

Zárójelentés az NK101680 számú OTKA pályázathoz

A kutatási tervben meghatározott fő kutatási irányok mindegyikében (saját képi adatbázis létrehozása, vizualizáció, matematikai modellek kidolgozása és azok felhasználása dermatoszkópiai képek feldolgozására) sikerült jelentősen előrelépni, számos új elméleti és gyakorlati jellegű eredményt elértünk. A kutatási időszakban összesen 79 az OTKA pályázat által támogatott közlemény született. Eredményeinket elsősorban magasan jegyzett nemzetközi folyóiratokban, illetve neves konferenciák konferencia-köteteiben publikáltuk. Emellett dermatoszkópiai adatbázisunk (egyelőre) a Debreceni Egyetemen belül történő megnyitásával, kereshetőségével az eredmények hozzáférhetőségét is biztosítottuk, és ahhoz a napi klinikai rutinban is jól használható webes alkalmazást is fejlesztettünk.

Az alábbiakban először tömören összefoglaljuk azokat a területeket, amelyeken eredményeket értünk el. Ezután az egyes részterületekről részletesen szólnunk majd.

1. Képi adatbázis létrehozása, fejlesztése, annotációja és publikálása. Létrehoztunk egy megfelelő nagyságú, megfelelően annotált képi adatbázist. Az adatbázis számos lekérdezési és keresési funkcióval rendelkezik. Emellett egy hatékony felületet is létrehoztunk bizonyos célfeladatok elvégzésére (információk kinyerésére) különböző algoritmusok flexibilis kombinációja segítségével. Az adatbázis megfelelő klinikai annotációját is elvégeztük. Kifejlesztettünk egy olyan adatbázis-kezelőt is, amely a klinikai igényekhez igazodva teszi lehetővé az adatbázis használatát. Az adatbázist olyan protokollal egészítettük ki, hogy az a mindennapos klinikai és házi orvosi gyakorlatból származható adatgyűjtéssel is bővíthető legyen. Az elkészített orvosi képkezelő alkalmazás egy könnyen kezelhető webes megoldás a bőrgyógyászok mindennapi munkájának támogatásához, a bőrelváltozásokról készített felvételek tárolásához, letöltéséhez, elemzéséhez és megosztásához. A rendszer képes kezelni a betegek kórtörténetét a megfelelő metaadatok tárolásával együtt. Egy betegrekordhoz tetszőleges számú bőrelváltozás tartozhat, amelyek időben egymástól független vizsgálatok során követhetők. A rendszer egyaránt képes dermatoszkópos, normális és hisztopatológias képeket kezelni a hozzájuk tartozó fontosabb dermatológiai jellemzőkkel együtt. A rendszer kezelői felülete reszponzív webes felhasználó felületi elemekkel került implementálásra, ami így lehetővé teszi annak hatékony használatát mind asztali számítógépekről, mind pedig mobil eszközökről. Az adatbázist és a kezelőalkalmazást tesztüzemben a Debreceni Egyetemen belül nyilvánossá és szabadon kereshetővé tettük (alkalmazás elérhetősége: <http://www.inf.unideb.hu/~hajdua/DermDB.html>; adatbázis elérhetősége: <http://172.17.45.93:8080/imagebank/>).

2. A gyakorlati problémák matematikai modellezése. Olyan jól körülhatárolt részproblémákat foglalmaztunk meg, amelyek már matematikai értelemben is precíz kérdéseket jelentenek, és különböző matematikai eszközökkel jól modellezhetőek és vizsgálhatóak. Több esetben sikerült olyan matematikai modellt kifejlesztenünk, amelyhez implementációt is készítettünk, illetve amely a gyakorlati alkalmazások (algoritmusok) számára is hasznosnak bizonyult.

3. Vizualizációs technikák kidolgozása. Felhasználva a kidolgozott matematikai modellek nyújtotta támogatást, megadtunk egy vizualizációs technikát a melanoma vizsgálatához nagy mennyiségű adatot igénylő klinikai vizsgálatok, illetve a szavazórendszereken alapuló összetett rendszerek jobb átláthatóságának és generálhatóságának támogatásához.

4. Speciális képfeldolgozó algoritmusok fejlesztése. Részben a fent említett matematikai modellekre támaszkodva, több és többféle algoritmust fejlesztettünk és implementáltunk a dermatoszkópiai

képek feldolgozására (szegmentálás, klinikai szempontból fontos sajátságok kinyerése, képek osztályozása, stb.), többek között a pigmenthálózat detektálásával kapcsolatban.

