

Kísérletes modell hagyományos és digitális szkeletális röntgenfelvételek diagnosztikus pontosságának összehasonlításához

Forgács Balázs dr., Mester Ádám dr., Györke Tibor dr. és Makó Ernő dr.

Semmelweis Egyetem Radiológiai és Onkoterápiás Klinika, Budapest

Összefoglalás: Cél: hagyományos és digitális röntgen berendezések diagnosztikus pontosságának összehasonlítására használt kísérleti modell bemutatása a szkeletális radiológiai terén.

Anyag és módszer: 70, a lágyrészekről megtisztított sertés femurról, kétirányú, hagyományos és digitális felvételeket készítettünk. 35-35 véletlenszerűen kiválasztott csonton vékony fűrészszel fissurát, fracturát utánzó bemetszéseket ejtettünk, illetve 10 mm átmérőjű fúróval csonthiányos területeket állítottunk elő, amelyek litikus eltérést voltak hivatva utánozni. Egy bizonyos csont tartalmazhatott egyfajta eltérést, mindkettőt, vagy egyet sem. A felvételezés idejére a lágyrészek szimulálására a csontokat vízfürdőbe helyeztük melynek mélysége minden alkalommal azonos volt. A digitális felvételekről 1:1 arányú hardcopykat készítettünk, az értékeléskor ezek lettek felhasználva.

Az összesen 280 darab felvételt négy, egyenként 70 fel-

vételt tartalmazó csoportba osztottunk. Egy felvételecsoporton belül egy bizonyos csontról mindig csak egy kép volt található, a hagyományos/digitális illetve kóros/ép esetek aránya azonos volt az egyes csoportokon belül.

Négy értékelő négy ülésben letelezte a felvételeket, minden egyes kép esetében egy 5-ös skála segítségével adta meg külön-külön a valószínűségét fissurának illetve litikus eltérésnek. A nyert adatokból Charles Metz és munkatársai által kifejlesztett LabMrc szoftverrel készítettük el a ROC görbékét.

Eredmények: Sem az egyes értékelőktől nyert a két különböző technikára vonatkozó görbék, sem a valamennyi értékelőtől nyert összesített görbék alatti területek nagysága között nem volt szignifikáns a különbség.

Következtetés: Az általunk használt modell jól bevált a különböző eljárások pontosságának összehasonlítására, bizonyos kiegészítésekkel a jövőben tovább használható.

AN EXPERIMENTAL MODEL FOR THE ROC ANALYSIS OF THE DIAGNOSTIC ACCURACY OF CONVENTIONAL AND DIGITAL SKELETAL RADIOLOGY

Aim: to present an experimental model for the comparison of the diagnostic accuracy of conventional and digital skeletal radiology. Methods and materials: bidirectional digital and conventional X rays were performed of 70 femoral shafts of swine which were completely cleaned from soft tissues. On 35-35 randomly chosen bones we prepared cuts with a subtle saw to imitate fissures or fractures meanwhile with a drill we made holes to model lytic lesions. One bone could contain one type, both or none of the lesions. To simulate soft tissues, the bone were submerged in a standard depth water basin during the preparation of the X rays. The hardcopies of the digital studies were printed and were used for image analysis. The totally 280 X rays were divided into four groups each containing 70 studies. The groups contained always only 1 image of a certain bone. The proportion of conventional/digital and negative/positive cases within the groups was the same. Four reader interpreted the images in four session using a five grade scale to express their certainty about the existence of both type of the lesion on each image. After the collection of data the LabMrc software was used to prepare the Roc curves. Results: No significant difference could be detected either by the individual or by the curves based on the summarized data of all readers. Conclusions: Our model performed well, with some further consideration it could be useful in further studies for comparing the diagnostic performance of different systems.

