



Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar

Mélytanulás a Melanóma Diagnosztikában

TDK dolgozat

Készítette:

Heizer Levente

Konzulens:

Békési Gergő Bendegúz

2024

Tartalomjegyzék

Kivonat	i
Abstract	ii
1. Bevezetés	1
2. Problémafelvetés	3
2.1. ISIC adathalmaz bemutatása	3
2.2. Az adathalmazban előforduló zavaró tényezők hatása	3
2.3. Módszerek zavaró tényezők eltávolítására	6
2.3.1. Klasszikus képfeldolgozási módszerek	6
2.3.2. Neurális háló alapú módszerek	7
3. Kapcsolódó munkák	9
3.1. Generatív Versengő Hálózatok	9
3.2. A GAN-ok alkalmazási területei	11
4. Módszertan	12
4.1. Adathalmaz létrehozása	12
4.1.1. Szörszálak szimulációja	12
4.1.2. Vonalzó szimulációja	13
4.1.3. Tintafoltok szimulációja	14
4.1.4. Dermatoszkóp árnyékának és fényének szimulációja	14
4.1.5. Keretek szimulációja	14
4.1.6. Minták generálása	15
4.2. Architektúra	16
4.2.1. Generatív modell	16
4.2.2. Generátor modell	16
4.2.3. Diszkriminátor modell	19
4.3. Tanítás	20
4.3.1. Veszteségfüggvények	20
4.3.2. Tanítás folyamata	22
4.4. Kiértékelés	24
4.4.1. Kiértékelési stratégia	24
4.4.1.1. Kvantitatív kiértékelés	24
4.4.1.2. Kvalitatív kiértékelés	27
4.4.2. Kiértékelési terv	27
5. Eredmények	29
5.1. Kvantitatív kiértékelés	29
5.2. Kvalitatív kiértékelés	31

5.3. Magyarázhatóság kiértékelése	33
6. Összefoglaló	35
Irodalomjegyzék	37
Függelék	41
F.1. Adathalmaz	41
F.2. A GAN tanítása	42
F.3. Kiértékelés eredmények	44

Kivonat

A dermatoszkópos képeken gyakran előforduló zavaró tényezők, mint a vonalzó, tollal rajzolt jelzések, szőrszálak és tintafoltok, jelentősen csökkentik a neurális hálózat alapú osztályozók megbízhatóságát és magyarázhatóságát. A jelen dolgozatban bemutatok egy innovatív, mélytanulás alapú megközelítést, amely egy generatív versengő hálózat (GAN) segítségével hatékonyan eltávolítja ezeket a zavaró tényezőket anélkül, hogy a bőrelváltozások diagnosztikailag releváns jellemzőit befolyásolná. Ehhez egy 10.676 képpárt tartalmazó szintetikus adatbázist hoztam létre, amely lehetővé tette a GAN felügyelt tanítását. A generátorhálózat U-Net architektúrát alkalmaz, figyelmi mechanizmusokkal és egy összetett veszteségfüggvénnyel (L1, adversariális és perceptuális veszteség), biztosítva a pixel-szintű pontosságot és a magasabb szintű jellemzők megőrzését.

A módszer hatékonyságát kvantitatív és kvalitatív értékelési módszerekkel vizsgáltam. Kvantitatív kiértékelésem során olyan metrikákat alkalmaztam, mint a Fréchet Inception Distance (FID; 17,45), a Structural Similarity Index Measure (SSIM; 0,976), a Peak Signal-to-Noise Ratio (PSNR; 37,5) és a Learned Perceptual Image Patch Similarity (LPIPS; 0,018), amelyek eredményei kimutatták, hogy a fejlesztett módszer jobb teljesítményt nyújt a jelenlegi megoldásokhoz képest, különösen a szőrszálak eltávolításánál. Továbbá, a zavaró tényezők eltávolításának hatását olyan klasszifikációs modellek teljesítményén mértem, mint a ResNet, EfficientNet és más modellek. Az eltávolításra a ResNet18 modell reagált a legjobban, ami minden vizsgált metrikában jobban tudta elkülöníteni a jó- és rosszindulatú elváltozásokat az AUC értékben mintegy 4,8%-os fejlődést mutatva. A kvalitatív értékelés során kilenc bőrgyógyász bevonásával összehasonlítottam a rekonstrukciókat a jelenlegi legjobb módszerekkel szemben. Az eredmények statisztikailag szignifikáns előnyöket mutattak: 99,99% és 98,21% konfidenciával állítható, hogy a szakértők a dolgozat keretében bemutatott módszert preferálják a referencia módszerekkel szemben. A szakértői tesztelés során kiderült, hogy a módszer magas minőségben megőrzi a bőrelváltozások szimmetriáját, pigmentációját és széleit, kevés torzítás és zaj keletkezése mellett. Ezek pedig kritikus képi jellemzők a helyes diagnózis meghozásakor.

Összességében a dolgozat egy újszerű, mélytanulás-alapú megközelítést kínál a dermatoszkópos képek zavaró tényezőinek eltávolítására, amely nemcsak hatékonyabban működik, mint a jelenlegi megoldások, hanem jelentősen növeli a neurális hálózatok osztályozási megbízhatóságát és magyarázhatóságát. Ez a módszer elősegíti a bőrrák korai és pontos diagnosztizálását, ami életmentő lehet, és jelentős előrelépést képvisel a dermatoszkópos képalkotás területén mind tudományos, mind ipari szempontból.

Abstract

Artifacts frequently present in dermoscopic images, such as rulers, ink markings, hair strands significantly reduce the reliability and explainability of neural network-based classifiers. This thesis introduces an innovative deep learning-based approach that effectively removes these confounding factors using a Generative Adversarial Network (GAN) without affecting the diagnostically relevant features of skin lesions. To achieve this, I created a synthetic dataset comprising 10,676 image pairs, enabling the supervised training of the GAN. The generator network employs a U-Net architecture enhanced with attention mechanisms and a complex set of loss functions (L1, adversarial, and perceptual losses), ensuring pixel-level accuracy and the preservation of higher-level features.

The effectiveness of the method was evaluated using both quantitative and qualitative assessment techniques. In the quantitative evaluation, metrics such as Fréchet Inception Distance (FID; 17.45), Structural Similarity Index Measure (SSIM; 0.976), Peak Signal-to-Noise Ratio (PSNR; 37.5), and Learned Perceptual Image Patch Similarity (LPIPS; 0.018) were utilized. The results demonstrated that the developed method outperforms current solutions, particularly in the removal of hair strands. Additionally, the impact of removing confounding factors was assessed on the performance of classification models, including ResNet, EfficientNet, and others. The ResNet18 model responded best to the removal, showing approximately a 4.8% improvement in the Area Under the Curve (AUC) metric across all evaluated metrics, enhancing its ability to distinguish between benign and malignant lesions.

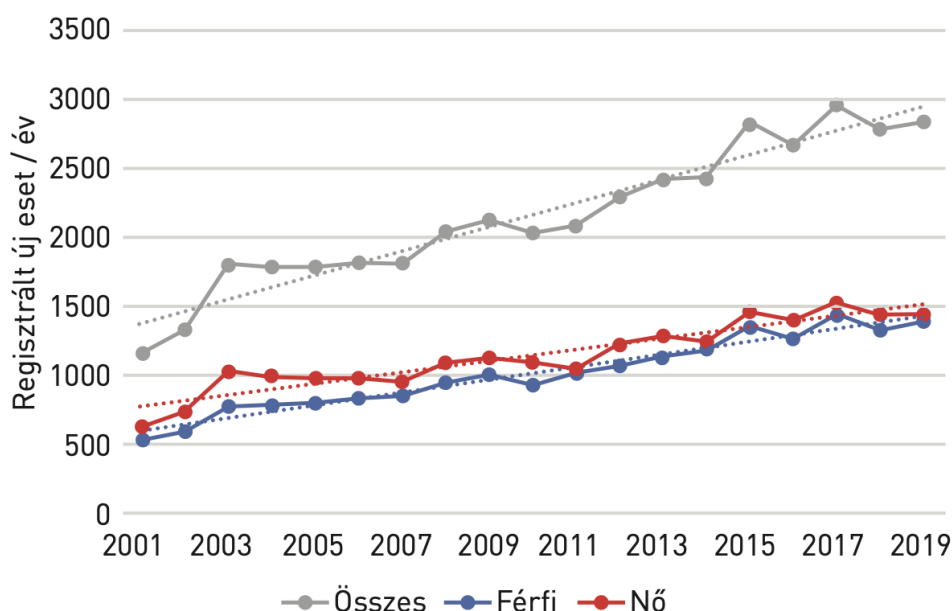
In the qualitative evaluation, reconstructions were compared against current state-of-the-art methods with the involvement of nine dermatology experts. The results revealed statistically significant advantages: with 99.99% and 98.21% confidence levels, experts preferred the method presented in this thesis over the reference methods. Expert testing indicated that the method preserves the symmetry, pigmentation, and edges of skin lesions with high quality, while introducing minimal distortion and noise, which are critical image characteristics for accurate diagnosis.

In summary, this thesis presents a novel deep learning-based approach for removing confounding factors from dermoscopic images, which not only operates more effectively than existing solutions but also significantly enhances the reliability and explainability of neural network classifiers. This method facilitates the early and accurate diagnosis of skin cancer, which can be life-saving, and represents a substantial advancement in the field of dermoscopic imaging from both scientific and industrial perspectives.

1. fejezet

Bevezetés

A melanóma a bőr rákos megbetegedésének legveszélyesebb formája, aminek esetszáma Magyarországon 2001 és 2019 között megkétszereződött, így az incidenciákat tekintve a 9. leggyakoribb rosszindulatú daganatos megbetegedéssé vált a Nemzeti Rákregiszter 2019-es adatai alapján [1].



1.1. ábra. Melanóma előfordulási arányának változása Magyarországon 2001 és 2019 között. Forrás: Nemzeti Rákregiszter

Bár az új terápiás módszerek egyre pozitívabb hatást eredményeztek a melanóma mortalitás incidencia arányának csökkentésében, a korai diagnózis fontossága és a prevenció továbbra is kulcsfontosságú tényező a betegség kimenete szempontjából. Ugyanis a korai stádiumban felismert melanómák esetében a túlélési ráta jóval magasabb, akár 98% [2].

A jelenkor gyors technológiai fejlődésével olyan klinikai képfeldolgozó megoldások jelentek meg, amik széles körben elérhetővé tették a gyors és költséghatékony szűrést. Bár ezek a szolgáltatások potenciálisan pozitívan hatnak a melanómával kapcsolatos edukációra, prevencióra és korai diagnózisra, mégis számos kritika érte őket a pontatlan és megbízhatatlan működésük miatt [3, 4, 5]. Napjainkra a rendelkezésre álló diverz melanóma adatbázisok lehetővé tették olyan mélyneurális háló alapú diagnosztikai rendszerek

fejlesztését, amik képesek cáfolni a korábbi kritikákat megfelelő szakértői szupervízió mellett.

2022-ben a Semmelweis Egyetem és az AIP Labs együttműködésével Magyarországon is megjelent egy hasonló szolgáltatás, írja a Semmelweis Hírek egyik cikke: „Az AIP Derm az ellátás elejétől végéig terjedő digitális kórház rendszer, ami nemcsak az esetek feldolgozási sebességét növeli többszörösére, de az adminisztratív terheket is csökkenti az egészségügyi dolgozók oldalán, és kényelmes felhasználói élményt nyújt mind számukra, mind a pácienseknek.” A szolgáltatást hamar igen nagy érdeklődés övezte, ugyanis a beiktatást követő 2 és fél hónapban, több mint 13 ezer páciens kapott diagnózist, és előfordult, hogy 24 óra leforgása alatt 2.000 képet töltöttek fel bőrelváltozásokról [6].

Az AIP Derm-hez hasonló klinikai döntéstámogató szolgáltatások hatékony és megbízható működéséhez nélkülözhetetlen a mesterséges intelligencia és mélytanulás alapú megoldások integrálása. Ezekről már számos tanulmány állította, hogy képesek legalább olyan hatékonysággal és pontossággal felismerni a melanómát, mint egy bőrgyógyász [7, 8]. Ezzel együtt a mélytanulás alapú szoftverek használata jelenleg számos kompromisszummal jár együtt klinikai kontextusban. Ezek között van a feketedoboz alapú működés, ami azt jelenti, hogy a szoftver nem ad rálátást arra, hogy mi alapján döntött egy adott diagnózis mellett. Pedig ennek érthetősége klinikai környezetben kritikus, ugyanis ez végső soron az orvosi döntéshozatalt akár nehezítheti is, ahelyett hogy támogatná. A neurális hálózatok egy másik problémája, hogy nem ugyanazon jellemzők alapján hoz meg egy döntést, mint egy képzett szakorvos. Sőt, egy bőrgyógyászati elváltozásról készített képen lehetnek olyan objektumok, amik míg egy szakorvos munkáját támogatják, addig egy neurális hálózat döntéshozatalába torzítást vihetnek be [9, 10].

Ezek megoldásával viszont egy háziorvos is gyorsan kaphat egy kiinduló diagnózist pácienséről, ami alapján továbbírányíthatja bőrgyógyászhoz. Továbbá, akár a saját otthonukból is könnyedén fordulhatunk szakemberhez egy fotó elkészítésével, akitől rövid időn belül, akár egy kezelési tervet is kaphatunk anélkül, hogy végig kellene mennünk a teljes orvoslátogatási folyamaton.

A fő célom, a mélyneurális hálókat potenciálisan zavaró objektumok eltávolítása mellett a bőrgyógyász szakorvosok munkájának abbéli segítése, hogy az egyes, a diagnosztizációjukat akadályozó tényezőket minél jobb képi minőségmegőrzés mellett tudjam eltávolítani az egyes képekről, generatív technológia alkalmazásával.

A munkám hátralévő része a következőképpen épül fel: a 2. fejezetben bemutatom a dermatoszkópos képek feldolgozása során felmerülő problémákat, felvezetve a generatív neurális hálózatok alkalmazásának szükségességét. A 3. fejezetben ismertetésre kerülnek a korábban publikált, a dolgozathoz szorosan kapcsolódó kutatások, amik mérföldkövet jelentettek a dermatoszkópos képeken megjelenő zavaró tényezők neurális hálózat alapú eltávolításában. Az azt követő 4. fejezetben a dolgozat keretében megvalósított megoldás módszertana és fejlesztési döntései kerülnek bemutatásra. A módszertan bemutatását az eredmények értékelése követi az 5. fejezetben. A dolgozatot a 6. fejezet zárja, ami az eredményeket összegzi.

2. fejezet

Problémafelvetés

Ebben a fejezetben a dolgozat fő problémájának kontextusát, és annak a feloldására megjelent klasszikus képfeldolgozási eszközöket és mély neurális hálózat alapú módszereket fogom bemutatni, ezzel felvezetve a generatív modellek alkalmazásának szükségességét.

2.1. ISIC adathalmaz bemutatása

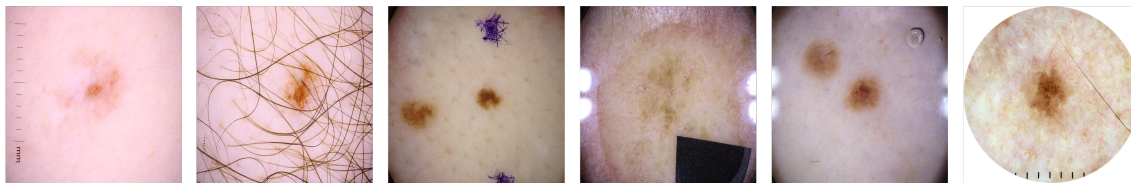
Az ISIC (International Skin Imaging Collaboration) egy akadémiai és ipari intézmények közti összefogás, aminek a célja első körben a melanóma korai felismeréséhez szükséges fejlesztések támogatása, ezzel elérve a melanóma miatti felesleges biopsziák elvégzését és a halálozások számának csökkentését. Éppen ezért, az ISIC szervezet olyan erőfeszítéseket tesz, mint a minőségbiztosítási és adatvédelmi szabványok fejlesztése, Kaggle Challenge-ek szervezése a tehetséges kutatók bevonása érdekében, valamint olyan erőforrások létrehozása és karbantartása, mint az ISIC Archive, ami a világ legnagyobb nyilvánosan elérhető bőrdaganat adatbázisa [11].

Az ISIC Archive 2020-ig¹ megjelent verziói dermatoszkóppal készült képeket tartalmaztak, amik közül az utolsó már összesen 33.126 címkézett tanító képet tartalmazott, amelyek közül a melanómát ábrázoló képeket biopsziával, míg a többi képet orvosi szakértők véleményével, longitudinális követéssel vagy biopsziával igazoltak. Az adathalmazt két fő kihívás jellemzi. Az egyik az, hogy a melanómát és az egészséges bőrelváltozásokat tartalmazó képek aránya igen kiegyensúlyozatlan, ugyanis a melanómát tartalmazó képek a teljes adathalmaznak csupán az 1,67%-át teszik ki. Egy másik probléma, hogy az adathalmaz képein a bőr és a bőrelváltozások mellett számos olyan tényező is jelen van, mint például a szőrszálak, tetoválások vagy olyan vizuális jelenségek, amik az orvosi feladatvégzés során kerülhettek a képekre. Orvosok közt bevett szokás, hogy tollal rajzolt kontúrral jelölik a bőrelváltozások helyét, vagy éppen vonalzóval annak méretét. E két probléma együttese megnehezíti az olyan mélytanulás alapú klasszifikációs modellek fejlesztését, amelyek képesek a melanómát az egészséges bőrtől és egyéb bőrelváltozásoktól hatékonyan és megbízhatóan megkülönböztetni.

2.2. Az adathalmazban előforduló zavaró tényezők hatása

A. Bissoto és tsai. [12] hívták fel a figyelmet arra a tényre, hogy az ISIC adathalmaz számos olyan tényezőt tartalmaz, amelyek potenciálisan torzítást és megtévesztő minta variációkat vihetnek be a mélytanulás alapú klasszifikációs modellek tanítási folyamatába.

¹2024-től az ISIC Archive a dermaszkóppal készített képek helyett a teljes test fotózásával készült és egységesen megvágott képeket tartalmaz. A változás oka, hogy a távorvoslás megjelenésével fontossá vált a gyenge minőségű kamerával készült képek pontos és megbízható elemzése is.



2.1. ábra. A dermatoszkóp képeken előforduló zavaró tényezők: vonalzó, szőrszál, tintafolt, marker, dermatoszkóp, keret stb.

