

Budapesti Műszaki Egyetem, Kandó Kálmán Villamosipari
Műszaki Főiskola, Weil Emil kórház

Egy diagnosztikus célt szolgáló tanulóprogram

Dr. Fenyő István, Dr. Sima Dezső, Dr. Siminszky Máriaó

1. Bevezetés

Az alább ismertetendő tanulóprogramot leukoplákiás betegek korai carcinoma veszélyeztettségének differenciál diagnosztikai céljaira fejlesztettük ki, de természetesen más betegségfajták diagnosztikai céljaira is felhasználható bizonyos, alább ismertetendő feltételek fennállása esetén. Nagy előnye minden eddig alkalmazott számítógépes kiértékelő eljárással szemben, hogy ha a tanulóprogramot egyszer már elkészítettük, akkor egy konkrét esetben a döntéshez nem szükséges számítógép vagy speciális célgép, a döntést egy zseb kalkulátoron is meghatározhatjuk.

2. Az eljárás lényege

Minden beteget egy-egy tünetvektorával jellemzünk. Ha a diagnózist n -tünet felhasználásával végezzük, akkor minden beteghez egy n -dimenziós tünetvektort rendelünk hozzá. Eljárásunk akkor alkalmazható, ha azonos kórképet

mutató betegek /esetünkben carcinomás illetve nem carcinomás/ tünetvektorai közel vannak egymáshoz és az egyik és másik csoportba tartozó betegeknek megfelelő n -dimenziós vektorok egy n -dimenziós polyéderrel egymástól elválaszthatók. Ha ez a feltétel teljesül, akkor a számítógépes feladat tulajdonképpen az elválasztó polyéder meghatározása. Ha ez már megvan, akkor egy ismeretlen csoportba tartozó beteg tünetvektoráról csak annyit kell megállapítani, hogy az a polyéder melyik oldalára esik, amivel a helyes diagnózist nagy valószínűséggel megállapítottuk. Az eljárás használhatóságának további feltétele, hogy csupán két egymást kizáró alternatíva, diagnózis lehetséges legyen.

3. Az elválasztó sík

A két halmazt, melyekre betegpopulációinkat szétválasztani kívánjuk, jelöljük E és R -el /ezek felelnek meg a két egymást kizáró alternatívának, illetve diagnózisnak/. Ezekről feltesszük, hogy diszjunktak, azaz $E \cap R = \emptyset$. A fent leírt eljárás akkor látszik célszerűnek és egyszerűnek, ha az előző pontban leírt nyílt polyéder egyetlen n -dimenziós hypersikből áll. Ha ez nem így van, akkor az eljárás hosszadalmassá és bonyolulttá válhat és esetünkre való gazdaságos használhatóságát kérdésessé teheti.

Feltételezzük tehát, hogy létezik egy S hypersík, mely az E és R halmazokat elválasztja. S -et elválasztó

siknak nevezzük /1. ábra/.

Az S hypersikot az archivumban lefektetett betegség tünetvektoraiból állapítjuk meg. Egy újabb B betegnél a tünetvektor meghatározása után a diagnózis felállítása abban áll, hogy megállapítjuk vajon B az E vagy az R halmazba tartozik-e. Ez pedig egyenértékű az-
zal, hogy megállapítjuk, hogy B tünetvektora az S elválasztó sík egyik vagy másik oldalára esik-e. Ha sikerül S egyenletét normálalakban meghatározni, akkor B tünetvektorát ebbe behelyettesítve eredményül vagy pozitív, vagy negatív számot, vagy pedig nullát kapunk. Ha az eredmény pozitív, akkor B az E csoportba tartozik, ha pedig 0 /vagy nagyon közel 0 /, akkor B a síkon van, tehát döntés nem születik. Ez utóbbi előjelmeghatározás zseb kalkulátoron is elvégezhető, ebben látjuk eljárásunk nagy előnyét.

Ha az E és R halmazok nem teljesítik az 1. ábrán látható feltételeket, akkor eljárásunk vagy nem, vagy nem gazdaságosan alkalmazható /2., 3. ábra/.

