**Tüdőgyulladás detektálás**

A téma teljes megnevezése: Tüdőgyulladás detektálása röntgen képek alapján, osztályozó modell segítségével.

Készítette: Halász Adrián, [halaszadrian09@gmail.com](mailto:halaszadrian09@gmail.com)

Konzulens: Pašić Azra, [pasic.azra@gmail.com](mailto:pasic.azra@gmail.com)

**A tüdőgyulladás története**

A legrégebbi feljegyzések

A nevét a görögök adták neki, jelentése „állapot a tüdőről”. A tüdőgyulladás az egyik legrégebbi olyan betegségek közé tartozik, melynek van diagnózisa, illetve neve. Kutatók már az egyiptomi múmiáknál is véltek felfedezni olyan tüdőbetegségeket, mint például az antracózis vagy a szilikózis melyek hasonlítottak a tüdőgyulladásra.

Korai kezelések

A korai időszakban még a tüdőgyulladás és a mellhártyagyulladás nem volt megkülönböztetve (A két betegséget egyébként gyakran ugyanaz a kórokozó okozza). Hippocrates Görög orvos (Kr.e 460-370) véreztetést és beöntést ajánlott a tüdőgyulladás és mellhártyagyulladás kezelésére, tekintettel a fájdalom helyére, a beteg életkorára és bőrszínére és az adott évszakra.

Aretaues Cappadociából görög orvos (Kr.u 2.század) észrevette hogy a halál igen gyakran a 7. napon áll be. Aretaus hasonló kezelést ajánlott, mint Hippocrates ugyanakkor ő még bőségesebb véreztetést javasolt, illetve tisztító, higító és csillapító italokat alkalmazott, továbbá izsóp virággal dörzsölte be a páciens mellkasát.

A 17. századig és még egy kicsivel azután is a tüdőgyulladás kezelése Hippocrates és Aretaues feljebb tárgyalt, alapvető módszereit alkalmazta.

Egészen 18.század végéig és 19.század elejéig a véreztetés az egyik fő kúra volt a betegségre. Először egy fiatal tantestületi tag, Hughes Bennett (1812-1875) kérdőjelezte meg a módszer helyességét. Úgy gondolta, hogy a véreztetés inkább árt, mintsem segít a pácienseken.

A 19. század közepére már nyilvánvaló volt, hogy a kevésbé drasztikus kezelések sokkal hatékonyabbak, mint a régi, agresszívak. Például Hughes benett által javasolt helyreállító terápia, mely borból és marhahús teából állt, elérte a 3%-os halálozási rátát, vagy a homeopátiás gyógymódok, melyekkel 6%-ot értek el. Egy 1863-as feljegyzés szerint ezzel szemben azok a kezelések, amelyek véreztetést alkalmaztak 20.4%-os halálozással jártak.

A 19.század végére a mikrobiológiában elért haladásnak köszönhetően elkezdték felismerni a tüdőgyulladás jellemzőit. 1891-ben George és Felix Klemperer sikeresen létrehoztak egy vakcinát, mellyel sikeresen tudtak immunitást biztosítani a pneumococcus baktériumfaj ellen. Ez megalapozta a vakcinák fejlesztését a következő századra is.

Az első világháború és az influenza

Már az első világháború kezdeténél komoly aggodalmat keltett a tüdőgyulladás és az egyéb fertőzések, ugyanis az előző háborúkban több ember halt meg betegségben, mint golyó által. Sir William Osler kanadai orvos azt jósolta 1916-ban, hogy a harctéren szerzett sérülések sokkal kisebb halálozással járnak majd, mint a tüdőgyulladás általi. A betegséget a legnagyobb háborús gyilkosnak nevezte. Ahogy a háború a vége felé közeledett igaza is lett, ugyanis az Atlanti tengerről elinduló hajókban elterjedt az influenza (melynek a szövődménye igen gyakran tüdőgyulladás) ami végül 30 millió ember halálát okozta világszerte. A vírus eredete ismeretlen volt, de katonai feljegyzések szerint Kansas államban, Fort Riley-ban látták először. Érdekességként ez a vírus volt, ami a spanyolnáthát is okozta.

