

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

Absztrakt

Egy koraszülött csecsemő légzési rátái, valamint aktív és alvó óráinak száma olyan paraméterek, amelyek elengedhetetlenek a fejlődéstámogató ellátáshoz a koraszülött klinikákon. Mindeddig ezen paraméterek figyelemmel követése vagy ritkán használt vezetékes rendszerek segítségével vagy a kórházi személyzet érzékszerveire támaszkodva történt. Most azonban nagy előrelépés figyelhető meg a gépi látási, képfeldolgozási és gépi tanulási tudományok széles körű, akár orvosi alkalmazásának terén, ami lehetőséget biztosíthat számunkra, hogy egész napos felvételekből származó adatokkal lássuk el az orvosi személyzetet és az egészségügyi intézményeket.

Ebben a munkában arra vállalkozom, hogy új algoritmusokat mutassak be koraszülött csecsemők légzésének érintésmentes kameraalapú monitorozására és a csecsemők aktivitási és alvási állapotainak osztályozására. A cél, hogy ezen bemutatott algoritmusok képesek legyenek a csecsemők 24 órás viselkedési állapotairól és légzési rátájáról adatokat szolgáltatni, függetlenül attól, hogy milyen orvosi vagy egyéb beavatkozások történnek a mérések közben.

Reményeim szerint a bemutatott algoritmusok nemsokára ipari alkalmazásokban kerülnek felhasználásra és új típusú adatokat szolgáltatnak az orvosok számára, valamint tehermentesítik a személyzetet a folyamatos megfigyelés alól.

Köszönetnyilvánítás

Mindenekelőtt szeretném kifejezni köszönetemet családomnak, különösen édesanyámnak Volner Nagy Melindának és édesapámnak Nagy Ivánnak szeretetükért, türelmükért és megingathatatlan támogatásukért. A bátorításuk és a belém vetett hitük volt a legnagyobb erőm és motivációm.

Továbbá szeretném megköszönni a SZTAKI és PPKE ITK támogatását és segítségét, amiért biztosították számomra a kutatásom elvégzéséhez szükséges forrásokat és finanszírozást.

Nagyon köszönöm témavezetőmnek Prof. Dr. Zarándy Ákosnak a folyamatos útmutatást és támogatást.

Köszönöm Csetverikov Dmitrij professzor úrnak, aki megalapozta a kutatásom irányát és nagyon sokat segített az első működő algoritmusaim elkészítésében.

Külön szeretném kifejezni mélyszéges köszönetemet a SZTAKI-ban dolgozó kollégáimnak, akik nélkül ez a munka ebben a formában nem valósulhatott volna meg. Különös tekintettel köszönöm (az általam elnyert KDP pályázatban) vállalati szakértőmnek Dr. Földesy Péternek, Jánoki Imre, Siket Máté és Terbe Dániel kollégáimnak, valamint Róka Lilla Zitának és Kiss Dorottyanak. Szintén külön köszönetet mondanék orvosi partnereinknek Dr. Szabó Miklósnak és Dr. Varga Juditnak a Semmelweis Egyetemről.

Ezúton is szeretném kifejezni őszinte köszönetemet bizottságom tagjainak, élelátó megjegyzéseikért, építő kritikáikért és értékes javaslataikért, amelyek segítettek javítani a munkámat.

Végül köszönöm barátaimnak és támogatóimnak az egyetemen és azon kívül, akik mindig rendelkezésre álltak, ha segítségre volt szükségem.

Tartalomjegyzék

Absztrakt	1
Köszönetnyilvánítás	2
Rövidítések	4
1. Bevezetés	6
1.1. Koraszülött csecsemők fiziológiai jeleinek folyamatos monitorozása . .	7
1.2. Érintésmentes monitorozás	9
1.3. A csecsemők cirkadián ciklusainak monitorozása egész napos felvéte- lekből	11
1.4. Kutatási előzmények	12
1.4.1. Saját terület definiálása a projekten belül	14
1.4.2. Kezdeti kísérletek a kutatás korai szakaszában	15
2. Szakirodalom áttekítése	17
3. Adatgyűjtés és annotáció	21
3.0.1. Adatgyűjtő rendszer viselkedési állapotok klasszifikációjához	25
4. Mozgás alapú mérések orvosi területen és általánosságban	29
5. Érintésmentes légzésfigyelő algoritmusok: előzmények és a java- solt algoritmus	34
5.1. Csecsemők légzésfigyelésének relevanciája	34
5.1.1. Kapcsolódó szakirodalom	36
5.1.2. Adatgyűjtés	37
5.1.3. Előzmények	38

5.1.4.	Sűrű optikai áramlás alapú megközelítés és a végponttól-végpontig "end-to-end" neurális hálózat alapú megoldás	42
5.1.5.	A probléma megoldására javasolt algoritmus: vegyes megközelítés (gépi tanulás alapú ROI detektor és sűrű optikai áramlás alapú algoritmus)	48
6.	Automatikus annotált adathalmaz generálás	64
6.1.	Kapcsolódó szakirodalom	66
6.2.	Adatgyűjtő	67
6.3.	A javasolt algoritmus leírása	67
6.3.1.	Mozgás becslés	69
6.3.2.	Terület alapú szűrés	69
6.3.3.	Frekvencia alapú szűrés	71
6.3.4.	Elmozdulás alapú szűrés	72
6.3.5.	Követési modul	73
6.4.	Értékelés	74
6.4.1.	Pixel szintű értékelés	75
6.4.2.	Légzési ráta becslő teljesítményének értékelése	76
6.4.3.	A javasolt algoritmus összehasonlítása a szakirodalomban fellelhető eljárásokkal	80
7.	Alvási és aktivitási (viselkedési) fázisok monitorozása	82
7.1.	Motiváció a csecsemők viselkedési állapotainak klasszifikációjához . .	82
7.2.	Kapcsolódó irodalom	84
7.3.	Adatgyűjtő	85
7.4.	A probléma definiálása	85
7.5.	Az ajánlott algoritmus: viselkedési állapot becslő	86
7.6.	Az ajánlott algoritmus eredményei és azok értékelése	93
7.7.	Összehasonlítás a versenytársakkal	97
7.8.	A fejezet konklúziója	99
8.	Diszkusszió és tézisek	100

8.1. Új tudományos eredmények	100
8.2. Jövőbeni tervek	102

Jelölések

A munka során használt képletek és szimbólumok jobb megértése érdekében a következőképpen különböztetjük meg a különböző matematikai objektumokat: Az állandó skalárokat görög vagy dőlt kisbetűvel jelöljük (pl. α vagy a), az egydimenziós tömböket pedig félkövér kisbetűvel (e.g. **a**), a 2-dimenziós mátrixokat nagybetűvel (pl. **A**), háromdimenziós tenzorokat pedig félkövér nagybetűvel jelöljük (pl. **A**). A fentiek bármelyike felülvonallal alkalmazva a megfelelő elemek halmazát jelöli (pl. $\overline{A} = \{A_1, A_2, A_3\}$).

Rövidítések

1D	egy dimenziós	17
2D	két dimenziós	30
AI	Artificial Intelligence	23
ALA	Automatikus Annotáló Algoritmus - Automatic Labelling Alorithm	76
CNN	Convolutional Neural Network	17
EEG	elektro-enkefalogram	6
EKG	elektro-kardiogram	6
EMG	elektro-miogram	6
EOG	elektro-okulogram	6
FFT	Gyors Fourier Transzformáció	14
GRU	Gated Recurrent Unit	87
HSV	HSV (hue, saturation, value) színtér	88
LPP	Légzés/perc	17
LR	Légzési ráta	35
LSTM	Long Short Term Memory	43
MAE	Mean Absolute Error	60
NICU	Neonatal Intensive Care Unit	34

NIDCAP	Newborn Individualized Developmental Care and Assessment Program - Újszülött Egyéni Fejlesztési Gondozási Program	6
OF	Optical Flow	38
OSA	Obstruktív Alvási Apnoe	34
PCA	Principal Component Analysis	71
Philips IntelliVue MP20/MP50	A kórházakban használt életjel monitor	21
PIC	Koraszülött Intenzív Osztály	14
POS	Plane Orthogonal to Skin	13
PPG	A PPG a véráramlás változásait mutatja hullámformában	89
PRV	Pulse Rate Variability	89
REM	Rapid Eye Movement Sleep	19
RMSE	Root Mean Squared Error	60
RNN	Rekurrens Neurális Hálózat	86
ROI	Region Of Interest	2
SVD	Singular Value Decomposition	39
U-Net	A U-Net szegmentáló konvolúciós hálózat	49

1. Fejezet

Bevezetés

Az olyan fiziológiai jelek, mint a pulzus, a légzés vagy a véroxigén szint rutinszerűen mért paraméterek az intenzív osztályokon és a koraszülött klinikákon is. Ezek az adatok elengedhetetlenek az orvosok számára ahhoz, hogy figyelemmel kövessék a páciensek állapotát és megfelelő diagnózist állíthassanak fel. Az e fiziológiai jelek mérésére alkalmas szenzorok elterjedésével lehetőség nyílt az orvosok számára digitális eszközöket használni az említett paraméterek gyűjtésére és elemzésére. Így egyfajta személyre szabott kezelést képesek nyújtani a pácienseik számára. A szenzorok olcsóbb, orvosi tanúsítvánnyal nem rendelkező változatai pedig megtalálhatóak már a háztartásokban bárhol, így fitness célokra is felhasználhatóak ezen eszközök. Számos applikáció segíti az orvosi képzettséggel nem rendelkező embereket fiziológiai jeleik mérésében és értékelésében.

Ugyanilyen, 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 skála szerint osztályozást követik. Ezek az állapotok annak mérésére 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 mérése egyre fontosabb az öregotthonokban is. Az aktivitás mérése is egy olyan terület, ahol az egész napos megfigyelések esetén nem várhatjuk el, hogy a páciensekhez állandóan poliszomnográfias szenzorok (EEG, EKG, EMG, EOG, stb.) legyenek

csatlakoztatva.

Bizonyos területeken a fiziológiai jelek és az aktivitás mérése sem engedhető meg olyan szenzorokkal, amelyek a bőrfelülettel közvetlenül érintkeznek. Ilyen területek a sérülékeny vékony bőrrel rendelkező koraszülött csecsemők vagy égési sérülésekkel rendelkező páciensek megfigyelése. Ezen felül a testfelülethez csatlakoztatott vezetékes szenzorok kényelmetlenek és nem teszik lehetővé a hosszú távú megfigyelést. Éppen ezért az olyan esetekben, ahol fekvő betegeket vagy koraszülött csecsemőket figyelnek meg egyre népszerűbbé válik az olyan (pl. kamera alapú) mérési módszerek fejlesztése, amelyek nem igényelnek vezetékes érintkezést a bőrfelülettel.

A mesterséges intelligencia és a gépi látás elterjedése egy új világot nyitott meg a különböző mozgások megfigyelésében és értékelésében. Elég csak a két és háromdimenziós "póz becslésre" vagy a különböző mozgásbecslő (pl. optikai áramlás alapú) algoritmusok fejlődésére gondolni, amelyeknek óriási irodalma van. Ugyanígy nagyon sok cikk látott napvilágot a fiziológiai jelek érintésmentes, mesterséges intelligencia vagy gépi látás alapú méréséről, amelyekről bővebben fogok beszélni a 2. alfejezetben. Az új technológiák és megközelítések most juttatják el ezt a területet arra a szintre, hogy már a kórházakban rutinszerűen felhasználható, orvosi tanúsítvánnyal rendelkező algoritmusok születhessenek.

1.1. Koraszülött csecsemők fiziológiai jeleinek folyamatos monitorozása

Koraszülöttnak azokat a csecsemőket nevezzük, amelyek 37 hetesnél fiatalabban születnek meg, míg a normális születés 40 hetes korban történik [81]. Az ilyen csecsemőkre jellemző az aránytalanul nagy fej, a vékony bőr és végtagok. A koraszülött csecsemőknél fennáll a nem megfelelő fejlődés (pl. agyi fejlődés kockázata), amely hosszútávú problémákhoz vezethet, mint amilyenek az agyi bénulás, fejlődési késések, mozgás és testtartás szabályozási problémák, valamint hallás és látásproblémák [3].

Mint ahogy fentebb említettem, a fiziológiai jeleket rutinszerűen mérik a kórházakban, (különösen a koraszülött osztályokon), olyan orvosi monitorok segítségével,

mint a "Phillips IntelliVue MP40/MP50". 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 [118], hogy fejlődéstámogató kezelést tudjanak nyújtani. Erre azonban a legtöbb klinikán nincs elegendő erőforrás. Érdekes lenne a folyamatot valamilyen módon automatizálni. A fiziológiai jelek esetében más a helyzet. Ezeket a jeleket elektromos szenzorokkal méri, ahogyan említettem. A pulzus és a légzés ugyanis fontos információkat szolgáltat az orvosok számára a csecsemők keringési és légzési rendszerének fejlődéséről. Az olyan rendellenességek például, mint a "légzési apnea" (rövid légzéskimaradás) felismerése fontos a korábban említett fejlődési rendellenességek elkerüléséhez. A légzés újraindításához a kórházi személyzetnek sok esetben egy különleges masszázst kell alkalmaznia. A csecsemők aktivitásának és alvásának a megfigyelése pedig fontos a csecsemők alvásmennyiségének és aktív óráinak optimalizálásához, amely kritikus szempont ahhoz, hogy a megfelelő agyi fejlődésüket biztosíthassuk [37].

Egyértelmű, hogy az említett fiziológiai jelek és alvási aktivitási fázisok monitorozása elengedhetetlen a modern orvostudományban a koraszülött csecsemők megfelelő ellátásához és a manapság egyre népszerűbbé váló "személyre szabott ellátás" biztosításához.

A fiziológiai jeleket folyamatosan monitorozzák és kijelzik a kórházi személyzet számára. A pulzus és a véroxigénszint esetében az bőrön áthaladó vagy az arról visszanyert fény elemzése segít a pulzus és a véroxigénszint aktuális értékének "nem-invazív" módon történő meghatározásában. Itt a transzmissziós fotopletizmográfiát alkalmazzák, amelynél a szenzort egy vékony testrésze (pl. az ujjbegyre kell helyezni), hogy a szenzor két speciális hullámhosszúságú fényt bocsásson át az adott testrészen, a testrésztől a másik oldalán található fotódiodába. A szenzor közben képes mérni a változó abszorbanciát az egyes hullámhosszokon [79]. A pulzus esetében itt csak a vér mennyiségének és az ebből adódó abszorbancia változásának a mérése történik. A véroxigénszint (SaO_2) esetében azonban az oxigenizált hemoglobin mennyiségéből következtethetünk az SaO_2 értékére, ugyanis az oxigenizált hemoglobin mennyisége arányos az SaO_2 értékével, valamint az adott hullámhosszokon (pl.

767 és 811 nm között) más az elnyelése, mint a nem oxigenizált hemoglobin-énak. Következésképpen az abszorbeált fény mennyisége alapján következtethetünk vér oxigéntartalmára és az SaO_2 értékére.

A légzés esetében EKG-t használnak, ugyanis az EKG képes mérni az egyes elektródái között rezisztencia változást és ebből következtetni a tüdő térfogatváltozására, vagyis a légzésre. Így tehát a légzést a bőrfelülettel közvetlenül érintkező elektródák közötti átvezetett áram segítségével lehet mérni, amely elektródákat a bőrfelületre kell ragasztania a megfelelő szakértelemmel rendelkező személynek. Csecsemő EKG-ból is létezik 3 és 5 elektródás rendszer.

1.2. Érintésmentes monitorozás

A koraszülött osztályokon azonban nem minden esetben problémamentes a fiziológiai jelek mérése, ezek a mérések ugyanis általában a bőrfelülettel közvetlenül érintkező szenzorok segítségével történnek, mint ahogyan azt fentebb említettem. A pulzus esetében a pulzoximéter is közvetlenül érintkezik a csecsemő kezén vagy lábán a bőrfelülettel, azonban ezen szenzor esetében egyetlen vezetékről van szó, nincs elektróda, amit fel kellene tapasztani a bőrre és a készülék megbízhatósága megkérdőjelezhetetlen. Így a pulzus és véroxigénszint esetében kevésbé merül fel a jelenlegi "kontakt-szenzorok" lecserélésének igénye. Ettől függetlenül mi is foglalkoztunk a pulzust érintés-mentesen mérni képes algoritmusok fejlesztésével, amelyben én is részt vettem. De az említett okok miatt nem a pulzusmérés képezte a kutatásom fő irányvonalát.

Ezzel szemben a légzés mérése már komolyabb problémákkal jár koraszülött csecsemők esetében. Mivel a kórházakban jelenleg EKG-t alkalmaznak erre. Ennek vezetékei azonban zavarhatja a kórházi személyzetet munkájuk elvégzésében és komoly stressz alá helyezheti a szülőket is a látvány. Valamint az elektródák akár fel is sérthetik a csecsemők vékony bőrét. Az már csak másodlagos szempont, hogy az egyszer használatos elektródák nagy tömegben való vásárlása nem olcsó és megfelelő felhelyezésük szakértelmet igényel. Az EKG-t ezért nem is használják olyan gyakorisággal a koraszülött klinikákon, mint a pulzoximétert, csupán indokolt esetben,



1.1. Ábra: A 3D nyomtatott kamera amelyet a MEDICOR beépített az inkubátor falába. Így a kamera fix pozícióból nézheti az inkubátorban lévő csecsemőt.

és ezért a légzést sem mérik olyan gyakran, mint a pulzust pedig ennek folyamatos mérése is értékes információt hordozhat. A légzési "apnoe" detektálásában és kezelésében pl. komoly szerepe lehet az állandó légzés megfigyelésnek.

Az említett probléma megoldásának egy lehetséges módja, ha az EKG alapú légzésmérést kamera alapú érintésmentes módszerre cseréljük. Így mozgásalapú algoritmusok vagy hőkamerák segítségével lehetünk képesek meghatározni a csecsemők pulzusát. Ezen módszerek nagy előnye, hogy nem igényelnek semmilyen vezetékot, a szenzor nem érintkezik a csecsemők bőrfelületével, lehetőleg egyáltalán nem zavarja az orvosi személyzetet a munkájuk elvégzésében és a kamerakép a mért légzési jelen kívül egyéb információkat is szolgáltathat az orvosok számára, mint amilyenek a csecsemők bőrének színváltozása vagy a különböző kezelések (pl. UV megvilágításos kezelések) időpontja, vagy a mozgási aktivitási adatok.

A kamera alapú módszerek tekintetében azonban szembesülnünk kell azzal a kihívással, hogy a csecsemő látható kell legyen a képen. Ennek legegyszerűbb módja, hogy a kamerát az inkubátorba szereljük. Ahogy a MEDICOR is tette a velünk kooperációban folytatott kutatás sikerének érdekében (lásd az 1.1. ábrát). Az ilyen mérési elrendezések mellett is fennáll azonban annak a lehetősége, hogy valamilyen objektum például a kezelést végző ápoló vagy orvos keze kitakarja a csecsemőt. Az ilyen szituációk kezelésére egy külön érintésmentes kamera alapú algoritmus terveztünk a "szituáció analízálót", melynek szükségességét és működését a későbbiekben

be fogom mutatni.

Az érintésmentes monitorozás ezenfelül lehetővé teszi, hogy a csecsemők viselkedési állapotairól, aktivitási és alvási fázisairól adjunk információt. Ez történhet tisztán kameraalapon vagy egyes a bőrfelülettel közvetlenül érintkező szenzorok (pl. pulzoximéter) adatainak és a kameraképekből kinyert információk együttes felhasználásával is. Ahogy azt korábban említettem, ezen aktivitási fázisok meghatározása jelenleg a kórházi személyzet feladata és manuálisan történik. Nagy lépés lenne, ha a jelenlegi módszert lecserélhetnénk egy teljesen automatizált kameraalapú megoldásra, mellyel tehermentesíthetnénk az orvosokat és ápolókat és megbízható nagy mennyiségű adathoz juttatnánk a koraszülött klinikákat. Éppen ezért kutatásom az ilyen jellegű érintésmentes algoritmusok fejlesztését is célozza. A viselkedési állapotok folyamatos és pontos klasszifikációja nem csak tehermentesíti a kórházi személyzetet ezen feladatok elvégzése alól, de lehetőséget adhat, hogy elég adat birtokában a koraszülött klinika úgy ütemezze a különböző kezeléseket, hogy az szinkronban legyen a csecsemők cirkadián és ultradián ritmusaival, így biztosítva a megfelelő mennyiségű pihenést a megfelelő időben.

1.3. A csecsemők cirkadián ciklusainak monitorozása egész napos felvételekből

Alapvető kitűzött célunk volt egy olyan mesterséges intelligencia és gépi látás alapú rendszer létrehozása, amely érintésmentesen tudja mérni az inkubátorokban lévő koraszülött csecsemők életjeleit, mint amilyenek a pulzus, a légzés, a bőr színváltozás, a mozgási aktivitás, a fényerő és a csecsemők alvás mennyisége. Alapvetően a teljesség igénye nélkül 5 olyan területet azonosítottam, amelyek esetében biztosan indokolt a hagyományos vezetékes rendszerek helyett érintésmentes kameraalapú algoritmusokat alkalmazni. Ezek a következők:

- Légzés monitorozása
- Csecsemők bőrszínváltozásának megfigyelése

- Szituációanalízis: Nyugodt periódusok és kezelések idejének megkülönböztetése
- Csecsemők érzelmi állapotának felmérése
- Csecsemők aktív és alvási periódusainak megkülönböztetése

Ezen területekből én a légzés monitorozásával, a szituációanalízissel és a csecsemők aktív és alvási periódusainak megkülönböztetésével foglalkoztam behatóan a munkám során. Leginkább mozgás alapú és mesterséges intelligencia alapú megoldásokat alkalmaztam és mivel ezek adatvezérelt algoritmusok ezért tanításukhoz, optimalizációjukhoz nagy mennyiségű adatra volt szükségünk. Ezenfelül a csecsemők különböző napi nyugodt és kezelési, valamint aktív és alvó periódusainak megfigyelése is azt igényli, hogy nagy mennyiségű felvétellel rendelkezünk, mind éjszakai, mind pedig nappali periódusokról. Így tehát, mint azt később bővebben is kifejtem majd egy általunk fejlesztett adatgyűjtő segítségével 24 órás felvételt készítettünk a csecsemőkről. Így pedig lehetőségünk nyílt, hogy a különböző érintésmentes algoritmusainkat együtt alkalmazva és "kontakt-szenzoros" adatokkal kiegészítve magasabb szintű információkat nyújthassunk a csecsemők napi cirkadián és ultradián ritmusairól, becsülhessük alvásmennyiségüket és megfigyelhessük aktív és kezeléssel töltött idejüket új típusú komplex információt szolgáltatva a kórházi személyzet számára. Nevezetesen a csecsemők egész napos megfigyeléséből származó cirkadián ciklushoz köthető információkat.

1.4. Kutatási előzmények

A munkámat 2018-ban kezdtem a Számítástechnikai és Automatizálási Kutatóintézet Számítógépes Optikai Érzékelés és Feldolgozás Kutatólaboratóriumában. Ebben a fejezetben bemutatásra kerülnek a kutatási előzmények, kitérve arra, hogy pontosan milyen előzményekkel indult a kutatási projekt, milyen kész megoldásokra építettem és hogyan rajzolható körül az általam kutatott terület a projekten belül. A fejezet végén ezen felül röviden bemutatom a kutatási témám irodalmi hátterét.

A disszertációm témája, mint azt említettem, nem saját kútfőből származó ötlet, hanem egy nagyobb, évek óta zajló projekt része, amelynek keretein belül ipari partnerek és a klinika segítségével történik a csecsemők érintésmentes megfigyelésére alkalmas szoftverek és hardverek fejlesztése. Gondolok itt a kiépített adatgyűjtő és megfigyelő rendszerre vagy inkubátorba építhető kamerákra és olyan algoritmusokra, amelyek képesek érintésmentesen mérni a csecsemők pulzusát, vér-oxigénszintjét, légzését, színváltozását, aktivitását, hőmérsékletváltozását, valamint automatikusan naplózni képesek a csecsemők körül történő összes eseményt, mint a fényváltozások, ápolási és kezelési szituációk, alvás, ébrenlét, vagy a baba inkubátorból történő kivétele.

Az én feladatom az érintésmentes légzés monitorozó algoritmus fejlesztése és a mozgási aktivitás érzékelése és mozgások követése volt, valamint az ezek algoritmusok fejlesztésére és tanítására szükséges infrastruktúra kiépítése. Az én kutatásom fókusza tehát a csecsemő mozgási jeleinek (légzés, mozgásaktivitás) megfigyelésére koncentrált már a kezdetektől fogva.

Amikor elkezdtem a munkát a laborban a kollégáim már implementáltak egy klasszikus algoritmust, a pulzust érintésmentes mérésére. A klasszikus szót itt abban az értelemben használom, hogy nem a modern adat-vezérelt eljárásokhoz hasonló, amelynek paramétereit a bemeneti adatok vagy felügyelt tanulás esetén a bemeneti és a kimeneti adatok alapján állítjuk be. Egyszerűbben megfogalmazva nem adat-vezérelt eljárásról beszélünk.

Ezenfelül szintén rendelkezünk már egy-egy érintésmentes klasszikus légzésfigyelő algoritmussal, amelyet Dmitry Chetverikov professzor úr és G. Morse publikáltak is a "Video-based Respiration Monitoring of Newborn Babies: a Feasibility Study" című konferencia-közleményben.

A fentebb említett légzés és pulzus figyelő algoritmusok:

- Klasszikus Pulzus Figyelő Algoritmus: Ebben az eljárásban bőrszín alapú ROI detekció történik, majd az egymást követő képkockákat tartalmazó csúszó ablakon folytatódik a feldolgozás. A ROI-ba tartozó minden egyes pixelen az időtengely mentén kerül végrehajtásra a "Plane Orthogonal to skin" (POS) algoritmus, hogy egy egydimenziós jelet kapjunk minden egyes pixelhez. Ezekre

az egydimenziós jelekre aztán FFT kerül alkalmazásra.

- Klasszikus Légzés Figyelő Algoritmus: Ebben az eljárásban jellemzőpontokat detektáltak és egy vagy több jellemzőpont mozgását követték. Ha a csecsemő nem mozdult el, illetve nem mozdították el akkor ebben az esetben az önkényesen kiválasztott ROI-n belüli jellemzőpontmozgás a csecsemő légzéséhez lesz köthető. Az algoritmusban a jellemző pontok követéséhez a Kanade-Lucas-Tomasi jellemző követés eljárást alkalmazták, hogy számszerűsíthessék a pontok mozgását.

1.4.1. Saját terület definiálása a projekten belül

A csatlakozásomkor a projekt egy új fázisba lépett. Elkezdődött egy olyan rendszer kiépítése, amely nagyobb mennyiségű adat felvételére képes és olyan szoftverelemek megalkotása, amelyek pedig adatelemzésre és annotálásra képesek. Valamint új, optimalizáltabb (mély tanulást is alkalmazó) adat-vezérelt algoritmusokat kezdtünk el fejleszteni érintésmentes monitorozáshoz.

Az én kutatásom fókusza ezen projekten belül a csecsemő aktivitásának és légzésének érintésmentes mérésére lokalizálódott, de részt vettem az érintésmentes neurális hálózat alapú pulzus monitorozó algoritmus fejlesztésében is, valamint annak a rendszernek a kiépítésében, amellyel képesek voltunk a SOTE Koraszülött Intenzív (PIC) osztályain egész napos felvételeket is készíteni. Mindazonáltal, ahogyan említettem az én kutatásaim legfőképp a különböző mozgási aktivitásokból származtatható mutatók kinyerésére és értelmezésére koncentrált (mint amilyenek a légzés vagy a csecsemő cirkadián ritmusa). Ennek szellemében pedig ebben a disszertációban az ezen jelek vagy mutatók érintésmentes mérésének problémakörét fogom körbejárni. Bemutatva a probléma megoldására alkalmas lehetséges módszereket és az ezen módszerek tanításához és/vagy értékeléséhez szükséges adatokat és az ezen adatok gyűjtésére létrehozott rendszert.

1.4.2. Kezdeti kísérletek a kutatás korai szakaszában

A SZTAKI Számítógépes Optikai Érzékelés és Feldolgozás Kutatólaboratórium-ban már a doktori tanulmányaim megkezdése előtt is elkezdtem dolgozni a témában. Ebben az alfejezetben röviden leírnám milyen kísérletek történtek a laborban még mielőtt a doktori kutatás megkezdődött. A laborban a munkám megkezdésekor először a légzés érintésmentes monitorozásának problémakörével kezdtem el foglalkozni. A kísérleteim során felmerült a kérdés, hogy milyen szenzor lenne a legalkalmasabb a csecsemők légzésének érintésmentes monitorozására. Itt számos szempontot figyelembe kellett vennem, mint amilyenek a következők:

- a szenzor ára,
- a szenzor megbízhatósága,
- felhasználható-e a szenzor egyéb a légzéstől eltérő mutatók monitorozására,
- a csecsemők szempontjából biztonságos-e a szenzor,
- a kórház engedélyezi-e a szenzor használatát vagy tudja-e engedélyezni,
- egyéb

Az orvosi területen történő fejlesztések egyik nehézsége, hogy nem használhatunk olyan eszközt, amely nem rendelkezik a megfelelő orvosi tanúsítványokkal. Ezért ebben a szakaszban még el is vetettük a csecsemők légzésének infrakamerás megfigyelését, mivel még nem volt birtokunkban olyan infravörös megvilágító, amely megfelelő orvosi tanúsítvánnyal rendelkezik. Mivel érintésmentes monitorozásról van szó ezért természetesen kamerák alkalmazásában kezdtünk el gondolkodni. Felmerült még ezen kívül a hőkamera alkalmazása, amely bár drágább, mint a hagyományos kamerák, azonban további adatokat szolgáltat a csecsemők hőeloszlásáról. Egy "Seek Thermal Compact" típusú hőkamerával kezdtem el kísérletezni. Termográfia alkalmazására koraszülötteknél voltak példák az irodalomban is (lásd [5]), ezeknél azonban nagyon nagy hőfelbontású kamerákat alkalmaztak. Az említett cikkben például egy olyan kamerát, amelynek 0.05°C a szenzitivitása 30°C -on. Valamint a cikkben közölt eredményeket minden valószínűség szerint csak a megfelelő szögből lehetett

volna reprodukálni, ahol a képen alulról látszanak a csecsemő orrlyukai. Az ehhez hasonló póz pedig nem mindig biztosítható. Így az általunk vizionált 24 órás valós idejű értékelés nem lett volna lehetséges. Valamint az általam alkalmazott hőkamera felbontása nem volt elégséges ahhoz, hogy a légzést megfelelően ki lehessen vonni az inkubátorokban történt felvételekből. Ennek oka, hogy az inkubátorokban meleg van (34-39°C). A hagyományos kamerák ráadásul olcsóbbak és könnyebben használhatóak voltak. Ezen okok miatt a hőkamera alkalmazását elvetettem és áttértem a Chetverikov professzor úr által fejlesztett hagyományos kamera alapú légzés figyelő eljárás továbbfejlesztése és sűrű optikai áramlás alapú algoritmussal való kiegészítése felé, amelyekről a későbbiekben írok bővebben.

2. Fejezet

Szakirodalom áttekítése

Ebben az fejezetben röviden a teljesség igénye nélkül áttekintem a témával kapcsolatos szakirodalmat legalábbis amely azokat a területeket érinti amelyekre én magam ebben a munkában kiemelten fókuszáltam.

Maga az érintésmentes monitorozás nem kifejezetten csecsemőkre vonatkozóan is rendkívül széles irodalommal rendelkezik. A kamera alapú pulzusmérés és légzés mérés is elterjedt technológiáknak számítanak már a klinikumban és a fitnessz irányú alkalmazások terén egyaránt [43, 50, 59, 62, 88, 117].

Ezek közül külön kiemelném a neurális hálózat alapú megközelítéseket, amelyek jobb teljesítményt érnek el általánosságban, mint a klasszikus megközelítések, főként az érintésmentes pulzusmérés tekintetében, amelynél az alkalmazott neurális hálózatok gyakran légzéskinyerésre is használhatóak [21, 33, 104, 124]. Kifejezetten légzés érintésmentes mérésére kifejlesztett mesterséges intelligencia alapú algoritmusok terén pedig (az 1D konvolúciós hálózatoktól kezdve az optikai áramlást elemző CNN-ig) találunk irodalmat [40, 47, 85].