4. Algorithmus az elválasztó sík meghatározására

Abból indulunk tehát ki, hogy E és R halmazaink teljesítik azokat a feltételeket, melyeket az 1. ábra szemléltet.

Első lépésként megkeressük az E és R halmazok

egymáshoz legközelebbi pontjait, ezek által meghatározott szakasz felezőpontja legyen O /4. ábra/. Ezután egy eltolásos koordináta transzformációt hajtunk végre úgy, hogy O legyen az új koordináta középpont. A keresett S hypersík nyilván O -n megy keresztül. /Ha több pontpár létezik, melyeknek egymástól való távolsága minimális, akkor bármelyik pár felezőpontja megfelel új origónak./

S meghatározására szolgáló algoritmus most már a következő: A szóbanforgó betegpopulációhoz tartozó tünetvektorok legyenek x_1, x_2, \dots, x_N , ezek mindegyikéhez hozzárendeljük a p_1, p_2, \dots, p_N számokat. $p_k = 1$, ha $x_k \in E$, $p_k = -1$, ha $x_k \in R$ ($k = 1, 2, \dots, N$).

Válasszunk egy tetszőleges n_1 vektort és képezzük azt az S_1 hypersíkot, melynek egyenlete $(x, n_1) = 0$ /(...) a skaláris szorzat/. Most sorban képezzük az (x_k, n_1) számokat ($k = 1, 2, \dots, N$). k_1 legyen az első index, melyre (x_{k_1}, n_1) nem előjelhelyes /tehát $(x_{k_1}, n_1) > 0$ holott $x_{k_1} \in R$ vagy $(x_{k_1}, n_1) < 0$, holott $x_{k_1} \in E$ /. Ekkor n_1 helyébe $n_2 = n_1 + p_{k_1} x_{k_1}$ lép.

Következő lépés: Képezzük azt az S_2 hypersíkot, melynek egyenlete $(x, n_2) = 0$. Ezzel megismételjük a fenti vizsgálatot, s.i.t. Véges sok módosító lépés után megkapjuk az elválasztó hypersík egyenletének normál alakját /5. ábra/.

5. Egy szemléltető számpélda

A 4. alatt ismertetett eljárást egy számpéldán illusztráljuk. Legyen most $n=3$ és a tünetvektorok

$$\begin{aligned} r_1 &= (2; 2; 2) & e_1 &= (3; -4; 7) \\ r_2 &= (4; 4; 4) & e_2 &= (6; -6; 10) & N=8 \\ r_3 &= (6; 9; 10) & e_3 &= (4; -2; 5) \\ r_4 &= (5; 13; 8) \\ r_5 &= (5; -3; 10) & r_k &\in R \ (k=1, 2, 3, 4, 5), \quad e_k \in E \ (k=1, 2, 3) \end{aligned}$$

Annak érdekében, hogy viszonylag kevés számolással meghatározzuk az elválasztó síkot, meghatározzuk mindkét halmaz súlypontját és az S_1 sík normálisaként e két pont összekötő egyenesét vesszük:

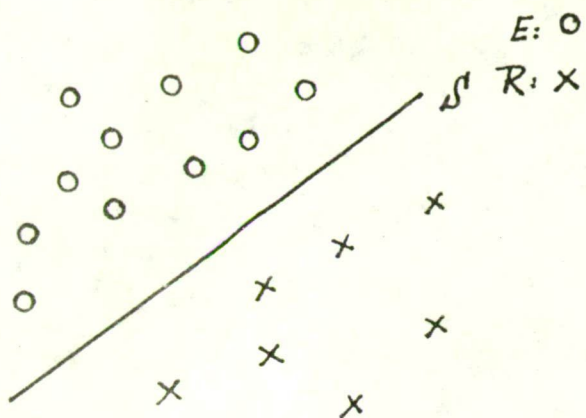
$$\underline{C}_R = (4, 4; 5; 6, 8), \quad \underline{C}_E = (4, 53; -4; 7, 33).$$

