



Budapesti Műszaki és Gazdaságtudományi Egyetem

Villamosmérnöki és Informatikai Kar

Automatizálási és Alkalmazott Informatikai Tanszék

Mozgásszekvenciák felismerése mesterséges intelligencia alapú pózdetektálás segítségével

TDK DOLGOZAT

Készítette

Stuksza Gábor

mérnök-informatikus MSc (II. évfolyam)

Konzulens

Dr. Forstner Bertalan

egyetemi docens

2022. november 1.

Tartalomjegyzék

Kivonat	i
Abstract	ii
1. Bevezetés	1
1.1. A dolgozat felépítése	2
2. Kapcsolódó kutatások, különböző felismerési módszerek	3
2.1. Lólengés gyakorlatok valós idejű észlelése és teljesítményelemzése . .	3
2.2. Gyakorlatfelismerés viselhető szenzorok segítségével	5
2.3. Tornagyakorlatok felismerése és számolása nyomásérzékelő segítségével	6
2.4. A dolgozat célja	7
3. Megközelítés	10
4. Pózfelismerés	11
4.1. Adathalmaz kialakítása	11
4.2. MoveNet: Ultragyors és pontos pózfelismerő modell	12
4.3. Tanítás	14
5. Mozcásszekvenciák meghatározása	16
6. Prototípus megvalósítása	23
6.1. Kamera használata	23
6.2. Referenciapont felismerés	24
6.3. Gyakorlatok felismerése és elemzése	25
6.4. Gyakorlatok bővítése	28
7. Összegzés	29
7.1. Továbbfejlesztési lehetőségek	29
7.1.1. Adathalmaz bővítése	30
7.1.2. További gyakorlatok felvétele	30

7.1.3. Végrehajtás helyességének részletes elemzése	30
7.1.4. Okoseszközök bővítése	30

Irodalomjegyzék	32
------------------------	-----------

Kivonat

A mai modern világban az informatika számos területén van szükség különböző mesterséges intelligencia és gépi tanulás alapú megoldásokra, a szövegfelismeréstől kezdve az önvezető autókig.

A mesterséges intelligencia segítségével valós időben észlelhetjük az embereket, vagy azonosíthatjuk be az alany főbb testrészeit, mint például a vállát, a csípőjét vagy térdét. Ezen testrészekre úgynevezett referenciapontokat illeszthetünk, amelyek egymáshoz viszonyított helyzetéből következtethetünk a felvett pózra vagy testhelyzetre. Ezekre a problémákra már léteznek hatékony megvalósítások, ezzel szemben viszont a dinamikus mozgások felismerése még csak speciális felszerelésekkel lehetséges, mint például RGBD vagy 3D kamera segítségével. Továbbá még nem létezik nagy méretű publikus tanító adathalmaz az emberi pózokra, testhelyzetekre, emiatt a sportolókat különböző szenzorokkal látják el adatgyűjtés gyanánt, amely plusz kényelmetlenséget okoz a mozgások során. Az ilyen speciális elemzések használata legfőképpen az élsportban terjedtek el, viszont a hétköznapi ember számára még nem készült hatékony megoldás.

Ez a dolgozat bemutatja annak a kutatásnak az eredményét, amelynek célja a mesterséges intelligencián alapuló pózfelismerést olyan további algoritmusokkal kiegészíteni, amelyek lehetővé teszik, hogy a felvett pózok váltakozásai alapján detektálni lehessen, hogy az alany milyen mozdulatsort, tornagyakorlatot hajt végre, mindezt speciális eszközök használata nélkül. A dolgozat kitér arra is, hogy miképpen különböztethetők meg az ugyanazokat a pózokat tartalmazó gyakorlatok, valamint hogyan lehet csúszóablak segítségével észlelni a másik gyakorlatra váltást.

A megoldás hatékonysága egy olyan kísérleti prototípussal kerül igazolásra, amely az eszköz kameráján keresztül azonosítja be a cselekvéseket.

Abstract

In today's modern world, various fields of IT require different solutions based on artificial intelligence and machine learning, from text recognition to self-driving cars.

Artificial intelligence can be used to detect people in real time or to identify key parts of a subject's body, such as their shoulders, hips, or knees. So-called reference points can be attached to these body parts, which can be used to infer the pose or posture of the subject. Effective implementations already exist for these problems, but the recognition of dynamic movements is only possible with special equipment, such as RGBD or 3D cameras. Furthermore, there is still no large-scale public teaching data set for human poses and postures, which is why athletes are equipped with various sensors for data collection, but this often causes discomfort to the wearer. Special analyzes like this can be seen almost exclusively in elite sports, and no effective solution has yet been developed for ordinary people.

This thesis presents the results of research aimed at extending artificial intelligence-based pose recognition with additional algorithms that make it possible to detect sequences of movements based on the alternations of the recorded poses – all without the use of specialized equipment. The paper also covers how exercises with identical poses can be distinguished, and how the transition from one exercise to another can be detected using a sliding window.

The effectiveness of the solution is demonstrated with an experimental prototype that identifies actions through the camera of the device.

1. fejezet

Bevezetés

A rendszeres testmozgás és sportolás elengedhetetlen, hogy hosszútávon megőrizzük egészségünket, azonban rohanó életmódunkkal ez nem mindig összeegyeztethető. Emiatt különböző alternatívák jelentek meg, hogy edzéseinket otthonról, vagy online formában bonyolíthassuk, azonban a visszajelzések hiánya miatt kevésbé hatékonyak, akár különböző sérülésekhez is vezethetnek.

Számos olyan mozgásforma létezik, amit néhány okoseszköz segítségével egyszerűen kielemezhetünk, többek között az olyan kardió edzéseket, mint a futás, kérpározás, gyaloglás vagy túrázás. Ezek előnye, hogy a különböző szenzorok (GPS, gyorsulásmérő, giroszkóp stb.) adatai és az eltelt idő alapján könnyen megállapítható a megtett táv, a sebesség, a szintemelkedés, valamint az ehhez hasonló statisztikák. Ezen megoldások viszont nem alkalmazhatók olyan esetekben, mint például a tornagyakorlatok elemzése.

A rohamosan fejlődő okoseszközök – okostelefonok, okosórák – kamerája, illetve szenzorai számos új lehetőséget biztosítanak számunkra, hogy edzéseink hatékonyságát elemezzük. Ehhez viszont szükség van arra, hogy az alany mozgása során felvett pózokat beazonosítsuk és azok szabályszerű egymásutániságából következtetni tudjunk, hogy pontosan milyen mozgásformát hajtanak végre. A pózfelismerés már régóta aktív területét képi a gépi látásnak, mely lehetővé teszi, hogy nyomon követhessük az alany mozgásának változásait a testére helyezett referenciapontok segítségével.

A pózokból összeálló dinamikus mozgások felismerésére is léteznek megoldások, viszont ezek még csak speciális felszerelésekkel vagy szenzorokkal lehetséges, amely nem található meg minden háztartásban. Emiatt az ilyen speciális elemzések használata legfőképpen az élsportban terjedtek el, a hétköznapi ember számára még nem készült hatékony megoldás.

1.1. A dolgozat felépítése

A bevezetést követően részletesen bemutatok néhány merőben eltérő megközelítést a mozgásfelismerési probléma megoldására, valamint, hogy ezek milyen előnyökkel és hátrányokkal rendelkeznek. Ezenkívül további kutatások alapján összefoglalom, hogy milyen gyakori kihívásokkal kell szembenéznünk egy ilyen rendszer tervezése és megvalósítása során, valamint kitérek a feladat fő indokoltságára és motivációjára is.

Ezután röviden ismertetem a feladat megközelítését és a kutatás során alkalmazott munkamódszereket.

Ezt követően bemutatom, hogyan lehet gépi látáson alapuló módszerekkel pózokat megkülönböztetni, valamint, hogy ehhez milyen adatokra van szükségünk és azok hogyan előállíthatók.

A következő fejezetben ismertetem, hogyan lehet a pózdetektálást olyan algoritmusokkal és módszerekkel kiegészíteni, amelyek segítségével már összetett mozgásszekvenciák azonosítására is képesek lehetünk. Ugyanebben a fejezetben kitérek a leggyakoribb problémákra és azok lehetséges kiküszöbölésére.

A 6. fejezetben bemutatok egy olyan prototípust, amely az említett megoldásokat alkalmazva bebizonyítja, hogy azok a gyakorlatban is jól használhatók.

Végül összefoglalom az elért eredményeket, értékelem a megoldást, valamint bemutatok néhány továbbfejlesztési lehetőséget a jövőre tekintettel.

2. fejezet

Kapcsolódó kutatások, különböző felismerési módszerek

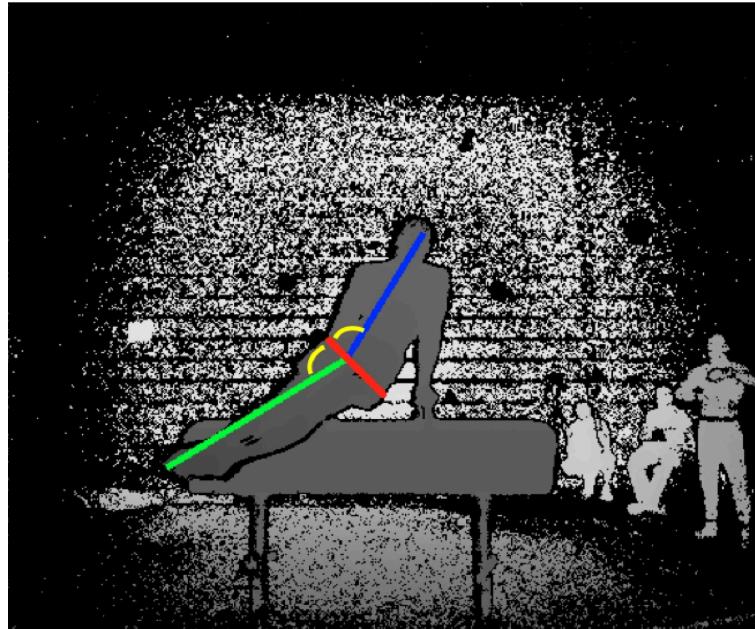
A pózokból összeálló dinamikus mozgások felismerésére már léteznek megoldások, viszont ezek túlnyomó többsége olyan speciális felszerelést és szenzort használ, amelyek nem találhatók meg minden háztartásban. Emiatt az ilyen elemzések használata leginkább az élsportban terjedt el, a hétköznapi ember számára még nem készült hatékony megvalósítás. Ebben a fejezetben ismertetem az adott problémakör legelterjedtebb módszereit és megoldásait, valamint kitérek arra, hogy ezek milyen előnyökkel és hátrányokkal rendelkeznek.