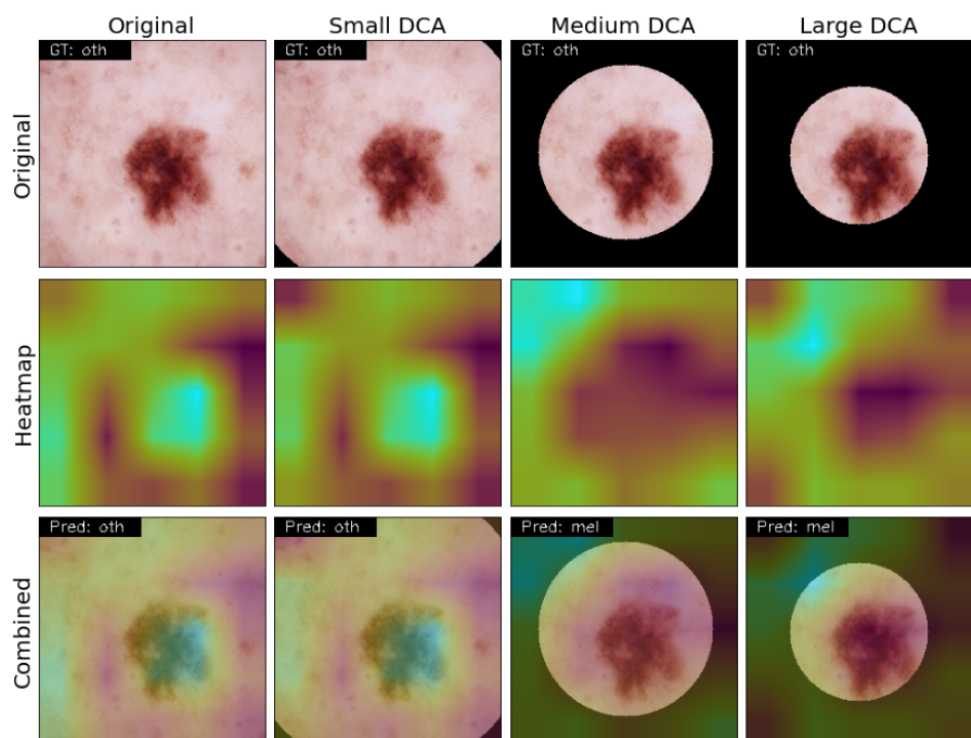
A cikk szerzői a torzító tényezőket először 7 kategóriába különítették el, majd különböző kísérletekkel igazolták, hogy ezek a zavaró tényezők befolyásolják az egyes kép régiókra vonatkozó figyelmet a modellek klasszifikációs döntései során. Emellett a zavaró tényezők okozta torzítás megszüntetésére is igyekeztek valamilyen módszert javasolni. Ehhez a Kim és tsai. [13] által bemutatott torzításelimináló tanítási algoritmust, a Learning Not To Learn (LNTL) módszert alkalmazták, ám ez a kísérlet sikertelen volt. P. J. Bevan és A. Atapour-Abarghouei [14] egy későbbi munkában rámutatott arra, hogy A. Bissoto és tsai. nem voltak teljesen körültekintőek, ugyanis a 7 kategória eltávolításának hatását külön-külön nem vizsgálták. Emellett hiányolták az LNTL módszer más torzításelimináló módszerekkel történő összehasonlítását is. Ezért egy újabb kiértékelést végeztek a Turning a Blind Eye (TABE) és egy hibrid megoldás (CLGR) bevonásával, amelyeknek eredményei a 2.1 táblázatban látványosan demonstrálják, hogy a vonalzó jelenléte összefüggésben áll a neurális háló AUC pontszámának romlásával.

Kísérlet	Vonalzó torzításának eltávolítása	
	Tiszta	Vonalzó
Referencia alap	0.999±0.000	0.831±0.022
LNTL	0.997±0.001	0.874±0.031
TABE	0.992±0.002	0.938±0.017
CLGR	0.999±0.010	0.958±0.018

2.1. táblázat. A vonalzó eltávolításának hatása a modell klasszifikációs teljesítményét jellemző AUC pontszámokra.

S. W. Passton [15] és tsai. az olyan sötét kereteknek vizsgálták a klasszifikációs modellek döntésére gyakorolt hatását, amelyek a bőrelváltozás körül jelennek meg körkörösén egészen a kép széléig, ezzel egyfajta fókuszot helyezve a bőrelváltozásra. A sötét keret hatásának megértése és megmagyarázhatósága érdekében a szerzők először egy eltávolítási módszert terveztek, amit később egy egyfajta előfeldolgozási lépésként alkalmaztak a modellek tanításánál. Ezután két modellt tanítottak: egyet a sötét kerettel rendelkező képeken, egy másikat pedig olyan képeken, amelyek ezektől mentesek. A modellek figyelmi mintázatának vizsgálatához Grad-CAM módszert alkalmaztak, ami a hálózat végső konvolúciós rétegéből nyert gradiens információk alapján hőképeket generált a végleges döntést leginkább meghatározó képi jellemzők szerint. Az elemzés során azt tapasztalták, hogy a sötét keret jelenléte torzította a modell figyelmét, ami a bőrelváltozás helyett nagyobb hangsúlyt helyezett a keretre. Ezzel szemben a zavaró tényezőtől mentes képeken tanított modell a bőrelváltozás releváns régióira koncentrált.

Kim és Hong [16] a szőrszálak hatását vizsgálta népszerű klasszifikációs modellekre. A szőrszálak eltávolítását egy nem felügyelt módon tanított GAN generátorával végezték, ami számos esetben jelentősen megváltoztatta a bőrelváltozások képi jellemzőit, ám mégis pozitív hatással volt a klasszifikációs modell teljesítményére. A 2.2 táblázatban lát-



2.2. ábra. Különböző méretű sötét keretek hatása a modell döntésére.

ható, hogy a szőrszálak eltávolítása egyértelműen pozitív hatással van a klasszifikációra. A szőrszálak eltávolításával átlagosan 0,08-0,09-el nőtt az ACC, AUC, F1, PRE, REC és SPEC metrikák értéke. A legnagyobb növekedés az ACC, AUC és F1 metrikákban érzékelhető, ami azt jelenti, hogy a szőrszálak eltávolítása jelentősen hozzájárult a modellek általánosító képességének javulásához.

Bemenet	Metrika	Res18	Res152	RX101	EC-B3	EC-B4
Eredeti	acc	0.7400	0.7667	0.7133	0.6467	0.6267
	auc	0.8186	0.7840	0.7556	0.6844	0.6951
	f1	0.7447	0.7683	0.7200	0.6503	0.6317
	pre	0.7400	0.7667	0.7133	0.6467	0.6267
	rec	0.7387	0.7663	0.7110	0.6447	0.6229
	spec	0.7467	0.8000	0.8000	0.6533	0.5733
D. Kim B.-W. Hong	acc	0.8267	0.7800	0.8133	0.7533	0.7444
	auc	0.9093	0.8196	0.8489	0.8124	0.8111
	f1	0.8288	0.7422	0.7327	0.7618	0.7489
	pre	0.8267	0.7333	0.7276	0.7533	0.7444
	rec	0.8263	0.7305	0.7246	0.7517	0.7433
	spec	0.8000	0.8267	0.7733	0.7067	0.7111

2.2. táblázat. A szőrszál eltávolítás hatása a modell döntésére.

A Heidelbergi Egyetem Dermatológiai Tanszékének orvosprofesszorai, J. Winkler és tsai. 2021-re felhívták a figyelmet arra, hogy a konvolúciós neurális hálózatok számára szánt minták esetében elkerülendő a tollal rajzolt jelölések vagy a vonalzők felvétele. Ugyanis

ezekről a tényezőkről egyértelműen kimutatták és klinikailag igazolták, hogy a jóindulatú bőrelváltozások rosszindulatúnak történő osztályozásához vezethetnek [9, 10].

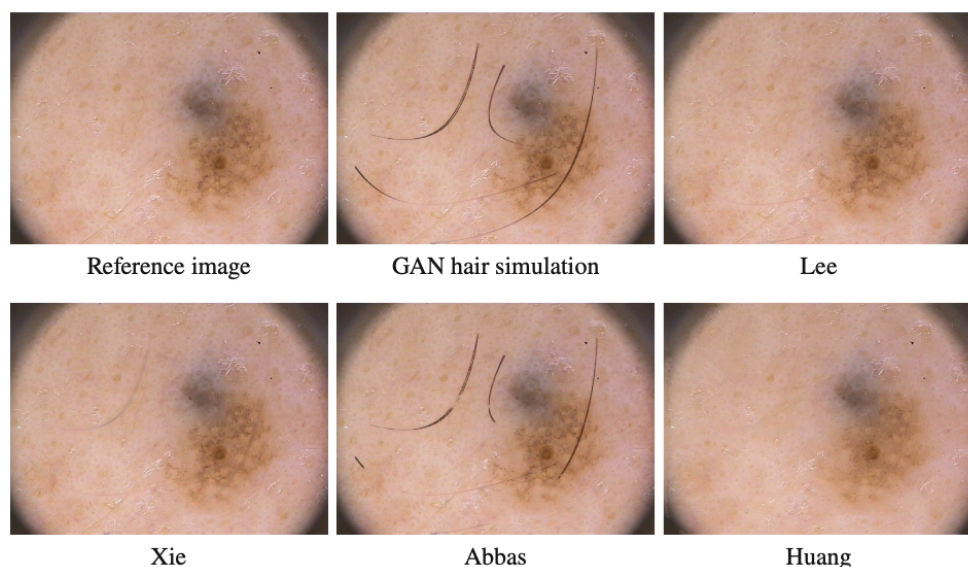
2.3. Módszerek zavaró tényezők eltávolítására

A 2.2 alfejezetben bemutatott kutatások eredményei alapján a dermatoszkóp képeket feldolgozó neurális hálózatok tanítása során kritikus, hogy a tanító adathalmaz torzítást okozó tényezőktől mentes legyen. A torzítások eltávolítására, főleg a szőrszálakat és vonalzókat megcélozva, már számos módszer megjelent a szakirodalomban, amelyek közül sokáig a legnépszerűbb megközelítés egy szegmentációs és egy rekonstrukciós algoritmus együttes alkalmazása volt. Ez a kétlépésű megközelítés klasszikus képfeldolgozási módszerekkel is megvalósítható, de a mélytanulás térnyerésével egyre inkább a neurális háló alapú megoldások felé tolódtak el a kutatások.

2.3.1. Klasszikus képfeldolgozási módszerek

A klasszikus képfeldolgozási módszerek szegmentációs lépése leginkább matematikai morfológián [17, 18, 19, 20] alapul, esetleg kiegészítve szűrőkkel [21] [22]. A kihívást a szegmentálás után eltávolított részek helyreállítása jelentette, amihez interpolációs [17], iteratív kitöltési és egyéb matematikai módszereket alkalmaztak [18] [21] [22] [23] [19] [20], de a 2010-es évek közepére már megjelentek olyan módszerek is, amelyek képesek voltak a bőrelváltozásokat elenyésző veszteséggel rekonstruálni.

Ezek a módszerek bár igen gyorsak, a szegmentációs lépést csak akkor tudják hatékonyan elvégezni, ha a bőrelváltozások és zavaró tényezők közötti kontraszt nagy, ami már nem teljesül a sűrűbb testszőrzet vagy a sötétebb bőrszínek esetén sem [21]. Ezért a kielégítő rekonstrukciós eredmény eléréséhez a különböző kontrasztarányok miatt sokszor egyedi hiperparaméterekre van szükség, így a klasszikus módszerek nem automatizálhatók.



2.3. ábra. Klasszikus szőrszál eltávolítási módszerek eredményei szimulált szőrszálakon.

Módszer	Év	Szegmentáció	Rekonstrukció
Lee et al.	[17] 1997	Morph. closing	Bilinear interpolation
Xie et al.	[18] 2009	Grayscale top-hat	Non-linear PDE
Abbas et al.	[21] 2011	Gaussian filters	Coherence transport
Huang et al.	[22] 2013	Conventional matched filters	Median filtering
Toossi et al.	[23] 2013	Canny edge detection Morph. operatorrs	Multi-resolution coherence transport
Koehoorn et al.	[24] 2015	Gap-detection by multiscale skeletons	Coherence transport
Bibiloni et al.	[19] 2017	Color-sensitive top-hat	Morph. inpainting
Kasmi et al.	[20] 2021	Adaptive thresholding	Color averaging

2.3. táblázat. Klasszikus szőrszál eltávolítási módszerek.

2.3.2. Neurális háló alapú módszerek

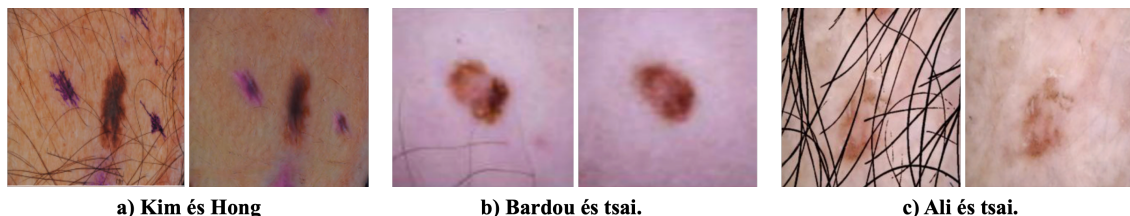
A neurális háló alapú megoldásokkal való kísérletezésben a kihívást elsősorban az jelentette, hogy a szegmentációs és kitöltő feladat magas precizitású elvégzéséhez szükséges tanításhoz nem állt rendelkezésre megfelelő adathalmaz, ugyanis a melanóma adatbázisok jobb esetben is csak a bőrelváltozások maszkját tartalmazzák a képekhez, de a zavaró tényezőket nem. Ezért a kutatók továbbra is a korábban említett klasszikus képfeldolgozási módszerekkel állították elő a zavaró tényezők maszkjait. Ezeknek a maszkoknak a felhasználása azonban problémát jelenthet a neurális hálózatok tanításánál, ugyanis a kép és maszk párok csak nagy bőr-szőrszál kontrasztarányok mellett állíthatók elő[25]. 2023-ra már számos olyan adatbázis vált publikusan elérhetővé, amelyekben dermatoszkópos képekhez tartozó szőrszál maszkok is rendelkezésre állnak.

A zavaró tényezők zajszerű megközelítésére épülő megoldásokhoz elsőként Mirzaalian és tsai. [26] mutattak be egy olyan módszert, amivel lehetségessé vált tiszta és szintetizált szőrszálakkal rendelkező képpárokból álló tanítóhalmaz létrehozása. A módszerrel másodfokú polinomokat alkalmazva hozhatók létre szőrszál jellegű objektumok, amiken Gauss-szűrőt alkalmazva elérhető, hogy jobban illeszkedjenek az eredeti kép pixeleihez. Bár ez a módszer lehetővé tette a vastagság, szín és irány szerint is változatos szőrszálstruktúrák létrehozását, mégis azzal kritizálták, hogy a generált szőrszálak nem elég változatosak és valósághiűek. Éppen ezért Attia és tsai. egy innovatívabb megközelítéssel álltak elő: létrehozták az első GAN alapú módszert, ami már élethűbb szőrszálakat hoz létre [27].

Martínez és tsai [28] elsőként alkalmaztak neurális hálózatokat szőrszálak eltávolítására. Ők egy enkóder-dekóder hálózatot alkalmaztak, amik konvolúciós és transzponált konvolúciók segítségével rekonstruálták a képeket szőrszálak nélkül. Eredményeik magas minőségben megőrizték az eredeti bőrelváltozás struktúráját, ami egy jól megtervezett veszteségnek és felügyelt tanításnak köszönhető.

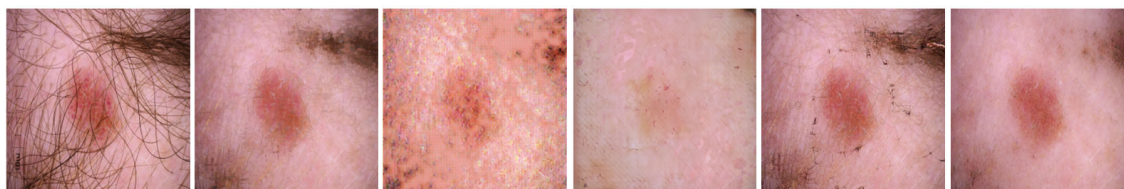
Kim és Hong [16] az ISIC 2020 adathalmaz egy részét használták fel arra, hogy egy olyan adathalmazt hozzanak létre, amelyeknek mintáit a szőrszálak jelenlétének függvényében osztottak fel két osztályra. Az adathalmaz segítségével elsőként tanítottak nem felügyelt módon WGAN-t szőrszálak eliminálására. Bardou és tsai. [29] pedig variációs autoenkódert alkalmazva javasoltak hasonló megoldást. Mindkét módszerrel sikerült a szőrszálakat hatékonyan eltávolítani még sűrű testszőrzet esetén is, viszont a bőrelváltozások rekonstrukciója még mindig nem volt tökéletes, még a veszteségfüggvény megfelelő megválasztásával sem. Ali és tsai. [30] egy újabb megközelítéssel álltak elő, aminek keretében szintetizált adathalmazon egy iteratív diffúziós modellel (IDM) oldották meg a

szőrszálak eltávolítását. Az IDM iteratív zajhozzáadási és eltávolítási folyamata lehetővé teszi a szőrszálak hatékony izolálását és eltávolítását. A módszer jelentős előrelépést jelent, mivel képes különböző változatos szőrszálstruktúrák eltávolítására nem felügyelt módon anélkül, hogy a bőrelváltozásokat torzítaná. Ali és tsai. publikációjában azonban nem tértek ki arra, hogy a modell hogyan teljesít valódi szőrszálak és sűrűbb testszőrzet esetén, aminek mulasztása egy kritikus hiányosság a módszer alkalmazhatóságának igazolását illetően.



2.4. ábra. Nem felügyelt módon tanított GAN eredményei szőrszálak eltávolítására.

Lee és You [31] a sűrű szőrszálstruktúrák ellen egy továbbfejlesztett GAN-t alkalmaztak, amiben a rekonstrukcióért egy kétirányú autoregresszív transzformer (enhanced bidirectional autoregressive transformer, EBAT) felel. Ez a módszer felhasználja a transzformerekben megjelent figyelmi mechanizmus sajátosságait a minél pontosabb helyreállítás érdekében. A módszerrel sikerült eltávolítani sűrű testszőrzettel kitakart bőrelváltozásokat alacsony mértékű diagnosztikai információ elvesztése mellett. A szőrszálak eltávolítása során a modell jól teljesít, akár sűrű testszőrzet esetén is. A módszerük hátránya viszont, hogy a szőrszálak eltávolítása nem sikeres azokban az esetekben, amikor a szőrszálak színei hasonlóak a bőr vagy a bőrelváltozás színeihez.



2.5. ábra. Lee és You [31] eredményének összehasonlítása más módszerekével. Balról jobbra: eredeti kép, Koehoorn és társai, Shift-Net, DeppFill v2, BAT-Fill, EBAT.

Az eddig megjelent adathalmazok és módszerek egy komoly hiányossága, hogy az olyan tényezőkre, mint a vonalzó, tinta markerek vagy keretek, amelyek orvosi munkavégzés során kerültek a képekre, nem fókuszáltak. Ez annak ellenére van így, hogy már számos publikációban felhívták a figyelmet arra, hogy ezen tényezők negatívan hatnak a modell döntéseire magyarázhatóság szempontjából. Emellett a szőrszálak eltávolítása még a legújabb neurális háló architektúrák alkalmazása mellett is egy bizonyos mértékű információvesztéssel jár, ami az orvosi döntéstámogatást kevésbé segíti. Éppen ezért hiánypótló egy olyan adathalmaz létrehozására, ami lehetővé teszi egy olyan modell tanítását, ami képes ezen tényezők eltávolítására is. Így a jelen munka egyik célja egy olyan adathalmaz létrehozása és modell architektúra tervezése, ami a szőrszálak mellett az egyéb zavaró tényezők eltávolításának tanítását is lehetővé teszi a bőrelváltozások lehető legjobb megőrzése mellett.

3. fejezet

Kapcsolódó munkák

A generatív modellezés a gépi tanulás egy olyan részterülete, ami egy adathalmazból vett mintavételezés alapján képes új, az eredeti mintákhoz hasonló mintákat létrehozni az adathalmazban rejlő jellemzők és struktúrák alapján. A generatív modellek tanítása során a tanítási mintákat egy ismeretlen eloszlásból mintavételezzük, és a célunk az, hogy a modell megtanulja minél pontosabban közelíteni az adathalmazt leíró eloszlást. A képfeldolgozás területén az egyik generatív modellarchitektúra az ún. generatív versengő hálózat, amit Goodfellow és tsai. [32] mutattak be [33].