$$n_1 = \underline{C}_E - \underline{C}_R = (-0, 07; -9; 0, 53)$$

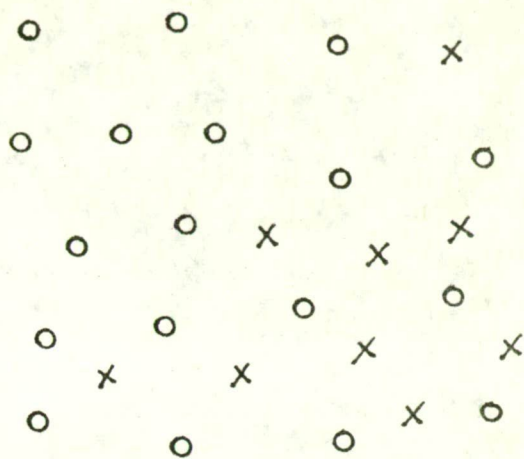
$$\begin{aligned} (r_1, n_1) &= -0, 14 - 18 + 1, 06 < 0 & (e_1, n_1) &= -0, 21 + 36 + 3, 71 > 0 \\ (r_2, n_1) &= -0, 28 - 36 + 2, 12 < 0 & (e_2, n_1) &= -0, 42 + 54 + 5, 3 > 0 \\ (r_3, n_1) &= -0, 42 - 81 + 5, 3 < 0 & (e_3, n_1) &= -0, 28 + 18 + 2, 65 > 0 \\ (r_4, n_1) &= -0, 35 - 117 + 4, 24 < 0 \\ (r_5, n_1) &= -0, 35 + 27 + 5, 3 > 0 \quad // \end{aligned}$$

Nem előjelhelyes az (r_5, n_1) skaláris szorzat. A következő lépés n_1 helyett az $n_2 = n_1 - r_5 = (-0, 07; -6; -9, 47)$ normálist venni. Ekkor

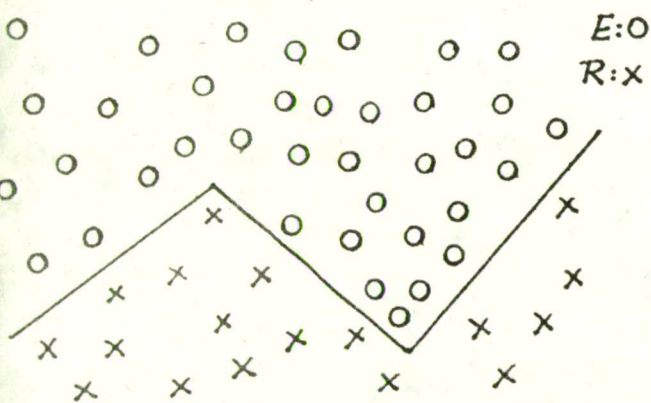
E: O R: X



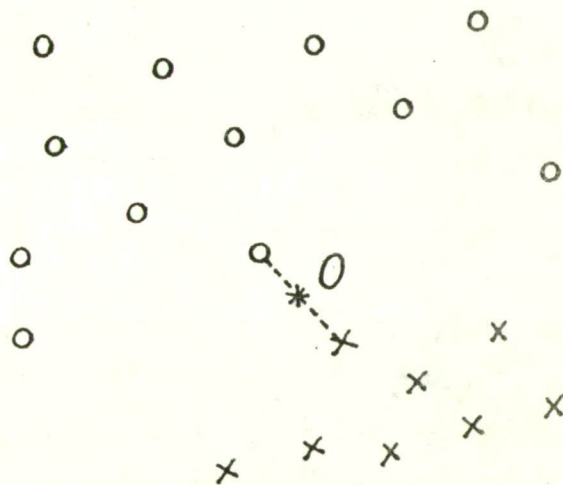
1. ábra.