2.1. Lólengés gyakorlatok valós idejű észlelése és teljesítményelemzése

A 3D kamerák megjelenésének köszönhetően számos új megközelítés jelent meg a mozdulatsorok elemzéséhez, mint például a lólengés gyakorlatok felismerésére és automatikus kiértékelésére, különösképpen az időzítésre és a testszögek konzisztenciájára vonatkozóan. Ezen paraméterek elemzése egy Microsoft Kinect 2 kamera segítségével történik, amely képes arra, hogy a kép minden egyes pixelének mélységét meghatározza. A gyakorlatok elemzése három fő lépésből áll: [1]

1. A kamera segítségével kinyert mélységkép feldolgozásra kerül úgy, hogy olyan mélységtartományokat keres, amelyeken valószínűsíthetően a tornász és a lólengés található. Ezt követően egy úgynevezett Depth of Interest (DOI) Segmentation segítségével eltávolítják a kép irreleváns területeit annak érdekében, hogy növeljék az ember felismerésének hatékonyságát, valamint, hogy tisztább képet biztosítson a későbbi elemzések számára.

2. Az előzőleg lokalizált sziluett alapján megtörténik a tevékenységfelismerés, amely meghatározza, hogy a tornász éppen forog-e vagy sem. Erre azért van szükség, hogy ne kerüljenek elemzésre olyan mozzanatok, amikor a személy felszáll a lóra, hiszen ezek elronthatnák a pontos elemzést.



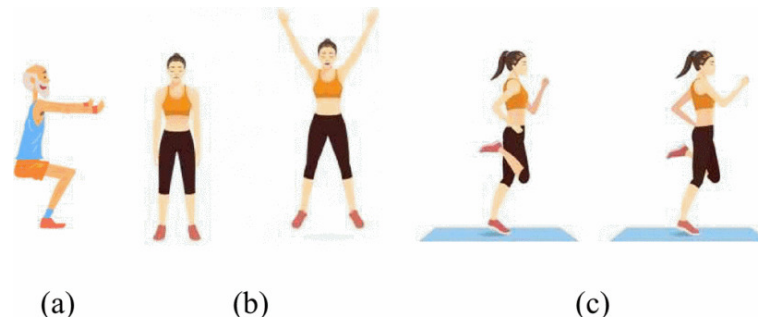
2.1. ábra. Mélységkép a tornász helyzetének követésére használt vektorok illusztrációjával [1]

3. Ha a tornász pörög, akkor az utolsó szakaszban kielemezzük a pörgés sebességét és annak konzisztenciáját. Ezek a mutatók a teljesítmény kvalitatív értékelésére szolgálnak és a tornaedzők számára edzési adatokként is rendelkezésre állnak. A pörgés pontos időbeli mérését a tornász testközéppontjának, valamint a fej- és lábvektorok segítségével határozták meg. Ezek alapján kerül validálásra a tornász testhelyzetének konzisztenciája is.

A megoldás egyik hátránya, hogy kizárólag a mélységi információra támaszkodnak ahelyett, hogy a színadatokat is figyelembe vennék. Ennek oka, hogy ez a megközelítés csupán a mélységi adatok felhasználásával képes nagy pontossággal működni, valamint a Kinect 2 színes adatainak használata sokkal magasabb követelményeket támasztana a hardverrel szemben, ami az edzők és a sportolók számára kevésbé tenné hozzáférhetővé a rendszert. Másik probléma a mélységi kamerás felismeréssel, hogy plusz eszközként be kell szerezni, hiszen még csak kevés hétköznapi készülékben található olyan kamera, amely lehetővé teszi az ilyen elemzések hatékony megvalósítását. Továbbá instruktori megközelítésből az új gyakorlatokkal bővítés is körülményes lehet, mivel a tanítás mélységi kamerás képeken alapul.

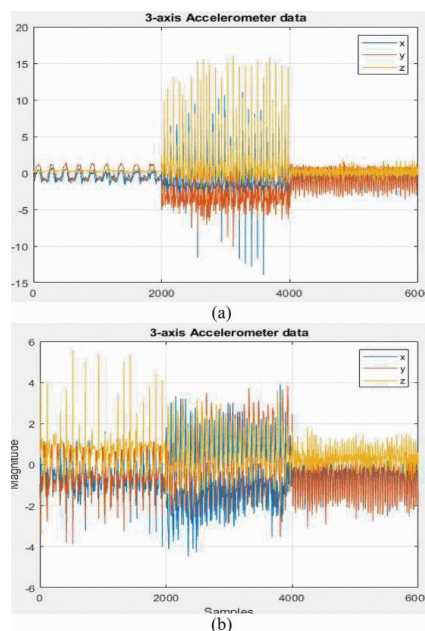
2.2. Gyakorlatfelismerés viselhető szenzorok segítségével

Léteznek olyan megoldások is, amelyek különböző viselhető szenzorok segítségével kísérlik meg felismerni a gyakorlatokat, valamint mérni azok hatékonyságát. Ebben a kutatásban [2] a guggolást, a terpeszugrást és a sarokemelést vizsgálják viselhető 3 tengelyes gyorsulásmérő szenzorral, amelyek a mellkason, a csuklón, a derékon és a bokán helyezkednek el. [2]



2.2. ábra. A kutatás során vizsgált három gyakorlat: (a) guggolás (b) terpeszugrás (c) sarokemelés [2]

A rendszer nyomon követi és elemzi, hogy a cselekvések megfelelnek-e az elvárt végrehajtásnak, mindezt a szenzorokból kinyert adatoknak az ideális eredményekkel való összevetésével, valamint gépi tanuláson alapuló felismerési technikák alkalmazásával.



2.3. ábra. (a) Csuklószenzor segítségével rögzített ideális eredmények az edzőtől és (b) az azt megismétlő résztvevőtől [2]

2.1. táblázat. Pozíciók és gyakorlatok felismerési pontossága [2]

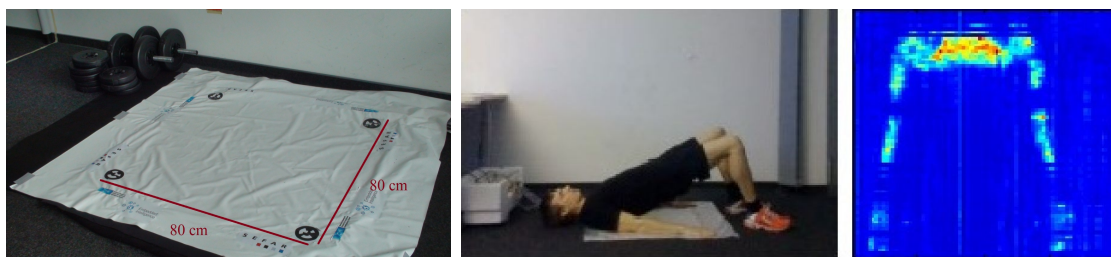
	Guggolás	Terpeszugrás	Sarokemelés
Derék	85,9%	78,3%	71,9%
Csukló	100%	80%	93%
Mellkas	100%	93,33%	84.23%
Boka	98,23%	80%	91,36%

Az olyan megoldásoknak, amelyek egy vagy több különböző szenzor adatain alapulnak a fő hátrányuk, hogy az edzés közbeni viselésük kényelmetlenné teheti és jelentősen megnehezítheti a végrehajtást. A 2.1 táblázat eredményei alapján jól látható, hogy egy gyorsulásmérő szenzorral meglehetősen körülményes a mozdulat pontos beazonosítása és már három gyakorlat esetén is megfigyelhető, hogy azokat időnként tévesen azonosítja. Továbbá a 3 tengelyes gyorsulásmérő adataiból nem feltétlenül meghatározható, hogy egy adott személy pontosan hajtja-e végre a gyakorlatot, például, hogy mekkora terpeszben áll vagy milyen mélyre guggol. A kiugró szenzoradatokból sem könnyen megállapítható, hogy az adott illető esetlegesen rosszul csinálja a gyakorlatot (például a mozdulat során rásegít a karjával), vagy csak a csuklószensor kevésbé szoros felszerelése végett keletkező nagyobb rázkódások okozhatják a kiugrásokat. Az edzők számára is körülményes lehet az újabb gyakorlatokkal való bővítés, hiszen minden esetben szükség lenne a szenzorok segítségével kinyert ideális adatokra, valamint azok átfogó elemzésére, amihez viszonyítani lehetne a végrehajtóét.

2.3. Tornagyakorlatok felismerése és számolása nyomásérzékelő segítségével

Ez a kutatás [3] egy teljesen új és egyedi megközelítéssel képes felismerni olyan tornagyakorlatokat, mint a felülés, ellentétes kar- és lábemelés térdelőtámaszban, hasprés, mellkas nyomás súlyzóval, híd, sarokemelés, guggolás, bicepsz gyakorlat súlyzóval vagy a kitörés. A különböző testrészek összehangolt mozgását igénylő fizikai gyakorlatok nem ismerhetők fel egyetlen testre szerelhető szenzor segítségével, emiatt egy új érzékelő, a rezisztív nyomásszőnyeg létrehozásával oldották meg a feladatot. Ennek előnye, hogy puha, vékony és alacsony költségű, így könnyen helyettesítheti a hagyományos tornaszőnyeget. A felismerés során kihasználják azt, hogy a gyakorlatok végrehajtása során a szőnyegen egyedi nyomásminták keletkeznek, amelyek alapján beazonosíthatóak a mozgások. A kutatásban 10 gyakorlat megkülönböztetését tették lehetővé, amely kNN osztályozás segítségével közel 90%-os pontosságot

ért el. Ezenkívül a rendszer képes az ismétlések számolására is, mindezt kizárólag az "okos" szőnyeg segítségével. [3]



2.4. ábra. Balra a rezisztív nyomásszőnyeg, jobbra pedig a híd gyakorlathoz tartozó nyomásminta [3]

Ennél a kutatásnál felismerték, hogy a bonyolult gyakorlatok magabiztos felismeréséhez több szenzorra lenne szükség, azonban ez rengeteg kényelmetlenséget okozna a végrehajtás során. Emiatt egy teljesen új koncepcióval próbálkoztak, amely szép eredményeket ért el, viszont továbbra is nyitva hagy kérdéseket az új érzékelő beszerzésével, a gyakorlatok könnyű bővíthetőségével, a hasonló mozgások összetévesztésével és a végrehajtott mozdulatok helyességének vizsgálatával kapcsolatban.