3.1. Generatív Versengő Hálózatok

Az általános GAN architektúra (Vanilla GAN) [32] egy g generátor és egy d diszkriminátor modelltől áll, ami egy bináris osztályozó. A generátor az $\mathbf{x}$ minták létrehozásáért, míg a diszkriminátor annak az $\mathbf{y}$ valószínűségeknek a meghatározásáért felel, ami szerint az $\mathbf{x}$ minták az eredeti eloszlásból származnak.

$$\mathbf{x} = g(\mathbf{z}, \theta_g), \text{ ahol } \mathbf{z} \sim p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \text{ és } \theta_g \text{ a generátor paraméterei} \quad (3.1)$$

$$\mathbf{y} = d(\mathbf{x}, \theta_d), \text{ ahol } \theta_d \text{ a diszkriminátor paraméterei} \quad (3.2)$$

Ez a felállás lehetővé teszi a generálási folyamat felügyelt tanítását, aminek során a cél az, hogy a generátor olyan képeket hozzon létre, ami maximalizálja a diszkriminátor veszteségét, miközben a diszkriminátor minimalizálni próbálja ugyanezt a veszteséget, a 3.3-as optimalizálási feladatot megoldva.

$$\theta_g^* = \min_{\theta_g} \max_{\theta_d} \underbrace{\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\ln d(\mathbf{x}, \theta_d)]}_{\text{veszteség a tanító adathalmaz eloszlásából}} + \underbrace{\mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} [\ln(1 - d(g(\mathbf{z}, \theta_g), \theta_d))]}_{\text{veszteség a generált adathalmaz eloszlásából}} \quad (3.3)$$

Ennek eredményeként a generátor és a diszkriminátor egymás ellen versenyzik a tanítás során, amiből a módszer neve is ered.

A GAN-ok képesek magas minőségű képeket generálni, azonban tanításuk során számos kihívással küzdhetünk a hiperparaméterekre való érzékenység és a tanítás versengő jellegéből adódó instabilitás miatt. Az instabilitásból eredő látványos probléma az ún. mode collapse, amikor a generátor az adathalmaz csak egy kisebb részhalmazának eloszlását képes megragadni, ami miatt nem képes a teljes adathalmaz alapján új minták generálására.

Ezen problémák orvoslására több megoldást is publikáltak, amelyek főként a veszteségfüggvény számításának módját befolyásolták. A GAN eredeti verziójában 3.3-as opti-

malizálási feladatban a diszkriminátor veszteségét a bináris keresztentropia veszteségével számolták. Mao és tsai. [34] bizonyították, hogy a keresztentropia alkalmazása az ún. eltűnő gradiens problémához vezet a generátor veszteségének visszaterjesztése során, ami végső soron az instabilitást okozza. Ennek feloldása érdekében a keresztentropia helyett a legkisebb négyzetes hibát alkalmazták veszteségfüggvényként. A veszteség így a döntési határtól való távolság alapján számítható, ami lehetővé teszi, hogy a generátor által létrehozott minták közelebb kerüljenek a valódi adatokhoz, mivel a generátor veszteségének számításakor az is büntetve lesz, ha egy generált minta túl messze esik a döntési határtól [33].

Arjovsky és tsai. [35, 33] az optimalizációs célként azt fogalmazták meg, hogy a generátor által létrehozott minták eloszlása és a valódi adatok eloszlása közötti Wasserstein-távolság minimális legyen a reprezentációs térben. Ebben a megközelítésben, a Wasserstein GAN-ban (WGAN), a diszkriminátor egy új szerepet tölt be, már nem egy bináris osztályozó, hanem azt méri, hogy mennyire vannak közel egymáshoz a generált és a valódi minták eloszlásai.

$$\theta_g^* = \min_{\theta_g} \max_{\theta_d} \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [d(\mathbf{x}, \theta_d)] - \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} [d(g(\mathbf{z}, \theta_g), \theta_d)] \quad (3.4)$$

A távolság minimalizáláshoz arra van szükség, hogy a diszkriminátor veszteségének gradiense csak egy bizonyos korláton belül növekedhessen vagy csökkenhessen, így a 3.4 optimalizációs problémában szereplő θ_d egy olyan diszkriminátor paraméterezés, ami ennek eleget tesz. Ezt a diszkriminátor súlyainak egy korlátozott tartományban tartásával érték el (weight clipping), amiről már a publikációjukban is írták, hogy nem a legjobb megoldás, mivel ha a meghatározott tartomány túl szűk, akkor diszkriminátor tanítása nehezzé válik, ha pedig túl nagy, akkor már nem garantált a veszteség gradiensváltozásának korlátozottsága. A korlát biztosításához egy új megközelítést Gulrajani és tsai. [36] mutattak be. Módszerük egyfajta regularizációs módszer, aminek keretében a diszkriminátor veszteségéhez egy új tagot adtak hozzá, ami gradient penalty néven ismert. Ez a tényező a valódi és a generált minták $\tilde{\mathbf{x}}$ interpolációján kapott veszteséggradiens egytől való eltérését méri L2 normával, ami garantálja, hogy a meredekség ne változzon túl gyorsan [33].

$$R_{GP}(\mathbf{x}, \tilde{\mathbf{x}}) = \lambda \cdot \mathbb{E}_{\tilde{\mathbf{x}} \sim p_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}})} [(\|\nabla_{\tilde{\mathbf{x}}} d(\tilde{\mathbf{x}}, \theta_d)\|_2 - 1)^2], \text{ ahol } \lambda \text{ a regularizációs tényező} \quad (3.5)$$

$$\tilde{\mathbf{x}} = \epsilon \mathbf{x} + (1 - \epsilon) \tilde{\mathbf{x}}, \text{ ahol } \epsilon \sim \mathcal{U}(0, 1), \tilde{\mathbf{x}} \text{ és } \mathbf{x} \text{ a generált és valódi minták} \quad (3.6)$$

A gradient penalty előnye a weight clippinggel szemben, hogy gyorsabbá teszi a tanítást, és lehetővé teszi magas minőségű minták generálását. Egy másik pozitívum, hogy a módszer nem zárja ki az olyan momentum alapú optimalizációs algoritmusok használatát, mint az Adam, ami Arjovsky és tsai. [35] megfigyelései szerint a weight clippingről nem volt elmondható. Zhang és tsai [37] azonban felhívták három problémára a figyelmet a gradient penalty alkalmazásával kapcsolatban. Az egyik probléma az, hogy réteg használat szempontjából homogénnek mondható GAN-ok esetében, amik normalizációs rétegek nélkül csak konvolúciós rétegekből állnak, a tanítás továbbra is instabil marad. Egy másik probléma, hogy még olyan körülmények között megtervezett hálózatok esetén is, mint a WGAN, megfigyelhető instabilitás, ha a tanítási idő túllép egy bizonyos küszöbértéket. A harmadik probléma pedig, hogy a WGAN kifejezetten érzékeny a tanulási rátára, és olyan értékek mellett, amelyeknél más GAN architektúrák jól konvergáltak, a WGAN instabillá vált és a veszteség gradiensei intenzív változásokat produkáltak. Ennek feloldása érdekében a szerzők egy új regularizációs módszert javasoltak, ami a totális variációt (total variation, TV) alkalmazta. A TV alkalmazásának előnye, hogy egyáltalán nem igényel gradiens számítást, így a tanítás nem igényel extra erőforrásokat. A számítási kapacitás mellett, a TV kevésbé érzékeny a GAN architektúra és a tanulási ráta megválasztására. Ennek

köszönhetően a gradiensnek nem változnak hirtelen, így stabilabb lesz a tanítás folyamata, még bonyolultabb vagy homogén hálózatok esetében is. A TV a valódi és generált minták közötti távolságot egy δ paraméterrel korlátozza, ami a regularizáció hiperparamétere a λ regularizációs együttható mellett.

$$R_{TV}(\mathbf{x}) = \lambda \cdot \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x}), \mathbf{z} \sim p(\mathbf{z})} [|d(\mathbf{x}, \theta_d) - d(g(\mathbf{z}, \theta_g), \theta_d) - \delta|] \quad (3.7)$$

3.2. A GAN-ok alkalmazási területei

A GAN-ok alkalmazási lehetőségei igen széleskörűek. Mára már számos javaslat megjelent különböző architektúrákra. Mirza és Ossindro [38] a Goodfellow és tsai. által publikált GAN generátorának és diszkriminátorának bemenetét kibővítették egy kondicionáló paraméterrel. A kondicionáló paraméter lehet egy címkézés, szöveges leírás vagy egy másik kép, ami segítségével szabályozhatóvá válik a generátor által létrehozott minták tartalma. Isola és tsai. [39] mutatták be az első olyan CGAN architektúrát, amit direkt kép-kép transzformációra terveztek, és aminek optimalizációs célját a 3.9 képlet mutatja be, ahol a hozzáadott regularizációs tag a generált és a valódi képek közötti L1 távolságot méri.

$$\theta_g^* = \min_{\theta_g} \max_{\theta_d} \mathbb{E}_{\mathbf{x}, \mathbf{y}} [\ln d(\mathbf{x}, \mathbf{y}, \theta_d)] + \mathbb{E}_{\mathbf{x}, \mathbf{z}} [\ln(1 - d(\mathbf{x}, g(\mathbf{z}, \theta_g), \theta_d))] \quad (3.8)$$

$$+ \lambda \cdot \mathbb{E}_{\mathbf{x}, \mathbf{z}} [||\mathbf{y} - g(\mathbf{x}, \theta_g)||_1] \quad (3.9)$$

Ezzel a módszerrel lehetségessé válik az ún. reprezentáció tanulás, ami különösen hasznos, ha egy adathalmazt leíró eloszlás több alacsonyabb rendű eloszlásból tevődik össze. Az alacsonyabb rendű eloszlásokra való érzékenységgel az adat augmentáció egy hatékonyabb módját érték el, ami megoldást jelentett olyan gyakori problémákra, mint a kicsi vagy kiegyensúlyozatlan adathalmazok. A generatív modelleknek köszönhetően, így már nem csak transzformatív leképezésekkel lehet minták sokaságát előállítani – amelyek valójában nem változtatnak érdemileg a minták jellemzőin –, hanem olyan új mintákat is létrehozhatunk, amelyeket a tanító adathalmaz eredetileg semmilyen formában nem tartalmazott, de a valóságban előfordulhatnak. Isola és tsai. megfigyelték, hogy a kép-kép transzformáció megvalósítható a $\mathbf{z}$ zajból való mintavételezés nélkül is, ám ebben az esetben a generátor determinisztikussá válik, tehát a tanítás után egy-egy bizonyos bemenetre mindig ugyanazt a kimenetet fogja adni. Ebben az esetben a generátor csak a zaj kiszűrését lesz képes megvalósítani, de nem lesz képes új mintákat generálni. Ez a felfedezés azonban a GAN-ok alkalmazási területét tovább bővíti, mivel ez a determinisztikus működés lehetővé teszi a képek zajtól vagy zajnak értelmezhető zavaró tényezőktől mentes rekonstrukcióját.

4. fejezet

Módszertan

A következő fejezetben a dolgozat keretében fejlesztett zavaró tényezők eltávolítását megcélzó megoldás módszertana kerül bemutatásra. A fejezetben részletesebben kifejtem a modell betanításához szükséges adathalmazt, az architektúrát, a modell tanítási algoritmusát meghatározó tervezési döntéseket, valamint a modell teljesítményét finomhangoló hiperparamétereket és a kiértékeléshez alkalmazott metrikákat.

4.1. Adathalmaz létrehozása

Az adathalmaz létrehozása során megfogalmazott kritérium volt, hogy a dermoszkópon leggyakrabban előforduló zavaró tényezőket lefedje. Éppen ezért a jelen dolgozat keretében is szintetikus adathalmaz létrehozására került sor. Ennek érdekében az ISIC 2020 adathalmazból manuálisan került kiválasztásra 68 db olyan kép, amik jóindulatú és rosszindulatú bőrelváltozásokat egyaránt tartalmaznak, kerülve rajtuk a zavaró tényezők előfordulását. A szintetizált objektumok az OpenCV keretrendszer segítségével lettek a képekbe beillesztve.

4.1.1. Szőrszálak szimulációja

A szőrszál struktúrák generálásához a Mirzaalian és tsai. [26] által publikált módszerhez hasonló stratégia lett alkalmazva, külön figyelve arra, hogy az általános szőrszálak mellett akár borotvált esetek is előfordulhatnak. Az általános szőrszálak harmadfokú Bézier-görbékkel (4.1) lettek modellezve, rajtuk Gauss-szűrő (4.1) alkalmazásával elérve, hogy jobban illeszkedjenek az eredeti képbe. A Gauss-szűrő egy lineáris konvolúciós szűrő, ami egy kernel segítségével végighalad a képen, és az adott területen lévő értékeket a kernel szerinti súlyokkal összegzi. A Gauss-szűrő esetében a kernel középpontjától távol eső súlyok kevésbé érvényesülnek (a Gauss-görbe szerint), aminek hatására a szőrszálak a képbe simíthatók. A Mirzaalian és tsai. [26] által publikált módszer kritikái miatt a szőrszálak a képek ötszörösen felnagyított változatára lettek felrajzolva, és vastagságuk és színük a szőrszálak hosszával arányosan változik. A szőrszálak színe, vastagsága és lokalizációja paraméterként állítható, és opcionálisan beállítható, hogy a szőrszálak közé kerüljenek bizonyos mennyiségben ősz szőrszálak is. A lokalizáció esetében lehetőség van annak beállítására is, hogy a szőrszálak a kép jobb és bal oldaláról induljanak a kép középső része felé.

$$P(t) = (1 - t)^3 \cdot P_0 + 3 \cdot (1 - t)^2 \cdot t \cdot P_1 + 3 \cdot (1 - t) \cdot t^2 \cdot P_2 + t^3 \cdot P_3 \quad (4.1)$$

Ahol:

- $P_i, i \in \{0, 1, 2, 3\}$: a Bézier-görbe kontrollpontjai,
- $t \in [0, 1]$: a görbe interpolációs paramétere.



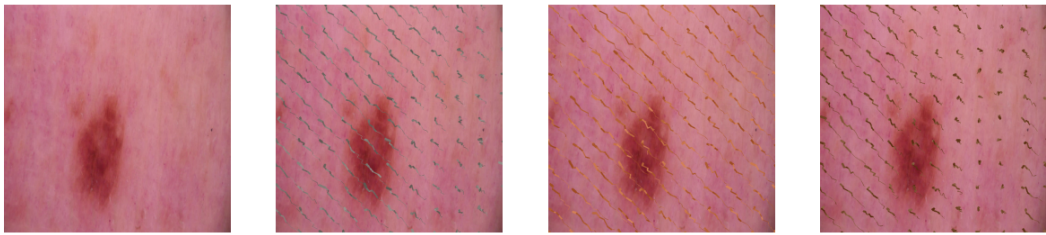
4.1. ábra. Szőrszálak szimulációja. Balról jobbra: eredeti kép, szimulált szőrszálváltozatok.

A borotvált szőrszálak generálása során használt eljárás elsőfokú Bézier-görbéken alapult, kiegészítve egy hullámgeneráló függvényvel, így egyszerűbb, kicsit hullámos vonalakat eredményezve. Az általános szőrszál generáláshoz hasonlóan a borotvált szőrszálak esetén Gauss-szűrőt alkalmazva került megvalósításra a szőrszálak illesztése az eredeti képhez. Az általános szőrszálstruktúrákhoz képest itt a szőrszálak egymáshoz képesti elhelyezkedése is ki lett kényszerítve, amit egy négyzetrácsos elrendezés határoz meg.

$$P(t, \omega, \phi) = (1 - t) \cdot P_0 + t \cdot P_1 + A \cdot \sin(\omega \cdot t + \phi) \quad (4.2)$$

Ahol:

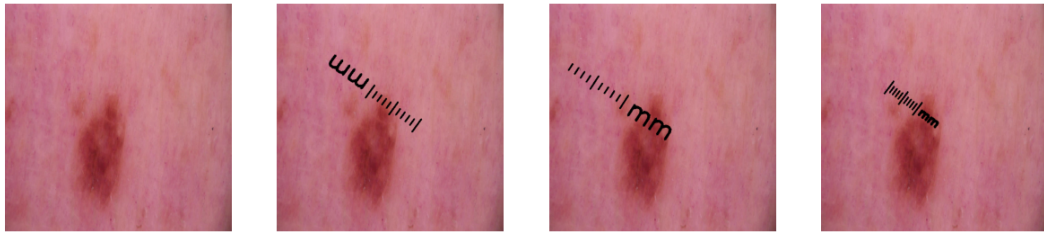
- $P_i, i \in \{0, 1, 2, 3\}$: a Bézier-görbe kontrollpontjai,
- $t \in [0, 1]$: a görbe interpolációs paramétere,
- A : a hullám amplitúdója,
- ω : a hullám frekvenciája,
- ϕ : a hullám fázisa.



4.2. ábra. Borotvált szőrszálak szimulációja. Balról jobbra: eredeti kép, szimulált borotvált szőrszálváltozatok.

4.1.2. Vonalzó szimulációja

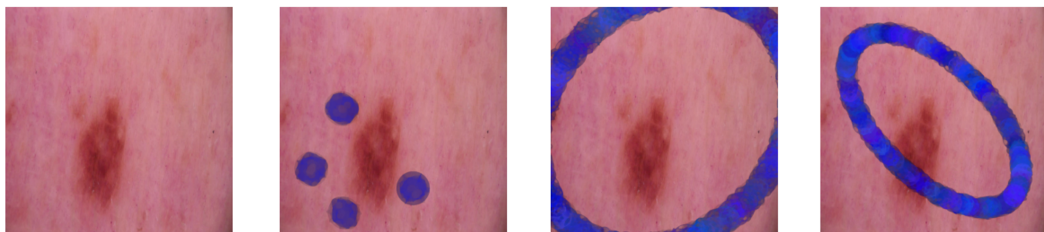
A vonalzó szimulációja során a vonalzó struktúrák különféle méretekben és orientációkban kerülnek beillesztésre a képekbe. Az algoritmus a bemeneti kép szélességétől és magasságától függően generál egy véletlenszerű méretű, rotációjú és pozíciójú vonalzót. A vonalzó 11 db skála beosztással is rendelkezik, hosszú és rövid vonalakból áll, amik között minden ötödik jelölés hosszabb. A vonalzóhoz tartozó “mm” felirat mérete a skála hosszával arányosan változik, biztosítva a realisztikus megjelenést.



4.3. ábra. Vonalzó szimulációja. Balról jobbra: eredeti kép, szimulált vonalzóváltozatok.

4.1.3. Tintafoltok szimulációja

A tintafoltok esetében fontos volt, hogy a foltok elrendezése, mérete és intenzitása is változatos legyen. A tintafoltok alapegységei egymásra rajzolt ellipszisekből állnak, amelyek egy véletlenszerű szöggel el vannak forgatva. Az algoritmus ebben az esetben is a képek egy felnagyított változatára helyezi fel a tintafoltokat véletlenszerűen, vagy egy ellipszis alakban elosztva a képen. Ahhoz, hogy a foltok a tinta elkenődött hatását is keltsék a Gauss-szűrő mellett egy ellipszis alakú kernel alapú elmosás került alkalmazásra a foltokon. Ebben az esetben nem volt cél, hogy a foltok teljes mértékben a valóságot tükrözzék, ugyanis a tintafoltok célja, hogy bőrön megjelenő elváltozásokat lokalizálják egyfajta kontrasztanyagként. Éppen ezért inkább a tintafoltok színbeli és lokalizációbeli változatosságára került a hangsúly, ami elégséges ahhoz, hogy ezeket a tényezőket hatékonyan el lehessen távolítani. A foltok felvétele során tetszőlegesen állítható a foltok száma, mérete, színe, az elkenődés intenzitása, a foltok elhelyezkedése, és hogy kilógjanak-e a képről.