5. *Egyéb kutatások.* A pályázat résztvevői folytatták alapkutatói tevékenységüket, szakértői területeiknek megfelelően. Ezek a kutatások is sok esetben szervesen kapcsolódnak a pályázat fő irányához.

Megemlítjük, hogy a fenti pontok között nagy az átjárhatóság és a visszacsatolás: például a vizsgált matematikai problémákat éppen a 3. és 4. pontban fellépő gyakorlati kérdések minél pontosabb, általánosabb és hatékonyabb megválaszolása motiválta.

1. Képi adatbázis létrehozása. A bőrgyógyászati betegellátásban alapvető fontosságú a páciensek bőrelváltozásainak pontos dokumentálása, melanoma szűréséhez pedig elengedhetetlen az anyajegyek változásainak időben történő rögzítése. A rendelkezésre álló direkt bőrgyógyászati képek dokumentálására szolgáló számítógépes adatbázis-kezelő szoftverek többsége csak professzionális digitális video-dermatoszkóp rendszerek részeként érhető el. Olyan központi adatbázist használó, az adott intézmény összes szakembere által hozzáférhető képi adatkezelő rendszer (Image Management System, IMS), mely célzottan bőrgyógyászati felhasználáshoz készült, tudomásunk szerint jelenleg a nemzetközi piacon sem létezik. A hazai klinikai gyakorlatban jellemző eljárás, hogy a bőrelváltozásokról készült képeket gyakran a beteg tartja magánál a következő vizsgálatig, mely ugyancsak rámutat egy ilyen rendszer szükségességére. Erre a problémára kínál megoldást a pályázat keretében fejlesztett webes alkalmazás.

Vizsgálatainkhoz nélkülözhetetlen volt egy megfelelő minőségű és nagyságú képi adatbázis összeállítása. Ennek megfelelően létrehoztunk egy olyan képadatbázist, amelynek elsődleges célja, hogy lehetővé tegye a fejlesztők és az orvos szakértők számára a rendelkezésre álló tesztadatok központi tárolását, könnyű elérését és azok céljaiknak megfelelő szűrését. A tesztképek részben a pályázat DEOEC Bőrgyógyászati Klinikán dolgozó résztvevőitől, részben validált, publikusan is elérhető, kereskedelmi forgalomban beszerezhető gyűjteményekből (Interactive CD-ROM of Dermoscopy: Instructional Exercises for Practitioners és Interactive Atlas of Dermoscopy) származnak. Az így összegyűjtött képek száma együttesen már elegendő arra, hogy az algoritmusok tesztelése során releváns eredmények álljanak elő, illetve kellően változatos bemenetet biztosítsanak. A jól felhasználhatóság érdekében az adatbázisban szereplő dermatoszkópos képek az elváltozások jellemző tulajdonságai alapján a következő címkékkel vannak ellátva: típusos vagy atípusos pigmentháló, típusos vagy atípusos globulus, típusos vagy atípusos szín. A képekhez opcionálisan további metaadatok is társíthatóak, amelyek kategorizáláshoz, kereséshez vagy a tesztelés során használhatóak fel. A fejlesztők számára egy webes felület áll rendelkezésre, amely lehetővé teszi a képek és az azokhoz kapcsolódó egyéb adatok programokból történő közvetlen elérését és manipulálását. Az alkalmazás egyaránt kényelmesen használható asztali számítógépről illetve mobil eszközökről is, ezáltal is elősegítve a mindennapi klinikai felhasználást. A rendszer eszközöket biztosít a különböző algoritmusok kombinálásához is. Erre két módon van lehetőség: azonos funkcionalitású algoritmusok kombinálása (több, önálló algoritmus szavazórendszerbe történő szervezése növeli a pontosságot), illetve különböző, de összekapcsolható funkcionalitású algoritmusok kombinálása (az algoritmusok optimális kombinációjának kiválasztásával jelentősen javítható azok pontossága). Az adatbázis folyamatosan bővíthető. Erre szükség is van, hiszen megkezdtük a naevusok és melanomák klinikai és dermatoszkópos képeinek gyűjtését a DEOEC Bőrgyógyászati Klinikán működő digitális képarchiváló rendszerben.

2. A gyakorlati problémák matematikai modellezése. Sikerült több problémát olyan precizitással megfogalmaznunk, amelyek már matematikai eszközökkel is modellezhetők, kezelhetők. Az alábbiakban csupán azokról a területekről szólnunk, ahol eredményeink kézzelfogható eredményhez

(publikáció, algoritmus, implementáció) vezettek. Megemlítjük, hogy ezeken túl több olyan (például fraktálokra alapuló) modellt sikerült felvázolnunk, amely a későbbiekben hasonló jellegű alkalmazásokhoz (illetve publikációhoz) vezethet.