A legújabb érintésmentes légzésfigyelő rendszereket leíró publikációk ([75],[65],[15]) tanulsága szerint napjainkban a legjobb eredmények a referenciával számított (Mean Absolute Error - MAE) tekintetében: 0.1004, 1.7 és 2.807 LPP körül helyezkednek el. Azonban a teljesítmény nyilvánvalóan erősen függ az alkalmazott módszer attribútumaitól és az értékelésnél használt adathalmaztól. (Pl. a [75] publikáció esetében, ahol a legjobb eredmény a többi esettől jóval erősebb volt, hőkamera alapú módszert alkalmaztak és a 0.1004-es MAE érték esetében a vizsgált személyek egy

széken ültek a kamera előtt kényelmesen és saját ütemben lélegezve.) Az alkalmazott szenzor és az adathalmazban lévő felvételek alatt történő szituációk tehát nagyban befolyásolják, hogy milyen eredmény érhető el, ahogy az várható is.

Mielőtt rátérnék a kifejezetten a csecsemők megfigyelésére alkalmazott algoritmusokra, szeretnék kitérni röviden az aktivitás megfigyelésére általánosságban. Az aktivitás mérését általában alvási állapotok meghatározására szokták alkalmazni [28, 48, 63, 69]. Az aktigráfia a mozgás számszerűsítésére alkalmas mérés és általában végtagokra rögzíthető gyorsulásmérő szenzorokkal (aktigráfokkal) szokás mérni de vannak példák, kamera alapú aktigráfiára [42, 125], vagy kamera alapú mozgás detekcióra is [1].

A koraszülött csecsemők érintésmentes monitorozása egy nagyon népszerű téma napjainkban ugyanis a kórházakban egyre nagyobb hangsúlyt fektetnek a személyre szabott kezelésekre és a komfortérzet növelésére. A fiziológiai paraméterek folyamatos monitorozása elengedhetetlen a személyre szabott kezelések kialakításához a vezetékes érintkező szenzorok azonban nem minden esetben helyezhetők fel vagy tarthatók folyamatosan a csecsemőn, valamint a vezetékek nagyban csökkentik a komfortérzetet is, ezért egy új irányelvként jelent meg, hogy próbáljon minél inkább elmozdulni a klinikum az érintésmentes monitorozás felé. Erről természetesen rengeteg publikáció született. Külön foglalkoznak cikkek csak a légzés mérésével [52, 74, 106] [Au1], de olyan publikációk is vannak, amelyek inkább a pulzus mérését vizsgálják [44, 80, 103, 104]. Természetesen olyan cikkeket is publikáltak, amelyek teljes rendszereket írnak le, olyanokat, amelyek nem egy, hanem többféle fiziológiai jel érintésmentes monitorozására képesek párhuzamosan valós időben kórházi körülmények között [114, 115] [C1] [J1]. Az újabb publikációkat tekintve csecsemők érintésmentes légzés monitorozásával kapcsolatban érdemes megemlíteni Maurya et al. és Rajalakshmi et al. munkáit ([74], [91]), amelyek áttekintését nyújtják a koraszülöttek érintésmentes légzési ráta monitorozásának és az inkubátorban lévő koraszülöttek megfigyelő rendszereinek. A 2020 - 2022 időszakból származó érintésmentes csecsemő légzésfigyeléssel foglalkozó publikációk közül pedig kiemelném a következőket: [96], [57], [55].

A fiziológiai jelek érintésmentes monitorozásával szemben a csecsemők aktivi-

tásának kamera alapú monitorozásáról már valamivel kevesebb cikk lelhető fel az irodalomban [32, 78, 99, 111]. Mivel a kamerák használata az inkubátorokban még nem rutinszerű dolog, ezért sok esetben kamerák helyett aktigráfokat használnak. Ezek fitneszórákhoz hasonló (általában a végtagokra erősíthető) accelerométerrel vagy esetleg pulzoximéterrel felszerelt szenzoreszközök, amelyek a bőrfelülettel közvetlenül érintkezve mérik a csecsemők mozgási aktivitását.

Mivel a csecsemők aktivitásán én nem csak a végtagmozgásaik, hanem a cirkadián ritmusuk megfigyelését, vagyis a viselkedési fázisaiknak (csendes alvás fázis, aktív alvás fázis, átmeneti fázis, csendes ébrenlét fázis, aktív ébrenlét fázis, sírás fázis) megfigyelését is értem, ezért most az ezzel kapcsolatos cikkeket is áttekinteném. Alvási fázisokon azonban nem a szokásos „REM” és „mély alvás” fázisokat értem, hanem azt az öt aktivitási vagy viselkedési fázist, amelyet a csecsemőt megfigyelő orvosi személyzet képes megállapítani a NIDCAP skála szerint, (az Újszülött Egyéni Fejlesztési Gondozási Programnak megfelelően [118]): [121],[41],[16].

Az orvosi partnereink is nyilvánvalóvá tették, hogy egy nagyon hasznos információ annak megbecsülése, hogy a csecsemők mennyit alszanak. A csecsemők alvás megfigyelésével nyilvánvalóan számtalan cikk foglalkozik legyen szó "kontakt-szenzor" alapú vagy "non-kontakt-szenzor" alapú eljárásokról [10, 67, 76]. De akad olyan cikk is, amely kifejezetten az egész napos cirkadián ritmusuk megfigyelését mutatja be [56]. Érdemes továbbá kiemelni Muhammad Awais et al. és Xi Long et al. munkáját [9], [68], mivel ezekben videóalapon osztályozzák a csecsemők alvási állapotait. Muhammad Awais et al. az arcról készült videókban lévő arkifejezések alapján osztályoz, míg Xi Long et al. munkája videóalapú aktigráfia alapján. Az említett publikációk hasonló témát dolgoznak fel, mint amellyel mi is foglalkozunk a 7. fejezetben és amely a 3. tézisem témáját adja, azonban kissé eltérő módon közelít meg a problémát. Továbbá, (a NIDCAP skála alapú) kategorizálás tekintetében érdemes megvizsgálni S. Cabon et al. ([18]) valamint Peter W. Fuller et al. ([30]) munkáit. Előbbi képi és hang adatokat használ fel a NIDCAP skála szerinti klasszifikációhoz, míg utóbbi a viselkedési állapotokba történő kamera alapú klasszifikációt hasonlítja össze a poliszomnográfia alapú klasszifikációval.

Végezetül mivel az egyik tézisem érinti az automatikus annotáció problémakörét,

ezért szeretném röviden bemutatni, hogy milyen irodalmat találhatunk ezen a téren. Néhány publikáció magának az automatikus annotációnak a témakörében: [93], [95], [73]. Természetesen az automatikus annotáció témakörében nem kizárólag kamera alapú eljárásokkal foglalkoznak. A címkék sok esetben egy más típusú szenzor (pl. egy hőkamerából származó adatok) alapján vannak meghatározva a hagyományos kamerából származó képekhez. Mivel a későbbiekben általam bemutatott szintetikus címke generáló eljárás orvosi területen történő alkalmazás, ezért külön kiemelnék néhány olyan publikációt, ahol szintén orvosi területen alkalmazták a szintetikus adat generálást: [20], [110], [102], [109], [128]. A 6. fejezetben bemutatok majd egy automatikus címkéző eljárást, amely a címkéket a képek sorozatán lévő mozgások frekvenciája és kiterjedése alapján hozza létre. A periodicitás alapú ROI detekció maga természetesen nem újkeletű gondolat. Számos ezzel foglalkozó publikációt találhatunk az irodalomban: [12], [7], [129], [64].

A legújabb típusú generatív neurális hálózat alapú automatikus adat generálási és annotálási módszereket tekintve érdemes megemlíteni néhány GAN (Generative Adversarial Network) alapú eljárást ([113], [122] és [83]). Ezek esetében a generatív algoritmus képes megtanulni egy adott adathalmazba tartozó mintákat generálni, majd különböző klasszikus módszerek segítségével alakítható ki az annotáció ezen mintákra automatikusan. Pl. a [122] esetében hibás vasúti síneket tartalmazó képeket generáltak hibamentes képekből, majd ezek különbségéből hozzák létre az a "szemantikus szegmentáció feladat" típusú annotációt. Ezzel szemben Daniil et al. munkájában a generátor bizonyos "feature map"-jeit használja fel a generált kép összes pixelének klaszterezésére, véges számú klasztert alkalmazva.

3. Fejezet

Adatgyűjtés és annotáció

Ahhoz, hogy a csecsemők fiziológiai jeleit és viselkedési állapotait becslő érintésmentes algoritmusokat tervezhessünk és értékelhessünk, nagy mennyiségű annotált adatra van szükség. Mivel etikai és egyéb okok miatt nyilvános adatbázisok nem érhetők el a témában ezért saját adatbázist kellett kiépítenünk. Ehhez először is ki kellett építenünk egy adatgyűjtő rendszert, amely képes a képi adatok és a referencia fiziológiai adatok gyűjtésére egyaránt. Ráadásul egy olyan mobil rendszerre volt szükség, amelyet ki tudunk helyezni a Semmelweis Egyetem II. számú Szülészeti és Nőgyógyászati Klinikájának, Újszülött Intenzív osztályára. Ugyanis az Újszülött Intenzív osztállyal együttműködésben végeztük az algoritmusaink fejlesztését. Ők az orvosi hátteret biztosították, míg mi magát a fejlesztést végeztük.

Az első adatgyűjtő rendszer amit terveztünk, még nem egész napos vagy több napos felvételek készítésére lett tervezve, ezért csak nappali fényben történő felvételek készítésére kellett alkalmassá tenni. Ez a rendszer 5 részből állt:

- A csecsemő az inkubátorban (az ábrán kék)
- Basler kamera (az ábrán piros)
- Laptop vagy mini-Pc (az ábrán zöld)
- Philips IntelliVue MP20/MP50 monitor (az ábrán fekete)
- Adatbázis



3.1. Ábra: Egy látható fény tartományban működő Basler kamerával (piros) figyeli meg az inkubátort és a benne lévő csecsemőt (kék) és továbbítja az adatokat egy laptopnak vagy mini PC-nek (zöld). Eközben csatlakoztatva van a kórházakban használt Philips IntelliVue MP20/MP50 monitorra (fekete) amely pulseoxymeter-rel és EKG-val méri a pulzust és a légzést. Majd a mért adatokat feltöltjük egy adatbázisba.

Az említett adatgyűjtő rendszert és annak részeit a 3.1 ábra mutatja be.

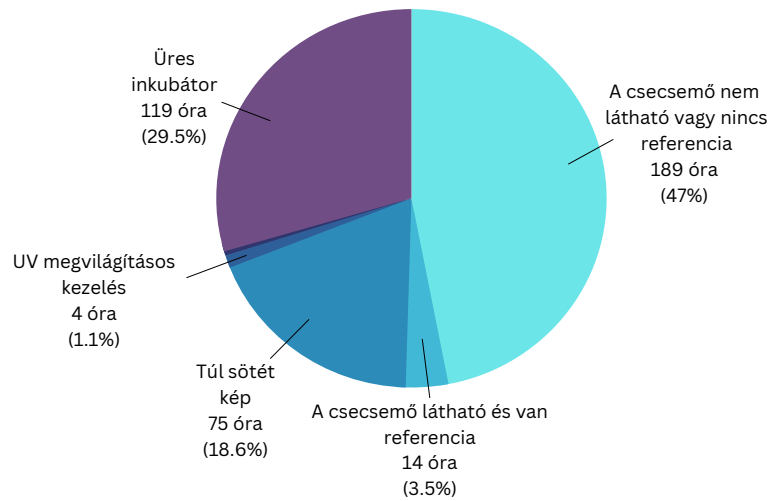
Mint a későbbiekben kifejtem, a munka során fejlesztettünk érintésmentes légzés figyelő algoritmusokat. Ennek az algoritmusnak a tanításához és értékeléséhez az említett adatgyűjtő rendszerrel a képi adatokon túl, hullámalakot és légzési rátát is gyűjtöttünk referencia adatként az impedancia alapú EKG szenzorokból, amelyek a Philips IntelliVue MP20/MP50 monitorra vannak csatlakoztatva. Erre a feladatra a fentebb leírt adatgyűjtő rendszer kielégítőnek bizonyult. A bemutatott rendszer kialakítása biztosította a hatékony és megbízható adatgyűjtést, lehetővé téve számunkra és az egészségügyi szakemberek számára a csecsemők légzésének és pulzusának hatékony nyomon követését. A rendszerrel összesen 700 órányi adatot gyűjtöttünk.

Természetesen az ehhez hasonló méréseknek szigorú etikai követelményeknek kell megfelelnie. Az ilyen klinikai vizsgálatok két nagy csoportra oszthatók: a klinikai intervenciós vizsgálatokra ("clinical trial") és a klinikai megfigyelésekre ("clinical observation"). Az első esetben a vizsgáló beavatkozik valamilyen betegség vagy ke-

zelés érdekében. Általában valamilyen új kezelést tesztel és ezért az ilyen vizsgálatok eltérnek a páciensek szokásos kezelési tervétől. Ez az első eset szigorúbb elbírálás alá kerül általában. A második esetben a résztvevők a szokásos kezelési tervük mellett maradnak és a vizsgálatot végzők csak megfigyelik, hogy az idő múlásával miként változik a betegek egészségi állapota. A mi méréseink a második csoportba tartoznak. Nem vizsgáltunk új kezelést csak megfigyeltük kamera alapú rendszerrel a betegek egészségi állapotát. Ilyen esetben a résztvevők beleegyezése szükséges. Mivel azonban a csecsemők nem tudnak még beleegyező nyilatkozatot kitölteni, ezért ezt szüleiknek kellett megtennie. Ezt minden résztvevőnk esetében meg is tették.

Az adatgyűjtés nem volt egyszerű. Mivel kamera alapú adatgyűjtő rendszerrel van szó, több probléma is felmerülhet. A kamera látómezejét időnként elhagyták a csecsemők, azaz kivették őket az inkubátorból, továbbá megváltoztak a fényviszonyok, vagy egyéb tényezők zavarták a mérést. A légzési referenciákhoz EKG-re volt szükség, ami sokszor nem volt felhelyezve. Meg kellett kérni egy nővért, hogy tartsa fel az elektródákat a csecsemőre. A mérések kezdetén még kamera állvánnyal kellett dolgoznunk, amelyet a kórházi személyzetnek néha el kellett mozdítania munkájuk elvégzéséhez, melynek következtében a kamera látóteréből kikerült az inkubátor. Valamint ezeken felül is számos zavaró tényezővel kellett számolnunk a mérések során. Az ilyen zavaró szituációk meglepően gyakran fordultak elő. A mérések kezdeti szakaszában, amikor már 400 órányi adattal rendelkezünk viszonylag sok adat bizonyult problémásnak vagy használhatatlannak (lásd a 3.2 ábrát). Mint a későbbiekben látni fogjuk az ehhez hasonló problémák megoldására kifejlesztettünk egy AI alapú eljárást, amely felismeri ezeket a zavaró helyzeteket. Ezt "szituációanalízisnek" neveztük. Kezdetben azonban kategorizálnunk kellett a méréseinket, hogy ne használjuk az összes adatot algoritmusok tanítására, hiszen nyilvánvaló, hogy pl. olyan esetben amikor a csecsemőt kivették az inkubátorból nincs értelme légzésfigyelő algoritmust tanítanunk az üres inkubátorról készült képeket használva.

Később szükségessé vált, hogy az adathalmazt kiegészítsük infra megvilágítású éjszakai felvételekkel és egész napos 24 és 48 órás mérésekkel. Erre azért volt szükség mert, amint a későbbi fejezetekben bemutatom, foglalkoztunk a csecsemő aktivitási és alvási állapotainak becslésével. Erre is terveztünk mesterséges intelligencia alapú



3.2. Ábra: Az itt látható kördiagramm mutatja, hogy a felvett összes órák mely részei voltak használhatóak és mely részein voltak zavaró körülmények, amikor 400 órával rendelkezünk.

algoritmust, amelyet tanítani és értékelni kellett. Ehhez pedig annotált adatokra volt szükség. A 24 órás felvételek készítéséhez terveztünk egy az inkubátor állványához rögzíthető 3D nyomtatott adatgyűjtő rendszert infra megvilágító eszközzel, amely éjszaka bekapcsolt. Valamint egy zárt inkubátorra szerelhető 3D nyomtatott adatgyűjtő eszközt (lásd a 3.3. ábrán), amely szintén rendelkezett infra megvilágítóval, kamerával és a zárt inkubátor tetejéhez rögzítő tapadókorongokkal.

A fentebb leírt adatgyűjtő rendszerekkel képesek voltunk megfelelő mennyiségű és minőségű adatokat gyűjteni azonban a nyers adatok még nem voltak alkalmasak arra, hogy egyből az algoritmusaink tanítására és értékelésére használjuk őket. Előbb előkategorizáláson és annotáción kellett átesniük. A fentebb említett problémás szituációkat is be kellett jelölnünk, ahogyan azokat az időperiódusokat is, amikor a kamerakép megfelelő volt és rendelkezünk légzési és pulzus referenciával. A pulzusról azért esik szó mert a laborban érintésmentes pulzusfigyeléssel is foglalkoztunk. Én azonban ezen algoritmus fejlesztésében keveset segítettem. A kutatás során és ebben a munkában is inkább a légzés, a mozgások és viselkedési fázisok becslésére koncentráltam. Az adatgyűjtőnket azonban úgy terveztük meg, hogy képes legyen, mind légzés, mind pedig pulzus referencia gyűjtésére. A kategorizáláshoz és annotáláshoz saját fejlesztésű QT grafikus felhatalmzott felhasználói környezet és C++ alapú

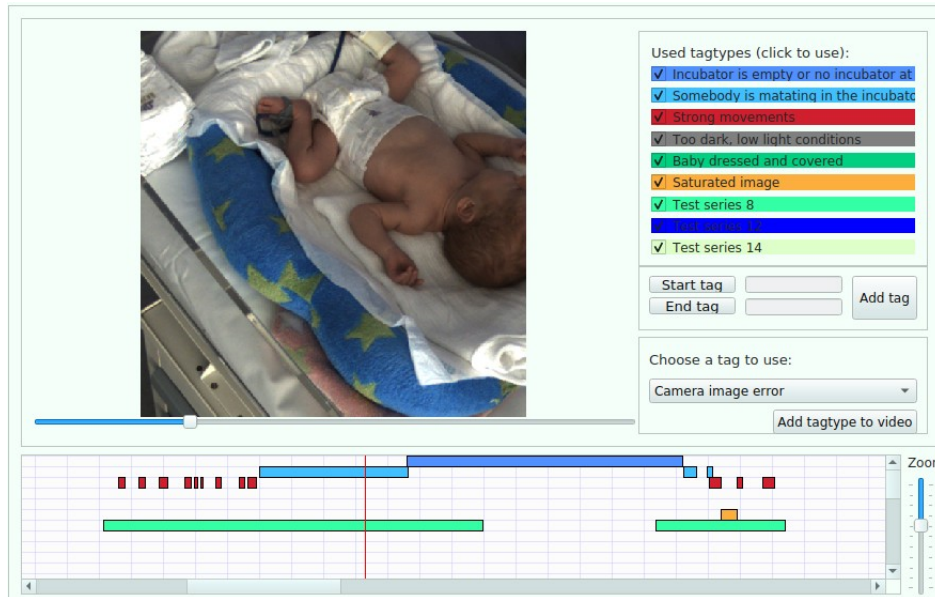


3.3. Ábra: A 3D nyomtatott burkolattal rendelkező adatgyűjtő rendszer, amely tartalmaz egy kamerát, egy infravörös megvilágítót és az ezek vezérlésére alkalmas Raspberry Pi-t. A bemutatott rendszer alkalmas volt akár egész napos felvételek elkészítésére is.

adatkezelő és annotáló szoftvereket hoztunk létre. Ez egy adatbázisból töltötte le az adatokat, hogy megjelenítse a képi adatot, a referencia légzési és pulzus jeleket, valamint a kategorizáló annotációkat. A szoftvert úgy kell elképzelni, hogy az adott t időpillanathoz tartozó képet látjuk és a hozzá tartozó referenciákat és kategóriákat egy adott időszávon jeleníti meg $(t - w)$ és $(t + w)$ között. Minden egyes kategóriának külön szint tudtunk beállítani. Vagyis a kategóriák sávjában az adott színű téglalapok jelezték, hogy az adott időperiódus milyen kategóriába esik. Ennek illusztrációját lásd a 3.4. ábrán.

3.0.1. Adatgyűjtő rendszer viselkedési állapotok klasszifikációjához

A 7. fejezetben bemutatok egy olyan eljárást, amelynek segítségével klasszifikálni tudjuk a megfigyelt csecsemők aktivitási és alvási állapotait, másképpen viselkedési állapotait. A bemutatott eljárás tanításához és értékeléséhez szintén szükségünk lehet annotált adathalmazra. Mint azt a későbbiekben bemutatom az eljárás bemeneteit képi adatokból és az IntelliVue MP20/MP50 monitorból kinyerhető életjelek-

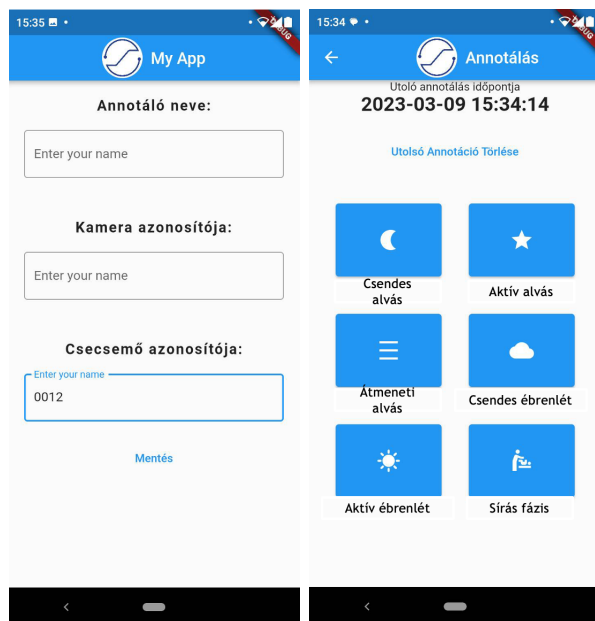


3.4. Ábra: Ezen az ábrán a használt annotáló és kategorizáló szoftvert láthatjuk működés közben. A bal felső sarokban az aktuális időpillanathoz tartozó kép látható. Alatta láthatók a $(t - w)$ és $(t + w)$ közötti tartományban bejelölt kategóriák színes téglalapokkal jelölve. Jobb oldalt látható a vezérlés.

ből állítjuk elő. Ez azt jelenti, hogy a bemenetek felvételére a fentebb bemutatott adatgyűjtő rendszer is tökéletesen alkalmas.

Annotált adathalmaz előállításához azonban szükségünk van az elvárt kimenetekre is. Ezek felvételére azonban új eljárást kellett kidolgozni. Mivel az elvárt kimenetek a viselkedési állapotok klasszifikációjának esetében a NIDCAP skála szerinti fázisok, amelyeket a kórházi személyzet tud megállapítani, ezért olyan módszert kellett kialakítanunk, amellyel képesek lehetünk (a bemeneti adatokkal idő szinkronizáltan) rögzíteni az orvosok döntését. Erre létrehoztunk egy mobilapplikációt, amellyel az orvosi személyzet képes volt manuálisan bevinni a NIDCAP skála szerinti osztályozásukat az egyes csecsemőkhöz. Az applikáció első lapján az annotáló megadhatja a saját nevét, a megfigyelt csecsemő és az arra használt kamera azonosítóját. Aztán a következő lapon 6 gomb van elhelyezve, amellyel az annotáló személy be tudja vinni a 6 NIDCAP skála szerinti állapot egyikét.

Ezzel a rendszerrel 10 babáról gyűjtöttünk (egy vagy több napnyi) felvételt a PIC klinikán referenciával szinkronizálva, amelyeket a 7. fejezetben leírt viselkedési állapot osztályozó betanítására használtunk fel. Vagyis pontosabban az ezekből a felvételekből kinyert különböző videóalapú jellemzőket használjuk bemenetként és



3.5. Ábra: Az első és második lapja annak a mobiltelefonos applikációnak, amelynek segítségével az orvosi személyzet képes manuálisan bevinni és menteni a különböző időperiódusokhoz tartozó NIDCAP skála szerinti állapotokat.

a NIDCAP szerinti állapotokat elvárt kimenetként. Az adatgyűjtésben résztvevő csecsemők adatait pedig a 3.1 táblázat foglalja össze. Az adatokat természetesen anonimizált módon tároltuk.

3.1. Táblázat: Az adatgyűjtésünkben részt vevő csecsemők adatai, akikről felvételt készítettünk.

Alany	1	2	3	4	5	6	7	8	9	10
Felvétel ideje (órák)	96.7	5.5	39.4	27.4	51	105.5	50.1	36.4	56	38.2
Nem	F	M	M	F	F	M	F	F	F	M
Terhességi kor (hetek)	32	32+3	31+4	35+4	39	32	33	38+6	24+2	33+4
Születési súly (g)	2020	1840	1850	1870	3150	2120	2080	2840	760	2100
Szülés utáni kor (napok)	4	4	10	8	4	7	2	7	11	1
Valós súly (g)	1900	1850	1680	1820	2905	2040	1960	3150	750	-
Hossz (cm)	46	44	-	45	57	45	44	48	46	45
Fejkörfogat (cm)	32	29.5	-	32	34	30	32	33	22	-
Légzéstámogatás	no	no	no	no	no	no	no	yes	yes	no
Gyógyszeres kezelés	no	no	no	no	yes	no	no	yes	yes	yes
Fitzpatrick skála	2	3	2	2	2	2	2	2	2	2

4. Fejezet

Mozgás alapú mérések orvosi területen és általánosságban

Mielőtt elmélyülnénk a konkrét érintésmentes monitorozó algoritmusok világában, szeretném megvilágítani általánosságban, hogy mik a mozgás alapú mérések az orvosi területen történő alkalmazásokra koncentrálva. Mit lehet mozgás alapon mérni? Hogyan lehet számszerűsíteni a mozgást? Milyen módszereket találhatunk az irodalomban? Hogyan közelítik meg más alkalmazások esetében ezt a területet?

A disszertációm témájának megfelelően abból indulok ki, hogy a mérésekhez szenzorként kamerát alkalmazunk és érintésmentesen monitorozunk. A mozgás alapú mérés alatt azt értem, hogy a képen látható mozgásokat és a mozgásból adódó pixel intenzitás változásokat használjuk fel, valamilyen a mozgással összefüggő érték mérésére. Logikailag ez elkülönítendő attól az esettől, amikor nemcsak a mozgásból adódó pixelintenzitásváltozást, hanem a színváltozást is figyelembe vesszük. Mint pl. az érintésmentes pulzus figyelő algoritmus esetén, amely a bőr apró színváltozásait megfigyelve becsüli a képen lévő személy pulzusát. Fejlesztői szempontból ez azt jelenti, hogy elég a szürkeárnyaltos képet használni a méréshez és nem szükséges a színcsatornákat figyelembe venni. Természetesen ez nem mindig teljesül ilyen szigorúan. Egyes mozgásokat megfigyelő algoritmusok (főként a neurális hálózat alapúak pl. a "póz becslők") figyelembe veszik a színcsatornákat is. Valamint a ROI ("Region of Interest") detektorok is használják a szín információkat az alkalmazások szempontjából fontos régiók detektálásakor. Az olyan pixel elmozdulási irányokat

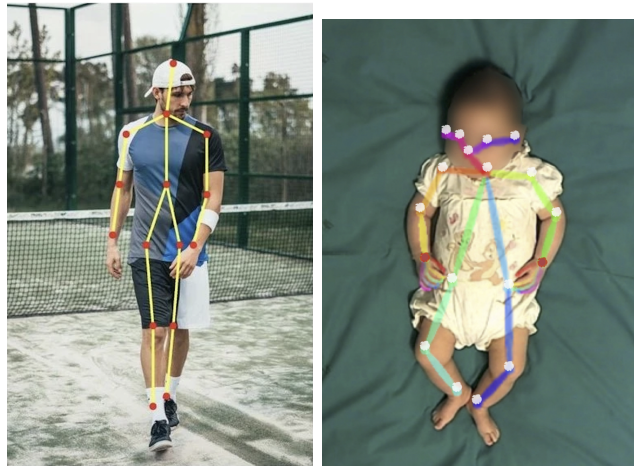
és sebességeket számszerűsítő, avagy mozgásbecslő algoritmusok, mint amilyenek az "optikai áramlás", vagy a "téridő alapú rekurzív keresés" (spatiotemporal-based recursive search) algoritmusok azonban csak a szürkeárnyaltos képeket használják fel.

Egy másik fontos aspektusa a mozgás alapú mérésnek, hogy mindig pontosan érteni kell, hogy mi az, amit mérünk. A fiziológiai jelek kameraalapú mérése besorolható a pletizmográfia fogalmkörébe, amely a görög "plethysmos" szóból származik (jelentése: növekszik, megtelt). A pletizmográf egy olyan eszköz, amely egy szerv vagy test térfogatváltozását méri, amely általában a vér vagy a levegő mennyiség változásának köszönhetően alakul ki az adott szervben. A fotopletizmográfia pl. az erekben történő vér mennyiség ingadozása következtében kialakuló térfogatváltozást méri meg optikai szenzorok segítségével és becsüli meg a pulzust. A légzés esetén a beáramló és kiáramló levegő mennyiségváltozásának köszönhető térfogatváltozást mérjük, ami egy adott frekvenciatartományban történő térfogatváltozás. Ebben az esetben a pletizmográf szenzora lehet egy kamera, amely a felvett képkockákon detektálja a térfogatváltozást és ennek alapján becsüli meg a pletizmográf a légzési rátát, vagyis a légzés frekvenciáját LPP-ben (légzés per perc). Ez esetben a kamerával egy mozgás alapú mérést végzünk a légzés megfigyelésére. Hasonló módon pletizmográfiaival megmérhetjük akár a pulzust is de a fotoplethysmográfia esetében a szín is fontos szerepet játszik a mérésben nem csak az intenzitás. Mivel azonban itt az erek térfogatváltozását mérjük, ezért szintúgy pletizmográfiaőról van szó.

Azonban a mozgás alapú mérések a pletizmográfianál szélesebb területre is kiterjednek, még ha az orvosi alkalmazásoknál maradunk is. Nem csak szervek térfogatváltozásából adódó periodikus mozgást tudjuk mérni, hanem pl. a megfigyelt alany végtagmozgásait is. A végtagmozgások kvantifikálásának problémáját oldják meg a 2D és 3D póz becslők.

A 2D pózbecslők a képkockákon kulcspontokat detektálnak és követnek. Ezek a pontok a végtagmozgások követéséhez kulcsfontosságú "ízületi" pontok, ahol az emberi test nagyobb szögű elhajlásra képes (térd, könyök, váll, csípő, nyak) valamint jól felismerhető és követhető végpontok (lábfejek, kézfejek, fej, has, stb). A legújabb és legsikeresebb változatoknál (pl. OpenPose) finomabb felbontást igénylő

kulcspontok detektálására is van lehetőség, (mint pl. szemek, orr, száj stb.). A póz becslést sport és fitness alkalmazásokban is előszeretettel használják, mint pl. a yoga applikációk. De orvosi területeken és biztonsági rendszerek esetén is gyakran alkalmazzák. Nálam a használata a csecsemők végtagmozgásainak elemzésénél merült fel. A végtagok mozgásainak ilyen pontosságú követése azonban nem volt szükséges a mi céljainkhoz, valamint a problémára már voltak az irodalomban fellelhető megoldások ([70]).



4.1. Ábra: Ezen a képen OpenPose nevű pózbecslő eljárás eredményét láthatjuk. Baloldalt egy felnőltre adott eredmény. Jobboldalt pedig csecsemőre adott eredmény látható.

Az eddigiekben csak olyan esetekről írtam, ahol kamerán lévő emberek mozgásait, szerveik térfogatváltozásait és végtagjaik mozgásait figyeltük meg. A mozgás alapú algoritmusok azonban még ennél is szélesebb területen alkalmazhatóak. Nemcsak a képen lévő alany testének mozgását vagyunk képesek mérni, hanem bármely mozgó objektumot képesek vagyunk megfigyelni. Egészségügyi területen megfigyelhetjük például a műtétet végző orvos szikéjének mozgását vagy a mentőautók mozgását a kórház területén. Egyéb területeken pedig rendkívül nagy mennyiségű alkalmazás nyílik meg a közlekedésben, vagy akár az úrkutatásban. Ilyen alkalmazások lehetnek az autópályákat megfigyelő kamerák képeinek értékelése vagy az önjáró autók kamerás rendszereinek működtetése. Mindezekre kitérni ebben az értekezésben nincs lehetőség. Azonban egy dolog közös bennük. Minden mozgást értelmező alkalmazás alapvetése, hogy valahogy számszerűsíteniük kell a mozgást.

