

Koraszülött csecsemők légzésének
és aktív, valamint alvó fázisainak
érintésmentes kamera alapú monitorozása



Pázmány Péter Katolikus Egyetem
Információs Technológiai és Bionikai Kar
Roska Tamás Műszaki és
Természettudományi Doktori Iskola

Nagy Ádám

Témavezető: Dr. Zarándy Ákos

A PhD fokozat megszerzésére benyújtott dolgozat

2023

I. Bevezetés

Az olyan fiziológiai jelek, mint a pulzus, a légzés vagy a véroxigén szint rutinszerűen mért életjelek az intenzív osztályokon és a koraszülött klinikákon is. Ugyanígy, ha nem fontosabb paraméterek, a koraszülött csecsemők és a felnőtt páciensek esetén a viselkedési állapotok (csendes alvás, aktív alvás, átmeneti, csendes ébrenlét, aktív ébrenlét), amelyek a csecsemők aktív és alvó fázisait hivatottak leírni és amelyek a NIDCAP (Newborn Individualized Developmental Care and Assess) skála szerint osztályozást követik [17]. Ezek az állapotok annak becslése szolgálnak, hogy a páciens mennyit alszik, és mennyi mozgást végez. A csecsemők esetén az alvás elengedhetetlen a babák megfelelő fejlődéséhez. De a felnőttek esetében is rendkívül fontos a megfelelő mennyiségű és minőségű alvás.

Mind eddig ezen paraméterek figyelemmel követése vagy ritkán használt vezetékes rendszerek segítségével vagy a kórházi személyzet érzékszerveire támaszkodva történt. Mostanában azonban nagy előrelépés figyelhető meg a gépi látási, képfeldolgozási és gépi tanulási tudományok területén. Azt is mondhatnánk, hogy technológiai forradalom történik ezeken a tudományterületen. Elég csak a most divatossá váló generatív algoritmusokra gondolni, mint amilyeneket a Dall·E 2 vagy a ChatGPT alkalmaznak. Ezen népszerű közel múltbeli alkalmazások mellett a képfeldolgozási algoritmusok is komoly fejlődésen mentek keresztül. Új típusú konvolúciós hálózatok tömege jelent meg az elmúlt években, amelyeket orvosi területeken is előszeretettel alkalmaztak.

Ebben a dolgozatban én is arra vállalkozok, hogy orvosi területen, pontosabban koraszülött csecsemők megfigyelésének területén hozzak létre új gépi látáson és gépi tanuláson alapuló algoritmusokat. Három olyan konkrét problémát vizsgáltam meg ebben a dolgozatban, amelyek kihívást jelentenek az említett területen. Ezen

problémák a csecsemők légzésének érintésmentes monitorozása, az ehhez szükséges adathalmaz automatizálása és végül a megfigyelt csecsemők viselkedési állapotainak klasszifikációja.

A fiziológiai jeleke, mint amilyen a légzés, mint azt tudjuk rutinszerűen mérik a kórházakban, (különösen a koraszülött osztályokon), olyan orvosi monitorok segítségével, mint a "Phillips IntelliVue MP40/MP50". Az ilyen monitorokhoz csatlakoztatott vezetékes szenzorok azonban sok problémát okoznak a csecsemőknek és az orvosi személyzetnek is. A csecsemők vékony bőrét akár meg is sérthetik a "kontakt" szenzorok, de mindenképp stressz alá helyezik a csecsemőt. Az orvosokat és ápolókat pedig a vezetékek zavarhatják munkájuk elvégzésében. Ezért is egyfajta trend-ként jelent meg ezen vezetékes szenzorok érintésmentes módszerekkel való lecserélése.

Az orvosi és ápolói személyzetnek pedig folyamatosan manuálisan kellene megfigyelnie a csecsemők aktivitását és alvását a "Gyakorlati Készségek a Családcentrikus Fejlődéstámogató Ellátás Terén" című "FINE" által jóváhagyott oktatóanyagban leírtak szerint [17], hogy fejlődéstámogató kezelést tudjanak nyújtani. Erre azonban a legtöbb kórházban nincs elegendő erőforrás. Nagy igény van a viselkedési állapotok osztályozásának automatizálására.

Látható tehát, hogy mindkét esetben van igény a jelenlegi módszerek fejlesztésére, vagy akár mesterséges intelligencia és gépi látás alapú algoritmusokkal való lecserélésére. A mesterséges intelligencia alapú algoritmusok használatának közös nehézsége azonban, hogy nagyon nagy mennyiségű annotált adathalmazra van szükség. Ebben a dolgozatban ennek megoldására is kínálunk egy módszert, legalábbis a légzés érintésmentes monitorozását végző algoritmus tekintetében.

II. Alkalmazott módszerek, eszközök

Ahhoz, hogy a csecsemők fiziológiai jeleit és viselkedési állapotait becsló érintésmentes algoritmusokat tervezhessünk és értékelhessünk, nagy mennyiségű annotált adatra van szükség. Mivel etikai és egyéb okok miatt nyilvános adatbázisok nem érhetőek el a témában ezért saját adatbázist kellett kiépítenünk.

Ehhez először is ki kellett építenünk egy adatgyűjtő rendszert, amely az adat vezérelt algoritmusaink bemeneteit és az elvárt kimeneteit szinkronizáltan képes gyűjteni. A légzésfigyelés esetében ez egyszerűbb ugyanis a légzést a kórházakban jelenleg EKG-val (elektro-kardiogram) monitorozzák, amelyet szeretnénk érintésmentes módszerrel felcserélni. Ebben az esetben tehát a képi adatok gyűjtése egy színes kamerával történik, míg a légzési jel gyűjtése az EKG-val, amely, mint azt fentebb említettem, egy "Phillips IntelliVue MP40/MP50" monitorhoz van csatlakoztatva. Mi kiépítettünk egy olyan adatgyűjtő rendszert, amely rendelkezik egy infravörös megvilágítóval és egy kamerával, valamint képes rácsatlakozni az említett monitorra. Így képes a képi adatokat (az infravörös megvilágításnak köszönhetően) akár éjjel is gyűjteni a légzési referencia jellel szinkronizáltan.

A csecsemők viselkedési állapotainak osztályozása esetében a képi adatokat ugyanúgy egy infravörös megvilágítóval rendelkező színes kamera rögzítette. Az elvárt kimeneteket, (a viselkedési állapotokat) azonban ez esetben mobiltelefonos alkalmazás segítségével rögzítette a csecsemőt megfigyelő orvosi személyzet.

