említett anyagok kiváló fényabszorpciós tulajdonságokkal rendelkeznek a látható és a közeli infravörös tartományban [40]. Hosseinnezhad és kutatótársai áttekintést adnak a különböző természetes festékanyagokról. Munkájukban több, mint húsz különböző természetes festék kivonat összehasonlítását vizsgálták, melyek olyan növények és virágok kivonataiból nyertek, melyek őshonosak Iránban, például vöröshagyma, narancs és dióhéj kivonat, gránátalma. Munkájuk kiderül, hogy a gránátalma kivonatból készült festékérzékenyített napelem cella hatásfoka a legmagasabb [41].

A harmadik nagyobb alkotóelem az elektrolit oldal, mely biztosítja az elektromos áramkör lezárását annak érdekében, hogy regenerálja a festékanyagot és transzportálja a festékanyag elvesztett elektronját. Az elektrolitnak hatékony töltésszállítással kell rendelkezzen egy nem korrozív környezetben, anélkül, hogy jelentős fényelnyelést okozna [42]. Az eredeti festékérzékenyített napelem cellában alkalmazott jodid/trijodid oldat továbbra is népszerű választásnak számít a kedvező töltéstranszfer jellemző miatt [43]. Azonban ennek a redox rendszernek az egyik fő hátránya a két-elektronnal járó redox folyamat, amely nagy túlfeszültséget eredményez [44]. Ez a túlfeszültség, alacsonyabb üresjárási feszültség értéket eredményez, mivel az üresjárási feszültség az elektronok Fermi-szintjének különbségével van meghatározva a megvilágított félvezető réteg és a redox potenciálban [45]. A festékérzékenyített napelem cella felépítésének és működésének rövid bemutatásából jól látható, hogy a kutatók jelentős innovációs eredményeket publikáltak és kerültek a tudományos kutatások fókuszába. A következő szakaszban bemutatom a DSSC hatásfokát meghatározó és befolyásoló fő tényezőket annak érdekében, hogy későbbi fejezetben a tudományos kérdésfeltevéseimet meg tudjam alapozni.

2.2 DSSC hatásfokát meghatározó és befolyásoló fő tényezők kísérleti úton való vizsgálata

A napelem hatásfoka definíció szerint számítható az alábbi képlet segítségével:

$$\eta = \frac{U_0 \cdot I_0 \cdot FF}{P_{in}} \quad (1)$$

ahol az U_0 a napelem üresjárási feszültsége, az I_0 a napelem rövidzárási árama, az FF a napelem kitöltési tényezője és a P_{in} a napelemet ért fénysugárzás teljesítménye. A fenti képletben

látszódik, hogy az intenzitás jellemzőn kívül a többi paraméter a napelemre jellemző mennyiség, ezek mindegyike hőmérsékletfüggő. A hatásfok esetleges növelése, tulajdonképpen a fenti definiáló egyenlőség jobb oldalán álló tört számlálójának alapos vizsgálatát, minél részletesebb megértését követeli meg. A kitöltési tényező az alábbi egyenlőséggel definiáljuk:

$$FF = \frac{U_{MPP} \cdot I_{MPP}}{U_0 \cdot I_0} \quad (2)$$

ahol U_{MPP} a napelem villamos teljesítmény munkapontjának feszültség koordinátája és az I_{MPP} a munkapontjának áram koordinátája. Az U_{MPP} és I_{MPP} szorzata a maximális teljesítményt vagy más szóval a munkaponti teljesítményt (MPP = Maximum Power Point) fejezi ki. A (2) egyenletet behelyettesítve az (1) egyenletbe:

$$\eta = \frac{U_{MPP} \cdot I_{MPP}}{P_{in}} = \frac{MPP}{P_{in}} \quad (3)$$

Tehát a (3) egyenlet szerint a napelem hatásfokának vizsgálata a munkapont feszültség és áram koordinátájának vizsgálatával ekvivalens. (A nevező egy emberi léttől független, természetből kapott érték.) A munkapont vizsgálata a napelem villamos áram-feszültség jelleggörbéjének tanulmányozását és annak felvételét követelik meg. Mindazonáltal, a fentieket a hőmérséklet-függvényében vizsgálom meg.

2.2.1 Pilot kutatásaim eredményei

Pilot kutatás keretein belül megvizsgáltam a festékérzékenyített napelem cella villamos paramétereit úgy, mint a fenti (1) képletben definiált $U_0, I_0, U_{MPP}, I_{MPP}, MPP$ a napelem cella hőmérséklet-függvényében. Munkám kezdeti szakaszában a trend jellegű változást szerettem volna kimutatni a villamos paraméterekben a hőmérséklet-függvényében. Ennek érdekében a fentebb említett villamos paraméterekre lineáris egyenest illesztettem a hőmérséklet-függvényében [46], [47], [48], [49]. Munkám második szakaszában a hőmérséklet-változás mellett bevezettem még egy fix paramétert, a beeső fény foton beütésszám numerikus integrálját [50]. A pilot kutatásom harmadik szakaszában pedig korrelációs vizsgálatot végeztem, hogy megvizsgáltam, van-e szignifikáns kapcsolat a hőmérséklet (fix paraméter) és a céltárgyként használt DSSC I_0 és U_0 értékei között. Az U_0 esetén azt az eredményt kaptam, hogy a null hipotézist elvetem és az ellenhipotézis lép életbe, vagyis van szignifikáns kapcsolat a hőmérséklet és az üresjárási feszültség között. Az I_0 esetén elvégzett hipotézis vizsgálatra azt kaptam, hogy a null hipotézis megmarad. Vagyis nincs szignifikáns kapcsolat a hőmérséklet és a rövidzárási áram között. Itt fontosnak tartom megjegyezni, hogy a pilot miatt kis mérési

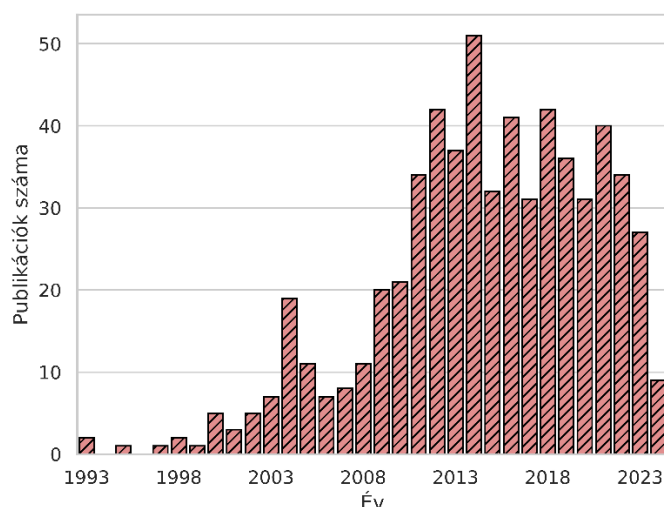
mintával dolgoztam. A Kendall-korreláció alapján az $U_0(T)$ -re közepes erősséget ($\tau_B = -0,528$) kaptam a vizsgált hőmérsékleti tartományon belül [51]. A pilot kutatás alapján elmondható, hogy további vizsgálatok elvégzése szükséges a felmerült kérdésfeltevések megválaszolására.

2.2.2 Pilot kutatásaim eredményei elválasztása más tudományos munkáktól

Chen-Ching Ting és Wei-Shi Chao munkájukban a festékérzékenyített napelem cella villamos áram-feszültség karakterisztikáját, maximális teljesítményét, üresjárási feszültségét és rövidzárási áramát vizsgálták $5\text{ }^{\circ}\text{C}$ és $80\text{ }^{\circ}\text{C}$ hőmérsékleti tartományban $10\text{ }^{\circ}\text{C}$ -os intervallumonként. Az eredményekből látszódik, hogy a cella hőmérséklet növekedés hatására az áram-feszültség karakterisztika feszültség értékei, üresjárási feszültség és rövidzárási áram, maximális teljesítmény értékek csökkennek. Valamint megállapítható, hogy a teljesítmény és az üresjárási feszültség esetén ez a változás közel lineáris [52]. Laboratóriumi körülmények között vizsgálva a napelemet, más kutatók is hasonló eredményekhez jutottak [53], [54]. Usami és munkatársai által megállapításra került, hogy az üresjárási feszültség hőmérsékletfüggését az elektronok élettartama és a rekombinációs áram mértéke befolyásolja, amely a félvezető réteg vezetési sáv szélének potenciáljában következik be. Azáltal, hogy a hőmérséklet növekedés hatására az elektronok élettartama csökken, a rekombinációs áram növekszik, ami magával vonja az üresjárási feszültség csökkenését [54]. A tudományos adatbázisokban a laboratóriumi körülmények között elvégzett kísérletek mellett fellelhetők a kültéri vizsgálatok is. Érdemes megemlíteni Lepikko és munkatársai munkáját, akik hat hetes kültéri vizsgálatokat végeztek északi klimatikai viszonyok mellett, mely csapadékos, fagypont közeli hőmérsékletet és magas páratartamot foglalt magába. Eredményül azt kapták, hogy a magas páratartalom esetén intenzív degradáció lépett fel a cellák esetén [55]. Kültéri vizsgálatot más kutatók is végeztek más éghajlati viszonyok mellett [56], [57]. A pilot kutatás során kapott eredményeim összhangban állnak a szakirodalomban fellelhető adatokkal. Ez az egyezés megerősíti az alkalmazott módszertan helyességét és segíti a jelenségek könnyebb interpretálását. Más szavakkal, az összehasonlítás során megállapításra került, hogy a kapott eredmények konzisztensek az áttanulmányozott irodalmakban megtalálható tudományos megfigyelésekkel.

2.3 Festékképzékenyített napelemekkel kapcsolatos predikciós modellek

A SCOPUS adatbázis keresőmotorja által a festékképzékenyített napelem cella és statisztikai módszerek és azzal analóg keresési szavakra számos publikáció érhető el, melyek időbeli eloszlását szemlélteti a következő, 2. ábra. Az ábráról jól látható, hogy a közel 30 éves időszakban a statisztikai jellegű publikációk nagy mennyiségű sokaságát 2010 és 2020 között publikálták. Az elmúlt időben a statisztikai jellegű tudományos munkák mennyisége enyhén csökkenő tendenciát, negatív gradiens változást mutat.



2. ábra. Az 1993 és 2024 közötti időszakban regisztrált publikációk számának év szerinti lebontásban való oszlopdiagram szerű ábrázolása, különös tekintettel a statisztikai módszertan alkalmazásán alapuló és az ahhoz terminológiailag kapcsolódó kapcsolatos munkák.

Minden esetre, az említett publikációk klasszifikálhatók a szerzők által megadott kulcsszavak (keyword), a rövid tartalmi kivonat (absztrakt) és cím alapján. Az említett időszakban megjelent, statisztikai módszertan alkalmazásán alapuló és ahhoz terminológiailag kapcsolódó publikációkat táblázatba foglaltam össze. Az irodalom áttanulmányozása során a magas citációval (hivatkozással) rendelkező munkákat vettem figyelembe. Ennek alapján készítettem el az 1. táblázatot, mely két külön kategóriába sorolja a tudományos publikációkat: a klasszifikáció egyik kategóriája a materiális, anyagi vizsgálatok helyezi előtérbe; míg a másik kategóriába kifejezetten a DSSC villamos paramétereinek elemzésére specializálódott publikációkat tartalmazza. Továbbá, a publikációk még további két másik faktor szerint is klasszifikálásra kerültek, mely két faktor a statisztikai módszerrel kapcsolatos: korreláció, regressziós; valamint varianciaanalízis (ANOVA). Dürr és munkatársai azt vizsgálták, hogy a cirkónium koncentrációjának növelése a titán-dioxidhoz hogyan befolyásolja a titándioxid sánszélességét és milyen hatást eredményez a DSSC teljesítményére. Vizsgálatukból

megállapításra került, hogy az említett anyag koncentrációjának növelésével lineárisan növekszik a vegyes oxidok atomi szerkezete, ami az üresjárási feszültség növekedésével jár együtt [58]. Saito és munkatársai korrelációs vizsgálatokkal megállapították, hogy a durvasági faktor, mely a félvezető rétegre jellemző paraméter és az elektromos áramsűrűség között pozitív összefüggés van. Mégpedig, a durvasági faktor növekedésével az áramsűrűség is növekszik [59]. Kebede és munkatársai az előzőkhöz hasonlóan korrelációs és regressziós vizsgálatokat alkalmaztak az I_2 (jodid) abszorpció csúcsinak helyváltozása és az oldószerek donor száma között [60]. Ghosh és kutatócsoportja arra tudományos megfigyelésre jutott, hogy szignifikáns korrelációs figyelhető meg a színhőmérséklet és a színmegjelenítési index között azoknál a festékérzékenyített napelem celláknál, melyek különböző vastagságú félvezető rétegeket tartalmaznak [61]. Al-Alwani és munkatársai varianciaanalízis (ANOVA) segítségével meghatározták az Areca catechu gyümölcséből kivont természetes festék kedvező paramétereit és megállapították, hogy a kinyerési hőmérséklet, a pH érték és a kivonószerként alkalmazott etanol jelentős mértékben befolyásolják a DSSC hatásfokát [62]. Természetes fényérzékeny festékanyag felhasználásával készült DSSC gyártási folyamat releváns tényezőinek feltárására és optimalizálására ugyancsak varianciaanalízist használt Leyrer és csapata. Megállapították, hogy az antocianin koncentrációja a hatásfokra 43,9%-os hatást, az impregnációs idő 4,55%-ot, míg a festéktípusa pedig 51,22%-ot gyakorolt [63]. Továbbá, Abdollahi és munkatársainak a kutatási célja az volt, hogy meghatározzák azokat a paramétereket, melyek mellett a legkisebb átmérőjű, homogén nanoszálakat elő lehet állítani a festékérzékenyített napelem cellához, melyhez ANOVA vizsgálatot végeztek [64]. Al-Alwani és munkatársai további vizsgálatokat végeztek varianciaanalízis segítségével, melyet a cordyline fruticosa leveleiből származó klorofill kivonásának optimalizálására alkalmaztak; és megállapították a szerves oldószer forráspontja, a pH értéke és a kivonás hőmérséklete esetén az optimális értékeket [65]. Patni és Pillai varianciaanalízis alkalmazásával megvizsgálták a vanádium-pentoxid százalékos arányának és a polimer kompozit vezetőképességének változását a hatásfokra nézve [66]. Zainudin és munkatársai a titán-dioxid nanorészecskék előállítási folyamatával kapcsolatban végeztek ANOVA vizsgálatot, ahol három független változó – a víz moláris aránya, az öregedési hőmérséklet és a kalcinálási hőmérséklet – hatását vizsgálták a teljesítményre nézve és szignifikáns kapcsolatot mutattak ki a három paraméter és teljesítmény között [67]. Mindezek mellett, Lepikko és munkatársai korrelációs és regressziós vizsgálatokat végeztek a villamos paraméterek vizsgálatára az időjárási paraméterek függvényében [55]. A táblázatból

megállapítható az áttanulmányozott irodalmak alapján, hogy a varianciaanalízis statisztikai módszert az elektromos paraméterek vizsgálatára a szakirodalomban nem alkalmazták.

1. táblázat: A feldolgozott tudományos publikációk kategorizálása adatértékelési módszerek alapján úgy, mint a korreláció, regresszió és varianciaanalízis.