Ehhez alapvetően abból kell kiindulnunk, hogy fejlesztői szemszögből a képek

mátrixoknak felelnek meg amelyek intenzitás értékeket tartalmaznak. Ha színes képről beszélünk, akkor 3 csatornán (pl. RGB) találhatunk intenzitás értékeket minden egyes pixelhez. (Természetesen léteznek olyan színes kép kódolások is, ahol 3-nál több csatornán tároljuk egy pixel színét. Pl. az RGBA, ahol az A alpha csatorna azt jelzi, hogy mennyire átlátszatlanok az egyes képpontok.) A szürkeárnyaltos kép azonban (amit a legtöbb mozgásbecslő algoritmus alkalmaz) ellenben egycsatornás 2D mátrix. Minden képkockája tehát egy videófelvételnek egy 2D mátrix, amely a képek típusának és bitmélységének megfelelő méretű számokat tartalmazza. A leggyakrabban használt típus a uint8-as pl., amelynél 8 bites egész típusú számok találhatóak a mátrixban. Miután a képkockáinkat egyenként kvantifikáltuk és mátrixokként értelmezzük, lehetőségünk van a mozgás becslésére két egymást követő képkocka között. Erre sok féle lehetőségünk van:

- Differenciakép (az egymást követő képek különbsége) kiszámítása
- Blokk Illesztő (Block Matching) algoritmusok kiszámítása
- Sűrű Optikai Áramlás (Dense-Optical-Flow) algoritmusok kiszámítása
- Mély Folyam (Deep-Flow) algoritmus kiszámítása
- Tér-idő Alapú Rekurzív Keresés (Spatiotemporal-based Recursive Search) kiszámítása

Bármely eljárást is alkalmazzuk, a végeredmény egy mozgási kép lesz, amelyen a pixelek értékei a mozgások intenzitását vagy irányát kódolják. Bizonyos esetekben, (mint amilyen differenciakép vagy a háttértől való eltérés kiszámítása) a kimenet egy csatornás és csak a kép különböző részein fellelhető mozgások intenzitásáról hordoz információt. Míg más esetekben, (mint amilyen a Sűrű Optikai Áramlás) a kimenet 2 csatornás, ahol az első csatorna a vertikális, míg a második csatorna a horizontális mozgások intenzitásának felel meg az egyes pixeleknél.

A mozgások számszerűsítése után (sok alkalmazás esetében) a képfeldolgozásból jól ismert morfológiai eljárások kerülnek alkalmazásra (mint amilyenek a nyitás, a zárás, dilatació vagy az erózió). Ezek célja, hogy javítsák a jel-zaj viszonyt.

Ha az adott problémánál valamilyen ROI detektor kerül alkalmazásra akkor a feldolgozandó mozgási képeken tudjuk hol kell keresnünk a szükséges információt. A légzési ráta becslő eljárás esetén pl. felnőtteknél a mellkas körüli mozgások kötődnek a légzéshez, míg a csecsemők esetében a has körül található meg ugyanez az információ.

A megfelelő helyen keresve tehát mozgás alapú algoritmusokkal mérhetőek olyan mutatók, mint a légzés, szemhéjzárttság, mozgási aktivitás vagy ultrahang felvételeket elemezve a szívverés.

5. Fejezet

Érintésmentes légzésfigyelő algoritmusok: előzmények és a javasolt algoritmus

5.1. Csecsemők légzésfigyelésének relevanciája

Az újszülöttek légzésszámának ellenőrzése kulcsfontosságú az újszülött intenzív osztályokon (NICU - Neonatal Intensive Care Unit) és érintésmentes mérése erősen kutatott téma napjainkban [74], [52], [106], [Au1], [C1], [J1]. Monitorozásának fontos szerepe van az olyan feladatokban, mint például a légzési rendellenességek korai felismerése az inkubátorban lévő csecsemőknél. Az olyan rendellenességek, mint "obstruktív alvási apnoe" (Obstructive Sleep Apnea - OSA) viszonylag gyakoriak koraszülött csecsemők esetében. Világszerte a csecsemők 40%-ánál az orr elzáródása szájlégzésre való átálláshoz vezet, ami OSA-t eredményez [53]. Továbbá az OSA tünete lehet számos egyéb betegségnek, pl. okozhat növekedési nehézségeket, viselkedési zavarokat, vagy akár csecsemő halálhoz is vezethet [54]. Gyakorlatilag minden szervrendszerre negatív hatása lehet [26]. Sőt egyes kutatások például már kimutatták a felnőttkori reflux és az OSA közötti kapcsolatot is [25]. Ha az OSA-t diagnosztizálták akkor a "folyamatos pozitív légúti nyomás (continuous positive airway pressure - CPAP)" alkalmazásával kezelik. Fontos azonban, hogy minél előbb fedezzük fel az OSA-t és minél előbb kezdjük meg a kezelést.

A légzés mozgás alapon, kamera segítségével viszonylag könnyen mérhető fiziológiai jel (lásd [97,107,108]), amely fontos információkat nyújthat a kórházi személyzet számára nemcsak a csecsemők légzőrendszerének megfelelő fejlődéséről, de az olyan származtatott mutatókról is mint amilyen az alvás vagy a csecsemők érzelmi állapota. Mégis ezt a fiziológiai jelet viszonylag ritkán mérik rutinszerűen a kórházakban (eltekintve a nagyon indokolt esetektől), mivel a légzést tradicionálisan EKG-val szokás mérni a koraszülött klinikákon.

Az EKG a légzés mérés során az elektródái között a légzési mozgás következtében kialakuló impedancia változást méri [84]. Ez a módszer azonban azt igényli, hogy az elektródák közvetlenül érintkezzenek a csecsemők bőrfelületével, ami zavarhatja a csecsemőt, a szülőket és a kórházi személyzetet is, akiket a sok vezeték korlátozhatja a munkájuk elvégzésében. Sőt a koraszülött csecsemők vékony bőrét az elektródák akár fel is sérthetik. Bár a mai speciálisan babákra fejlesztett elektródák kíméletesebbek, ezekkel is vannak problémák, ugyanis egyszer használatosak és felhelyezésük szakértelmet igényel.

Amikor a csecsemők légzésének az érintésmentes monitorozásáról beszélünk akkor az ennek megfigyelésére alkalmazott szenzor a hagyományos színes kamera. A későbbiekben minden esetben ezt értjük alatta. Ezt a szenzort azért választottuk ki, mert az általunk végzett irodalomkutatásból azt a következtetést vontuk le, hogy az általunk megoldani kívánt érintésmentes légzésfigyelési probléma megoldására ez a legalkalmasabb tekintve, hogy mennyire vagyunk képesek vele adatot rögzíteni kórházi körülmények között egy inkubátorban lévő csecsemőkről. Vannak egyéb lehetőségek, mint pl. radar rendszerek alkalmazása (többek között) a légzési ráta (LR) becsléshez [72, 90], amely a Doppler-hatás alapján működik vagy a hő-képalkotás használata az LR becslésére az orr körüli hőmérséklet-változásokból ([87]).

Azonban a radar alapú módszerek is (sok más módszerhez hasonlóan) érzékenyek az intenzív végtagmozgásra és a saját jeleire, amelyek a környezetről verődnek vissza és amelyek interferenciát okozhatnak a radarjelben, míg a hagyományos RGB kamera esetében a végtagmozgások kezelésére lehetnek megoldások és egyéb a képfeldolgozás terén használt algoritmusok (pl. ROI detektorok) is könnyen lefuttathatók a kamera által szolgáltatott képekre. A hő-képalkotás alapú módszerek alkalmazása

(fűtött inkubátorban) pedig milikelvin-es tartományba eső pontosságot igényel, ami nagyon költségessé teszi ennek a módszernek az inkubátorokban történő alkalmazását. A költséghatékonysági és praktikus szempontok alapján így a hagyományos színes kamera alkalmazása mellett döntöttünk.

Ahogy korábban már említettem az első olyan mozgásalapon érintésmentesen mérhető életjel, amelynek monitorozásával tüzetesebben foglalkoztam a légzés volt. Ennek a fiziológiai jelnek a kamera alapú mérésére több lehetőséget is körbejártam a hagyományos nem adat-vezérelt algoritmusoktól kezdve az "végponttól végpontig (end-to-end)" neurális hálózatokig. Végül egy olyan megoldás mellett döntöttem, amely tartalmaz egy mesterséges intelligencia alapú modult, de alapvetően klasszikus gépi látási algoritmusokon alapszik.

Ebben a fejezetben ismertetem a légzés érintésmentes monitorozásának megoldására történt kutatásaimat. Azt, hogy mely algoritmusokkal próbálkoztam és végül melyik mellett döntöttem és miért.

Elöljáróban hangsúlyoznám, hogy a csecsemők érintésmentes légzés monitorozására nem alkalmazhatóak módosítás nélkül egy az egyben a felnőttek számára kifejlesztett módszerek ahogyan ezt Chetverikov et al. is leírta ([31]). Ugyanis a felnőttek és a koraszülött csecsemők légzési mintája és rátája között különbségek vannak. A csecsemők légzési jelei kevésbé szabályosak, és az LR-jük jelentősen magasabb, jellemzően 30-40 légzés-per-perc (LPP) közötti tartományban helyezkedik el az egészséges csecsemők esetében.

5.1.1. Kapcsolódó szakirodalom

Ebben az alfejezetben bemutatok néhány olyan említésre méltó publikációt, amelyek a 2020 és 2022 közötti időszakból származnak és kifejezetten csecsemők érintésmentes légzés monitorozásával foglalkoznak ([96], [57], [55]).

Rossol et al. munkájában ([96]) egy tanulmányt végeztek, amelynek célja egy videóalapú légzéssel kapcsolatos monitorrendszer kifejlesztése volt a koraszülött intenzív osztályok (NICU) számára, amely nem igényel érintkező szenzorokat és javíthatja a betegellátást. Az algoritmus mikromozgások és állapotállandóság észlelése révén volt képes a légzési mozgásokat felismerni és rögzíteni a stabil állapotú, 33-40

hetes korrekciós gesztációs korig eljutott NICU betegek videofelvételeiből. Az eredmények jelentős korrelációt mutattak az algoritmus és az aktuális kórházi standard, az elektrokardiográfiai impedancia mérés között. A videóalapú rendszer 6.36 LPP négyzetes középhibát mutatott a folyamatos monitorozással összehasonlítva.

Kyrollos at al. pedig arról számol be munkájában ([57]), hogy hogyan fejlesztettek ki és hasonlítottak össze különböző technikákat az érintésmentes légzésszám becslésére, neonatológiai betegek videofelvételeinek felhasználásával. Az adatokat egy fentről rögzített szín- és mélységkamera (RGB-D) segítségével gyűjtötték, miközben a kórház betegmonitorjából rögzítették a szinkronizált fiziológiai adatokat. Egy mélytanulási algoritmust fejlesztettek az újszülött arcának és mellkasának automatikus detektálására. Ezután a légzési jelet a mellkas mozgásának kvantifikálásával hozták létre. Ezzel az algoritmussal Kyrollos at al. kevesebb mint 3.5 LPP-s átlagos abszolút hibát ért el.

A harmadik cikkben ([55]) bemutatnak egy olyan módszert, amely távolról monitorozza az újszülöttek szívritmusát és légzésszámát az érintésmentes érzékelés előnyeinek kihasználásával. A kutatás során egyetlen központban vizsgáltak hét újszülöttet a koraszülött intenzív osztályon (NICU), és egy digitális kamerát használtak a mérésre. A ROI automatikus kiválasztására egy konvolúciós neurális hálózatot alkalmaztak, majd a jellemzőkül szolgáló jelek előállításához különböző felbontási módszereket és spektrális analízist használtak. Az eredményeket az EKG monitor adataival validálták, és erős korrelációt tapasztaltak az algoritmusból kapott adatok és a referencia adatok között (MAE - Mean Absolute Error: 2.13 LPP).

5.1.2. Adatgyűjtés

Az adatgyűjtés a légzés monitorozás esetében egy olyan általunk tervezett adatgyűjtő eszközzel történt, amely tartalmaz egy színes kamerát, egy merev lemezt, egy infravörös megvilágítót és képes vezetékkel kapcsolódni a Philips IntelliVue MP20/MP50-es orvosi monitorhoz, ezáltal képes a képi adatok mellett a szinkronizált EKG légzés adatokat is menteni, amelyek referencia jelként szolgálnak. Az adatgyűjtő rendszernek 3D nyomtatott burkolata van és tapadókorongokkal kapcsolódik a zárt inkubátor tetejéhez. Az orvosi monitor a hullám alakokon kívül légzési

rátákat is küld az adatgyűjtőknek. Mentésre kerülnek tehát a bemenetként használt képsorozatok és az elvárt kimenetként használt légzési ráták. Ezek után saját szoftverekkel kategorizáltuk a felvett adathalmazt hogy mely részei használhatóak fel a tanításhoz. Az adatgyűjtő rendszer részletesebben is leírom a 3. fejezetben.

5.1.3. Előzmények

Az érintésmentes légzésfigyelés terén végzett munkám közvetlen előzményének az a publikáció tekinthető [31], amelyben Dmitry Chetverikov egy rövid áttekintést ad a felnőttek és csecsemők érintésmentes LR-mérésével kapcsolatos munkákról és egy új módszert is javasol.

A fentebb említett publikáció ([31]) közvetlen folytatásaként készített munkában [Au1] két új megközelítést alkalmaztunk a csecsemők légzésének érintésmentes monitorozásához:

- 1) Ritka optikai áramlás alapú módszer
- 2) Szinguláris érték dekompozíció alapú módszer

Mindkét bemutatott módszer azon a naiv feltételezésen alapszik, hogy az egész képet elfoglalja a csecsemő és az egymást követő képeken és többségében nyugodt légzés figyelhető meg, így nincs szükség ROI detekcióra és a képen látható bármely olyan mozgás, amelynek amplitúdója és frekvenciája a légzési tartományba esik légző mozgásnak tekinthető.

1) Ritka optikai áramlás alapú légzési jel (OF - "Optical Flow")

Az első bemutatott eljárás a ritka optikai áramláson alapuló módszer az OpenCV [14] által biztosított algoritmusokat használja, mely esetében a színes videó szürkeárnyaltos változatát használjuk bemenetként. Az optikai áramlás 2D elmozdulási vektorok halmaza, amelyek az aktuális (t időpillanathoz tartozó) és az előző ($t-1$ időpillanathoz tartozó) képkockák közötti mozgást mutatják. Az optikai áramlásnak két fő típusa ismert: a "sűrű" változat feltételezi, hogy az áramlás kiszámítható az összes pixelre és nincsenek követendő pontok a képkockasorozatokban, míg a "ritka" változat kiválasztott pixelekre számítja ki az optikai áramlást. A hivatkozott

tanulmányban ([Au1]) a ritka optikai áramlást alkalmaztuk, ahol jellemző pontokat követünk és a jellemző kiválasztás két különböző módját is kipróbáltuk:

- a) Sarokszerű jellemzőpontot kiválasztása, a "Good Features to Track" című munkában javasoltak szerint [101].
- b) A lehetséges terület nagyobb részét lefedő, szabályos elhelyezkedésű rácson lévő pontok halmazának követése. (Lásd a 5.1. ábrát.)

A kiválasztott helyeken az elmozdulás vektorok (u, v) kiszámítása után megkapjuk az átlagos elmozdulás vektort, és ezután annak hosszát (amit a 5.1. egyenlet alapján számolunk) az aktuális képkockához tartozó jelértékként használjuk. Így minden képkockához kapunk egy numerikus jelértéket. Az így kapott egydimenziós mozgási jelet a 5.2. ábra mutatja be. A jelet kettős csúcsok sorozata jellemzi, amelyek mindegyike egyetlen légzésnek felel meg. (A levegő be és kiáramlásának megfelelő események adják a két csúcsot egy légzési ciklus esetén). A légzési események közötti mély minimumok a lélegzetvételek közötti szünetek.

$$||v|| = \sqrt{u^2 + v^2}; u = \frac{dx}{dt}; v = \frac{dy}{dt} \quad (5.1)$$

ahol dx és dy a pixelelmozdulások horizontális és vertikális összetevői. Ugyanis:

$$\mathbf{I}(x, y, t) = \mathbf{I}(x + dx, y + dy, t + dt) \quad (5.2)$$

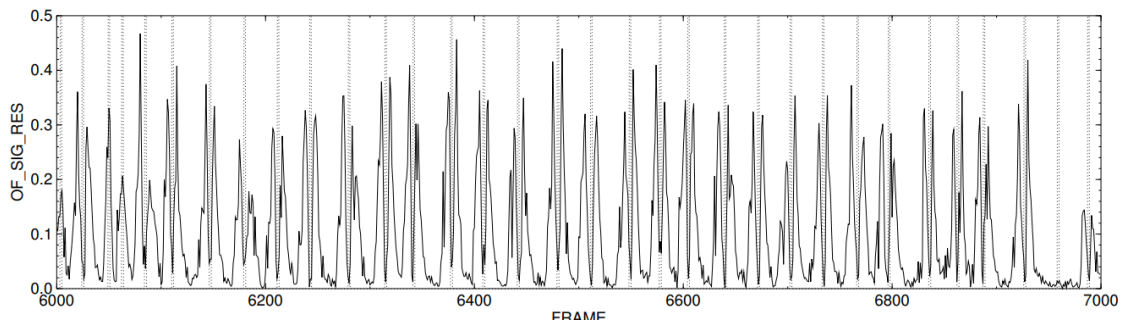
írja le egy pixel intenzitásának megváltozását az dt idő elteltével. Ahol $\mathbf{I}(x, y, t)$ az (x, y) helyen lévő pixel intenzitása t időpillanatban. Ennek alapján az optikai áramlás feladata $v = \frac{dy}{dt}$ és $u = \frac{dx}{dt}$ elmozdulások kiszámítása.

2) Szinguláris érték dekompozíció alapú légzési jel (SVD - "Singular Value Decomposition")

Második módszerünk a légzési mozgások közötti szünetek észlelésére irányul, nem pedig maguknak a légzéseknek a detektálására. A módszer a szinguláris dekompozíciót (SVD) alkalmazza a képek sorozatára. A szünetek jobban megkülönböztethetők



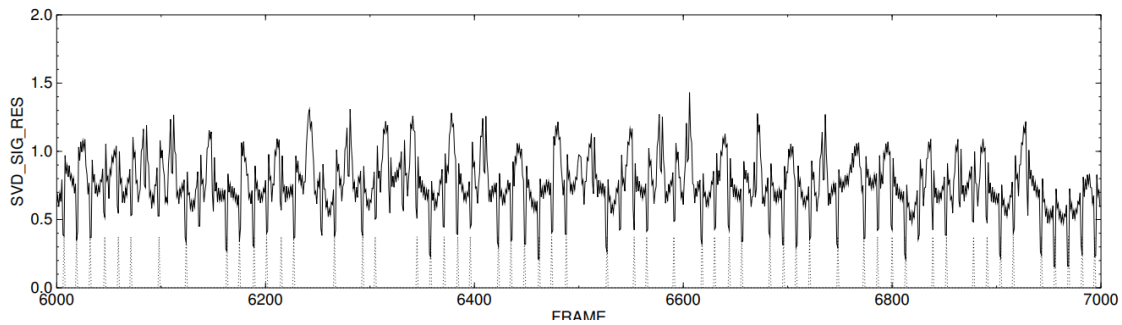
5.1. Ábra: Egy példa a ritka optikai áramlás esetén követett pontok elhelyezkedésére. A követett képpontok egy szabályos rácson helyezkednek el. A pontokból kiinduló vonalak mutatják a követett pontok mozgásának irányát.



5.2. Ábra: Ezen az ábrán az optikai áramlás alapú eljárás eredményeként kapott légzési jelet láthatjuk. Kettős csúcsok jelölnék egy légzést. Az első csúcs a belégzésnek a második pedig a kilégzésnek felel meg. A grafikon egy tesztvideó 1000 képkockájára terjed ki (40 másodpercnyi felvétel 25 fps sebességgel).

és könnyebben észlelhetők (az ezen módszer alkalmazásával kapott) légzési jelekben, ha az SVD korábban kifejlesztett gyors inkrementális (futó) megvalósítását alkalmazzuk [22]. Ez a megvalósítás elkerüli, hogy az SVD minden alkalommal újra ki legyen számítva, miközben az időbeli ablak előrehalad. Ehelyett úgy módosítja az előző eredményt, hogy azt korrigálja a belépő képkockákkal miután törölte a csúszó ablakból a kilépőket.

Az "inkrementális" SVD minden 25 egymást követő képkockára lefut. A 25 kép egy nagy mátrixba kerül, amelynek a legnagyobb értékű szinguláris vektora adja a (statikus) háttér. Minden képkockánál a légzés jelérték a képkocka átlagos abszolút eltérése a háttértől. Az így kapott egydimenziós légzési jelet a 5.3. ábra mutatja be.



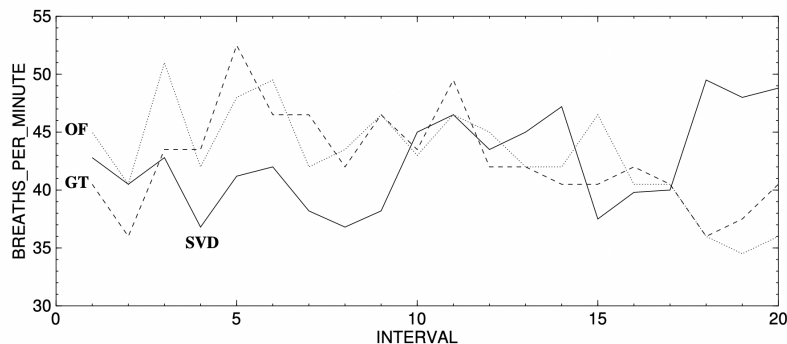
5.3. Ábra: Ezen az ábrán a "szinguláris érték dekompozíció" alapú eljárás eredményeként kapott légzési jelet láthatjuk. A módszer nem a légzési jel csúcsait, hanem a légzések közötti szüneteket detektálja, ugyanis ezen a típusú jelalakon ezeket könnyebb detektálni. Az ábrán a pontozott vonalak ezeket a szüneteket (amelyek lokális minimumok) mutatják.

A két légzésfigyelő módszer értékelése

A fentebb leírt két módszer légzési hullámalakokat von ki a képek sorozatából. A légzési ráták meghatározásához azonban légzés számlálásra is szükség van. Az általunk alkalmazott légzés számláló módszer mindkét jel esetében a következő lépésekkel foglalható össze:

1. Először egy átlagszűrőt ("mean filter") használtunk a referenciaszint megállapításához.
2. A referenciaszint ezután kivonásra kerül a hullámalakból. (Az optikai áramlás alapú jel esetén ez nem szükséges.)
3. A következő lépés a jelalak simítása w ablakhosszúságú átlagszűrővel, (ahol $w < hullmalakhossza$).
4. Ezután következik a jel adaptív küszöbölése az OTSU algoritmus segítségével ([82]).
5. Majd eltávolítjuk a Th_{lower} értéknél vékonyabb csúcsokat és összeolvasztjuk a Th_{upper} értéknél vastagabb csúcsokat. (Ezek empirikusan beállított paraméterek voltak.)
6. Végül megszámoljuk a lefutó éleket. A légzési rátát pedig úgy kapjuk meg, hogy a csúcsok számát elosztjuk a csúszó ablak hosszával.

A két módszer értékeléséhez egy manuálisan előállított referencia adathalmazt használtunk. Ez azt jelenti, hogy az adathalmazt egy emberi megfigyelő generálta, akinek figyelnie kellett a csecsemő légzését és minden alkalommal amikor légzőmozgást figyelt meg kattintott egyet az egérrel. A szoftverünk eltárolta a kattintás időpontjait, így generáltuk a referencia adathalmazt. A 5.4. ábra egy 40 másodperces példán mutatja be a kétféle jelnek a referenciával való korrelációját. Amint láthatjuk ezen példa esetében az SVD alapú jel is átlagosan követi a referenciát, viszont összességében az OF alapú jel jobban korrelál, mivel az egyes értékeik közötti abszolút különbség kisebb.



5.4. Ábra: A két bemutatott módszer korrelációja a manuálisan előállított referenciával. Amint az ábrán láthatjuk, az "optikai eljárás" (OF) alapú módszer értékei és a referencia (ground-truth - GT) értékei közötti különbség kisebb, mint a "szinguláris érték dekompozíció" alapú módszer értékei és a GT értékei közötti különbség.

5.1.4. Sűrű optikai áramlás alapú megközelítés és a végponttól-végpontig "end-to-end" neurális hálózat alapú megoldás

Mivel az előző alfejezetben bemutatott tanulmányban ([Au1]) az optikai áramlás alapú módszer teljesített jobban a referencia légzési rátákkal való korrelációt tekintve, ezért ebbe az irányba folytattuk az érintésmentes légzésfigyelő fejlesztését. A ritka optikai áramlás azonban rendelkezett egy hátránnyal, mégpedig azzal, hogy a jellemzőpontok kiválasztása nagyban függött a környezettől. Ha pedig a követendő pontokat szabályos rácson helyeztük el akkor nem biztos, hogy a pontok jól követhető pixeleknél helyezkedtek el. Tehát általánosságban robusztusság szempontjából

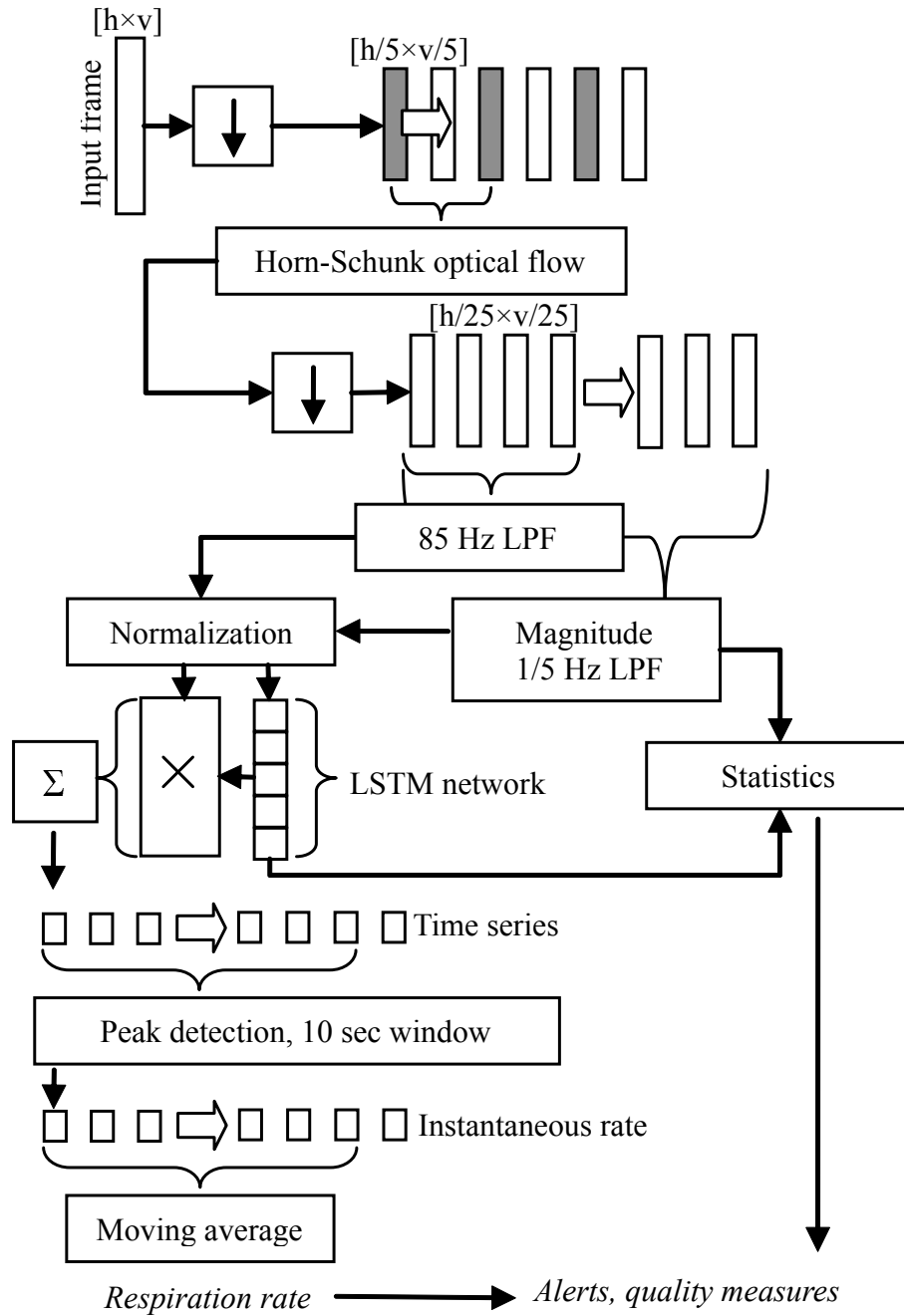
merült fel az ötlet, hogy nem kiválasztott pontokat kellene követnünk, hanem az összes pixel elmozdulását. Azaz, hogy ritka optikai áramlás helyett sűrű optikai áramlást alkalmazzunk, melynek kiszámításához a Farneback által leírt módszert alkalmazzuk [27].

A fő nehézség, amelyet meg kellett oldani a sűrű optikai áramlás alkalmazása esetén az volt, hogy nagyobb (pl. 1920x1080) felbontású képek esetén az optikai áramlás kiszámítása lassabb volt. Emiatt az algoritmusnak bizonyos (alacsonyabb számítási kapacitású) hardvereszközökre való implementálása esetén nem feltétlenül tudtuk volna biztosítani a valós idejű futást.

Ennek feloldására hoztunk létre egy olyan érintésmentes légzésfigyelő módszert, amely leskálázott képekkel dolgozik és arra optimalizált, hogy valós idejű futásra legyen képes alacsonyabb számítási kapacitással rendelkező hardvereken is, mint amilyen egy "Raspberry Pi 3/4" ([C1]).

Az eredményt a [C1]-ös közleményben ismertettük. Ebben az "RPE"-nek nevezett algoritmus (lásd a 5.5. ábrát) ahogyan említettem a sűrű optikai áramláson dolgozik, azaz a képek sorozatából előállított 2D mozgási képekből számítja ki a légzési jelet. Az algoritmus tartalmaz egy csúszó ablakot, amely úgy működik, hogy a kamerából érkező belépő képkockákat hozzáadja az ablakhoz és a kilépőket törli. Így lesz az ablak fix hosszú, amely mindig adott N számú képkockát tartalmaz. A képkockák sorozata mind térbeli mind időbeli felbontásban leskálázásra kerül. Ezután az eredeti méretük ötödére leskálázott képekre kiszámításra kerül a sűrű optikai áramlás. Az ennek eredményeként kapott elmozdulás vektorok ezután áthaladnak egy időbeli exponenciális aluláteresztő szűrőn. Ezen a ponton még minden egyes pixelhez egy elmozdulásvektor tartozik, az algoritmusunk azonban kiválasztja minden elmozdulásvektor y irányú összetevőjét és hozzárendeli az adott pixelhez, hogy minden pixelhez egy 8 bites, előjel nélküli egész érték tartozzon.

Ezután az "RPE" 20x20-as blokkokra osztja a térbeli felbontás mentén a sűrű optikai áramlásból kapott mozgási képeket. Majd a 20x20-as blokkok által kijelölt területekhez időbeli jeleket rendel oly módon, hogy az időbeli ablakban lévő mozgási képek adott blokk által lefedett részét átlagolja mind térben mind időben. Az algoritmus következő lépése egy (saját C++ implementáció alapú) LSTM (Long Short



5.5. Ábra: A bemutatott sűrű optikai áramlás alapú légzés figyelő eljárás ami alacsony számítási kapacitású hardware eszközökre lett optimalizálva. A $[h, v]$ a bemeneti kép magasságát és szélességét jelenti.