2.4. A dolgozat célja

A bemutatott megoldások rávilágítanak arra, hogy több, merőben eltérő megközelítéssel próbálják megoldani a problémát. Mindegyik koncepciónak megvan a maga előnye és hátránya, valamint az adott alkalmazási területe. Számos egyéb kutatást – mint például ritmikus gimnasztika gyakorlatok értékelése videófelvétel vagy videóstream alapján [4], vagy betanított mozgások felismerése és azok reprodukálása egy robottal [5] – is figyelembe véve arra a megállapításra jutottam, hogy a póz- és mozdulatfelismerés során olyan – leggyakrabban előforduló – kihívásokkal nézünk szembe, mint a

- *Speciális felszerelések:* Legtöbbször 3D-, RGBD kamerák vagy különböző gyorsulásmérő szenzorok használatával próbálják felismerni a mozgásokat. Ezek egyrészt költséget jelentenek azok számára, akik el szeretnének kezdeni edzeni, valamint kényelmetlen lehet egyes szenzorok viselése. A legtöbb gyakorlatra jellemző, hogy több testrész összehangolt, egyidejű mozgása szükséges hozzá, amely kizárólag egy szenzor testre szerelésével nem megoldható mivel a karra helyezett szenzor nem érzékeli a láb mozgásának adatait, így a későbbi elemzések lehetővé tételét is nagyban gátolják. Emiatt a különböző testrészeken elhelyezett érzékelők hálózata lenne a legmegfelelőbb, ami viszont nem mindig kivitelezhető és még több kényelmetlenséget okozna. [1, 2, 3, 6, 7, 8]



2.5. ábra. Számos szenzorral ellátott személy [8]

- *Valósídejűség biztosítása:* A rendszerek nagy része a valósídejűséget célozza meg, hiszen a mozgás pillanatában szeretnénk érzékelni a felvett pózokat és minél előbb visszajelzéseket kapni a végrehajtott tevékenységről. Ahhoz, hogy ez megvalósulhasson, másodpercenként rengeteg mintát kell feldolgoznunk nagyon rövid idő alatt, amely komoly kihívásokat támaszthat a megvalósítandó rendszerrel szemben. Az olyan alkalmazások, amelyek nem képesek gyorsan reagálni – például másodpercekkel később szól, hogy nem jól hajtottuk végre a gyakorlatot, vagy csak az 50 fekvőtámasz befejezése után szól, hogy a kezünket nem megfelelő távolságra helyeztük – kimondottan taszítóak lehetnek a felhasználók számára. [1, 4, 5, 6, 9]
- *Több személy mozgásának egyidejű észlelése:* Néhány kutatás célul tűzte ki, hogy ne csak egy ember mozgását lehessen egyidejűleg nyomon követni, hanem akár egy egész edzőterem kamerán keresztűli megfigyelése és elemzése is megvalósítható legyen. Már egy személy mozgásának pontos felismerése és elemzése is kihívást okoz, több személyre skálázva viszont biztosan komoly megpróbáltatást jelent. [9]
- *Váltások nehézkes felismerése:* Szintén egy összetett feladat annak detektálása, hogy egy adott személy mikor kezd másik gyakorlatot csinálni. Például, ha az alany több guggolás után néhányat fekvőtámaszozik, majd ismét elkezdi guggolni. Nagyon fontos, hogy a felismerés viszonylag gyors maradjon, hiszen a túl lassú érzékelés következtében nagyon sok ismétlésről és elemzésről lemaradhat az alany.

- *Hasonló gyakorlatok megkülönböztetése:* A kutatások sokszor merőben eltérő gyakorlatok és pózok felismerését teszik lehetővé, ahol vagy nem mutatkoznak ilyen problémák, vagy ha vannak hasonló gyakorlatok, akkor rendszerint tévesztik őket. Ilyen például a hatütemű fekvőtámaszban is szereplő egy darab karhajlítás (fekvőtámasz), amely önálló gyakorlatként is megállná a helyét. Ezek megkülönböztetése is felvet néhány komoly problémát. [3, 7]
- *Hatékony megvalósíthatóság:* A felismerési módszerek megalkotása során figyelembe kell venni azt, hogy annak megvalósítása ne igényeljen rendkívüli erőforrásokat. Tehát nem előnyös, ha az alkalmazás csak erőforrás-igényes hardveren tud futni, lassan reagál, bizonyos idő után túlmelegszik, vagy nagyon hamar lemerül.

Jól látható, hogy ezen megoldások egyike sem érhető el az átlag ember számára, hiszen valamilyen különleges tudást vagy elvárást támasztanak velünk szemben, ezáltal nem növelik a sportolni és mozogni kívánók körét. Kutatásom célja, hogy olyan megoldást dolgozzak ki, amely olyan megközelítésből vizsgálja a problémát, hogy az a hétköznapi emberek számára elérhető és használható legyen. Ennek során olyan algoritmusok és módszerek kidolgozását kísérelem meg, amelyek a kapcsolódó kutatásokban legtöbbször problémaként, vagy nyitott kérdésként jelennek meg. A célkitűzés, hogy az említett nehézségek döntő többségét, mint

- a speciális felszerelések mellőzését,
- a valósidejűséget,
- a hasonló gyakorlatok megkülönböztethetőségét,
- a hatékony megvalósítást,
- valamint a további mozdulatokkal való egyszerű bővítést

lehetővé tegyem.

3. fejezet

Megközelítés

A 2. fejezetben bemutatott megoldások rávilágítanak arra, hogy számos különböző módon igyekeznek megközelíteni ezt a problémakört.

A megoldásom során igyekszem kihasználni azt a tényt, hogy a legtöbb gyakorlat jól elkülöníthető pózokból épül fel. Vegyünk egy egyszerű fekvőtámaszt, ahol két testhelyzet folyamatos váltakozása alkotja az adott mozdulatsort: az egyenes karral kinyomott és a derékszögben behajlított karokkal végzett karhajlítás pozíció. A két állapot között nyilvánvalóan van egy átmenet, amelynek során az egyik helyzetből a másikba kerül.

Ezt a megfigyelést felhasználva először létrehozok egy mesterséges intelligencia alapú pózfelismerőt, amely nagy pontossággal képes megkülönböztetni az eltérő testhelyzeteket. Ennek kivitelezéséhez adatokat kell gyűjtenem, amelyek lehetővé teszik a tanítást és a felismerés pontosságának mérését.

Amennyiben a pózfelismerés sikeresnek és hatékonnak bizonyul, úgy a betanított testhelyzetek lehetőséget nyújtanak, hogy azokból mozgásszekvenciákat alkossak. Ezek meghatározása után megkísérlem olyan algoritmus elkészítését, amely képes a referenciapontok alapján detektálni a pózokat és azok nyilvántartásával beazonosítani a végrehajtott gyakorlatokat. A módszer meghatározása során megvizsgálom a főbb nehézségeket, valamint egy prototípus elkészítésével ellenőrzéseket végzek annak hatékonyságára.

4. fejezet

Pózfelismerés

A dinamikus mozgások felismeréséhez szükség van arra, hogy az alany mozgása során felvett pózokat beazonosítsuk és azok szabályszerű egymásutánosságából következtetni tudjunk, hogy pontosan milyen mozgásformát hajtanak végre. A pózfelismerés már régóta aktív területét képzi a gépi látásnak, mely lehetővé teszi, hogy nyomon követhessük az alany mozgásának változásait a testére helyezett referenciapontok segítségével. Ebben a szakaszban bemutatom az általam készített TFLite modellt, amely az emberi test referenciapontjai alapján meghatározza, hogy a személy melyik előre betanított pozíciót veszi fel. A tanítási folyamat három fő részből áll:

1. A képek előfeldolgozása során egy MoveNet TFLite modell segítségével (lásd 4.2. fejezet) felismeri az emberi test 17 referenciapontját, majd ezek x és y koordinátáit egy táblázatba összegyűjti.
2. Létrehozza és betanítja a modellt (lásd 4.3. fejezet), ami bemenetként megkapja a felismert koordináta információkat, a kimenete pedig az adott pózok valószínűsége.
3. A létrejött modellt TFLite modellé konvertálok, ami a későbbiekben könnyen felhasználható.

4.1. Adathalmaz kialakítása

Két alternatíva volt, vagy képek, vagy referenciapontok alapján tanítom be a modellt. Előbbinek a hátránya, hogy a modell jóval nagyobb méretű lehet (ami például egy telefonon nem túl ideális), valamint sokkal nehezebb és erőforrás-igényesebb képekkel dolgozni. Ebből kifolyólag választottam a referenciapont alapú tanítást. A tanításhoz szükséges kép- vagy referenciapont adatokhoz nem létezik publikusan elérhető adathalmaz, ami gyakorlatok pózait tartalmazná, emiatt kénytelen voltam ezt saját magam létrehozni, amit kétféleképpen kíséreltem meg:

- *Referenciapontok alapján:* Először megpróbáltam gyakorlóvideókból kinyerni az adott póz referenciapontjait, azonban ez nem bizonyult elég hatékonynak. A fő problémát az okozta, hogy az adatkészletet nehezen lehetett sokszínűvé alakítani, például, hogy a testhelyzetet különböző pozíciókból (szemből, oldalról, hátulról), vagy eltérő módon (más mértékű guggolások stb.) ábrázolja, valamint a kinyert adatokat nehezen lehetett validálni, hogy tényleg megfelelően írja-e le a pózt.
- *Képek alapján:* A következő próbálkozás során nagy mennyiségű stock fotót gyűjtöttem minden testhelyzetre. Ennek hatalmas előnye, hogy egyszerűen kinyerhetők belőle a referenciapont információk, amelyek könnyen validálhatók a pontháló képen való ábrázolásával, valamint meglehetősen sokszínű adathalmaz kialakítását teszi lehetővé.