A háború utáni időszak és a penicillin

A háború után még sokáig próbálkoztak különböző vakcinák fejlesztésével több kevesebb sikerrel, de igazi áttörést a penicillin feltalálása hozott. 1928-ban Alexander Fleming fedezte fel. Az asztalon hagyott lefedetlen baktériumtenyészetei közül az egyik megfertőződött egy penészgombával. Miután izolálta utána azonosítani tudta, hogy a penicillium családba tartozik. Ezt a „penészlevet” különböző baktériumokon tesztelve rájött, hogy az összes Gram-pozitív patogén ellen hatásos (ilyen például a tüdőgyulladás, de a skarlát és az agyhártyagyulladás is). A felfedezésének fontosságát csak 1940-ben ismerte fel két oxfordi kutató, akik sikeresen tömeggyártani tudták a második világháború során. A felfedezésért Fleming 1945-ben fiziológiai és orvostudományi Nobel-díjat kapott.

**A betegség**

A betegség definíciója úgy írható le, mint egy fertőzés, mely jellemzően a gázcserezónát (vagyis az alveoláris teret) érinti. Ugyanakkor a mikroorganizmusok jelenléte az alveoláris térben, kísérő gyulladás nélkül csak kolonizációt jelent, nem tüdőgyulladást.

Csoportosítás

A betegség lehetséges fajtáinak meghatározása többféle módon is történhet. A legáltalánosabb csoportosítás azon alapszik, hogy a páciens hol tartózkodott, amikor elkapta. Például, ha a beteg kórházban fertőződött meg, akkor számítani kell arra, hogy a kórokozók ellenállóbbak lesznek a gyógyszerekkel szemben, azokhoz képest, amelyeket egy csoportból kapott el az ápolt. Ezen kívül még kétfajta csoportosítás szokott szóba jönni.

Az egyik az immunállapot alapján:

* Van immunműködési zavar
* Nincs immunműködési zavar

Az utolsó a mikrobiológiai csoportosítás, ez alapján lehet:

* Vírusos
* Bakteriális
* Gombás
* Mikobakteriális
* Parazita

Diagnosztikája

A tüdőgyulladás végleges diagnosztikája 4 szempontból áll.

* A legúti terület fertőzöttségére utaló tünetek és jelek.
* Radiológiai változások
* Egy vélt kórokozó felfedezése és
* A tüdőgyulladásnak megfelelő kezelési reakció vagy klinikai lefolyás

**felépítés: adatbányászat, MI, gépi tanulás, területen aktuális eredmények.**

**Az informatika és a tüdőgyulladás**

Egy 2016-os cikk részletesen ír az adatbányászat fontosságáról az 5 évnél fiatalabb gyermekeknél jelentkező tüdőgyulladásról. A motiváció erre az volt, hogy ezeknél a kisgyermekeknél vezető halálozási ok (1,1millió halál évente) a tüdőgyulladás, melyeknek több, mint a 95 százaléka a fejlődő országokban történik. A probléma abból fakadt, hogy a pontos diagnosztizáláshoz az orvosoknak a létfontosságú jelek és karakterisztikák kombinációját kell felmérniük. Az ilyen magas minőségű ellátásra viszont ezeknek az országoknak nincs mindig lehetősége. Szükség volt egy olyan eszközre, mely olcsóbban tud hasonló pontossággal összefüggéseket találni a rendelkezésre álló adatokban és diagnosztikát végezni. A gépi tanulás sikeres eszköznek bizonyult különösen arra, hogy meghatározza mely páciensek alkalmasak otthoni kezelésre, ezzel csökkentve a kiadásokat. A modell készítésénél az adatok nagyon hiányosak voltak (egyes oszlopoknak akár 42%-a is hiányzott) ezeket imputáltálással pótolták, de csak azoknál az oszlopoknál, ahol az adatok legalább 85%-a megvolt, az ennél hiányosabb részeket eldobták. Az imputálás medián stratégiával történt. Három modellt használtak fel, logisztikus regressziót, randomforestet és support vector machine-t. Végül a randomforest tudott a legjobban teljesíteni 98,2%-os eredménnyel.

(A fentebb leirtak igazabol tudomanyos eredmenyek nemtudom ez igy jo e.)