	Korreláció, regresszió	Varianciaanalízis
Anyag	[58], [59], [60], [61]	[62], [63], [64], [65], [66], [67]
Elektromos paraméter	[55]	

Mindezek mellett, a napelem kutatás az utóbbi években a mesterséges intelligencia felé mozdult el. Egyre több rangos folyóiratban jelennek meg olyan tudományos publikációk, amelyek a neurális hálózatok, döntési fák és további gépi tanulási algoritmusok alkalmazásán alapszanak. Hisham A. Maddah predikciós modellt készített a természetes fényérzékeny festékanyag szerkezeti, elektronikai és molekuláris tulajdonságainak vizsgálatára, ahol a bemeneti rétegen a festékszerkezet π -kötéseinek száma, a festékek rögzítőcsoportjának száma, HOMO-LUMO és a sáv szélességi energia volt. Munkájában két predikciós modellt alkalmazott és hasonlított össze, melyek a döntési fát (Decision Tree, DT) és Support Vector Machine (SVM). Az összehasonlítás eredményéből azt kapta, hogy a döntési fák jobban teljesítenek, mint a SVM modell, ami a 20%-al jobb pontosságban mutatkozik meg [8]. A festékanyag jellemzőit gépi tanulási algoritmusok felhasználásával több kutató és kutatócsoport is tanulmányozta. Wen és munkatársai az organikus fényérzékeny festékanyag predikciós vizsgálatára fejlesztettek modellt [7]. Sutar és munkatársai a cink-oxid alapú DSSC hatásfokának vizsgálatát végezték, amely a gyártástechnológiai oldalról nézve tölt be fontos szerepet. Bemeneti réteggént a félvezető réteg struktúráját, a prekursor, a fényérzékeny festékanyag típusát, a szintézis hőmérsékletét és a szintézis idejét választották; melyekből a kimeneti rétegen lévő DSSC hatásfokát jelezték elő. Továbbá, munkájukban kiemelik, hogy a hidrotermálisan előállított félvezető réteg kialakítása a leghatékonyabb módszer a további technikákkal szemben. Mindezek mellett, predikciós modellként a döntési fát, neurális hálózatot, véletlen erdőt (Random Forest, RF) alkalmaztak. A predikciós modellek közül a legnagyobb pontosságot a DT mutatta [68]. Xu és munkatársai neurális hálózatok segítségével tanulmányozták az organikus fényérzékeny festékanyag abszorpciós maximumát [69]. Ezen publikációk rámutatnak arra, hogy ezen módszerek nem csak elméleti jelentőséggel bírnak, hanem gyakorlati alkalmazásuk is lehetővé teszi az energiahatékony fejlesztést, hiszen a sok kísérleti

próbálkozás helyett, a módszerek alkalmazásával fontos következtetések vonhatunk le a cella készítése előtt, ezáltal időt és energiát spórolhatunk meg.

3 Alkalmazott kutatás módszertan

Az alkalmazott kutatás módszertani részben bemutatom célkitűzésem, a kutatómódszertant, valamint tudományos munkám gazdasági és társadalmi hasznosulásának lehetőségeit.

Az áttanulmányozott irodalmak alapján célomlul tűztem ki, hogy kísérleti adataimat feldolgozom és a cella teljesítményét a hőmérséklet függvényében statisztikai eljárás alkalmazásával értékelem. További céloml még, hogy átfogó irodalomkutatást hajtok végre, melynek célja, hogy a releváns eredményeket szisztematikusan összegyűjtsem. Ezen adatgyűjtési folyamat eredményeképpen adatbázist fejleszték, majd a létrehozott adatbázist felhasználva gépi tanulási modellt implementálok, hogy predikciót végezzek a festékérzékenyített napelem cella villamos paramétereinek meghatározására.

Tudományos kérdésfeltevések:

- (1) A kísérleti adathalmazon végzett statisztikai vizsgálatokból milyen következtetések vonhatók le a festékérzékenyített napelem cella villamos teljesítményére nézve?
- (2) Milyen pontosságú modellt lehet interpretálni a festékérzékenyített napelem cellára gyártástechnológiáját figyelembe véve?

3.1 Kutatómódszertan

Az említett fő célkitűzés megvalósításához első körben adatgyűjtési és adatfeldolgozási folyamatot végzek, majd korrelációs stratégiával alkalmazok a begyűjtött adataimon. Munkám második lépéseként a felépített adatbázison modellelmet tanítom és tesztelem. A logikai következtetési formák esetén pedig deduktív módszert alkalmazok, mert az egyes esetekben az előre meghatározott premisszák alapján határozom meg a logikai szabályokat.

3.2 Tevékenység során elvárt eredmények

A statisztikai vizsgálatok eredményeként azt várom, hogy a nullhipotézist (H_0) elvetem és az ellenhipotézist (H_1) tartom meg, vagyis szignifikáns kapcsolat kimutatását várom a festékérzékenyített napelem cella villamos teljesítménye és a hőmérséklet között. Továbbá, hogy a vizsgált hőmérsékleti tartomány esetén pár Celsius fokos hőmérséklet-különbség esetén van szignifikáns változás a teljesítmény és hatásfok értékekben. Mindezek mellett, a

modellemtől azt várom, hogy a teszt adatokra a predikciós modell közel 60%-os választ ad. Továbbá, eredményeimet összehasonlítom és elválasztom más kutatók által bemutatott eredményektől.

Tudományos munkám első részének megvalósítása

A kísérleti adataimat feldolgozom és az eredményeket értékelem. Az eredmények értékelése során a csoportok közötti különbségek elemzéséhez az átlagok páronkénti összehasonlítását veszem figyelembe, 5%-os szignifikancia szint mellett. A kapcsolatvizsgálatra varianciaanalízist használok, mely keretein belül legelsőnek normalitást, szóráshomogenitást vizsgálom. A szóráshomogenitásra Levene's Test-et alkalmazok. A normalitás és szóráshomogenitás vizsgálatoktól függően további teszteket fogok végezni a csoportok közötti különbségek elemzéséhez. Amennyiben a vizsgálat indokoltá teszi, akkor további robosztus teszt futását végzem el.

Tudományos munkám második részének megvalósítása

Első lépésként a SCOPUS és WoS adatbázisokban irodalomkutatást végez a téma mélyebb megismerése érdekében. Az irodalomkutatás során realizálok, hogy az első körben kiválasztott bemenő paraméterek mellett mely más paraméterek megléte szükséges és elengedhetetlen. Munkám következő lépéseként, a szisztematikus irodalomkutatás követően adatbázist fejleszték, amely a tudományos publikációkban fellelhető festékérzékenyített napelem cella adatokra alapoz. A tudományos munkámhoz felügyelt tanulási algoritmust alkalmazok. A kiválasztás módja az adatok jellegétől függ. A tanító adatok mellett a modellt teszt adatokkal tesztelem és a módszer jóságát mutatók alapján ellenőrzöm.

3.3 A tudományos munka gazdasági és társadalmi haszna

A kutatási eredményeim nem csak a hazai és a nemzetközi tudományos közegnek nyújtanak új ismereteket a festékérzékenyített napelemek területén, de a piaci igények kiszolgálásának is eleget tesz a DSSC gyártástechnológiáját tekintve, hozzájárulva a gazdasági kihívások eliminálásaihoz és megoldásaihoz. Ezenfelül, kutató munkám támogatja az Óbudai Egyetemen BSc, MSc képzésen tanuló, illetve szakmérnök és szakember képzéseken részvevő hallgatók alkalmazott informatikai diszciplína terén a tudásuk bővülését.

4 Elemzési módszerek és szimulációs eljárások

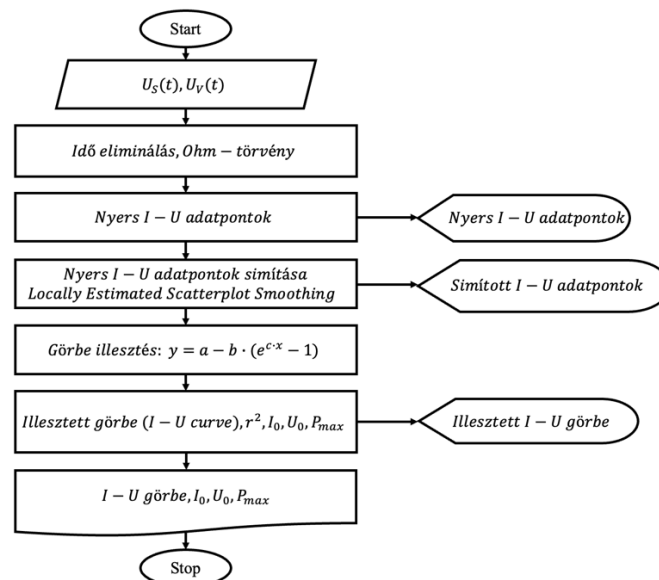
A tudományos munkám céljainak megvalósítása érdekében első lépésként kísérleteket terveztem meg és hajtottam végre. Az adatokat begyűjtöttem és értékeltem annak érdekében, hogy az előzetesen megfogalmazott hipotézisek vizsgálatát elvégezzem. Majd munkám második fázisában adatbázist fejlesztettem és predikciós modellt alkalmaztam.

4.1 Kísérletek megtervezése és elvégzése

Első lépésként, az optimális kísérleti elrendezés megtervezése során szükséges volt a napelem hőgenerátor kettős rendszer pontos, reprodukálható és könnyen monitorozható fűtését és hűtését megterveznem, melyhez egy $6,2\text{ cm} \times 6,2\text{ cm}$ TEC1-12730 típusszámú Peltier-elemet alkalmaztam. Az eljárás működését, pontosságát, reprodukálhatóságát ellenőriztem a rendszer időállandóival. A kísérleti eljárást korábbi publikációimban részletesen kifejtettem [70], [71]. A kísérleti vizsgálatokat saját tervezésű és konstrukciójú napszimulátor segítségével végeztem el, mely az American Society for Testing and Material (ASTM) E972 szabvány besorolás alapján C-kategóriájú minősítésbe sorolható [72]. A festékérzékenyített napelem cellát a céltárgytartó üvegtálcára helyezést követően a napszimulátort aktiváltam és a cella elektromos feszültség-idő ($U_s(t)$), feszültség-idő ($U_v(t)$) adatpontjait rögzítettem, mely eljárás időfüggő méréseket tartalmaz. Majd az idő eliminálásával és az Ohm-törvényt felhasználva nyers villamos áram és villamos feszültség koordinátákat határoztam meg. Ezzel párhuzamosan monitoroztam a napelem cella hőmérsékletét, a beeső elektromágneses sugárzás intenzitás-hullámhossz spektrumát és a fotonbeütésszám-hullámhossz spektrumát. A napszimulátor spektruma, mely a DSSC céltárgyat megvilágítja LED és halogén fényforrás szuperpozíciójából tevődik össze. A halogén fényforrás miatt az elektromágneses sugárzásnak nem csak fotovillamos hanem hőhatással is rendelkezik a napelemre vonatkozólag. Az IEC 60904-1 szabványban rögzített szabványosított vizsgálati körülményeket a kísérletek megtervezése és végrehajtása során figyelembe vettem. Mindezek mellett, a kísérleteim során a DSSC öregedési folyamatának lassítása céljából a festékérzékenyített napelem cellát $31\text{ }^{\circ}\text{C}$ és $40\text{ }^{\circ}\text{C}$ fok közötti hőmérséklet tartományban vizsgáltam $n = 10$ mérési mintán.

4.2 Nyers adatpontok elemzése

A kísérletek elvégzéséhez alkalmazott mérés technikai eljárást a [73] publikációban bemutatott eljárásra alapoztam, de ezen kísérleti eljárás adatelemzés részét munkám során továbbfejlesztettem. Mivel a fenti kísérleti eljárás időfüggő méréseket tartalmaz, ezért egy olyan MatLab rutint terveztem meg és készítettem el, amely a kísérletekből származó adatokon az időt eliminálva villamos áram-feszültség adatpár (koordinátájú) pontokat hoz létre. Az általam készített rutin folyamatábráját a 3. ábra szemlélteti. A villamos áram-feszültség nyers adatpontokat megkapva a rutin e nyers adatpontokon adatsimítást végez, a véletlenszerű, nem szisztematikus zaj kiküszöbölése végett. Az adatsimításhoz locally estimated scatterplot-ot (LOESS) alkalmaztam. A LOESS egy olyan nem-parametrikus regressziós eljárás, amely simítja az adatpontokat azáltal, hogy egy alacsony fokszámú polinomot illeszt az adatok lokalizált részhalmazára. Mindazonáltal, a LOESS alapvetően két változó közötti kapcsolat vizuális értékelésére szolgál. A programom következő lépéseként a villamos áram-feszültség adatpontokra a villamos áram-feszültség elméleti karakterisztika egyszerűsített egyenletét illeszti, meghatározva a szükséges illesztési állandókat és az illesztés hibáját. Az illesztett görbét követve a kívánt villamos paraméterek (U_0, I_0, MPP) a program a futása során meghatározza és táblázatba kiírja.



3. ábra. A napelem villamos áram-feszültség karakterisztikáját kiolvasó MatLab rutinom folyamatábrája, ahol az $y = a - b \cdot (e^{c \cdot x} - 1)$ a napelem elektromos áram-feszültség karakterisztikájának egyszerűsített egyenlete.

Annak ellenére, hogy a kutatásom során a fentebb említésre kerülő LOESS adatsimítási módszert alkalmaztam, számos más adatsimítási technika is elérhető, mint például a

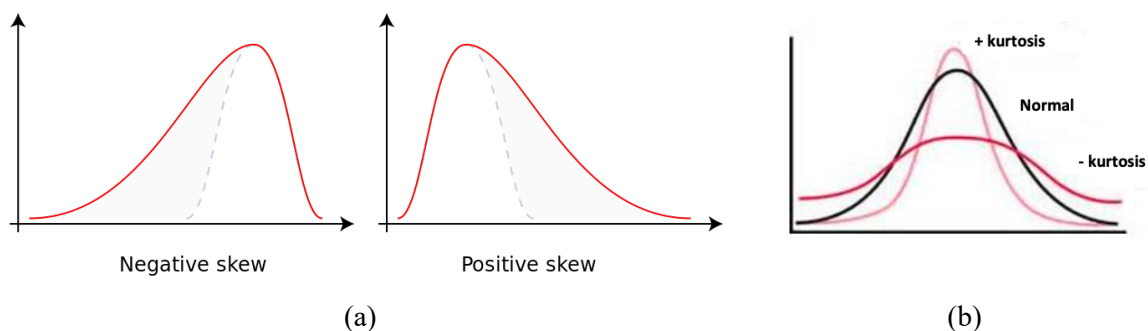
mozgóátlag-módszer. Adatsimításra azért a Locally Estimated Scatterplot Smoothing-ot használtam a mozgóátlag-módszer helyett, mert (i) jól működik nem-lineáris trendek esetén is; (ii) a helyi súlyozások miatt megfelelő pontossággal kezeli a különböző sűrűségű adatokat, ezáltal az egyenlőtlen időközökkel rendelkező adatokat is jól kezeli [74].