Az említett fennakadások miatt végül a képalapú tanítást választottam, ahol 8 pozíció betanítását tettem lehetővé több mint 1000 kép segítségével:

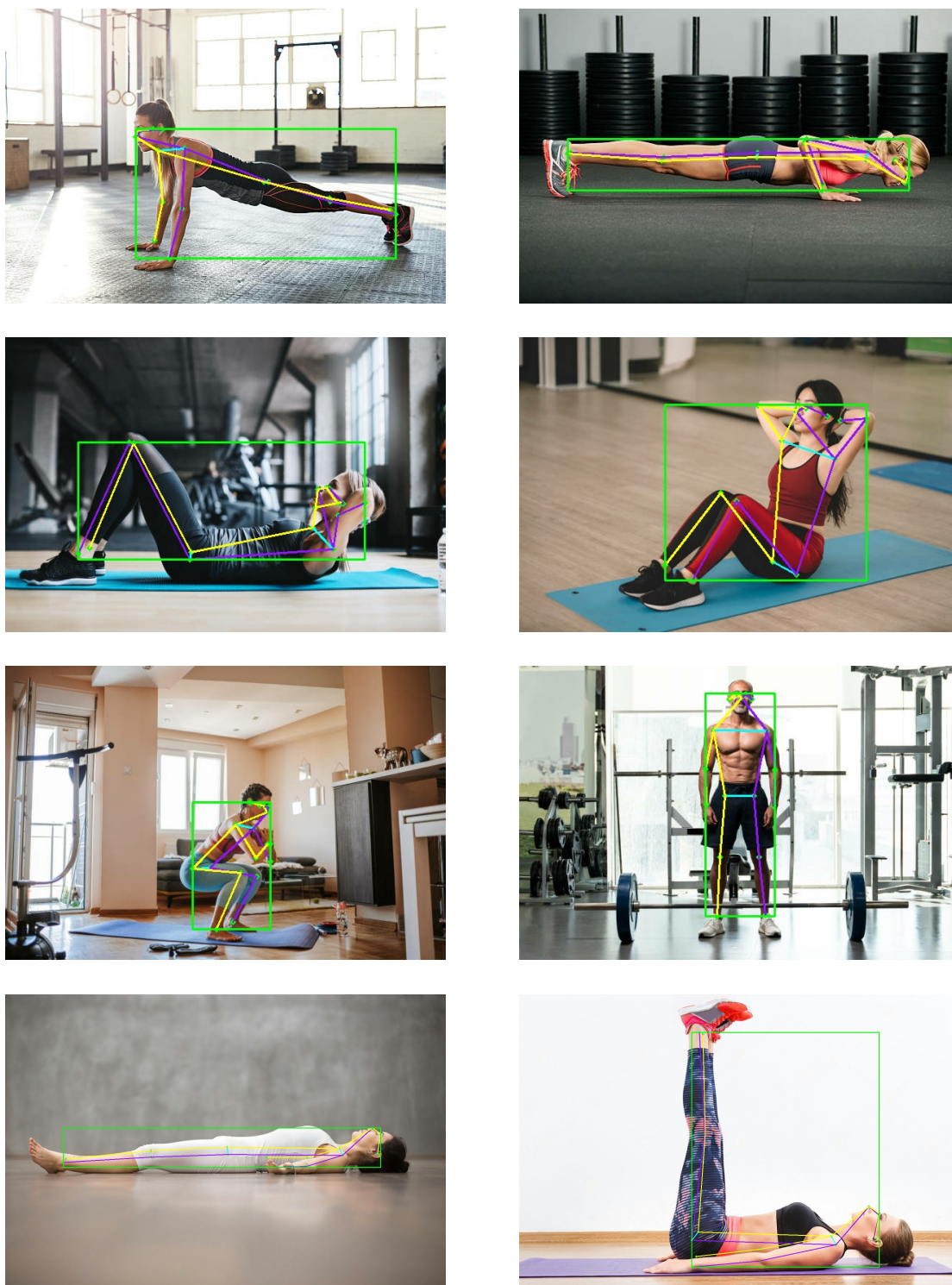
- 121 db fekvőtámasz pozíció, karhajlítással (lenti) – `pushup_down`
- 215 db fekvőtámasz pozíció, egyenes karral kinyomva (fenti) – `pushup_up`
- 93 db felülés pozíció, felhúzott lábakkal fekvés (lenti) – `situp_down`
- 142 db felülés pozíció, felhúzott lábakkal felemelkedés (fenti) – `situp_up`
- 323 db guggolás pozíció – `squat`
- 63 db álló pozíció – `standing`
- 54 db hanyattfekvés pozíció – `lie`
- 50 db lábemelés pozíció – `raised_leg`

4.2. MoveNet: Ultragyors és pontos pózfelismerő modell

A MoveNet egy ultragyors és pontos pózfelismerő modell, ami képes az emberi test 17 fő kulcspontját beazonosítani. Két változata érhető el számunkra, az úgynevezett **Lightning**, amely a késleltetéskritikus rendszerekhez ajánlott, valamint a **Thunder**, ami pedig a nagy pontosságot igénylő alkalmazásokhoz készült. Mindkét modell a valós időnél gyorsabban fut, azaz 30+ FPS¹-en képes működni, ami döntő fontosságú az élő fitness, sport- és egészségügyi alkalmazások számára. Míg a legtöbb

¹Frames per second: Képkocka/másodperc

pózfelismerő képes könnyen beazonosítani az olyan egyszerű mozgásokat, mint a guggolás vagy a karok emelése, addig problémába ütköznek a fekvő pozíció vagy az ülés közben végzett térdhajlítás esetében, ahol nem feltétlenül tudják a lábak helyzetét egyértelműen megkülönböztetni. [10]

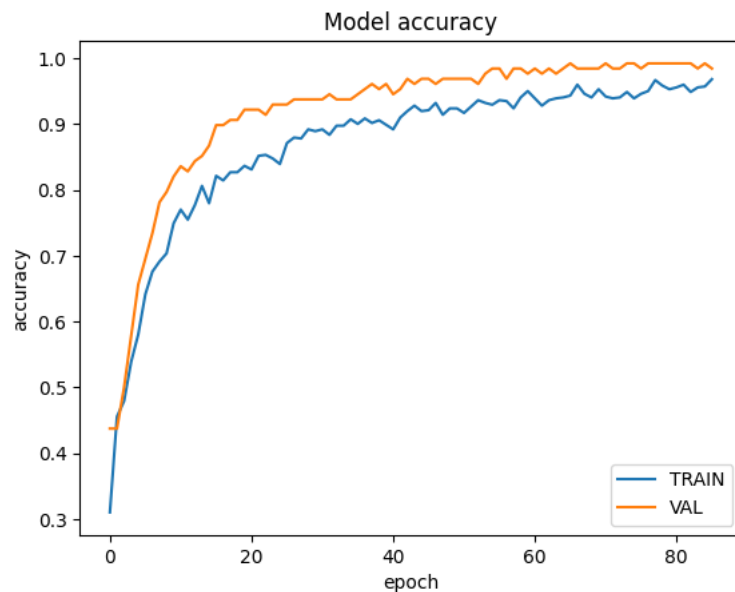


4.1. ábra. Az adathalmaz képein futtatott referenciapont felismerés eredménye [11, 12, 13, 14, 15, 16, 17, 18]

4.3. Tanítás

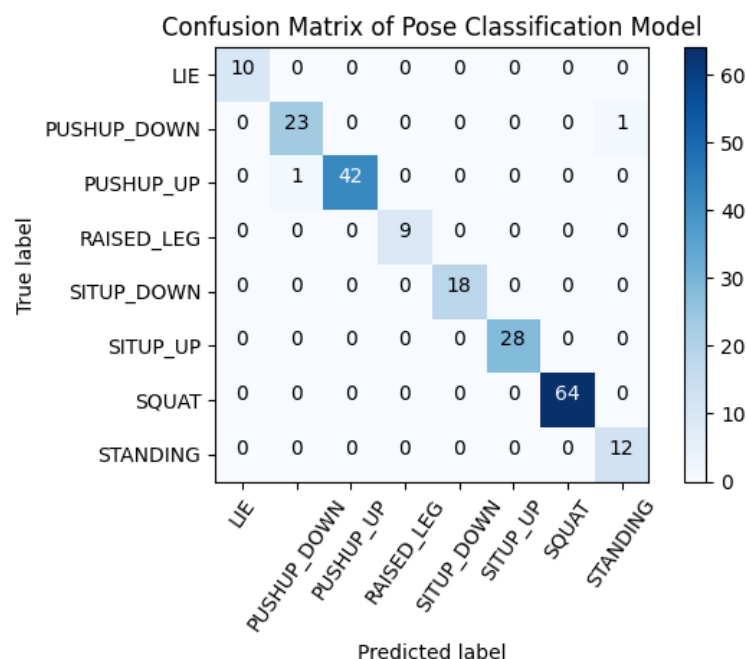
A tanítás során létrehoztam a modellt, ami a referenciapontok koordinátáit alapul véve megjósolja, hogy a bemeneti képen szereplő személy milyen pózt vesz fel. A modell két almodelltől áll:

1. Az első almodell kiszámítja az úgynevezett tulajdonság vektorokat (feature vector), amely leírja az adott póz tulajdonságait a referenciapontok alapján. A folyamat során meghatározom az adott testhelyzet középpontját és eltolom a (0,0) pontba, majd normalizálom a pózt annak érdekében, hogy a pózméret konstans 1 értéket vegyen fel. Ha mindezeket végrehajtotta, akkor a különálló koordinátákat egyszerűen tulajdonság vektorra alakítom.
2. Az előzőekben elkészített tulajdonság vektorok felhasználásával több Dense rétegen keresztül betanítom a neurális hálózatot, majd elkészítem a modellt.