Diagnosztikus módszerek pontosságának összehasonlítása során gondot okoz, hogy különböző eljárások még teljesen megegyező vizsgálati alanyok esetében is rendszerint eltérő szenzitivitást és specificitást mutatnak, miáltal a módszerek pontosságának objektív összehasonlítása nem lehetséges. A Receiver Operating Characteristic eljárás előnye abban rejlik, hogy az adatok feldolgozása után az adott módszerre jellemző görbe segítségével bármely tetszőleges specificitás esetén leolvasható az eljárás szenzitivitása, így a különböző eljárások diagnosztikus pontosságának objektív összehasonlítását teszi lehetővé. Az összetartozó értékeket koordináta-rendszeren ábrázolhatjuk, a függőleges tengelyen a szenzitivitás értéke True Positive fractio (TPF), a vízszintes tengelyen az 1-specificitás, false positive fraction (FPF) értéke ábrázolódik. Az egymáshoz tartozó specificitás-szenzitivitás párok mellett kiszámolható a görbe alatti terület nagysága – A_z – is, ami szintén használható az adott módszer pontosságának jellemezésére. A görbe előállításához szükséges referencia pontok – rendszerint 5 – meghatározásához a vizsgálatban résztvevők a vizsgált jelenség meglétét nem egyértelműen igennel vagy nemmel jellemzik, hanem egy valószínűségi skálán fejezik ki meggyőződésüket az adott jelenség előfordulásával kapcsolatban. Természetesen a ROC technika nemcsak a radiológiában használt képalkotó eljárásokra alkalmazható, hanem bármilyen vizsgálati eljárásra, a meteorológiától a radar rendszerek pontosságának meghatározásáig (1). Egy diagnosztikus eljárás pontosságának megállapításához alapvetően fontos, hogy az adott módszer mellett rendelkezésre álljon egyéb információforrás is a vizsgált objektumról, amelyet „abszolút” igazságnak fogadunk el. Egyszerű radiológiai példánál maradva egy olyan eljárás, ami egyértelmű diagnózist ad amennyiben mondjuk törés kimutatásáról van szó. Radiológia esetében ez a módszer, az „abszolút” igazság forrása lehet képalkotó módszer, műtéti, boncolási vagy szövettani lelet, illetve több szakértő együttes konszenzussal kimondott véleménye. Ez utóbbi módszer pontossága azonban maga is sok szempontból kétséges.

A radiológiában újabb és újabb képalkotó módszerek jelennek meg, illetve egyre tökéletesebb eljárások, algoritmusok, szoftverek állnak rendelkezésre, melyek ismételt felvetik az igényt az újabb és újabb technikák pontosságának meghatározására. A radiológiai eljárások sajátossága, költségességük mellett, hogy nagyrésztük ionizáló – az élő szervezetet terhelő – sugárzás felhasználásán alapszik. Noha az egyes vizsgálatok során az esetek jelentős részében az egyes egyént érő sugárterhelés csaknem elhanyagolható, azonban az egyes módszerek pontosságának összehasonlításához végzett párhuzamos vizsgálatok sok esetben feleslegesen növelik mind az egyén, mind a populáció sugárterhelését. Modell használatával ez a felesleges dózisznövekedés kiküszöbölhető.

Jelen munkánk célja egy általunk alkalmazott in vitro modell bemutatása, mely segítségével hagyományos és digitális skeletális röntgenfelvételek pontosságát hasonlítottuk össze ROC technikával.

ANYAG ÉS MÓDSZER

Hetven sertés femurról, melyek egészséges, vágóhídon levágott egyedekből származtak, hosszas detergenst tartalmazó lében való főzéssel teljesen eltávolítottuk a lágyrészeket. Mivel fiatal egyedek kerültek felhasználásra, a főzés során az epiphysealis rész is levált, így a továbbiakban csak a diaphysist használtuk fel. A megtisztított csontokat, hogy optikailag előnyösebb legyen megjelenésük. 24 órán keresztül 1%-os hidrogénperoxidos fürdőbe helyeztük, majd teljesen kiszáritottuk. Valamennyi csontsorszámot kapott. Vizsgálatainkhoz két, a skeletális diagnosztikában alapvető patológiás eltérést akartunk modellezni. Vékony fűrésszel fissurát, fracturát utánozó bemetszéseket ejtettünk, illetve 10 mm átmérőjű fűrésszel csonthiányos területeket állítottunk elő, amelyek litikus eltérést voltak hivatva utánozni (1., 2. sz. ábra). Mindkét eltérést 35-35 véletlen szerűen kiválasztott csonton végeztük el, így egy bizonyos csont tartalmazhatott egyfajta eltérést, mindkettőt, vagy egyet sem. Valamennyi csontról kétirányú felvétel készült mind hagyományos, mind digitális technikával a beállítások változtatása nélkül rögtön egymást követően. A felvételezés idejére a lágyrészek szimulálására a csontokat vízfürdőbe helyeztük, melynek mélysége minden alkalommal azonos volt. A digitális felvételekről 1:1 arányú hardcopykat készítettünk, az értékeléskor ezek lettek felhasználva.