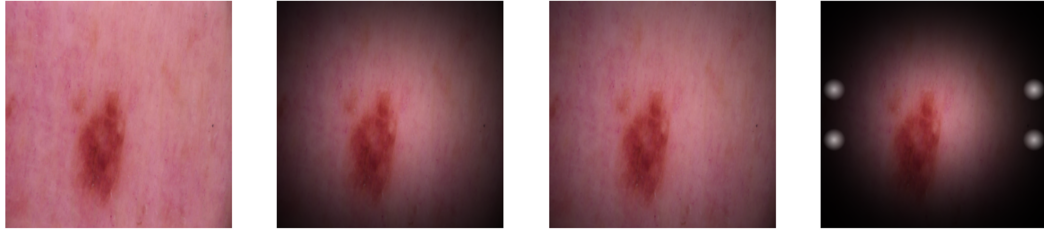
4.4. ábra. Tinta foltok szimulációja. Balról jobbra: eredeti kép, szimulált tinta foltváltozatok.

4.1.4. Dermatoszkóp árnyékának és fényének szimulációja

A dermatoszkópos képeken gyakran előforduló képi jellemzők közé tartozik a műszer által vetett árnyék, illetve a műszer lámpáinak visszatükröződése is. Az árnyék körkörösén helyezkedik el a bőrelváltozások körül a képek szélei felé folyamatosan sötétedve. Ebben az esetben az árnyék mellé opcionálisan felvehető négy fénypont, ami a képek bal és jobb oldalán helyezkedik el párosával. Az árnyék Gauss-kernellel került felvitelre, ami lehetővé teszi, hogy a központi elem, esetünkben a bőrelváltozás megőrizze az eredeti részleteit, míg a szélek felé a kép többi része folyamatosan elsötétedik. A fénypontok a kör képletének segítségével kerültek megrajzolásra, amiket egy maszkként változtatható intenzitással lehet hozzáadni a képekhez.

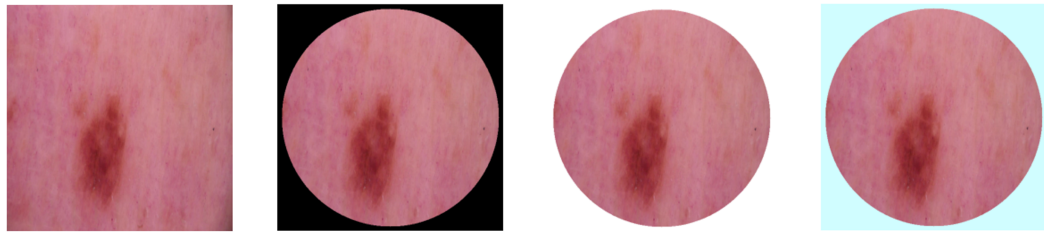
4.1.5. Keretek szimulációja

A dermatoszkópos képen egy másik előforduló tényező a bőrelváltozást kiemelő kör alakú keret. A keret felvétele során először a kép opcionálisan négyzet alakúra átméretezhető,



4.5. ábra. Dermatoszkóp árnyékának és fényének szimulációja. Balról jobbra: eredeti kép, szimulált dermoszkóp variációk.

ugyanis számos esetben az ilyen zavaró tényezővel rendelkező képek 1:1 arányúak. Ezt követően egy bináris maszk lett létrehozva, majd invertálva, aminek hatására egy fehér háttér közepén egy olyan fekete kör látható, aminek átmérője majdnem megegyezik a kép dimenzionalitásával, egy kis teret hagyva a keretnek. Ezt követően a körön kívüli rész egy tetszőleges színnel, míg a belső rész az eredeti képpel lett kitöltve.



4.6. ábra. Keret szimulációja. Balról jobbra: eredeti kép, szimulált keretváltozatok.

4.1.6. Minták generálása

Az adathalmaz létrehozása során fontos volt, hogy a zavaró tényezők változatosan jelenjenek meg a képeken. Éppen ezért a rajzoló algoritmus bemenete dict objektumok listája, amelyek a különböző zavaró tényezők felrajzolási elveit tartalmazzák felparaméterezve.

```
{
  "hair_strands": {
    "num_strands": 90,
    "min_thickness": 4,
    "max_thickness": 18,
    "mixed_color": False,
    "num_segments": 30
  },
  "ruler": {},
  "border": {
    "crop_to_rectangle": True,
    "spotlight_color": [255, 250, 240] # Linen white
  }
}
```

4.1. lista. Zavaró tényezők felrajzolási szabálya.

A generáló algoritmus minden ilyen tényező-kombináción végigiterál a következő sorrendben: vágás, tintafoltok, hajszálak, vonalzó majd keret vagy árnyék. A sorrend azért lett így meghatározva, hogy a tényezők átfedése a lehető legjobban illeszkedjen a természetes előfordulásukhoz. A generálást követően a végső adathalmaz 68 bőrelváltozásnak 157 tényező-variációját tartalmazza, ezzel közel 10.676 képpárt tartalmazó adathalmazt létrehozva. A generálás során a tényezők jelenlétének metaadata one-hot enkódolva került

rögzítésre az eredeti és generált kép azonosítója mellett, ami lehetővé teszi a tényezők hatékonyabb felismerésének biztosítását. Az adathalmaz mintáinak egy nagyobb elemszámú reprezentációja a F.1.1. mellékletben megtekinthető.

4.2. Architektúra

Az adathalmaz létrehozását a generátor és a diszkriminátor architektúrájának és hiperparamétereinek megtervezése követte. Ez kerül bemutatásra a következő alfejezetben.

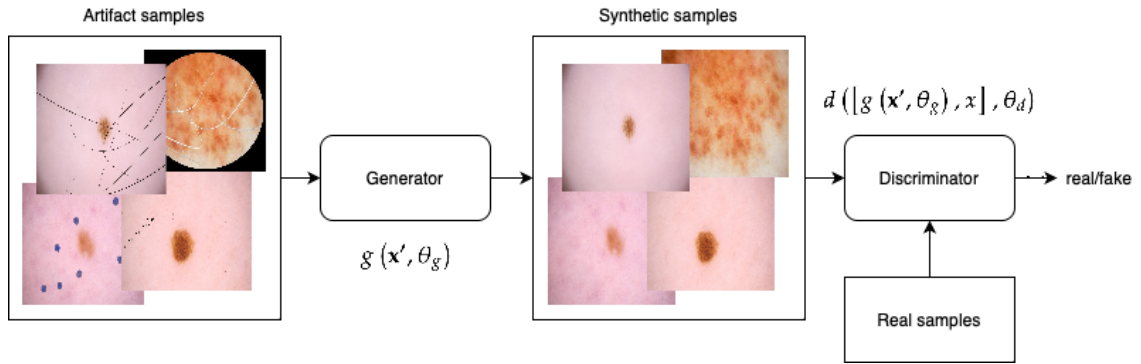
4.2.1. Generatív modell

A zavaró tényezők eltávolításához a 3. fejezetben tárgyalt generatív versengő hálózatoknak egy módosított változatát alkalmaztam. A módosítás lényege, hogy a generátor háló nem zajból állít elő egy képet, hanem már egy meglévő képből, ami tartalmaz zavaró tényezőket. A 2. fejezetben leírtak alapján a dermatoszkópos képeken előforduló szőrszálak, vonalazók és egyéb objektumok értelmezhetők egyfajta speciális zajként. Ez az intuíció a GAN optimalizációs problémáját a következőképp módosította:

$$\hat{\mathbf{x}} = g(\mathbf{x}', \theta_g), \text{ ahol } \theta_g \text{ a generátor paraméterei és } \mathbf{x}' \sim D_{\text{artifacts}} \subseteq D \text{ minta} \quad (4.3)$$

$$\mathbf{y} = d([g(\mathbf{x}', \theta_g), x], \theta_d), \text{ ahol } \theta_d \text{ a diszkr. paraméterei és } \mathbf{x}', \mathbf{x} \sim D \text{ minta} \quad (4.4)$$

$$\min_{\theta_g} \max_{\theta_d} \mathbb{E}_{\mathbf{x} \sim D} [d(\mathbf{x}, \theta_d)] - \mathbb{E}_{\mathbf{x}' \sim D_{\text{artifacts}}} [d(g(\mathbf{x}', \theta_g), \theta_d)] \quad (4.5)$$



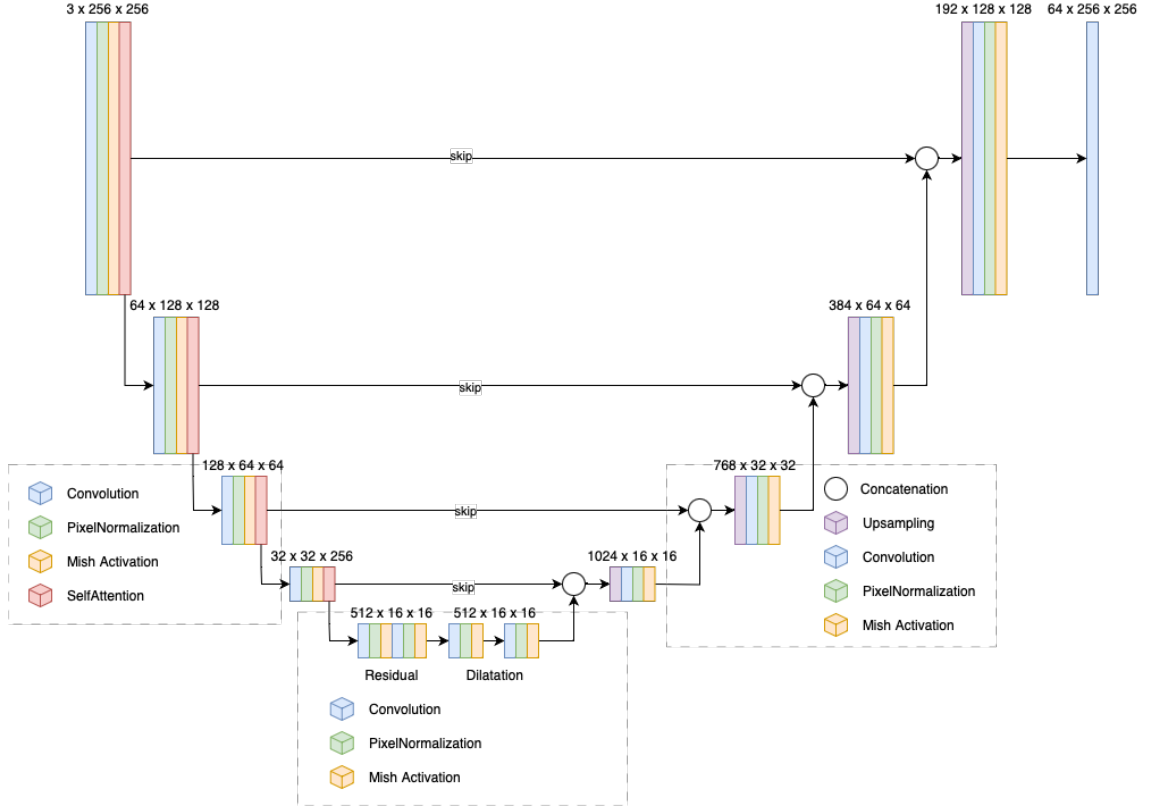
4.7. ábra. Generatív modell magasszintű áttekintése.

A 3. fejezetben volt szó arról, hogy a generátor bemenetének egy x kép is adható a z zaj helyet, ami viszont azt eredményezi, hogy a GAN determinisztikus működést mutat majd. Esetünkben ez nem probléma, ugyanis a zavaró tényezők rekonstrukciója esetén a cél az, hogy a kimenet az eredeti kép tiszta változata legyen. A generátor veszteségfüggvényének számítása is módosult annak érdekében, hogy a rekonstruált kép minél nagyobb mértékben egyezzen az eredeti tiszta képpel. A veszteségfüggvények a későbbiekben részletesebben bemutatásra kerülnek.

4.2.2. Generátor modell

A generátor esetében a cél az volt, hogy a bemenetként kapott zavaró tényezőkkel rendelkező képekről úgy távolítsa el a nem kívánt komponenseket, hogy közben a bőrelváltozások

képi jellemzői a lehető legkevésbé sérüljenek. Így a generátor architektúrája egy U-Net [40] alapú hálózat lett, ami már számos képfeldolgozási, köztük objektum eltávolítási feladatban bizonyította hatékonyságát [13, 31]. A generátor végső architektúráját a 4.8. ábra szemlélteti.



4.8. ábra. A generátor architektúrája, ahol minden blokkok felett a bemenetek dimenzionalitása látható.

A generátor hálózat egy enkóder ággal kezdődik, ami blokkról blokkra a bemeneti kép legfontosabb jellemzőit nyeri ki egyre nagyobb részletességgel. Az enkóder ág minden blokkja egy 3×3 kernelméretű és egy 2-es lépésszámú konvolúciós rétegből, egy normalizációs rétegből, egy aktivációból és végezetül figyelmi rétegből tevődik össze. A figyelmi mechanizmus esetében a Vaswani és tsai. [41] által publikált Transzformer architektúrából ismert önfigyelmi módszer (4.6) került bevezetésre. Ahol a Q lekérdezést, K kulcsot és V értéket minden dimenziócsökkentés eredményén vett 1×1 kernelű konvolúciós réteggel állítjuk elő. Az 1×1 kernel lehetővé teszi, hogy a bemeneti képből egyfajta V figyelmi térképet kapjunk. Ezt követően vesszük a Q és K tenzorok skaláris szorzatát, amit a K kulcs dimenzionalitásának gyökével normalizálunk. Majd az így kapott figyelmi súlyokon egy softmax aktivációt alkalmazunk, ami a V figyelmi térképpel összeszerozva megadja a fontos jellemzőket.

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (4.6)$$

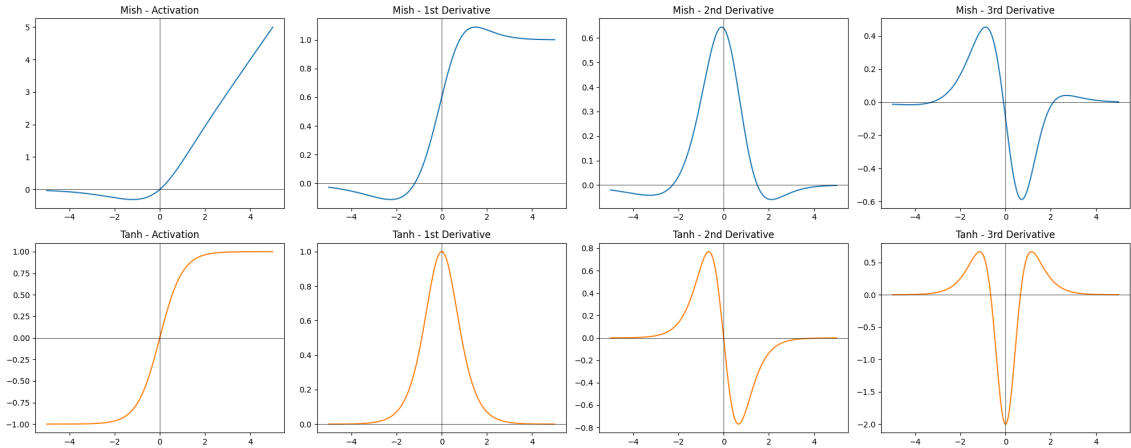
A generátor mélységenkénti csatornaszáma a modell mélységével folyamatosan megkétszereződik, ami a receptív mező megnövelésével hozzájárul ahhoz, hogy a modell képes legyen jobban megragadni a finom anatómiai információkat. Így a generátor 64-es csatornaszámmal indít, ami a mélységi szintenként először 128-ra, majd 256-ra és végül 512-re növekszik.

Kim és Hong [16] hasonlóan jártak el az enkóderblokkok csatornaszámát illetően, annyi különbséggel, hogy modelljük csak három mélységi szinten végezte a rekonstrukciót, így a legalsó rétegekben csak 256-os csatornaszámot alkalmaztak.

Az így megvalósított dimenziócsökkentést a hálózat bottleneck komponense követi, ami egy reziduális blokkból és két dilatációs blokkból áll, hogy a hálózat képes legyen a nagy területet lefedő zavaró tényezők eltávolítására is. A reziduális blokkokban 3×3 kernelméretű kétszeres konvolúciók szerepelnek, ahol minden konvolúciós művelet eredményét normalizáljuk. A reziduális és dilatációs blokkokat több szegmentációs cikkben is alkalmazták azzal a céllal, hogy növeljék a hálózat receptív mezejét, ami lehetővé teszi hogy a hálózat képes legyen megragadni globális információkat, így hatékonyabban rekonstruálva a képeket [42],[43].

Ezt követően rögtön átlépünk a dekóder ágba, ami blokkról blokkra visszaállítja az eredeti dimenzionalitására a bemenetnek a zavaró tényezőktől megtisztított változatát. Az első dekóder blokk esetében az utolsó enkóder blokk kimenete kerül összefűzésre, míg a többi enkóder blokk esetében a megfelelő szinten lévő enkóder blokk kimenete és a korábbi dekóder blokk által meghatározott magasabb szintű jellemzőkkel kerülnek katenálásra a U-Net architektúrára jellemzőt skip connection mechanizmust megvalósítva. Minden dekóder blokk esetében a rekonstrukciót egy upsampling, egy 3×3 kernelméretű konvolúciós és egy aktivációs réteg végzi. Az összefűzés miatt a dekóder blokkok magasabb bemeneti csatornaszámmal rendelkeznek, amelyek a konvolúciós rétegek kimeneti csatornáinak segítségével visszaállnak a megfelelő szintű enkóder blokkok értékeihez. A pixelértékek visszaállítása esetében egyszerű interpolációs algoritmusokkal szemben lehetett volna transzponált konvolúciós rétegeket is alkalmazni, amelyek a konvolúciós rétegekkel ellentétben, a bemenetnek egy magasabb dimenzionalitású változatát hozzák létre. Ebben az esetben a konvolúciós rétegek súlyai, az interpolációs technikához hasonló algoritmust határoznak meg, amely a modell tanítása során folyamatosan optimalizálódik. A transzponált konvolúcióval potenciálisan minőségibb rekonstruált képek hozhatók létre, viszont ezek a rétegek a modell paraméterszámát is növelik, ami a szükséges idő- és hardver erőforrásokat is növeli. Kolarik és tsai. [44] igazolták, hogy az egyszerűbb interpolációs algoritmusok hatékonyabbak mind az idő, mind az erőforrás-felhasználás szempontjából a transzponált konvolúcióval szemben. A tanulmányban a nearest neighbour módszert alkalmazó hálózat 99,47%-os pixelpontosságot ért el, és gyorsabb konvergenciát mutatott a transzponált konvolúciós rétegeket használó modellhez képest. Kundu és tsai. [45] pedig a transzponált konvolúció helyett figyelmi mechanizmust alkalmazó upsamplinget alkalmaztak, aminek paraméterszáma, a figyelmi mechanizmusa alkalmazásának ellenére is kevesebb volt a transzponált konvolúciós rétegekénél, ezzel csökkentve a hardveres erőforrásigényt és a futási időt. A két tanulmány eredményei egyértelműsítik, hogy az interpolációs upsampling rétegek összességében hatékonyabbak. Az upsampling rétegek esetében a nearest neighbor, bilinear és bicubic interpolációs módszerek választhatók az új pixelértékek kiszámításához. A réteg előnye, hogy tisztán előre definiált matematikai alapon működik, így nem vezet be új tanulható paramétereket. Mindhárom technika az új pixeleket a környező pixelek függvényében mintavételezi. A három módszer közül a nearest neighbor a leggyorsabb, ami az interpolált pixel értékeket egy konstans függvénnyel közelíti. Ennél a módszernél egy kicsit költségesebb megoldás a bilinear interpoláció, ami a környező négy pixel értékének súlyozott átlagát veszi az új pixel értékének meghatározásához. A legpontosabb, de legköltségesebb megoldás a bicubic interpoláció, ami a bilinear interpolációhoz hasonlóan a környező pixel értékek súlyozott átlagát veszi, de itt a súlyokat egy harmadfokú polinommal közelíti. A korábban említett cikkekből kiindulva a dolgozat generátora is csak nearest neighbour interpolációs módszert alkalmaz a rekonstrukció során. A generátor hálózat utolsó blokkja pedig a képet újra RGB csatornákra alakítja.