4.2. ábra. A modell pontossága a tanítási folyamat során

A tanító- és teszhalmazt az eredeti adathalmaz 80-20%-os felosztásával határoztam meg, valamint a tanító halmaz 15%-át használom validációs halmaznak. A 4.2. ábra alapján megállapítható, hogy a modellnek sikerült megfelelően rátanulnia a testhelyzetekre, amelyet a teszhalmazzal ellenőrizve 99,04%-os pontosságot ért el.



4.3. ábra. A pózfelismerő modell konfúziós (tévesztési) mátrixa

A 4.3. ábra bemutatja, hogy a teszhalmazon futtatott ellenőrzés során hányszor sikerült beazonosítani a megfelelő pózokat, vagy ha nem, akkor mivel és hányszor tévesztette össze. Az ilyen esetekhez tartozó kép – valamint a rárajzolt referenciapont háló – alapján könnyen ellenőrizni tudom, hogy az emberi test felismerése és a kép besorolása megfelelően történt-e.

4.1. táblázat. A tanítás leggyakoribb kiértékelési metrikái, mint a precizitás, a felidézés vagy a pontosság

	precision	recall	f1-score	support
lie	1,00	1,00	1,00	10
pushup_down	0,96	0,96	0,96	24
pushup_up	1,00	1,00	1,00	43
raised_leg	1,00	1,00	1,00	9
situp_down	1,00	1,00	1,00	18
situp_up	1,00	1,00	1,00	28
squat	1,00	1,00	1,00	64
standing	0,92	1,00	0,96	12
accuracy			0,99	208
macro avg	0,99	0,99	0,99	208
weighted avg	0,99	0,99	0,99	208

5. fejezet

Mozgásszekvenciák meghatározása

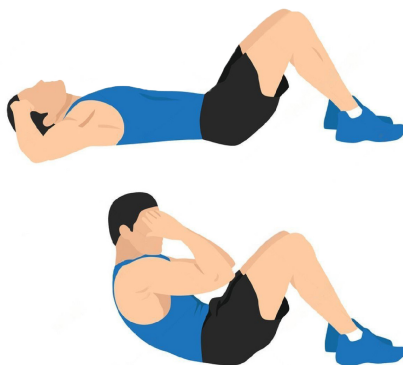
A kapcsolódó munkák során nyilvánvalóvá vált, hogy a tématerületet kutatásának egyik nagy kihívása a mozgásszekvenciák megfelelő azonosítása és kódolása. Ebben a fejezetben bemutatásra kerül hogyan lehet a mesterséges intelligencián alapuló pózfelismerést olyan további algoritmusokkal kiegészíteni, amelyek lehetővé teszik, hogy a pózok váltakozásai alapján detektálni lehessen, hogy az alany milyen mozgólatsort, tornagyakorlatot hajt végre, mindezt speciális eszközök használata nélkül. Ezáltal a gyakorlatfelismerés már nem csak az élsportolóknak és a speciális felszerelésekkel rendelkező személyeknek lesz elérhető, hanem bárki számára. Ahhoz, hogy a mozgások beazonosíthatók legyenek, meg kellett határoznom különböző, a felismerhető pózokból felépülő mozgólatsorokat. A 4.1. ábrán lévő pozíciókból 6 gyakorlatot határoztam meg, amelyek a következők:

- **Guggolás:** squat, standing



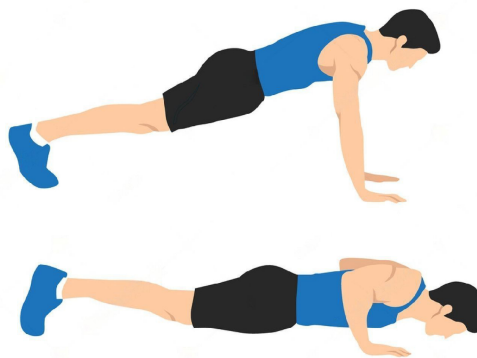
5.1. ábra. A guggolás illusztrációja [19]

- **Felülés:** situp_up, situp_down



5.2. ábra. A felülés illusztrációja [20]

- **Fekvőtámasz:** pushup_down, pushup_up



5.3. ábra. A fekvőtámasz illusztrációja [21]

- **Négyütemű fekvőtámasz:** squat, pushup_up, squat, standing



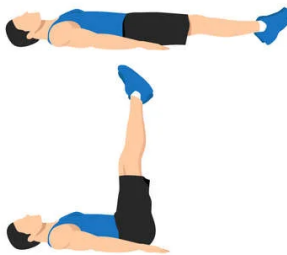
5.4. ábra. A négyütemű fekvőtámasz illusztrációja [22]

- **Hatütemű fekvőtámasz:** squat, pushup_up, pushup_down, pushup_up, squat, standing



5.5. ábra. A hatütemű fekvőtámasz illusztrációja [23]

- **Lábemelés:** raised_leg, lie



5.6. ábra. A lábemelés illusztrációja [24]

A több pőzból felépülő mozgásszekvenciák felismeréséhez szükséges, hogy folyamatosan nyilvántartsuk a 4. fejezetben bemutatott módszerrel meghatározott testhelyzeteket. Ennek egy lehetséges módja, hogy egy listában kerülnek tárolásra a felismert pőzíciók. Fontos, hogy másodpercenként akár 30-45 pőzfelismerés is történhet, amely során a testhelyzet nem feltétlenül változik. Mivel egy olyan listából nehézkes lenne dolgozni, amely minden képkockához tárolja az adott pillanatban felvett pőzitúrát és mivel az adatok ilyen részletességű tárolására nincs szükségem, így csak akkor tárolom el az új pőzt, ha az az utolsóhoz képest megváltozott. Nyilvánvaló, hogy ha az illető nem váltott pőzítóit, akkor a gyakorlat sem változhatott meg. Abban az esetben, ha megváltozott a pőz, akkor hozzáadjuk a listához, így az egymást követő elemek mindig eltérőek lesznek.

Példa: 2 guggolást és 1 fekvőtámaszt végrehajtva a várt eredmény:

1. squat
2. standing
3. squat
4. standing
5. pushup_down
6. pushup_up

Az algoritmus, amely a póztörténetből meghatározza, hogy melyik gyakorlat végrehajtása történik, egy egyszerű számláláson alapul. Meghatározom, hogy melyik gyakorlatból hányat hajtottak végre, majd kiválasztom a legnagyobb ismétlésszámút. Ebben az esetben egyértelműen látszódik, hogy a 2 guggolás és az 1 fekvőtámasz közül automatikusan a guggolást választaná, hiszen abból eggyel többet végzett. Ez a módszer azonban több lehetséges problémát is felvet, amelyek a tématerület irodalmában is gyakran előkerülnek.

1. **probléma:** Mi történik akkor, ha a póztörténet üres, vagy csak minimális adatot tartalmaz?

Példa: Az összes gyakorlatból felismer egy-egy darabot.

Megoldást célzó hipotézis: Ennek a problémának a megoldása egyszerű. Amíg nem lehet egyértelműen eldönteni, hogy pontosan melyik gyakorlatot hajtják végre – mert mindegyikből talál 1 db egyezést – addig nem választunk. Az ilyet érdemes úgy kezelni, hogy kijelölünk egy minimálisan meghatározott ismétlésszámot, amiből a minta már kétségtelenül felismerhető. Ennek az értékét alaposan át kell gondolni, mivel, ha túl alacsonyra választjuk, akkor előfordulhat, hogy túl sok gyakorlatot ismerünk fel egyszerre. Ezzel szemben, ha túl magas ismétlésszámot várunk el, akkor az illető csak későn kap pontos visszajelzést arról, hogy felismertük a mozgását. Tapasztalati méréseim alapján érdemes 2-3 ismétlésre korlátozni, amelyből már határozottan elkülöníthető a minta.

2. **probléma:** Hogyan észlelhető, hogy a tornázó személy elkezd egy másik gyakorlatot csinálni?

Példa: 20 fekvőtámasz után elkezdünk guggolásokat végrehajtani.

Megoldást célzó hipotézis: Erre nyújt megoldást a csúszóablak bevezetése, aminek lényege, hogy a mozdulatsor meghatározása során nem a teljes előzménysorozatot vesszük figyelembe, hanem kizárólag egy bizonyos időintervallumot. Ennek oka nyilvánvalóan az, hogy a nagyon régi adatok már irrelevánsak számunkra, hiszen folyamatosan belekezdhet újabbnál-újabb mozgásokba. A gyors észlelés érdekében érdemes az időablakot alacsonyan tartani. Azonban hasonlóan az előző problémához, ha túl alacsonyra választjuk, akkor nagyon hamar kicsúsznak az ablakból az értékek és túl kevés adat alapján kell meghatároznunk a gyakorlatot. Ha viszont túl magas időablakot szabunk meg, akkor pedig nagyon lassú marad a reakcióidő. A csúszóablak idejének megfelelő meghatározására az előzőleg kiválasztott minimum ismétlésszám is hatással van, mivel csak akkor felismerhető a gyakorlat, ha az időablakon belül elvégezhető a kívánt ismétlésszám. Emiatt érdemes úgy meghatározni az ablak idejét, hogy a leghosszabb gyakorlat minimum végrehajtásánál 2-3-szor nagyobb legyen. Az említett gyakorlatok esetén egy szintén tapasztalati – viszonylag nagy – 20 másodperces időablak megfelelőnek bizonyult. Annak érdekében, hogy az időablak alkalmazható legyen, a korábban említett pózelözményeket ki kell egészíteni egy eltárolt dátummal is, amikor a felismerés történt.

3. **probléma:** Hogyan ismerhetők fel az olyan bonyolultabb gyakorlatok is, amelyek magukba foglalnak más gyakorlatokra illeszkedő mozdulatsorokat is?

Példa: A négy- és hatütemű fekvőtámaszban is van guggolás, valamint a hatütemű fekvőtámaszban szerepel még az egyszerű fekvőtámasz is.