Term Memory) használata, amelynek feladata, hogy osztályozza az egyes időbeli hullámalakokat aszerint, hogy a hullámalakok által leírt lokális mozgások a következő kategóriák melyikébe tartoznak:

- 1) Légzésszerű mozgás
- 2) Hirtelen végtagmozgás
- 3) Mozgásmentes terület

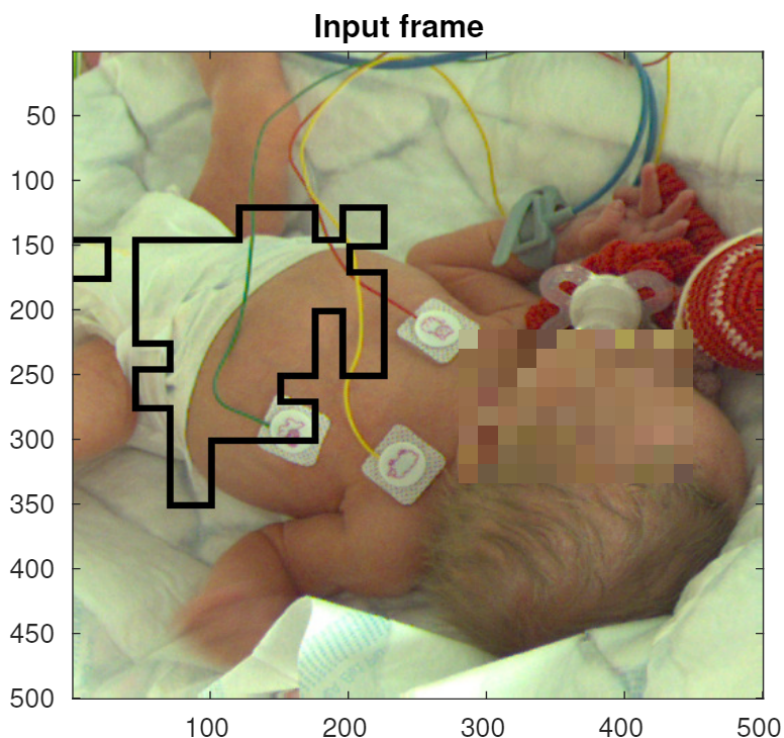
Az ily módon klasszifikált hullámalakok megsűrűsíthetők aszerint, hogy melyek tartoznak az 1. osztályba. Így egyrészt kapunk egy ROI-t, amely megmutatja, hogy melyik 20x20-as blokkon történik légzőmozgás és melyeken nem (Lásd a 5.6. ábrát). Másrészt a "légzés szerű" mozgásokat összegezve kapunk egy olyan mozgási hullámalakot, amely jól korrelál a csecsemő légzésével. A hivatkozott közleményben olyan módszert akartunk bemutatni, amely kis számítási kapacitású hardveren is képes valós időben futni. Ezért készítettünk egy saját C++ alapú LSTM -et itt, amelynek rövidebb a futási ideje, mint egy "pytorch" vagy "keras" alapú "deployed" verziónak.

Az "RPE" utolsó lépése, hogy a fentebb leírt módon kapott légzési egydimenziós hullámalakot felhasználva kiszámolja a légzési rátát, oly módon, hogy egyszerű csúcs detektort alkalmaz (amelyet a "python" alapú "SciPy" csomag "signal.find_peaks" függvénye valósít meg) és kiszámolja, hogy a 10 másodperces időablakban hány légzés történt.

Teljesen neurális hálózat alapú megközelítés

Meg kell még említeni, hogy a fentebb hivatkozott publikációban ([C1]) leírtuk azt is, hogy alkottunk egy másik "végponttól végpontig (end-to-end)" neurális hálózat alapú megközelítést is az érintésmentes légzésfigyelés elvégzésére, amelynek bemenete közvetlenül a színes képek sorozata, kimenete pedig maga a légzési ráta. Ez egy olyan konvolúciós neurális hálózat, amely 3D konvolúciókat alkalmaz a képek sorozatán és a következő két modulból épül fel:

- A) Hullámalak extraháló modul



5.6. Ábra: A bemutatott sűrű optikai áramlás alapú légzés figyelő eljárás pillanatképe, ahol ki vannak jelölve a "1) Légzés szerű mozgás" kategóriába tartozó blokkok.

- B) Légzési ráta számoló modul

Az A) modul feladata a hullámalak kivonása a képek sorozatából. A B) modul feladata pedig, hogy kiszámolja a légzési rátát az A) modul eredményeként kapott hullámalakra. Mivel mind a két modulja neurális hálózat, ezért az egész struktúra egy "végponttól végpontig (end-to-end)" neurális architektúra, amelyet együtt is lehet tanítani.

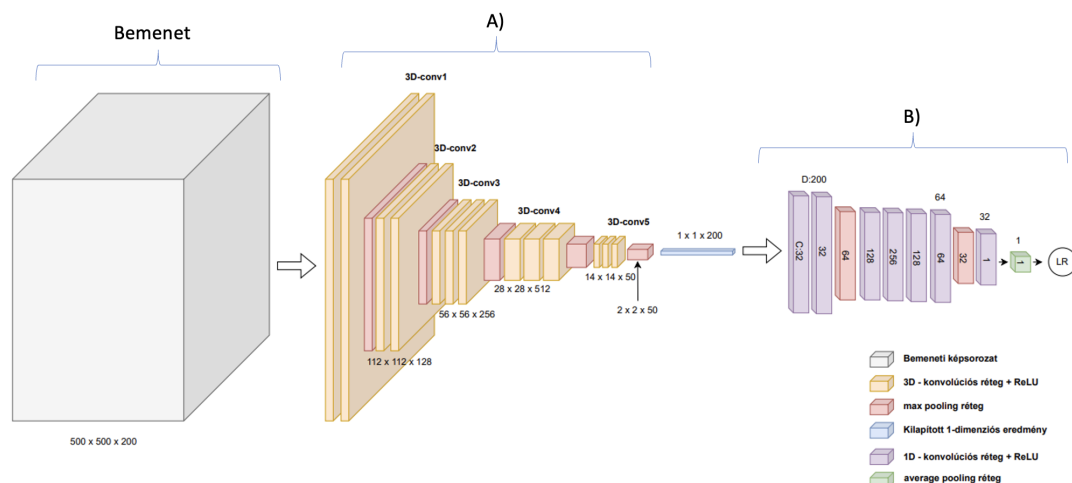
Az ötlet abból adódott, hogy a pulzus érintésmentes monitorozása során is alkalmazták. Példaként érdemes megemlíteni a PhysNet ([124]) és a DeepPhys ([21]) ("End-to-end") architektúrákat. Mi magunk is terveztünk egy az említettekhez hasonló konvolúciós hálózatot (pulzus helyett) a légzési ráta becslésére, amelyet a 5.7. ábra mutat be.

Ez a bemutatott hálózat képes volt megtanulni a légzési ráták képsorozatból történő meghatározását, azonban ennek a neurális hálózatnak a futási ideje jóval nagyobb volt (72.9-195.4 s), mint a korábban említett "RPE" algoritmusnak (57.9 s), ezért a fentebb említett publikációban ([C1]) ez utóbbit alkalmaztuk. Továbbá a

későbbiekben bemutatott "vegyes megközelítés" teljesítménye meghaladta az ("End-to-end") hálózataink által elért légzési ráta becslési teljesítményt. Ugyanez igaz a pulzus becslésre használt neurális architektúráknak (mint pl. a DeepPhys-nek [21]) a légzési ráta becslésre való áttanítására. A teljesítményük nem haladta meg a sűrű optikai áramlás alapú megoldások teljesítményét. Főként a kinyert légzési hullámalak minőségét tekintve.

Ennek oka minden valószínűség szerint abban keresendő, hogy ezeknek a hálózatoknak nagyságrendileg nagyobb tanító adathalmazra van szükségük a megfelelő teljesítmény eléréséhez, mint amivel mi rendelkezünk. (A 402 órányi felvételtől csupán 40-50 órányi adat volt alkalmas az érintésmentes légzésfigyelő tanítására.) Általánosan is elmondható, hogy a mély tanuló eljárások akkor teljesítenek jól és haladják meg a klasszikus algoritmusok teljesítményét, ha nagy mennyiségű annotált adathalmaz áll rendelkezésre. Elég csak megemlíteni a mély tanulás "állatorvosi lovának" tekintett MNIST adathalmazt, amely kézzel írott számokról (0-9) tartalmaz 32x32 felbontású szürkeárnyalatos képeket. Ezen jól ismert adatbázis esetében a tanuló halmaz 60000 képet tartalmaz és itt a feladat a képek 10 osztályba való besorolása volt. Általában komplexebb feladatok és nagyobb hálózatok esetén a megfelelő teljesítmény eléréséhez szükséges adathalmaz nagyobb. A légzési ráta becslésénél mi azonban a fentebb példaként említett adathalmazban lévőknel kevesebb mintával rendelkezünk a tanítások időpontjában.

Az adathalmaz méretén túl az is problémát jelenthetett a neurális hálózat számára, hogy a referenciaként használt impedancia-alapú eljárás is különböző frekvenciaszűréseknek vetette alá az elektromosan mért jelalakot, amelyeknek különböző frekvenciaeltolásai lehetnek és maga a jelalak feldolgozása és a légzési ráta kiszámítása is eltér. Pontosabban megfogalmazva az optikai és impedancia-alapú mérések különböző jelalakokat eredményeznek és a két módszer által generált jelalakok csúcsai között nem állandó időbeli eltolások lehetnek. Továbbá a szűrések hatására egyes csúcsok átcsúszhatnak az egyik időablakból a másikba. Így nem várhatunk tökéletes egyezést az optikai és az impedancia alapú módszerek által szolgáltatott légzési jelek és légzési ráták között. Ez a limitáció a későbbiekben leírt módszerek (pl. a "vegyes megoldás") esetében is fennáll. Az "End-to-end" eljárások esetén



5.7. Ábra: Az általunk érintésmentes légzés monitorozásra tervezett "end-to-end" neurális hálózat, amelynek A) modulja a légzés extrakciójáért felelős, B) modulja pedig a légzési ráta meghatározásáért. Az A) modul egy VGG inspirált konvolúciós hálózat, amiben a VGG-vel ellentétben 3D konvolúciókat alkalmazunk. A B) modul pedig egy 1D konvolúciós hálózat, amely az A) modul által előállított légzési jelen dolgozik. Ezt a jelet az ábrán kékkel jelöltük (és az alakja $[1 \times 1 \times 200]$).

azonban hatványozottan igaz, hogy az optikai és az impedancia alapú mérések közötti eltérés problémát okoz, ugyanis ez nagyon megnehezíti a hálózat számára a megfelelő paraméterek beállítását a tanulás során, különösen olyan esetben, ha a tanító adatok különböző forrásból jönnek, és a mérőeszközök különböző fáziseltolású szűrőket alkalmaznak.

5.1.5. A probléma megoldására javasolt algoritmus: vegyes megközelítés (gépi tanulás alapú ROI detektor és sűrű optikai áramlás alapú algoritmus)

Az optikai áramlás alapú módszerek jó hatásfokkal képesek kivonni a képek sorozatából azt az egydimenziós mozgási jelet, amely korrelál a légzéssel, ha a videón dominánsan légzésből eredő mozgás látható. Ha azonban egyéb a légzéshez nem kapcsolódó mozgások is megfigyelhetők a kameraképen akkor hatékonyságuk csökken. Leginkább olyan esetekre gondolok itt, mint hogy valaki benyúl a kép szélén vagy valamilyen egyéb mozgó objektum (pl. felakasztott belógó sztetoszkóp) van a képen. Józan ésszel is belátható azonban, hogy az optikai eljárást alkalmazó mód-

szerek jobb hatásfokúak és robusztusabbak lesznek, ha azokon a területeken vonjuk ki a mozgást a képek sorozatából, ahol a csecsemők légzése általában megfigyelhető. Mivel a csecsemők hasból lélegeznek, így fiziológiai szempontok alapján is beazonosítható, hogy a csecsemők hasánál vagy törzsénél kell a légzéssel kapcsolatos mozgást megfigyelni.

Bár korábban, például az "RPE" algoritmus esetében már alkalmaztunk egy olyan LSTM alapú módszert, amely ROI detektorként is funkcionált, ez a módszer nem a csecsemő hasát kereste, hanem a légzés-szerű mozgások helyeit és ily módon ez a módszer becsapható volt bármely olyan légzési frekvenciával mozgó objektum által, amely nem a csecsemő része, de hasonlóan mozog. A módszer akkor is tudott légzési rátát megállapítani, ha a csecsemő nem is volt a képen. Ezért jobbnak láttuk a ROI detekciót egy robusztusabb, különálló, statikus képjellemző analízis alapú eljárásra bízni. Algoritmusunkat tehát elláttuk egy különálló, mély tanulás alapú ROI detektorral. A U-Net nevű architektúra egy jól ismert neurális hálózat ([94]), amelyet alkalmazások széles köre használt szemantikus szegmentációs problémák megoldására [46, 58, 89]. Legjobb tudomásunk szerint először mi alkalmaztuk ezt a neurális hálózatot ROI detekcióra érintésmentes légzésfigyelőben.

A ROI detekció önmagában azonban nem oldja meg az összes problémát. Ha például ápolás történik és az ápolást végző személy eltakarja a U-Net által detektált területet vagy egy csecsemő nincs az inkubátorban akkor nem tudunk megfelelő minőségű légzési rátát biztosítani. Az előbbiekhöz hasonló szituációk kiszűrésére hoztunk létre egy olyan neurális hálózatot, amely különböző statikus és dinamikus jellemzők alapján képes megállapítani, ha olyan helyzet áll fenn ami nem teszi lehetővé az érintésmentes monitorozást. Az ezekhez hasonló helyzetek listája kiegészítve a megfelelő szituációval, amikor az általunk adott légzési ráta megbízható (1. osztály) a következő kategóriákat tartalmazza:

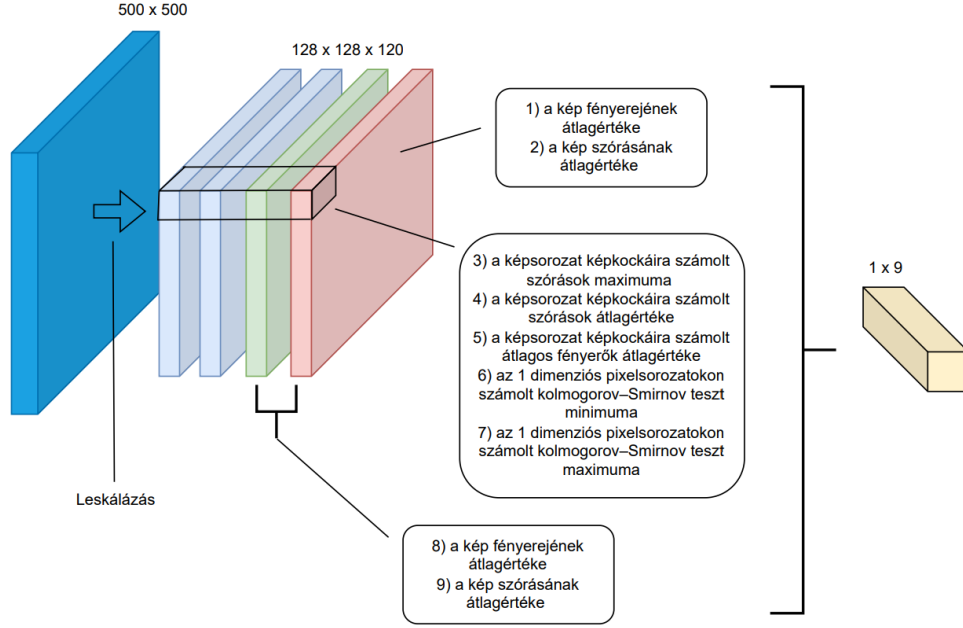
1. A baba jelen van, és légzésszerű jelet észlel.
2. A baba jelen van, intenzív mozgást mutat, amelyet véletlenszerű és gyakori végtagmozgásként értelmezünk.
3. Gondozás vagy egyéb beavatkozás történik.

4. Az inkubátorban nincs mozgás vagy minimális mozgás tapasztalható, de a baba észlelhető.
5. A csecsemő nem érzékelhető a jelenetben, az inkubátor üres.
6. A kamera képminősége nem megfelelő vagy egyéb hiba merült fel.

A fentebb bemutatott helyzetek klasszifikációjára a továbbiakban "szituáció analízisként" fogok hivatkozni. Az 1. osztályba tartozó szituációkban vagyunk képesek megbízhatóan légzést becsülni, míg a többi osztályba tartozó szituációk "problémás" szituációk amikor nem tudunk megfelelő minőségű légzési rátát mérni vagy nem is lehetséges a mérés (pl. amikor csecsemő nem látható a képen). Ennek az osztályozási problémának a megoldására egy LSTM ([38]) alapú neurális hálózatot alkalmaztunk, amelynek összefoglalása a 5.9. ábrán látható. A szituáció analízis bemenete a videófolyam, a kimenete pedig egy döntés arról, hogy az érintésmentes megfigyelés szempontjából "problémás" szituáció áll-e fent. A szituáció analízis "statikus" és "dinamikus" analízist is végez a bejövő képkockákon. A bemeneti képkockák átméretezésre kerülnek az algoritmus kezdetén $[128 \times 128 \times 3]$ alakúra lineáris interpolációval.

A statikus analízis csakis az aktuális képkockán dolgozik. Ennek során egy YOLO-V3 [92] architektúra azonosítja a csecsemő testét (fej és törzs), a szülők jelenlétét (kesztyű nélküli kezek) vagy az egészségügyi személyzet (kék kesztyű) és szoptatós üvegek jelenlétét. Az erre alkalmazott YOLO-V3 architektúrát szintén mi tanítottuk be.

A dinamikus analízis azonban az aktuális képkockán és az azt megelőző $N=119$ darab képkockán, vagyis egy $[128 \times 128 \times 120]$ alakú tenzoron dolgozik. A kinyert dinamikai tulajdonságok a videó alacsony szintű jellemzői, például a kamerazaj, az átlagos fényerő és a periodikus mozgások. Egyes jellemzők, mint pl. a kép fényerejének középértéke (5.3. egyenlet) vagy a kép szórásának középértéke (5.4. egyenlet), az individuális képkockákra lettek kiszámolva, míg más jellemzők, mint az egydimenziós pixelvektorok (időbeli) szórásának maximuma (5.5. egyenlet) pedig a képek sorozatára, pontosabban a képek sorozatának minden egyes pixele mentén ki lettek kiszámolva. A dinamikus jellemzők kiszámolásának mikéntjét a 5.8. ábra szemlélteti.



5.8. Ábra: Az egyes dinamikus jellemzőket a képek sorozatából háromféleképpen számoljuk ki: a) egyes jellemzőket az aktuális legújabb képből számolunk ki (piros kép). b) Más jellemzőket az aktuális időablakban lévő képsorozat egyes pixelein végiglépve számolunk ki. c) Végül egyes jellemzőket az aktuális (piros) képből és az azt megelőző (zöld) képből számolunk ki.

$$F_1 = \frac{1}{N} \sum_{i=1}^N \overline{X_i^b}, \quad (5.3)$$

$$F_2 = \frac{1}{N} \sum_{i=1}^N \sigma(X_i), \quad (5.4)$$

$$F_5 = \max_{x,y} \{s(X)_{x,y}\} \quad (5.5)$$

A statikus és dinamikus analízisből összesen 13 jellemzőt számolunk ki, amelyek az LSTM bemenetét fogják képezni. A kivont jellemzők teljes listája a 5.1. táblázat mutatja be.

A statikus analízis csakis az aktuális képkockán dolgozik. Ennek során egy YOLO-V3 ([92]) architektúra azonosítja a csecsemő testét (fej és törzs), a szülők jelenlétét (kesztyű nélküli kezek) vagy az egészségügyi személyzet jelenlétét (kék kesztyű) és szoptatós üvegek jelenlétét (magának az üvegnek a jelenléte).

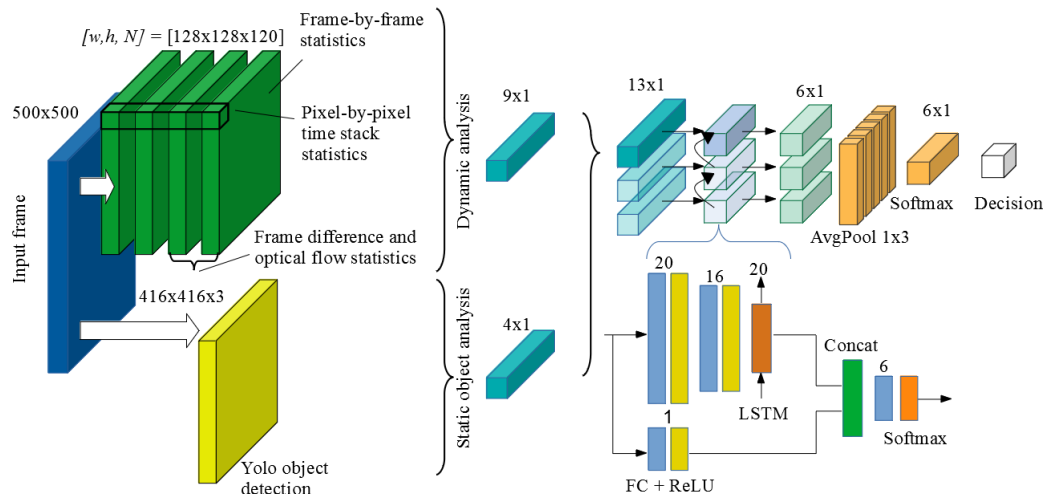
Az első tételtemet alátámasztó publikációban ([J1]) mutattunk be egy olyan légsfigyelő algoritmust, amely egy U-Net neurális architektúrát használ ROI-detektorként

1) a kép fényerejének átlagértéke
2) a kép szórásának átlagértéke
3) a képsorozat képkockáira számolt szórások maximuma
4) a képsorozat képkockáira számolt szórások átlagértéke
5) a képsorozat képkockáira számolt átlagos fényerők átlagértéke
6) az egydimenziós pixelsorozatokon számolt kolmogorov–Smirnov teszt ([34]) minimuma
7) az egydimenziós pixelsorozatokon számolt kolmogorov–Smirnov teszt ([34]) maximuma
8) differenciakép abszolút átlaga az aktuális és az azt megelőző képkockára számolva
9) a képsorozat pixeleire számolt egydimenziós vektorok szórásának maximuma
10) csecsemő jelenlétének detektálása YOLO-V3 által
11) kesztyű nélküli kezek jelenlétének detektálása YOLO-V3 által
12) kék kesztyű jelenlétének detektálása YOLO-V3 által
13) szoptatós üvegek jelenlétének detektálása YOLO-V3 által

5.1. Táblázat: Az LSTM alapú szituációanalízis által használt bemenetet alkotó statikus és dinamikus jellemzők teljes listája.

és a korábban leírt szituációanalízis döntését is felhasználja az olyan szituációk detektálásához amikor a bejövő képekből nem állapítható meg a légzési ráta. Ennek az érintésmentes légzésfigyelő algoritmusnak a működését a 5.10. ábra foglalja össze. Ez az algoritmus, amelyet az idézett cikkben "Respiration Calculator Block"-nak nevezünk először megkeresi a csecsemő törzsét, majd kivonja a hullámszerű légzés jelet a videófolyamból a törzset tartalmazó területen. Végül kiszámítja a légzési rátát. Ennek az algoritmusnak egy python alapú implementációja megtalálható a következő github tárhelyen: [77].

Fontos megjegyezni, hogy amíg a törzs megtalálása egyetlen képkockán is megtörténhet, addig az optikai áramlásnak szüksége van nemcsak az aktuális t időpillanathoz tartozó képkockára, hanem az azt megelőző $t - 1$ időpillanathoz tartozó képkockára is, hogy a két képkocka közötti pixel elmozdulásokat becsülni tudja. Ezen felül a mozgási jel extraháló modul is egynél több képkockán dolgozik. Ez a modul 200 képkockából, vagyis pontosabban 200 (az optikai áramlás alkalmazásával generált) mozgási képből állít elő egy egydimenziós mozgási jelet, amely korrelál a



5.9. Ábra: Az ábra bal oldala a csúszó ablak és az annak tartalmát alkotó képkockák méreteit mutatja be. A középső ábrán láthatók a dinamikus és statikus objektumelemzésből származó jellemzővektorok. Míg a jobboldali ábra az ezen vektorokat feldolgozó LSTM alapú hálózat látható. A jobboldal alulsó részén látható, hogy egy modul miként épül fel.

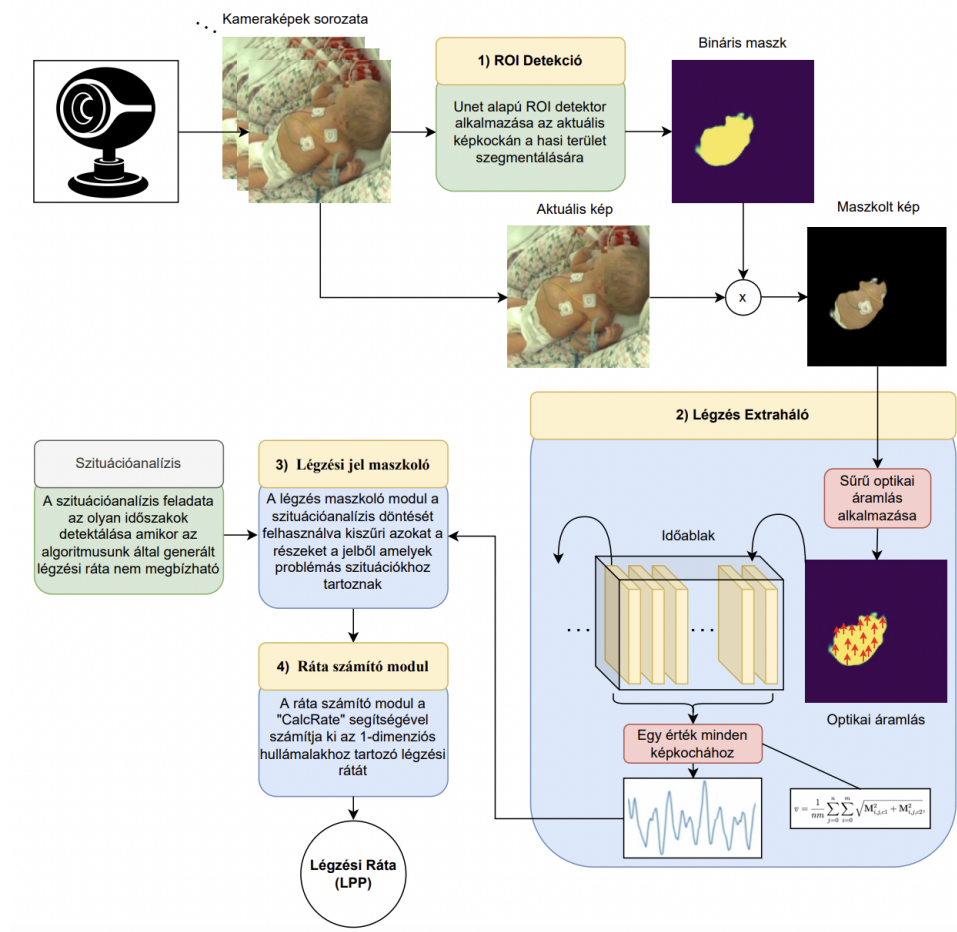
légzéssel.

Az algoritmus leírása

A következőkben bemutatom azt a korábban hivatkozott közleményünkben publikált, (gépi tanulás alapú ROI detektorral ellátott) érintésmentes légzésfigyelő algoritmusomat, amely az 1. tézisemben szereplő algoritmus. Ennek bemenetei a szituáció analízis döntése és a videófolyam (a kamerából érkező képek sorozata). A kimenete pedig maga a légzési ráta. Az algoritmus 4 alrészre osztható, amelyek a következők:

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

A ROI detektor minden n . bejövő képkockára meghatározza a képkockán lévő csecsemő törzsének helyét ($n=20$). (Nem fordulhat elő, hogy a képen nincs csecsemő,

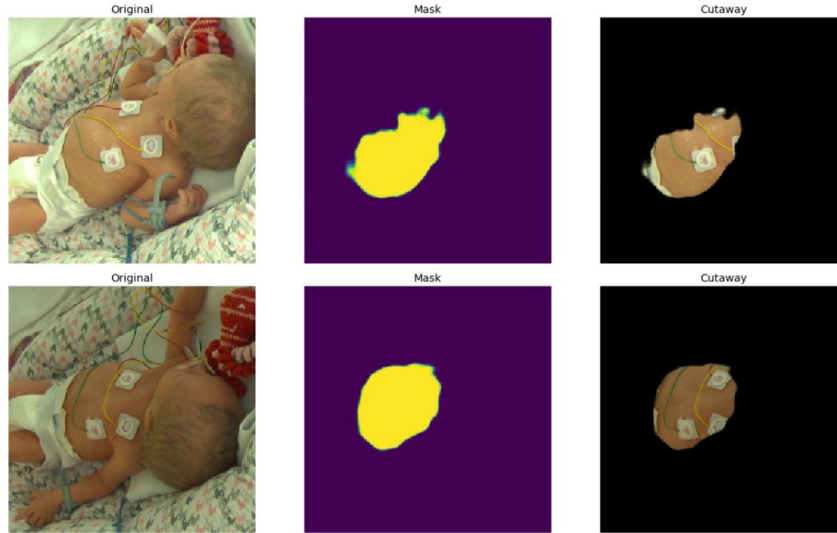


5.10. Ábra: A korábban hivatkozott közleményben ([J1]) publikált érintésmentes légzés monitorozó algoritmus, amely egy U-Net alapú ROI detektorral (1) indul és egy "sűrű" optikai áramlás alapú légzés extraktorral (2) rendelkezik és egy csúcsdetektor alapú légzési ráta kiszámolóval végződik (4). Ebbe az algoritmusba be van kötve a "szituációanalízis" eredménye is ami segít a problémás szituációk megtalálásában (3), hogy az algoritmus meg tudja állapítani mikor képes megbízható légzési rátát szolgáltatni.

mert ezesetben a szituáció analízis döntése jelzi számunkra, hogy az adott szituáció nem megfelelő légzési ráta becslésére.)

1) ROI Deketor A ROI Detektor, ahogyan már említettem egy U-Net [36], 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. Intenzív mozgások esetén a csecsemő törzse is a légzőmozgástól független mozgást végezhet. Vannak azonban olyan esetek amikor a csecsemő csak a kézfejét mozdítja meg vagy a lábát és a törzsét nem. Ilyen-

kor a ROI detektor által megtalált területen a légzés zavartalanul figyelhető meg. A ROI detekció eredménye egy bináris maszk, ahol a törzshöz tartozó képpontok 1-ek, míg a törzsön kívüli képpontok 0 értéket vesznek fel (Lásd az 5.11. ábrát). Ez a ROI-t kijelölő bináris kép lesz a "Légzés extraháló modul" egyik bemenete.



5.11. Ábra: Ezen az ábrán a U-net alapú ROI detektor bemeneteit és az azokra szolgáltatott eredményeit láthatjuk. A baloldali képeken a két szokványos színes kamerából kapott példabemenet látható. Míg a középső képeken a (U-net által) ezen példákra adott bináris kép látható. A jobboldali képeken láthatjuk, hogy mennyire pontosan szegmentálta a háló a csecsemők hasát.

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 (Lásd a 5.12. ábrát) a következőképpen: Az új beérkező képből és az azt megelőző képből kiszámítjuk a mozgási képet (Farneback optikai áramlásának segítségével [27]) és az ennek eredményül kapott mozgási képet elemenkénti szorzással megszorozzuk a "ROI Detektor"-ból származó bináris képpel a következő egyenletnek megfelelően:

$$\mathbf{M} = \text{U-Net}(\mathbf{I}(t)) \odot \text{Dense}(\mathbf{I}(t), \mathbf{I}(t-1)), \quad (5.6)$$

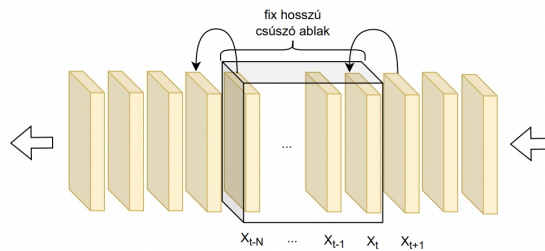
ahol $\mathbf{I}(t) \in \mathbb{R}^{m \times n}$ az aktuális időpillanathoz tartozó kamerából érkező képkocka, ahol felvételeink során $m=500$ és $n=500$ volt jellemző. A "U-Net()" metódus a ROI

detektor alkalmazása az aktuális képkockán, míg a "Dense()" metódus Farneback "sűrű" optikai áramlását számítja ki az aktuális és az azt megelőző képkockákra. **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 (5.12. ábra). Ezek után az ablak minden egyes képéhez egy (v) értéket rendelünk a 5.7. 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 ([17]) sáváteresztő szűrővel (20–120 RPM) frekvenciaértékek között, hogy végeredményként egy olyan hullámalakot kapjunk kimenetként, amelynek frekvenciája a légzés fiziológiai frekvenciatartományába esik és korrelál a képen lévő csecsemő légzésével.

$$v = \frac{1}{nm} \sum_{j=0}^n \sum_{i=0}^m \sqrt{M_{i,j,c1}^2 + M_{i,j,c2}^2}, \quad (5.7)$$

ahol $M_{i,j,c1}$ 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: $\mathbf{s} = [v_1, v_2, \dots, v_N]$.



5.12. Ábra: Egy N hosszú csúszó ablak, amely az aktuális kamerából bejövő $\mathbf{X}(t)$ képkockát és az azt időben megelőző $N-1$ darab képkockát tartalmazza. Minden egyes újonnan beérkező képkockánál törli az ablak első elemét és a végéhez hozzáadja az új képkockát.