A felvételek a csukló vizsgálatára használt paraméterekkel, Bucky rács használata nélkül készítettük el.

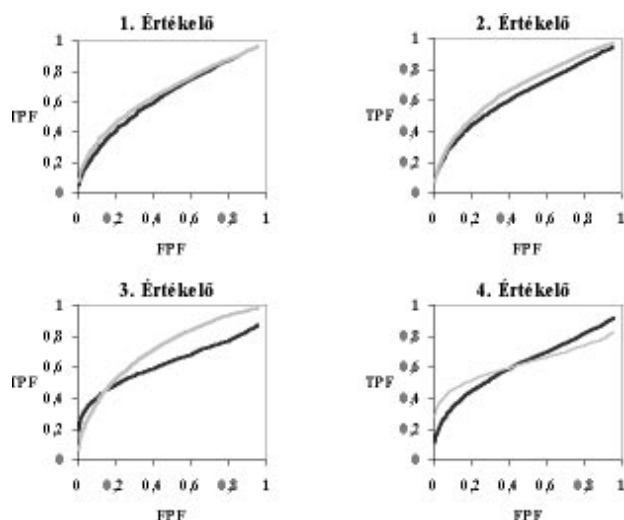
Valamennyi csont vizsgálata után összesen 280 darab felvétel – 140 hagyományos technikával készült és 140



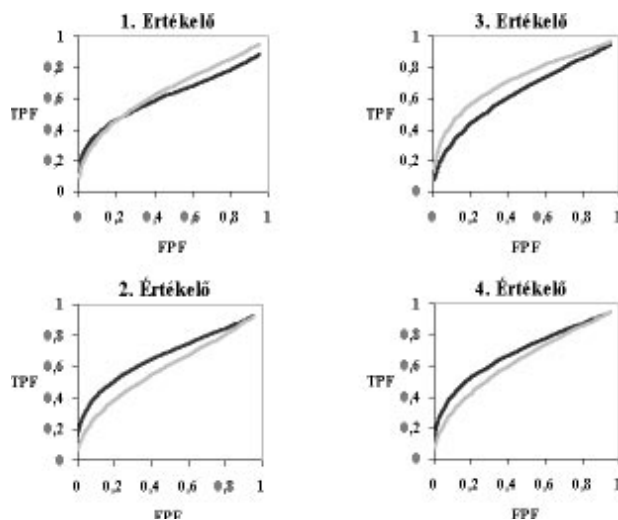
1. ábra. Litikus elváltozást utánozó lyuk a csonton és RTG képen



2. ábra. Fissurát utánozó bemetszés a csonton és RTG képen



3. ábra. Az értékelők görbéi fissura kimutatása esetén



4. ábra. Az értékelők görbéi litikus elváltozás kimutatása esetén

1. táblázat

A görbe alatti területek nagysága fissura esetében értékelőként

	hagyományos	digitális	95%-os CI A_zh-A_{zd}
1 értékelő	.6543	.6651	-.1513, .1299
2 értékelő	.6632	.6899	-.1666, .1131
3 értékelő	.6268	.7237	-.1737, .0139
4 értékelő	.6309	.6281	-.1350, .1407

CI= konfidencia intervallum, A_zh = görbe alatti terület hagyományos módszerrel,

A_{zd} = görbe alatti terület digitális módszerrel

3. táblázat

Az értékelők összesített eredménye

Elváltozás típusa	hagyományos	digitális	95%-os CI A_zh-A_{zd}
Fissura	.6483	.6777	-.0988, .0400
Litikus elv.	.6606	.6552	-.1061, .1169

CI= konfidencia intervallum, A_zh = görbe alatti terület hagyományos módszerrel,

A_{zd} = görbe alatti terület digitális módszerrel

digitális – állt rendelkezésünkre, melyeket négy, egyenként 70 felvételt tartalmazó csoportba osztottunk. Egy felvételszámra belül egy bizonyos csontból mindig csak egy kép volt található, a hagyományos/digitális illetve kóros/ép esetek aránya azonos volt az egyes csoportokon belül.