Megoldást célzó hipotézis: Az összetettebb gyakorlatokat sokkal nehezebb felismerni, mint az abban lévő esetleges részgyakorlatokat. Ennek oka, hogy a könnyebb, kevesebb pózból álló mozdulatoknál kisebb az esély, hogy egy hibás póz azonosítása miatt drasztikusan csökkenne a meghatározott ismétlésszám. Ezt a problémát úgy lehet áthidalni, hogy a gyakorlatok ismétlésszám alapú felismerését kiegészítjük még egy komplexitás paraméterrel is, melynek lényege, hogy egy több pózból álló mozdulatsorhoz nagyobb komplexitást rendeljünk, mint egy egyszerűbbhöz.

Példa: Legyen 4 felismerhető gyakorlat:

- Guggolás: **squat, standing**
- Fekvőtámasz: **pushup_down, pushup_up**
- Négyütemű fekvőtámasz: **squat, pushup_up, squat, standing**
- Hatütemű fekvőtámasz: **squat, pushup_up, pushup_down, pushup_up, squat, standing**

Ha végrehajtanak 3 ismétlést a hatütemű fekvőtámaszból, akkor ezt eltároljuk a pózelőzmények listájában. Ezt követően megvizsgáljuk, hogy hány guggolást, hány fekvőtámaszt és hány hatütemű fekvőtámaszt hajtottak végre, az eredmény pedig mindegyik esetben 3 lesz, azaz nem egyértelműen eldönthető. Emiatt minden gyakorlathoz egy egyedi pontszámot rendelék, amely kezeli ezt a problémát:

$$\text{Pontszám} = \text{Ismétlésszám} * \text{Komplexitás}$$

ahol az ismétlésszám a pózelőzményeket és a megadott csúszóablakot figyelembe vett ismétlések száma, a komplexitás pedig az adott gyakorlatot alkotó pózok száma. Miután minden gyakorlatnál megállapítottam a pontszámot, kiválasztom a legnagyobb rendelkezőt. Ezáltal az összetettebb gyakorlatok nyilvánvalóan előnyt élveznek még akkor is, ha abból esetleg néhányal kevesebb ismétlésszám történt.

4. probléma: Előfordulhatnak-e téves pózfelismerések és ezek hogyan szűrhetők?

Megoldást célzó hipotézis: Nyilvánvalóan előfordulhatnak téves pózfelismerések több okból kifolyólag is. Az egyik, hogy nem készíthető tökéletes modell, amely minden testhelyzetet mindig helyesen azonosít, ezáltal csúszhatnak be hibák. Legfőképpen akkor, ha két póz közötti átmenet (például guggolásból fekvőtámasz pozícióba váltás) során felvett pozíció egyik pózra sem hasonlít vagy esetleg egy létező másik pózt ismer fel. Mivel valósidejű elemzésről beszélünk, ahol másodpercenként 30-nál is több felismerést hajthatunk végre, így mindenképp számolnunk kell ezzel a problémával. Ezt kétféleképpen szűrhetjük:

- A nagyon rövid ideig megjelenő pózok könnyen bezavarhatnak az ismétlésszám meghatározásánál, ezért őket úgy kezeljük, mintha meg sem történtek volna. A gyakorlatok során felvett pozíciók viszonylag hosszabb ideig állnak fent, így időtartam alapján kiszűrhetők a hamis pózok. Szintén tapasztalati mérések alapján a 200-400 milliszekundumnál rövidebb

ideig azonosított pózok általában ilyenek. Ennek fényében az algoritmus úgy módosítható, hogy csak akkor fűzünk hozzá a pózelőzményekhez, ha az új póz legalább egy minimális ideig fennállt.

- Még mindig előfordulhatnak olyan esetek, hogy két pozíció közötti átmenet során hosszabb ideig azonosít egy téves pózt, vagy esetleg a személy fáradtan a földön feküdve pihen, emiatt az átmegy a szűrőn. Az ilyet úgy tudjuk kiküszöbölni, hogy az ismétlésszámon és a komplexitáson alapuló pontszámok kiszámításánál nem a teljes pózelőzményt vesszük figyelembe, hanem előtte kiszűrjük azokat a pózokat, amelyek nem találhatók meg az adott gyakorlatban. Így egy 6-8 pózból álló gyakorlat során bekövetkező egy-egy téves póz nem rontja el a meghatározást. A plusz szűrés hozzáadása okán is keletkezhetnek téves felismerések, viszont ezek kis valószínűségűek és a komplexitással súlyozva jelentéktelennek tekinthetők.

6. fejezet

Prototípus megvalósítása

Készítettem egy prototípust, hogy az algoritmusok működését és hatékonyságát a gyakorlatban is ellenőrizni tudjam. Egy olyan Android szoftvercsomagot valósítottam meg, amely az eszköz kameráján keresztül képes felismerni a különböző gyakorlatokat, valamint az instruktor által bemutatott videó elemzése után meg tudja állapítani, hogy a gyakorlatot a saját eszközének kameráján át felvevő alany mennyire pontosan tudja leképezni azt. Az adatokat – mint a mozdulatsorok és az azokat alkotó pózokat – egy MySQL adatbázisban tárolom. Ezek egy Kotlin Spring alapú REST API-n keresztül kerülnek kiszolgálásra. Az adatok feltöltéséhez pedig készítettem egy React admin felületet, ahol a gyakorlatok menedzselhetők.

6.1. Kamera használata

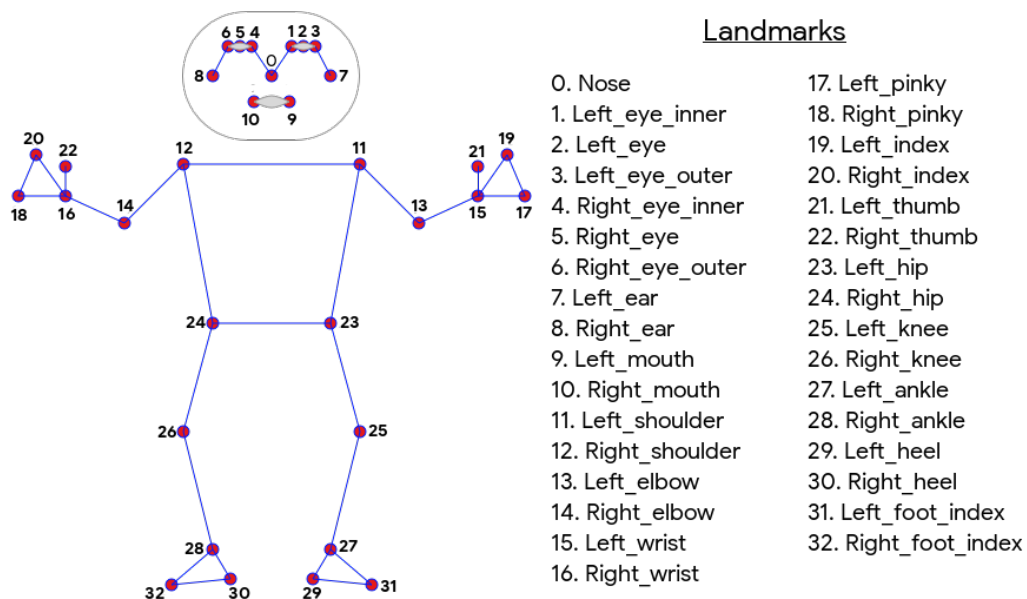
Az alkalmazás alapját a kamera használata jelenti, amin keresztül a felhasználó mozgását detektálom, így annak bekötése egy nélkülözhetetlen folyamat volt. A kamera használata Androidon engedélyhez kötött, így először azt futásidőben el kell kérni a felhasználótól. Ennek hiányában az alkalmazás nem képes elemezni a gyakorlatokat, hiszen nem látja az alanyt. Készítettem az alkalmazásba egy beépített kamera nézetet a CameraX könyvtár segítségével, amely a kamera kezelésével kapcsolatos műveletek fejlesztését könnyíti meg, mindezt életciklus-tudatos és használati eseteken alapuló megközelítésekkel. A könyvtár olyan használati eseteket valósít meg, mint az [25]

- Előnézet: Előnézeti képet jeleníti meg az alkalmazásban.
- Képelemzés: Támogatja, hogy az előállított képeken különböző analíziseket hajtsunk végre, mint például az ML Kit segítségével való póz detektálás.
- Képkészítés: Biztosítja számunkra a képek jó minőségben való rögzítésének lehetőségét memóriába vagy fájlba.

6.2. Referenciapont felismerés

A mai modern világban az informatika egyre több területén van szükségünk különböző mesterséges intelligencia és gépi tanulás alapú megoldásokra. Az ML Kit egy olyan SDK, amelyet a Google hozott létre, annak érdekében, hogy a leggyakoribb gépi tanulásos elemeket beépíthessünk Android és iOS alapú alkalmazásainkba. [26]

Az ML Kit Pose Detection API egy könnyű és sokoldalú megoldás az alkalmazásfejlesztők számára, hogy akár egy folyamatos videófolyamon, akár egy statikus képen valós időben észleljük a felvett pozíciót. Mindezt úgy éri el, hogy minden időpillanatban különböző referenciapontokat helyez el az alany testén, amelyek a különböző testrészeket jelölik, mint például a vállakat vagy a csípőt. A referenciapontok relatív helyzete alapján pedig megkülönböztethetjük az alany által felvett pózokat. Az API 33 referenciapontot biztosít számunkra, amellyel az emberek legfontosabb testrészeit lefedi. Minden referenciaponthoz meghatározható, hogy mekkora valószínűséggel esik bele a képbe, valamint elérhető hozzájuk a képen meghatározott X és Y koordinátájuk. Z koordinátával is rendelkezik, amellyel a csípőtől való irányt határozhatjuk meg. Ezekből az adatokból könnyen számolhatunk bizonyos távolságokat, valamint szögeket. [27]



6.1. ábra. A biztosított referenciapontok listája [27]

A MoveNethez hasonlóan az ML Kit Pose Detection is kétféle megoldást kínál számunkra, egy gyorsat és egy pontosat. A prototípus jelenleg a pontos (accurate) verziót használja, mivel ez is másodpercenként 20-30 felismerést képes végrehajtani, amely a célnak bőven megfelel.