A források tekintetében először áttekintő cikkeket néztem át, amelyekben a szerzők általánosságban nézték át a csecsemők érintésmentes légzés monitorozásával kapcsolatos publikációkat [9] és az alvási állapotaikkal kapcsolatos publikációkat [19]. Ezután kifejezetten arra koncentráltam, hogy mélyebben áttekintsem azt az irodalmat, amely az színes kamerával történő mozgáselemzésen alapuló légzés fi-

gyelő eljárásokról [8–10, 14], valamint a csecsemők NIDCAP szerinti viselkedési állapotainak megfigyeléséről szól [2, 7, 17, 20]). Természetesen a csecsemő megfigyelés témakörének irodalmát magasabb perspektívából is áttekintettem. Külön foglalkoztam a "kontakt" szenzoros és az érintésmentes megfigyelés minden formájával.

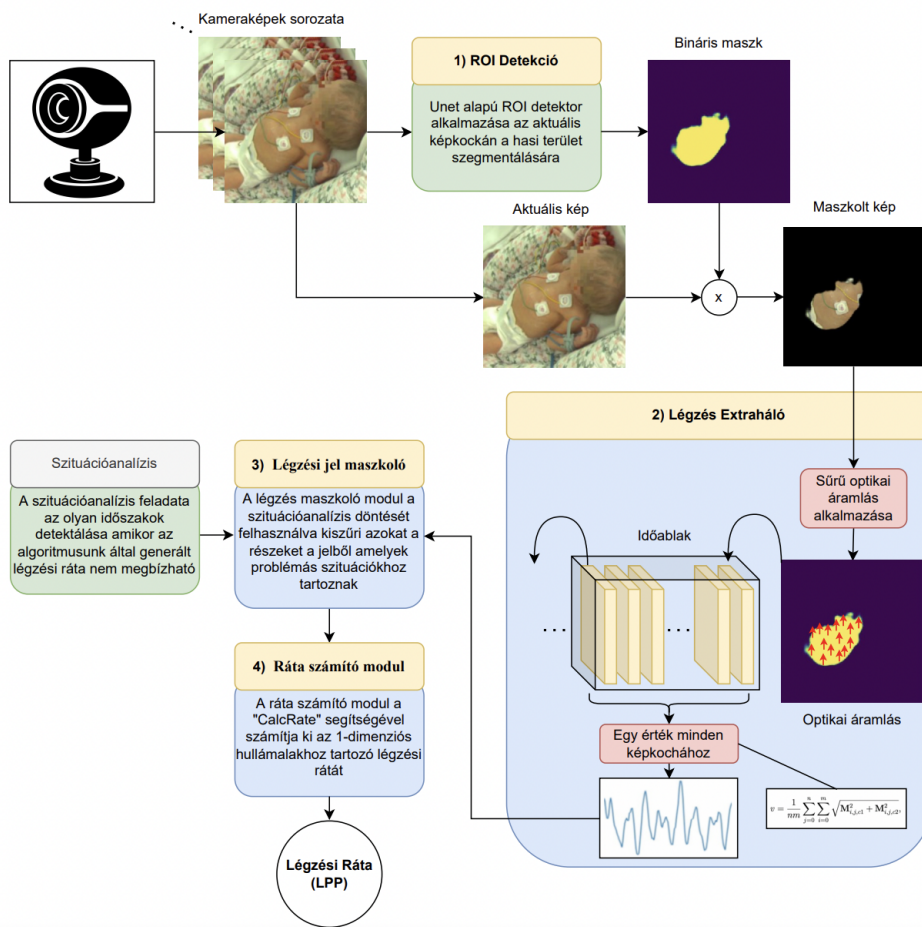
Az adatok és az annotációk kezelésére saját egyéni igényeinknek megfelelő C++ alapú szoftvereket hoztunk létre, amelyek lehetőséget adtak az adatok adatbázisból történő lekérdezésére és kategorizálására, valamint meg is jelenítették a képi adatokat és a referencia jeleket egyaránt.

A fentebb definiált problémák megoldására gépi tanulási és gépi látási eljárásokat alkalmaztam. Az elmúlt években a konvolúciós neurális hálózatok nagy fejlődésen mentek át és számos területen alkalmazták őket. Mozgásbecslés terén pedig a különböző típusú optikai áramlások már régóta jól bevált módszernek számítanak. Az érintésmentes légzésfigyelő összefoglalása az 1. ábrán látható.

Az olyan RNN-ek (recurrent neural network), mint az LSTM (long short-term memory) és a GRU (gated recurrent unit) pedig már számos alkalmazásban bizonyítottak az időszerelemzésben és predikcióban egyaránt. Én az általam tervezett légzésfigyelő algoritmusban alkalmaztam konvolúciós hálózatot ROI (region of interest) detektorként és a "sűrű" optikai áramlást mozgásbecslőként. A viselkedési állapotok klasszifikációjához pedig egy GRU alapú klasszifikátort alkalmaztam, amelynek bemeneteit különböző gépi tanulási és képfeldolgozási eljárások segítségével állítottam elő. Az automatikus adat annotáció generálásához szintén gépi látási és képfeldolgozási eljárásokat (pl. egyszerű morfológiai metódusokat) és differenciálkép elemzést használtam.

A saját algoritmusok megírása a Python nyelven és a Python alapú Pytorch és Pytorch-Lightning programcsomagok segítségével történt, majd a leimplementált és tesztelt algoritmusok telepítésre kerültek C++-ra is. A gépi tanulási algoritmusokhoz szükséges adatokat a fentebb leírt adatgyűjtő rendszerrel rögzítettük a Semmelweis Egyetem II. számú Szülészeti és Nőgyógyászati Klinikájának, Újszülött Intenzív osztályán. A gyűjtött adatokat ezután saját szoftverjeinkkel kategorizáltuk, hogy tudjuk mely adatok alkalmasak tanításra.

A validációs lépések szintén python-ban és C++-ban történtek független teszthal-



Ábra 1: Az általam tervezett és publikált érintésmentes légzés monitorozó algoritmus, amely egy U-Net alapú ROI detektorral (1) indul és egy "sűrű" optikai áramlás alapú légzés extraktorral (2) rendelkezik és egy csúcsetektor alapú légzési ráta kiszámolóval végződik (4). Ebben az algoritmusban van köztve a "szituációanalízis" eredménye is, ami segít a problémás szituációk megtalálásában (3), hogy az algoritmus automatikusan képes legyen megállapítani mikor tud megbízható légzési rátát szolgáltatni.

mazon, ami azt jelenti, hogy a halmaz olyan csecsemőkről készült felvételeket tartalmazott, amelyekről a tanító halmazban egyáltalán nem szerepelt képi adat. Valamint a bemutatott algoritmusok telepítésre kerültek olyan (a normál PC-vel összehasonlítva) kisebb számítási kapacitással rendelkező beágyazott rendszerekre, mint a Raspberry Pi vagy a Jetson Nano, amelyekre valós idejű futásra voltak képesek.