3) Légzési jel maszkoló modul

A korábban bemutatott "szituációanalízisből" 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 0-ák tartoznak azokhoz, amikor problémás szituáció lép fel. A Légzési jel maszkoló modul kiszűri a problémás szituációhoz tartozó mozgási jeleket azáltal, hogy a kapott hullámalakot 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, 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 többféle megközelítést vizsgáltam. Triviális megoldásnak tűnt, a légzési frekvencia megállapításához egy gyors Fourier-transzformációt (FFT - "Fast Fourier Transformation") alkalmazni. A csecsemők légzési görbéje azonban olyan jellegű, hogy hirtelen megváltozhat mind amplitúdójában, mind frekvenciájában, illetve rövid, fél periódus légzéskimaradás miatt ellentétes fázisba ugrik a légzés. Ezek a hirtelen változások megzavarhatják az FFT-t a légzési ráta megállapításában. Ezért az FFT alkalmazását néhány sikertelen kísérlet után elvetettük. Ahogyan a különböző "wavelet" alapú megoldásokat is, mivel azok sem nyújtottak meggyőző eredményt a referenciával való összehasonlítás során. Próbálkoztunk még továbbá az olyan programcsomagok által szolgáltatott csúcsdetektorokkal is, mint amilyen a "SciPy" (<https://scipy.org/>) által szolgáltatott "find_peaks", de a legjobb légzési ráta számítást mégsem a fentebb említett algoritmusok, hanem az általam tervezett adaptív küszöbölés alapú csúcsdetekciós algoritmus adta, melyet

"CalcRate"-nek neveztünk a fentebb hivatkozott publikációban ([J1]).

Ez az algoritmus a jel megfordításával (1) és egy sáváteresztő szűrő alkalmazásával (2) indul, amely a 20 és 120 közötti frekvenciákat ereszti át. Ezt követi egy adaptív küszöbölési lépés, ahol a küszöbszint a bemeneti jel adaptív (a jel aktuális környezetében számolt) középértéke (3). Ily módon bináris jelet kapunk, ahol az 1-es érték a légzéseket (a pozitív csúcsokat) jelenti, a légzések kezdetét pedig a felfutó élek jelzik (4). A felfutó élek megszámlálásával (5) megállapíthatjuk az adott "idő ablakban" történt légzés számot. Az utolsó lépésben a 0,3 másodpercnél közelebbi légzések összevonásra kerülnek (6), és a csúcsok közötti időintervallumok átlagolásával és az eredmény frekvenciaértékké alakításával számítjuk ki a légzési frekvenciát (7). A folyamat összefoglalása a 5.13. ábrán látható. A bemeneti jel megfordítására azért volt szükség, mert a kapott jelalakok olyan formájúak voltak, hogy a negatív csúcsok hangsúlyosabban jelentkeztek. Azaz a lokális minimum értékek mindig nagyobb távolságra voltak a referenciaszinttől, mint a lokális maximumok. Így az adaptív küszöbölési lépés (3) könnyebben detektálta őket.

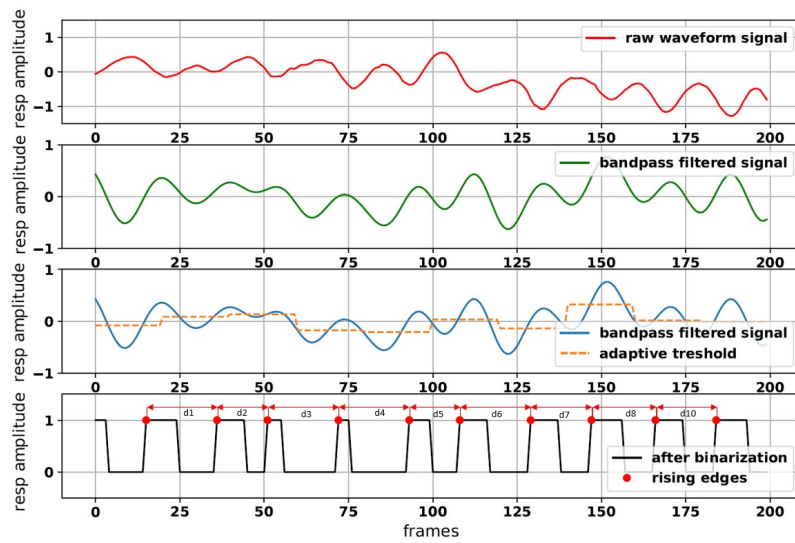
Az algoritmus értékelése

A fentebb bemutatott (U-Net alapú ROI detektorral ellátott) légzésfigyelő algoritmust több szinten is kiértékeltek a korábban referált munkában ([J1]) és össze is hasonlítottuk egy kifejezetten az (érintésmentes és csecsemőkön történő) légzésfigyelés területén "state-of-the-art"-nak tekinthető munka eredményeivel. Hangsúlyozni szeretném, hogy nem olyan munkával akartuk összehasonlítani az eljárásunkat, amely pulzus és légzés figyelésére létrehozott algoritmus, vagy egy pulzusfigyelő algoritmus légzésfigyelésre történő adaptálásával, hanem kifejezetten olyan munkával amely koraszülött csecsemők légzésének érintésmentes hagyományos kamerával történő monitorozásával foglalkozik.

Az értékelés első szintjén a "légzés extraháló modulból" kapott hullámalakot hasonlítjuk össze a referencia hullámalakkal.

A második szinten a hullámalakokból számított LPP-ben kifejezett légzési ráták összehasonlítása történik.

Természetesen az értékelés olyan felvételeken történt amelyeken az adathalmaz-



5.13. Ábra: A "CalcRate" algoritmus működése: A felső sorban látható az invertált egydimenziós mozgási jel, amelyet a ROI területén nyertünk ki (1). A második sorban a sáváteresztő szűrőn átengedett jelet láthatjuk (2). A harmadik sor mutatja szaggatott vonallal a középértékek helyének meghatározását (3). Az utolsó sorban láthatjuk az adaptív küszöbölt jelet (4) és a felfutó élek megszámlálását (5). Szintén az utolsó sorban ábrázoljuk a csúcsok közötti távolságok megállapítását. Ha a csúcsok közötti távolság időben kifejezve kisebb, mint 0,3 másodperc, akkor a csúcsokat egyesítjük (6). A légzési ráta kiszámítása a csúcsok közötti időintervallumok átlagolásával és az idődifferenciák átlagának reciprokértékét kiszámítva történik (7).

tól független babák szerepelnek. Ez alatt a függetlenség alatt azt értjük, hogy a teszthalmazba tartozó csecsemőkről készült felvételek nem kerültek a tanító halmazba.

A felhasznált referencia

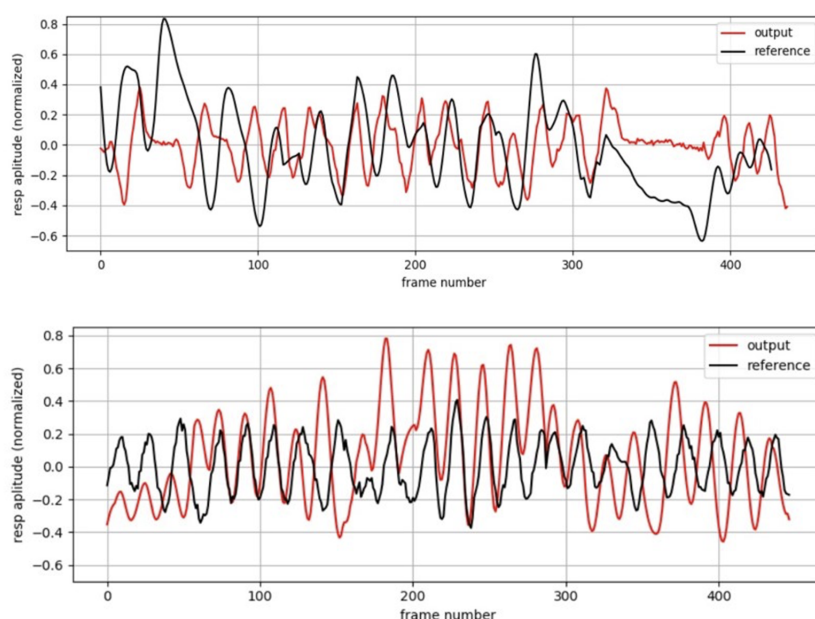
A korábban bemutatott eljárások esetén alkalmaztunk manuálisan rögzített referenciaadatokat, EKG jelalakokat és légzési rátákat is. A most bemutatott vegyes megközelítés esetében a referenciát a "Philips IntelliVue MP20/MP50" orvosi monitor szolgáltatja, ami egy a koraszükkött klinikákon rutinszerűen használt eszköz. A méréseink során csatlakoztatva volt a monitorhoz egy EKG, amelynek 3 elektródáját az orvosi személyzetnek egy tagja a csecsemő bőrének megfelelő pontjaira ragasztotta fel. A monitor hullámalakot és numerikus légzési ráta adatokat is biztosított a számunkra, (amely LPP-ben volt kifejezve). Az eszköz dokumentációját tanulmányozva megállapítottuk, hogy egy légzési ráta érték egy 10 másodperces időablakban található légzési hullám frekvenciájának felel meg.

Az értékelés első szintje: a hullámalakok összehasonlítása

Az első szintű értékelés során a légzési extrahálóból származó hullámalakokat hasonlítottuk össze a "Philips IntelliVue MP20/MP50" monitorokból származó hullámalakkal. A hullámalakok közötti összefüggés méréséhez abszolút hibát és négyzetes hibát számoltunk a két jelalak között.

Bár az algoritmusunk által szolgáltatott és a referencia jel is tükrözi a mellkas légzés során történő mozgását, kinyerésük és ezért reprezentációjuk is eltérő, ezért nem számíthatunk arra, hogy a két jelalak tökéletesen illeszkedni fog egymásra. A jelek különböző módon reagálnak a végtagok mozgására és a kinyerésük különbözik. Sőt, különböző frekvenciaszűréseken esnek át. A vizualizáció alapján látható, hogy a jelalakok között jelentős, időben változó fáziseltolás tapasztalható. Emiatt az sem biztosítható, hogy mindkét hullámalak ugyanannyi csúccsal rendelkezik.

A görbék összehasonlítását a 5.14. ábra szemlélteti, ahol a két jel közti összefüggés átlagos abszolút hibában - Mean Absolut Error (MAE = 0,5494) és négyzetes hibában - Root Mean Square Error (RMSE = 0,6768) fejezhető ki.



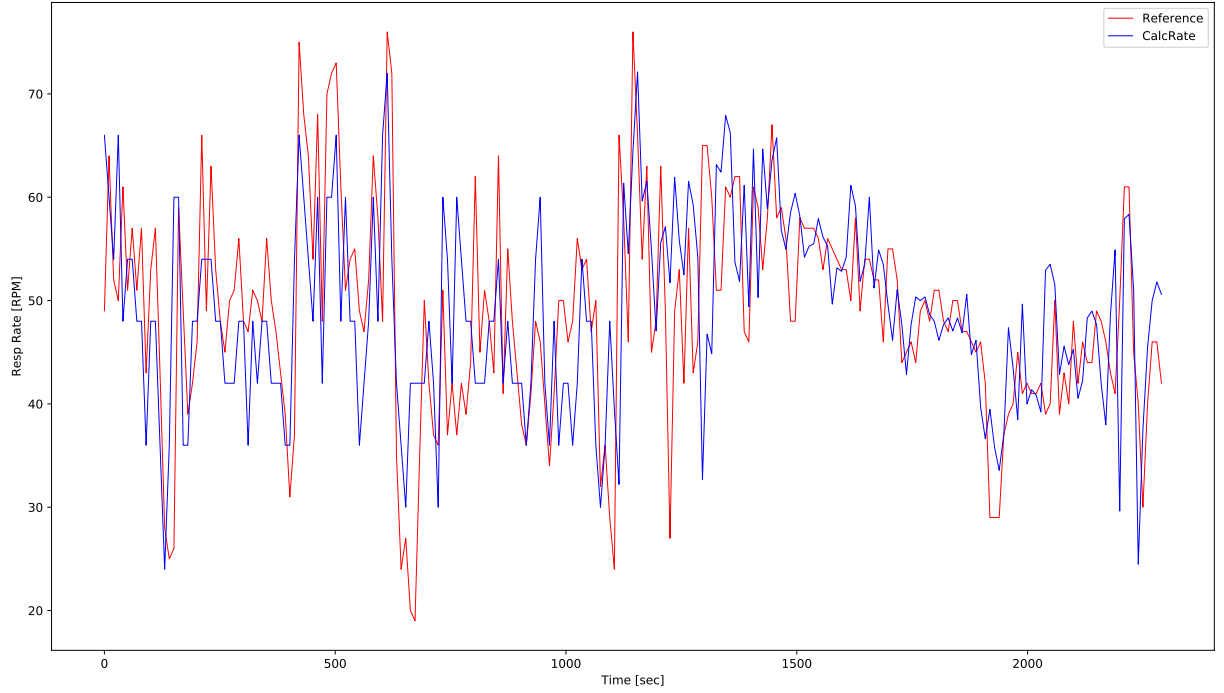
5.14. Ábra: A számított légzési jel (piros görbe) összehasonlítása a referencia értékkel (fekete görbe) két jellemző szakaszon. A tözs mozgása és az alkalmazott szűrők miatt mind a referencia számításnál, mind a tárgyalt algoritmus esetében megváltozik a jelek fázisa, de a csúcsok száma közel ugyanaz (MAE: 0,5494, RMSE: 0,6768).

Az értékelés második szintje: a numerikus összehasonlítás, a légzési ráták összehasonlítása

A második értékelési szint a kimeneti légzési ráták összehasonlítása volt a referencia rátákkal. A Philips IntelliVue MP20/MP50 monitor által szolgáltatott légzési ráták és a tárgyalt érintésmentes légzésfigyelő által a 3 órás független szakaszra adott légzési ráták összehasonlítását a 5.15. ábra foglalja össze.

Amint látható, a számított értékek és a referencia értékek közötti különbségek leginkább a szélsőséges értékek körül fordulnak elő. A többi időpillanatban viszonylag jól korrelál a tárgyalt algoritmus által generált ráták görbéje a referencia ráták görbéjével. Azt, hogy pontosan milyen mértékben korrelálnak, a statisztikai analízis mutatja meg.

A statisztikai értékelés eredményét a 5.2. táblázat mutatja be. Az értékelés ezen szintjéhez abszolút hibát (Mean Absolute Error - MAE) és négyzetes hibát (Root



5.15. Ábra: A referencia ráták (piros) és a tárgyalt algoritmus által generált légzési ráták (kék) összehasonlítása. A két görbe tredje láthatóan illeszkedik egymásra.

Mean Square Error - RMSE) számoltunk a javasolt algoritmus teljesítményének értékeléséhez.

$$MAE(\mathbf{d}, \mathbf{f}) = \sum_{i=1}^D |\mathbf{d}_i - \mathbf{f}_i| \quad (5.8)$$

$$RMSE(\mathbf{d}, \mathbf{f}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mathbf{d}_i - \mathbf{f}_i)^2} \quad (5.9)$$

A legfelső sorban látható az ajánlott algoritmus teljesítménye. A táblázatban ezen felül feltüntettem, hogy a "DeepFlow" nevezetű algoritmus miként teljesít az általam tervezett algoritmussal összevetve. Ugyanis a fentebb hivatkozott publikációban ([J1]) egy olyan munkával hasonlítom össze a légzésfigyelő algoritmusomat, amely ezt a módszert ajánlja a légzés képsorozatból történő extrahálásához.

Ahogy a 5.2 táblázatban is láthatjuk a CalcRate jobban teljesített a vizsgált 3 órás adathalmazon. Kísérleteink azt mutatták, hogy ennek több oka is van. Egyrészt a "DeepFlow" algoritmus által történő légzési mozgás számítás érzékenyebbnek

Architektúra	MAE	RMSE
<i>CalcRate</i>	5.080 LPP	6.618 LPP
DeepFlow + findPeaks	14.196 LPP	17.122 LPP
DeepFlow + diffRate	5.848 LPP	7.553 LPP

5.2. Táblázat: A bemutatott (U-Net alapú ROI detektorral) ellátott érintésmentes légzésfigyelő statisztikai értékelése. Késsel jelöltük az általunk tervezett algoritmust, míg a másik két sorban a "DeepFlow" alapú algoritmus eredményét láthatjuk különböző csúcsdetekciós eljárásokat alkalmazva.

bizonyult a kisebb mozgásokra. Legalábbis az ezen algoritmus által generált mozgási jelben több hamis csúcs alakult ki mint a Farneback sűrű optikai áramlás által generált mozgási jelben.

Másrészt a számított légzési ráták kiszámítása nagyon érzékeny arra, hogy milyen csúcsdetekciós eljárást alkalmazunk. A DeepFlow algoritmus által generált mozgási jelekből számított ráták és a referenciaértékek között 14.196 LPP abszolút hiba mérhető, ha egyszerűen a "SciPy" (<https://scipy.org/>) programcsomag által szolgáltatott "find_peaks" csúcsdetektort alkalmazzuk. Ez ugyanis sok hamis csúcsot is detektál a mozgási jelekben. Ha egy szofisztikáltabb módszer segítségével az alapvonalon áthaladó éleket vagyis a hullámok leszálló és felszálló ágait keressük és az így meghatározott csúcsok közötti távolságokat tekintjük (diffRate) a légzési frekvencia kiszámításához akkor már csak 5.848 LPP abszolút hibát kapunk. A legjobb eredményt mégis az általunk bemutatott algoritmus adta a "Ráta számító modul"-nál alkalmazott csúcsdetekciós eljárással. Az általunk használt "CalcRate" nevezetű eljárást kifejezetten a "légzési extraháló modul" által generált 200 adatpont hosszú mozgási jeleken történő csúcsdetekcióra fejlesztettük ki.

Ezen felül megvizsgáltuk, hogy az implementált és tesztelt algoritmus milyen futási sebességre képes Intel(R) Xeon(R) CPU @ 2.20GHz processzorral rendelkező PC-n. Ezen az eszközön a futási idő 720 ms volt az összes lépést bele értve.

6. Fejezet

Automatikus annotált adathalmaz generálás

Az előző fejezetben leírt légzésfigyelő fontos eleme volt a ROI detektor. Ez ugyanis mind a zavaró végtag mozgások kizárásával, mind pedig a jel felerősítésében, a jel-zaj viszony javításában fontos szerepet játszik. A [J1]-es publikációban leírt vegyes megoldásban ROI detektorként egy neurális hálózatot, pontosabban az U-Net nevű hálózatot használtunk. Ami azonban az ilyen neurális hálózatoknak közös tulajdonsága, hogy viszonylag nagy mennyiségű annotált halmazra van szükség a betanításukhoz. Kezdetben magunk annotáltunk a „labelme” nevű annotáló program segítségével. Ez azonban hosszadalmas és nagy mennyiségű emberi erőforrást vesz igénybe. A hosszadalmas manuális annotáció elkerülése érdekében merült fel, hogy olyan megoldások felé forduljunk, amelyek részben vagy egészében automatizálhatják az annotálási folyamatot, a manapság nagy népszerűségnek örvendő "self-supervised" vagy "semi-supervised" eljárásokhoz hasonlóan. Egyszerűbb megoldásként azonban egy ezektől kissé eltérő (mozgási kiterjedés és frekvencia alapú) automatikus annotációt végző algoritmust hoztunk létre (ALA néven), hogy emberi beavatkozás nélkül annotált adathalmazt generáljon a számunkra bármely előtanítás, vagy részben annotált minták megléte nélkül is. Ezt úgy tettük meg, hogy létrehoztunk egy eljárást, amely a U-Net által figyelembe vett geometriai és színjellemzők helyett a képkockák sorozatán a légzési frekvencia tartományban lévő mozgásokat kereste, és látható mozgások kiterjedése és frekvenciája alapján határozta

meg a légzés helyét, így detektálva a csecsemő hasa körüli területet. Ez a módszer számos tényezőre érzékeny, mint például a fény-árnyék változások, végtag mozgások stb. Így valós idejű ROI detekcióra nem alkalmas, ugyanakkor nagy mennyiségű annotált adat generálására igen. (A 6.1. ábrán látható néhány példa elvárt kimenetekről, amelyek annotált adatok, tehát be van rajtuk jelölve a csecsemő hasa körüli terület.) A megoldandó probléma tehát ebben a fejezetben az, hogy egy olyan algoritmust alkossunk, amely képes automatikusan felannotálni a csecsemőkről készült képeket az érintésmentes légzésfigyelő algoritmushoz.



6.1. Ábra: Néhány példa, hogyan néz ki az annotáció az eredeti képen. Pirossal láthatóak azok a területek, amelyeken az elvárt kimenetek esetében a bináris képen 1-esek találhatók. Mint látjuk ezek a babák hasa vagy háta körüli területen találhatók.

6.1. Kapcsolódó szakirodalom

A szakirodalomban található automatikus annotációs megközelítéseket tekintve, érdemes megemlíteni a [113], [122] és [83] publikációkat. Az első kettő cikk tanulsága szerint, a CycleGAN-nak vagy annak módosulatainak alkalmazása jó alternatívát jelenthet egyszerűbb adatkészletek automatikus annotálására. Ezen publikációkban ugyanis a U-Net-et tanították be a generált adatkészleten annak érdekében, hogy konkrét szegmentációs problémákat oldhasson meg és az ehhez szükséges annotációt különböző automatikus módszerekkel állították elő. A [113] és az [122] eredményeire alapozva elmondható, hogy a CycleGAN és a RailGAN, még viszonylag kevés kiindulási mintából tanulva is, képesek olyan adatkészletek generálására, amelyek anyag felületi hibákról és vasúti hibákról tartalmaznak képeket. A [113] esetében a CycleGAN hibás képet generál egy hibamentes képből, és a generált hibás kép és a hibamentes kép közötti különbség szolgáltatja a pixel-szintű szegmentációs annotációt. A RailGAN esetében pedig a RailGAN által generált hibás képeket a Roboflow segítségével címkézték. A CycleGAN-hoz hasonló generatív modellek kiválóak az ilyen egyszerűbb hibás képek és szegmentációs annotációk generálására, mivel a domain transzformációra lettek kifejlesztve. Azonban valószínűleg kevésbé lennének sikeresek a képeken ábrázolt összetettebb helyzetek kezelésében vagy az ilyen bizonytalanabb szegmentációs feladatokban, mint például a csecsemő törzsének megtalálása és az ehhez szükséges annotált adatkészlet generálása. Ráadásul valószínűleg több címkézett kiindulási adatra lenne szükségük ahhoz, hogy ezt megtanulják. Az esetünkben nincs szükség kiindulási annotációkra, mivel klasszikus algoritmusokat alkalmazunk az automatikus annotáció végrehajtásához. Ráadásul az általunk használt automatikus annotációs módszer kifejezetten erre az alkalmazásra lett kifejlesztve.

Szintén egy generatív eljárás alapú megoldást mutat be szemantikus szegmentáció automatikus annotációjára Daniil et al munkája ([83]). Ebben a StyleGAN2 kerül alkalmazásra. Ebben az esetben pl. emberi arcokat, kutyákat és macskákat ábrázoló képeket generáltak. Az automatikus annotáció pedig ezután úgy történt, hogy egy klaszterező algoritmus a kép minden egyes pixelét különböző összefüggő

régiók klasztereihez rendelte hozzá. Véges számú klaszter volt. Ehhez a klaszterezéshez pedig a konvolúciós hálózatalapú generátor egyes rétegei által adott "feature map"-eket használta fel. Ez az eljárás is kiválóan teljesített azon alkalmazás esetén amelyhez kapcsolódóan bemutatták.

A mi alkalmazásunk esetében azonban néha olyan nem összefüggő régiókat kell megtalálni, mint a csecsemő pelenkájának és a bőrfelületnek a találkozása, amelyeket a klaszterező eljárás minden valószínűség szerint külön klaszterbe szervezne. Így egy az egyben nem használható a módszer ezen alkalmazás esetében. Nyilván bizonyos utófeldolgozással ezt az annotációs eljárást is be lehetne vetni, de az eljárás kevésbé van erre az alkalmazásra rá szabva, mint az általunk javasolt módszer.

6.2. Adatgyűjtő

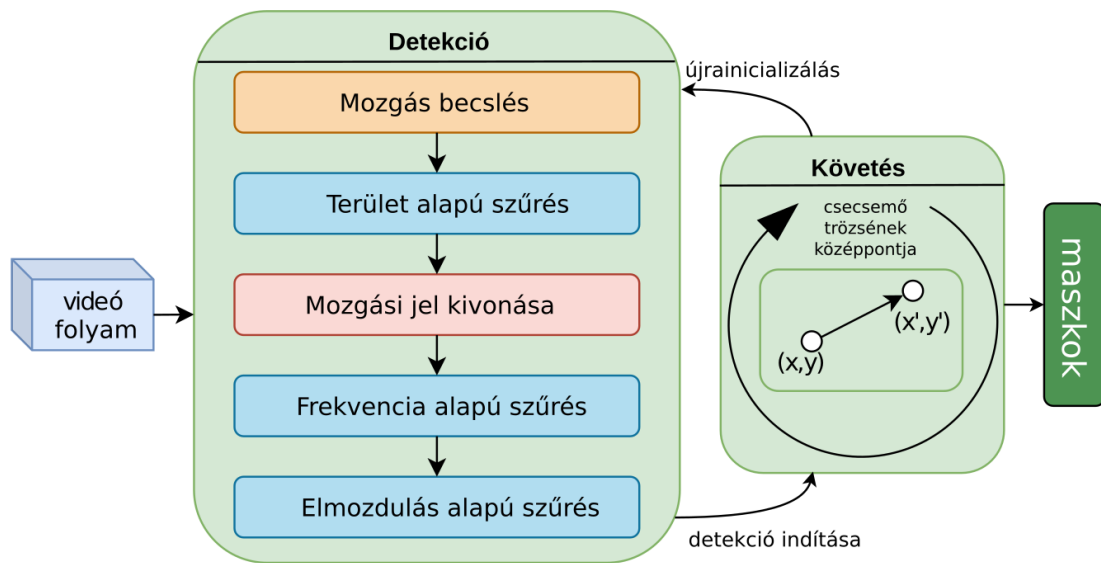
Az adatgyűjtéshez ugyanazt a rendszert alkalmaztuk, melyet a légzés monitorozás esetében is alkalmaztunk. Az elvárt kimenetek mentésére itt nincs szükség, ugyanis egy automatikus annotálóról van szó, amelynek pont az a lényege, hogy annotációt generál. Ettől függetlenül később alkalmaztunk manuálisan előállított annotációt is összehasonlításuképpen, amelyeket a LabelMe nyílt forráskódú annotáló segítségével készítettünk el (<https://github.com/wkentaro/labelme>). Itt a szemantikus szegmentáció példa mutatja be, hogyan kell előállítani a VOC típusú adathalmazt, amelyet mi is alkalmaztunk.

6.3. A javasolt algoritmus leírása

Mint már említettem a probléma megoldására egy mozgási kiterjedés és mozgási frekvencia alapján dolgozó algoritmust javaslok. Ez az algoritmus alapvetően két fő részre osztható:

- Detekciós modul
- Követési modul

A két modult és az együttes működésüket a 6.2. ábra foglalja össze. Az algoritmus működése azon a feltételezésen alapszik, hogy ha nincs más interakció és a bejövő kameraképeken csak a csecsemő látszik, akkor a képen lévő legnagyobb kiterjedésű mozgás a hasi terület körül fog látszani, amennyiben a mozgás frekvenciája a légzés fiziológiai tartományába esik (20 és 120 LPP közé). Az ilyen területeket nem kell folyamatosan detektálni. Egy sikeres detekció után megtalált területet követhetjük egyszerű „tracking” eljárások segítségével.



6.2. Ábra: Az automatikus annotációt végző algoritmus (ALA – Automatic Labelling Algorithm) összefoglaló ábrája. Az algoritmus bemenete a csecsemőről kamerával felvett képek sorozata. Kimenete pedig bináris képek (maszkok) halmaza, amelyek ugyanakkorák, mint a bementi képek és ahol az R-ROI, vagyis a légzőmozgás helye maszkolva van.

Az algoritmus bonyolultabb része a detekciós modul, amely a képek sorozatának fix hosszú vektorát kapja meg egy csúszó ablakban és a detektált terület középpontjával tér vissza. A detekció 4 fő lépésből áll:

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

6.3.1. Mozgás becslés

Ennek a lépésnek a célja egy mozgási kép előállítása, amelynek segítségével meg tudjuk adni, hogy melyik pixelek körül milyen intenzitású mozgás történt az adott t pillanathoz tartozó képkocka és az előző $t - 1$ időpillanathoz tartozó képkocka között. Az ilyen mozgási képek létrehozásához számos algoritmus létezik, mint pl. a korábban már bemutatott sűrű optikai áramlás, az úgynevezett "block-matching" algoritmusok ([24]), vagy a "DeepFlow" ([119]), amelyet Yue Sun et al. alkalmazott légzési mozgás becslésére [107]. 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 differenciakép 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 (6.1) egyenletben). A négyzetreemelés felerősíti az erősebb mozgásokat és gyengíti a zajokat. Tehát a jel-zaj viszony javításához szükséges. Ezen a módon egy egysatornás elforgatás invariáns mozgási képet kapunk.

$$\mathbf{D}(x, y, t) = (\mathbf{I}(x, y, t) - \mathbf{I}(x, y, t - 1))^2, \quad (6.1)$$

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

6.3.2. 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 [13]. A spagetti pixel klaszterek listáját adja vissza. Egy klaszter olyan pixeleket tartalmaz, amelyek egymással összeköttetésben vannak. A spagetti egy olyan képpel tér vissza, ahol minden pixel a hozzá tartozó klaszter sorszámaival lesz egyenlő. A háttér is egy külön klaszter, amivel nem foglalkozunk. Ezután kiválasztjuk a héttértől eltérő legnagyobb összefüggő komponens (lásd a (6.2). egyenletet).

$$g(x, y) = \begin{cases} 1, & \text{if } l(x, y) = \phi \\ 0, & \text{otherwise} \end{cases} \quad (6.2)$$

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

Ily módon kiválasztottuk a legnagyobb összefüggő komponens (g), amely esetében a képen csak 0-ák és 1-esek szerepelnek. Ezután megnézzük, hogy az LR 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 beleesik a megengedett sávba, akkor az algoritmus továbbléphet a következő lépésre. Ellenkező esetben az algoritmus visszatér az első lépéshez, amely a mozgásbecslés alkalmazása a következő bejövő képre.

6.3.3. 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 frekvenciája. Ezután tehát az időbeli mintázatát vizsgáltuk a mozgásnak. Ehhez azonban először időbeli egydimenziós jellé kellett alakítanunk a mozgási képek sorozatát.

Mint már sokszor korábban, most is egy csúszó ablakot alkalmaztunk, mint amilyen a 5.12. ábrán látható, csak ebben az esetben képkockák helyett a nyers mozgási képek voltak hozzáfűzve az ablakhoz, amely 300 mozgási képkockát tartalmazott, amelyeket a 6.1. egyenlet szerint számítottunk ki, majd sorfolytonosan kiolvastva az értékeket egydimenziós vektorokká alakítottuk, amelyeket mintáknak nevezünk és amelyeknek hossza $H=N \times M$, amennyiben N a mozgási képek horizontális M pedig vertikális kiterjedése. Így tehát a csúszó ablakban ezúttal nem 2D mozgási képek, hanem egydimenzióssá transzformált mozgási képek szerepeltek, amely egy $(300 \times H)$ alakú mátrix. Ezután alkalmazhatjuk erre a mátrixra a "főkomponens analízis - Principal Component Analysis (PCA)" eljárást [112], amely a sajátvektorok terébe képezi le a mintáinkat. Ezekből az első komponenst 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 jelkinyerés 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 [29], 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 egydimenziós 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:

$$f = \mathbf{b}[\max_f \{FFT(\mathbf{s}_M)\}], \quad (6.3)$$

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 továbbmehet a következő lépésre. Ellenkező esetben az algoritmus visszatér az első lépéshez, amely a mozgásbecslés alkalmazása a következő bejövő képre.

6.3.4. Elmozdulás alapú szűrés

Az előzőleg bemutatott szűrések már alkalmasak lehetnek a légzési mozgást tartalmazó régió detektálásához. Bizonyos esetekben azonban 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. Meg kell jegyezni, hogy ilyen az esetek nagyon kis százalékában fordulnak elő, viszont mi robusztus algoritmust akartunk tervezni, amely szigorú feltételeket támaszt a lélegzőmozgáshoz köthető régió megtalálásához. Nekünk az a jó, ha kifejezetten csak légzőmozgást végző területet találunk meg. Mert abból indulunk ki, hogy a videófelvevételeken van elég képkocka de a rossz annotáció nagyon megnehezítheti a későbbi neurális hálózat tanítást. Így jobb elveszteni néhány képet, mint rossz annotációt generálni. Ennek alapján a fentebb leírt ritka esetre is fel kellett készülni. Épp ezért vezettük

be az elmozdulás alapú szűrést, amely a legnagyobb összefüggő terület geometriai középpontjának mozgását követte figyelemmel és ha hirtelen túl nagy vagy 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:

$$\forall \mathbf{c}_i \in \mathbf{c}, d(\mathbf{c}_i, \mathbf{c}_0) < Th_{st} \quad (6.4)$$

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 $\mathbf{c}_i \in \mathbb{R}^2$.