2. táblázat

A görbe alatti területek nagysága litikus elváltozás esetében értékelőként

	hagyományos	digitális	95%-os CI A_zh-A_{zd}
1 értékelő	.6223	.6546	-.1475, .0829
2 értékelő	.6770	.5970	-.0481, .2080
3 értékelő	.6447	.7278	-.1807, .0147
4 értékelő	.6984	.6414	-.0945, .2085

CI= konfidencia intervallum, A_zh = görbe alatti terület hagyományos módszerrel,

A_{zd} = görbe alatti terület digitális módszerrel

Négy értékelő négy ülésben letelezte a felvételeket, minden egyes kép esetében egy 5-ös skála segítségével adta meg külön-külön a valószínűségét fissurának illetve litikus eltérésnek, egy erre a célra elkészített kérdőív kitöltésével. A tanulási effektusok kiküszöbölésére legalább 3 hetes intervallum volt minden egyes értékelés között. A nyert adatokból Charles Metz és munkatársai által kifejlesztett LabMrc szoftverrel készítettük el a ROC görbéket. A statisztikai analízishez a jackknife módszert használtuk (2, 3, 4).

EREDMÉNYEK

Sem az egyes értékelőktől nyert a két különböző technikára vonatkozó görbék, sem a valamennyi értékelőtől nyert összesített görbék alatti területek nagysága között nem volt szignifikáns a különbség (1., 2., 3. sz. táblázat, 3., 4. sz. ábra).

MEGBESZÉLÉS

A digitális készülékek használata illetve a felvételek digitalizált tárolása és kiértékelése az utóbbi években a nagy részletgazdagságot igénylő szkeletális radiológia terén is általánosan elfogadottá vált. A hagyományos és digitális rendszerek pontosságának összehasonlítása több módon kivitelezhető: a vizsgált testrészről mind hagyományos, mind digitális felvételek készülnek, így az összehasonlítani kívánt képanyag tökéletesen megegyezik azonban a vizsgált személyt többlet dózis éri ami etikai következményeket von magával (5, 6). Az összehasonlítás elvégezhető úgy is, hogy a két különböző technikával készült felvételekből utólag válogatják ki a felvételeket, úgy, hogy azok nagyjából azonosak legyenek egymással mind a betegpopulációt, mind a patológiás eltéréseket illetően. A fenti módszer előnye, hogy nincs többlet dózis azonban a vizsgált sokaság különbözősége miatt az összehasonlítás objektivitása sem tökéletes (7). Digitalizált felvételek esetében a felvételek természetesen megegyeznek egymással (8). A vizsgálatok kísérleti modell segítségével is elvégezhetők. Felhasználhatók emberi csontok (9) – melyek alkalmazása szintén bizonyos körülményt igényel –, illetve állati eredetűek is (10). Munkánk célja volt bemutatni, egy kísérleti modellt, amely alkalmas a két rendszer pontosságának összehasonlítására, egyúttal igazolni, hogy a filmre nyomtatott digitális felvételek kiértékelése során sincs szignifikáns különbség a két eljárás között. A hardcopyk használata mellett több szempont miatt döntöttünk: a klinikusok, traumatológusok, ortopédusok továbbra is ezek használatát részesítik gyakran előnyben, számos országban jelenleg is csak a filmre nyomtatott felvétel a törvényesen elfogadott eszköz az archiválásra, a betegek a hardcopykat kapják kézhez, ezek állnak a konziliárusok rendelkezésére. Radiológiai vizsgálatok pontosságának mérésre számos érv szól mind modell használata mellett, mind ellene. Modell alkalmazásakor nem éri felesleges dózis a vizsgált alanyokat. A felvételek tetszőleges alkalommal elvégezhetők a beállítások legkisebb változása nélkül, így a minden felvételt biztosan ugyanazokat a részleteket tartalmazza. Amennyiben modellt használunk, vagyis magunk állítjuk elő azokat az elváltozásokat, melyek igazolásában a más és más eljárások pontosságát akarjuk vizsgálni nem mérülhetnek fel kétes esetek, mindig rendelkezésünkre áll az az abszolút diagnosztikus igazság, amely elengedhetetlen kelléke a különböző módszerek összehasonlításának. Ugyanakkor a modell sajátos tulajdonsága, hogy csak korlátozottan képes a valóságot utánozni, mindig egyszerűíti az élő szervezet bonyolultságát, annak csak bizonyos aspektusairól ad többé-kevésbé hű képet, például esetünkben a csontszerkezet és az izomzat modellezése kielégítő volt, de a csontthártya nem volt reprodukálható.