1. *Többségi szavazáson alapuló szakértői rendszerek.* Több algoritmus egyidejű használata nagyobb pontosságot eredményezhet, mintha csak egy algoritmust használnánk egy adott feladatra. Az irodalomban számos ilyen jellegű rendszert leírtak, ám ezek alapos, képfeldolgozási problémákra való, matematikai igényű tárgyalása meglehetősen hiányos. Különböző valószínűségszámítási, lineáris optimalizációs illetve egyéb diszkrét és folytonos matematikai eszközökkel sikerült több elméleti eredményt bizonyítanunk klasszikus szavazórendszerek pontosságával és térbeli problémákra való kibővíthetőségével kapcsolatban. Az elvet egy orvosi képfeldolgozási problémán keresztül a gyakorlatban is alkalmaztuk. Emellett a többségi szavazás modelljének feltételtől függő általánosításával sikerült pontszerű kimeneteket képező algoritmusok eredményeinek összevonhatóságára modellt alkotnunk, amit sikeresen publikáltunk is. Eredményeinket független és függő algoritmusokra is közöltük, továbbá foglalkoztunk azzal is, hogy egy már rendelkezésre álló rendszer miként bővíthető tovább a pontosság növelését szem előtt tartva. További eredményeket is közöltünk arra nézve, hogy a modellünk hogyan pontosítható tovább súlyozott többségi szavazómodellre általánosítva. Mivel a szavazórendszerek komponensei függőségének teljes feltárása kombinatorikailag nagyon összetett, ezért kiterjesztettük a klasszikus egyszerű/súlyozott többségi szavazásnál a függőségi viszonyok becslésére használt ún. diverzitási mérőszámokat. A pontszerű kimeneteket generáló algoritmusok mellett foglalkoztunk a több pontú, illetve tartomány kimenetekkel rendelkező algoritmusok eredményeinek fúziójával is. Ezen a területen sikeresen publikáltunk algoritmusokat a klasszikus többségi szavazás elvét követve. Elkészítettünk és sikeresen publikáltunk továbbá egy olyan komplex rendszert klinikai szűrőrendszerekhez, amelyek a fúziót több szinten (képfeldolgozó algoritmusok kimenete, osztályozók kimenete) is megvalósítják.
2. *Paraméteroptimalizáció.* Szinte minden képfeldolgozó algoritmus rendelkezik egy vagy több paraméterrel (vágási küszöb, elfogadási érték, stb.). Ezeknek a paramétereknek az optimális (minél hatékonyabb futási eredményt adó) beállítása, különösen több algoritmus kombinálása esetén rendkívül nehéz feladat. Az irodalomban van néhány elszórt idevágó esettanulmány, viszont átfogó vizsgálatra nem került sor. A problémát különböző iterációs és egyéb eljárások segítségével vizsgáltuk, sikeresen alkalmazva a magas dimenziós állapottér hatékony átvizsgálásához szükséges sztochasztikus keresési technikákat (pl. szimulált hűtés). Kapcsolódó eredményeinkből egy elsősorban empirikus eredményeken alapuló publikációt is készítettünk. Mivel ez a megközelítésünk nagyon erőforrás-igényes, ezért további vizsgálatot végeztünk arra nézve, hogyan építhetők be gyorsító lépések a sztochasztikus keresőalgoritmusba annak globális optimumhoz való konvergenciáját megőrizve. Az eljárásban használt energiafüggvény zajos kiértékelésére vonatkozó irodalmi eredményekre alapozva egy mintavételezési stratégiát dolgoztunk ki az összetett rendszer pontosságának egy adott képi adatbázisra való meghatározásában. Ezen eredményeinket szintén tudományos közleményben publikáltuk.
3. *Anyajegyek rácstruktúrájának vizsgálata.* Az anyajegyek egy része rácsszerű (szabályos mintázatú) pigmentációt mutat. Sok esetben egy ilyen struktúra kóros elváltozásra utalhat – ugyanakkor ez a jelenség klinikailag még nem teljesen tisztázott. Így egy fontos és érdekes probléma egyrészt annak eldöntése, hogy egy anyajegy mutat-