2. ábra

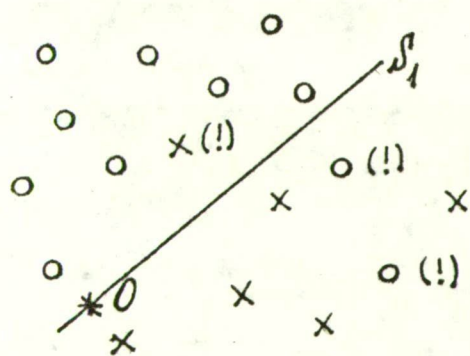


3. ábra



E: O
R: X

4. ábra



5. ábra

$$\begin{aligned}(r_1, n_2) &= -10,14 - 12 - 18,94 < 0 & (e_1, n_2) &= -15,21 + 24 - 66,29 < 0 & /!// \\(r_2, n_2) &= -20,28 - 24 - 37,88 < 0 & (e_2, n_2) &= -30,42 + 36 - 94,7 < 0 & /!// \\(r_3, n_2) &= -30,42 - 54 - 94,7 < 0 & (e_3, n_2) &= -20,28 + 12 - 47,35 < 0 & /!// \\(r_4, n_2) &= -25,35 - 78 - 75,76 < 0 \\(r_5, n_2) &= -25,35 + 18 - 94,7 < 0\end{aligned}$$

Most R és E halmazok S_2 ugyanazon oldalán állnak. Az első nem előjelhelyes vektor e_1 , ezért a következő lépés n_2 helyébe

$$n_3 = n_2 + e_1 = (-2,07; -10,00; -2,47)$$

vektort írni.

Ekkor

$$\begin{aligned}(r_1, n_3) &= -4,14 - 20,00 - 4,94 < 0 & (e_1, n_3) &= -6,21 + 40,00 - 17,29 < 0 \\(r_2, n_3) &= -8,28 - 40,00 - 9,88 < 0 & (e_2, n_3) &= -12,42 + 60,00 - 24,71 > 0 \\(r_3, n_3) &= -12,42 - 90,00 - 24,7 < 0 & (e_3, n_3) &= -8,28 + 20,00 - 12,35 < 0 \\(r_4, n_3) &= -10,35 - 130,00 - 19,76 < 0 \\(r_5, n_3) &= -10,35 - 30,00 - 24,71 < 0\end{aligned}$$

A következő lépés:

$$n_4 = n_3 + e_3 = (1,93; -12,01; 2,53)$$

Ekkor

$$\begin{aligned}(r_1, n_4) &= 3,86 - 24,02 + 5,06 < 0 & (e_1, n_4) &= 5,79 + 48,31 + 17,71 > 0 \\(r_2, n_4) &= 7,72 - 48,11 + 10,12 < 0 & (e_2, n_4) &= 11,58 + 72,00 + 25,30 > 0 \\(r_3, n_4) &= 11,58 - 108,14 + 50,60 < 0 & (e_3, n_4) &= 7,72 + 24,00 + 12,65 > 0 \\(r_4, n_4) &= 9,65 - 156,00 + 20,24 < 0 \\(r_5, n_4) &= 9,65 - 36,00 + 25,30 < 0\end{aligned}$$

Egyik elválasztó sík egyenlete

$$1,93x - 12y + 2,53z = 0 .$$

Ha például a B pontról, melynek koordinátái $x = 0,1$; $y = 10$; $z = 0$ melyik oldalra esik, akkor látjuk, hogy $0,193 - 120 < 0$, tehát B az R oldalán van.

6. A gépi program

A programban használt tömbök, változók, állandók:

- AR : Az R halmaz azonosítójának tömbje
- AE : Az E halmaz azonosítójának tömbje
- IR : Az R adatainak tömbje
- IE : Az E adatainak tömbje
- R : Az R adatainak transzformált tömbje
- E : Az E adatainak transzformált tömbje
- MR : Az R adatkártyáinak száma
- ME : Az E adatkártyáinak száma
- Z : Az aktuális sík koordinátáit tartalmazó tömb
- ZK : A kezdő sík koordinátáit tartalmazó tömb
- ZR : Az utolsó R szempontjából még jó sík
- OJ : Az új origó koordinátáit tartalmazó tömb
- ML : Az aktuális lépésszám
- MLS : Maximális lépésszám
- DDLK : A két halmaz pontjainak legkisebb távolsága
- RRLN : Az új origó és a tőle legtávolabbi pont távolsága

A program folyamatábrája a 6., 7., 8. és 9. ábrákon látható.