III. Új tudományos eredmények

Összességében elmondható, hogy a doktori munkám során számos olyan problémát megvizsgáltam, amelyek inkubátorban lévő koraszülött csecsemők megfigyelésével kapcsolatban merülnek fel.

Körbejártam a csecsemő mozgás és légzés érintésmentes monitorozásának témakörét, valamint a csecsemők viselkedési állapotainak automatizált osztályozását. Ezen problémák megoldására új, (hagyományos színes kamera alapú) algoritmusokat hoztam létre.

Adatbázist építettünk és az adatbázis kezelésére és (akár automatikus) annotálására alkalmas szoftver elemeket készítettünk el.

Születtek tehát adatgyűjtő rendszerek, adatannotációs eszközök és szoftverek, valamint olyan algoritmusok, amelyek újszerű módon közelítik meg a csecsemők bizonyos életjeleinek és állapotainak érintésmentes megfigyelését.

Ez a dolgozat koraszülött csecsemők érintésmentes, kamera alapú monitorozásával foglalkozik. Három kihívást jelentő kutatási témát vizsgáltam meg ezen a területen. Nevezetesen (i) a koraszülött csecsemők érintésmentes légzés monitorozását, (ii) az ehhez szükséges ROI detektor tanítására használt adathalmaz automatikus annotációját, és (iii) a csecsemők viselkedési állapotainak érintésmentes monitorozását.

A problémák megoldásához gépi látási és gépi tanulási algoritmusokat alkalmaztam. Ezeket használva mozgás alapon számítottam ki a képen lévő csecsemők légzését és részben mozgás alapon, részben a csecsemők bizonyos életjeleit felhasználva számítottam ki a viselkedési állapotait a NIDCAP skálának megfelelően.

Az alábbiakban ismertetem az ebben a dolgozatban lévő új tudományos eredményeimet, melyeket 3 tézispontban fogalmaztam meg.

Az első olyan probléma amelynek megoldására színes kamera alapú algoritmus készült, az érintésmentes légzésfigyelés. Ennek megoldásánál egy olyan módszer került alkalmazásra, amely egy neurális hálózat alapú ROI detektorral megkeresi a képen lévő csecsemő hasát, számszerűsíti az ezen a területen fellelhető mozgást és meghatározza a légzési rátát. Ez a folyamat 4 fő részre osztható fel:

1. ROI Detektor
2. Légzés extraháló vagy légzés kivonó modul
3. Légzési jel maszkoló modul
4. Ráta számító modul

1) ROI Deketor

A ROI Detektor maga egy U-Net ([11]), amelyet a törzs szegmentálására alkalmazunk, mivel a képeknek ez a része hordozza a legtöbb információt a légzéssel kapcsolatban. Ily módon, a végtagok zavaró mozgáskomponenseit is ki tudjuk küszöbölni. A ROI detekció eredménye egy bináris maszk, ahol a törzshöz tartozó képpontok 1-ek, míg a törzsön kívüli képpontok 0 értéket vesznek fel. Ez a ROI-t kijelölő bináris kép lesz a "Légzés extraháló modul" egyik bemenete.

2) Légzés extraháló modul

A Légzés extraháló feladata az egydimenziós légzési hullámalak előállítása a videóból. Az algoritmus tartalmaz egy csúszó ablakot, amely a kamerából érkező képekből számolt mozgási képekkel lesz feltöltve a következőképpen: Az új beérkező képből és az azt megelőző képből kiszámítjuk a mozgási képet (Farneback optikai áramlásának segítségével [5]) és az ennek eredményéül kapott mozgási képet elemenkénti szorzással megszorozzuk a "ROI Detektor"-ból származó bináris képpel a következő egyenletnek megfelelően:

$$\mathbf{M} = \text{U-Net}(\mathbf{I}(t)) \odot \text{Dense}(\mathbf{I}(t), \mathbf{I}(t - 1)), \quad (1)$$

ahol $\mathbf{I}(t) \in \mathbb{R}^{m \times n}$ az aktuális időpillanathoz tartozó kamerából érkező képkocka, ahol felvételeink során $m=500$ és $n=500$ volt jellemző. A "U-Net()" metódus a ROI

detektor alkalmazása az aktuális képkockán, míg a "Dense()" metódus Farneback "sűrű" optikai áramlását számítja ki az aktuális és az azt megelőző képkockákra. \mathbf{M} pedig a bináris ROI képpel megszorozott mozgási kép.

A kapott mozgási képet betesszük a csúszó ablak elejére. Mivel a csúszó ablak fix hosszú kell maradjon ezért az utolsó elemet töröljük. Ezek után az ablak minden egyes képéhez egy (v) értéket rendelünk a 2. egyenlet alkalmazásával, így ezekből az értékekből egy $N=200$ hosszú egydimenziós vektort kapunk, amely "float-64" típusú elemeket tartalmaz. A következő lépésben az egydimenziós hullámalakot szűrjük egy "Butterworth" típusú ([3]) sáváteresztő szűrővel (20–120 LPP) frekvenciaértékek között, hogy végeredményként egy olyan hullámalakot kapjunk kimenetként, amelynek frekvenciája a légzés fiziológiai frekvenciatartományába esik és korrelál a képen lévő csecsemő légzésével.

$$v = \frac{1}{nm} \sum_{j=0}^n \sum_{i=0}^m \sqrt{\mathbf{M}_{i,j,c1}^2 + \mathbf{M}_{i,j,c2}^2}, \quad (2)$$

ahol $\mathbf{M}_{i,j,c1}$ a fentebb leírt módon kiszámított aktuális mozgási kép (i, j) pixele a c csatornán. A mozgási képnek két csatornája van. A $c1$ a horizontális, míg $c2$ a vertikális irányú mozgásokat mutatja. A mozgási kép horizontális és vertikális méreteit n és m adja meg. Az egyes maszkolt mozgási képkockákhoz rendelt v értékek sorozata adja meg a légzési jellel korreláló egy dimenziós mozgási jelet: $\mathbf{s} = [v_1, v_2, \dots, v_N]$.