6.1. táblázat. Pózfelismerő könyvtárak összehasonlítása [27]

Könyvtár neve	pose-detection	pose-detection-accurate
Implementáció	A kódok és eszközök fordítási időben kapcsolódnak az alkalmazáshoz.	A kódok és eszközök fordítási időben kapcsolódnak az alkalmazáshoz.
Alkalmazás méretének növekedése	~11 MB	~14 MB
Teljesítmény	~30 FPS	~23 FPS (CPU) ~30 FPS (GPU)

Ahhoz, hogy a kameraképen szereplő póz pontosan felismerhető legyen, a könyvtárnak minimum 480x360 pixel méretű képre van szüksége, amelyen a személynek legalább 256x256 pixel nagyságúnak kell lennie. A megvalósításnál ügyelnem kellett arra, hogy a képméretet minél alacsonyabban tartsam, hiszen a nagyobb képek feldolgozása tovább tarthat, amely növelné a késleltetést. Az alkalmazásban jelenleg 1280x720 pixel méretű képeket használok, amely megfelelőnek bizonyult.

6.3. Gyakorlatok felismerése és elemzése

A prototípusban nem csak a gyakorlatok felismerését, hanem azok értékelésére is készítettem egy kezdetleges megoldást. A teljes gyakorlatfelismerési és elemzési folyamat 4 fő részből áll:

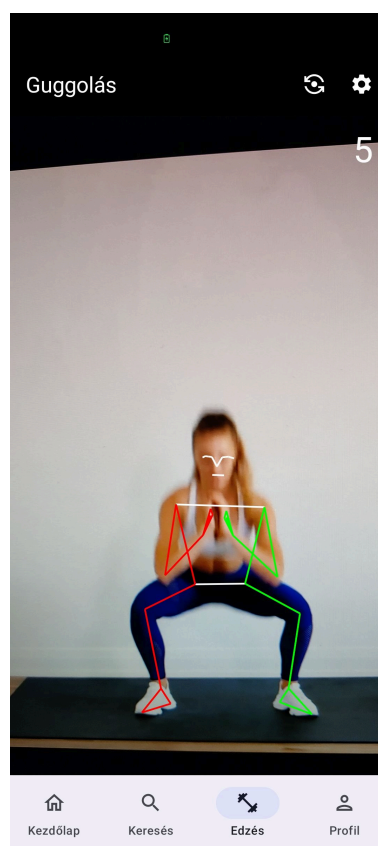
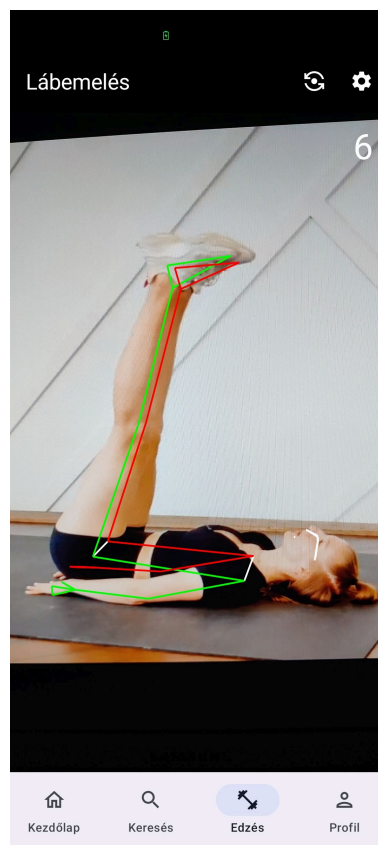
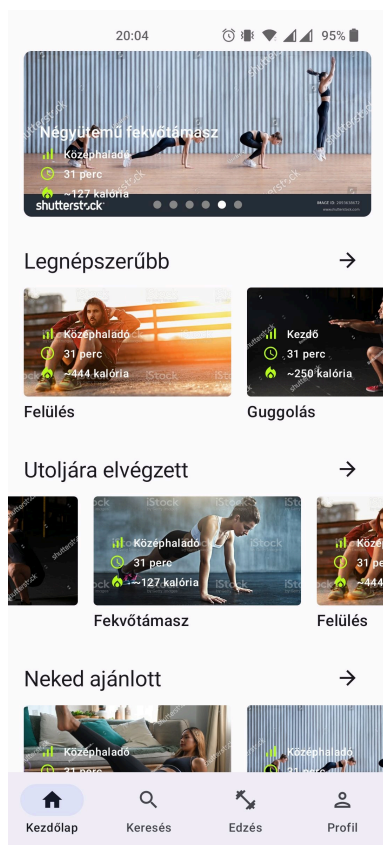
1. *A póz felismerése:* Az ML Kit Pose Detection segítségével kinyert referencia-pontokkal meghívom az alkalmazáshoz hozzáadott – a korábban bemutatott és elkészített – TFLite modellt, amely meghatározza az egyes pózok valószínűségeit, majd kiválasztom a legnagyobbat közülük.
2. *A gyakorlat felismerése:* A 5. fejezetben bemutatott módszerrel folyamatosan nyilvántartom a pózelőzményeket, valamint annak változása esetén meghatározom, hogy melyik gyakorlatot hajtják végre.
3. *A gyakorlat ismétlésszámának meghatározása:* Mivel a mozdulatsorok felismerése az ismétlésszám meghatározásán alapul, így ennek nyilvántartása csak kisebb módosításokat igényel. Ha sikeresen felismerhető volt egy gyakorlat, akkor a felismeréstől kezdve a pózelőzmények figyelembevételével számolhatók az ismétlésszámok. Fontos kiemelni, hogy ez a megoldás a gyakorlat felismerésétől kezdve számol, tehát azt nem kalkulálja hozzá, hogy a felismeréshez hány ismétlés kellett. Ennek oka, hogy a felismerés folyamatos csúszóablakkal működik és abban vizsgálódik.

Példa: A csúszóablakba belefért 5 guggolás, 5 fekvőtámasz, majd újra 5 guggolás.

A felismerés során a guggolást választja, hiszen abból 10 ismétlést hajtott végre, viszont a legutolsó gyakorlatfelismerés óta csak 5 ismétlést végzett. Az ilyen és hasonló esetek kiszűrése kisebb bonyolultságot okoz, amely a prototípus lényegét nem befolyásolja.

4. *A gyakorlat tempójának mérése:* Arra is kíváncsi voltam, hogy idővel hogyan alakul a gyakorlat végzésének tempója, hiszen amikor az ember már fárad, az könnyen lelassulhat és elvesztheti hatékonyságát. Mivel az ismétlések számolásánál pontosan rendelkezésemre áll, hogy mikor kezd el az új ismétlést, így jelenleg itt kerül mérésre a tempó is. Egyszerűen csak az ismétlésszám növekedésnél elmentem az időbélyeget, amikből később kiszámolhatjuk a tempót, akár egy hasonló csúszóablakos megoldással, a kiugró értékek kiszűréssel együtt.

A gyakorlatok végrehajtása során igyekeztem visszajelzéseket adni a végzés helyességéről is. Ezt úgy kell elképzelni, hogy ha guggolás közben nem áll vállszéles terpeszben, vagy ha ugrás közben nem emeli fel a kezét, akkor erre figyelmeztetjük. Ehhez először szükség volt arra, hogy az admin felületen keresztül lehetővé tegyem, hogy a pózokhoz ellenőrzéseket rendelhessünk. Például a guggolás esetében a vállszéles terpeszt, amelyet az alkalmazás a referenciapontok összehasonlításának segítségével ellenőrizni tud. Ezt olyan módon teszi, hogy a bal- és jobb váll távolságát összeveti a bal- és jobb boka távolságával. Ha a kettő merőben eltér, akkor egy TextToSpeech API használatával felolvassa az előre meghatározott figyelmeztető szöveget.

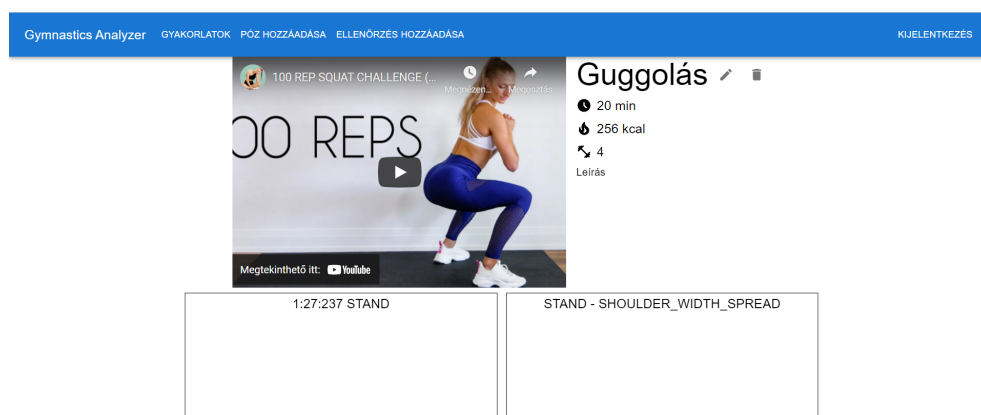


6.2. ábra. Képek az alkalmazás kezdőlapjáról és az edzés funkcióról a referenciapontok megjelenítésével

Az alkalmazás ”edzés” funkciója lehetővé teszi, hogy szabadon mozoghassunk a kamera előtt és automatikusan felismeri, hogy éppen melyik gyakorlatot hajtjuk végre. Van lehetőség arra is, hogy például a kezdőlapról egy specifikus mozgást kiválasztva kizárólag azt elemezzük, ilyenkor a mozdulatsor felismerése kimarad. A kameraképen opcionálisan megjeleníthető a referenciapont háló, ami nagy segítséget nyújt a felismeréssel kapcsolatos hibák azonosításában. A jobb felső sarokban lévő szám a végrehajtott ismétlésszámot jelzi.