Az aktivációs függvények helyes megválasztása is kiemelten fontos a GAN tanításnak abban szakaszában, amikor az ún. backpropagation folyamat során a hálózat súlyainak gradienseivel frissülnek a hálózatok paraméterei. A neurális hálózatokban eleinte közkedvelt aktivációs függvényválasztás volt a Tanh és a Sigmoid függvény, amiknek problémája, hogy a gradiensek visszaterjesztése során exponenciálisan közelítik a nulla értéket, ha a bemenetek értékei egy széles számtartományban mozognak. Ez a jelenség vezethet az eltűnő gradiensek problémaköréhez, ami olyan mély neurális hálózatok esetében, mint a GAN generátora és diszkriminátora igen könnyedén előfordulhat, nem megfelelő aktivációs függvény választása mellett, ami ráadásul a GAN tanításának eleve is nagy instabilitással küzdő folyamatát tovább ronthatja. Az eltűnő gradiens egy olyan jelenség, amikor a gradiensek elérnek egy lokális minimumot, amiből nem tudnak tovább haladni, ezzel a hálózat tanulási folyamatát leállítva vagy instabillá téve [46]. Ezért érdemes olyan ak-



4.9. ábra. Mish, Tanh aktivációs függvények és deriváltjaik¹.

tívációs függvényeket alkalmazni, amelyeknek deriváltjai nem nullák semmilyen esetben sem. Éppen ezért a generátor esetében a Mish aktiváció került alkalmazásra, ami a bemenetén egy Softplus után egy Tanh aktivációt is alkalmaz, amivel majd beszorozza a bemenetét [47]. A 4.9 ábra látványosan szemlélteti az aktivációs függvények és deriváltjainak viselkedését és a nullához való konvergenciájuknak sebességét. A Mish aktiváció eleve rendelkezik normalizációs képességekkel, ezért lehetővé teszi, hogy kevésbé legyen szükséges olyan normalizációs rétegek alkalmazására is, mint a BatchNormalization [47]. Ezért a generátor hálózatban PixelNormalization rétegeket alkalmaztunk, ami bemenetét az RGB tartományba normalizálja, hogy minden kimenet a képekre jellemző értéktartományban maradjon. A PixelNormalization-nek nincsenek tanulható paraméterei, de a Progressive Growing GAN-ok esetében hatékonyan bizonyult alkalmazásuk [48]. A PixelNormalization rétegek kimenete a (4.7) képlet szerint számítható ki, ahol ϵ egy konstans, ami a numerikus stabilitást biztosítja.

$$y = \frac{x}{\sqrt{\mu + \epsilon}} \quad (4.7)$$

4.2.3. Diszkriminátor modell

A diszkriminátor esetében a cél az volt, hogy egy olyan architektúra kerüljön létrehozásra, ami megfelelő kihívója a 4.2.2. alfejezetben bemutatott generátornak, ezzel rákényszerítve arra, hogy minél jobb minőségű rekonstrukciókat hozzon létre.

¹A függvények képei a Pytorch könyvtár és Matplotlib könyvtárak segítségével készültek.

A hálózat mélysége és az alulmintavételezési blokkjainak csatornaszáma a modell hiperparaméterei, amik más melanóma rekonstrukciót megcélzó hálózatokhoz hasonlóan lettek konfigurálva. Kim és Hong [16] diszkriminátorával szemben, akik 5 rétegmélységben 64, 128, 256, 512 és 1024 csatornaszámú rétegeket alkalmaztak, a dolgozat diszkriminátorában csak 4 rétegmélységben, 32-től kétszereződő csatornaszámú rétegek kerültek bevezetésre, amelyek reziduális és kétszeres konvolúciós blokkokkal lettek kompenzálva. Így a háló összesen 4 reziduális alulmintavételező blokkból áll, amelyek először egy konvolúciós művelettel alulmintavételezik a bemenetet egy 4×4 kernelmérettel és 2-es stride mérettel. Ezt egy kétszeres konvolúció követi, aminek mindkét tagját normalizáljuk, mielőtt kimenetük tovább kerülne a következő réteg számára. Ezt követően a kétszeres konvolúció eredményéhez hozzáadódik az alulmintavételezett eredmény, ami egy aktivációt követően átadásra kerül a következő blokk számára. A diszkriminátort felépítő blokkok hasonlóak a VGG [49] hálózatokban alkalmazott sémához, ami konvolúciós rétegek helyett MaxPooling-ot használ a kép dimenziócsökkentéséhez, amit további két konvolúciós réteg követ. A VGG hálózattal szemben a diszkriminátor blokkjai reziduális kapcsolatokat is tartalmaznak, amik lehetővé teszik a túlillesztés elkerülését, és a hálózat tanulási folyamatát is optimálisabbá teszik [50].

A diszkriminátor esetében a LeakyReLU aktiváció lett bevezetve, ugyanis Xu és tsai. [51] különböző ReLU alapú aktivációk összehasonlításának eredményeként arra jutottak, hogy a negatív bemeneteknek egy ún. negative slope-pal (ns) való szorzása pozitívan hat a klasszifikációs eredményre. Az aktiváció számítása a (4.8) képlet szerint történik.

$$f(x) = \begin{cases} x, & \text{ha } x > 0 \\ ns \cdot x, & \text{ha } x \leq 0 \end{cases} \quad (4.8)$$

A normalizáláshoz a BatchNormalization réteg lett alkalmazva, ami a bemenetét a batchen belüli minták pixeleinek μ átlagából és σ szórásából megkapott eloszlás segítségével normalizálja a (4.9) képlet szerint, ahol α és β a réteg paraméterei, amik a tanítás során optimalizálódnak. Az ϵ paraméter pedig egy konstans, ami a numerikus stabilitást biztosítja.

$$y = \alpha \cdot \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (4.9)$$

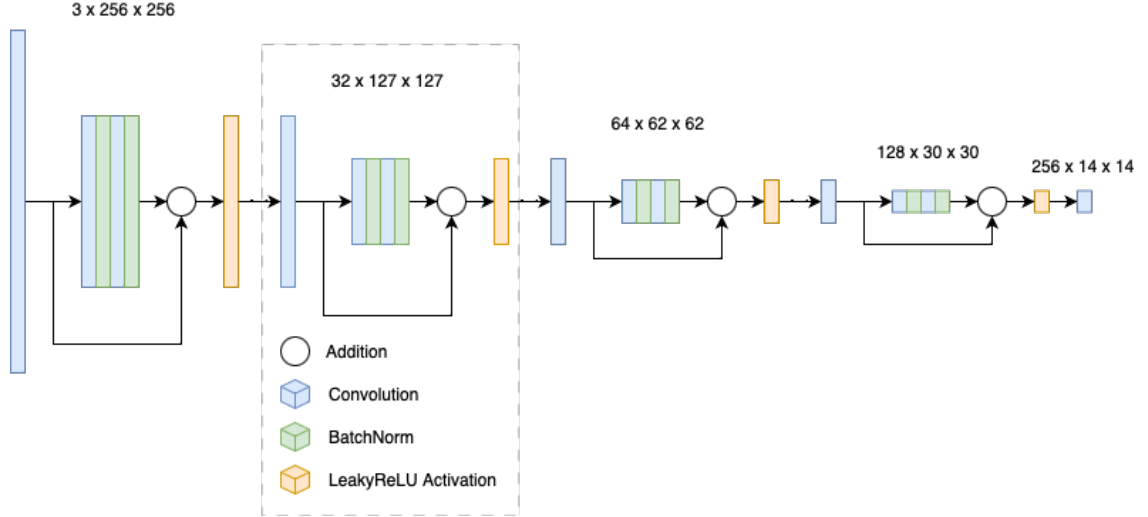
A reziduális alulmintavételezés lehetővé teszi, hogy a hálózat a rekonstruált kép finom jellemzőit – bár kisebb felbontásban –, de mélyebb struktúrák mellett legyen képes megfigyelni. A 4 alulmintavételezési szint végül egy diszkriminátor fejjel zárul, ami egy egyszerű konvolúciós rétegből áll. Ez a réteg a korábbi blokkokból származó magas szintű jellemzőket egyetlen értékre sűríti, amely jelzi, hogy a bemenet valós vagy mesterségesen generált kép-e.

4.3. Tanítás

Az architektúra megtervezését a megfelelő veszteségfüggvények és tanító algoritmus implementálása követte.

4.3.1. Veszteségfüggvények

A tanítás stabilitása mind a zavaró tényezők eltávolításában, mind a magas minőségű rekonstrukció szempontjából meghatározó. Éppen ezért a veszteségfüggvények helyes megválasztása egy kulcsfontosságú szempont. Ezért a 4.10 veszteség számításához a Wasserstein-hiba volt a kiinduló pont. A tanítás stabilitása érdekében a diszkriminátor



4.10. ábra. A diszkriminátor architektúrája, ahol minden blokk felett a bemenetek dimenzionalitása látható.

veszteségéhez egy TV regularizációs tag is hozzáadásra került, ahol λ és δ egy optimalizálandó hiperparaméter. A λ súlyt Zhang és tsai [37] ajánlásával 1-nek vettem, míg az elfogadható marginális különbséget szabályozó δ értékét $1e - 4$ -nek választottam, hogy a diszkriminátor szigorúan kezelje a különbséget.

$$\mathcal{L}_d(\hat{x}, x) = \mathbb{E}_{\hat{x}} [d(\hat{x})] - \mathbb{E}_x [d(x)] + \lambda \cdot \mathbb{E}_{\mathbf{x}, \hat{\mathbf{x}}} [|d(\mathbf{x}, \theta_d) - d(g(\hat{\mathbf{x}}, \theta_g), \theta_d) - \delta|] \quad (4.10)$$

A generátor esetében a veszteségfüggvények közül a legfontosabb a rekonstrukciós veszteség, ami a generált kép és az eredeti kép közötti L1 távolságot méri a (4.11) képlet szerint, ahol N a minták száma.

$$\mathcal{L}_{L1}(\hat{x}, x) = \frac{1}{N} \sum_{i=1}^N |\hat{x}_i - x_i| \quad (4.11)$$

Az $L1$ veszteség a két kép közötti eltérést pixel szinten méri 1-es normával, ami a képek közötti átlagos abszolút eltérést adja meg. A generátor veszteségfüggvénye a rekonstrukciós veszteség mellett tartalmazza az adversariális veszteséget is, ami a Wasserstein-hibának a rekonstruált képekre vonatkozó alakja (4.12).

$$\mathcal{L}_{\text{adversarial}} = -\mathbb{E}_{\hat{x}} [d(\hat{x})] \quad (4.12)$$

A generátor veszteségfüggvényéhez egy harmadik tag is hozzáadásra került, ami a rekonstruált és az elvárt kimenet közti hasonlóságot nem pixel szinten, hanem magasabb szintű vizualizációs jellemzők alapján méri. Az ilyen veszteséget perceptuális veszteségnek hívjuk, amit elsőként Johnson és tsai. [52] alkalmaztak a stílus átvitel problémaköréhez, ami a jelen dolgozat szempontjából is hasznos a bőrelváltozások jellemzőinek megtartásához. A perceptuális veszteség a (4.13) képlet szerint számítható, ami rekonstruált és tiszta képekre egy Φ előtanított VGG hálózat kimeneteinek különbségét méri 1-es norma szerint.

$$\mathcal{L}_{\text{perceptual}} = \frac{1}{N} \sum_{i=1}^N |\phi(\hat{\mathbf{x}}_i) - \phi(\mathbf{x}_i)| \quad (4.13)$$

A generátor vesztesége így három komponens kombinációjából tevődik össze a (4.14) képlet alapján, ahol λ_{L1} , $\lambda_{\text{adversarial}}$ és $\lambda_{\text{perceptual}}$ a hiperparaméterek, amik az egyes veszteségek fontosságát szabályozzák, amikhez fontossági sorrend szerint 100, 10 és 1-es súlyt rendeltem, hogy a generátor első körben a célkép minél pontosabb rekonstrukciójára koncentráljon.

$$\mathcal{L}_G = \lambda_{L1} \cdot \mathcal{L}_{L1} + \lambda_{\text{adversarial}} \cdot \mathcal{L}_{\text{adversarial}} + \lambda_{\text{perceptual}} \cdot \mathcal{L}_{\text{perceptual}} \quad (4.14)$$

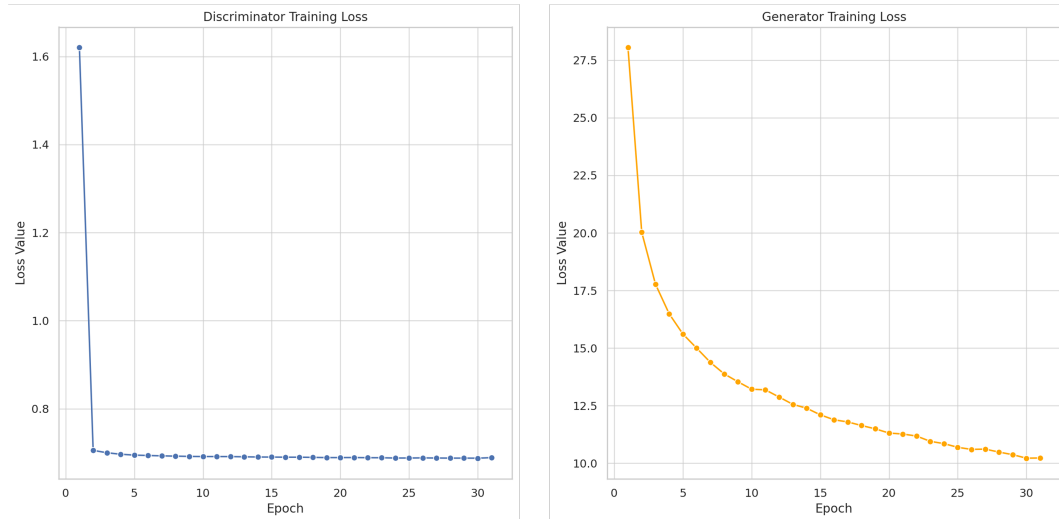
4.3.2. Tanítás folyamata

A tanításhoz először a rendelkezésre álló teljes adathalmaz 80%-a került alkalmazásra, amit további 80%-ban a tanító adatok alkotják, a maradék 20% pedig a validációra és tesztelésre lett felosztva, figyelve arra, hogy az eredeti bőrelváltozást ábrázoló képek és a zavaró tényezőkkel ellátott verziók ne lapolódjanak át a három adathalmaz között. A tanító adatokon kétszeres augmentálás is alkalmazásra került, ami a képek forgatását, tükrözését, kontrasztjának és színtartományának megváltoztatását, illetve egyéb transzformációkat jelent. Ennek köszönhetően a tanító adatok még diverzebbek lettek, ami a modell tanítási folyamatának eredményességéhez járul hozzá. A tanító adatok így közel 15.192, míg a validációs és tesztelési adatok 768 és 696 képpárt tartalmaznak.

A tanítás első szakaszának mind a 32 epoch-ja egy tanítási és egy validációs fázisból áll, amik során a modellek 4-es batch mérettel kapják a párosított mintákat. A tanítási szakaszban a diszkriminátor és a generátor felváltva tanulnak a következő módon: először a generátor rekonstruálja a zavaró tényezőkkel rendelkező képeket, majd a diszkriminátornak átadjuk a rekonstruált és valódi képeket, hogy azokat osztályozza. Ezt követően veszteséget számítunk a (4.10) képlet szerint, aminek eredményei alapján a diszkriminátor paramétereit frissítjük. A diszkriminátor tanítása során fontos, hogy a generátor paramétere ne változzanak, ezért a diszkriminátor tanítása során a generátor paramétereit befagyasztjuk. A generátor esetében nem generálunk újabb képet, hanem a diszkriminátorral osztályozzuk a rekonstruált képet, majd a veszteséget a (4.14) képlet szerint számítjuk, majd visszatértesztjük a hálózatban a gradiensket frissítve a generátor modell paramétereit. A generátor tanítása során hasonlóan fontos, hogy a diszkriminátor paramétereit ne változtassuk, ezért a diszkriminátor paramétereit befagyasztjuk.

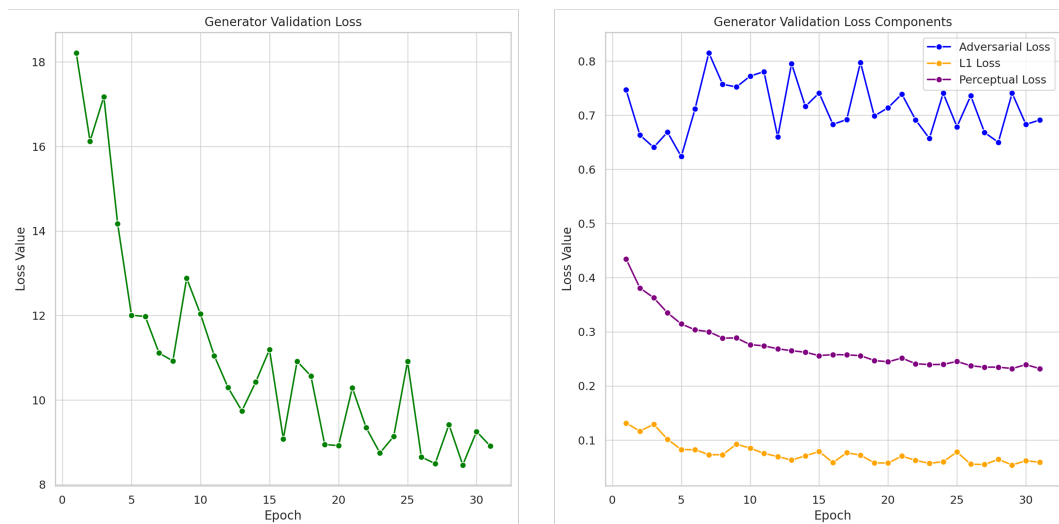
A tanítást követően a generátor validációs szakasza zajlik, amelynek során a generátor validációs adathalmazán kapott veszteség mellett az SSIM és Blocked PSNR értékek voltak mérve, amik részletesen a kiértékelésről szóló 4.4.1.1. alfejezetben vannak bemutatva. A tanítás akkor ér véget, ha a validációs veszteség nem csökken tovább egy bizonyos számú epochon belül, vagy ha elértük a maximális epoch számot. Ha a validációs veszteség javul, akkor elmentésre kerülnek a modellek, az optimalizációs és tanulási ráta ütemező paramétere, valamint a tanítás aktuális epoch száma. A modellek tanítása mindkét esetben Adam optimalizálóval történt, a generátor esetében $2e-4$, a diszkriminátor esetében pedig $4e-4$ tanulási rátával, amiknek értéke a tanítás minden tizedik epochja után csökkentve lett egy 0,5-ös faktorról, ha a veszteség nem csökkent.

A tanítás második szakasza során a dermatoszkópos és keretes esetek elhagyásra kerültek, így a teljes adathalmazból 7638 kép lett szétosztva a megfelelő adathalmazokra. Ebben az esetben ötszörös augmentáció volt alkalmazva, hogy a tanító adathalmaz mérete a korábbihoz hasonló dimenzionalitású maradjon. Így a tanító adatok száma 12.648, a tesztelési és validációs adatok pedig 333 és 351 képpárból álltak. Ezen a GAN további 26 epoch-nyit tanult, hogy a kevésbé nagy területet elfoglaló tényezőket hatékonyabban tudja eltávolítani.



4.11. ábra. A generátor és diszkriminátor veszteségének alakulása a tanítás első szakasza során.

A tanítás veszteségeinek alakulását a 4.11 ábra mutatja. A diszkriminátor vesztesége rögtön a tanítás esetén jelentősen csökken, majd stabilizálódik. Ezzel szemben a generátor vesztesége fokozatosan csökken, ami arra utal, hogy a generátor egyre jobb minőségű rekonstrukciókat hoz létre. A fokozatos csökkenés azt mutatja, hogy a generátor egyre nagyobb kihívást jelent a diszkriminátor számára. Ezt a diszkriminátor veszteségének stabilizálódása is igazolja, ugyanis abból látszik, hogy egy darabig hatékonyan tanulja a valódi és a generált képek megkülönböztetését, de ahogy a generátor fejlődik, ez a két képfajta közti eloszlás megkülönböztetése egyre nehezebbé válik a diszkriminátor számára.