6.3.5. Követési modul

A sikeres detekció után követést ("tracking"-et) alkalmaztunk, ugyanis nem volt szükség rá, hogy a detekciót minden esetben lefuttassuk, csak abban az esetben, ha a követett régiót valamilyen oknál fogva elvesztettük, pl. ha valami kitakarta a képen. A követésre számos "tracking" eljárást alkalmazhattunk volna, mint amilyenek az OpenCV csomag által kínált BOOSTING Tracker, MIL Tracker, KCF Tracker stb. algoritmusok. Mi azonban a ritka optikai áramlást alkalmaztuk a kiválasztott régió geometriai középpontjának követésére. Ez tűnt a legegyszerűbb megoldásnak, valamint nehezebb volt benyúló objektumok vagy végtagmozgás következtében elveszteni a követett pontot. Ha ugyanis a zavaró mozgások nem keresztezték a követett pont útját, akkor ezek nem tudták megzavarni a követést. A sűrű optikai áramlás kiválasztott jellemző pontokat követ, mint ahogyan azt a 5. fejezetben már kifejtettük. Esetünkben egy darab követendő pont lesz, amely, mint azt fentebb említettem a kiválasztott hasi régió geometriai középpontja. A követés során az annotációs maszkokat úgy generáljuk, hogy a követett pont köré egy R sugarú kitöltött kört rajzolunk fel. A sugarat az alapján számoljuk ki, hogy a csecsemőtől milyen távol található a kamera. Természetesen a csecsemő méretétől is függ a megfelelő

sugár kiválasztása, így valójában az igazán megfelelő sugár beállítását egy operátornak érdemes elvégezni az algoritmus indításakor. A maszkolt képek ezután mentésre kerülnek, úgy, hogy a fájlok nevei tartalmazzák az aktuális bejövő képkocka nevét is. És ezen felül a maszk szóval egészülnek ki. Így könnyen megválaszolható az a kérdés, hogy a szegmentációs maszk melyik képkockához tartozik. Ez a folyamat addig zajlik amíg sikeresen követni tudjuk a kiválasztott középpontot vagy amíg videót nem vált az algoritmus. Ezen esetekben újra meghívjuk a detekciót, hogy újabb követhető pontot találjon.

6.4. Értékelés

Ebben az alfejezetben a fentebb leírt automatikusan annotáló algoritmusunk értékelését mutatom be. Az automatikus annotáció értékelése nem egyszerű feladat, ugyanis a generált maszkokat össze kellene hasonlítani már létező hasonló maszkokkal, amelyek nincsenek a birtokunkban. Éppen maszkokat gyártunk automatikusan mesterséges intelligencia alapú algoritmus tanításához. Mindenekelőtt tehát megvizsgáltuk szabad szemmel, hogy a generált maszkok hol találhatóak és megállapítottuk, hogy mindegyik a csecsemőn, a hasi régió körül helyezkedik el. Ezen felül arra van lehetőség, hogy pixel szinten összehasonlítsuk a generált maszkokat az általunk manuálisan önkényes módon szegmentált maszkokkal.

A másik, talán még fontosabb szempont az, hogy a mesterséges intelligencia alapú szegmentáló betanítására mennyire alkalmas ez az automatikusan generált adathalmaz. Ezért az algoritmusunkat két szinten értékeltük ki. Egyrészt megvizsgáltuk, hogy a generált maszkok hogyan viszonyulnak az általunk manuálisan és önkényesen bejelölt elvárt maszkokhoz pixel szinten. Másrészt megvizsgáltuk, hogyha a generált adathalmazt egy U-Net betanítására alkalmazzuk, akkor az U-Net által szegmentált terület mennyire alkalmas a légzés kinyerésére. Azért, hogy jobban megértsük a továbbiakban bemutatott eredményeket néhány dolgot definiálnunk kell. Először is az értékelésnél használt adathalmazokat a következő nevezéktannal látjuk el:

- D_{man} : Egy olyan adathalmaz, ahol a maszk generálás manuálisan történt emberi annotálók által.

- D_{ala} : Egy olyan adathalmaz, ahol a maszk generálás a fentebb bemutatott automatikusan annotáló algoritmus segítségével történt.

Továbbá mivel az U-Net több típusát is betanítottuk az automatikusan generált halmazon, és ezek eredményeit megvizsgáltuk, ezért (a jobb megkülönböztethetőség érdekében) a különböző típusokat a következő elnevezésekkel láttuk el:

- Hagyományos U-Net architektúra, amely a D_{man} adathalmaz segítségével lett betanítva. ($U-Net_0$)
- Hagyományos U-Net architektúra, amely ez esetben a D_{ala} adathalmaz segítségével lett betanítva. ($U-Net_1$)
- Végül egy $U-Net++$ [127] architektúra, amely a D_{ala} adathalmazon lett betanítva. ($U-Net++$)

6.4.1. Pixel szintű értékelés

Az automatikusan generált annotációk értékeléséhez létrehoztunk egy mérőszámot, amely azt méri, hogy ha betanítunk egy U-Net architektúrát az automatikusan generált adathalmazon, akkor a egy független teszhalmaz képeire adott szegmentációk hogyan viszonyulnak az ezen a független adathalmazon kapott elvárt annotációkhoz.

$$S_{geo}(\mathbf{M}_U, \mathbf{M}_A) = \frac{\text{number of pixels in } (\mathbf{M}_U \cap \mathbf{M}_A)}{\text{number of pixels in } (\mathbf{M}_A)}, \quad (6.5)$$

ahol $\mathbf{M}_U$ az U-Net által eredményként adott annotáció, amely egy bináris maszk. Míg $\mathbf{M}_A$ szintén egy bináris maszk, amelyet az automatikusan annotáló algoritmus generált, vagy az annotálónk, manuálisan szoftver segítségével. A 6.3. ábrán láthatjuk egy példán, hogy miként viszonyulnak egymáshoz az összehasonlított annotációk. S_{geo} azt mutatja meg, hogy az átfedés (kék pixelek a 6.3. ábrán) hogyan aránylik az elvárt maszkkal (piros pixelek a 6.3. ábrán). Vagyis másképpen megfogalmazva azt, hogy az U-Net-ből kapott maszk hányad része fed át az elvárt maszkkal.

A fentebb definiált mérőszámot alkalmazva numerikusan értékeltük a kapott annotációkat. Először beállítottunk egy küszöbértéket (0.5), és azt vizsgáltuk meg,

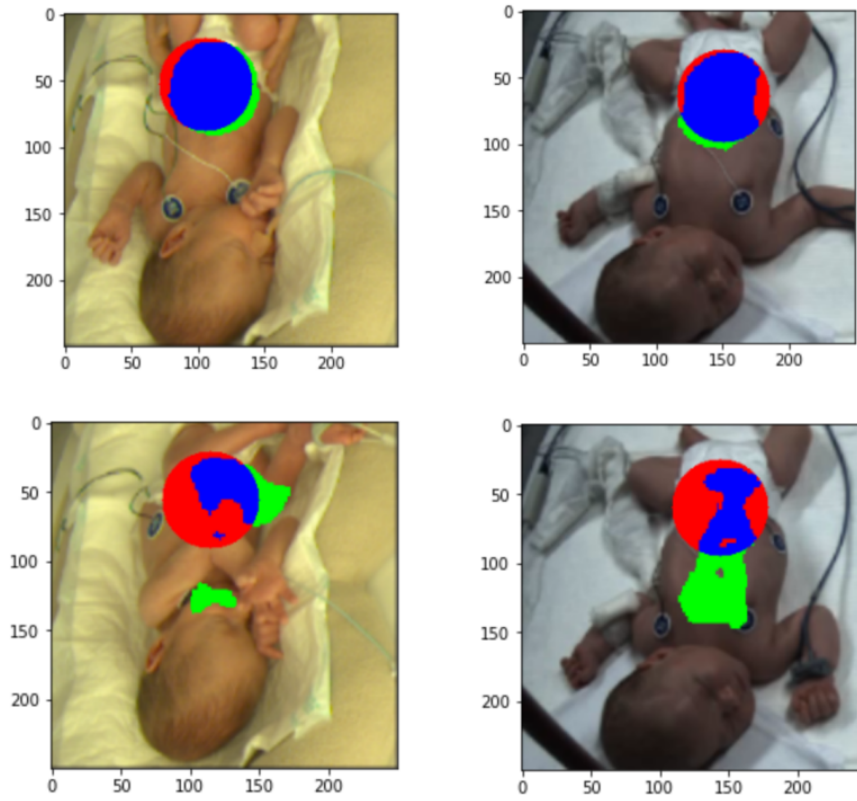
hogy az adathalmaz képeinek hány százalékánál lesz magasabb az S_{geo} mint 50%. Vagyis hány százalék fed át jobban az 50%-nál az elvárt maszkkal. Ebben az esetben 96%-os egyezést kaptunk. A laikus szemlélőnek ez a 0.5-ös küszöbérték kevésnek tűnhet és jobban megbízhat a másik típusú mérőszámunkban, amely az átlagos S_{geo} a kapott és az elvárt maszkok között (ami 78%-os egyezést mutat). Ez azonban elharmarkodott döntés lehet, ugyanis azt is megvizsgáltuk, milyen eredményt kapunk, ha a küszöbértéket 0.9-re állítjuk, és ebben az esetben sem kaptunk 93%-os eredménynél kevesebbet. Továbbá kételkedésünk azon a feltételezésen alapulhat, hogy az elvárt maszkok helyesen mutatják pixelszinten a légzés helyét. Ez azonban nem igaz. A légzési mozgás régiója a kapott kör alakú elvárt régiókon kívül is helyesen ismerhető fel és ott is helyes jel mérhető, ha a csecsemő testén helyezkedik el. Rossz annotációt csak a csecsemőn kívüli részekben vagy végtagokon kaphatunk. Hogy ez alapján is végezzünk numerikus mérést ezért manuálisan bejelöltük a csecsemők teljes torzóját. Vagyis az összes olyan pixelt, amely helyes lehet. Az ezeken a helyeken mért S_{geo} pedig 98% feletti eredményt adott.

6.1. Táblázat: Az U-Net alapú ROI generáló ($U-Net_1$) pixel szintű értékelése, ahol a hálózat a D_{ala} adathalmazon lett betanítva. Az első két sorban képkocka alapú statisztikák (jó képkockák / összes képkocka) láthatóak. A 3. sorban a teszhalmazra kapott átlagos S_{geo} érték látható. A negyedik sorban szintén az S_{geo} értékek láthatóak, viszont itt olyan manuális elvárt kimenetekkel (D_{man}) lettek összehasonlítva a kimenetek, ahol a csecsemők teljes torzója maszkolva volt.

Az alkalmazott U-Net alapú ROI detektor	Elért Teljesítmény
$S_{geo} > 0.5$ az ALA által generált maszkokon	96%
$S_{geo} > 0.9$ az ALA által generált maszkokon	93.5%
Avg S_{geo} az ALA által generált maszkokon	0.787
Avg S_{geo} manuálisan készített (teljes torzó) maszkokon	0.987

6.4.2. Légzési ráta becslő teljesítményének értékelése

Alapvető célunk az volt az ALA algoritmussal, hogy az általa generált adathalmazzal mesterséges intelligencia alapú ROI detektort legyünk képesek betanítani, amely ezután megfelelően képes szegmentálni a bejövő képeken azt a területet, ahol a képen lévő csecsemők légzése leginkább kinyerhető.



6.3. Ábra: Ezen az ábrán egy olyan ROI detektor eredményei láthatóak, ahol a detekciót egy $U-Net_1$ hajtotta végre, amely a D_{ala} adathalmazon lett betanítva. A ($U-Net_1$ által adott) kimeneteket zöld színnel ábrázoljuk, míg az (az ALA által generált) elvárt maszkokat piros színnel. (A kék szín pedig a metszetüket mutatja.) Az első sorban láthatunk néhány jó példát, ahol a kapott detekció több mint 50%-a (kék pixelekkkel jelölve) átfedésben van az elvárt maszkokkal (a képek 96%-án), míg a második sorban néhány rossz példát láthatunk, ahol az átfedés 50%-nál kisebb.

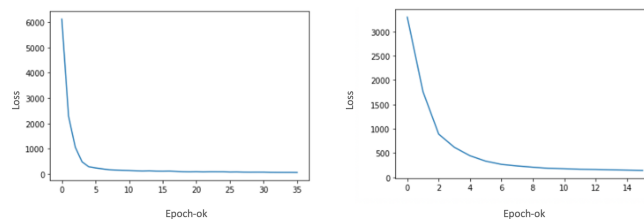
Ezen cél teljesülésének értékelésére a legjobb mód nem is a kapott maszkok pixel szintű értékelése, hanem annak vizsgálata, hogy a D_{ala} adathalmazon tanított ROI detektorok ($U-Net_0$ és $U-Net_1$) által szolgáltatott területeken mért légzési ráták minősége (a referenciákhoz hasonlítva) hogyan viszonyul a D_{man} adathalmazon tanított ROI detektor által talált területen kinyerhető légzési rátákhoz (szintén a referenciárátkhoz hasonlítva). Nyilvánvalóan az lenne a pozitív eredmény számunkra, ha az előbbi esetben kapott légzési ráták legalább annyira hasonlítanának a referenciára, mint az utóbbi esetben kapott légzési ráták.

A ROI detektor tanítása egy AMD EPYC processzorral (2.25 Ghz) és 12 Gi-gabájt RAM-al rendelkező rendszeren történt. A tanító script-eknek és a neurális hálózatnak az implementálásához a PyTorch keretrendszert alkalmaztuk. Az alkal-

mazott "optimizer" az "RMSprop" volt, míg az alkalmazott "loss" függvény pedig a "BCEWithLogitsLoss()". Néhány fontosabb példa a beállított hiperparaméterek közül:

- learning rate: $1e-5$
- momentum: 0.999
- weight_decay: $1e-8$
- gradient_clipping: 1.0
- img_scale: 0.5
- batch_size: 32

A konvergencia jellegét a 6.4 ábrán mutatom be egy-két jellemző példa ábrázolásával.



6.4. Ábra: A konvergencia jellegének bemutatása egy-két jellemző példával. A loss értékek epoch-onként lettek ábrázolva. Amint láthatjuk a loss szépen lecsökkent.

Ha megvizsgáljuk a különböző ROI detektorok által szolgáltatott területeken kapott légzési rátákat a referenciához viszonyítva (6.2. táblázat), akkor láthatjuk, hogy a számok a vártaknak megfelelően alakulnak, ugyanis az MAE nem haladja meg a 2.0-át. A szakirodalmat áttekintve több esetben ekörüli eredményeket értek el légzésfigyelő algoritmusokkal ([65],[15]). A légzési rátának a képek sorozatából történő kinyeréséhez a már korábban bemutatott légzésmonitorozó eljárást alkalmaztuk, amelynek összefoglalása az 5.10. ábrán látható. Ebben a légzésfigyelőben az 1. lépés a "ROI detekció" U-Net segítségével. Az értékelés során pedig az adott ROI detektor értékeléséhez a légzésfigyelő monitor U-Net alapú detektorát az adott adathalmazon betanított értékelendő ROI detektorra cseréltük. Ezzel légzést

nyertünk ki, mindegyik esetben ugyanabból a független teszhalmazból. Amint azt láthatjuk a 6.2. táblázatban, a generált adathalmazon tanított ROI detektorok által szolgáltatott területen kinyert légzési ráták nemcsak elérik a manuális adathalmazon tanított ROI detektorokkal kapott légzési ráták minőségét, de meg is haladják azt. Az első sorban lévő $U-Net_0$ -val olyan légzési rátákat kaptunk, amelyek csak 1.762 RPM négyzetes hibával illeszkednek a referenciára, míg az $U-Net_1$ esetében a négyzetes hiba már csak 1.507 RPM. Ha az U-Net-et egy újabb típusú reziduális konvolúciós rétegekkel ellátott verzióra ([127]) cseréljük, akkor pedig még nagyobb egyezést tapasztalhatunk a kapott légzési rátáink és a referencia ráták között (1.348 RPM). Ha az abszolút átlagos hibát vizsgáljuk, akkor is hasonló következtetésre juthatunk.

6.2. Táblázat: Egy olyan U-Net alapú ROI-detektorral ellátott légzési ráta becselő algoritmus értékelése, amelyben a neurális hálózat a D_{ala} generált adathalmazon lett betanítva.

Algoritmus	MAE	RMSE
ROI detektor nélküli eredmény	1.488 RPM	1.808 RPM
$U-Net_0$ detektor, a D_{man} adathalmazon tanítva	1.348 RPM	1.762 RPM
$U-Net_1$ detektor, a D_{ala} adathalmazon tanítva	1.223 RPM	1.507 RPM
$U-Net++$ detektor, a D_{ala} adathalmazon tanítva	1.094 RPM	1.348 RPM

Amint említettük, az $U-Net_1$ és $U-Net++$ ROI detektorok teljesítménye jelentősen meghaladják a manuálisan annotált esetben kapott teljesítményt, és minden vizsgált neurális architektúra jobb teljesítményt nyújt, mint amikor nem használunk ROI detektort. Fontos megjegyezni, hogy az értékeléshez a független teszt sorozat videófelvevételeiből egy tipikus folyamatos időszakot választottunk ki, ahol nem figyelhető meg orvosi vagy egyéb beavatkozás, csak a légzés és rövid végtagmozgások láthatóak a képen.

Ezen felül ennél az algoritmusnál is megvizsgáltuk, hogy a futási sebességet egy AMD EPYC @ 2.25Ghz processzorral és 12 Gigabájt RAM-al rendelkező PC-n. A javasolt algoritmusunk sebessége itt 0.0645 ± 0.02 s volt. Azaz kb. 64 ms.

6.4.3. A javasolt algoritmus összehasonlítása a szakirodalomban fellelhető eljárásokkal

Minden erőfeszitésem ellenére olyan publikációt, amely csecsemőkről készült képek szemantikus szegmentációjához mutat be automatikus annotációs eljárást, nem találtam. A legtöbb ezzel foglalkozó cikkben olyan módszerekről írnak, ahol hagyományos módon állították elő az annotációt vagy "transfer learning"-et alkalmaztak. Magának az automatikus annotációnak a témaköre azonban nagy irodalommal rendelkezik.

Ahogy a "Kapcsolódó szakirodalom" alfejezetben bemutattam, a manapság legnépszerűbb automatikus annotálási eljárások a generatív algoritmus alapú automatikus annotáló eljárások. Erre több példát is találunk a irodalomban, ahogyan ezt a "Kapcsolódó szakirodalom" alfejezetben bemutattam ([113], [122], [83]).

Ezek esetében egy generatív (általában Generative Adversarial Network alapú) eljárás van alkalmazva, amely képes mintákat generálni egy adott adathalmazból. Van, hogy a bemenet ehhez random zaj, vagy (domain transzformációt végrehajtó) GAN-ok (pl CycleGAN) esetén egy másik típusú adathalmaz mintái. Mint korábban említettem, a RailGAN esetén a normál sínekről készült képek domain-jéből viszik át a RailGAN a hibás sínek képeinek domain-jébe. A [113] publikációban leírt CycleGAN pedig a hibamentes felületek domain-jéből a hibás felületek domain-jébe. Ez domain transzformáció. Majd a valós és a hibás képek egymásból való különbsége szolgáltatja a szegmentációs annotációt. Míg a StyleGAN2 esetében random zaj a bemenet, és emberi vagy állati arcokat generáltak vele. Ezután pedig a generátor alacsonyabb szintű rétegeiből kapott "feature map"-eket felhasználva klaszterezi a kép pixeleit. Így annotálva be különböző régiókat.

Ezek az eljárások hatékonyak és jól használhatóak több alkalmazásban. Beszédesebb azonban, hogy általában viszonylag egyszerűbb vagy jól ismert adathalmazok esetében alkalmazzák ezeket az eljárásokat, (mint sínek vagy arcok). Az inkubátorokban lévő csecsemőkről készült képek komplex szituációkat mutatnak, amelyek esetében a az előbbiekhöz hasonló generatív algoritmusoknak jóval nehezebb dolga van. Itt a keresett terület, ahol jól megfigyelhető a légzőmozgás (mint

a pelenka és a bőrfelület találkozása), nem annyira jól elkülönülő régió, hogy az könnyen kalszterelhető lenne bizonyos jellemzők alapján. Ezért úgy gondolom, ezek a módszerek erre az alkalmazásra nehezen lennének egy az egyben adaptálhatók. Daniil et al. munkájának implementációja megtalálható publikus könyvtárban (https://github.com/warmspringwinds/segmentation_in_style/tree/master), amelynek linkje a cikkben is szerepel. Itt láthatjuk, hogy az ezen algoritmus fejlesztői előtanított modelleket használtak. Vannak olyan generatív modellek amelyek emberi arcokat vagy kutyákat ábrázoló képeket tudnak generálni. Ahhoz azonban, hogy a mi alkalmazásunknál is használható legyen a szóbanforgó eljárás, újra kell tanítanunk a StyleGAN2 modellt a csecsemőkről készült képeket tartalmazó adathalmazra, ami egy hosszabb folyamat, és nem garantálható, hogy a modell újratanítása után a klaszterezés elkülönítené a megfelelő légzéshez tartozó régiót. Ráadásul eléggé diverzek a képek. Különböző szögekből készültek a felvételek, háton és hason fekvő babákról egyaránt.

A generatív modelleket alkalmazó módszerek tehát először megtanulják generálni az adott képeket és utána a generált képekhez készítik el az annotciót, különböző módszereket alkalmazva. A mi általunk javasolt módszer ezzel szemben a légzési mozgás kiterjedése és frekvenciája alapján találja meg a keresett területet. Vagyis olyan attribútumait használja a légzőmozgásnak, amelyek alapján leginkább elkülöníthető, felismerhető az. Továbbá az általunk javasolt módszer a csecsemőkről készült folyamatos videófelveteleket feldolgozva annotál, figyelembe véve a képek egymásutániságát. Miközben a generatív eljárás alapú annotálók képek halmazából dolgoznak, figyelmen kívül hagyva a képek sorrendjét.

7. Fejezet

Alvási és aktivitási (viselkedési) fázisok monitorozása

7.1. Motiváció a csecsemők viselkedési állapotainak klasszifikációjához

Korábban bemutattuk a mozgási aktivitást, légzést és orvosi vagy egyéb beavatkozásokat becslő algoritmusokat. Ezen felül a kutatóintézetben, ahol dolgozom (a SZTAKI-ban) foglalkoztunk a csecsemők pulzusának mérésével is. Kollégáim kifejlesztettek egy érintésmentes pulzus figyelő algoritmust, amely a PhysNet nevű neurális hálózaton alapult ([124]). Én a pulzusfigyelő rendszer fejlesztésében kevésbé, inkább az ehhez szükséges pulzus referencia gyűjtésében vettem részt, valamint az algoritmusuk tesztelésében. Kollégáim bizonyították azonban, hogy klinikai környezetben főleg nyitott inkubátorok esetében lehetséges a csecsemők pulzusának érintésmentes monitorozása (és erről publikáltunk is [J1]). Fentebb pedig már bemutattam, hogy miként becsülhető a csecsemők mozgási aktivitása és a számított mozgási képekből, hogyan lehet légzési rátát becsülni. Felmerült azonban a kérdés, hogy ha kamera vagy vezetékes alapú módszerekkel megbízhatóan tudjuk mérni a légzést, a pulzust és a mozgási aktivitást, akkor ezen mutatókat elemezve milyen további információt szolgáltatathatunk az orvosok és az ápolást végző személyek számára? A kérdésről konzultáltunk orvosi tanácsadóinkkal is és egyértelművé vált,

hogy a légzésnek és a mozgásnak a becslése jó kiindulóalap lehet a csecsemők alvási és viselkedési fázisainak elkülönítéséhez, amelynek automatizálásában orvosi tanácsadóink is rendkívül érdekeltek. Az ő igényeiken felül az irodalomkutatásunk is azt támasztotta alá, hogy a csecsemők alvási fázisainak klasszifikációja és alvásmennyiségük becslése a koraszülött klinikák számára hasznos és elvárt dolog ([11,61,66,120]). Az alvás mennyisége kritikus szerepet játszik az újszülöttek megfelelő fejlődésében, mivel az alvás és az alvási ciklusok és a megfelelő mennyiségű alvás biztosítása elengedhetetlenek az érzékszervek fejlődéséhez, a korai idegrendszer, a tanulás, a memória kialakulásához és az agyi plaszticitás megőrzéséhez [23], [34]. Az alváshiány vagy egyéb alvási rendellenességek gátolhatják az elsődleges szenzoros rendszerek neurális áramköreinek kialakulását, és negatív hatással lehetnek a babák agyi fejlődésére[35], [36]. Az alvási és az aktív fázisok (vagy más néven viselkedési állapotok) megfigyelése és az alvásidő becslése tehát fontos a koraszülött klinikákon, ahol ez sok esetben (még automatizálás nélkül) közvetlenül az orvosi személyzet által történik.

A folyamatos megfigyelés és a megfelelő gondozási ütemterv kialakítása létfontosságú a családcentrikus fejlődésbarát gondozásban. Azokban a kórházakban, ahol fejlődésbarát kezelés lehetséges, a kórházi személyzet folyamatosan figyelemmel kíséri a babák aktivitását, alvási állapotait. Az alvó és aktív fázisok klasszifikációjához az Újszülött Egyéni Fejlesztési Gondozási Program (NIDCAP) skáláját alkalmazzák. A NIDCAP skála szerint a következő állapotokat különítjük el:

1. csendes alvás fázis
2. aktív alvás fázis
3. átmeneti fázis
4. csendes ébrenlét fázis
5. aktív ébrenlét fázis
6. sírás fázis

A fentebb leírt fázisokat viselkedési állapotoknak nevezzük. Ez az osztályozás összhangban áll a FINE („Family Infant Neurodevelopmental Education”) oktató-

anyagban ajánlott osztályozási sémával, amelynek anyagait a NIDCAP szakemberei és oktatói készítették el londoni és hollandiai képzési központokban [118].

A csecsemők alvásának megfigyelésére a legelterjedtebb módszer a poliszomnográfia (PSG), amely a csecsemő testére erősített több érzékelő és elektróda használatát jelenti a különböző fiziológiai paraméterek megfigyelésére. Ezeknek az érzékelőknek a vezetékai azonban zavarhatják az orvosokat, a nővéreket és stresszelhetik a családot, a csecsemőt magát, sőt az érzékelők és elektródák akár fel is sérthetik a koraszülöttek vékony bőrét, ami fertőzésveszélyhez vezethet.

7.2. Kapcsolódó irodalom

Az elmúlt évtizedben számos kísérlet történt a csecsemők alvásának és aktivitásának monitorozására szolgáló nem invazív vagy minimálisan invazív megoldások kifejlesztésére [9], [68], [49], [98], [8], [60], [51], [39], [71], [6], [4]. Ezek közül érdemes kiemelni Muhammad Awais et al. és Xi Long et al. munkáját [9], [68], mivel ezekben videóalapon osztályozzák a csecsemők alvási állapotait. Muhammad Awais et al. az arcról készült videókon lévő arc kifejezések alapján osztályoz, míg Xi Long et al. munkája videóalapú aktigráfia alapján. Továbbá, (a NIDCAP skála alapú) kategorizálás tekintetében érdemes megvizsgálni S. Cabon et al. ([18]) valamint Peter W. Fuller et al. ([30]) munkáit. Az első esetben egy olyan tanulmányról van szó, amely nagyon hasonló a miénkhez, és bemutat egy új viselkedési állapot becslési módszert, amely a NIDCAP skála szerinti kategorizálást nemcsak kamera képadatok, hanem hangadatok felhasználásával is végzi el egy kisebb adathalmazon, mint amit mi használunk. Mi viszont a kamera képi adatok és a pulzus jel adatok kombinációját használjuk fel a NIDCAP állapotok kategorizálására, ami egy másik megközelítés. Mi a (PRV - Pulse Rate Variability) adatok használata mellett döntöttünk a hang adatok helyett, mert a mi adathalmazunk esetében azok alapértelmezetten rendelkezésre álltak a csecsemőknél, míg a mikrofon használata esetleg adatvédelmi problémát okozhatott volna. A ([30]) publikációban pedig Peter W. Fuller et al. a viselkedési állapotokat (amelyeket szem- és testmozgás videó felvételeinek felhasználásával becsültek meg) a koraszülött csecsemők poligráfiai módszerrel

meghatározott állapotaival hasonlítja össze.

7.3. Adatgyűjtő

A viselkedési állapotok osztályozásának esetén az adatgyűjtés hasonló a korábban leírtakhoz a bemeneteket tekintve. A viselkedési osztályozó képsorozatok alapján végzi a klasszifikációt. Így a képi adatokat felvehetjük a légzési monitorozásnál alkalmazott adatgyűjtő rendszerrel, ami tartalmaz egy színes kamerát, egy merevlemez, egy infravörös megvilágítót és tapadókorongokkal rögzíthető a zárt inkubátor tetejéhez. Az elvárt kimeneteket azonban máshogy kellett rögzíteni. A viselkedési állapotokat alapártelmezetten az orvosi személyzet rögzíti ha van rá erőforrás. Így ezen viselkedési állapot jegyzetek digitalizálásához létrehoztunk egy mobil applikációt, amellyel az orvosi személyzet rögzítheti az aktuális viselkedési állapotot. Ezt rendszert bővebben bemutatatom a 3. fejezetben.

7.4. A probléma definiálása

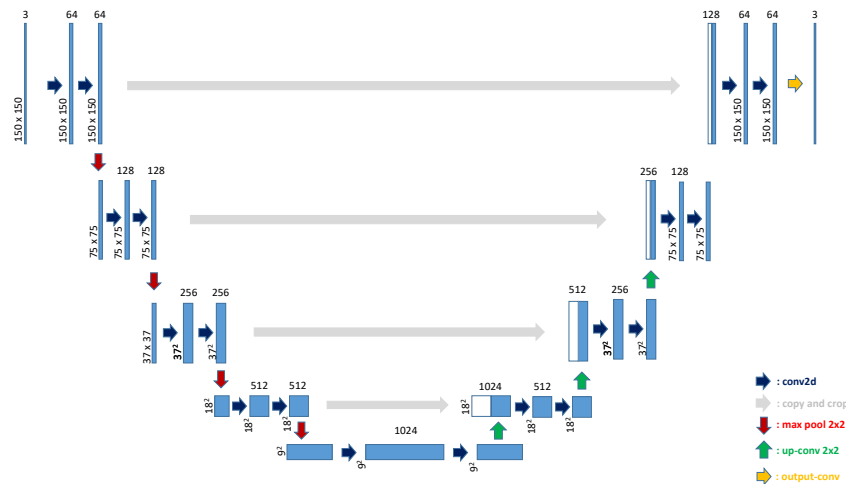
A megoldandó probléma tehát úgy foglалható össze, hogy az ajánlott algoritmusnak a NIDCAP skála szerinti viselkedési fázisait kell megállapítania csecsemőknek, vagyis a kimenet a felsorolt viselkedési fázisok egyike lesz egy adott időpillanatban. Abból indulunk ki, hogy az algoritmust koraszülött klinikákon alkalmazzák, azaz a kísérleti (setup) rendszer tehát az "Adatgyűjtő és Annotáció" fejezetben leírtaknak felel meg, attól eltekintve, hogy nem garantálhatjuk, hogy az elektródák minden esetben rögzítve vannak a csecsemők bőréhez. Az azonban bizonyos, hogy a pulzoximéter a csecsemőre van rögzítve, ugyanis a koraszülött csecsemők pulzusát a kórházakban folyamatosan mérik. Továbbá feltételezzük, hogy a kamerás megfigyelő és feldolgozó rendszerünk rögzítve van az inkubátorhoz. Ez azt jelenti, hogy az ajánlott algoritmusnak bemenetként rendelkezésére állnak a Philips IntelliVue MP20/MP50 pulzoximétere által biztosított numerikus pulzus értékek és hullámalakok, valamint a kamerából érkező képkockák és ezen bemenetek sorozata alapján kell meghatároznia az aktuális időpillanathoz tartozó NIDCAP skála szerinti viselkedési

fázist.