e rácsszerű struktúrát, másrészt, hogy egy ilyen struktúra megléte hogyan korrelál a melanoma jelenlétével. Készítettünk egy sztochasztikus keresésen alapuló implementációt, amely időben hatékonyan (bár esetleges hibával) képes meghatározni a legjobban illeszkedő rács(ok) paramétereit. Ugyanakkor különböző diofantikus approximációelméleti eszközök segítségével megkezdtük a probléma matematikai vizsgálatát, több elméleti eredményt is igazolva. Ezen a ponton jól bemutatható a modellalkotás folyamata és nehézsége, így erről kicsit részletesebben szólnunk. A kérdés első közelítésben így tehető fel: keressünk adott R^n -beli A véges ponthalmazhoz azt minél jobban közelítő rácsot. Világos, hogy ez a kérdés egyrészt nem elég precíz, másrészt tetszőleges ponthalmazhoz található azt „nagyon jól” közelítő (sűrű) rács (bármit jelentsen is a „nagyon jó” közelítés). Így definiálnunk kellett egy adott rács esetén a közelítés „jóságát”, figyelembe véve a rács „sűrűségét”. Ez egy megfelelően megválasztott hibafüggvénnyel jól sikerült (a „sűrűséget” n dimenzió esetén a rács n -edik szukcesszív minimumának segítségével meghatározva). A modellbe azt is beépítettük, hogy valójában nem ráccsal, hanem eltolt ráccsal közelítünk (hiszen az origó helye szabadon megválasztható). A megfelelően megválasztott (és gyakorlati szempontból is használható) definíciók után megkezdhattuk a probléma matematikai vizsgálatát. Ehhez a diofantikus approximáció elmélete bizonyult a legmegfelelőbb eszköznek. Megmutattuk, hogy bármely A ponthalmaz közelíthető ráccsal úgy, hogy a közelítés hibája egy explicit módon megadott konstans alatt marad. Azt is megmutattuk, hogy (bizonyos feltételek mellett) minden egy adott konstansnál jobban közelítő rács szükségképpen egy diofantikus approximációs probléma megoldásából származik. Végül, kidolgoztunk egy az LLL-algoritmuson alapuló hatékony eljárást adott A halmazt jól közelíthető rácsok konstrukciójára. Eredményeink közlésre előkészítés alatt állnak.

3. Vizualizációs technikák kidolgozása. A pályázat témaköréhez kapcsolódva két területen folytattunk vizualizációval kapcsolatos kutatásokat. Egyrészt megvizsgáltuk, hogy a melanomához kapcsolódó nagy mennyiségű (pl. genetikai) adat milyen matematikai (fraktál)modellekkel írható le. Másrészt, modellt adtunk meg, melynek segítségével a használt szavazórendszereket tekintő komplex rendszer jól áttekinthetővé tehető.

Korábban is ismert volt, hogy léteznek a melanóma kialakulásával kapcsolatba hozható génmutációk, azonban azokat a DNS fehérjekódoló szakaszain találták. A melanómákból vett minták szekvenálásával nyert legfrissebb eredmények viszont azt mutatják, hogy a DNS nemkódoló szakaszain bekövetkező mutációknak is van jelentőségük a bőr daganatos megbetegedéseiben. A genetikai elemzések és összehasonlítások során fontos az adatok könnyen interpretálható formában történő vizualizációja, így segítve az összefüggések feltárását. Ezekhez a vizsgálatokhoz a gyakorlatban legtöbbször különböző génbőngésző alkalmazásokat használnak, amelyek szekvenciák és hozzájuk kapcsolódó annotációk megjelenítését teszik lehetővé, általában lineáris formátumban ábrázolva. Azonban nagy mennyiségű adat esetén ezek a hagyományos vizualizációs technológiák nem hatékonyak, és nem veszik figyelembe a skálázhatóság kérdését sem. Nagy mennyiségű genetikai adat vizualizációjára többek között térkitöltő fraktálgörbék alkalmazása javasolt, viszont ilyen ábrázolási mód jelenleg nem található meg egyetlen génbőngészőben sem. Ennek megoldására az egyik legszélesebb körben használt génbőngészőbe, a UCSC Genome Browser-be integráltunk egy Hilbert térkitöltő görbén alapuló vizualizációs eszközt. A Hilbert görbék alkalmazásának előnye, hogy azok önaffin viselkedése lehetővé teszi az adatok tetszőleges szintű skálázhatóságát, illetve hogy a hagyományos szekvenciális reprezentációktól jobban tükrözik a vizualizált gének térbeli elhelyezkedését. Eredményünket publikáltuk, a vizualizációt támogató

génbongészőt pedig elérhetővé tettük a Debreceni Egyetemen (az alkalmazás elérhetősége: <http://www.inf.unideb.hu/~hajdua/DermDB.html>).