3) Légzési jel maszkoló modul

A "szituációanalíz" nevű RNN alapú modulból kapunk egy döntést minden egyes bejövő képkockához arról, hogy az adott szituációban nyugodt légzés történik-e az aktuális és az azt megelőző képkockákból számolt jellemzők alapján. A döntések bináris értékek 1-ek vagy 0-ák lehetnek, amelyekkel szintén feltölthetünk egy döntéseket tartalmazó csúszó ablakot (döntési ablak). Ez az ablak tehát 1-eseket és 0-akat fog tartalmazni. Egyesek tartoznak azokhoz az időpillanatokhoz amikor az érintésmentes légzésfigyelő megbízhatóan tud légzési rátát szolgáltatni és nullák tartoznak azokhoz, amikor problémás szituáció lép fel. A Légzési jel maszkoló modul kiszűri a problémás szituációhoz tartozó mozgási jeleket azáltal, hogy a kapott hul-

lámalakot tartalmazó vektor és a "döntési ablak" elemeit tartalmazó vektor között elemenkénti szorzást hajt végre. Ezáltal kinullázza azokat a részeket, amelyek a problémás szituációkhoz tartozó időtartományba esnek.

4) Ráta számító modul

Ez az érintésmentes légzésfigyelő utolsó modulja. A feladata (nevéből következően) a légzési ráta meghatározása. Bemenete a légzési hullámalak (s). (Amely egy egydimenziós "float-64" típusú 200 hosszú vektor.) A kimenete pedig a légzési ráta LPP-ben (légzés/perc), amelyet egy "uint8" érték reprezentál 20 és 120 között. (Azért ezen két érték között, mivel ez a tartomány felel meg a csecsemők légzésének fiziológiai tartományának az irodalomkutatásunk és kísérleteink szerint.)

A ráta kiszámítására egy CalcRate nevű csúcsetekció alapú eljárást fejlesztettünk. Ez az algoritmus a jel megfordításával és egy sáváteresztő szűrő alkalmazásával indul, amely a 20 és 120 közötti frekvenciákat ereszt át. Ezt követi egy adaptív küszöbölési lépés, ahol a küszöbszint a bemeneti jel adaptív (a jel aktuális környezetében számolt) középértéke. Ily módon bináris jelet kapunk, ahol az 1-es érték a légzéseket (a pozitív csúcsokat) jelenti, a légzések kezdetét pedig a felfutó élek jelzik. A felfutó élek megszámlálásával megállapíthatjuk az adott "időablakban" történt légzés számot. Az utolsó lépésben a 0,3 másodpercnél közelebbi légzések összevonásra kerülnek, és a csúcsok közötti időintervallumok átlagolásával és az eredmény frekvenciaértékké alakításával számítjuk ki a légzési frekvenciát.

I. Tézispont: *Javaslatot tettem egy újszerű érintésmentes, kameraalapú légzés monitorozó algoritmusra, amely színes kamerával felvett képkockák sorozatát elemezve becsüli meg a képen lévő koraszülött csecsemők légzési rátáját és valós kórházi körülmények között is megbízhatóan működik, ugyanis képes felismerni az olyan szituációkat amikor a légzési ráta becslése nem lehetséges, mint pl. amikor kezelés/etetés zajlik vagy a csecsemőt kiveszik az inkubátorból.*

Kapcsolódó publikáció: [J1]

A második olyan általam tervezett algoritmus, amelyet szeretnék részletesebben bemutatni erősen kapcsolódik az első érintésmentes légzés figyelő algoritmushoz,

ugyanis annak az algoritmusnak az első moduljának, a ROI detektornak a be-
tantásához szükséges adathalmaz automatikus annotációját végzi el. Az annotá-
ció, különösen a szemantikus szegmentáció esetén szükséges pixel szintű annotáció
nagyon fáradságos munka. Valamint kevés számú annotáló esetén problémát jelen-
thet az annotálók szubjektivitása. Az emberi annotálók egyéni megítélése szetint
kerülhet annotálásra az adathalmaz. Az automatikus gépi annotálás esetén ez a
probléma nem áll fenn és megkímél minket a fáradságos annotálási munkától.

Enek az automatikus annotáló algoritmus két fő részre osztható fel:

- Detekciós modul
- Követési modul

A detekciós modul feladata a légzési mozgáshoz tartozó terület megtalálása a
mozgás kiterjedése és frekvenciája alapján. A Követési modul feladata, hogy kövesse
a megtalált terület esetleges mozgását, hogy ne kelljen folyton lefuttatni a detekciós
modult.

A komplexebb rész a detekciós modul, amelynek működése 4 fő lépésben foglal-
ható össze, melyek:

- Mozgás becslés
- Terület alapú szűrés
- Frekvencia alapú szűrés
- Elmozdulás alapú szűrés

Mozgás Becslés

Ennek a lépésnek a célja egy mozgási kép előállítás, amelynek segítségével meg-
tudjuk adni, hogy melyik pixelek körül milyen intenzitású mozgás történt az adott t
pillanat hoz tartozó képkocka és az előző $t-1$ idő pillanathoz tartozó képkocka között.
az ilyen mozgási képek létrehozásához számos algoritmus létezik, mint pl. a sűrű
optikai áramlás, az úgy nevezett "block-matching" algoritmusok ([4]), vagy a "Deep-
Flow" ([18]), amelyet Yue Sun et al. alkalmazott légzési mozgás becslésére [15].
A legegyszerűbb megoldás azonban nyilvánvalóan a két említett képkocka közötti

differenciakép kiszámítása. Ez az eljárás nem ad információt a mozgások irányáról csak az intenzitásáról. Mivel azonban nekünk nincs szükségünk az irányokról információra ennél az alkalmazásnál és a differenciálkép négyzete (\mathbf{D}) jól megmutatja a mozgás kontrasztjait, amelyen nagyon könnyen detektálhatók a nagy területű hasi mozgások, (főként a pelenka és a bőr találkozásának mozgása), ezért ez a módszer tökéletes számunkra az itt bemutatott alkalmazáshoz. A differenciakép önmagában még zajos lenne. De a differenciakép négyzetén már nagyon szépen detektálható a legnagyobb területű mozgás és ez is könnyen és nagyon gyorsan számolható (lásd a (3) egyenletben). A négyzetre emelés felerősíti az erősebb mozgásokat és gyengíti a zajokat. Tehát a jel zaj viszony javításához szükséges. Ezen a módon egy csatornás elforgatás invariáns mozgási képet kapunk.

$$\mathbf{D}(x, y, t) = (\mathbf{I}(x, y, t) - \mathbf{I}(x, y, t - 1))^2, \quad (3)$$

ahol $\mathbf{I}(x, y, t)$ az (x, y) koordinátához tartozó pixel intenzitása a t időpillanatban. A $\mathbf{D}(x, y, t)$ pedig a t . időpillanathoz tartozó négyzetes differenciakép (x, y) koordinátákhoz tartozó pixelének intenzitása.