7.5. Az ajánlott algoritmus: viselkedési állapot becs- lő

A fentebb leírt probléma megoldására egy mély tanulás alapú rekurrens neurális hálózatot (Recurrent Neural Network - RNN), pontosabban ilyen hálózatok kaszkádját ajánljuk, amely bemenetként mind a videó alapon generált jellemzőket, mind pedig a pulzusszám változékonyság alapján generált jellemzőket megkapja és felhasználja. Kimenetként pedig viselkedési fázisok egyikével tér vissza. Pontosabban egy 0-kat és egy darab 1-est tartalmazó vektorral, amelynek annyi eleme van ahány kimeneti osztállyal rendelkezünk. Az adott időablakban lévő bemeneti időperiódusra az 1-es helyéhez tartozó fázist adja az algoritmus.

A javasolt algoritmus felhasználja videóalapú jellemzőket és a Philips IntelliVue MP20/MP50 monitor által szolgáltatott pulzusjelből kiszámolt jellemzőket is. Az ajánlott viselkedési állapot becslő videóalapú részének alapja, hogy a képi jellemzőket a képek különböző (heurisztikus információk alapján kiválasztott) területein nyerjük ki a már korábbi feladatokban jól bevált algoritmusok és statisztikai számítások alapján. Ezeket a régiókat individuális képkockákon ismerjük fel míg a kiválasztott régiókon a jellemzőket képek sorozatából nyerjük ki. (Az aktuális képből és az azt megelőző $W=200$ képből.) Mint említettem a ROI halmazt, amit használtunk heurisztikus megfontolások alapján válogattuk össze. A kiválasztási logikánk azon a feltételezésen alapszik, hogy a csecsemőkről készült képeknek vannak olyan területei, ahol a helyi mozgások különösen fontos szerepet játszanak az alvási és aktivitási fázisok meghatározásában. Az ajánlott algoritmusunk esetében 3 olyan régiót választottunk ki, amelyek fontos információkat hordoznak a baba viselkedési állapotáról. Ezen informatív régiók különböző színekkel kijelölve láthatók egy példa képkockán a 7.3 ábrán. Az első ilyen régió a csecsemő hasának régiója, mivel a has körüli mozgások általában a légzéssel együtt figyelhetők meg, ami egy élettani jel, amely fontos szerepet játszik az alvás-ébrenlét szakaszainak megkülönböztetésében. Ennek vagy az erős légzési mozgások középső részének észlelésére a U-Net (7.1. áb-



7.1. Ábra: Ezen az ábrán a U-Net architektúrája látható, amely szemantikus szegmentációra létrehozott neurális hálózat. A mi esetünkben a hasi régió és annak környezetének szegmentálását végzi, hogy ROI-t szolgáltatson a légzés kinyeréséhez.

ra) architektúrát használunk, amely a képen látható csecsemő hasát szegmentálja (7.3.B ábra). Ha megvan ez a szegmentálás, maszkoljuk az adott sugarú körbe tartozó környező területet, hogy rögzítsük a csecsemők végtagjainak aktivitását (7.3.C ábra). Erről a területről tehát a légzéstől független végtagmozgásokról kaphatunk információkat. Végül a csecsemő testétől távolabb eső terület (7.3.D ábra) segít elnyomni a külső eseményeket.

A ROI-halmazon történő jellemző kinyeréshez használt „intra-„ és „iter-frame” statisztikák és algoritmusok kiválasztásakor abból indultunk ki, hogy mit használtunk a korábban bemutatott "szituációanalízis" során. (Az ott használt statisztikákat a 5.8. ábra és a 5.1. táblázat foglalja össze.) A viselkedési állapotok klasszifikációjánál azonban nemcsak ugyanazokat a jellemzőket használtuk és más módon dolgoztuk fel őket. Az elsődleges különbség a feldolgozás módja, vagyis az, hogy míg a "szituációanalízis" a statisztikák segítségével skalárok halmazát állította elő addig a viselkedési állapot becslő egydimenziós jelek sorozatát állítja elő. Majd a NID-CAP skála szerinti állapotok klasszifikációjához egy GRU-kaszád dolgozza majd fel az egydimenziós jelek halmazát. A "szituációanalízis"-hoz hasonlóan a viselkedési állapot becslő is alkalmazza a "sűrű" optikai áramlás és a differenciakép kiszámítását végző algoritmusokat, amelyek két egymást követő képből állítanak elő mozgási

képeket.

Az „intra-frame” statisztikák azonban, mint a képkockák pixeleinek fényerejének átlaga viszont az individuális képkockákon számítandók ki.

Fontos videóalapú jellemző még az alvás megállapításában a légzési jel, amelyet optikai alapon az előző fejezetben leírt módszerrel számítottunk ki. A hosszú nyugodt légzés általánosságban az alvó fázisokra jellemző. Míg az éber fázisokban a nyugodt légzést sok esetben végtagmozgások szakítják meg valamint a légzési min-tázat és a frekvenciája is különböző lehet az alvásközbéli állapot jeleitől.

A számított jellemzők teljes listája a következő:

1. Képkocka pixel értékeinek átlaga.
2. Képkocka pixel értékeinek szórása.
3. Képkocka HSV transzformáltjának átlaga
4. Képkocka HSV transzformáltjának szórása
5. Sűrű optikai áramlásból kapott mozgási kép átlaga
6. Sűrű optikai áramlásból kapott mozgási kép szórása
7. Négyzetes differenciakép átlaga
8. Négyzetes differenciakép szórása
9. Légzés (optikai alapon számolva)

Ezt a 9 jellemzőt a ROI-halmaz minden elemére kiszámítjuk, leszámítva a légzési jelet, amelyet a csecsemő hasa környékén számítunk ki. Így összesen 25 videóalapú jellemzővel fogunk rendelkezni. A videó alapú jellemzők értékkészletéről elmondható, hogy bármely videóalapú jellemző f értékére $f \in \mathbb{R}^2$. Kivétel ez alól a légzési jel, mert annak értékkészlete a 20–120 LPP közötti tartományra van korlátozva, ugyanis ez a légzés fiziológiai tartománya.

A javasolt algoritmus ezen felül felhasznál olyan jellemzőket is, amelyeket a pulzusjelből számítottunk ki. 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. Ezeket a PRV mérőszámokat a [19] csúcsdetektor segítségével számítjuk a PPG jelből. A PRV-t többféle módszerrel is értékelhetjük a csúcsdetekció után [2]. A legkézenfekvőbb, ha az időtartománybeli metrikákat alkalmazzuk. Ha a csúcsokat detektáltuk, akkor kiszámíthatjuk a köztük lévő időintervallumok hosszát. (Ezek az NN ("normal to normal R-R peak") intervallumok és másodpercben fejezendők ki.) Az ilyen intervallumok sorozatából összetettebb metrikák számolhatók ki a különböző fázisok összehasonlításához. Az általunk felhasznált összetettebb jellemzők kiszámítását a 7.1 és 7.2 egyenletek írják le:

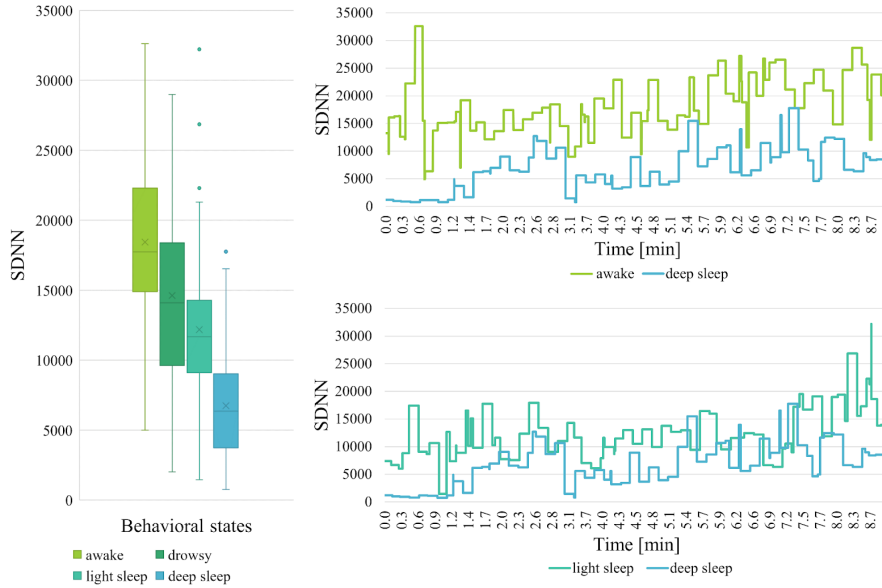
$$RMSSD = \sqrt{\frac{\sum_{i=1}^{N-1} (\mathbf{PP}_{i+1} - \mathbf{PP}_i)^2}{N-1}} \quad (7.1)$$

$$SDNN = \sqrt{\frac{\sum_{i=1}^N (\mathbf{PP}_i - \mathbf{mPP})^2}{N}} \quad (7.2)$$

ahol $\mathbf{PP}$ a két egymást követő impulzuscsúcs közötti időkülönbség, a $\mathbf{PP}_i$ az i -edik impulzuscsúcs párhoz tartozó időkülönbség, az $\mathbf{N}$ pedig az impulzuscsúcsok száma. Az $\mathbf{mPP}$ a PP intervallumok átlagát jelenti, az $\mathbf{SDNN}$ ([105]) 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.

Ezen felül kiszámításra kerül még egy frekvencia alapú pulzus aktivitás jellemző ($\mathbf{LFHF}$) szintén 5 perces csúszóablakon amely úgy képezhető, hogy a PPG jelen végrehatjunk két féle frekvencia szűrést. Egy magas frekvenciájú szűrést (0.154Hz - 0.4Hz) és egy alacsony frekvenciájú szűrést (0.04Hz - 0.15Hz). A kétféle jelnek ezután kiszámoljuk az energiaspektrumát. Az alacsonyabb frekvenciájú energia spektrum integrálját (lf) és a magasabb frekvenciájú spektrum integrálját (hf) is kiszámítjuk. Majd az $\mathbf{LFHF}$ jellemző értékét ezek aránya adja meg (lf / hf).

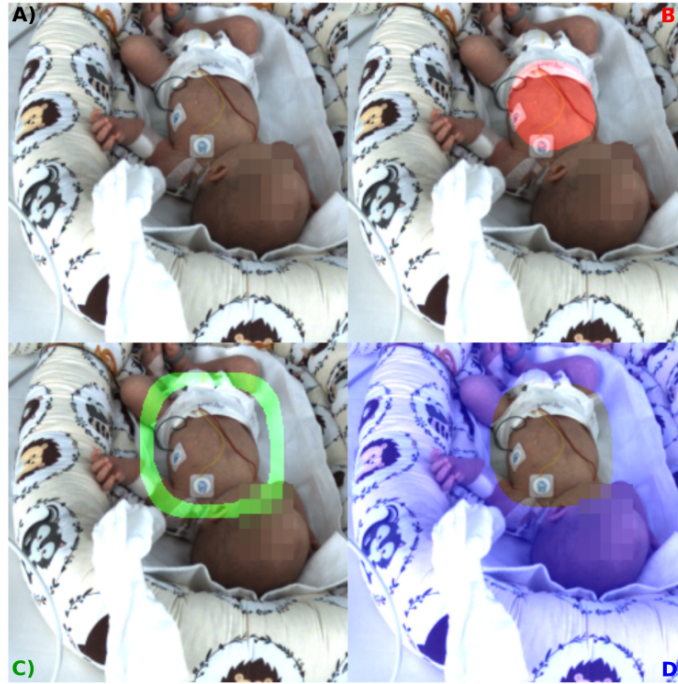
Értékkészletüket tekintve a PRV alapú jellemzők a következőképpen alakulak: $SDNN \in [32, 93]$, $RMSSD \in [19, 75]$, $LFHF \in [1.6, 11.6]$.



7.2. Ábra: A különböző viselkedési állapotokhoz tartozó SDNN metrikák doboz diagrammon ábrázolva. Jobb oldalt a kiszámított SDNN értékek időbeli görbéi látszanak különböző 8.7 perces reprezentatív szakaszokon kiszámítva és időtengely eltolás után egymásra téve éber, mély alvó és aktív alvó állapotok esetében. A különböző színekkel a különböző viselkedési fázisok vannak jelölve.

A számítás egy 5 perces csúszó ablakban történik a PRV-metrikák [100] áttekintésében leírtaknak megfelelően. A [86] közleményben leírt algoritmust használtuk a frekvenciatartomány-leírók kiszámításához. A 7.2 ábra szemlélteti, hogy az SDNN jelek miként viszonyulnak egymáshoz a különböző viselkedési állapotokban. Amint láthatjuk az ábrán, a mély alvó és éber állapotok jól megkülönböztethetők az SDNN jellemzőt figyelembe véve.

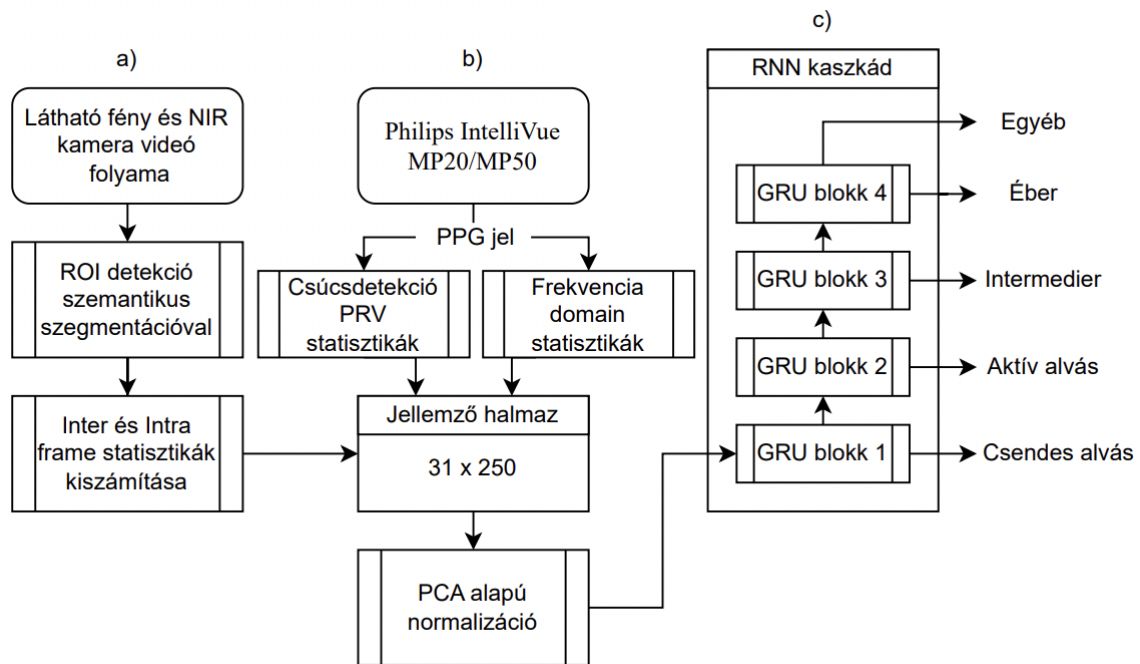
A fentebb leírt módon kapott PRV alapú jellemzőket hozzáfűzzük a videóalapú jellemzők halmazához. Rendelkezünk 25 féle videóalapú jellemzővel és 6 féle pulzus variabilitás alapú jellemzővel, mert az SDNN és a RMSSD két féle csúcsetektorral is kiszámításra került, valamint hozzáfűzésre kerül még az LFHF és maga a pulzus rátákat tartalmazó egydimenziós jel. Így egy 31×250 -es tenzort hozunk létre. Ez azt jelenti, hogy a készlet 31 sornyi 250 hosszú egydimenziós jelből áll. Ezt követi egy főkomponens-analízis (PCA) transzformációs lépés, ahol a transzformáció paramétereit előzőleg a teljes adatkészletre számítottuk ki. A PCA-t ezúttal nem használjuk dimenziócsökkentésre, tulajdonképpen egyfajta normalizációt hajt végre a bemeneten. A transzformáció eredménye ugyanolyan alakú. Továbbra is egy



7.3. Ábra: Ezen az ábrán különböző ROI-k láthatóak, ahol a helyi mozgások különböző információkat hordoznak a csecsemő viselkedési állapotairól. A) - a kamerából érkező bemeneti képkocka; B) - a hasi területet jelöli (amely a légzés kiszámításához kell); C) - egy gyűrű alakú terület a hastól adott távolságra (amely áthalad a végtagokon); D) - a (B) rész inverz területe.

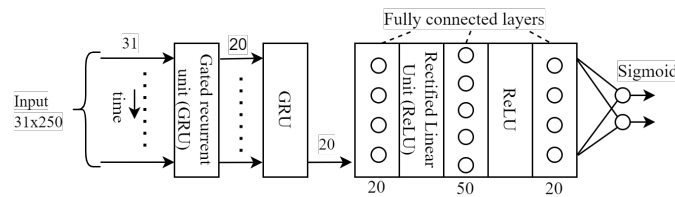
31x250-es tenzor lesz. A PCA transzformációval véget ér a bemenet előfeldolgozása. Az így kapott, normalizált egydimenziós jel halmaz lesz az RNN alapú klasszifikációs modul bemenete.

Az algoritmus utolsó lépése az előfeldolgozott bemenetnek a klasszifikációja a NIDCAP skála szerinti fázisok egyikébe. Klasszifikációs modulként ismétlődő neurális hálózatok (RNN-ek) kaszkádját használjuk, mivel ezt az architektúrát számos alkalmazásban sikeresen használták már különböző fajta idősorok osztályozására ([116], [126], [45]). Pontosabban a GRU (Gated Recurrent Unit) típusú hálózatot használjuk, mivel ez a típusú RNN (amely a hagyományos LSTM struktúrából kihagyja a cellaállapotot, és csak a rejtett állapotot használja) jobban teljesít a viszonylag hosszú szekvenciákat tartalmazó kisebb méretű adat készleteken [123]. Úgy találtuk, hogy kísérleteink alapján egy négyosztályú kimenettel rendelkező GRU (ahol a négy kimeneti osztály a négy viselkedési állapotnak felel meg) viszonylag gyenge eredményeket adott a pontosság tekintetében minden kimeneti osztály esetében. A teljesítmény fokozása érdekében a GRU osztályozókat (blokkokat) bináris



7.4. Ábra: A viselkedési állapot becselő algoritmus működését bemutató összefoglaló ábra. Itt a) rész mutatja a videóalapú jellemzők kiszámítását végző modulokat. A b) rész pedig a pulzus alapú jellemzők kiszámítását végző modulokat és a videóalapú jellemzőkkel való összefűzést. A c) mutatja a PCA alkalmazását. A d) pedig az aktuális viselkedési állapotot meghatározó klasszifikációs modult.

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. Az éber fázisokat összevontuk egy fázissá. Ezen fázisok megkülönböztetése ugyanis egyrészt nem létfontosságú a csecsemők alvásmennyiségének megbecsléséhez, másrészt kísérleteink szerint jóval nehezebbnek bizonyult, mint az alvó és átmeneti fázisok elkülönítése. A [C2] publikálásának idején ezért ezeket összevontan kezeltük és az éber szakaszok elkülönítését egy későbbi cikk témájául tűztük ki. Minden osztályozást végző blokkhoz tartozik egy kétrétegű GRU-blokk és egy háromrétegű "fully-connected" (FC) klasszifikációs neurális hálózat, ahol az első és a második FC réteget ReLU aktivációs függvények követik (7.5. ábra.). A kaszkádban minden fázisba való bináris klasszifikációt egy a fentebb leírt felépítéssel rendelkező GRU-blokk végzi. A kaszkád felépítés miatt minden bináris klasszifikációs blokk csak akkor fut le a bemeneten, ha az előző GRU-blokkok 0-val tértek vissza. Az



7.5. Ábra: Egyetlen GRU blokk szerkezete, amely egy blokk felépítését és egy ilyen blokkban lévő hálózat méretét mutatja.

adott időperiódushoz tartozó jelhalmaz csak akkor lesz az adott viselkedési fázisba klasszifikálva, ha a hozzá tartozó GRU-blokk tér vissza 1-gyel.

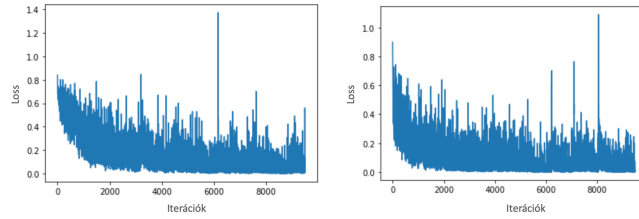
Az ajánlott algoritmus fentebb leírt részeit: az előfeldolgozást, a jellemzőhalmaz felépítését és a klasszifikációs modult a 7.4. ábra foglalja össze.

7.6. Az ajánlott algoritmus eredményei és azok értékelése

Az ajánlott algoritmust ki is értékeltük egy olyan teszhalmazon, amelyben lévő felvételeken olyan csecsemők szerepeltek, amelyek a tanítóhalmazban lévő felvételeken nem. Ebben az alfejezetben bemutatom az értékelés folyamatát és az elért eredményeket.

A tanítás maga itt egy AMD EPYC processzorral (2.25 Ghz) és 12 Gigabájt RAM-al rendelkező rendszeren történt. A PyTorch keretrendszert alkalmaztuk. Abban írodott maga a neurális hálózat és tanító és értékelő "script"-ek is. Az alkalmazott "optimizer" a nagyon népszerű "ADAM optimizer", az alkalmazott "loss" függvény pedig a jól ismert "CrossEntropy()", amelyet általában osztályozási feladatok esetén használnak. A hiperparaméterek optimalizációja a "Grid Search" eljárást alkalmazva történt. A konvergencia jellegét a 7.6 ábrán mutatom be egy-két jellemző példa ábázolásával.

Bináris vagy nem bináris klasszifikátorok értékelésére is ismerünk több jól bevált egymást kiegészítő mérőszámot az irodalomból, amelyeket érdemes együtt tekinteni, ahhoz, hogy megbízható képet kaphassunk a vizsgált klasszifikátor teljesítményéről. A pontosságot tekintve a javasolt algoritmus több mint 85 %-ot ért el minden vi-



7.6. Ábra: Két jellemző példa mutatja, hogyan változik a loss függvény, milyen a konvergencia jellege. Az ábrázolás során a minden iterációra kapott loss érték felkerült a "plot"-ra. Lassan ugyan, de a loss átlagosan lecsökkent. Az "early stopping" technikáját alkalmazva pedig végül sikerül visszatérnünk a legjobb epoch-hoz, amikor a legkisebb volt a loss.

selkedési fázisban egy olyan független tesztkészleten kiértékelve, amely különböző csecsemőkről, különböző körülmények között végzett felvételekből származnak és a tanító halmazban nem találhatóak meg. A pontosság mellett a következő mutatókat és mérőszámokat is megvizsgáltuk:

Pontosság:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7.3)$$

Specificitás — Ez megmutatja, hogy a modell mennyire tudja elkerülni a téves riasztásokat (angolul "false alarm"-okat):

$$Specificity = \frac{TN}{FP + TN} \quad (7.4)$$

Recall vagy Szenzitivitás — Ez a mérőszám megmutatja, hogy a bináris osztályozó mennyire jó a pozitív minták észlelésében. Valódi pozitív árfolyamnak is nevezzük:

$$Recall = \frac{TP}{TP + FN} \quad (7.5)$$

ROC görbe alatti terület — Ez a mérőszám is az osztályozók értékelésére szolgál:

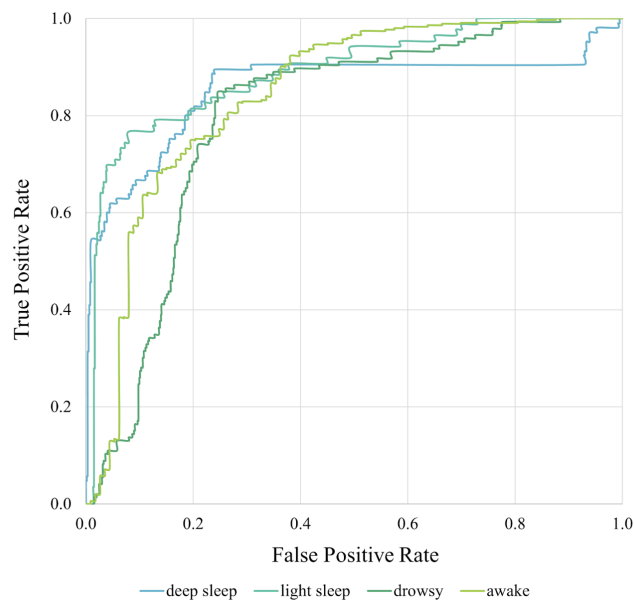
$$ROC = \frac{1 - Specificity}{Recall} \quad (7.6)$$

ROC görbe és az alatta lévő terület a gépi tanulásban használt fontos mérőszámok. Az ROC görbe egy grafikus reprezentáció, amely az osztályozó rendszer

teljesítményét mutatja az érzékenység és a szpecificitás függvényében. Ahogyan fentebb említettem, az érzékenység azt mutatja meg, hogy az osztályozó mennyire képes helyesen azonosítani a pozitív eseteket, míg a szpecificitás azt mutatja meg, hogy mennyire képes helyesen azonosítani a negatív eseteket.

Az ROC görbe alatti terület (Area Under the Curve, AUC) pedig a görbe alatti terület mértéke, és egy összefoglaló érték a teljesítményről. Ez az érték 0 és 1 közötti szám, ahol 0 jelenti a legrosszabb, 1 pedig a legjobb teljesítményt. Minél közelebb van az AUC 1-hez, annál jobb az osztályozó rendszer.

Az AUC értékeket az ajánlott algoritmus esetén is kiszámítottuk minden egyes fázishoz tartozó bináris klasszifikátorhoz. A ROC görbék a 7.7. ábrán láthatóak a különböző fázisokra bontva. Az AUC értékek pedig a 7.1. táblázatban.



7.7. Ábra: Ezen az ábrán a különböző bináris osztályozók ROC görbéi láthatók, amelyek alatti területek adják meg az AUC értékeket. Az ROC görbe lényege, hogy vizualizálja egy osztályozó rendszer teljesítményét az érzékenység és a szpecificitás függvényében. Egy jó osztályozó ROC görbélye a bal alsó sarokból a jobb felső sarok felé ível. Ez azt jelzi, hogy az osztályozó rendszer képes megfelelően elkülöníteni a pozitív és negatív eseteket, és növekvő döntési küszöbértékek mellett növekvő érzékenységet és szpecificitást mutat.

A fentebb leírt specificitás és szenzitivitás értékelési mutatók egymást kiegészítő jellege abban áll, hogy számolnunk kell azzal az eshetőséggel, ha csak az egyiket tekintjük a szenzitivitás és specificitás mutatókból, ami lehetőséget adna a csalásra.

7.1. Táblázat: Ez a táblázat foglalja össze a fentebb leírt metrikák eredményeit a különböző bináris osztályozókra vonatkozóan.

Adott bináris klasszifikátor cél osztálya	Pontosság	Szenzitivitás	Specifititás	AUC
Csendes Alvás fázis	93.5±4%	0.76	0.99	0.89
Aktív Alvás fázis	95.8±0.8%	0.76	0.98	0.85
Átmeneti fázis	91.13±6%	0.81	0.90	0.81
Ébrenlét fázis	85.51±2%	0.88	0.65	0.84

Olyan értelemben, hogy ezzel elérhetjük, hogy egy rossz teljesítményű klasszifikátort is jónak értékeljünk. Ezt a csalást vagy inkább önbecsapást nyilvánvalóan szeretnénk elkerülni. Mint azt fentebb említettem, a szenzitivitás azt méri, hogy az osztályozó mennyire tudja megtalálni a pozitív mintákat. Egy teszt itt csalhat, és ezt a mutatót maximalizálhatja, ha az osztályozó mindig „pozitív” értéket ad vissza. A specificitás ellenben azt méri, hogy mennyire tudja detektálni a vizsgált osztályozó a téves riasztásokat, vagy más néven "true negative" mintákat. Egy teszt itt is csalhat, és maximalizálhatja ezt a mutatót, ha az osztályozó mindig „negatív” értéket ad vissza. Tehát mindkét mutató félrevezető lehet önmagában. A megoldás, hogy nem egyiket vagy másikat, hanem mind a két mutatót egyszerre tekintjük. Ha mindkét mutatóban jól teljesít az osztályozónk, akkor egy jó teljesítményű osztályozóról van szó. Az általunk fentebb használt pontosságot bizonyos értelemben tekinthetjük a két mérőszám egyesítésének, ugyanis a számlálójában a specificitás és a szenzitivitás számlálóinak összege van. Nevezőjében pedig ezek nevezőinek összege. Csalásra azonban itt is van lehetőség. Egy osztályozó ezt a mutatót is maximalizálhatja, ha csak egy olyan eredményt ad vissza, amelyben a legbiztosabb. A legjobb megoldás, ha minden mutatót figyelemmel követünk és ezek együttese alapján ítéljük meg a klasszifikátorok teljesítményét. Ennek szellemében tekintsük a 7.1. táblázat által összefoglalt eredményeket.

A bemutatott algoritmus pontossága a fentebbi eredmények fényében viszonylag meggyőzőnek látszik. A pontosság tekintetében az eredmények szórását úgy kaptuk meg, hogy a hiperparaméter optimalizáció után az RNN kaszkádot 6 alkalommal újra tanítottuk. Ha jobban megvizsgáljuk a kapott értékeket, akkor láthatjuk, hogy elsősorban a mély alvó és az aktív alvó állapotok esetén magas az ajánlott algoritmus

pontossága, ami magyarázza, hogy miért volt érdemes a kaszkád elejére helyezni az ezeket a fázisokat detektáló bináris klasszifikátorokat. A magasabb pontosság oka ezeknél a klasszifikátoroknál valószínűleg abban keresendő, hogy az ajánlott algoritmusnak egyik fontos eleme a korábban bemutatott érintésmentes légzésbecslő. A mozgásalapú légzés alapján pedig az alvó fázisokat könnyebb elkülöníteni ugyanis alvás közben viszonylag folyamatos és nyugodt légzés figyelhető meg. Sőt ez igaz a többi mozgásalapú jel esetében is, hogy alvás közben ritkábbak a szélsőértékek és kisebb a különböző mozgások frekvenciája. Mivel a videóalapú jellemzők nagyobb része mozgásalapon van kiszámítva, ezért elsősorban mozgási mintázatok felismerésén alapszik az ajánlott algoritmus működése. A nyugodt mozgások jobban felismerhető mintázatok mutatnak alvó fázisban, mint a szabálytalan végtagmozgások. Ez lehet a magyarázata annak is, hogy a magas pontosság alól kivételt jelent az éber fázist detektáló bináris klasszifikátor. Ennek esetében a pontossága csak 85.51 % volt. Ezen az értéken a jövőben szeretnénk javítani, amire jó esélyünk lehet további jellemzők megfigyelésével és ezeknek a bemeneti jel halmazhoz való hozzáadásával. Erre lehet egy példa a babák arcának és szemének detektálása és az ezeken a területeken történő mozgások és egyéb jellemzők megfigyelése, amelyekkel reményeink szerint javítani tudjuk a teljesítményt az összes mutató terén. A csukott vagy nyitott szem ugyanis minden valószínűség szerint nagyban segíthet a viselkedési fázisok szétválasztásában. Ezt az is bizonyítja, hogy ennek (a szem nyitottságának) a megfigyelése is szerepet játszik a kórházi dolgozók döntésében, amikor automatizálás nélkül meghatározzák a babák viselkedési állapotát a kórházakban – erről a megfigyelésről bővebben Inga Warren et al. ír [118].

Ennek az algoritmusnak az esetében is megvizsgáltuk azt is, hogy mennyi a javasolt algoritmus futási ideje, ami egy AMD EPYC processzorral (2.25 Ghz) és 12 Gigabájt RAM-al rendelkező rendszeren futtatva): kb. $5.1 \text{ ms} \pm 4.7$.

7.7. Összehasonlítás a versenytársakkal

A korábban bemutatott kapcsolódó szakirodalomban találunk a javasolt algoritmushoz hasonló eljárásokat, vagy olyanokat amelyek ugyanúgy NIDCAP skála

szerint osztályoznak. Teljesen megegyező kimenetekkel és bemenetekkel rendelkező algoritmust azonban nem találtam a szakirodalomban. Korábban kiemeltem Muhammad Awais et al. munkáját ([9]), amely esetében a csecsemők arcát figyelték meg alvási állapotaik megállapításához. Ennek alkalmazása azonban jelenlegi formájában nem tűnik megvalósíthatónak klinikai körülmények között. Az arc nem mindig, vagy nem mindig egészen jól látszik. Klinikai körülmények között történő alvási állapot klasszifikációra vannak egyéb módszerek, amelyek kifejezetten erre lettek kifejlesztve, mint Xi Long et al. munkája ([68]), amelynél hozzánk hasonlóan aktigráfiát, ráadásul videóalapú aktigráfiát alkalmaztak. Bemenetként elsősorban mi is mozgási jeleket használtunk, különböző informatív régiókból kivonva, és a pulzus variabilitás alapú jelekkel kiegészítve. Kimenetként azonban Xi Long et al. munkájával ellentétben az általunk javasolt algoritmus nem csupán az alvó és nem alvó szakaszokat különítette el, hanem a NIDCAP szerinti osztályokba sorolta be az aktuális periódust.