4.3 Alkalmazott statisztikai eljárás

Leíró statisztika

A leíró statisztikák azt a célt szolgálják, hogy az olvasó megértését segítse a nyers adatok bemutatásában, hiszen a kapcsolódó ábrák egy tömör módja az adatok leírásának. Hozzá kell tenni, hogy a leíró statisztikák esetén még hipotetikus tesztek futtatása nem történik. Az adatok leírása során a tapasztalati jellemzők úgy, mint a középérték, szórás, percentilis értékek, eloszlások mértékei szokás megadni. A központi érték leírására a jól ismert három módszert szokás alkalmazni, melyek (i) átlag ($\bar{x}$), mely az összes érték összegének és az adatkészletben lévő értékek számának hányadosa; (ii) medián, mely a nagyság szerint sorrendbe rendezett adatkészletben a középső érték; (iii) módusz, mely az adatkészletben a leggyakrabban előforduló érték. Mindezek mellett, az adatpontok középértéktől való szóródását, szétterülését is fontos megvizsgálni, melyeket az alábbi mutatókkal jellemezhetők: (i) standard hiba (SE), mely azt méri, hogy a különböző minták átlagai várhatóan mennyire térnek el a populáció valódi átlagától (szórás (s) és a minta méretének (n) négyzetgyökével való hányadosa, azaz $SE = \frac{SD}{\sqrt{n}}$); (ii) szórás (SD), mely az átlagtól való átlagos eltérés mértéke, $SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$, ahol x_i az egyes adatpontok értéke; (iii) medián abszolút eltérése (median absolute deviation, (MAD)), mely az adatok szóródásának mértékét írja le a medián körül; (iv) interkvartilis tartomány (Interquartile Range, (IQR)), mely kvantitatívan jellemzi az adatkészlet középső 50%-ának eloszlását. Mindezek mellett, a leíró statisztikák esetén érdemes megjegyezni az adatok eloszlásának vizsgálatát, mert ezáltal betekintést kapunk az adatok viselkedésébe. Két mutató – ferdeség és csúcsosság – segítségével meghatározható, hogy a kísérletekből származó adatok eloszlása mennyire tér el a normális eloszlásától. A normális eloszlást, más néven Gauss-eloszlást számos természetes folyamat és jelenség követi, így ezen paraméter vizsgálata kritériális a későbbi tesztek elvégzése céljából. A ferdeségi mutató (angolul: skewness) megmutatja (4.a ábra), hogy az adathalmaz eloszlása mennyire tolódik el a normális eloszlásától. A ferdeségi mutató előjele reprezentálja, hogy a módusz milyen irányba

mozdult el. Abban az esetben, ha a ferdeség előjele negatív, a módusz jobbra mozdul el, ezáltal az eloszlásnál domináns bal farkat láthatunk. Mindezek mellett, ha a ferdeség előjele pozitív, a módusz balra mozdul el, domináns jobb farkat eredményezve. A csúcsossági mutató (angolul: kurtosis) megmutatja (4.b. ábra), hogy az adathalmaz eloszlása mennyi laposodik el (negatív előjel) vagy csúcsosodik ki (pozitív előjel) a normális eloszláshoz képest [75]. Továbbá, érdemes megjegyezni, hogy az adatok vizuális megjelenítésével is nagyon sok információ nyerhető ki a mintából. A boxplot (dobozdiagram) öt mutató értéket ábrázol, melyek a minimum, első kvartilis (Q1), medián (második kvartilis, Q2), harmadik kvartilis (Q3) és a maximum. Az ábra a nevét a dobozról kapta, amely a kvartilisek között területet jelenti. Pontosabban, a kvartilisek (Q1-Q3) közötti területe az adatok felét tartalmazza. A boxplot ábrán szerepelnek azon adatpontok, melyek a küszöbértékeken kívül esnek, melyeket angolul outline-eknek hívnak. A kis adathalmaz esetén a küszöbértékeken kívül eső adatpontok, outline-ek, torzíthatják az adatok eloszlását és szóráshomogenitását.



4. ábra: A ferdeség és csúcsosság bemutatása, azaz a normális eloszláshoz képest az adatok hogyan változnak, ahol (a) a ferdeségi mutató (skewness, rövidítve: skew) asszimetriája ábrázolva a negatív és pozitív ferdeséget; (b) a csúcsossági mutató (kurtosis) ábrázolása pozitív és negatív esetekben.
forrás: Dr. Mark A Goss-Sampson: JASP Guide book v0.16 (2022)

Homogenitás vizsgálat

A leíró statisztikákat követően, fontos ellenőrizni az adatok konzisztenciáját a különböző csoportok között. A csoportok közötti konzisztenciák vizsgálatát, azaz megvizsgálni, hogy az adatok változásának mértéke hogyan oszlik meg a különböző csoportok között, a szóráshomogenitás vizsgálata nyújt segítséget. A homogenitás vizsgálatra két tesztet (i) Bartlett tesztet (4) és (ii) Levene tesztet (5) mutatom be részletesebben, melyeket a gyakorlatban elterjedtek. A (4) egyenlet szemléletli a Bartlett statisztikát és az (5) egyenlet pedig a Levene statisztikát.

$$T = \frac{(N - k) \ln(S^2) - \sum_{i=1}^k (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{N - k} \right)} \quad (4)$$

ahol T : Bartlett statisztika; N : összes megfigyelések száma; k : csoportok száma; n_i : i -edik csoport mérete; S^2 : teljes minta korrigált mintavarianciája; S_i^2 : i -edik csoport korrigált mintavarianciája.

$$W = \frac{(N - k) \sum_{i=1}^k n_i (\bar{Z}_i - \bar{Z})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2} \quad (5)$$

ahol W : Levene statisztika; N : összes megfigyelések száma; k : csoportok száma; n_i : i -edik részminta nagysága; $\bar{Z}_i$: Z_{ij} -k csoportjainak egyenkénti átlaga, $\bar{Z}$: a Z_{ij} -k főátlaga, $Z_{ij} = |Y_{ij} - \bar{Y}_i|$, ahol $\bar{Y}_i$ az i -edik részminta átlaga [76].

A Bartlett teszt és a Levene teszt közül a Levene teszt a leggyakrabban használt módszer a szóráshomogenitás vizsgálatára, mert nem igényli az adatok normális eloszlását, ezáltal kevésbé érzékeny a normalitásra [77]. A Levene teszt esetén a nullhipotézis azt mondja ki, hogy a csoportok variációi egyenlők, azaz $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$, ahol σ_i^2 az i -edik csoport variációját jelöli, és a k a csoportok száma.

Varianciaanalízis

A varianciaanalízis, más néven ANOVA, egy olyan statisztikai módszer, melynek célja a különböző csoportok közötti átlagértékek vizsgálata. Azon vizsgálatok során, ahol több csoportot hasonlítunk össze, például több hőmérsékleti csoportban lévő villamos teljesítményt, elengedhetetlen. A normalitás és a szóráshomogenitás függvényében választunk tesztet a varianciaanalízisre: (i) ha az eloszlás normális és a szóráshomogenitás nem sérül szignifikánsan, akkor a hagyományos, klasszikus egyirányú ANOVA; (ii) ha az eloszlás normális eloszlású és a szóráshomogenitás sérül szignifikánsan, akkor Brown-Forsythe-próba; (iii) ha a normális eloszlás feltétele nem teljesül és a szóráshomogenitás nem sérül szignifikánsan, akkor Kruskal-Wallis próba; (iv) ha egyik feltétel sem teljesül, akkor Welch-próba. Dolgozatomban a tartalmi korlátokból kifolyólag csak a Kruskal-Wallis nem parametrikus statisztikai próbát mutatom be, amit három vagy több független csoport közötti középértékek különbségeinek vizsgálatára alkalmaznak. A teszt az egyutas varianciaelemzés (one-way ANOVA) nemparaméteres verziója, amely Mann-Whitney statisztika kiterjesztett változata több csoportra nézve. Az említett teszt alkalmazása azért előnyös, mert robusztus a

kiugró értékekkel szemben, azaz kevesebb megszorító feltétel mellett ad választ a probléma megválaszolására. A próba a nullhipotézis (H_0) vizsgálatára alkalmas, mely H_0 : 'A villamos teljesítmény értékek nem különböznek szignifikánsan egymástól', azzal a megjegyzéssel, hogy a középértékek eltérésére érzékeny, nem pedig a szóródási mutatókéra. Más szavakkal, ha a $p < 0.05$, akkor a nullhipotézist megtagadjuk, ellenkező esetben az ellenhipotézist (H_1) tekintjük, azaz H_1 : 'A villamos teljesítmény értékek szignifikánsan különböznek egymástól'. A (6) egyenletet reprezentálja a Kruskal-Wallis próba egyenletét, ahol N : az összes adatpontok száma; k : a csoportok száma; R_i : az i -edik csoport rangösszege; n_i : az i -edik csoport elemszáma.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (6)$$

A hipotézisvizsgálatot követően további elemzések szükségesek, melyek a post-hoc teszt, mely segítségével azonosítható, hogy mely csoportok között vannak jelentős (szignifikáns) eltérések. A post hoc teszt kiválasztása során figyelembe vettem a normalitás és a szóráshomogenitás változását is. A paraméteres próbák esetén (pl.: Wells próba) a hatásméretet is fontos meghatározni, mely segítségével kvantifikáljuk, mennyire magas a független változó magyarázó ereje a függő változóra nézve. A kutatási gyakorlatban széles körben a parciális eta négyzetet (η_p^2) használják és elfogadott, hogy ha η_p^2 alacsonyabb, mint 0,1, akkor a hatásméretet alacsony; ha η_p^2 0,1 és 0,35 között helyezkedik el, akkor a hatásmérete közepes; ha η_p^2 magasabb, mint 0,35, akkor a hatásméretet magasnak értékeljük [77]. Tekintettel arra, hogy a Kruskal-Wallis teszt nemparaméteres vizsgálat, hatásméret nem kerül számításba.

4.4 Gépi tanulási modellek

Az irodalomkutatásból kiderült, hogy a kutatások egy jelentős része a gépi tanulási algoritmusok (angolul: Machine Learning, rövidítve ML) irányába mozdult el, amely lehetővé teszi, hogy a fontos következtetéseket vonjunk le a cella készítése előtt. Az alábbi fejezetben részletesen bemutatom a döntési fát, k-legközelebbi szomszédok, support vector machine, véletlen erdő és neurális háló módszereket.

Döntési fa (Decision Tree, DT)

A gépi tanulási algoritmusok közül legelsőnek a döntési fát (angolul: Decision Tree, rövidítve DT) mutatom be, mely az osztályozási feladatoknak az egyik legismertebb és széles körben alkalmazott eszköz. A DT-ről elmondható, hogy egy fa formájú folyamatábrát eredményez, melynek gyökeréből kiindulva, fentről lefelé a csúcsok közötti éleken haladva a belső csúcsok egy-egy attribútum értékének vizsgálatával a döntéshez (levélcsúcsokhoz) eljutunk. A fa struktúra miatt a DT osztályozási szabályokat foglal magába, mely összességét szabálybázisnak nevezzük. A gyakorlatban a bináris DT terjedt el, mely azt jelenti, hogy az egyes csomópontoknak két gyermeke van. Ugyanakkor, előfordulhat, hogy egyes attribútumok nem jelennek meg a döntési fában, mert nem jelennek meg a döntésben, döntéshozatalban. Ezen elemeket invariánsnak tekintjük. A fa készítése során kezdetben egyetlen csúcsból indulunk ki, majd az az attribútum kerül kiválasztásra, mely a legjobban szeparálja (szétosztja) a halmazt. Az elágazások elkészítését követően az algoritmus rekurzívan elvégzi az eljárást a minták minden egyes partíciójánál. A rekurzív folyamat leállításának feltétele a következő: (i) nem maradt attribútum, ami alapján tovább bonthatnánk a mintát; (ii) az összes adatpontot osztályba soroltuk; (iii) egy adott csomópont tartalmazza az adatpontok olyan halmazát, melyek ugyanabba az osztályba esnek [78]. Léteznek különböző típusú döntési fa algoritmusok, köztük népszerű az iteratív dikotomizáló 3 (ID3), az ID3 későbbi iterációja (C4.5), valamint a klasszifikációs és regressziós (CART) [8]. Az egyik leggyakoribb osztályozó algoritmus a CART (Classification And Regression Tree) algoritmus, mely a Gini index alapján választja ki az attribútumot, melyet a (7) egyenlet segítségével adhatjuk meg [79].

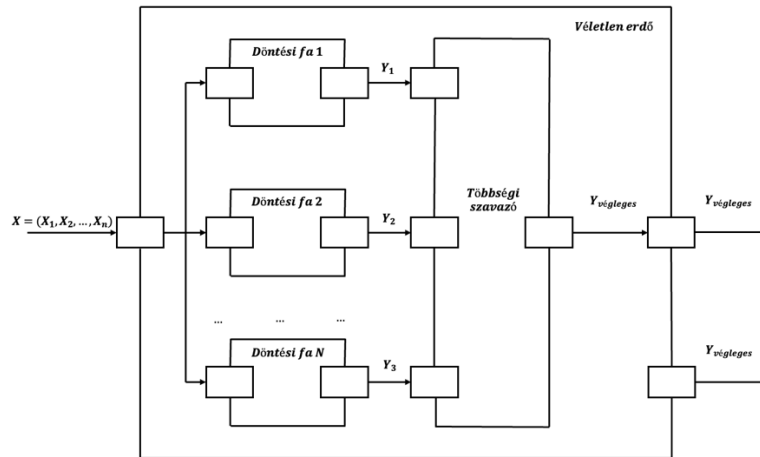
$$Gini(D) = 1 - \sum_{i=1}^m p_i^2 \quad (7)$$

ahol p_j a j-edik osztály relatív előfordulását jelöli a D halmazban, mely M mintát tartalmaz és a mintákat soroljuk m különböző osztályba.

Véletlen Erdő (Random Forest, RF)

A véletlen erdő (angolul: Random Forest, rövidítve RF) modell a döntési fák koncepciójára épül, de a túltanulás – másképpen túlillesztés – veszélyes csökken. Az RF bemeneti adatai a különálló döntési fák részhalmaza, melyek külön-külön adnak egy-egy eredményt. Az egyes fák által kapott egyedi előrejelzések aggregációját követően, többségi szavazás – azaz a

leggyakoribb eredmény – segítségével megkapjuk a végső eredményt. Az 5. ábra szemlélteti a véletlen erdő működési elvét, ahol megfigyelhető a közvetlen a bemenetnél lévő döntési fák.



5. ábra. Véletlen erdő működési elve

A véletlen erdő előnyei, hogy a túltanulásra kevésbé érzékeny, pontosabban képes elkerülni a túlilleszkedést, ami a döntési fák esetén jelentkezhet [80].

K-legközelebbi szomszéd (k-NN)

A k-legközelebbi szomszéd (angolul: k-nearest neighbors, k-NN) algoritmus egy „lusta” klasszifikációs és regressziós módszer a gépi tanulási algoritmusok között. Az algoritmus azokra a mintákra támaszkodik, melyek az adathalmazban, mint tanuló adatok megtalálhatók, vagyis nem igényel explicit modellépítést [81]. Az algoritmus három nagyobb lépésből áll, melyek a következők:

1. *Távolságmérés:* a tesztelő és a tanító adathalmaz pontjai közötti előre definiált távolság meghatározása. Továbbá, feltesszük, hogy a tanuló mintákat n -dimenziós numerikus attribútumok írják le, melyek n -dimenziós térben helyezkednek el. A távolság meghatározására a leggyakrabban az euklideszi távolságot és a Manhattan távolságot alkalmazzák, melyek az alábbi képlettel definiáljuk, ahol (8) egyenlet az euklideszi (másnéven L2 norma) távolság, amely az attribútumok közötti legközvetlenebb távolságot méri az n -dimenziós térben és a (9) egyenlet a Manhattan (másnéven L1 norma) metrika, amely az attribútumok értékkülönbségeinek abszolút összegét veszi.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (8)$$

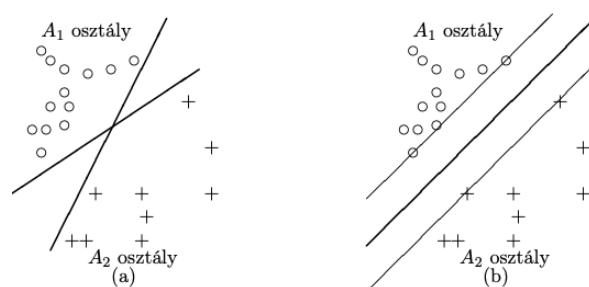
$$d(X, Y) = \sum_{i=1}^n |x_i - y_i| \quad (9)$$

2. *Legközelebbi szomszédok kiválasztása:* A k minimális távolság meghatározása a tanító adathalmaz alapján. Nevezetesen, kiválasztjuk a tanító adathalmazból azokat a k darab szomszédokat, melyek a teszt adatbázisból származó teszt adatpontjához a legkisebb távolságra helyezkednek el figyelembe véve az alkalmazott metrikát.
3. *Osztályozás:* A legközelebbi szomszédok kiválasztását követően, a k leggyakoribb osztálycímke alapján történik a döntés meghozatala. A szavazási mechanizmus esetén súlyozási módszert is lehet alkalmazni, mely esetén a k szomszéd minden tagjának a súlyát úgy határozzuk meg, hogy a prediktálandó (osztályozandó) ponttól mért távolságának vesszük az inverzét [82].