7. Alkalmazás leukoplakiás betegek carcinoma veszélyeztettségének korai megállapítására

A betegpopuláció leukoplakiás betegekből áll. R az a beteghalmaz, melynek tagjainál carcinoma bekövetkezik, E akiknél nem. A tünetvektorok adatai:

Életkor /6 csoport/

Megfigyelési /észlelési/ idő /5 csoport/

Localisatio /szájzug, bucca, alsó ajak, felső ajak, nyelv,

kemény szájpád, légyszájpad, szájfenék, gerinc/

Klinikai typus /leukoplakia simplex, l. verrucosa, l. erosiva/

Dohányzás mértéke /5 csoport/

Dohányzás ideje /5 csoport/

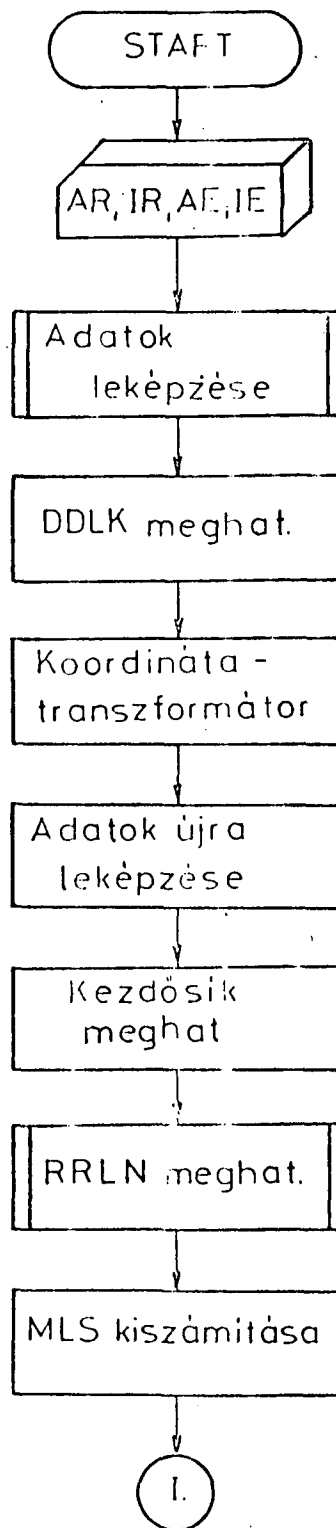
Aetiológiai tényezők /alkohol, mechanikai tényezők, syphilis, electropotenciál, candida/