Terület alapú szűrés

A fentebb említett mozgási kép kiszámítása után, ha a csecsemő nyugodt légzést mutat és nincs más mozgás a képen, akkor meg tudjuk találni a hasi terület körüli légzőmozgás helyét. Ha azonban nagyobb kiterjedésű összefüggő mozgási területek láthatóak a $\mathbf{D}(x, y, t)$ képen vagy irreálisan kicsik, akkor tudhatjuk, hogy ezek nem a has légzőmozgásából adódnak. Ezért, ha a kamera csecsemőtől való távolsága fix, akkor ennek függvényében meg tudunk állapítani egy alsó és egy felső határt a $\mathbf{D}(x, y, t)$ képen található legnagyobb összefüggő komponens pixelszámára. Ha a legnagyobb komponens pixelszáma nem ebben a sávban található, akkor a detekció sikertelen.

Ennek kiszámításához először egy küszöbölésnek kell alávetnünk a $\mathbf{D}(x, y, t)$ mozgási képet. Az így kapott bináris képen ezután már megtalálhatjuk az összefüggő területet a spagetti algoritmus segítségével [1]. A spagetti pixel klaszterek listáját

adja vissza. Egy klaszter olyan pixeleket tartalmaz, amelyek egymással összekötésben vannak. A spagetti egy olyan képpel tér vissza, ahol minden pixel a hozzá tartozó klaszter sorszámával lesz egyenlő. A háttér is egy külön klaszter, amivel nem foglalkozunk. Ezután kiválasztjuk a héttértől eltérő legnagyobb összefüggő komponst (lásd a (4). egyenletet).

$$g(x, y) = \begin{cases} 1, & \text{if } l(x, y) = \phi \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

ahol $l(x, y)$ a spagetti algoritmus által (x, y) koordinátájú pixelhez hozzárendelt csoport sorszáma. Az ϕ pedig a legtöbb pixelt magába foglaló csoport sorszáma.

Ily módon kiválasztottuk a legnagyobb összefüggő komponst (g), amely esetében a képen csak 0-ák és 1-esek szerepelnek. Ezután megnézzük, hogy g pixelszáma benne van-e a megengedett sávban: $Th_{lower} < \sum_{x,y=1,1}^{M,N} g(x, y) < Th_{upper}$. Ezek empirikusan beállított küszöbértékek. Ha ugyanis a legnagyobb összefüggő terület túl kevés pixelből áll, akkor valószínűleg nem a hasi területhez tartozó mozgást választottuk ki. Ha pedig túl nagy akkor valószínűleg intenzív végtagmozgást vagy egyéb zavaró mozgó objektumot találunk meg. Ha az említett pixelszám bele esik a megengedett sávba, akkor az algoritmus tovább léphet a következő lépésre. Ellenkező esetben az algoritmus visszatér az első lépéshez, amely a mozgásbecslés alkalmazása a következő bejövő képre.

Frekvencia alapú szűrés

A terület alapú szűrés önmagában nem elég ahhoz, hogy megtaláljuk a légzőmozgáshoz tartozó régiót a képen, ugyanis vannak olyan egyéb (pl. végtagmozgások), amelyek kiterjedése a hasi légzőmozgás kiterjedésével többé kevésbé megegyező a mozgási képen. Ezért szükség van további szűrésre. A légzőmozgás egy másik fontos attribútuma a frekveníája. Ezután tehát az időbeli mintázatát vizsgáltuk a mozgásnak. Ehhez azonban először időbeli egy dimenziós jellé kellett alakítanunk a mozgási képek sorozatát.

Mint már sokszor korábban, most is egy csúszó ablakot alkalmaztunk, csak ebben az esetben képkockák helyett a nyers mozgási képek voltak hozzáfűzve az ablakhoz,

amely 300 mozgási képkockát tartalmazott, amelyeket a 3. egyenlet szerint számítottunk ki, majd sorfolytonosan kiolvastva az értékeket egydimenziós vektorokká alakítottuk, amelyeket mintáknak nevezünk és amelyeknek hossza $H=N \times M$, amelyben N a mozgási képek horizontális M pedig vertikális kiterjedése. Így tehát a csúszó ablakban ezúttal nem 2D mozgási képek, hanem 1D-á transzformált mozgási képek szerepeltek, amely egy $(300 \times H)$ alakú mátrix. Ezután alkalmazhatjuk erre a mátrixra a "főkomponens analízis - Principal Component Analysis (PCA)" eljárást [16], amely a sajátvektorok terébe képezi le a mintáinkat. Ezekből az első komponens kiválasztva végzünk dimenziócsökkentést. Azaz egydimenzióba képezzük le a mintáinkat. Így egy 1D mozgási jelet kapunk, amely 300 hosszú és a mozgási képen fellelhető mozgások közül a PCA által leginkább hangsúlyosnak talált mozgást írja le.

A jel kinyerés megtörténte után már meg tudjuk vizsgálni, hogy a kapott egydimenziós jel frekvenciája azon tartományban található-e, amely a légzés fiziológiai frekvencia tartománya. Az egészséges újszülöttek a 25 LPP és 68 LPP közötti frekvencia tartományban lélegeznek [6], de a méréseink alapján bizonyos (akár rendellenes) esetekben magasabb frekvenciájú légzés is előfordulhat, ezért a lehetséges légzési frekvenciatartományt kitágítottuk (20 - 120). A fentebb említett csúszó ablak aktuális tartalmához tartozó 1 dimenziós jelet tehát annak alapján értékeljük, hogy a frekvenciája benne van-e a légzés fiziológiai tartományában. Ehhez kiszámítjuk az említett jelhez tartozó FFT spektrumot és kiválasztjuk a spektrum legmagasabb csúcsához tartozó frekvenciát:

$$f = \mathbf{b}[\max_f \{FFT(\mathbf{s}_M)\}], \quad (5)$$

ahol \mathbf{s}_M az említett csúszó ablakhoz tartozó mozgási jelalak, míg \mathbf{b} egy vektor, amely az FFT spektrum frekvencia helyeit tartalmazza.

Ezután megvizsgáljuk, hogy a kérdéses sávban (20 LPP - 120 LPP) található-e ez a frekvencia. Ha a számított frekvenciaspektrum legnagyobb csúcsához tartozó frekvencia (f) nincs az említett sávban, akkor a detekció nem sikeres az aktuális mozgási képen, mert ez esetben nem tudjuk garantálni, hogy a detektált mozgás

a légzéshez kapcsolódik. Ha pedig benne van akkor az algoritmus tovább mehet a következő lépésre. Ellenkező esetben az algoritmus visszatér az első lépéshez, amely a mozgásbecslés alkalmazása a következő bejövő képre.