Fontos megjegyezni, hogy a módszer érzékeny a metrikára, vagyis a távolság helyes megválasztásával a modell becslése növelhető.

Tartóvektor-gép (Support Vector Machine, SVM)

A tartóvektor-gépek (angolul: support vector machine, rövidítve: SVM) célja, hogy megtalálja a legnagyobb távolságú hipersíkot (döntési határt) a különböző osztályokhoz tartozó adatpontok között. Az SVM széles körben elterjedt a regressziós és klasszifikáció problémák megoldására, alkalmazzák karakterfelismerésre, képfeldolgozásra, adatbányászatra és még bioinformatika területén is fontos szerepet tölt be.



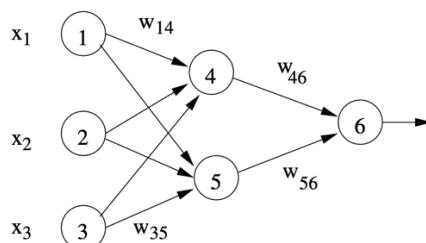
6. ábra. Hipersíkkal történő elválasztásra példa, ahol a hipersík kétdimenziós egyenes. (a) az elválasztás nem optimális; (b) maximális margójú elválasztás, ahol vékony vonallal a margók és a két margó között elhelyezett vastag vonal a határsávot reprezentálja [83].

Ahogy fentebb említésre került, az SVM eljárás során a legjobb elválasztó hipersíkot keressük, melyre egy szemléltetes példát láthatunk a 6. ábrán, ahol (a) a nem optimális esetet láthatunk, mert a margók keskenyek és (b) optimális szeparálás figyelhető meg, ugyanis az elválasztó sík a lehető legtávolabb van a minta pontokról [83]. A támogató vektorok azok az

adatpontok, melyek a döntési határtól a legközelebb esnek; a 6.(b) ábrán a margókra illeszkedő minta pontok.

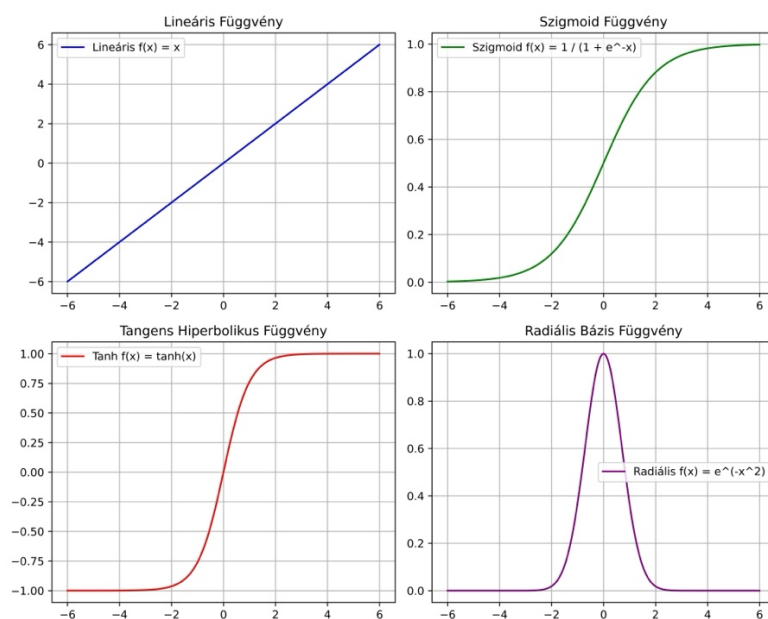
Neurális hálózat (Artificial Neural Network, ANN)

A mesterséges neurális hálózatoknak az agyműködést utánzó analógiája van, mert a neurobiológia és az informatika összefonódása bonyolultabb modellezési technikákat tesz lehetővé. Az ANN mesterséges neuronokból állnak, melyek az egymásnak küldött jeleken keresztül kommunikálnak egymással és a jelek nagyszámú súlyozott kapcsolatban áramlanak. Az egyik legnépszerűbb modell a többrétegű előrecsatolt neuronhálózat, melyet a 7. ábra szemléltet, ahol az első réteg csomópontjaiban helyezkednek el az inputok, melyek az ábrán 1, 2, 3 neuronokként vannak jelölve, és a legutolsó réteg kimenete az output (az ábrán 6. neuron).



7. ábra. Mesterséges neurális hálózat egy részlete: többrétegű előrecsatolt neurális hálózat [84].

Az input és az output neuronok közötti rétegek, másnéven a közbenső rétegeket rejtett rétegeknek (angolul: hidden layer) nevezzük, melyek a 7. ábrán 4. és 5. neuronokként szerepelnek. Az ábrán továbbá megfigyelhető, hogy minden réteg minden neuronjának kimenete a következő réteg neuronjának a bemenete lesz és a neuronok közötti kapcsolat szorosságát súlyok (w_{ij}) jellemzik. Az irodalmak azt mutatják, hogy a nemlineáris paraméterezett függvényekkel, vagyis súlyozás használatával jobb közelítés érhető el [85]. A kimenet generálása valamilyen transzferfüggvény (activation function) segítségével történik, melyek közül párat emelek ki, melyek a gyakorlatban elterjedtek: lineáris; szigmoid függvény; tangens hiperbolikus; logisztikus és radiális. A 8. ábra szemlélteti a gyakorlatban alkalmazott transzferfüggvényeket. A lineáris függvény hátránya, hogy nem képes a nem-lineáris összefüggések modellezésére, ezért a komplexebb feladatok modellezése során korlátozott a használata. A szigmoid függvény jól használható bináris osztályozási feladatoknál. A logisztikus függvény pedig gyakorlatilag megegyezik a szigmoid függvénnyel, ugyanis a logisztikus függvény a szigmoid függvény egy speciális esete. A tangens hiperbolikus függvény ugyancsak egy szigmoid típusú függvény, de a függvény -1 és 1 közötti értékeket vesz fel [86].

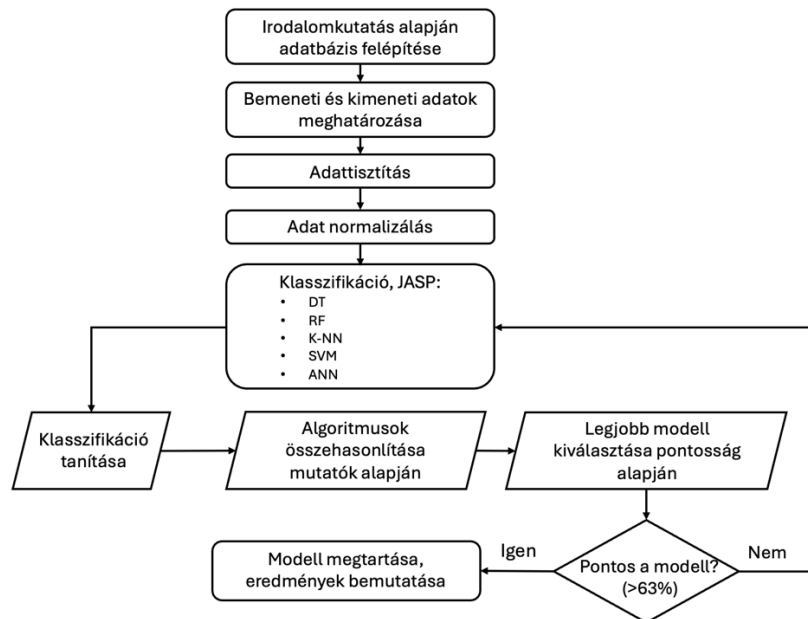


8. ábra. Mesterséges neurális hálózathoz alkalmazott leggyakoribb transzferfüggvények.

4.5 Munka folyamatábrája

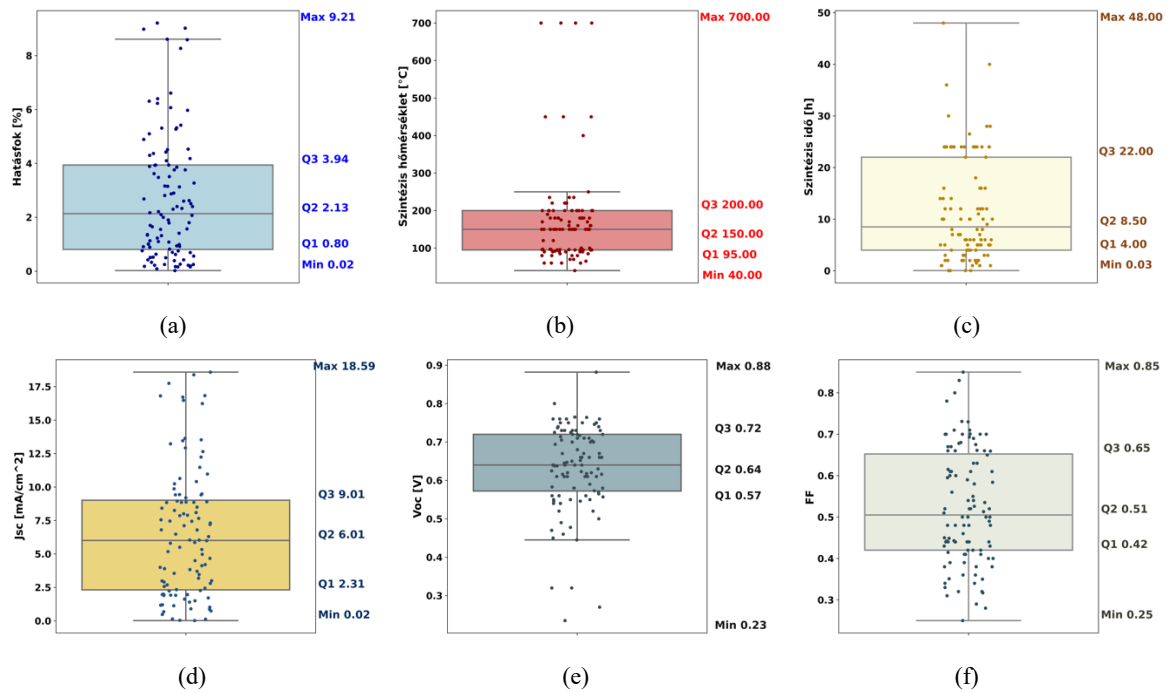
A statisztikai és a gépi tanulási módszerek bemutatását követően munkám ezen részfejezetében kifejtem a tudományos munkám gépi tanulási algoritmusára vonatkozó folyamatokat, mely folyamatábráját a modell fejlesztés fázisáról a modell elfogadásáig a 9. ábra szemlélteti. Az irodalomkutatás során összesen 68 darab publikációt dolgoztam fel, ezáltal 168 kísérleti megfigyelésből álló adathalmazt fejlesztettem. A feldolgozott publikációk 2005 és 2022 között kerültek publikálásra. Továbbá, az elemzés során kiemelt figyelmet fordítottam a 168 rekordom minőségének biztosítására, ennek érdekében adattisztítási folyamat alkalmazása elengedhetetlen volt. Az adattisztítás célja az adatok minőségének javítása, ugyanis az adatokban lévő hibák kiszűrése és eltávolítása növeli az adatok konzisztenciáját. A változóknak (attribútumoknak) három típusát különböztetjük meg, melyek (i) kategorikus attribútum, mely csak megadott értékek közül választhat; (ii) numerikus attribútum, mely számértéket (pl.: mérési eredményt) tartalma; és (iii) bináris attribútum, melyek csak két értéket vehetnek fel (pl.: igen – nem, melyet 1 – 0-ra lehet lefordítani). Az adattisztítás esetén fontos megemlíteni a hiányzó értékek kezelését, hiszen minden attribútumot ismerni szükséges a vizsgálatok elvégzése érdekében. A hiányzó értékek kezelésére számos megoldás áll rendelkezésre, melyek a következők: (i) azon sorokat, melyek hiányosak nem vesszük figyelembe; (ii) attribútum átlagértékével pótoljuk a hiányzó értéket; (iii) a legvalószínűbb értékkel pótoljuk a hiányzó

értéket. Az adattisztítás során a hiányos sorokat törtöltem, ezáltal összesen 106 rekordon végeztem a vizsgálataimat.



9. ábra. Tudományos munkám gépi tanulási alkalmazásának lépései folyamatábraként szemlélítve a modell fejlesztés fázisától a modell elfogadásáig, ahol DT: döntési fa; RF: véletlen erdő; K-NN: K-legközelebbi szomszéd; SVM: tartóvektor-gép; és ANN: mesterséges neurális hálózatot jelenti.

Sutar és munkatársai adatnormalizációs módszereket alkalmaztak a szintézis hőmérsékletén, szintézis idején és a hatásfokon [68]. Kutatásom során hasonló adatnormalizációs technikát alkalmaztam, melyet további változókon is implementáltam úgy, mint az üresjárási feszültség, rövidzárási áramsűrűség és kitöltési tényező. Az adatnormalizálási folyamat során a minimum, maximum és a kvantálási intervallumokat használtam fel, ugyanis a minimum és az első kvartilis (Q1) közötti értékek alacsony; a Q1 és a második kvartilis (Q2) közötti értékek közepes; Q2 és harmadik kvartilis (Q3) közötti értékek magas; és Q3 és maximum közötti értékek nagyon magas kategóriákba sorolódnak. Az adatok normalizálása után a kezdeti vizsgálatokból megállapítottam, hogy a szintézis hőmérsékletének és időtartamának normalizált értékei alacsony predikciós pontosságot eredményeznek, ennek érdekében folytonos értéként vettem ezeket figyelembe. A 10. ábra szemlélteti a hat paraméter adatnormalizálásának bemutatását dobozdiagramok segítségével, ahol (a) a hatásfok; (b) a szintézis hőmérséklete; (c) a szintézis ideje; (d) rövidzárási áramsűrűség (J_{sc}); (e) üresjárási feszültség (V_{oc}); (f) kitöltési tényező (FF).



10. ábra. Különböző paraméterek adatnormalizálását bemutató dobozdiagram, ahol (a) a hatásfok; (b) a szintézis hőmérséklete; (c) a szintézis ideje; (d) a rövidzárási áramsűrűség (J_{sc}); (e) a üresjárási feszültség (V_{oc}), (f) a kitöltési tényező (FF)

A gépi tanulási algoritmusok összehasonlítására különböző mérőszámokat alkalmaztam, melyek a precizitás (precision) a visszahívás (recall) és az F1 pontszám.

A precizitás egy olyan arányszám, mely a modell pozitívként helyesen azonosított tételének és az összes pozitív előrejelzés hányadosa, melyet a (10) egyenlet szemléltet.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

ahol TP (True Positives) azoknak a pozitív példának a száma, melyeket a modell helyesen azonosított pozitívként és FP (False Positives) azoknak a negatív példának a száma, melyeket a modell tévesen azonosított pozitívként. Más szavakkal, a precizitás megmondja, hogy a modell által pozitívként azonosított esetek mennyire pontosak. A visszahívás ((11) egyenlet) megmutatja, hogy a modell milyen arányban ismeri fel az összes lehetséges pozitív esetet. Más szavakkal, a helyesen azonosított tételek (TP) és az összes pozitív tételek hányadosa ($TP + FN$).

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

ahol FN (False Negatives), mely azoknak az eseteknek a száma, ahol a modell tévesen azonosított egy esetet negatívként, pedig pozitív volt.