Az általunk nyert görbék alatti területek nagysága alacsonyabb volt az irodalomban általában olvashatóknál (5, 7, 9, 10). Ez több okra vezethető vissza. A használt csontok fiatal állatokból származtak, a csontszerkezet még meglehetősen éretlen volt így az elváltozások sok esetben még azok pontos ismeretében is nehezen voltak identifikálhatók, illetve több esetben egyáltalán nem ábrázolódtak. Az emberi csontoktól, eltérő megjelenésű, csak a diaphysist ábrázoló csontok vizsgálat szokatlan

volt az értékelők számára. Ugyancsak nehézséget okozott némelykor a diagnosztikus kritériumok következetes használata, ami elengedhetetlen feltétele a helyes ROC analízisnek.

A görbék két esetben keresztezték egymást, ami elméletileg nem megengedhető: ha két módszer között különbség van, akkor a pontosabb eljárás szenzitivitásának valamennyi specifitás érték mellett magasabbnak kell lennie. A fentiek miatt érdemesebb a görbék grafikai megjelenése mellett az összehasonlítás során a görbe alatti terület nagyságát figyelembe venni.

Mindezek a vizsgálat értékét korlátozó tényezők mellett azonban a kísérlet sikeresnek tekinthető, habár a diagnosztikus pontosság mindkét eljárással alacsony volt azonban szignifikáns különbség nem volt, a két felvételi technika használata között az értékelők szubjektíve sem találtak különbséget.

KÖVETKEZTETÉSEK

Összességében az általunk alkalmazott modell használhatónak bizonyult. A jövőben az objektumok gondosabb kiválasztásával – idősebb, érettebb csontszerkezetű egyedek –, illetve a kiértékelendő felvételek gondos előszelektálásával – csak olyan felvétel használata, amely vagy negatív, vagy a modell készítője számára az eltérés ismeretében mindenképpen látható, egyértelmű elváltozást tartalmaz – tovább növelhető a modell használhatósága.

IRODALOM

1. Sweets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-1292
2. Dorfmann DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis-generalization to the population of readers and patients with the jackknife method. *Investigative Radiology* 1992;27:723-731
3. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagnostic Radiology* 1982;143:29-36
4. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology* 1989; 24:234-245
5. Wilson AJ, Mann FA, Murphy WA, Monsees BS, Linn MR. Photostimulable phosphor digital radiography of the extremities: diagnostic accuracy compared with conventional radiography. *AJR* 1991; 157: 533-538
6. Müller RD, Buddenbrock B, Kock HJ, et al. Digitale Lumineszenzradiographie (DLR) zur Skelettdiagnostik in der Traumatologie. *RöFo* 1991; 154: 575-581
7. Klein HM, Wein HJ, Langen KH, et al. Frakturdiagnostik mit der digitalen Lumineszenzradiographie. *RöFo* 1991; 154: 582-586
8. Wegryn SA, Piraino DW, Richmond BJ, et al. Comparison of digital and conventional musculoskeletal radiography: an observer performance study. *Radiology* 1990; 175: 225-228
9. Prokop M, Galanski M, Oestmann JW, et al. Storage phosphor versus screen-film radiography: Effect of varying exposure parameters and unsharp mask filtering on the detectability of cortical bone defects. *Radiology* 1990; 177: 109-113
10. Strotzer M, Gmeinwieser J, Spahn M, et al. Amorphous silicon, flat panel, X-ray detector versus screen-film radiography: Effect of dose reduction on the detectability of cortical bone defects and fractures. *Investigative Radiology* 1998;33: 33-38