Elmozdulás alapú szűrés

Ez előzőleg bemutatott szűrések már alkalmasak lehetnek a légzési mozgás tartalmazó régió detektálásához. Bizonyos esetekben azonban az a helyzet áll fenn, hogy a megfelelő frekvenciájú mozgás látható a mozgási képek sorozatán és az utolsó képen lévő legnagyobb összefüggő terület kiterjedése is megfelelő, mégsem a légzési mozgás következménye ez az összefüggő terület a mozgási képen. Ilyen esetekben nyugodt légzés figyelhető meg a képen viszont hirtelen történik valamilyen alacsony intenzitású mozgás a hasi régiótól eltérő területen pl. a végtagok környékén. Ritka esetben ennek a végtagmozgásnak a kiterjedése pont beleesik a keresési sávba és ilyenkor a végtag körüli területet detektáljuk a légzőmozgás forrásaként. Ennek elkerülése érdekében vezettük be az elmozdulás alapú szűrést, amely a legnagyobb összefüggő terület geometriai középpontjának mozgását követte figyelemmel és ha hirtelen túl nagy irreális elmozdulást tapasztaltunk akkor a detekció sikertelen volt.

Ehhez újra csúszó ablakot alkalmaztunk, amely ezúttal nem képkockákat, hanem $\mathbf{c}_i = (\mathbf{x}_i, \mathbf{y}_i)$ koordinátákat tartalmazott. Amennyiben a csúszó ablakban lévő középpontok közötti távolság soha nem haladja meg az empirikusan beállított Th_{st} threshold-ot a detekció sikeres:

$$\forall \mathbf{c}_i \in \mathbf{c}, d(\mathbf{c}_i, \mathbf{c}_0) < Th_{st} \quad (6)$$

ahol $i < N$ az i . indexet jelöli, N a középpontokat tartalmazó csúszó ablak hossza, $d(\cdot)$ a két bemeneti pont közötti euklideszi távolságot adja meg, \mathbf{c} pedig egy vektor, amely $N=300$ darab 2D pontot tartalmaz, ahol $\mathbf{c}_i \in \mathbb{R}^2$.

Követési modul

A követési modul feladata, hogy a detekció által megtalált terület középpontját kövesse "ritka" optikai áramlást alkalmazva. A követési modul a követés során menti a követett területet megmutató bináris képet és az ahhoz tartozó aktuális bemeneti

képet. Így egy annotált adathalmazt generál. A követés addig zajlik amíg valami el nem takarja a követett pontot vagy az ki nem mozog a képből vagy videót nem vált az algoritmus. Ezekben az esetekben vissza kell térnünk a detekciós model futtatásához. Ha a detekciós modul újra megtalálja a légzéshez kapcsolódó területet, akkor pedig folytatódhat az automatikus annotálás.

II. Tézispont:

Megmutattam, hogy amennyiben neurális hálózat alapú szegmentálót alkalmazunk a légzés várható helyének megtalálására, lehetséges automatikusan annotációs adatbázist generálni a koraszülött csecsemőkről készült folyamatos videófelvételeket felhasználva. Javaslatot tettem egy olyan algoritmusra, amely mozgási kiterjedés és frekvencia alapján detektálja és követi a légzőmozgás helyét és így annotálja fel a koraszülöttről készült folyamatos videófelvétel képkockáit.

Kapcsolódó publikáció: [C1]

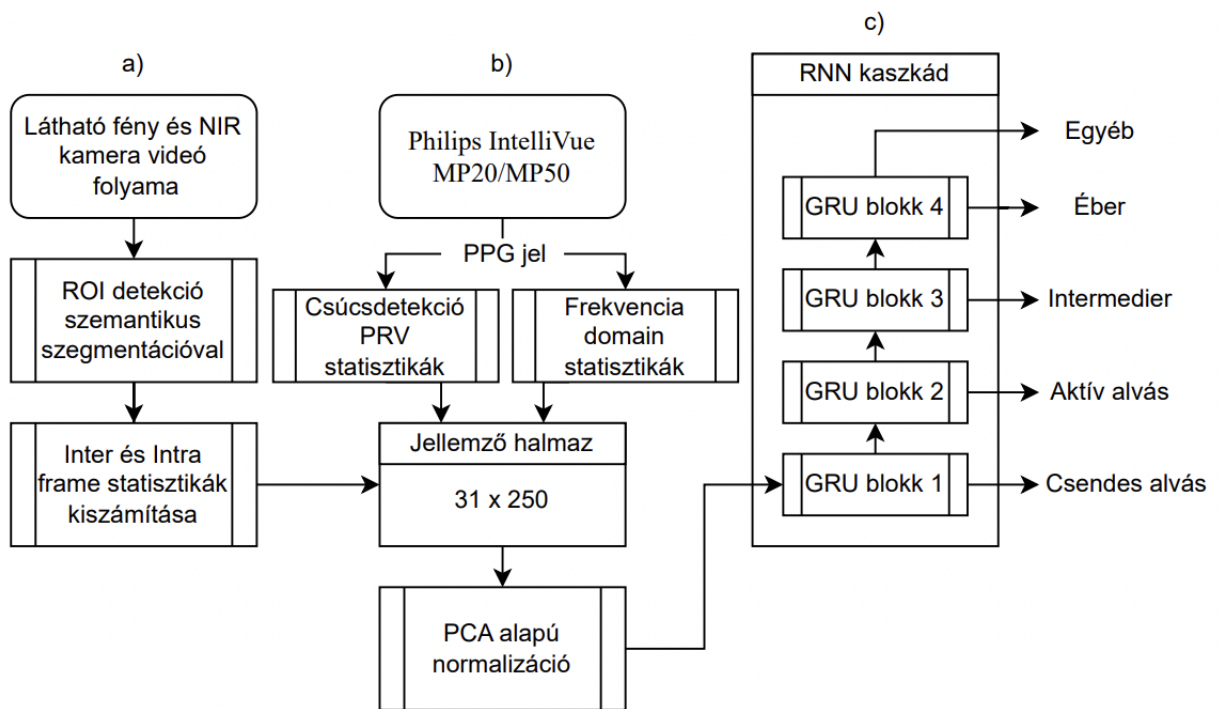
A harmadik algoritmus (2. ábra), amelyet szeretnék bemutatni, a NIDCAP szerinti viselkedési állapotokba történő klasszifikációt végzi el a csecsemőről készült videófelvételt és a Philips IntelliVue MP20/MP50 orvosi monitorból származó pulzusvariabilitási adatokat felhasználva.