Az F1 pontszámot (F1 score) az előző két metrika segítségével állítjuk el, mégpedig a (12) egyenlet alapján, mely az alábbi:

$$F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (12)$$

Az F1 mérőszám meghatározása azért fontos, mert előfordulhat, hogy az előző két mérőszám ellentétben áll egymással, azaz a precizitás növelése csökkentheti a visszahívást és fordítva. Ennek érdekében, az F1 érték megadása, ami a precizitás és a visszahívás harmonikus közepét adja, egyensúlyozza a két mérőszámot. A modellek összehasonlítását követően azon klasszifikációs algoritmust választottam ki, mely a legmagasabb mérőszámokkal rendelkezik. Majd a tesztalmazon mért pontossági ráta (angolul: test accuracy) alapján validáltam a modellt. Amennyiben a teszt pontossága meghaladja a 63%-os küszöbértéket, a modellt alkalmazhatónak ítélt meg, ellenkező esetben – avagy, ha a teszt pontosságára adott érték kisebb a küszöbértéknél – újraindítottam az algoritmust. A teszt pontossága a (12) egyenletben szemléltetett egyenlet alapján adható meg.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

A 63%-os küszöbértéket a Hisham A. Maddah munkája által határoztam meg, akik hasonlóképpen klasszifikációs módszereket alkalmazva vizsgálták a festékérzékenyített napelem cella természetes fényérzékeny festékanyagát [8].

5 Eredmények

Az alábbi fejezetben részletesen kifejtem a statisztikai vizsgálatokra és a gépi tanulási algoritmusok alkalmazására kapott eredményeimet. A statisztikai vizsgálatok során a villamos munkaponti teljesítményt vizsgáltam a hőmérséklet függvényében, hiszen közvetlenül meghatározza a napelem hatásfokát (lásd (3) egyenlet). Továbbá, a klasszikus statisztikai módszerek mellett, munkám során gépi tanulási algoritmusokat is alkalmazok a festékkérszenyített napelem cella hatásfokának előre jelzésére a gyártástechnológiai folyamatok figyelembevétele mellett. Ez a multidiszciplináris megközelítés lehetővé teszi, hogy részletesebb és több aspektusból elemezzem a festékkérszenyített napelem cella működését.

5.1 Statisztikai vizsgálat eredményei

A csoportok közötti különbségek elemzéséhez az átlagok páronkénti összehasonlítását vettem figyelembe, 5%-os szignifikancia szint mellett. Az elemzéshez a JASP 0.16.4. programcsomagot használtam. A kapcsolatvizsgálatra varianciaanalízist használtam (ANOVA), melynek keretein belül legelsőnek normalitást, szóráshomogenitást vizsgáltam, majd a normalitás sérülése és a szóráshomogenitás szignifikáns nem sérülése miatt Kruskal-Wallis tesztet végeztem, melyet Bonferroni post-hoc tesztel és parciális eta négyzet (η_p^2) hatásmérettel ellenőriztem.

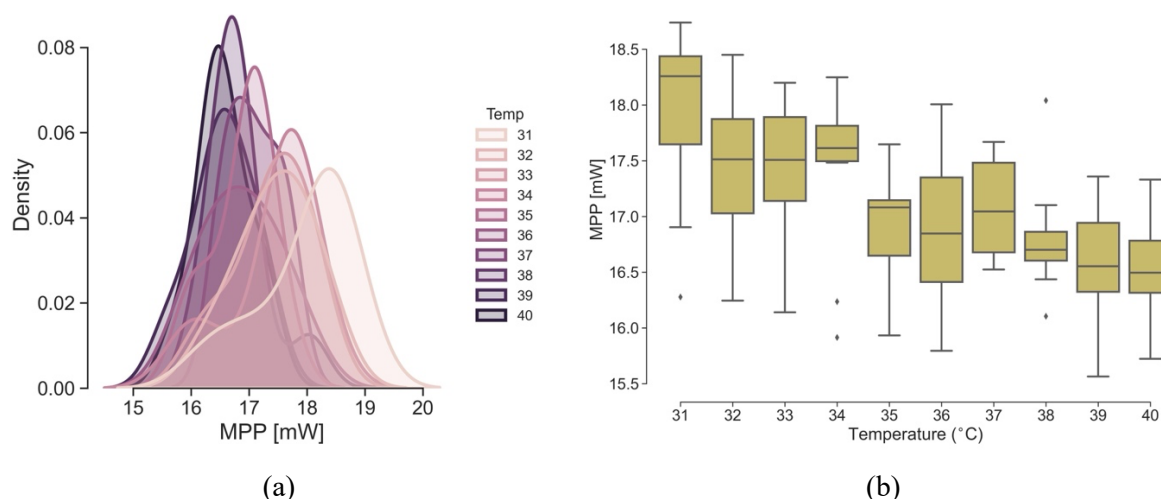
5.1.1 Leíró statisztika

Az alábbi szekcióban bemutatom a különböző mérőszámokat, melyeket vizsgáltam a statisztikai mintámon úgy, mint az átlag, medián, modus, minta szórásának standard hibája, szórás, ferdeség és csúcsosság. A különböző mérőszámokra kapott eredmények táblázatos összefoglalását a 2. táblázat szemlélteti. Az adatokból jól látszódik, hogy minden hőmérsékleten 10 darab mérés került elvégzésre. Továbbá, a ferdeségi és csúcsossági mutatókból az olvasható ki, hogy egyes csoportokban a normalitás sérül. Pontosabban, 31 °C, 34 °C, 38 °C csoportokban a ferdeség sérül és 34 °C, 37 °C, 38 °C csoportokban pedig a csúcsosság mutat sérülést. Az adatok táblázat összefoglalása után vizuális eszközökkel is megvizsgáltam az adathalmazt annak érdekében, hogy átfogó képet kapjak az eloszlásról.

2. táblázat: Különböző mérőszámok összefoglalása leíró statisztikához, ahol az SE a standard hiba és az SD a szórás.

	MPP [mW]									
	31	32	33	34	35	36	37	38	39	40
Darabszám	10	10	10	10	10	10	10	10	10	10
Átlag	17,964	17,466	17,385	17,419	16,883	16,864	17,073	16,803	16,553	16,56
Medián	18,259	17,512	17,507	17,613	17,082	16,847	17,045	16,701	16,553	16,496
Modus	16,278	16,244	16,14	15,914	15,933	15,795	16,524	16,105	15,564	15,723
SE	0,259	0,211	0,208	0,236	0,167	0,216	0,141	0,161	0,17	0,148
SD	0,818	0,667	0,656	0,746	0,527	0,682	0,445	0,51	0,537	0,468
MAD	0,329	0,479	0,439	0,191	0,162	0,535	0,465	0,153	0,385	0,241
Ferdeség	-1,257	-0,376	-0,75	-1,412	-0,761	0,098	0,155	1,584	-0,353	0,081
Csúcsosság	0,635	-0,223	-0,044	1,07	-0,01	-0,629	-1,572	4,064	-0,03	0,26

Az 11. ábra szemlélteti a DSSC napelem villamos teljesítmény értékek vizuális megjelenítését a hőmérséklet függvényében, ahol (a) reprezentálja a sűrűségfüggvényt különböző színárnyalatokat alkalmazva az intervallumok jelölésére és (b) mutatja a boxplot ábrát, ahol a dobozok a Q1 és Q3 kvartilisek közötti értékeket mutatják. Továbbá, az ábrán a küszöbértékeken kívül eső adatpontok is megfigyelhetők. A kis mintaszám miatt ezen adatpontok eliminálása nem volt indokolt, ugyanakkor valószínűsíthető, hogy előfordulása a mérési pontatlanságra vezethető vissza.



11. ábra: A festékképzékenyített napelem cella villamos teljesítményének vizsgálata a hőmérséklet függvényében, ahol (a) a sűrűségfüggvény grafikon különböző hőmérsékleti értékek esetén (b) boxplot ábra a maximális villamos teljesítményről a hőmérséklet függvényében.

5.1.2 Homogenitás vizsgálat

A normalitás ellenőrzését követően Levene tesztet futtattam annak érdekében, hogy a csoportok közötti konzisztenciát vizsgáljam. A nullhipotézis: a csoportok varianciái egyenlők, azaz H_0 : 'A csoportok közötti varianciák egyenlők'. A gyakorlatban legtöbbször alkalmazott Levene's tesztre az alábbi táblázatban összefoglalt eredményeket kaptam, ahol F a statisztika eredménye; a $df1$ a számláló szabadságfoka (angolul: degree of freedom of the numerator), ami $k - 1$ képlet segítségével számolható ki, ahol k a csoportok száma; $df2$ a nevező szabadságfoka (angolul: degree of freedom of the denominator), ami az összes megfigyelés száma mínusz a csoportok száma ($N - k$); és a p érték reprezentálja, hogy az eredmény statisztikailag szignifikáns-e vagy sem figyelembe véve a szignifikancia szintet.

3. táblázat: Szóráshomogenitás vizsgálatára kapott eredmények, ahol F a statisztikai számítás eredménye, $df1$ a számláló szabadsági foka, $df2$ a nevező szabadsági foka és a p az eredmény statisztika szignifikanciája.

F	df1	df2	p
0,811	9,000	90,000	0,607

A 3. táblázat alapján a nullhipotézist megtartjuk a 0,05 szignifikancia szint mellett, hiszen $F(9,90) = 0,811, p = 0,607$. Más szavakkal, nincs elegendő statisztikai bizonyíték arra nézve, hogy jelentős különbségek lennének az egyes csoportok varianciái között, melyből arra következtethetünk, hogy az ANOVA vizsgálat eredményei nem lesznek torzítottak.

5.1.3 ANOVA

Figyelembe véve a normalitás sérülését a ferdeségi és csúcsossági mutatók alapján – másképpen, a normális eloszlás nem realizálódik a függő változók esetében – és figyelembe véve a szóráshomogenitás konzisztenciáját, varianciaanalízisre Kruskal-Wallis próbát alkalmaztam. Az 4. táblázat mutatja a Kruskal-Wallis teszt eredményeit és a teszt alapján 0,05-ös szignifikancia szint mellett, statisztikailag különbség van a hőmérsékleti csoportok villamos teljesítményében, mivel $H(9) = 31,759, p < 0.001$. Más szavakkal, a nullhipotézist, (H_0)-t, elutasítjuk és az ellenhipotézis, (H_1) teljesül, azaz található legalább egy hőmérsékleti csoport, amely esetén a villamos teljesítmény értékei szignifikánsak eltérnek a többi csoporttól.

4. táblázat: Kruskal-Wallis teszt eredményei a varianciaanalízis esetén, ahol N a minta nagysága, a $Statistic$ a Kruskal-Wallis teszt statisztikája, a df a szabadsági fok és a p a valószínűségi szint.

N	Statistic	df	p
100	31,759	9	<.001

A nemparaméteres vizsgálatot követően a csoportok közötti különbségek pontosabb megismerése érdekében Bonferroni utótesztet alkalmaztam és a vizsgálat eredményei által meghatároztam, hogy mely csoportok között van szignifikáns eltérés:

- $31\text{ }^{\circ}\text{C}$ és $35\text{ }^{\circ}\text{C}$ ($p = 0.006$); $31\text{ }^{\circ}\text{C}$ és $36\text{ }^{\circ}\text{C}$ ($p = 0,005$); $31\text{ }^{\circ}\text{C}$ és $37\text{ }^{\circ}\text{C}$ ($p = 0,05$); $31\text{ }^{\circ}\text{C}$ és $38\text{ }^{\circ}\text{C}$ ($p = 0,002$); $31\text{ }^{\circ}\text{C}$ és $39\text{ }^{\circ}\text{C}$ ($p < 0,001$); $31\text{ }^{\circ}\text{C}$ és $40\text{ }^{\circ}\text{C}$ ($p < 0,001$).
- $32\text{ }^{\circ}\text{C}$ és $39\text{ }^{\circ}\text{C}$ ($p = 0,042$); $32\text{ }^{\circ}\text{C}$ és $40\text{ }^{\circ}\text{C}$ ($p = 0,045$).

Az eredmények alapján a magasabb hőmérsékleti csoportokban, $33\text{ }^{\circ}\text{C}$ -tól, az adatok eltérő tendenciát mutatnak. Továbbá, a Bonferroni utóteszt jól alátámasztja, hogy a megfigyelt különbségek nem véletlen fluktuációk. A mintaszám növelésével és a hőmérsékleti tartomány kiszélesítésével az eredmények pontosíthatók.

5.2 Gépi tanulási algoritmusok eredményei

A szisztematikus irodalomkutatásból adatbázist fejlesztettem, majd az adathalmazt alapos előfeldolgozásnak vettem alá, mely magába foglalta az adattisztítást és normalizálást. A bemeneti paramétereim: félvezető típusa; félvezető réteg vastagsága; morfológiai szerkezet; prekursor; fényérzékeny festékanyag típusa; szintézis hőmérséklete; szintézis ideje. Az adatfeldolgozást követően különböző klasszifikáló algoritmusok – úgy, mint döntési fa (DF), véletlen erdő (RF), K-legközelebbi szomszédok (K-NN), vektortartó-gép (SVM) és mesterséges neurális hálózat (ANN) – tesztelését végeztem a JASP statisztikai szoftver mellett. A következő alfejezetekben részletesen bemutatom a különböző osztályozó algoritmusok teljesítményeit, majd azokat összehasonlítom három mérőszám alapján minden egyes villamos paraméterre (hatásfok, rövidzárási áramsűrűség, üresjárási feszültség és kitöltési tényező) nézve. Az 5. táblázatba összefoglaltam a predikciós modell esetén alkalmazott kategorikus jellemzőket, azok változóit és a kísérleti megfigyelések számát.

5. táblázat: DSSC-ket alkotó különböző jellemzők kategorikus összefoglalása feltüntetve a megfigyelések számával, ahol Ti-dipac a titán diizopropoxid bisz(acetilacetónát); Ti-butox a titán butoxid; TTIP a titán(IV)-izopropoxid; Ti-tetrabut a tetrabutyl-titanát; TiCl₄ a titán tetraklorid

Kategorikus jellemző	A DSSC-k változói	Kísérleti megfigyelések száma
Félvezető réteg típusa	1: cink-oxid (ZnO)	55
	2: titánium-dioxid (TiO ₂)	49
Félvezető réteg vastagsága		104
Morfológiai szerkezet	4: Nanorod	70
	7: Nanowire	11
	9: Nanoszemcse	23
Prekursor	1: cink-nitrát	30
	2: cink-acetát	19
	3: cink-klorid	6
	4: Ti-dipac	4
	5: Ti-butox	17
	6: Ti-tetrabut	19
	7: TiCl ₄	9
Fényérzékeny festékanyag	1: N719	67
	2: N3	26
	3: D102	3
	5: természetes festékanyag, Z907, C-218, D205	8
Szintézis hőmérséklete		104
Szintézis ideje		104
Hatásfok [%]	1: alacsony: 0,02 – 0,8	26
	2: közepes: 0,8 – 2,13	26
	3: magas: 2,13 – 3,94	26
	4: nagyon magas: 3,94 – 9,21	26
Rövidzárási áramsűrűség [mA/cm ²]	1: alacsony: 0,02 – 2,31	26
	2: közepes: 2,31 – 6,01	26
	3: magas: 6,01 – 9,01	26
	4: nagyon magas: 9,01 – 18,59	26
Üresjárási feszültség [V]	1: alacsony: 0,23 – 0,57	26
	2: közepes: 0,57 – 0,64	22
	3: magas: 0,64 – 0,72	29
	4: nagyon magas: 0,72 – 0,88	27
Kitöltési tényező	1: alacsony: 0,25 – 0,42	25
	2: közepes: 0,42 – 0,51	27
	3: magas: 0,51 – 0,65	26
	4: nagyon magas: 0,65 – 0,85	26

5.2.1 Hatásfok

Az 6. táblázat szemlélteti a különböző algoritmusok teljesítményeinek összehasonlítását a precizitás (precision), visszahívás (recall) és F1 pontszám (F1 score) mérőszámok alapján a hatásfok villamos paraméterre nézve. A táblázatból jól látható, hogy a precizitás mutató átlagára a véletlen erdő (RF) algoritmus adja a legnagyobb értéket (0,912), majd ezt követi a vektortartó-gép (0,855), K-legközelebbi szomszédok (0,817), majd a neurális hálózat (0,723) és végül a döntési fa (0,615).

6. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján a hatásfok paraméterre nézve.

PRECIZITÁS					
	1	2	3	4	Átlag
DT	0,667	0,5	0,667	1	0,675
RF	1	0,5	0,75	1	0,912
K-NN	0,333	0,667	1	1	0,817
SVM	0,75	0,5	1	0,9	0,855
ANN	0,333	1	0,667	0,6	0,723
VISSZAHÍVÁS					
	1	2	3	4	Átlag
DT	0,75	0,2	1	1	0,7
RF	0,714	1	1	0,875	0,85
K-NN	0,5	0,857	0,6	0,833	0,75
SVM	0,75	0,5	0,8	1	0,85
ANN	0,667	0,429	0,571	1	0,6
F1					
	1	2	3	4	Átlag
DT	0,706	0,286	0,8	1	0,664
RF	0,833	0,667	0,857	0,933	0,86
K-NN	0,4	0,75	0,75	0,909	0,763
SVM	0,75	0,5	0,889	0,947	0,849
ANN	0,444	0,6	0,615	0,75	0,605

Továbbá, a visszahívás (recall) mutatóra a DF (0,85) és a SVM adja a legnagyobb értéket (0,85), majd ezt követően a K-NN (0,75), DT (0,7) és ANN (0,6). Figyelembe véve az F1 értéket, a legjobb predikciós modell a hatásfok paraméterre nézve megadható. Tekintettel arra, hogy mind a három mutató esetén a véletlen erdő adja a legmagasabb értéket, ezért a hatásfok vizsgálata

esetén a véletlen erdő algoritmust választottam ki. A hatásfok villamos paraméterre nézve a véletlen erdő 13 fából áll és a modell pontossága 85%, ami magasabb a korábbi fejezetben bemutatott küszöbértéknél (>63%). Továbbá, a véletlen erdőnél, mely egy bagging-alapú modell, az Out-of-Bag, rövidítve OOB, pontosságot is érdemes figyelembe venni, mely egy specifikus metrika. Az OOB pontosság az összes fa átlagos teljesítményét mutatja meg azokon a mintákon, amelyeket tanítási fázisban a döntési fák nem használtak fel. A véletlen erdő esetén az OOB pontosság 72,2%.

5.2.2 Rövidzárási áramsűrűség

7. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján a rövidzárási áramsűrűség paraméterre nézve.

PRECIZITÁS					
	1	2	3	4	Átlag
DT	0,714	1	0,2	0,857	0,777
RF	0,4	0,833	0,75	0,8	0,749
K-NN	1	0,667	0,833	0,857	0,826
SVM	1	1	0,75	0,6	0,865
ANN	0,571	0,5	0,286	1	0,723
VISSZAHÍVÁS					
	1	2	3	4	Átlag
DT	0,833	0,25	0,5	0,75	0,65
RF	0,667	0,5	1	1	0,7
K-NN	0,5	1	0,833	0,75	0,8
SVM	0,833	0,4	1	1	0,8
ANN	0,667	0,333	0,444	0,615	0,605
F1					
	1	2	3	4	Átlag
DT	0,769	0,4	0,286	0,8	0,659
RF	0,5	0,625	0,857	0,889	0,694
K-NN	0,667	0,8	0,833	0,8	0,797
SVM	0,909	0,571	0,857	0,75	0,785
ANN	0,667	0,333	0,444	0,615	0,605

Az 7. táblázat szemlélteti a különböző algoritmusok teljesítményeinek összehasonlítását a precizitás (precision), visszahívás (recall) és F1 pontszám (F1 score) mérőszámok alapján a

rövidzárási áramsűrűség villamos paraméterre nézve. A táblázat alapján elmondható, hogy a precizitás mutató átlagára az SVM adta a legnagyobb értéket (0,865), majd a K-NN (0,826), ezt követi DT (0,777), RF (0,749) és ANN (0,723). A precizitást követően a visszahívás mutatót megvizsgáltam. A 6. táblázatból feltűnik, hogy a K-legközelebbi szomszéd és vektortartó-gép esetén kapjuk a legnagyobb értéket (0,8) a visszahívás átlag értékére, majd ezt követi a véletlen erdő 0,7 értékkel, döntési fa (0,659) és neurális hálózat (0,605). Továbbá, az F1 érték átlagai további segítséget nyújt a modell kiválasztása során. A legnagyobb érték a K-NN algoritmus esetén látható (0,797), majd SVM algoritmus (0,785), ezt követi RF (0,694), DT (0,659) és ANN (0,605). Figyelembe véve a precizitási, visszahívási és F1 paramétereket a legjobb modell a vektortartó-gép, SVM algoritmus. A vektortartó-gép esetén a modell 74 vektorból áll, melynél a súlyozás lineáris és a modell pontosság 80%, mely magasabb a meghatározott küszöbindexnél. A lineáris súlyozás azt jelenti, hogy az adatpontok közötti határt egy hipersík reprezentálja.

5.2.3 Üresjárási feszültség

A 8. táblázat szemlélteti a különböző algoritmusok teljesítményeinek összehasonlítását a precizitás (precision), visszahívás (recall) és F1 pontszám (F1 score) mérőszámok alapján az üresjárási feszültség paraméterre nézve.

8. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján az üresjárási feszültség paraméterre nézve.

PRECIZITÁS					
	1	2	3	4	Átlag
DT	0,667	1	0,8	0,333	0,763
RF	0,75	0,667	0,6	0,6	0,643
K-NN	0,5	0,6	0,75	0,571	0,625
SVM	1	0,667	0,375	1	0,856
ANN	0,638	1	0,5	0,4	0,66
VISSZAHÍVÁS					
	1	2	3	4	Átlag
DT	1	0,5	0,571	0,667	0,66
RF	0,75	1	0,5	0,5	0,65
K-NN	0,5	0,75	0,429	0,8	0,6
SVM	0,667	0,667	1	0,625	0,7
ANN	0,875	0,4	0,333	0,5	0,6

F1					
	1	2	3	4	Átlag
DT	0,8	0,667	0,667	0,444	0,66
RF	0,75	0,8	0,545	0,545	0,637
K-NN	0,5	0,667	0,545	0,667	0,591
SVM	0,8	0,667	0,545	0,769	0,73
ANN	0,737	0,571	0,4	0,444	0,586

A rövidzárási áram modellhez hasonlóan a precizitási, visszahívási és F1 paraméterek figyelembevétele mellett, a legjobb modell az SVM, mely esetén a pontosság érték 71%, a tartó vektorok száma 71 és a modellnél alkalmazott súlyozás polinomiális. A polinomiális súlyozás azt jelenti, hogy polinom függvénnyel történik az osztályozó funkció, ami magasabb rendű kapcsolatokat is képes megadni az adatokban.

5.2.4 Kitöltési tényező

A 9. táblázatban látható a DT, RF, K-NN, SVM, és ANN algoritmusok teljesítményeinek összehasonlítása a precizitás (precision), visszahívás (recall) és F1 pontszám (F1 score) mérőszámok alapján a kitöltési tényezőre nézve. A 9. táblázat alapján látható, hogy minden egyes mutató érték átlagát nézve, azaz mind a precizitás (0,819), mind a visszahívás (0,8) és az F1 pontszámra (0,798) a K-legközelebbi szomszéd algoritmus mutat maximális értéket. A K-NN modell euklideszi távolsággal került implementálásra, ahol a legközelebbi szomszédok értéke 10. Mindezek mellett a modell pontossága értéke 80%, ami meghaladja a dedikált küszöbértéket.

9. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján a kitöltési tényező paraméterre nézve.

PRECIZITÁS					
	1	2	3	4	Átlag
DT	0,714	1	0,5	0,625	0,716
RF	0,5	0,667	0,75	0,667	0,642
K-NN	0,8	0,714	1	0,8	0,819
SVM	0,571	0,5	0,833	1	0,776
ANN	0,333	1	0,625	0,75	0,7

VISSZAHÍVÁS					
	1	2	3	4	Átlag
DT	1	0,6	0,25	0,833	0,7
RF	1	0,286	0,75	0,5	0,6
K-NN	0,667	1	0,75	0,8	0,8
SVM	0,8	0,333	1	0,714	0,75
ANN	0,333	0,25	0,833	0,857	0,65
F1					
	1	2	3	4	Átlag
DT	0,833	0,75	0,333	0,714	0,677
RF	0,667	0,4	0,75	0,571	0,571
K-NN	0,727	0,833	0,857	0,8	0,798
SVM	0,667	0,4	0,909	0,833	0,746
ANN	0,333	0,4	0,714	0,8	0,624

Elmondható, hogy *a hatásfok esetén a véletlen erdő, az üresjárási feszültség és áramsűrűség esetén a SVM és a kitöltési tényező esetén a K-NN modell alkalmazásával értem el a legpontosabb predikciót.*

5.3 Tudományos munkám elválasztása más tudományos eredményektől

Kutatásom alapját Sutar és kollégáinak előzetes tanulmánya képezi, melyet kiterjesztettem további bemeneti és kimeneti paraméterekkel, nem korlátozva a vizsgálatomat kizárólag a cink-oxid alapú festékképzékenyített napelemre. Ugyanis a választott független változóim nem csak a DSSC hatásfokát foglalja magába, hanem további villamos paramétereket úgy, mint üresjárási feszültség, rövidzárási áramsűrűség és kitöltési tényező. Ugyancsak a bemeneti paraméter esetén a cink-oxid mellett, a titán-dioxid alapú rendszereket is megvizsgáltam. A tudományos publikációk között, ahogyan tudományos munkám bevezetőjében is részletesen kifejtettem, számos gépi tanulási algoritmus alkalmazásán alapuló publikáció született az elmúlt időben, de az áttanulmányozott irodalmak alapján ilyen típusú tudományos munka nem született. Sutar és munkatársainak tudományos publikációja jó alapot biztosított a DSSC gyártástechnológiájának optimalizációjára nézve, azonban kutatásommal új dimenziókat nyitottam meg ezen a területen és előrelépést értem el a festékképzékenyített napelemek gyártástechnológiájának területén, ami hozzájárul a hatékonyabb fejlesztéshez.

6 Kitekintés

Munkámra alapozva az irodalomkutatást tovább folytatom és a szisztematikus irodalomkutatást követően a korábbi adathalmazhoz képest egy nagyobb terjedelmű, több rekordú adatbázist fejleszték. A nagyobb méretű adathalmaznak köszönhetően a modellek pontosabb eredményeket produkálnak. Valamint, az adatoktól függően az adatbázist szintetikus adatokkal bővítem ki, ahol az eredeti adatok módosítása történik $\pm 1\%$ és $\pm 2\%$ -os változtatások mellett. Mindezek mellett, az adatértékelést python programozási környezetben tervezem végrehajtani. A gépi tanulási algoritmusok mellett, a statisztikus vizsgálatok lefuttatása is jövőbeli céljaim közé tartozik, ugyanis statisztikai vizsgálatból fontos számos információ vonható le a modell futtatása előtt.

Mindezek mellett, az elvárt eredmények szerint az adatok elemzését követően munkámat tervezetten IEEE nemzetközi konferencia publikációban és folyóirat publikációban tervezem közzéadni és bemutatni.

7 Összefoglaló

Az áttanulmányozott irodalmak alapján *célomul tűztem ki, hogy kísérleti adataimat feldolgozom és a cella teljesítményét a hőmérséklet függvényében statisztikai eljárás alkalmazásával értékelem. További célom még, hogy átfogó irodalomkutatást hajtok végre, melynek célja, hogy a releváns eredményeket szisztematikusan összegyűjtsen. Ezen adatgyűjtési folyamat eredményeképpen adatbázist fejlesztek, majd a létrehozott adatbázist felhasználva gépi tanulási modellt implementálok, hogy predikciót végezzek a festékkérszenyített napelem cella villamos paramétereinek meghatározására.*

Tudományos kérdésfeltevések:

- (1) A kísérleti adathalmazon végzett statisztikai vizsgálatokból milyen következtetések vonhatók le a festékkérszenyített napelem cella villamos teljesítményére nézve?
- (2) Milyen pontosságú modellt lehet interpretálni a festékkérszenyített napelem cellára gyártástechnológiáját figyelembe véve?

Tudományos munkám eredményeimet összefoglalva az alábbiakban mutatom be:

1. eredmény Olyan MatLab rutint fejlesztettem, mely a [73] publikációban leírt időfüggő mérési eljárást felhasználva a kísérletekből származó időfüggő $U_s(t)$ és $U_v(t)$ adatpontok beolvasását követően, az idő eliminálásával és az Ohm-törvényt felhasználva meghatározza a nyers villamos áram és villamos feszültség koordinátákat, majd ezen meghatározott koordinátákat simítja LOESS algoritmussal és a simított adatpontokra a napelem áram-feszültség karakterisztika függvényének görbét illeszti, melyből az $I - U$ jelleggörbe folytonos változását és a kívánt I_0 , U_0 és MPP megkapom és mindezeket a paramétereket a rutin táblázatba kiírja.

2. eredmény Kísérletet terveztem meg és hajtottam végre a napelem karakterisztika nyers felvételeire. $N=10$ mérési mintás korrelációs stratégiát követően a mérési adathalmaz kiértékelésén további adatelemzést végeztem varianciaanalízissel. A ferdeségi és csúcsossági mutatók alapján eredmény kaptam, hogy az adatok nem követik a normális eloszlást, azaz a normalitás sérül. A szóráshomogenitás eredményei alapján ($F(9,90) = 0,811, p = 0,607$) a nullhipotézis fenntartásra került 0,05 szignifikancia szinten. Más

szavakkal, nincs elegendő statisztikai bizonyíték arra nézve, hogy jelentős különbségek lennének az egyes csoportok varianciái között, melyből arra következtethetünk, hogy az ANOVA vizsgálat eredményei nem lesznek szignifikánsan torzítottak. Figyelembe véve a normalitás sérülését és a szóráshomogenitás konzisztenciáját, varianciaanalízisnek Kruskal-Wallis próbát alkalmaztam. Eredményül megkaptam, hogy statisztikailag különbség van a hőmérsékleti csoportok villamos teljesítményében, mivel $H(9) = 31,759, p < 0.001$. Más szavakkal, a nullhipotézist, (H_0)-t, elutasítottam és az ellenhipotézis, (H_1), teljesülését vettem figyelembe, azaz található legalább egy hőmérsékleti csoport, amely esetén a villamos teljesítmény értékei szignifikánsan eltérnek a többi csoportétól. A nemparaméteres vizsgálatot követően a csoportok közötti különbségek pontosabb megismerése érdekében Bonferroni utótesztet alkalmaztam. *Az eredmények alapján a magasabb hőmérsékleti csoportokban, 33 °C-tól, az adatok eltérő tendenciát mutatnak.*

3. eredmény A szisztematikus irodalomkutatásból adatbázist fejlesztettem, majd az adathalmazt alapos előfeldolgozásnak vettem alá, mely magába foglalta az adattisztítást és normalizálást. A bemeneti paramétereim: félvezető típusa; félvezető réteg vastagsága; morfológiai szerkezet; prekursor; fényérzékeny festékanyag típusa; szintézis hőmérséklete; szintézis ideje. Az adatfeldolgozást követően különböző klasszifikáló algoritmusok – úgy, mint döntési fa (DF), véletlen erdő (RF), K-legközelebbi szomszédok (K-NN), vektortartó-gép (SVM) és mesterséges neurális hálózat (ANN) – tesztelését végeztem a JASP statisztikai szoftver mellett. A különböző algoritmusok teljesítményeinek összehasonlítását a precizitás (precision), visszahívás (recall) és F1 pontszám (F1 score) mérőszámok alapján végeztem. Elmondható, hogy *a hatásfok esetén a véletlen erdő (13 fából álló modell, mely pontossága 85%), az üresjárási feszültség (71 darab tartóvektorból álló modell polinomiális súlyozást alkalmazva, melynek pontossága 71%) és áramsűrűség esetén a SVM (74 darab tartóvektorból álló modell lineáris súlyozást alkalmazva, melynek pontossága 80%) és a kitöltési tényező esetén a K-NN modell ($k=10$, modell pontossága 80%) alkalmazásával értem el a*

legpontosabb predikciót. Mindezek mellett, Sutar és munkatársainak tudományos publikációja jó alapot biztosított a DSSC gyártástechnológiájának optimalizációjára nézve, azonban kutatásommal új dimenziókat nyitottam meg ezen a területen és előrelépést értem el a festékérzékenyített napelemek gyártástechnológiájának területén, ami hozzájárul a hatékonyabb fejlesztéshez

8 Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani Dr. habil. Rácz Ervin tanár úrnak, hogy jelen TDK dolgozatom témavezetését elvállalta, segítette a munkámat. Köszönöm a Tanár Úrnak, hogy igazi mentorként áll mellettem. Köszönöm Rado Lászlónak és Erős Dánielnek a segítségüket és javaslataikat. A fentiekén kívül köszönöm mindazoknak, akik tudományos munkámat támogatták, segítették, korrigálták.