Mechanikai tényezők /élesszélű saját fogak, élesszélű hid,

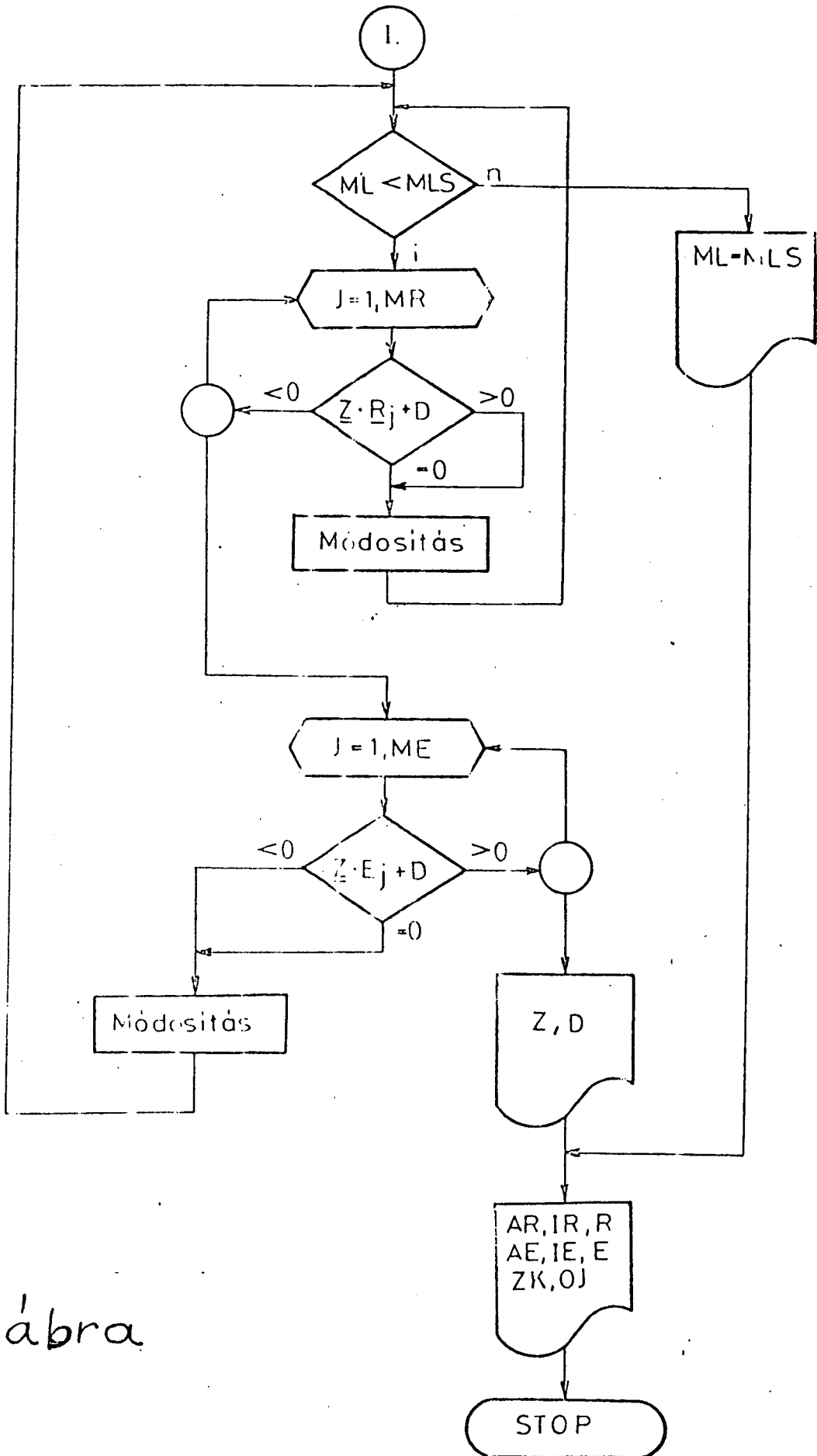
teljes protézis, részleges protézis/

Ez azt jelenti, hogy $n=8$.