4.12. ábra. A generátor veszteségének alakulása a validáció első szakasza során.

A validációs veszteség változása a 4.12 ábrán látható. A generátor validációs vesztesége instabilabb, mint a tanítás alatt, de jól látható, hogy a veszteség folyamatosan csökken. Ez a csökkenő trend azt mutatja, hogy a generátor egyre jobb minőségű rekonstrukciókat hoz létre, és nem lépett fel a túlillesztés jelensége. Ugyanezen az ábrán a validációs veszteség komponensei külön-külön is láthatók. Az adverziális veszteség az elején csökken, majd

elkezd stabilizálódni, ami azt jelenti, hogy generátor és a diszkriminátor egyensúlyban van. Ez egy pozitív jele annak, hogy egyik modell sem dominálja a másikat: a generátor képes olyan képeket előállítani, amelyek kihívást jelentenek a diszkriminátor számára, miközben a diszkriminátor még mindig elég erős ahhoz, hogy hasznos visszajelzést adjon a generátornak a hibáiról. A perceptuális veszteség fokozatosan csökken, ami azt mutatja, hogy a generált képek egyre inkább hasonlítanak a valódi képekre magas szintű jellemzők alapján. A pixelek azonosságát mérő L1 veszteség kapásból alacsony értékről indít, majd fokozatos csökkenést mutat. Ez arra utal, hogy a rekonstruált képek finom részletei is egyre inkább javulnak. A validációs veszteség komponenseinek alakulása azt is mutatja, hogy érdemes lehet az egyes veszteségek súlyait a (4.14) képletben változtatni, hogy a generátor még jobb minőségű képeket hozzon létre. A validáció során a PSNR és az SSIM metrikák is folyamatosan mérve voltak, amiknek változása az F.2.1 mellékletben látható. A PSNR és az SSIM értékek is erőteljesen növekednek, ami azt jelzi, hogy a rekonstruált képek egyre kevésbé zajosak és a zavaró tényezők kitöltése egyre inkább közelíti az elvárt képi sajátosságokat.

A tanítás folyamatát a validációs szakasz, korai megállás és tanulási ráta ütemezése nélkül az 1. algoritmus mutatja be, ami a F.2 mellékletben található.

4.4. Kiértékelés

A modell kiértékelése során a generatív modell képi rekonstrukciós képességét és a zavaró tényezők eltávolításának hatékonyságát vizsgáltam. A kiértékeléshez stratégiát és tervet kellett kidolgozni, ami kvantitatív és kvalitatív kiértékelési módszerek együttes alkalmazását tartalmazza.

4.4.1. Kiértékelési stratégia

A generatív modell kiértékelése során az egyik cél az volt, hogy átfogóan felmérjük, hogy a modell milyen mértékben képes a zavaró tényezők (szörszálak, vonalzó, keretek, tinta-foltok, dermatoszkóp által behozott jelenségek) eltávolítására a bőrelváltozások képi jellemzőinek megtartása mellett. Egy másik fontos szempont annak vizsgálata volt, hogy a tényezők hatása, hogyan befolyásolja a különböző melanóma osztályozók döntéseinek helyességét. Ehhez egy olyan átfogó kiértékelési stratégiára volt szükség, ami kvantitatív és kvalitatív kiértékelési módszereket egyaránt magába foglal.

4.4.1.1. Kvantitatív kiértékelés

A modell kvantitatív kiértékeléséhez megfelelő metrikák megválasztására volt szükség, amelyek kétféle csoportban kerültek elkülönítésre: az egyik csoport a rekonstruált képek minőségét határozza meg, míg a másik csoport a klasszifikációs modellek teljesítményre való hatását kívánja demonstrálni.

A képek minőségének megítéléséhez képgenerálási problémakörökben gyakran alkalmazott metrikák kerültek bevezetésre. A Fréchet Inception Distance (FID) esetünkben a rekonstruált képek minőségét vizsgálja úgy, hogy a generált zavaró tényezőktől mentesített képek $\hat{x} \sim \mathcal{N}(\mu_{\hat{x}}, \Sigma_{\hat{x}})$ eloszlását összehasonlítja az eredendően tiszta képek $x \sim \mathcal{N}(\mu, \Sigma)$ eloszlásával. A két képtípus eloszlásait egy Inception V3 neurális hálózat segítségével nyeri ki a képek reprezentációiból, majd a két eloszlás közötti hasonlóságot (4.15) képlet alapján számítja ki.

$$FID = \|\mu_{\mathbf{x}} - \mu_{\hat{\mathbf{x}}}\|^2 + \text{tr}(\Sigma_{\mathbf{x}} + \Sigma_{\hat{\mathbf{x}}} - 2(\Sigma_{\mathbf{x}}\Sigma_{\hat{\mathbf{x}}})^{1/2}) \quad (4.15)$$

Jelenleg a FID az egyik legelterjedtebb módszer szintetizált képek értékeléséhez, és hatékonyabban méri a hasonlóságot a vizsgált eloszlások között. Minél alacsonyabb értéket vesz fel, annál jobb a rekonstrukció minősége [53].

Önmagában a FID nem ad teljes képet a rekonstrukció hatékonyságáról. Ezért mellette olyan metrikák is alkalmazásra kerültek, mint a strukturális hasonlóság index (SSIM), és annak egy módosított változata (MS-SSIM), ami a képek strukturális hasonlóságát egy még finomabb szinten képes mérni, ami kifejezetten fontos egy olyan rekonstrukciós feladatban, ahol diagnosztikai szempontból kritikus a képek minőségének megőrzése. A SSIM a rekonstruált és az eredetileg tiszta képek hasonlóságát a fényerő, a kontraszt és a struktúrák alapján határozza meg úgy, hogy egy kernel segítségével végighalad a képeken, kiszámítva az egyes képrészletekhez tartozó pixelek átlagát (fényerő), szórását (kontraszt) és kovarianciáját (struktúra). Ezeket kiszámítva a (4.16) képlet szerint megkapható egy $[-1, 1]$ tartományba eső érték, ahol -1 a teljes eltérést, 1 pedig a teljes hasonlóságot jelenti.

$$\text{SSIM}(x, \hat{x}) = \left(\frac{2\mu_x\mu_{\hat{x}} + c_1}{\mu_x^2 + \mu_{\hat{x}}^2 + c_1} \right)^\alpha \cdot \left(\frac{2\sigma_x\sigma_{\hat{x}} + c_2}{\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2} \right)^\beta \cdot \left(\frac{\sigma_{x\hat{x}} + c_3}{\sigma_x\sigma_{\hat{x}} + c_3} \right)^\gamma \quad (4.16)$$

Ahol c_1 , c_2 és c_3 egy konstans, ami a numerikus stabilitást biztosítja. Az α , a β és a γ pedig az egyes komponensek fontosságát szabályozó konstansok. Az MS-SSIM a képek hasonlóságát egy még ennél is finomabb szinten képes mérni. Ezt úgy teszi, hogy iteratívan alulmintavételezi a képeket, majd a fényerőt kivéve az egyes szinteken kiszámítja a kontrasztra és a struktúrára vonatkozó hasonlóságot [53].

A képek rekonstrukciójakor közvetve a zavaró tényezőktől eltérő zajok is kialakulhatnak. Az ilyen zajok jelenlétének mérésére az egyik legnépszerűbb metrika a Peak Signal-to-Noise Ratio (PSNR), ami pixel szinten méri a rekonstruált képen kialakult $\hat{x}_i$ zajok és a tiszta kép megfelelő x_i pixeleinek átlagos eltérését (4.17).

$$\text{MSE}(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2 \quad (4.17)$$

Mivel a zaj és a tiszta képek viszonya nagy spektrumon mozoghat, a PSNR értékét decibelben adják meg a (4.18) képlet alapján.

$$\text{PSNR}(x, \hat{x}) = 10 \cdot \log_{10} \left(\frac{255^2}{\text{MSE}(x, \hat{x})} \right) \quad (4.18)$$

A PSNR kritikája, hogy az általa megadott érték nem tükrözi teljes mértékben az észlelt valóságot, mivel a módszer nem képes hatékonyan kezelni az olyan blokkos zajt, mint amiket a többek között a generátor modellben is alkalmazott nearest upsampling rétegek okozhatnak. Erre a problémára ad megoldást a PSNR egy továbbfejlesztése, a Blocked PSNR (PSNR-B), ami a metrika számításánál a blokkolás hatását is figyelembe veszi a teljes képre vonatkozóan (4.19) [54].

$$\text{PSNR-B}(x, \hat{x}) = 10 \cdot \log_{10} \left(\frac{255^2}{\text{MSE-B}(x, \hat{x})} \right) \quad (4.19)$$

A rekonstruált képek esetében nem csak a nem kívánt zaj kerülhet a képre, hanem könnyedén megtörténhet az is, hogy a kép által ábrázolt objektum vizuális jellemzői eltorzulnak. Ez a melanóma diagnosztika esetében igen nagy probléma, ugyanis a közönséges anyajegyektől a melanómát megkülönböztető jellemzők között a bőrelváltozások szimmetriája, széle, színe, homogenitása és mérete is szerepel. Az ilyen torzulások mérésére alkal-

mas a Spectral és Spatial Distortion Index, ami a színbeli (spektrális) és térbeli torzulásokat képes meghatározni. Ez a két metrika az SSIM-hez hasonlóan a $\hat{x}$ rekonstruált és x tiszta kép pixeleinek átlaga (fényerő), szórása (kontraszt) és kovarianciája (struktúra) alapján számít egy minőségi értéket, amik a $[0, 1]$ tartományba esnek. Ha a képek teljesen azonosak, akkor mindkét metrika értéke 0, ha a torzítás maximális, akkor pedig 1. A két torzítást a Quality With No Reference (QNR) metrika együttesen is képes mérni a (4.20) képlet alapján, ahol D_λ a spektrális torzítás, D_γ a térbeli torzítás, α és β a metrikák fontosságát szabályozó konstansok. A QNR értéke is a $[0, 1]$ tartományba esik, ahol 0 a teljes hasonlóságot, 1 pedig a maximális torzítást jelenti [55].

$$\text{QNR}(x, \hat{x}) = (1 - D_\lambda)^\alpha \cdot (1 - D_\gamma)^\beta \quad (4.20)$$

Végző soron a képi minőségre vonatkozó alkalmazott metrikák közül talán a legerősebb a Learned Perceptual Image Patch Similarity (LPIPS), ami a képek hasonlóságát egy neurális hálózat segítségével képes meghatározni. A hálózatot úgy tanították be, hogy a képek hasonlóságát emberi érzékelés alapján képes legyen megállapítani. Az LPIPS neurális hálózatokból kinyert jellemzőket használ fel a finom képrészletek összehasonlítására, ezért lehetővé teszi a képek hasonlóságának magasabb szintű meghatározását, a korábban említett metrikákkal szemben, amik főleg pixel szinten működnek. Mivel a módszer a finom részleteket vizsgálja, ezért alkalmas az emberi észlelés modellezésére. A metrika értéke a $[0, 1]$ tartományba esik, ahol a 0 a vizsgált képek finom részleteinek teljes hasonlóságát, míg az 1 pedig a maximális torzítást jelenti [56].

A zavaró tényezők eltávolításának hatása az osztályozó modellek teljesítményváltásával magyarázható, aminek mérésére a következő metrikák kerültek alkalmazásra: a pontosság (accuracy), a precizitás (precision), az érzékenység (recall), az F1 érték, a specifitás (specificity). A melanóma diagnosztikában kritikus, hogy ezek a metrikák a lehető legmagasabb értéket vegyék fel, mivel a rosszindulatú bőrelváltozások korai felismerése életmentő lehet, míg a jóindulatú bőrelváltozások rosszindulatúnak tévesztése olyan súlyos következményekkel járhat, mint a felesleges biopszia, műtétek, kezelések és stressz.

A pontosság mérésével arra kaphatunk választ, hogy a modell általánosságban mennyire képes a helyes osztályozásra. Ehhez a true positive (TP), a true negative (TN), a false positive (FP) és a false negative (FN) értékeket használjuk fel a (4.21) képlet alapján.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.21)$$

A precizitás azt mutatja meg, hogy a rosszindulatú osztályba sorolt bőrelváltozások közül milyen arányban vannak azok, amik valóban rosszindulatúak. Ez a (4.22) képlet alapján számítható ki.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.22)$$

Az érzékenység azt mutatja meg, hogy a valóban rosszindulatú bőrelváltozások közül mennyi lett helyesen azonosítva. A (4.23) képlet alapján számítható ki.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.23)$$

Az F1 érték a precizitás és az érzékenység harmonikus átlaga, ami arról hordoz információt, hogy a modell mennyire képes a rosszindulatú bőrelváltozásokat helyesen azonosítani. Ez

az érték a (4.24) képlet alapján számítható ki.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.24)$$

A specifitás azt mutatja meg, hogy a valóban jóindulatú bőrelváltozások közül mennyi lett helyesen azonosítva. A (4.25) képlet alapján számítható ki.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4.25)$$

A metrikák mellett GradCam módszer is alkalmazva volt, hogy a modell döntései mögött lévő jellemzők változása is megfigyelhető legyen. Ez a módszer a kvantitív kiértékelés mellett a kvalitatív kiértékelésben is fontos szerepet játszhat. Ugyanis a melanóma diagnosztikában a bőrelváltozások egyes vizuális jellemzői fontos támpontok a diagnosztikai döntés meghozatalában, éppen ezért fontos, hogy a modell döntése magyarázható legyen a modell által legfontosabbnak vélt jellemzők alapján.

4.4.1.2. Kvalitatív kiértékelés

A kvalitatív kiértékelés fő mutatója az ún. Mean Opinion Score (MOS), ami a rekonstruált képek minőségét emberi érzékelés alapján képes mérni. A MOS értéke a [1, 5] tartományba esik, ahol az egyes értékek jelentését a 4.1 táblázat foglalja össze.

Érték	Jelentés	Leírás
1	Nem diagnosztizálható	A kép használhatatlan, a bőrelváltozás fontos jellemzői elvesztek.
2	Nehezen diagnosztizálható	A kép diagnosztikai értéke alacsony, a bőrelváltozás jellemzői torzultak.
3	Diagnosztizálható, de kompromisszumokkal	A kép diagnosztikai értéke elfogadható, de kisebb torzulásokkal.
4	Nagyon jól diagnosztizálható	A kép diagnosztikai értéke jó, a zavaró tényezők eltávolítása sikeres.
5	Tökéletesen diagnosztizálható	A kép diagnosztikai értéke kiváló, a zavaró tényezők teljesen eltávolítva.

4.1. táblázat. Mean Opinion Score értékek jelentése

Ennek a módszernek az előnye, hogy a képek minőségét emberi érzékelés alapján képes mérni, ami a melanóma diagnosztikában kritikus fontosságú.

4.4.2. Kiértékelési terv

A kiértékelési terv a következő szakaszból áll. Először a rekonstrukció minőségét kvantitatív módszerekkel mérem, amihez elő kell állítani egy teszt adathalmazt olyan minta párokból, amelyek nem szerepeltek a tanító adathalmazban. Ezen adathalmaz segítségével vizsgálhatóak a rekonstruált képek minősége a korábban említett képi metrikák segítségével. A kvantitatív kiértékelés során a zavaró tényezők eltávolításának hatása is megvizsgálendő különböző osztályozó modell architektúrák teljesítményváltozására vonatkozó. Ehhez Kim és Hong [16] cikkében szereplő ResNet18, ResNet152, RXNet101, EfficientNet-B3 és EfficientNet-B4 modellek lettek betanítva kétféle módon. Az egyik esetben az ISIC 2020 adathalmazon tanultak a modellek, míg a másik esetben a generátor egyfajta előfeldolgozási lépésként igyekezett eltávolítani a zavaró tényezőket. A modell teljesítményének mérésére korábban említett klasszifikációs metrikák alkalmazandók.

A kiértékelés második szakaszában arra kerestem a választ, hogy a megoldásom, hogyan teljesíthet más módszerekhez képest diagnosztikai szempontból, így az 1 és a 2 hipotéziseket vizsgáltam.

1. Hipotézis (Nullhipotézis). A rekonstrukciós eredményeim nem térnek el szignifikánsan más módszerek eredményeitől, így a választások véletlenszerűek 0, 5. ■

2. Hipotézis (Ellenhipotézis). A rekonstrukciós eredményeim szignifikánsan eltérést mutatnak más módszerek eredményeihez képest. ■

A kvalitatív kiértékelés második szakaszához Kim és Hong [16], illetve Lee és You [31] rekonstrukcióit használtam fel referencia alapnak, aminek során azt vizsgáltam, hogy a dolgozat keretében implementált rekonstrukciós modell, hogyan teljesít más módszerekhez képest. Ennek megfelelően az ISIC 2020 adathalmazból kikerestem a megfelelő képeket, így összesen 21 kép rekonstrukciós eredményei kerültek összehasonlításra, amiből összesen 16 kép tartozott Kim és Hong munkájához, míg a fennmaradó 5 Lee és You kiértékelésében szerepelt. A rekonstruált és az eredeti képeket felhasználva egy kérdőívet állítottam össze, ami a 21 szakaszból állt. Minden szakaszban a zavaró tényezőkkel rendelkező képnek az általam javasolt módszerrel készített rekonstrukcióját kellett összevetni egy másik módszer eredményével, egyfajta A-B tesztet megvalósítva.

Minden A-B teszt után a jobb minőségűnek vélt mintát egy 1-5 terjedő MOS skálán is értékelni kellett a 4.1 táblázatban összefoglalt szempontok szerint. A kérdőív kitöltése során a résztvevőknek lehetőségük van a kiválasztott képek minőségének szöveges véleményezésére is, ami további információkat szolgáltat a rekonstrukció teljesítményéről. A kérdőív kitöltéséhez bőrgyógyász szakorvosokat toboroztam, így megfelelő szakértői véleményt kaphattam a rekonstrukció diagnosztikai értékéről.

A korábban említett hipotézisek vizsgálatához binomiális próbát [57] alkalmaztam, ami két lehetséges kimenet közül az egyik előfordulásának valószínűségét vizsgálja fix számú kísérlet során. A kísérleteknek egymástól függetleneknek kell lenniük és minden egyes megfigyelés két kategória valamelyikébe kell, hogy essen, amiknek valószínűsége $p_A + p_B = 1$, és a kísérletek során végig azonosnak kell maradniuk. Ezek a követelmények teljesülnek a kérdőív kitöltése során, hiszen minden bőrgyógyász egymástól függetlenül értékeli a rekonstrukciókat, és választásuk kimenetének valószínűsége a teljes kérdőív alatt 0, 5 marad az A-B teszt jellege miatt. A teszt elvégzése során azt a valószínűséget kerestem, ami a (4.26) képlet alapján számítható.

$$P(X \geq Z) = \sum_{i=Z}^k \binom{n}{i} p^i (1-p)^{n-i}, \text{ ahol } n \text{ az összes és } k \text{ a B kimenetet adó eset} \quad (4.26)$$

A binomiális próba alkalmazása így lehetővé teszi annak statisztikai elemzését, hogy a bőrgyógyászok mennyire részesítik előnyben más rekonstrukciós módszerek eredményeit az enyémmel szemben.

Ha a kiértékelés során a rekonstrukció minősége nem kielégítő, akkor a modell hibáinak okait feltárva, a modell továbbfejlesztésére is lehetőség nyílik. A kiértékelés során kapott eredmények az 5. fejezetben kerülnek ismertetésre.