9 Irodalomjegyzék

- [1] J. Gong, K. Sumathy, Q. Qiao, and Z. Zhou, ‘Review on dye-sensitized solar cells (DSSCs): Advanced techniques and research trends’, *Renewable and Sustainable Energy Reviews*, vol. 68, pp. 234–246, Feb. 2017, doi: 10.1016/j.rser.2016.09.097.
- [2] Z. Varga and E. Racz, ‘Machine Learning Analysis on the Performance of Dye-Sensitized Solar Cell—Thermoelectric Generator Hybrid System’, *Energies*, vol. 15, no. 19, p. 7222, Oct. 2022, doi: 10.3390/en15197222.
- [3] Z. Varga and E. Racz, ‘Application Development with Finite Element Method to Calculate Photogeneration Rate and Open-Circuit Voltage of Dye Sensitized Solar Cell’, *Syst. Theor. Control Comput. J.*, vol. 2, no. 2, pp. 17–24, Dec. 2022, doi: 10.52846/stccj.2022.2.2.39.
- [4] Z. Varga and E. Racz, ‘COMSOL 3D Simulation of Heat Distribution in a Simplified Dye-Sensitized Solar Cell - Thermoelectric Generator Hybrid System’, 2022.
- [5] Z. Varga and E. Racz, ‘A Numerical Model and a Code Development for Photogeneration Rate Calculation for a Dye Sensitized Solar Cell’, in *2022 IEEE 16th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, Timisoara, Romania: IEEE, May 2022, pp. 000341–000346. doi: 10.1109/SACI55618.2022.9919499.
- [6] P. Raccuglia *et al.*, ‘Machine-learning-assisted materials discovery using failed experiments’, *Nature*, vol. 533, no. 7601, pp. 73–76, May 2016, doi: 10.1038/nature17439.
- [7] Y. Wen, L. Fu, G. Li, J. Ma, and H. Ma, ‘Accelerated Discovery of Potential Organic Dyes for Dye-Sensitized Solar Cells by Interpretable Machine Learning Models and Virtual Screening’, *Sol. RRL*, vol. 4, no. 6, p. 2000110, Jun. 2020, doi: 10.1002/solr.202000110.
- [8] H. A. Maddah, ‘Machine learning analysis on performance of naturally-sensitized solar cells’, *Optical Materials*, vol. 128, p. 112343, Jun. 2022, doi: 10.1016/j.optmat.2022.112343.
- [9] E. Martinot, A. Chaurey, D. Lew, J. R. Moreira, and N. Wamukonya, ‘Renewable Energy Markets in Developing Countries’, *Annu. Rev. Energy. Environ.*, vol. 27, no. 1, pp. 309–348, Nov. 2002, doi: 10.1146/annurev.energy.27.122001.083444.
- [10] U. Blieske *et al.*, ‘Current and future trends in photovoltaic technology’, in *2019 International Energy and Sustainability Conference (IESC)*, Farmingdale, NY, USA: IEEE, Oct. 2019, pp. 1–5. doi: 10.1109/IESC47067.2019.8976871.
- [11] X. Li, H. Liao, Y.-F. Du, C. Wang, J.-W. Wang, and Y. Liu, ‘Carbon dioxide emissions from the electricity sector in major countries: a decomposition analysis’, *Environ Sci Pollut Res*, vol. 25, no. 7, pp. 6814–6825, Mar. 2018, doi: 10.1007/s11356-017-1013-z.
- [12] Zh. I. Alferov, V. M. Andreev, and V. D. Rumyantsev, ‘Solar photovoltaics: Trends and prospects’, *Semiconductors*, vol. 38, no. 8, pp. 899–908, Aug. 2004, doi: 10.1134/1.1787110.
- [13] M. Malinowski, J. I. Leon, and H. Abu-Rub, ‘Solar Photovoltaic and Thermal Energy Systems: Current Technology and Future Trends’, vol. 105, no. 11, pp. 2132–2146, Nov. 2017, doi: 10.1109/JPROC.2017.2690343.
- [14] U. K. Das *et al.*, ‘Forecasting of photovoltaic power generation and model optimization: A review’, *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, Jan. 2018, doi: 10.1016/j.rser.2017.08.017.
- [15] J. Li *et al.*, ‘A Brief Review of High Efficiency III-V Solar Cells for Space Application’, *Front. Phys.*, vol. 8, p. 631925, Feb. 2021, doi: 10.3389/fphy.2020.631925.
- [16] R. Verduci *et al.*, ‘Solar Energy in Space Applications: Review and Technology Perspectives’, *Advanced Energy Materials*, vol. 12, no. 29, p. 2200125, Aug. 2022, doi: 10.1002/aenm.202200125.
- [17] M. V. Dambhare, B. Butey, and S. V. Moharil, ‘Solar photovoltaic technology: A review

- of different types of solar cells and its future trends', *J. Phys.: Conf. Ser.*, vol. 1913, no. 1, p. 012053, May 2021, doi: 10.1088/1742-6596/1913/1/012053.
- [18] B. O'Regan and M. Grätzel, 'A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO₂ films', *Nature*, vol. 353, no. 6346, pp. 737–740, 0 1991, doi: 10.1038/353737a0.
- [19] N. Tomar, A. Agrawal, V. S. Dhaka, and P. K. Surolia, 'Ruthenium complexes based dye sensitized solar cells: Fundamentals and research trends', *Solar Energy*, vol. 207, pp. 59–76, Sep. 2020, doi: 10.1016/j.solener.2020.06.060.
- [20] R. Bashir, A. R. Makhdoom, M. K. Bilal, and M. Ahmad Badar, 'Comparative study of the photovoltaic behavior of ruthenium and the other organic and inorganic Dye-Sensitized Solar Cells (DSSC)', *Optik*, vol. 157, pp. 11–15, Mar. 2018, doi: 10.1016/j.ijleo.2017.11.064.
- [21] M. Grätzel, 'Photoelectrochemical cells', vol. 414, p. 7, 2001.
- [22] F.-T. Kong, S.-Y. Dai, and K.-J. Wang, 'Review of Recent Progress in Dye-Sensitized Solar Cells', *Advances in OptoElectronics*, vol. 2007, pp. 1–13, Aug. 2007, doi: 10.1155/2007/75384.
- [23] M. R. Narayan, 'Review: Dye sensitized solar cells based on natural photosensitizers', *Renewable and Sustainable Energy Reviews*, p. S1364032111003959, Sep. 2011, doi: 10.1016/j.rser.2011.07.148.
- [24] J. Maçaira, L. Andrade, and A. Mendes, 'Modeling, simulation and design of dye sensitized solar cells', *RSC Adv.*, vol. 4, no. 6, pp. 2830–2844, 2014, doi: 10.1039/C3RA46295A.
- [25] Z. Varga and E. Rácz, 'A szilícium alapú napelem cella összehasonlítása a festékkérszenyített napelem cellával', in *XXXVI. Kandó Konferencia 2020*, Budapest, pp. 54–65.
- [26] H. M. Upadhyaya, S. Senthilarasu, M.-H. Hsu, and D. K. Kumar, 'Recent progress and the status of dye-sensitized solar cell (DSSC) technology with state-of-the-art conversion efficiencies', *Solar Energy Materials and Solar Cells*, vol. 119, pp. 291–295, Dec. 2013, doi: 10.1016/j.solmat.2013.08.031.
- [27] A. Lopes, A. Araújo, A. Mendes, and L. Andrade, 'A dye-sensitized solar cell model implementable in electrical circuit simulators', *Solar Energy*, vol. 122, pp. 169–180, Dec. 2015, doi: 10.1016/j.solener.2015.08.021.
- [28] T. Bak, M. K. Nowotny, L. R. Sheppard, and J. Nowotny, 'Mobility of Electronic Charge Carriers in Titanium Dioxide', *J. Phys. Chem. C*, vol. 112, no. 33, pp. 12981–12987, Aug. 2008, doi: 10.1021/jp801028j.
- [29] M. Grätzel, 'Photoelectrochemical cells', *Nature*, vol. 414, no. 6861, pp. 338–344, 0 2001, doi: 10.1038/35104607.
- [30] B. Onwona-Agyeman *et al.*, 'Sensitization of Nanocrystalline SnO₂ Films with Indoline Dyes', *Jpn. J. Appl. Phys.*, vol. 44, no. 5L, p. L731, May 2005, doi: 10.1143/JJAP.44.L731.
- [31] P. Guo and M. A. Aegerter, 'RU(II) sensitized Nb₂O₅ solar cell made by the sol-gel process', *Thin Solid Films*, 1999.
- [32] Y. Ohsaki *et al.*, 'Dye-sensitized TiO₂ nanotube solar cells: fabrication and electronic characterization', *Phys. Chem. Chem. Phys.*, vol. 7, no. 24, p. 4157, 2005, doi: 10.1039/b511016e.
- [33] J. Jiu, S. Isoda, F. Wang, and M. Adachi, 'Dye-Sensitized Solar Cells Based on a Single-Crystalline TiO₂ Nanorod Film', *J. Phys. Chem. B*, vol. 110, no. 5, pp. 2087–2092, Feb. 2006, doi: 10.1021/jp055824n.
- [34] M. Law, L. E. Greene, J. C. Johnson, R. Saykally, and P. Yang, 'Nanowire dye-sensitized solar cells', *Nature Mater.*, vol. 4, no. 6, pp. 455–459, Jun. 2005, doi:

10.1038/nmat1387.

- [35] S. H. Ko *et al.*, 'Nanoforest of Hydrothermally Grown Hierarchical ZnO Nanowires for a High Efficiency Dye-Sensitized Solar Cell', *Nano Lett.*, vol. 11, no. 2, pp. 666–671, Feb. 2011, doi: 10.1021/nl1037962.
- [36] Md. F. Hossain and T. Takahashi, 'Hydrothermal Synthesis of Novel ZnO Nanomushrooms for Improving the Solar Cells Performance', *IEEE Trans. Nanotechnology*, vol. 13, no. 4, pp. 755–759, Jul. 2014, doi: 10.1109/TNANO.2014.2319097.
- [37] Md. K. Nazeeruddin, E. Baranoff, and M. Grätzel, 'Dye-sensitized solar cells: A brief overview', *Solar Energy*, vol. 85, no. 6, pp. 1172–1178, Jun. 2011, doi: 10.1016/j.solener.2011.01.018.
- [38] L.-S. Shi, K.-C. Lee, M.-S. He, and H. H. Yu, 'Plant Dyes as Photosensitizer for a Dye-Sensitized Solar Cell', *J Biobased Mat Bioenergy*, vol. 7, no. 4, pp. 430–438, Aug. 2013, doi: 10.1166/jbmb.2013.1341.
- [39] G. Hashmi *et al.*, 'Review of materials and manufacturing options for large area flexible dye solar cells', *Renewable and Sustainable Energy Reviews*, vol. 15, no. 8, pp. 3717–3732, Oct. 2011, doi: 10.1016/j.rser.2011.06.004.
- [40] A. Parisi, R. Pernice, A. Andò, A. C. Cino, V. Franzitta, and A. C. Busacca, 'Electro-optical characterization of ruthenium-based dye sensitized solar cells: A study of light soaking, ageing and temperature effects', *Optik*, vol. 135, pp. 227–237, Apr. 2017, doi: 10.1016/j.ijleo.2017.01.100.
- [41] M. Hosseinneshad *et al.*, 'Dye-sensitized solar cells based on natural photosensitizers: A green view from Iran', *Journal of Alloys and Compounds*, vol. 828, p. 154329, Jul. 2020, doi: 10.1016/j.jallcom.2020.154329.
- [42] M. Grätzel, 'Solar Energy Conversion by Dye-Sensitized Photovoltaic Cells', *Inorg. Chem.*, vol. 44, no. 20, pp. 6841–6851, Oct. 2005, doi: 10.1021/ic0508371.
- [43] M. Green, E. Dunlop, J. Hohl-Ebinger, M. Yoshita, N. Kopidakis, and X. Hao, 'Solar cell efficiency tables (version 57)', *Prog Photovolt Res Appl*, vol. 29, no. 1, pp. 3–15, Jan. 2021, doi: 10.1002/pip.3371.
- [44] G. Boschloo and A. Hagfeldt, 'Characteristics of the Iodide/Triiodide Redox Mediator in Dye-Sensitized Solar Cells', *Acc. Chem. Res.*, vol. 42, no. 11, pp. 1819–1826, Nov. 2009, doi: 10.1021/ar900138m.
- [45] I. Mora-Seró and J. Bisquert, 'Fermi Level of Surface States in TiO₂ Nanoparticles', *Nano Lett.*, vol. 3, no. 7, pp. 945–949, Jul. 2003, doi: 10.1021/nl0342390.
- [46] E. Racz and Z. Varga, 'Investigation of Electrical Parameters on a DSSC Solar Cell in Function of the Cell Temperature', in *2019 International IEEE Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE)*, Budapest, Hungary: IEEE, Nov. 2019, pp. 91–96. doi: 10.1109/CANDO-EPE47959.2019.9110998.
- [47] E. Racz and Z. Varga, 'Investigation of the Maximum Power Point Position on a DSSC Solar Cell Using Different Irradiations', in *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Herlany, Slovakia: IEEE, Jan. 2020, pp. 83–88. doi: 10.1109/SAMI48414.2020.9108726.
- [48] E. Racz and Z. Varga, 'Investigation of the Fill Factors for a DSSC Solar Cell on the Basis of Module Temperature Using Different Artificial Irradiations', in *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, Timisoara, Romania: IEEE, May 2020, pp. 000159–000164. doi: 10.1109/SACI49304.2020.9118828.
- [49] E. Racz and Z. Varga, 'Investigation of the Maximum Power Point on a DSSC Solar Cell based on the Incoming Light Irradiation and Temperature', in *2020 IEEE 3rd International*

Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), Budapest, Hungary: IEEE, Nov. 2020, pp. 000161–000166. doi: 10.1109/CANDO-EPE51100.2020.9337763.