Ez az algoritmus működése 3 fő lépésre osztható fel:

- Videó alapú jellemzők kinyerése
- Pulzusvariabilitás alapú jellemzők kinyerése
- Klasszifikációs modul alkalmazása

Videó alapú jellemzők kinyerése

A videó alapú jellemzők kinyerése úgy történik, hogy előre definiált informatív régiókat detektálunk a képen, mint a csecsemő hasa, a kép széle, vagy a végtagokat átfedő has körüli gyűrű. Ezekben a régiókon különböző mozgási jeleket vonunk ki, mint pl a differenciakép az adott területen, vagy a sűrű optikai áramlás vagy a kép HSV transzformáltjának adott régióba eső pixeleinek átlaga, stb. Három informatív



Ábra 2: A viselkedési állapot becselő algoritmus működését bemutató összefoglaló ábra. Itt a) rész mutatja a videó alapú jellemzők kiszámítását végző modulokat. A b) mutatja PPG jel alapú jellemzők kiszámítását és a PCA (Principal Component Analysis) alkalmazását. A c) pedig az aktuális periódushoz tartozó viselkedési állapotot meghatározó klasszifikációs modult.

régióon számítunk ki 8 féle mozgási jelet, amelyekhez hozzá fűzzük még a légzési jelet. Így egy 25 egydimenziós hullámalakot tartalmazó jelhalmazt kapunk.

Pulzus variabilitás alapú jellemzők kinyerése

Mivel minden újszülöttet folyamatosan monitoroznak pulzoximéterrel, de EKG-val nem, ezért úgy döntöttünk, hogy a kontakt-szenzorral mért adatok közül a pulzus jelet használjuk és pulzus ráta variabilitás (Pulse Rate Variability - PRV) alapú mérőszámokat számítunk ki. A legkézenfekvőbb, ha az időtartománybeli metrikákat alkalmazzuk. Ha a csúcsokat detektáltuk, akkor kiszámíthatjuk a köztük lévő időintervallumok hosszát. Az ilyen intervallumok sorozatából összetettebb metrikák számolhatók ki a különböző fázisok összehasonlításához. Az általunk felhasznált összetettebb jellemzők kiszámítását a 7 és 8 egyenletek írják le:

$$RMSSD = \sqrt{\frac{\sum_{i=1}^N (\mathbf{PP}_{i+1} - \mathbf{PP}_i)^2}{N}} \quad (7)$$

$$SDNN = \sqrt{\frac{\sum_{i=1}^N (\mathbf{PP}_i - \mathbf{mPP})^2}{N - 1}} \quad (8)$$

ahol \mathbf{PP} a két egymást követő impulzuscsúcs közötti időkülönbség, a \mathbf{PP}_i az i -edik impulzuscsúcs párhoz tartozó időkülönbség, az N pedig az impulzuscsúcsok száma. Az \mathbf{mPP} a PP intervallumok átlagát jelenti, az \mathbf{SDNN} ([13]) a fentebb említett PP intervallumok standard eltérése \mathbf{mPP} -től, ami a variancia négyzetgyöke, míg \mathbf{RMSSD} a normál szívverések közötti egymást követő különbségek négyzetének középértéke. A számítás egy 5 perces csúszó ablakban történik a HRV-metrikák [12] áttekintésében leírtak szerint.

A kiszámolt PRV alapú metrikákat, (amelyek szintén adott hosszúságú egydimenziós jelek), hozzá fűzzük a korábban létrehozott videó alapú jelek hamazához, majd a kapott jellemző halmazon végrehajtunk egy PCA normalizációt. Az így kapott jellemző halmaz lesz a klasszifikáció végző RNN kaszkád alapú modul bemenete.

Klasszifikációs modul

A klasszifikációt a NIDCAP alapú viselkedési osztályokba egy RNN (pontosabban egy GRU - Gated Recurrent Unit) kaszkád alapú modul végzi. A teljesítmény fokozása érdekében a GRU osztályozókat (blokkokat) bináris klasszifikátorok lépcsőzetes, döntési fa szerű kaszkádjába rendeztük. A blokkok megkülönböztetik a viselkedési állapotokat a mély alvás, az aktív alvás, az átmeneti és az ébrenlét sorrendjében. A sorrend fontos ugyanis a később lévő osztályba csak akkor kerül klasszifikálásra az adott periódushoz tartozó jellemzőhalmaz ha az azt megelőző blokkok nem trétek vissza pozitív klasszifikációval. Előre kerültek a legpontosabb bináris klasszifikátorok. A kaszkád pontossága így megnőtt a teljes osztályhalmazt tekintve. A bináris klasszifikáló blokkok GRU rétegekből és RELU aktivációs függvényekkel ellátott lineáris rétegekből állnak a végén sigmoid aktivációs függvényvel.

III. Tézispont: *Javaslatot tettem egy olyan algoritmusra, amely csecsemők (NIDCAP skála szerinti) viselkedési állapotait képes klasszifikálni, oly módon, hogy az egymást követő képkockák sorozatából 1D időbeli mozgási jeleket nyer ki (vizuális actigraph) és az így kapott jelsorozatot a pulzus variabilitás alapú jellemzőkkel kombinálva adja a visszacsatolásos neurális hálózat alapú osztályozó kaszkádnak bemenetként. Kimenete a viselkedési állapotok valamelyike és legjobb tudomásunk szerint ez az első olyan mesterséges intelligencia alapú osztályozó, amely ezen NIDCAP skála szerinti állapotokba osztályoz úgy, hogy bemenetként csak kamerából származó képi adatokat és a pulzus variabilitás alapú adatokat használja fel.*

Kapcsolódó publikáció: [C2], [Au1], [Au2]

IV. A szerző folyóirat publikációinak gyűjteménye

- [J1] **A. Nagy**, P. Földesy, I. Jánoki, D. Terbe, M. Siket, M. Szabó, J. Varga, and A. Zarándy, “Continuous camera-based premature-infant monitoring algorithms for nicu,” *Applied Sciences*, vol. 11, no. 16, pp. 1–24, 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/16/7215> 9)
- [J2] P. Földesy, I. Jánoki, **A. Nagy**, M. Siket, and A. Zarándy, “Temperature measurement with photodiodes: Application to laser diode temperature monitoring,” *Sensors and Actuators A: Physical*, vol. 337, p. 113441, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924424722000796> 19)
- [J3] P. Földesy, M. Siket, **A. Nagy**, and I. Jánoki, “Correction of overexposure in laser speckle contrast imaging,” *Opt. Express*, vol. 30, no. 12, pp. 21 523–21 534, Jun 2022. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-30-12-21523> 19)
- [J4] P. Földesy, M. Siket, I. Jánoki, K. Demeter, and **A. Nagy**, “Ensemble averaging laser speckle contrast imaging: statistical model of improvement as function of static scatterers,” *Opt. Express*, vol. 29, no. 18, pp. 29 366–29 377, Aug 2021. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-29-18-29366> 19)