5. fejezet

Eredmények

Ebben a fejezetben a dolgozat kertében végzett munka során elért eredményeket mutatom be a 4. fejezetben ismertetett kiértékelési terv alapján.

5.1. Kvantitatív kiértékelés

Először a rekonstrukció minőségét vizsgáltam, ahol kiértékelés során két fő fókusz volt. Az egyik annak vizsgálata, hogy a generátor képes-e hatékonyan eltávolítani az egyes zavaró tényezők kombinációt külön-külön, illetve együtt is. Ehhez a 4. fejezetben tárgyalt adathalmaz validációs részét további 23 részhalmazra bontottam az egyes tényezők előfordulása szerint. Ezt követően cél-rekonstruált képpárokból mértem a rekonstrukció minőségét a 4. fejezetben bemutatott képi metrikák szerint. A kiértékelt részhalmazok eredményeinek egy részét az 5.1 és 5.2 táblázatban gyűjtöttem össze. A rekonstrukció eredményei alapján megállapítható, hogy a generátor teljesítménye nagymértékben függ a zavaró tényezők típusától. A szőrszálak, a tintafoltok és a vonalzó esetében a modell kiválóan teljesít: hatékonyan eltávolítja ezeket a zavaró objektumokat, miközben megőrzi a bőrelváltozások strukturális és vizuális hasonlóságát. Ezeknél a zavaró tényezőknél a rekonstrukció során minimális torzítás és zaj keletkezik, amelyek azt jelzik, hogy a generátor képes megtartani a bőrelváltozások finom részleteit. Ez nem elmondható a keretek és dermoszkóp által felvitt zavaró tényezők esetében, amik többnyire körkörös árnyék és fényfoltok formájában jelennek meg a képeken. Ezek a zavaró tényezők a kép egy nagyobb részét fedik le, ezért a kép maradék része alapján kell rekonstruálni a kitakart területeket. Ebből kifolyólag a rekonstrukció során a kép minősége észrevehetően romlik. A bőr struktúrája torzul, magasabb a zajszint és torzítások lépnek fel. Érdekes jelenség, hogy ezeket a zavaró tényezőket más típusokkal kombinálva – például a szőrszál, keret, tintafolt kombináció (S+K+T) esetén –, a generátor teljesítménye esetenként pozitívan változik. Ez arra utal, hogy a generátor tanítása során azok képek, amik csak keretet vagy a dermoszkóp által keltett zajt tartalmaztak alulreprezentáltak voltak. Ez azért van így, mert az adatbázis generálása során azokról az egyedülálló esetekről – mint például csak egy keret jelenik meg – képenként

ID	Zavaró tényező	FID ↓	SSIM ↑	MS-SSIM ↑	PSNR ↑	PSNR-B ↑	SpeDI ↓	SpaDI ↓	QNR ↑	LPIPS ↓
S	Szőrszál	17.45	0.976	0.989	37.50	38.15	0.022	0.472	0.516	0.018
T	Tintafolt	10.98	0.986	0.992	38.65	39.35	0.016	0.466	0.526	0.017
D	Dermoszkóp	165.7	0.928	0.949	28.91	29.65	0.107	0.587	0.369	0.162
V	Vonalzó	2.540	0.995	0.999	50.64	52.22	0.196	0.464	0.526	0.003
K	Keret	153.4	0.919	0.951	32.71	43.24	0.067	0.571	0.400	0.164

5.1. táblázat. A modell rekonstrukciós képességének értékelése a különböző zavaró tényezők esetén, ahol ↑ a magasabb érték jobb, ↓ a magasabb érték rosszabb jelentésével bír.

Zavaró tényező	FID ↓	SSIM ↑	MS-SSIM ↑	PSNR ↑	PSNR-B ↑	SpeDI ↓	SpaDI ↓	QNR ↑	LPIPS ↓
S + K + T	70.00	0.934	0.975	35.38	36.06	0.022	0.557	0.443	0.086
S + V	14.92	0.981	0.992	36.95	37.50	0.021	0.467	0.522	0.022
V + T	5.962	0.991	0.997	45.04	46.04	0.012	0.467	0.523	0.007
S + D + V	112.0	0.931	0.958	29.09	29.63	0.082	0.557	0.407	0.118
D + T	172.5	0.915	0.932	28.28	29.25	0.105	0.607	0.352	0.185

5.2. táblázat. A modell rekonstrukciós képességének értékelése a különböző zavaró tényezők kombinációi esetén, ahol ↑ a magasabb érték jobb, ↓ a magasabb érték rosszabb jelentésével bír.

Szörszál lefedettség	PSNR ↑	SSIM ↑	FID ↓	LPIPS ↓
10%	43,07	0,969	16,42	0,020
30%	38,40	0,916	32,64	0,051
Random	40,80	0,942	35,50	0,037
Javasolt módszer	37,50	0,976	17,45	0,018

5.3. táblázat. Lee és You eredményeivel való összehasonlítás.

csak egy került létrehozásra. A többi esetben pedig más zavaró tényezővel együtt jelennek meg a képeken. Az eredeti adathalmazt átnézve se volt gyakori az önálló jelenléte a kereteknek bármely más zavaró tényező jelenléte nélkül. További minták generálása esetén ezen tényezők esetében a generátor teljesítménye javulhat. Lee és You [31] a saját megoldásukat különböző szörszál sűrűségi szinteken vizsgálták egy szintetikus adathalmazon. Az adathalmazukban a szörszálak reprezentációja valóságosabb, mint a dolgozat keretében létrehozott adathalmazban, ezért a két eredmény közvetlen összehasonlításához ezek a számok önmagukban nem alkalmasak. Ezért a két módszer eredményeit kvalitatívan is kiértékeltem szakértőket felkérve, aminek eredményeit a következő szakaszban tárgyalom. Lee és You eredményei azt a tendenciát mutatják, hogy a szörszálak lefedettségének növekedésével a rekonstrukció minősége romlik. A dolgozat generátorának a teljesítménye a random sűrűsűes esettel felveszi a versenyt. Bár Lee és You generátorának eredményei kevésbé zajosak, a dolgozat generátorának rekonstrukcióinak eloszlása közelebb esik a szörszálmentes változatokhoz. Az SSIM metrika értéke azt mutatja, hogy bőrelváltozások vizuális és struktúrális jellemzőit magasabb minőségben sikerült megőrizni, ami mellett az LPIPS érték pedig azt jelzi, hogy a dolgozat generátora potenciálisan az emberi észlelés esetén is jobban fog teljesíteni.

Kim és Hong [16] eredményeiket csak a klasszifikációs teljesítmény alapján vizsgálták, ezért a két módszer képi metrikákon keresztüli összehasonlítása megfelelő adatmennyiség hiányában nem volt lehetséges. Ezért én is elvégeztem a klasszifikációs teljesítményre való hatás vizsgálatát. Kim és Hong [16] az adathalmazuk összeállításának lépéseit csak leírták és nem tették elérhetővé publikusan. Ők az ISIC 2020 adathalmaz első negyedét válogatták szét szörszál jelenléte szerint manuálisan. Ennek eredményeként a tiszta osztály 5007, míg a szörszálal osztály 2989 képből állt. A kiértékeléshez ezt az adathalmazt tovább bontották 6396 tanító és 800-800 validációs és tesztelési mintára. Én is igyekeztem rekonstruálni az adathalmazukat, azonban én 5864 tiszta és 2132 szörszálal képet találtam. Így az adathalmazt leszámítva ugyanazt a tanítási stratégiát követtem. A kiértékelés végül az ISIC 2020 adatbázisnak duplikációtól mentes, jó- és rosszindulatú címkézés szerint kiegyenlített részhalmazán zajlott, ami a teljes adatbázisnak mindössze 3.4%-a. Az adathalmaz további 8:1:1 arányban lett felosztva tanítási, validációs és tesztelési célra. Így a tanítás 936, a validáció és tesztelés pedig 96-96 kép felhasználásával történt, amik minden esetben 256×256 dimezionalításra lettek méretezve, és az ImageNet adatbázis pixelértékeinek átlaga és szórása szerint normalizálva is voltak. Minden klasszifikációs modell tanítása 24 epoch-nyit futott, aminek minden negyedik epoch-jában egy validációs fázis is volt. Majd a folyamat

a teszt adatokon történő inferenciával zárult. A kiértékelési tervben megfogalmazott irányelv szerint a tanítást a rekonstrukcióval és anélkül lett elvégezve, aminek eredményeit a 5.4 táblázat mutatja be a vizsgált metrikák szerint.

Zavaró tényezők	Metrika	ResNet18	ResNet152	ResNext101	EfficientNetB3	EfficientNetB4
Eltávolítása nélkül	acc	0,8125	0,6771	0,7186	0,6771	0,6667
	auc	0,9204	0,7185	0,8110	0,6684	0,6228
	f1	0,8421	0,7395	0,7611	0,7304	0,7242
	pre	0,7272	0,6197	0,6620	0,6269	0,6176
	rec	1,0000	0,9167	0,8959	0,8750	0,8750
	spec	0,6250	0,4375	0,5417	0,4792	0,4583
Eltávolításával	acc	0,8542	0,6563	0,7083	0,6979	0,7292
	auc	0,9646	0,7554	0,7849	0,7789	0,7789
	f1	0,8727	0,7227	0,7308	0,7563	0,7547
	pre	0,7742	0,6060	0,6786	0,6338	0,6896
	rec	1,0000	0,8958	0,7917	0,9375	0,8333
	spec	0,7083	0,4167	0,6250	0,4583	0,6250

5.4. táblázat. A klasszifikációs modellek teljesítményének alakulása a tesztelés során.

Az eredmények alapján a zavaró tényezők eltávolítása változatosan hatott az egyes modellek diagnosztikai pontosságára. A ResNet18 reagált a legjobban, ahol az AUC értéke 4,8%-kal nőtt (0,9204-ről 0,9646-ra), a pontosság 5,1%-kal javult (0,8125-ről 0,8542-re), és a specifitás 13,3%-kal emelkedett. Ez azt mutatja, hogy a modell hatékonyabban ismeri fel a rosszindulatú eseteket, mivel csökken a hamis pozitív döntések száma. Az EfficientNetB4 modell szintén jelentős javulást mutatott, az AUC értéke 25,1%-kal nőtt (0,6228-ről 0,7789-re), a pontosság 9,4%-kal emelkedett, és a specifitás 36,4%-kal javult, ami azt jelenti, hogy a modell hatékonyabban tanulta meg elkülöníteni a jó- és rosszindulatú eseteket, bár a recall 4,8%-os csökkenése arra utal, hogy nőtt a hamis negatívok aránya. A ResNet152 teljesített a legrosszabbul, ahol a pontosság 3,1%-kal csökkent (0,6771-ről 0,6563-ra), az F1-érték 2,3%-kal esett vissza, és a specifitás 4,8%-kal romlott. Az AUC 5,1%-os javulást mutatott (0,7185-ről 0,7554-re), ami a többi metrika csökkenésével együtt azt jelenti, hogy a modell kevésbé preferál egy adott osztályt. Itt nagy eséllyel lehet arról szó, hogy a modell szempontjából éppen egy lokális minimumnál történt a kiértékelés. A ResNext101 modell vegyesen reagált: a specifitás 15,4%-kal javult, de a recall 11,6%-kal csökkent, ami a rosszindulatú esetek felismerésének romlását mutatja. Az EfficientNetB3 esetében az AUC 16,5%-kal nőtt, a recall 7,1%-kal javult, de a specifitás 4,4%-kal csökkent. Ez azt jelenti, hogy a modell kevesebb esetben téveszti el a ténylegesen rosszindulatú eseteket, ám sok esetben a jóindulatúakat is hamisan rosszindulatúnak osztályozza.

Összességében elmondható, hogy a zavaró tényezők eltávolítása pozitívan hatott a rosszindulatú elváltozások azonosítására, azonban a jóindulatú esetek sokszor hamisan rosszindulatúnak lettek osztályozva. A ResNet18 modell kiegyensúlyozott és megbízható tendenciákat mutatott, magas precizitás és recall értékekkel, ami ideálissá teszi a klinikai gyakorlatban való alkalmazásra további finomhangolás után egy nagyobb adatmintán.

5.2. Kvalitatív kiértékelés

A kvalitatív kiértékelést 9 bőrgyógyász közreműködésével végeztem összesen 21 mintán, amiket Kim és Hong [16] illetve Lee és You [31] is használtak kiértékelésük során. A kiértékeléshez kiválasztott minták a mellékletben láthatóak F.3.2, F.3.1 és F.3.3 ábrákon. Lee és You [31] is végeztek hasonló kiértékelést a saját megoldásukon egy 30 fős tesztelői csoporton, amiből 24 fő volt férfi és 6 nő, és mindegyikük a 20 év feletti volt. A szakértőknek megmutatott képeket a F.3.3, F.3.1 és F.3.2 ábrák mutatják be. A kvalitatív kiértékelés A-B tesztelés fázisának eredményét a 5.5 táblázat mutatja be. A 189 kísérlet-

Kitöltő	Kim és Hong	Javasolt módszer	Összes kísérlet	Kitöltő	Lee és You	Javasolt módszer	Összes kísérlet
Bőrgyógyász 1	2	14	16	Bőrgyógyász 1	2	3	5
Bőrgyógyász 2	1	15	16	Bőrgyógyász 2	1	4	5
Bőrgyógyász 3	0	16	16	Bőrgyógyász 3	3	2	5
Bőrgyógyász 4	0	16	16	Bőrgyógyász 4	0	5	5
Bőrgyógyász 5	1	15	16	Bőrgyógyász 5	1	4	5
Bőrgyógyász 6	0	16	16	Bőrgyógyász 6	0	5	5
Bőrgyógyász 7	0	16	16	Bőrgyógyász 7	4	1	5
Bőrgyógyász 8	1	15	16	Bőrgyógyász 8	4	1	5
Bőrgyógyász 9	0	16	16	Bőrgyógyász 9	0	5	5
Összesen	5	139	144	Összesen	15	30	45

5.5. táblázat. Az dolgozat módszerének eredményei a referencia módszerekkel szemben.

ből összesen 169 esetben választották a dolgozat keretében implementált B módszert. A maradék 20 esetből 5 kép tartozott Kim és Hong [16] módszeréhez, míg a fennmaradó 15 esetben Lee és You [31] módszerét választották. A binomiális próba alkalmazásával megállapítottam, hogy a bőrgyógyászok szignifikánsan Kim és Hong eredményeivel szemben a dolgozatban javasolt módszer rekonstrukcióit preferálták. A próba eredménye közel nulla lett, ami által elutasítható a nullhipotézis, és megerősíthető, hogy az orvosi szakvélemény alapján a dolgozatban bemutatott módszer rekonstrukciói jobbak diagnosztikai szempontból. A bőrgyógyászok közül volt aki megjegyezte, hogy Kim és Hong módszere esetenként olyan mértékben megváltoztatta a bőrelváltozások szerkezetét, hogy az asszimetrián kívül más dianosztikailag fontos jellemzők, mint például a homogenitás elvesztek. Ezt különösen problémának tartották a melanómák ún. incipinens eseteiben, amelyek alaphangon nehezebben diagnosztizálhatóak. Más esetben megjegyezték, hogy az eredeti képek minősége is rossz, és bár voltak esetek, amikor Kim és Hong módszere hatékonyabban eltávolította a zavaró tényezőket, mégis a dolgozat módszerét dicsérték, hogy segítette a diagnózist. Ilyen esetekben volt jellemző, hogy Kim és Hong módszere és a dolgozat rekonstrukció között jobban megoszlottak a szavazatok.

$$P(X \geq 139) = \sum_{i=139}^k \binom{144}{i} p^i (1-p)^{144-i} \approx 0 \quad (5.1)$$

Lee és You [31] eredményei alapján is sikeresen cáfoltam a nullhipotézist. Ebben az esetben 98, 22%-os konfidenciával jelenthető ki az, hogy a szakértők inkább a dolgozat módszerének rekonstrukcióit fogják választani.

$$P(X \geq 30) = \sum_{i=30}^k \binom{45}{i} p^i (1-p)^{45-i} \approx 0,017 \quad (5.2)$$

A szakértőket arra is megkértem, hogy diagnosztikai szempontok szerint értékeljék a jobbnak ítélt rekonstrukciókat egy 1-5-ig terjedő skálán. A két módszer MOS értékeit a F.2 táblázat foglalja össze. Ezekből az eredmény jól látható, hogy a szakorvosok egyáltalán nem értékelték olyan pozitívan Lee és You [31] rekonstrukcióit, mint a 30 fős tesztcsoportuk. Lee és You [31] cikkükben azonban nem írtak arról, hogy ezek a tesztelők szakértők voltak-e a melanóma diagnosztika terén. Így a megfelelő háttértudás hiányában a kapott értékeik nem feltétlenül adnak teljes képet a rekonstrukció diagnosztikai minőségéről. A dolgozatban javasolt módszer a MOS értékek alapján is jobban teljesített Lee és You [31] rekonstrukciójánál, amit a mellékletben szereplő F.2 táblázat foglal össze. Összesen 17 esetben érkezett olyan visszajelzés, ami szerint a dolgozat módszere nagyon jól vagy tökéletesen diagnosztizálható rekonstrukciókat készített. Lee és You esetében ez a szám csak 4. A MOS értékek alapján a kompromisszumos esetekben is jobban teljesített a dolgozatban bemutatott modell, és kevesebb esetben fordult elő, hogy a másik módszert választották volna, vagy rosszul diagnosztizálható eredményt adott volna.

Diagnosztikai érték	Szavazatok
Nem diagnosztizálható	1
Nehezen diagnosztizálható	8
Diagnosztizálható, de kompromisszumokkal	47
Nagyon jól diagnosztizálható	83
Tökéletesen diagnosztizálható	30
MOS	3,49 (1,94)

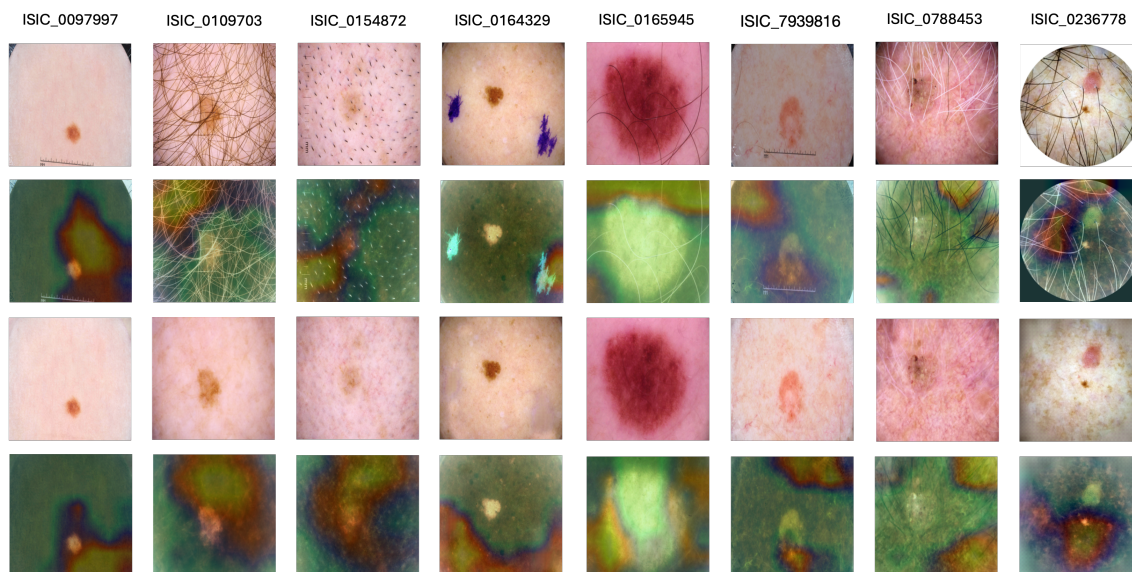
5.6. táblázat. A szakértők által adott diagnosztikai értékelések.