Egy több, esetleg egymásra épülő szavazórendszert is tartalmazó összetett rendszer működését a felhasználók sokkal jobban tudják vizuális eszközök segítségével interpretálni. Ehhez olyan újszerű 3D-s vizualizációs megoldásokat mutattunk be, amelyek egy rendszer komponenseinek (algoritmusok, adatbázisok, energiafüggvények, illetve szavazórendszerek), azok attribútumainak, valamint lehetséges kapcsolatainak reprezentációjára szolgálnak. Az így létrehozott virtuális tér támogatja a manuális alkalmazásgenerálást a rendelkezésre álló komponensekből, illetve ezen alkalmazások kiértékelését különböző adatbázisokkal és energiafüggvényekkel. Így lehetővé válik annak vizsgálata is, hogy az egyes szavazórendszerek algoritmusainak módosítása hogyan befolyásolja a döntés pontosságát. Ezen kívül a felhasználó megismerheti azt is, hogy a rendszer által automatikusan előállított alkalmazások hogyan épülnek fel.

4. Speciális képfeldolgozó algoritmusok fejlesztése. Részben a fent említett matematikai modellekre támaszkodva, több és többféle algoritmust fejlesztettünk és implementáltunk a dermatoszkópiai képek feldolgozására (szegmentálás, klinikai szempontból fontos sajátságok kinyerése, képek osztályozása, stb.), többek között a pigmenthálózat detektálásával kapcsolatban.

Kutatásokat végeztünk dermatoszkópos felvételeken történő pigmenthálózat szegmentálásával illetve típusosságának becslésével kapcsolatban is. A pigmenthálózat típusosságának ismerete az egyik legfontosabb információ a melanoma felismerésében. Az ezzel párhuzamosan a retina érhálózat szegmentálásához kifejlesztett saját algoritmus részeként használt irányonkénti másodrendű Gauss derivált maszkokat használó módszert sikerült alkalmaznunk a pigmenthálózat kinyeréséhez is. Ez a módszer abban tér el az irodalomban létező hasonló módszerektől, hogy minden képpont esetén a különböző irányokra kapott maszkválaszokat egységesen, mint vektort használjuk, szemben azzal, hogy csak a maximális választ tekintenénk. A vektorok statisztikai mértékeiből képzett ún. „feature”-ök és „score” értékek alapján történik a szegmentálás, ahol azt a megközelítést is felhasználjuk, hogy a szomszédos pixelek válaszvektorainak bizonyos szintű hasonlóságot kell mutatniuk, melyet egy forgatás-invariáns korreláció alapú mérőszámmal fejeztünk ki. A hálózat felismerése jelenleg a hálózat által közrefogott világosabb területek, vagyis lyukak alapján történik. A lyukak alakja és egymáshoz viszonyított elhelyezkedése alapján klikkeket tudunk kialakítani, amelyekből ha egy szintnél több található a lézió területén, akkor azt a pigmenthálózat jelenlétére utalónak tekintjük. A pigmenthálózat típusosságáról való döntés egyik megközelítése lehet a hálózat szabályosságának becslése. Itt nem feltétlenül a matematikai értelemben vett szabályosság a döntő, sokkal inkább az, hogy a struktúrák mekkora eltérést mutatnak a lézió belső területén, beleértve ebbe a lyukak méretét és alakját, illetve a vonalak vastagságát és színét. A lyukak szabályosságának méréséhez a komponens konvex burkát és az ebből számított jellemzőket vehetjük alapul.

Az anyajegyek mintázottságának vizsgálata magában foglalja számos kisebb, foltoszerű elem (globulus, pont, stb.) detektálását. Ezek az elemek a háttértől eltérő intenzitású, valamelyest kör alakú foltokként reprezentálhatók, a méretük változatos lehet. Aneurizma-detektáló kutatásaink során több olyan algoritmust, illetve kombinációs technikát dolgoztunk ki, amelyek közül többet sikerrel alkalmaztunk az anyajegyben előforduló hasonló foltok detektálásához.

5. Egyéb kutatások. A fentiek mellett a pályázat résztvevői folytatták alapkutatási tevékenységüket, a megfelelő szakterületeken (számelmélet, analízis, diszkrét tomográfia, statisztika, bőrgyógyászat, orvosi képfeldolgozás). Ezekről a kutatásokról ehelyütt nincs lehetőségünk beszélni, bár azok sok esetben szorosan kapcsolódnak a pályázat témájához.