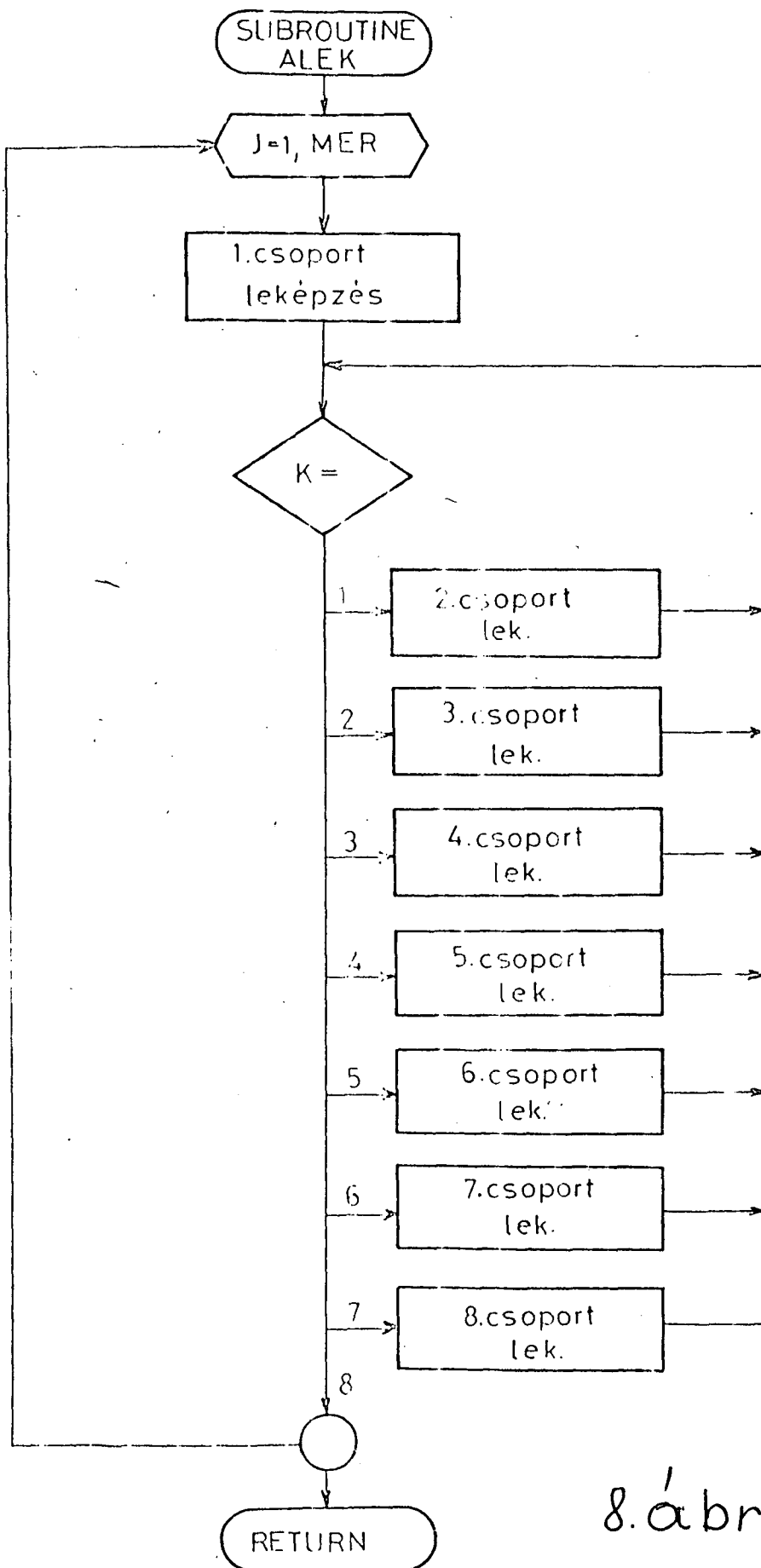
A vizsgálatok még folyamatban vannak, a kezdeti eredmények biztatóak.



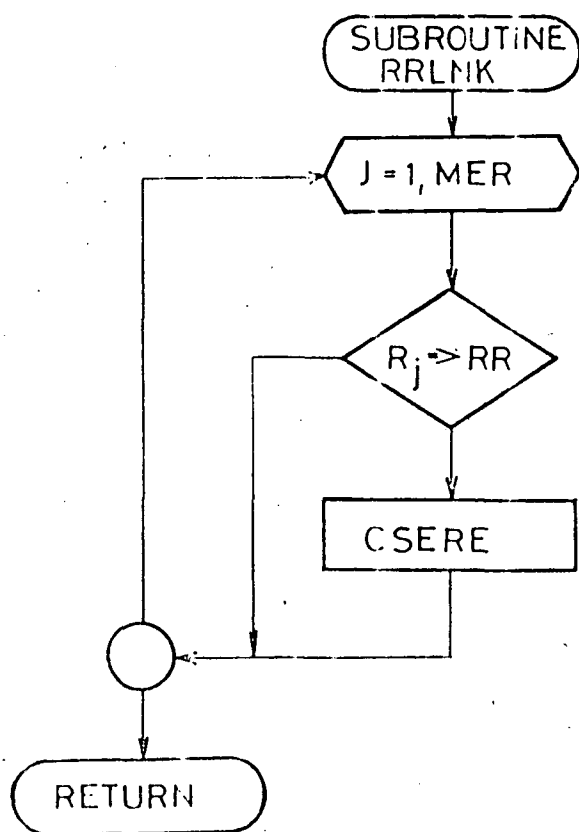
6. ábra



7. ábra



8. ábra



g.ábra