A teljes kísérletre vonatkozó eredményeket az 5.6 táblázat mutatja be. A dolgozat módszere esetében a szakértők összesen 21 esetben nem tudták diagnosztizálni a bőrelváltozást, vagy választották más módszer eredményét a rekonstrukció során megjelent zajok és torzítások miatt. Ezekben az esetekben jelentős veszteség keletkezett a bőrelváltozások struktúrájának olyan részeiben, amelyek fontosak a helyes diagnózis meghozásához. 8 esetben a rekonstruált képet nehezen diagnosztizálhatónak ítélték, mert a diagnózisban fontos jellemzők, mint a pigmentáció vagy a bőrelváltozás szélei elvesztek. 47 esetben a rekonstrukció során csak kisebb mértékű torzítások keletkeztek, amelyek jártak egy bizonyos mértékű információvesztéssel, de a kép diagnosztizálhatóbbá vált. Viszont a pontos diagnózishoz jobb minőségű rekonstrukcióra van szükség. A legtöbb esetben a rekonstrukciót nagyon jól diagnosztizálhatónak találták. Csak minimális torzítások keletkeztek, de ezek nem befolyásolják a diagnosztikai folyamatok hatékony elvégzését, és a bőrelváltozások jellemzői szinte változatlanok maradtak. Összesen 30 esetben a rekonstrukciót tökéletesnek találták, amelyeknél a bőrelváltozások változatlanok maradtak, és zavaró tényezőktől is mentessé váltak a képek.

A kiértékelés alapján elmondható, hogy a dolgozat modellje eredményesen alkalmazható a dermatoszkópos képek zavaró tényezőktől való megtisztítására, különösen a szőrszálak, tintafoltok és vonalazók esetében, ahol a diagnosztikai struktúrák megőrzése mellett minimális torzítás lépett fel. Bár a dermatoszkóp és a keret által okozott jelentősebb zavaró tényezők eltávolítása kihívást jelentett, ezen esetekben további minta előállítással a generátor teljesítménye potenciálisan tovább javítható. A kvalitatív és kvantitatív kiértékelés eredményei alapján a dolgozat módszere diagnosztikailag helyt áll a referencia megoldásokkal szemben.

5.3. Magyarázhatóság kiértékelése

A magyarázhatóságot a ResNet18 modell segítségével vizsgáltam, mivel ez a modell teljesített a legjobban a klasszifikációs feladatok mindkét változatában. Az eredményekről összességében elmondható, hogy a szőrszálak, a tintafoltok, a vonalazók és a kép szélei magas figyelmet kapnak a modell döntéshozatala során, sok esetben elvive a fókusz a bőrelváltozások fontos részeiről. Ezek eliminálását követően a modell döntéshozatala sokkal inkább az elváltozás és a körülötte lévő bőr képi jellemzői alapján történt. A 5.7 táblázatban jól látható, hogy a rekonstrukció a modell döntéshozatalát pozitívan befolyásolta. A FP értékek esetében a rekonstrukciót alkalmazó ResNet alacsonyabb konfidenciával állította, hogy egy bőrelváltozás rosszindulatú, mint az, ami nem alkalmazott előfeldolgozást. A rekonstrukciót alkalmazó modell a TN és TP esetekben is jobb konfidenciával hozott döntést. Érdekes, hogy voltak olyan esetek, amikor rekonstrukciós változat FN eredményt adott, míg a rekonstrukció nélküli 98.89%-os konfidenciával TP eredményt. Ebben az esetben a



5.1. ábra. A ResNet18 modell által készített Grad-CAM vizualizációk az eredeti dermoszkópos képeken (felső kétsor), és a rekonstruált változatokon (alsó két sor).

Image ID	Eltávolítás nélkül			Eltávolítással		
	Predikció	Döntés	Eredmény	Predikció	Döntés	Eredmény
ISIC_0097997	0,6141	rosszindulatú	FP	0,6867	rosszindulatú	FP
ISIC_0109703	0,0574	jóindulatú	TN	0,0021	jóindulatú	TN
ISIC_0154872	0,3085	jóindulatú	TN	0,0140	jóindulatú	TN
ISIC_0164329	0,9940	rosszindulatú	FP	0,7593	rosszindulatú	FP
ISIC_0165945	0,7990	rosszindulatú	FP	0,1479	jóindulatú	TN
ISIC_7939816	0,9984	rosszindulatú	TP	0,2200	jóindulatú	FN
ISIC_0788453	0,9893	rosszindulatú	FP	0,4691	jóindulatú	TN
ISIC_0236778	0,6844	rosszindulatú	FP	0,2578	jóindulatú	TN

5.7. táblázat. Jó- és rosszindulatú döntések predikcióinak változása a rekonstrukció alkalmazásával.

rekonstrukció nélkül működő modell magasabb mértékben támaszkodott zavaró tényezőkre, mint magára a bőrelváltozásra a döntés során.

Az eredmények azt igazolják, hogy a zavaró tényezők valóban torzítást visznek a melanóma osztályozó modellek döntéshozatalába, ezért ezek eltávolítása a megbízhatóbb és magyarázhatóbb döntésekhez szükséges.

6. fejezet

Összefoglaló

A dolgozatban egy újszerű, mélytanulás-alapú megközelítés kerül bemutatásra a dermatoszkópos képeken található zavaró tényezők—mint például vonalzó, tollal rajzolt jelzések, szőrszálak és tintafoltok—hatékony eltávolítására, anélkül hogy a bőrelváltozások fontos jellemzőit megváltoztatná. Ezek a zavaró objektumok jelentős negatív hatással vannak a neurális háló alapú osztályozók döntéseinek megbízhatóságára és magyarázhatóságára, ami különösen kritikus a melanóma diagnosztikában. A téma korszerű és nem lezárt; vizsgálata magas szintű, tudományos és elmélyült tudást igényel, hiszen a bőrrák korai felismerése életmentő lehet.

Innovációs értéként egy olyan generatív versengő hálózatot fejlesztettem ki, amely képes a zavaró tényezők eltávolítására, miközben megőrzi a bőrelváltozások diagnosztikai szempontból releváns jellemzőit. Ehhez létrehoztam egy szintetikus adathalmazt, amely 10.676 képpárt tartalmaz, különböző zavaró tényezők változatos kombinációival. Megközelítésem új a szakirodalomban, és növeli a klasszifikációs modellek robusztusságát és skálázhatóságát, mivel így az adott modellek képesek voltak különféle zavaró tényezők kombinációival is megbirkózni. A módszertan egy U-Net alapú generátorhálózatot alkalmaz, amely figyelmi mechanizmusokkal és speciális veszteségfüggvényekkel van ellátva. A veszteségfüggvények közé tartozik az L1 veszteség a pixel-szintű pontosságért, az adverzariális veszteség a GAN optimalizálásához, valamint a perceptuális veszteség a magasabb szintű jellemzők megőrzéséhez. A modell tanításához szigorú train-validation-test szétválasztást alkalmaztam, biztosítva az eredmények megalapozottságát és általánosíthatóságát.

Eredményeimet kvantitatív és kvalitatív módszerekkel értékeltem ki, amelyek egyaránt alátámasztják a módszer hatékonyságát. Kvantitatív kiértékelésem során a képi rekonstrukció minőségét olyan metrikákkal mértem, mint a FID, az SSIM, a PSNR és az LPIPS. Eredményeim azt mutatják, hogy módszerem általában jobban teljesít, mint a jelenlegi legjobb megoldások. Ezek közül kiemelném, az alkalmazási szempontból különösen fontos LPIPS metrikát a szőrszálak esetén, ami esetemben 0,018 volt, míg a sűrű szőrszálakat hatékonyan eltávolító kompetitornál [31] az 0,020-0,051 tartományban mozgott. Ahogy említettem, a zavaró tényezők eltávolításának hatását a klasszifikációs modellek teljesítményére is vizsgáltam. Ezek közül kiemelném a ResNet18, valamint az EfficientNet-B4 és B3 modelleket, amelyek esetében a legjobban teljesítő metrikák – mint az AUC és a specifitás – átlagosan 14,3%-os növekedést mutattak a zavaró tényezők eltávolítása után. Ez azt mutatja, hogy a modell jobban képes megkülönböztetni a jóindulatú és rosszindulatú elváltozásokat, ami kritikus a pontos diagnózis felállításához. A kvalitatív kiértékelés során kilenc bőrgyógyász bevonásával összehasonlítottam módszeremet a jelenlegi legjobb megoldásokkal. A szakértők szignifikánsan előnyben részesítették a rekonstrukciómát Kim és Hong [16] módszerével szemben, amit a binomiális próba eredménye is alátámasztott, a nullhipotézist közel nulla valószínűséggel elutasítottam, ami több mint 99.99%-os szigni-

fikanciaszintet jelent. Lee és You [31] módszerével szemben a megoldásom statisztikailag 98.21%-os szignifikanciaszinten bizonyult jobbnak a szakértők értékelése alapján. A MOS értékeink is magasabbak voltak, jelezve a rekonstrukciók diagnosztikai értékének kiválóságát. A bőrgyógyászok megjegyezték, hogy módszerem jobban megőrzi a bőrelváltozások fontos diagnosztikai jellemzőit, mint például a szimmetriát, a pigmentációt és a széleket. Ezen jellemzők megőrzése kritikus a melanóma korai és pontos felismerésében. Az is kiderült, hogy a módszerem kevésbé torzítja a képeket, és kevesebb zajt generál a rekonstrukció során, ami tovább növeli a diagnosztikai pontosságot.

Összességében a dolgozat újszerű ipari megoldásokat és kutatási eredményeket mutat be a dermatoszkópos képalkotás területén. A korszerű és nem lezárt téma vizsgálata magas szintű, tudományos és elmélyült munkát követelt meg a robusztusság elérésének érdekében. A bemutatott módszer nemcsak hatékonyabban távolítja el a zavaró tényezőket, mint a jelenlegi legjobb megoldások, hanem hozzájárul a neurális háló alapú osztályozók döntéseinek megbízhatóságához és magyarázhatóságához is, ezzel elősegítve a bőrrák korai és pontos diagnosztizálását.

Irodalomjegyzék

- [1] Petra Parrag, András Wéber, G. Liskay, P. Nagy, M. Kásler, C. Polgár, and I. Kennessey. Hungarian situation of melanoma incidence and mortality in the first two decades of 21st century. *Magyar onkologia*, 66 2:94–99, 2022.
- [2] Adnan R. Ahmad. Particularly convolutional neural network, improves melanoma detection. *Journal of Angiotherapy*, 8(6), June 2024.
- [3] J. Resneck, M. Abrouk, M. Steuer, A. Tam, Adam J. Yen, Ivy Lee, C. Kovarik, and Karen Edison. Choice, transparency, coordination, and quality among direct-to-consumer telemedicine websites and apps treating skin disease. *JAMA dermatology*, 152 7:768–75, 2016.
- [4] Elizabeth Chao, Chelsea K Meenan, and L. Ferris. Smartphone-based applications for skin monitoring and melanoma detection. *Dermatologic clinics*, 35 4:551–557, 2017.
- [5] J. Morand. Telemedicine in dermatology during external operations. *Medecine et sante tropicales*, 27 4:375–382, 2017.
- [6] Pálma Dobozi. A vártnál is nagyobb az érdeklődés a digitális bőrgyógyászat iránt, 09 2022. <https://semmelweis.hu/hirek/2022/06/16/a-vartnal-is-nagyobb-az-erdeklodes-a-digitalis-borgyogyaszat-irant/> [Accessed: (2024. 10. 19. 11:43)].
- [7] Titus J. Brinker, Achim Hekler, Alexander H. Enk, and ... Klode. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, 113:47–54, May 2019.
- [8] Tri-Cong Pham, Chi-Mai Luong, Van-Dung Hoang, and Antoine Doucet. Ai outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-cnn architecture with custom mini-batch logic and loss function. *Scientific Reports*, 11(1), September 2021.
- [9] Julia Winkler, Katharina Sies, Christine Fink, Ferdinand Toberer, Alexander Enk, Mohamed Abassi, Tobias Fuchs, and Holger Haenssle. Association between different scale bars in dermoscopic images and diagnostic performance of a market-approved deep learning convolutional neural network for melanoma recognition. *European Journal of Cancer*, 145:146–154, 03 2021.
- [10] Julia Winkler, Christine Fink, Ferdinand Toberer, Alexander Enk, Teresa Kränke, Rainer Hofmann-Wellenhof, Luc Thomas, Aimilios Lallas, Andreas Blum, Wilhelm Stolz, and Holger Haenssle. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA Dermatology*, 155, 08 2019.

- [11] International Skin Imaging Collaboration. About isic. <https://www.isic-archive.com/mission>, 2024. Accessed: (2024. 09. 20.).
- [12] Alceu Bissoto, Eduardo Valle, and Sandra Avila. Debiasing skin lesion datasets and models? not so fast. 2020.
- [13] Byungju Kim, Hyunwoo Kim, Kyungsu Kim, Sungjin Kim, and Junmo Kim. Learning not to learn: Training deep neural networks with biased data. 2019.
- [14] Peter J. Bevan and Amir Atapour-Abarghouei. Skin deep unlearning: Artefact and instrument debiasing in the context of melanoma classification. 2023.
- [15] Samuel William Pewton and Moi Hoon Yap. Dark corner on skin lesion image dataset: Does it matter? *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4830–4838, 2022.
- [16] Dahye Kim and Byung-Woo Hong. Unsupervised feature elimination via generative adversarial networks: Application to hair removal in melanoma classification. *IEEE Access*, 9:42610–42620, 2021.
- [17] T Lee, V Ng, R Gallagher, A Coldman, and D McLean. DullRazor: a software approach to hair removal from images. *Comput. Biol. Med.*, 27(6):533–543, November 1997.
- [18] Feng-Ying Xie, Shi-Yin Qin, Zhi-Guo Jiang, and Ru-Song Meng. PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma. *Comput. Med. Imaging Graph.*, 33(4):275–282, June 2009.
- [19] Pedro Bibiloni, Manuel González Hidalgo, and Sebastià Massanet. Skin hair removal in dermoscopic images using soft color morphology. In Annette ten Teije, Christian Popow, John H. Holmes, and Lucia Sacchi, editors, *Artificial Intelligence in Medicine - 16th Conference on Artificial Intelligence in Medicine, AIME 2017, Vienna, Austria, June 21-24, 2017, Proceedings*, volume 10259 of *Lecture Notes in Computer Science*, pages 322–326. Springer, 2017.
- [20] Reda Kasmi, Jason Hagerty, Reagan Young, Norsang Lama, Januka Nepal, Jessica Miinch, William Stoecker, and R Joe Stanley. SharpRazor: Automatic removal of hair and ruler marks from dermoscopy images. *Skin Res. Technol.*, 29(4):e13203, April 2023.
- [21] Qaisar Abbas, M E Celebi, and Irene Fondon Garcia. Hair removal methods: A comparative study for dermoscopy images. *Biomed. Signal Process. Control*, 6(4):395–404, October 2011.
- [22] Adam Huang, Shun-Yuen Kwan, Wen-Yu Chang, Min-Yin Liu, Min-Hsiu Chi, and Gwo-Shing Chen. A robust hair segmentation and removal approach for clinical images of skin lesions. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, July 2013.
- [23] Mohammad Taghi Bahreyni Toossi, Hamid Reza Pourreza, Hoda Zare, Mohamad-Hoseyn Sigari, Pouran Layegh, and Abbas Azimi. An effective hair removal algorithm for dermoscopy images. *Skin Res. Technol.*, 19(3):230–235, August 2013.
- [24] Joost Koehoorn, André C Sobiecki, Daniel Boda, Adriana Diaconeasa, Susan Doshi, Stephen Paisey, Andrei Jalba, and Alexandru Telea. Automated digital hair removal

- by threshold decomposition and morphological analysis. In *Lecture Notes in Computer Science*, Lecture notes in computer science, pages 15–26. Springer International Publishing, Cham, 2015.
- [25] A. Ashour, B. S. A. El-Wahab, Maram A. Wahba, D. Mansour, A. Hodeib, R. Khedr, and Ghada F R Hassan. Cascaded hough transform-based hair mask generation and harmonic inpainting for automated hair removal from dermoscopy images. *Diagnostics*, 12, 2022.
 - [26] Hengameh Mirzaalian, Tim K Lee, and Ghassan Hamarneh. Hair enhancement in dermoscopic images using dual-channel quaternion tubularness filters and MRF-based multilabel optimization. *IEEE Trans. Image Process.*, 23(12):5486–5496, December 2014.
 - [27] Mohamed Attia, Mohammed Hossny, Hailing Zhou, Saeid Nahavandi, Hamed Asadi, and Anousha Yazdabadi. Realistic hair simulator for skin lesion images: A novel benchmarking tool. *Artif. Intell. Med.*, 108(101933):101933, August 2020.
 - [28] Lidia Talavera-Martínez, Pedro Bibiloni, and Manuel González-Hidalgo. Hair segmentation and removal in dermoscopic images using deep learning. *IEEE Access*, 9:2694–2704, 2021.
 - [29] Dalal Bardou, Hamida Bouaziz, Laishui Lv, and Ting Zhang. Hair removal in dermoscopy images using variational autoencoders. *Skin Research and Technology*, 28(3):445–454, March 2022.
 - [30] Anas M. Ali, El-Sayed M. El-Rabaie, Khalil F. Ramadan, and Fathi E. Abd El-Samie. Enhancing dermoscopic skin cancer detection via hair artifact removal: An iterative diffusion model approach. In *2023 3rd International Conference on Electronic Engineering (ICEEM)*, page 1–4. IEEE, October 2023.
 - [31] Youngchan Lee and Wonsang You. EBAT: Enhanced bidirectional and autoregressive transformers for removing hairs in hairy dermoscopic images. *IEEE Access*, 11:14225–14235, 2023.
 - [32] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
 - [33] Christopher M. and Hugh. *Deep learning*. Springer International Publishing, Cham, Switzerland, 1 edition, November 2023.
 - [34] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks, 2017.
 - [35] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017.
 - [36] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans, 2017.
 - [37] Lijun Zhang, Yujin Zhang, and Yongbin Gao. A wasserstein gan model with the total variational regularization. *ArXiv*, abs/1812.00810, 2018.
 - [38] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014.
 - [39] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018.

- [40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.
- [42] Naga Raju Gudhe, Hamid Behravan, M. Sudah, H. Okuma, R. Vanninen, V. Kosma, and A. Mannermaa. Multi-level dilated residual network for biomedical image segmentation. *Scientific Reports*, 11, 2021.
- [43] Chen Yue, Mingquan Ye, Peipei Wang, Daobin Huang, and Xiaojie Lu. Generative adversarial network combined with se-resnet and dilated inception block for segmenting retinal vessels. *Computational Intelligence and Neuroscience*, 2022(1):3585506, 2022.
- [44] M. Kolarík, Radim Burget, and K. Říha. Upsampling algorithms for autoencoder segmentation neural networks: A comparison study. *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 1–5, 2019.
- [45] Souvik Kundu, Hesham Mostafa, Sharath Nittur Sridhar, and Sairam Sundaresan. Attention-based image upsampling, 2020.
- [46] Szemenyei Márton. Számítógépes látórendszerek, 2022.
- [47] Diganta Misra. Mish: A self regularized non-monotonic activation function, 2020.
- [48] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation, 2018.
- [49] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
- [50] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [51] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network, 2015.
- [52] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016.
- [53] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [54] Changhoon Yim and Alan Conrad Bovik. Quality assessment of deblocked images. *IEEE Transactions on Image Processing*, 20(1):88–98, 2011.
- [55] Luciano Alparone, Bruno Aiazzi, Stefano Baronti, Andrea Garzelli, Filippo Nencini, and Massimo Selva. Multispectral and panchromatic data fusion assessment without reference. *Photogrammetric Engineering and Remote Sensing*, 74:193–200, 2008.
- [56] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.
- [57] Michaela M. Wagner-Menghin. Binomial test, April 2005.