**Kiindulási állapot**

A modell röntgenképek alapján fogja eldönteni azt, hogy az adott páciensnek egészséges-e a tüdeje vagy tüdőgyulladásban szenved-e. Ehhez adatokra, jelen esetben röntgenképekre van szükségem. Az adatokat a Kaggle nevezetű oldalról szereztem be, így összesen 5856 képen tudok majd dolgozni. Az itt lévő képek három kategóriát tartalmaznak. Az első egy egészséges tüdő képe, a második egy tüdőgyulladásos tüdő képe, melyet vírus okozott. A harmadik pedig szintén egy tüdőgyulladásos tüdő, viszont ezt baktérium okozta.

Az alábbi három ábrán látható egy-egy példa minden kategóriára, az adathalmazból.

|  |  |  |
| --- | --- | --- |
| Normál | Vírusos tüdőgyulladás | Bakteriális tüdőgyulladás |
|  |  |  |

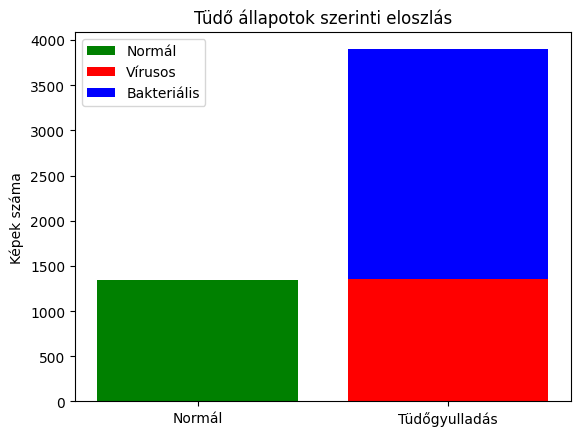
A fentebb említett 5856 kép összesen 3 mappára van szétbontva. Az első a train mely a modell tanítására szolgál. A második a val ami pedig a modell validálására való. A harmadik pedig a teszt, amely a már betanított modell hatékonyságának mérésére szolgál. Az én feladatom ezen képek beolvasása, értelmezése és feldolgozása, majd a megfelelő modell kiválasztása és rátanítása a feladatra a képek segítségével.

**Munkaelőzmény**

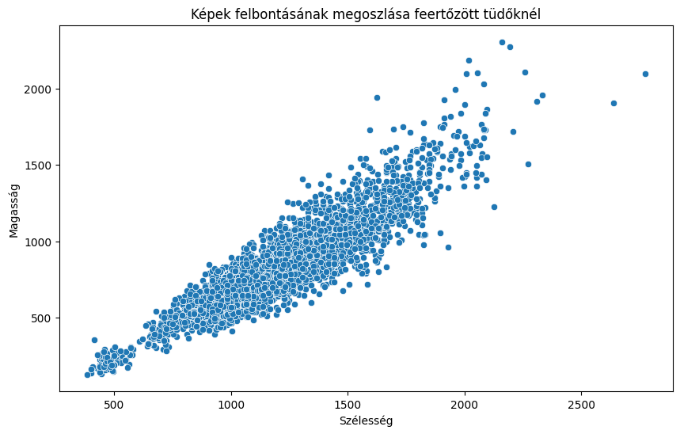
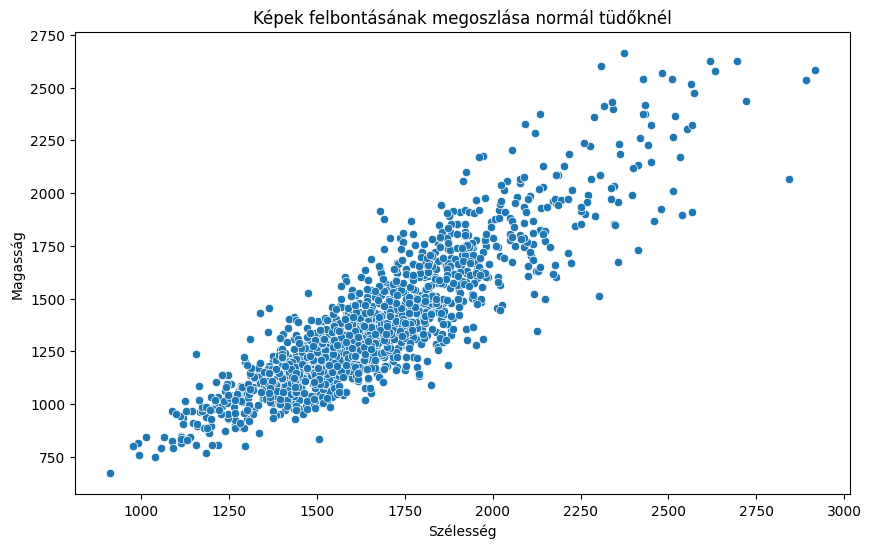
A programozás megkezdése előtt jobban elmélyültem a képalapú osztályozás elméleti hátterében. A számomra szükséges információknak utánanéztem és általánosságban végig vezettem, hogy pontosan hogyan is kell egy ilyen feladatot megoldani. (itt majd meg lehet folytatni a modell valasztasaval, a modszer valasztasaval jo lesz ez)