V. A szerző konferencia publikációinak gyűjteménye

- [C1] **A. Nagy**, P. Földesy, I. Jánoki, M. Siket, and A. Zarándy, “Automatic annotation to train roi detection algorithm for premature infant respiration monitoring in nicu,” in *2023 In: Képfeldolgozók, és Alakfelismerők Társasága Képfeldolgozók és Alakfelismerők Társaságának 14. Országos Konferenciája*, no. 24, 01 2023, pp. 1–27. 15)
- [C2] I. Jánoki, **A. Nagy**, P. Földesy, A. Zarándy, M. Siket, J. Varga, and M. Szabó, “Neonatal activity monitoring by camera-based multi-lstm network,” 10 2022, pp. 1–4. 18)
- [C3] A. Zarándy, P. Földesy, **A. Nagy**, I. Jánoki, D. Terbe, M. Siket, M. Szabó, and J. Varga, “Multi-level optimization for enabling life critical visual inspections of infants in resource limited environment,” in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1–5. 19)

VI. További publikációk

- [Au1] P. Földesy, **A. Nagy**, Z. Róka, Zarándy, I. Janoki, J. Varga, and M. Szabó, “Automatic nidcap sleeping stage monitoring by deep learning network using video actigraphy and heart rate variability,” in *Poster presented at 32nd Annual Meeting of the European Society of Paediatric and Neonatal Intensive Care*. ESPNIC, 2023, p. 1. 18, 19)
- [Au2] **A. Nagy**, I. Jánoki, A. Zarándy, P. Földesy, M. Szabó, and J. Varga, “Computer vision based neonatal behavior stage detection and classification,” in *Poster presented at 9th Congress of the European Academy of Paediatric Societies*. EAPS, 2022, pp. 1–2. 18, 19)
- [Au3] **A. Nagy**, D. Chetverikov, and A. Zarándy, “Novel methods for video-based respiration monitoring of newborn babies,” in *2019 In: Képfeldolgozók, és Alakfelismerők Társasága Képfeldolgozók és Alakfelismerők Társaságának 12. Országos Konferenciája*, no. 22, 08 2019, pp. 1–10. 19)

Hivatkozások

- [1] Federico Bolelli, Stefano Allegretti, Lorenzo Baraldi, and Costantino Grana. Spaghetti labeling: Directed acyclic graphs for block-based connected components labeling. *IEEE Transactions on Image Processing*, PP:1–1, 10 2019. 11)
- [2] Aude BRIS, Nadia Mazille-Orfanos, Pauline SIMONOT, Maude LUHERNE, Cyril Flamant, Geraldine Gascoin, Gearóid ÓLaighin, Richard Harte, and Patrick Pladys. Parents’ and healthcare professionals’ perceptions of the use of live video recording in neonatal units: a focus group study. *BMC Pediatrics*, 20, 04 2020. 4)
- [3] Stephen Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6):536–541, 1930. 8)
- [4] G. de Haan, P.W.A.C. Biezen, H. Huijgen, and O.A. Ojo. True-motion estimation with 3-d recursive search block matching. *IEEE Transactions on Circuits and Systems for Video Technology*, 3(5):368–379, 388, 1993. 10)
- [5] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. volume 2749, pages 363–370, 06 2003. 7)
- [6] Susannah Fleming, Matthew Thompson, Richard Stevens, Carl Heneghan, Annette Plüddemann, Ian Maconochie, Lionel Tarassenko, and David Mant. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *The Lancet*, 377(9770):1011–1018, 2011. 13)

- [7] Deanna Hanson-Abromeit. The newborn individualized developmental care and assessment program (nidcap) as a model for clinical music therapy interventions with premature infants. *Music Therapy Perspectives*, 21:60–68, 01 2003. 4)
- [8] João Jorge, Mauricio Villarroel, Sitthichok Chaichulee, Alessandro Guazzi, Sara Davis, Gabrielle Green, Kenny McCormick, and Lionel Tarassenko. Non-contact monitoring of respiration in the neonatal intensive care unit. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 286–293, 2017. 4)
- [9] Lalit Maurya, Pavleen Kaur, Deepak Chawla, and Prasant Mahapatra. Non-contact breathing rate monitoring in newborns: A review. *Computers in Biology and Medicine*, 132:104321, 03 2021. 3, 4)
- [10] Á Nagy, D Chetverikov, and Á Zarándy. Novel methods for video-based respiration monitoring of newborn babies. In *2019 In: Képfeldolgozók, és Alakfelismerők Társasága Képfeldolgozók és Alakfelismerők Társaságának 12. Országos Konferenciája*, number 22, pages 1–10, 2019. 4)
- [11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 7)
- [12] Fred Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, September 2017. 17)
- [13] Phyllis K Stein, Matthew S Bosner, Robert E Kleiger, and Brooke M Conger. Heart rate variability: a measure of cardiac autonomic tone. *American heart journal*, 127(5):1376–1381, 1994. 17)
- [14] Yue Sun, Wenjin Wang, Xi Long, Mohammed Meftah, Tao Tan, Caifeng Shan, Ronald M. Aarts, and Peter H. N. de With. Respiration monitoring for premature neonates in nicu. *Applied Sciences*, 9(23), 2019. 4)

- [15] Yue Sun, Wenjin Wang, Xi Long, Mohammed Meftah, Tao Tan, Caifeng Shan, Ronald M Aarts, et al. Respiration monitoring for premature neonates in nicu. *Applied Sciences*, 9(23):5246, 2019. 10)
- [16] Michael E. Tipping and Christopher M. Bishop. Mixtures of Probabilistic Principal Component Analyzers. *Neural Computation*, 11(2):443–482, 02 1999. 13)
- [17] Inga Warren, Ezam Mat-Ali, Mark Green, and Dumisani Nyathi. Evaluation of the family and infant neurodevelopmental education (fine) programme in the uk. *Journal of Neonatal Nursing*, 25:93–98, Nov. 2018. 1, 2, 4)
- [18] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid. Deepflow: Large displacement optical flow with deep matching. *2013 IEEE International Conference on Computer Vision*, 2013. 10)
- [19] Jan Werth, Louis Atallah, Peter Andriessen, Xi Long, Elly Zwartkruis-Pelgrim, and Ronald M. Aarts. Unobtrusive sleep state measurements in preterm infants – a review. *Sleep Medicine Reviews*, 32:109–122, 2017. 3)
- [20] Björn Westrup. Newborn individualized developmental care and assessment program (nidcap): Family-centered developmentally supportive care. *Early human development*, 83:443–9, 08 2007. 4)