A legkézenfekvőbb azonban valószínűleg az, ha S. Cabon et al. munkájával ([18]) hasonlítjuk össze az általunk javasolt algoritmust. Ebben ugyanis egy olyan módszer mutatnak be, ahol szintén a NIDCAP szerinti kategóriákba osztályoztak, még hozzá nagyon hasonló módon. A bemenetek az ő általuk vizsgált módszereknél szintén aktigráfia alapú adatok és hang alapú jellemzők voltak. Míg mi hang alapú jellemzők helyett pulzus variabilitás alapú jellemzőket alkalmaztunk. Ahogyan bemutattuk, a pontosság tekintetében az általunk javasolt algoritmus az alvó fázisok klasszifikációja esetén ért el jobb eredményeket. A csendes alvás fázis esetén 93.5% körüli, az aktív alvás esetén pedig 95.8% körüli eredményt. Ezzel ellentétben S. Cabon et al. munkájában az aktív ébrenlét és a csendes ébrenlét fázisok esetén érték el átlagosan 93.5% és 99.0% körüli pontosságokat, a legjobb eredményeket tekintve. Az alvó fázisok esetében viszont 85% alatti eredményeket. Tehát, mi pont azokban a kategóriákban voltunk pontosak amelyekben ők kevésbé és fordítva. Míg mi rekurzív neurális hálózatokat alkalmaztunk addig ők "SVM - Support Vector Machine"-el és "Random Forest"-el végeztek kísérleteket. További különbség még, hogy ők hang alapú adatokat alkalmaztak a mozgás alapú adatok mellett, míg mi pulzus adatokat, és a mozgásalapú jeleket különböző informatív régiókból vontuk ki. Elmondható,

hogy a két eljárás több részletében különbözik és az eredmények alapján másban erősek.

7.8. A fejezet konklúziója

A bemutatott algoritmus és az annak értékelése során elért teljesítmény bizonyítja a csecsemők viselkedési fázisainak automatikus osztályozásának megvalósíthatóságát a NIDCAP skála szerint. Legalábbis az általunk vizsgált csecsemők populációjához hasonló populációk esetében (lásd 3.1). Az algoritmus klinikákon való alkalmazása reményeink szerint tehermentesítheti a kórházi személyzetet a csecsemők viselkedési állapotainak folyamatos megfigyeléséről és mégis lehetőséget ad a családcentrikus fejlődéstámogató ellátásra. A megfelelő mennyiségű alvás biztosítása kritikus a csecsemők megfelelő fejlődése szempontjából, amint azt korábban említettük. Az algoritmusunk lehetőséget biztosíthat a koraszülött klinikák számára, hogy a különböző kezeléseket és családi látogatásokat úgy ütemezzék, hogy ne a csecsemők legmélyebben alvó fázisait kelljen megszakítani. Egy valós idejű alvás-értékelő rendszer kiépítéséhez azonban további értékelés, fejlesztés és persze nagyon nagy mennyiségű adat szükséges. A legújabb trendek a csecsemőmegfigyelés kutatása terén afelé mutatnak, hogy 5-10 résztvevő helyett 50-100 vagy még több résztvevőről készült felvételeket magába foglaló adathalmazokat hozzunk létre és 10-20 ezer órányi adattal rendelkezünk. Éppen ezért a Semmelweis Egyetem gyermekgyógyászati klinikáival, pontosabban ezen klinikák Koraszülött Intenzív Osztályaival közösen elindítottunk egy projektet, amelynek célja egy teljes 24 órás felvételek készítésére alkalmas csecsemő megfigyelő rendszer kiépítése, beleértve minden szükséges szoftver és hardver fejlesztést. A projekt a Transzlációs Idegtudományi Nemzeti Laboratóriumhoz tartozik. Ez a nemzeti labor biztosította a finanszírozást.

8. Fejezet

Diszkusszió és tézisek

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. Új, hagyományos színes kamera alapú algoritmusokat hoztam létre.

Adatbázist építettük és az adatbázis kezelésére és (akár automatikus) annotálására alkalmas szoftver elemeket hoztunk létre.

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

8.1. Új tudományos eredmények

Ez a dolgozat koraszülött csecsemők érintésmentes, kamera alapú monitorozásával foglalkozik.

Három kihívást jelentő kutatási témát vizsgáltam meg ezen a területen. Nevezetesen a koraszülött csecsemők érintésmentes légzés monitorozását, az ehhez szükséges ROI detektor tanítására használt adathalmaz automatikus annotációját és a csecsemők viselkedési állapotainak érintésmentes monitorozását.

A problémák megoldásához gépi látási és gépi tanulási algoritmusokat alkalmaz-

tam. Ezeket használva mozgás alapon számítottam ki a képen lévő csecsemők légzését és részben mozgás alapon, részben a csecsemők 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.

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]

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: [C3]

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], [P1], [P2]

8.2. Jövőbeni tervek

Az adatgyűjtés és a kutatások folytatásához szükséges etikai engedélyeket már megkaptuk és a Semmelweis Egyetemen dolgozó partnereinkkel közösen megterveztük a kutatás következő fázisát, amelynek keretein belül több orvos vagy nővér fog több csecsemőt megfigyelni és annotálni mobiltelefonos alkalmazás segítségével a NIDCAP skálájú viselkedési állapotok szerint. A korábban megvalósíthatósági tanulmányban bemutatott viselkedési állapot becslőt így egy nagyobb annotált adathalmaz segítségével továbbfejleszthetjük és validálhatjuk.

Együttműködések születtek a MEDICOR-ral és a Semmelweis Egyetemmel is, és olyan startup-okkal is, mint a CONSUSIS. Jelenleg már a módszerek validációja és ipari terméké alakítása van előttünk, ami egy évekig tartó hosszú folyamat lesz. De elmondhatjuk, hogy kutatási fázisból eljutottunk odáig, hogy termék validáció és fejlesztés történik.

Reményeink szerint az itt bemutatott algoritmusok egy az inkubátorba integrált komplex csecsemőmegfigyelő részét képezik majd amelyeket kiegészítő termékként vásárolhatnak majd meg az inkubátorokat megvásároló fogyasztók. Egy ilyen rendszer kamera alapon kiegészítő információkkal láthatja el a kórházi személyzetet a csecsemő állapotáról, amelyek sokat segíthetnek a születési rendellenességek elkerülésében és a csecsemők életminőségének javításában.

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, 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/16/7215> 18, 34, 51, 54, 58, 62, 64, 82, 101)
- [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> 103)
- [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> 103)
- [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> 103)

A szerző konferencia publikációinak gyűjteménye

- [C1] 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. 18, 34, 43, 45, 46, 103)
- [C2] J. Imre, **A. Nagy**, F. Péter, Z. Akos, S. Máté, V. Judit, and S. Miklós, „Neonatal activity monitoring by camera-based multi-lstm network,” 10 2022, pp. 1–4. 92, 102)
- [C3] **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. 101)

További publikációk

- [Au1] **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. 18, 34, 38, 39, 42, 103)

A szerző konferencia posztereinek gyűjteménye

- [P1] 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. 102)
- [P2] **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. 102)

Hivatkozások

- [1] Videoalapú újszülöttkori mozgásérzékelés. In *2020-ban az IEEE Engineering in Medicine Biology Society (EMBC) 42. éves nemzetközi konferenciája*, pages 6135–6138. 18)
- [2] Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. *Circulation*, 93(5):1043–1065, March 1996. 89)
- [3] Preterm labor and birth: Condition information, 2017. Accessed: 2022-09-20. 7)
- [4] Lonneke A M Aarts, Vincent Jeanne, John P Cleary, C Lieber, J Stuart Nelson, Sidarto Bambang Oetomo, and Wim Verkruysse. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - a pilot study. *Early Hum. Dev.*, 89(12):943–948, Dec. 2013. 84)
- [5] Abbas K. Abbas, Konrad Heimann, Katrin Jergus, Thorsten Orlikowsky, and Steffen Leonhardt. Neonatal non-contact respiratory monitoring based on real-time infrared thermography. *Biomedical engineering online*, 10:93, 10 2011. 15)
- [6] Abbas K Abbas and Steffen Leonhardt. Intelligent neonatal monitoring based on a virtual thermal sensor. *BMC Med. Imaging*, 14(1):9, March 2014. 84)
- [7] Khalid Al-mutib, Muhammad Emaduddin, Mansour Alsulaiman, Ramdane Hedjar, and Ebrahim Mattar. Motion periodicity based pedestrian detection and particle filter based pedestrian tracking using stereo vision camera. 03 2013. 20)

- [8] L Atallah, A Serteyn, M Meftah, M Schellekens, R Vullings, J W M Bergmans, A Osagiator, and S Bambang Oetomo. Unobtrusive ECG monitoring in the NICU using a capacitive sensing array. *Physiol. Meas.*, 35(5):895–913, May 2014. 84)
- [9] Muhammad Awais, Chen Chen, Xi Long, Bin Yin, Anum Nawaz, Saadullah Abbasi, Saeed Akbarzadeh, Linkai Tao, Chunmei Lu, Laishuan Wang, R. Aarts, and Wei Chen. Novel framework: Face feature selection algorithm for neonatal facial and related attributes recognition. *IEEE Access*, 8:59100–59113, March 2020. 19, 84, 98)
- [10] Muhammad Awais, Xi Long, Bin Yin, Saadullah Farooq Abbasi, Saeed Akbarzadeh, Chunmei Lu, Xinhua Wang, Laishuan Wang, Jiong Zhang, Jeroen Dudink, and Wei Chen. A hybrid dcnn-svm model for classifying neonatal sleep and wake states based on facial expressions in video. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1441–1449, 2021. 19)
- [11] Muhammad Awais, Xi Long, Bin Yin, Saadullah Farooq Abbasi, Saeed Akbarzadeh, Chunmei Lu, Xinhua Wang, Laishuan Wang, Jiong Zhang, Jeroen Dudink, and Wei Chen. A hybrid dcnn-svm model for classifying neonatal sleep and wake states based on facial expressions in video. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1441–1449, 2021. 83)
- [12] Chiraz Benabdelkader and L. Davis. Detection of people carrying objects : A motion-based recognition approach. pages 378 – 383, 06 2002. 20)
- [13] 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. 70)
- [14] J.-Y. Bouguet. Pyramidal implementation of the lucas kanade feature tracker. 1999. 38)

- [15] Jorge Brieva, Hiram Ponce, and Ernesto Moya-Albor. Non-contact breathing rate estimation using machine learning with an optimized architecture. *Mathematics*, 11:645, 01 2023. 17, 78)
- [16] 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. 19)
- [17] Stephen Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6):536–541, 1930. 56)
- [18] S. Cabon, F. Porée, A. Simon, B. Met-Montot, P. Pladys, O. Rosec, N. Nardi, and G. Carrault. Audio- and video-based estimation of the sleep stages of newborns in neonatal intensive care unit. *Biomedical Signal Processing and Control*, 52:362–370, 2019. 19, 84, 98)
- [19] Peter H Charlton, Kevin Kotzen, Elisa Mejía-Mejía, Philip J Aston, Karthik Budidha, Jonathan Mant, Callum Pettit, Joachim A Behar, and Panicos A Kyriacou. Detecting beats in the photoplethysmogram: benchmarking open-source algorithms. *Physiological Measurement*, 43(8):085007, Aug. 2022. 89)
- [20] Richard Chen, Ming Lu, Tiffany Chen, Drew Williamson, and Faisal Mahmood. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5:1–5, 06 2021. 20)
- [21] Weixuan Chen and Daniel McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 349–365, 2018. 17, 46, 47)
- [22] Dmitry Chetverikov and Attila Axt. Approximation-free running svd and its application to motion detection. *Pattern Recognition Letters*, 31(9):891–897, 2010. 40)

- [23] Ronald E. Dahl. The regulation of sleep and arousal: Development and psychopathology. *Development and Psychopathology*, 8(1):3–27, 1996. 83)
- [24] 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. 69)
- [25] Pál Demeter and Akos Pap. The relationship between gastroesophageal reflux disease and obstructive sleep apnea. *Journal of gastroenterology*, 39:815–20, 10 2004. 34)
- [26] Joseph Espiritu. *Health Consequences of Obstructive Sleep Apnea*, pages 23–43. 01 2021. 34)
- [27] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. volume 2749, pages 363–370, 06 2003. 43, 55)
- [28] Desta Fekedulegn, Michael E Andrew, Mingming Shi, John M Violanti, Sarah Knox, and Kim E Innes. Actigraphy-Based Assessment of Sleep Parameters. *Annals of Work Exposures and Health*, 64(4):350–367, 02 2020. 18)
- [29] 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. 71)
- [30] Peter Fuller, Waldemar Wenner, and Susan Blackburn. Comparison between time-lapse video recordings of behavior and polygraphic state determinations in premature infants. *Psychophysiology*, 15:594–8, 12 1978. 19, 84)
- [31] Morse G., Chetverikov D., and Egyed D. Video-based respiration monitoring of newborn babies: a feasibility study. In *Ninth Hungarian Conference on Computer Graphics and Geometry (Grafgeo)*, page 209–215, 2018. 36, 38)

- [32] Barbara Galland, Kim Meredith-Jones, Philip Terrill, and Rachael Taylor. Challenges and emerging technologies within the field of pediatric actigraphy. *Frontiers in psychiatry*, 5:99, 08 2014. 19)
- [33] Kim Gibson, Ali Al-Naji, Julie Fleet, Mary Steen, Adrian Esterman, Javaan Chahl, Jasmine Huynh, and Scott Morris. Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: a method comparison study. *Pediatric research*, 86(6):738–741, 2019. 17)
- [34] Stanley Graven. Sleep and brain development. *Clin. Perinatol.*, 33(3):693–706, Sept. 2006. 83)
- [35] Stanley Graven and Joy Browne. Sleep and brain development: The critical role of sleep in fetal and early neonatal brain development. *Newborn and Infant Nursing Reviews*, 8:173–179, Dec. 2008. 83)
- [36] Stanley N Graven and Joy V Browne. Sleep and brain development. *Newborn Infant Nurs. Rev.*, 8(4):173–179, December 2008. 83)
- [37] Stanley N. Graven and Joy V. Browne. Sleep and brain development: The critical role of sleep in fetal and early neonatal brain development. *Newborn and Infant Nursing Reviews*, 8(4):173–179, 2008. Brain Development of the Neonate. 8)
- [38] Alex Graves. *Long Short-Term Memory*, pages 37–45. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. 50)
- [39] M R Grubb, J Carpenter, J A Crowe, J Teoh, N Marlow, C Ward, C Mann, D Sharkey, and B R Hayes-Gill. Forehead reflectance photoplethysmography to monitor heart rate: preliminary results from neonatal patients. *Physiol. Meas.*, 35(5):881–893, May 2014. 84)
- [40] Tianqi Guo, Qian Lin, and Jan Allebach. Remote estimation of respiration rate by optical flow using convolutional neural networks. *Electronic Imaging*, 33:267–1, 01 2021. 17)

- [41] 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. 19)
- [42] Adrienne Heinrich, Frank Heesch, Bhargava Puvvula, and Mukul Rocque. Video based actigraphy and breathing monitoring from the bedside table of shared beds. *Journal of Ambient Intelligence and Humanized Computing*, 6:107–120, 02 2014. 18)
- [43] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 17)
- [44] Gee-Sern Hsu, ArulMurugan Ambikapathi, and Ming-Shiang Chen. Deep learning with time-frequency representation for pulse estimation from facial videos. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 383–389. IEEE, 2017. 18)
- [45] Xinmei Hu, Shasha Yuan, Fangzhou Xu, Yan Leng, and Kejiang Yuan. Scalp eeg classification using deep bi-lstm network for seizure detection. *Computers in Biology and Medicine*, 124:103919, July 2020. 91)
- [46] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059, 2020. 49)
- [47] Hyeon-Sang Hwang and Eui-Chul Lee. Non-contact respiration measurement method based on rgb camera using 1d convolutional neural networks. *Sensors*, 21(10), 2021. 17)
- [48] Yuki Ito, Kento Morita, Tetsushi Wakabayashi, Harumi Shinkoda, Asami Matsumoto, Yukari Noguchi, and Masako Shiramizu. Automatic estimation of neonatal sleep/wake states in the nicu using 3d cnn. In *2022 World Automation Congress (WAC)*, pages 163–168, 2022. 18)

- [49] Yuki Ito, Kento Morita, Tetsushi Wakabayashi, Harumi Shinkoda, Asami Matsumoto, Yukari Noguchi, and Masako Shiramizu. Automatic estimation of neonatal sleep/wake states in the nicu using 3d cnn. pages 163–168, 10 2022. 84)
- [50] Rik Janssen, Wenjin Wang, Andreia Moço, and Gerard de Haan. Video-based respiration monitoring with automatic region of interest detection. *Physiological measurement*, 37(1):100, 2015. 17)
- [51] A Johansson, P A Oberg, and G Sedin. Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique. *J. Clin. Monit. Comput.*, 15(7-8):461–467, Dec. 1999. 84)
- [52] 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. 18, 34)
- [53] Eliot Katz, Ron Mitchell, and Carolyn D’Ambrosio. Obstructive sleep apnea in infants. *American journal of respiratory and critical care medicine*, 185:805–16, 12 2011. 34)
- [54] Eliot Katz, Ron Mitchell, and Carolyn D’Ambrosio. Obstructive sleep apnea in infants. *American journal of respiratory and critical care medicine*, 185:805–16, 12 2011. 34)
- [55] Fatema-Tuz-Zohra Khanam, Asanka G. Perera, Ali Al-Naji, Kim Gibson, and Javaan Chahl. Non-contact automatic vital signs monitoring of infants in a neonatal intensive care unit based on neural networks. *Journal of Imaging*, 7(8), 2021. 18, 36, 37)
- [56] Janou Korte, Katharina Lyulff, Claudia Opp, Renate Siegrmund, and Katharina Wulff. Ultradian and circadian activity-rest rhythms of preterm neonates compared to full-term neonates using actigraphic monitoring, 01 2001. 19)

- [57] Daniel G. Kyrollos, Joshua B. Tanner, Kim Greenwood, JoAnn Harrold, and James R. Green. Noncontact neonatal respiration rate estimation using machine vision. In *2021 IEEE Sensors Applications Symposium (SAS)*, pages 1–6, 2021. 18, 36, 37)
- [58] Dmitry Lachinov, Evgeniy Vasilyev, and Vadim Turlapov. Glioma segmentation with cascaded unet, 10 2018. 49)
- [59] Hyunwoo Lee, Ayoung Cho, Seongwon Lee, and Mincheol Whang. Vision-based measurement of heart rate from ballistocardiographic head movements using unsupervised clustering. *Sensors*, 19(15):3263, 2019. 17)
- [60] W K Lee, H Yoon, D W Jung, S H Hwang, and K S Park. Ballistocardiogram of baby during sleep. *Annu Int Conf IEEE Eng Med Biol Soc*, 2015:7167–7170, 2015. 84)
- [61] Won Lee, Seung Kim, Jae Na, Young-Hyo Lim, Seok Cho, Sung Ho Cho, and Hyun-Kyung Park. Non-contact sleep/wake monitoring using impulse-radio ultrawideband radar in neonates. *Frontiers in Pediatrics*, 9:1–9, 12 2021. 83)
- [62] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jędrzej Nowak. Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity. In *2011 federated conference on computer science and information systems (FedCSIS)*, pages 405–410. IEEE, 2011. 17)
- [63] Claudio Liguori, Samantha Mombelli, Mariana Fernandes, Marco Zucconi, Giuseppe Plazzi, Luigi Ferini-Strambi, Giancarlo Logroscino, Nicola Biagio Mercuri, and Marco Filardi. The evolving role of quantitative actigraphy in clinical sleep medicine. *Sleep Medicine Reviews*, 68:101762, 2023. 18)
- [64] Yizhi Liu, Xiaoyan Gu, Lei Huang, Junlin Ouyang, Miao Liao, and Liangran Wu. Analyzing periodicity and saliency for adult video detection. *Multimedia Tools and Applications*, 79, 02 2020. 20)
- [65] Zimeng Liu, Bin Huang, Chun-Liang Lin, Chieh-Liang Wu, Changchen Zhao, Wen-Cheng Chao, Yu-Cheng Wu, Yadan Zheng, and Zhiru Wang. Contact-

- less respiratory rate monitoring for icu patients based on unsupervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6004–6013, June 2023. 17, 78)
- [66] Xi Long, R.A. Otte, Eric Sanden, Jan Werth, and Tao Tan. Video-based actigraphy for monitoring wake and sleep in healthy infants: A laboratory study. *MDPI Sensors*, 19:1075, 03 2019. 83)
- [67] Xi Long, Renée Otte, Eric van der Sanden, Jan Werth, and Tao Tan. Video-based actigraphy for monitoring wake and sleep in healthy infants: A laboratory study. *Sensors*, 19(5), 2019. 19)
- [68] Xi Long, Renée Otte, Eric van der Sanden, Jan Werth, and Tao Tan. Video-based actigraphy for monitoring wake and sleep in healthy infants: A laboratory study. *Sensors*, 19(5), 2019. 19, 84, 98)
- [69] Xi Long, Javier Perez, R.A. Otte, Wenjin Wang, R. Aarts, and Peter Andriessen. Video-based actigraphy is an effective contact-free method of assessing sleep in preterm infants. *Acta Paediatrica*, 110, 12 2020. 18)
- [70] V Marchi, A Hakala, A Knight, F D’Acunto, ML Scattoni, A Guzzetta, and S Vanhatalo. Automated pose estimation captures key aspects of general movements at eight to 17 weeks from conventional videos. *Acta Paediatr.*, 10 2019. 31)
- [71] P Marchionni, L Scalise, I Ercoli, and E P Tomasini. An optical measurement method for the simultaneous assessment of respiration and heart rates in preterm infants. *Rev. Sci. Instrum.*, 84(12):121705, Dec. 2013. 84)
- [72] P Marchionni, L Scalise, I Ercoli, and EP Tomasini. An optical measurement method for the simultaneous assessment of respiration and heart rates in preterm infants. *Review of Scientific Instruments*, 84(12):121705, 2013. 35)
- [73] Alina Marcu, Dragos Costea, Vlad Licaret, and Marius Leordeanu. Towards automatic annotation for semantic segmentation in drone videos. 10 2019. 20)

- [74] 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. 18, 34)
- [75] Lalit Maurya, Reyer Zwiggelaar, Deepak Chawla, and Prasant Mahapatra. Non-contact respiratory rate monitoring using thermal and visible imaging: a pilot study on neonates. *Journal of Clinical Monitoring and Computing*, 37, 12 2022. 17)
- [76] Yohei Mukai, Kento Morita, Nobu C. Shirai, Tetsushi Wakabayashi, Harumi Shinkoda, Asami Matsumoto, Yukari Noguchi, and Masako Shiramizu. Automatic classification of neonatal sleep-wake states based on facial video analysis. In *2019 4th International Conference on Information Technology Research (ICITR)*, pages 1–6, 2019. 19)
- [77] Adam Nagy. Neonatal-respiration-monitoring-algorithm. <https://github.com/cezius/Neonatal-Respiration-Monitoring-Algorithm>, 2022. 52)
- [78] Yunyoung Nam and Jung Park. Physical activity recognition using a single triaxial accelerometer and a barometric sensor for baby and child care in a home environment. *Journal of Ambient Intelligence and Smart Environments*, 5:381–402, 07 2013. 19)
- [79] Meir Nitzan, Ayal Romem, and Robert Koppel. Pulse oximetry: Fundamentals and technology update. *Medical devices (Auckland, N.Z.)*, 7:231–9, 07 2014. 8)
- [80] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. Synrhythm: Learning a deep heart rate estimator from general to specific. *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3580–3585, 2018. 18)
- [81] American College of Obstetricians and Gynecologists. Definition of term pregnancy. *Obstetrics and Gynecology*, page 1139–1140, 2013. 7)
- [82] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9:62–66, 01 1979. 41)

- [83] Daniil Pakhomov, Sanchit Hira, Narayani Wagle, Kemar Green, and Nassir Navab. Segmentation in style: Unsupervised semantic image segmentation with stylegan and clip, 07 2021. 20, 66, 80)
- [84] J PALLETT and J SCOPES. Recording respirations in newborn babies by measuring impedance of the chest. *Medical electronics biological engineering*, 3:161–8, 05 1965. 35)
- [85] Cheolhyeong Park and Deokwoo Lee. Classification of respiratory states using spectrogram with convolutional neural network. *Applied Sciences*, 12(4), 2022. 17)
- [86] Elena Peralta, Jesus Lazaro, Raquel Bailon, Vaidotas Marozas, and Eduardo Gil. Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals. *Physiological measurement*, 40(2):025007, 2019. 90)
- [87] Carina Barbosa Pereira, Konrad Heimann, Boudewijn Venema, Vladimir Blazek, Michael Czaplik, and Steffen Leonhardt. Estimation of respiratory rate from thermal videos of preterm infants. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3818–3821. IEEE, 2017. 35)
- [88] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2010. 17)
- [89] Anindya Pravitasari, Nur Iriawan, Mawanda Almuhayar, Taufik Azmi, Irhamah Irhamah, Kartika Fithriasari, Santi Purnami, and Widiani Ferriastuti. Unet-vgg16 with transfer learning for mri-based brain tumor segmentation. *TEL-KOMNIKA (Telecommunication Computing Electronics and Control)*, 18:1310, 06 2020. 49)
- [90] Tauhidur Rahman, Alexander T Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N Patel, Julie A Kientz, and Tanzeem Choudhury. Dopples-

- leep: A contactless unobtrusive sleep sensing system using short-range doppler radar. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 39–50, 2015. 35)
- [91] A Rajalakshmi, Sunitha K.A, and Revathi Venkataraman. A survey on neonatal incubator monitoring system. *Journal of Physics: Conference Series*, 1362:012128, 11 2019. 18)
- [92] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. 04 2018. 50, 51)
- [93] Md Reza, Akshay Naik, Chen Kai, and David Crandall. Automatic annotation for semantic segmentation in indoor scenes. 07 2019. 20)
- [94] 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. 49)
- [95] German Ros, Laura Sellart, Joanna Materzynska, David Vázquez, and Antonio López. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. pages 3234–3243, 06 2016. 20)
- [96] Scott Rossol, Jeffrey Yang, Caroline Toney-Noland, Janine Bergin, Chandan Basavaraju, Pavan Kumar, and Henry Lee. Non-contact video-based neonatal respiratory monitoring. *Children*, 7:171, 10 2020. 18, 36)
- [97] Scott Rossol, Jeffrey Yang, Caroline Toney-Noland, Janine Bergin, Chandan Basavaraju, Pavan Kumar, and Henry Lee. Non-contact video-based neonatal respiratory monitoring. *Children*, 7:171, 10 2020. 35)
- [98] A Sadeh, P Lavie, A Scher, E Tirosh, and R Epstein. Actigraphic home-monitoring sleep-disturbed and control infants and young children: a new method for pediatric assessment of sleep-wake patterns. *Pediatrics*, 87(4):494–499, Apr. 1991. 84)

- [99] Nadezhda Sazonova, Edward Sazonov, and S.A.C. Schuckers. Activity-based sleep-wake identification in infants. volume 29, pages 525– 528, 10 2002. 19)
- [100] Fred Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, September 2017. 90)
- [101] Jianbo Shi and Tomasi. Good features to track. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994. 39)
- [102] Hoo-Chang Shin, Matthew Orton, David Collins, Simon Doran, and Martin Leach. Autoencoder in time-series analysis for unsupervised tissues characterisation in a large unlabelled medical image dataset. volume 1, pages 259–, 12 2011. 20)
- [103] Rencheng Song, Senle Zhang, Chang Li, Yunfei Zhang, Juan Cheng, and Xun Chen. Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks. *IEEE Transactions on Instrumentation and Measurement*, 2020. 18)
- [104] Radim vŠpetlík, Vojtech Franc, and Jiri Matas. Visual heart rate estimation with convolutional neural network. In *Proceedings of the British Machine Vision Conference, Newcastle, UK*, pages 3–6, 2018. 17, 18)
- [105] 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. 89)
- [106] 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. 18, 34)
- [107] 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. 35, 69)

- [108] K.S. Tan, Reza Saatchi, Heather Elphick, and Derek Burke. Real-time vision based respiration monitoring system. pages 770 – 774, 08 2010. 35)
- [109] Qunfeng Tang, Zhencheng Chen, John Allen, Ayman Alian, Carlo Menin, Rabab Ward, and Mohamed Elgendi. Ppgsynth: An innovative toolbox for synthesizing regular and irregular photoplethysmography waveforms. *Frontiers in Medicine*, 11 2020. 20)
- [110] Qunfeng Tang, Zhencheng Chen, Rabab Ward, and Mohamed Elgendi. Synthetic photoplethysmogram generation using two gaussian functions. *Scientific Reports*, 10:13883, 08 2020. 20)
- [111] Joëlle Tilmanne, Jerome Urbain, Mayuresh Kothare, Alain Vande Wouwer, and Sanjeev Kothare. Algorithms for sleep-wake identification using actigraphy: A comparative study and new results. *Journal of sleep research*, 18:85–98, 03 2009. 19)
- [112] Michael E. Tipping and Christopher M. Bishop. Mixtures of Probabilistic Principal Component Analyzers. *Neural Computation*, 11(2):443–482, 02 1999. 71)
- [113] Du-Ming Tsai, Shu-Kai Fan, and Yi-Hsiang Chou. Auto-annotated deep segmentation for surface defect detection. *IEEE Transactions on Instrumentation and Measurement*, PP:1–1, 06 2021. 20, 66, 80)
- [114] Mauricio Villarroel, Sitthichok Chaichulee, Joao Jorge, Sara Davis, Gabrielle Green, Carlos Arteta, Andrew Zisserman, Kenny McCormick, Peter Watkinson, and L. Tarassenko. Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit. *npj Digital Medicine*, 2:128, 12 2019. 18)
- [115] Mauricio Villarroel, Alessandro Guazzi, Joao Jorge, Sara Davis, Peter Watkinson, Gabrielle Green, Asha Shenoi, Kenny McCormick, and L. Tarassenko. Continuous non-contact vital sign monitoring in neonatal intensive care unit. *Healthcare Technology Letters*, 1:87–91, 09 2014. 18)

- [116] Ping Wang, Aimin Jiang, Xiaofeng Liu, Jing Shang, and Li Zhang. Lstm-based eeg classification in motor imagery tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, PP:1–1, Oct. 2018. 91)
- [117] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2016. 17)
- [118] 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. 8, 19, 84, 97)
- [119] 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. 69)
- [120] Jan Werth, Xi Long, Elly Zwartkruis-Pelgrim, Hendrik Niemarkt, Wei Chen, Ronald M. Aarts, and Peter Andriessen. Unobtrusive assessment of neonatal sleep state based on heart rate variability retrieved from electrocardiography used for regular patient monitoring. *Early Human Development*, 113:104–113, 2017. 83)
- [121] Björn Westrup. Newborn individualized developmental care and assessment program (nidcap): Family-centered developmentally supportive care. *Early human development*, 83:443–9, 08 2007. 19)
- [122] Yuwei Xia, Sang Han, and Hyock Kwon. Image generation and recognition for railway surface defect detection. *Sensors*, 23:4793, 05 2023. 20, 66, 80)
- [123] Shudong Yang, Xueying Yu, and Ying Zhou. Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example. In *2020 International workshop on electronic communication and artificial intelligence (IWECAI)*, pages 98–101. IEEE, June 2020.)

- [124] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In *Proc. BMVC*, 2019. 17, 46, 82)
- [125] Luis Albert Zavala-Mondragón, Bishal Lamichhane, Lu Zhang, and Gerard Haan. Cnn-skelpose: a cnn-based skeleton estimation algorithm for clinical applications. *Journal of Ambient Intelligence and Humanized Computing*, 11, 06 2020. 18)
- [126] Hong Zeng, Chen Yang, Guojun Dai, Feiwei Qin, Jianhai Zhang, and Wanzeng Kong. Eeg classification of driver mental states by deep learning. *Cognitive Neurodynamics*, 12, 12 2018. 91)
- [127] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 12 2019. 75, 79)
- [128] Fei Zhu, Ye Fei, Yuchen Fu, Quan Liu, and Bairong Shen. Electrocardiogram generation with a bidirectional lstm-cnn generative adversarial network. *Scientific Reports*, 9, 05 2019. 20)
- [129] Liu Zijuan, Lin Wang, Wenyan Liu, and Binbin Li. Human movement detection and gait periodicity analysis using channel state information. pages 167–174, 12 2016. 20)