6.4. Gyakorlatok bővítése

Fontos volt annak szem előtt tartása, hogy a gyakorlatokat a későbbiekben bővíteni lehessen. A megoldás kialakítása során figyelembe vettem azt, hogy instruktori oldalról hogyan lehet kényelmesen új gyakorlatok hozzáadását megvalósítani. Ennek első lépéseként az admin felületen keresztül fel kell venni az esetleges új pózokat és meghatározni, hogy az új mozdulatsort milyen pozíciók meghatározott sorrendje alkotja. Új testhelyzetek hozzáadása esetén új modell betanítása is szükséges, amelyhez vagy kép, vagy referenciapont koordináták szükségesek. Ezek gyűjtése történhet edzések-ről készített képekkel, letöltött stock fotók segítségével, vagy akár referenciapontok gyűjtésével. Az új modell betanítása, valamint annak hozzáadása az alkalmazáshoz kézzel történik, de a folyamat könnyen automatizálható.



6.3. ábra. Gyakorlat részletes nézete

7. fejezet

Összegzés

A kutatásomban arra a kérdéskörre kerestem választ, hogy lehetséges-e speciális felszerelések és szenzorok használata nélkül felismerni a mozgásokat.

A dolgozatomban bemutattam, hogy a mesterséges intelligencián alapuló pózfelismerést ki lehet egészíteni olyan további algoritmusokkal, amelyek lehetővé tették, hogy a pózok váltakozásai alapján detektálhassuk, hogy az alany milyen mozgatlansort hajtott végre. Fontos kiemelni, hogy a megoldásom a 2. fejezetben bemutatott szakirodalmakkal és módszerekkel ellentétben mindezt speciális eszközök használata nélkül, valós időben teszi, így szinte bárki detektálhatja otthoni edzéseit.

Különböző gyakorlóvideókból összeállított elemzés bemutatója az alábbi linken érhető el: <https://www.tinyurl.com/StukszaTDK>. A videó bemutatja, hogy az egymás után következő gyakorlatok felismerése sikeresen megtörténik viszonylag rövid időn belül. A bemutatott módszerek segítségével az ismétlésszámolás is hatékonyan működik: a felülés során a kisebb mozgásokat helyesen nem számolja, a hatütemű fekvőtámasz során viszont az egyik ismétlést helytelenül nem érzékeli. Összességében elmondható, hogy a módszer hatékonyan működik és az ismétlések számlálása is kis hibahatáron belül funkcionál.

Úgy gondolom, hogy a kutatás rávilágít arra, hogy komoly potenciál van még ezen terület mélyebb elemzésében és remélhetőleg dolgozatom jó kiindulópontja lehet a további kutatások (mint például a mozgásfelismerésen alapuló helyesség visszajelzés) számára.

7.1. Továbbfejlesztési lehetőségek

Annak ellenére, hogy a kutatás a tervezett szerint sikerült és a gyakorlati prototípus kipróbálása meglehetősen sikeres eredményeket produkált, számos további potenciális továbbfejlesztési lehetőséget nyújt. Ezek közül emelnék ki néhányat.

7.1.1. Adathalmaz bővítése

Az elkészített TFLite modell kimondottan nagy pontosságot ért el annak ellenére is, hogy a tanításhoz használt képek száma csupán 50-300 között változott. A még pontosabb modell elkészítésének érdekében feltétlenül szükséges lenne egy jóval nagyobb adathalmaz kialakítása, amely pózonként akár 500 képet tartalmazna. Egy ekkora tanító halmaz határozottan jobb eredmények elérését tenné lehetővé, ezáltal magabiztosabban ismerhetné fel az adott testhelyzeteket.

7.1.2. További gyakorlatok felvétele

A jelenlegi 6 megvalósított gyakorlat azzal a céllal került kialakításra, hogy ténylegesen validálni tudjam az algoritmusokat működés közben. A kialakítás során igyekeztem teljesen eltérő és összetettebb mozdulatsorokat is felvenni, valamint ügyeltem arra, hogy azokban más részgyakorlatok is szerepeljenek. Hosszútávon viszont érdemes lenne még több szekvenciával bővíteni az alkalmazást, amely még inkább rávilágítana a megoldás erősségeire és az esetleges további kihívásokra.

7.1.3. Végrehajtás helyességének részletes elemzése

A gyakorlatok megbízható felismerése egy kétségtelenül fontos lépés, viszont ezek továbbgondolásával akár a mozgások helyességének ellenőrzése is ráépíthető. Erre a prototípus megvalósításánál mutattam egy egyszerű módszert, viszont ez a téma további összetett vizsgálódást igényel.

7.1.4. Okoseszközök bővítése

A jelenlegi megoldásomban kizárólag egy okostelefon kamerájával ismerem fel a gyakorlatokat és csupán ezek alapján adok visszajelzést azok helyességéről. További eszközök és szenzorok viselése viszont újfajta megközelítéseket tenne lehetővé. Olyan eszközökre gondolok, mint a pulzusmérő, vagy a modern okosórák, amelyek olyan funkciókat tesznek lehetővé, mint a

- pulzusmérés,
- szívritmusváltozások figyelése,
- légzésfigyelés,
- egészségi állapot nyomon követése,
- vagy a hidratálás nyomon követése.

Ezek olyan további hasznos információkat biztosítanak, amelyek alapján jobban előre jelezhetőek az edzések ideális időpontjai, valamint hozzájárul azok még pontosabb elemzéséhez.

Irodalomjegyzék

- [1] B. Reily, H. Zhang, and W. Hoff, „Real-time gymnast detection and performance analysis with a portable 3D camera,” *Computer Vision and Image Understanding*, vol. 159, pp. 154–163, 2017.
- [2] X. Mu and C.-H. Min, „Wearable Sensing and Physical Exercise Recognition,” in *2022 IEEE World AI IoT Congress (AIIoT)*, pp. 413–417, IEEE, 2022.
- [3] M. Sundholm, J. Cheng, B. Zhou, A. Sethi, and P. Lukowicz, „Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix,” in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pp. 373–382, 2014.
- [4] M. P. Díaz-Pereira, I. Gomez-Conde, M. Escalona, and D. N. Olivieri, „Automatic recognition and scoring of olympic rhythmic gymnastic movements,” *Human movement science*, vol. 34, pp. 63–80, 2014.
- [5] J. D. Domingo, J. Gómez-García-Bermejo, and E. Zalama, „Visual recognition of gymnastic exercise sequences. Application to supervision and robot learning by demonstration,” *Robotics and Autonomous Systems*, vol. 143, p. 103830, 2021.
- [6] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, „RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3225–3234, 2014.
- [7] H. Koskimäki and P. Siirtola, „Recognizing gym exercises using acceleration data from wearable sensors,” in *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pp. 321–328, IEEE, 2014.
- [8] I. Pernek, G. Kurillo, G. Stiglic, and R. Bajcsy, „Recognizing the intensity of strength training exercises with wearable sensors,” *Journal of biomedical informatics*, vol. 58, pp. 145–155, 2015.

- [9] R. Khurana, K. Ahuja, Z. Yu, J. Mankoff, C. Harrison, and M. Goel, „Gym-Cam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, pp. 1–17, 2018.
- [10] R. Votel and N. Li, „Next-generation pose detection with MoveNet and TensorFlow.js,” 2021.
- [11] <https://www.gettyimages.com/detail/photo/woman-making-pushups-in-gym-at-sunrise-royalty-free-image/559688781>. [Hozzáféres dátuma: 2022. október 31.].
- [12] <https://www.gettyimages.com/detail/photo/caucasian-woman-doing-push-ups-royalty-free-image/155770622>. [Hozzáféres dátuma: 2022. október 31.].
- [13] <https://www.gettyimages.com/detail/photo/witness-the-fitness-royalty-free-image/961254552>. [Hozzáféres dátuma: 2022. október 31.].
- [14] <https://www.gettyimages.com/detail/photo/young-asian-chinese-women-doing-sit-ups-exercise-in-royalty-free-image/1217960566>. [Hozzáféres dátuma: 2022. október 31.].
- [15] <https://www.gettyimages.com/detail/photo/home-exercising-royalty-free-image/1240696055>. [Hozzáféres dátuma: 2022. október 31.].
- [16] <https://www.gettyimages.com/detail/photo/full-length-of-shirtless-man-with-barbell-standing-royalty-free-image/1205514398>. [Hozzáféres dátuma: 2022. október 31.].
- [17] <https://www.shutterstock.com/hu/image-photo/side-view-portrait-beautiful-young-woman-528602410>. [Hozzáféres dátuma: 2022. október 31.].
- [18] <https://www.shutterstock.com/da/image-photo/beautiful-girl-dark-hair-wearing-pink-363953966>. [Hozzáféres dátuma: 2022. október 31.].
- [19] <https://www.shutterstock.com/hu/image-vector/woman-doing-bodyweight-squats-exercise-flat-1985220485>. [Hozzáféres dátuma: 2022. október 31.].
- [20] <https://www.shutterstock.com/hu/image-vector/woman-doing-bodyweight-squats-exercise-flat-1985220485>. [Hozzáféres dátuma: 2022. október 31.].

- [21] <https://www.shutterstock.com/hu/image-vector/woman-doing-bodyweight-squats-exercise-flat-1985220485>. [Hozzáféres dátuma: 2022. október 31.].
- [22] <https://www.shutterstock.com/hu/image-vector/woman-doing-bodyweight-squats-exercise-flat-1985220485>. [Hozzáféres dátuma: 2022. október 31.].
- [23] <https://www.shutterstock.com/hu/image-vector/woman-doing-bodyweight-squats-exercise-flat-1985220485>. [Hozzáféres dátuma: 2022. október 31.].
- [24] https://www.123rf.com/photo_158485148_man-doing-lying-leg-raises-exercise-abdominals-exercise-body-weight-lifts-flat-vector-illustration.html. [Hozzáféres dátuma: 2022. október 31.].
- [25] K. Mehmood, „Make A Camera App Using CameraX API.” <https://proandroiddev.com/make-a-camera-app-using-camerax-api-part-1-db8fa0e176ce>, 2021. [Hozzáféres dátuma: 2022. október 31.].
- [26] „ML Kit for Firebase.” <https://firebase.google.com/docs/ml-kit>. [Hozzáféres dátuma: 2022. október 31.].
- [27] „Pose detection.” <https://developers.google.com/ml-kit/vision/pose-detection>. [Hozzáféres dátuma: 2022. október 31.].