**Feltáró adatelemzés (EDA)**

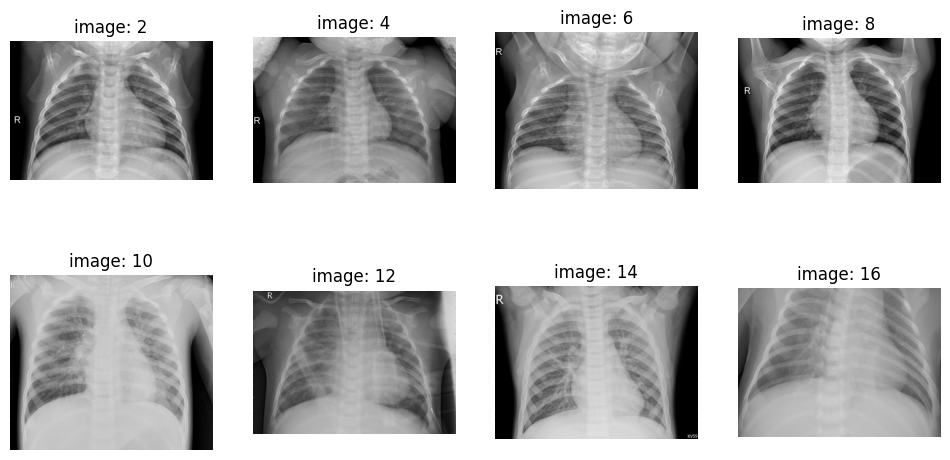
Első lépésként megvizsgáltam azt, hogy a két osztály külön külön hány képet tartalmaz. Az eredményt egy hisztogramon ábrázoltam:

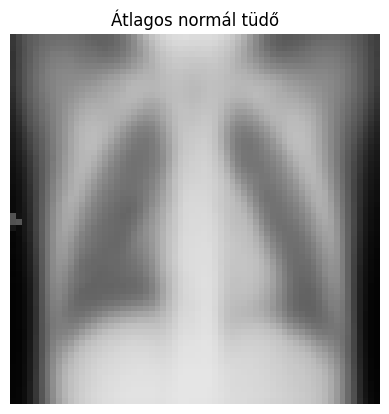


Az ábrán jól látható, hogy az egészséges tüdőkről jóval kevesebb kép áll rendelkezésre, mint a betegekről (1341 egészséges és 3895 beteg). Ezt az egyenlőtlenséget adat augmentációval fogom megoldani, melynek lényege, hogy mesterséges úton fogok képeket generálni (például a már meglévő képeket elforgatom, eltolom vagy eldöntöm). Ezután megnéztem, hogy milyen eloszlásban vannak a képek felbontásai. Az eredmény azt mutatta, hogy a képek egyáltalán nem egységesek, hanem sokféle méretben vannak jelen az adatbázisban. A modell tanításához már egy egységes méretre fogom alakítani őket.



A két osztály képeinek felbontásbeli megoszlása a fenti ábrákon látható. Ezt követően véletlenszerűen kiválasztottam pár képet mindkét osztályból és kirajzoltattam egy ábrára.



Végső lépésként minden képet átméreteztem 64\*64-esre és pixelenként kiszámoltam az átlagot mind az egészséges mind pedig a gyulladásos tüdőnél. Az így kapott átlag értékeket egy képként kirajzoltattam.

Eltérés:

Az eltérés értelmezése: A kék rész az, ahol a beteg tüdő pixel értéke nagyobb, mint az egészségesé (vagyis világosabb az a pixel), a piros ezzel szemben az, ahol az egészséges tüdő pixelértéke nagyobb. A fehér értéknél megegyeznek a pixelek. Látható, hogy a fertőzött tüdőknél jobban elhomályosodik a mellkasi rész.