[50] Z. Varga and E. Racz, 'Influence of the Cell Temperature on the Performance of a Dye Sensitized Solar Cell', in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Herl'any, Slovakia: IEEE, Jan. 2021, pp. 000175–000180. doi: 10.1109/SAMI50585.2021.9378631.

[51] Z. Varga and E. Racz, 'Influence of the Temperature on the Open-Circuit Voltage and Short-Circuit Current of a yellow colorized Dye Sensitized Solar Cell using Correlation Approach', in *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, Timisoara, Romania: IEEE, May 2021, pp. 000253–000258. doi: 10.1109/SACI51354.2021.9465600.

[52] C.-C. Ting and W.-S. Chao, 'Measuring temperature dependence of photoelectric conversion efficiency with dye-sensitized solar cells', *Measurement*, vol. 43, no. 10, pp. 1623–1627, Dec. 2010, doi: 10.1016/j.measurement.2010.09.025.

[53] H. J. Snaith, L. Schmidt-Mende, M. Grätzel, and M. Chiesa, 'Light intensity, temperature, and thickness dependence of the open-circuit voltage in solid-state dye-sensitized solar cells', *Phys. Rev. B*, vol. 74, no. 4, p. 045306, Jul. 2006, doi: 10.1103/PhysRevB.74.045306.

[54] A. Usami, S. Seki, Y. Mita, H. Kobayashi, H. Miyashiro, and N. Terada, 'Temperature dependence of open-circuit voltage in dye-sensitized solar cells', *Solar Energy Materials and Solar Cells*, vol. 93, no. 6–7, pp. 840–842, Jun. 2009, doi: 10.1016/j.solmat.2008.09.040.

[55] S. Lepikko, K. Miettinen, A. Poskela, A. Tiihonen, and P. D. Lund, 'Testing dye-sensitized solar cells in harsh northern outdoor conditions', *Energy Sci Eng*, vol. 6, no. 3, pp. 187–200, Jun. 2018, doi: 10.1002/ese3.195.

[56] C. Cornaro *et al.*, 'Comparative analysis of the outdoor performance of a dye solar cell mini-panel for building integrated photovoltaics applications', *Progress in Photovoltaics*, vol. 23, no. 2, pp. 215–225, Feb. 2015, doi: 10.1002/pip.2426.

[57] S. Dai *et al.*, 'The design and outdoor application of dye-sensitized solar cells', *Inorganica Chimica Acta*, vol. 361, no. 3, pp. 786–791, Feb. 2008, doi: 10.1016/j.ica.2007.04.018.

[58] M. Du, 'Band-Gap Engineering of Metal Oxides for Dye-Sensitized Solar Cells', p. 4.

[59] Y. Saito, S. Kambe, T. Kitamura, Y. Wada, and S. Yanagida, 'Morphology control of mesoporous TiO₂ nanocrystalline films for performance of dye-sensitized solar cells', *Solar Energy Materials and Solar Cells*, vol. 83, no. 1, pp. 1–13, Jun. 2004, doi: 10.1016/j.solmat.2004.02.010.

[60] Z. Kebede and S.-E. Lindquist, 'Donor—acceptor interaction between annodnitasqiumepoluicsastoiolvneinntsdaynedseIn stitoizgeednseoralatre cIe\ l,ls', *Solar Energy Materials*, p. 17, 1999.

[61] A. Ghosh, P. Selvaraj, S. Sundaram, and T. K. Mallick, 'The colour rendering index and correlated colour temperature of dye-sensitized solar cell for adaptive glazing application', *Solar Energy*, vol. 163, pp. 537–544, Mar. 2018, doi: 10.1016/j.solener.2018.02.021.

[62] M. A.M. Al-Alwani, H. Abu Hasan, N. Kaid Nasser Al-Shorgani, and A. B. S.A. Al-Mashaan, 'Natural dye extracted from Areca catechu fruits as a new sensitizer for dye-sensitized solar cell fabrication: Optimisation using D-Optimal design', *Materials Chemistry and Physics*, vol. 240, p. 122204, Jan. 2020, doi: 10.1016/j.matchemphys.2019.122204.

[63] J. Leyrer, M. Rubilar, E. Morales, B. Pavez, E. Leal, and R. Hunter, 'Factor Optimization in the Manufacturing Process of Dye-Sensitized Solar Cells Based on Naturally

Extracted Dye from a Maqui and Blackberry Mixture (Aristotelia Chilensis and Rubus Glaucus)', *Journal of Elec Materi*, vol. 47, no. 10, pp. 6136–6143, Oct. 2018, doi: 10.1007/s11664-018-6514-0.

[64] S. Abdollahi, M. Ehsani, J. Morshedian, H. A. Khonakdar, and E. Aram, 'Application of response surface methodology in assessing the effect of electrospinning parameters on the morphology of polyethylene oxide/polyacrylonitrile blend nanofibers containing graphene oxide', *Polym. Bull.*, vol. 76, no. 4, pp. 1755–1773, Apr. 2019, doi: 10.1007/s00289-018-2448-1.

[65] M. A. M. Al-Alwani, 'Optimization of dye extraction from Cordyline fruticosa via response surface methodology to produce a natural sensitizer for dye-sensitized solar cells', p. 10, 2016.

[66] N. Patni and S. G. Pillai, 'Optimization of the doping of polyaniline via response surface method to prepare polymer electrolytes for dye sensitized solar cells', *Journal of Environmental Chemical Engineering*, vol. 8, no. 2, p. 103709, Apr. 2020, doi: 10.1016/j.jece.2020.103709.

[67] S. N. F. Zainudin, M. Markom, H. Abdullah, R. Adami, and S. M. Tasirin, 'Optimized synthesis and photovoltaic performance of TiO₂ nanoparticles for dye-sensitized solar cell', *Particuology*, vol. 11, no. 6, pp. 753–759, Dec. 2013, doi: 10.1016/j.partic.2012.07.010.

[68] S. S. Sutar, S. M. Patil, S. J. Kadam, R. K. Kamat, D. Kim, and T. D. Dongale, 'Analysis and Prediction of Hydrothermally Synthesized ZnO-Based Dye-Sensitized Solar Cell Properties Using Statistical and Machine-Learning Techniques', *ACS Omega*, vol. 6, no. 44, pp. 29982–29992, Nov. 2021, doi: 10.1021/acsomega.1c04521.

[69] J. Xu, H. Zhang, L. Wang, G. Liang, L. Wang, and X. Shen, 'Artificial neural network-based QSPR study on absorption maxima of organic dyes for dye-sensitised solar cells', *Molecular Simulation*, vol. 37, no. 1, pp. 1–10, Jan. 2011, doi: 10.1080/08927022.2010.506513.

[70] 'Racz és Varga - 2019 - Investigation of Electrical Parameters on a DSSC S.pdf'.

[71] Z. Varga and E. Racz, 'Statistical Analysis of the Performance of the Dye-Sensitized Solar Cell – Thermoelectric Generator', in *2023 IEEE 21st World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Herl'any, Slovakia: IEEE, Jan. 2023, pp. 000267–000272. doi: 10.1109/SAMI58000.2023.10044524.

[72] Z. Varga and E. Racz, 'Sun Simulator for Laboratory Testing', in *2022 IEEE 5th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE)*, Budapest, Hungary: IEEE, Nov. 2022, pp. 000107–000112. doi: 10.1109/CANDO-EPE57516.2022.10046374.

[73] A. Varga, E. Racz, and P. Kadar, 'New experimental method for measuring power characteristics of photovoltaic cells at given light irradiation', in *2013 IEEE 8th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, Timisoara, Romania: IEEE, May 2013, pp. 405–409. doi: 10.1109/SACI.2013.6609008.

[74] A. Phinikarides, G. Makrides, B. Zinsser, M. Schubert, and G. E. Georghiou, 'Analysis of photovoltaic system performance time series: Seasonality and performance loss', *Renewable Energy*, vol. 77, pp. 51–63, May 2015, doi: 10.1016/j.renene.2014.11.091.

[75] G. Marshall and L. Jonker, 'An introduction to descriptive statistics: A review and practical guide', *Radiography*, vol. 16, no. 4, pp. e1–e7, Nov. 2010, doi: 10.1016/j.radi.2010.01.001.

[76] Szabolcs T., 'ÉRZÉKENYSÉGVIZSGÁLATOK A STATISZTIKAI ELJÁRÁSOKBAN', *Alkalmazott Matematikai Lapok*, 2012.

[77] B. G. Tabachnick and L. S. Fidell, *Computer-assisted research design and analysis*. Boston: Allyn and Bacon, 2001.

[78] B. Charbuty and A. Abdulazeez, 'Classification Based on Decision Tree Algorithm for

- Machine Learning’, *JASTT*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- [79] A. Khandakar *et al.*, ‘Machine Learning Based Photovoltaics (PV) Power Prediction Using Different Environmental Parameters of Qatar’, *Energies*, vol. 12, no. 14, p. 2782, Jul. 2019, doi: 10.3390/en12142782.
- [80] L. E. Ordoñez Palacios, V. Bucheli Guerrero, and H. Ordoñez, ‘Machine Learning for Solar Resource Assessment Using Satellite Images’, *Energies*, vol. 15, no. 11, p. 3985, May 2022, doi: 10.3390/en15113985.
- [81] U. Singh, M. Rizwan, M. Alaraj, and I. Alsaidan, ‘A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments’, *Energies*, vol. 14, no. 16, p. 5196, Aug. 2021, doi: 10.3390/en14165196.
- [82] H. A. Alghamdi, ‘A Time Series Forecasting of Global Horizontal Irradiance on Geographical Data of Najran Saudi Arabia’, *Energies*, vol. 15, no. 3, p. 928, Jan. 2022, doi: 10.3390/en15030928.
- [83] I. Fazekas, *Nurális hálózatok*, Debreceni Egyetem Informatikai Kar. Debrecen, 2013.
- [84] Ferenc D. B., ‘Adatbányászati algoritmusok’.
- [85] F. Li *et al.*, ‘Machine Learning (ML)-Assisted Design and Fabrication for Solar Cells’, *Energy Environ. Mater.*, vol. 2, no. 4, pp. 280–291, Dec. 2019, doi: 10.1002/eem2.12049.
- [86] M. S. Bakay and Ü. Ağbulut, ‘Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms’, *Journal of Cleaner Production*, vol. 285, p. 125324, Feb. 2021, doi: 10.1016/j.jclepro.2020.125324.

10 Ábrajegyzék

1. ábra. Energiadiagram a festékérzékenyített napelem cella (DSSC) működésének bemutatására az elektrontranszfer folyamatot is szemlélítve. A sötét kékkel jelölt rész az üveghordozót jelenti (glass felirat az ábrán); a szürke rész az átlátszó vezetőképes oxid réteget szemlélteti (Transparent Conductive Oxid, TCO); a világos kék félvezető réteg, bal oldalon jellemzően titán-dioxid (jelölve Porous TiO_2), míg az ábra jobb oldalán pedig platina (Pt); piros színnel a fényérzékeny festékanyag látható (ábrán jelölve, mint Sensitizing Dye), ahol a HOMO: Highest Occupied Molecular Orbital, azaz legmagasabb molekula orbitál, és a LUMO: Lowest Unoccupied Molecular Orbital, azaz legalacsonyabb nem foglalt molekula orbitál; és a világossárga rész pedig az elektrolit oldal. (Az ábra saját konferencia publikációból került másolásra) [5].....	6
2. ábra. Az 1993 és 2024 közötti időszakban regisztrált publikációk számának év szerinti lebontásban való oszlopdiagram szerű ábrázolása, különös tekintettel a statisztikai módszertan alkalmazásán alapuló és az ahhoz terminológiailag kapcsolódó kapcsolatos munkák.....	11
3. ábra. A napelem villamos áram-feszültség karakterisztikáját kiolvasó MatLab rutinom folyamatábrája, ahol az $y = a - b \cdot ec \cdot x - 1$ a napelem elektromos áram-feszültség karakterisztikájának egyszerűsített egyenlete.	18
4. ábra: A ferdeség és csúcsosság bemutatása, azaz a normális eloszláshoz képest az adatok hogyan változnak, ahol (a) a ferdeségi mutató (skewness, rövidítve: skew) asszimetriája ábrázolva a negatív és pozitív ferdeséget; (b) a csúcsossági mutató (kurtosis) ábrázolása pozitív és negatív esetekben. forrás: Dr. Mark A Goss-Sampson: JASP Guide book v0.16 (2022) ...	20
5. ábra. Véletlen erdő működési elve	24
6. ábra. Hipersíkkal történő elválasztásra példa, ahol a hipersík kétdimenziós egyenes. (a) az elválasztás nem optimális; (b) maximális margójú elválasztás, ahol vékony vonallal a margók és a két margó között elhelyezett vastag vonal a határsávot reprezentálja [83].....	25
7. ábra. Mesterséges neurális hálózat egy részlete: többrétegű előreccsatolt neurális hálózat [84].	26
8. ábra. Mesterséges neurális hálózatnál alkalmazott leggyakoribb transzferfüggvények.....	27
9. ábra. Tudományos munkám gépi tanulási alkalmazásának lépései folyamatábraként szemlélítve a modell fejlesztés fázisától a modell elfogadásáig, ahol DT: döntési fa; RF: véletlen erdő; K-NN: K-legközelebbi szomszéd; SVM: tartóvektor-gép; és ANN: mesterséges neurális hálózatot jelenti.....	28

10. ábra. Különböző paraméterek adatnormalizálását bemutató dobozdiagram, ahol (a) a hatásfok; (b) a szintézis hőmérséklete; (c) a szintézis ideje; (d) a rövidzárási áramsűrűség J_{sc} ; (e) a üresjárási feszültség (V_{oc}), (f) a kitöltési tényező FF 29
11. ábra: A festékérzékenyített napelem cella villamos teljesítményének vizsgálata a hőmérséklet függvényében, ahol (a) a sűrűségfüggvény grafikon különböző hőmérsékleti értékek esetén (b) boxplot ábra a maximális villamos teljesítményről a hőmérséklet függvényében. 32

11 Táblázatjegyzék

1. táblázat: A feldolgozott tudományos publikációk kategorizálása adatértékelési módszerek alapján úgy, mint a korreláció, regressziós és varianciaanalízis.	13
2. táblázat: Különböző mérőszámok összefoglalása leíró statisztikához, ahol az SE a standard hiba és az SD a szórás.	32
3. táblázat: Szóráshomogenitás vizsgálatára kapott eredmények, ahol F a statisztikai számítás eredménye, df1 a számláló szabadsági foka, df2 a nevező szabadsági foka és a p az eredmény statisztika szignifikanciája.....	33
4. táblázat: Kruskal-Wallis teszt eredményei a varianciaanalízis esetén, ahol N a minta nagysága, a Statistic a Kruskal-Wallis teszt statisztikája, a df a szabadsági fok és a p a valószínűségi szint.....	34
5. táblázat: DSSC-ket alkotó különböző jellemzők kategorikus összefoglalása feltüntetve a megfigyelések számával, ahol Ti-dipac a titán diizopropoxid bisz(acetilacetonát); Ti-butox a titán butoxid; TTIP a titán(IV)-izopropoxid; Ti-tetrabut a tetrabutyl-titanát; TiCl4 a titán tetraklorid	35
6. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján a határfok paraméterre nézve.	36
7. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján a rövidzárási áramsűrűség paraméterre nézve.	37
8. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján az üresjárási feszültség paraméterre nézve.	38
9. táblázat: A különböző algoritmusok (DT – döntési fa; RF – véletlen erdő; K-NN – K-legközelebbi szomszédok; SVM – vektortartó-gép; ANN – mesterséges neurális hálózat) teljesítményének összehasonlítása precizitás, visszahívás és F1 pontszám alapján a kitöltési tényező paraméterre nézve.	39

