Computer-aided Design and Manufacturing of a Patented, Left Ventricle Assist Device Positioning Tool – 3D Navigated Surgical Treatment of End-Stage Heart Failure

Imre Janos Barabas

Heart and Vascular Center, 3D Center, Semmelweis University
Városmajor u. 68, 1122 Budapest, Hungary
barabas.janos_imre@med.semmelweis-univ.hu

Bela Merkely, Istvan Hartyanszky

Heart and Vascular Center, Semmelweis University
Városmajor u. 68, 1122 Budapest, Hungary
rektor@semmelweis-univ.hu; hartyanszky.istvan@med.semmelweis-univ.hu

Daniel Palkovics

3D Center, Semmelweis University, Városmajor u. 68, 1122 Budapest, Hungary
Department of Periodontology, Semmelweis University, Szentkirályi u. 47, 1088 Budapest, Hungary
palkovics.daniel@dent.semmelweis-univ.hu

Abstract: The Left Ventricular Assist Device (LVAD), is a leading therapy option for patients with end-stage heart failure. A successful LVAD implantation is influenced by various factors, which are difficult or impossible to control during the operation. For optimal functioning of the device, its axis should be close to parallel with the septum, facing the mitral valve within the left ventricle. Our aim is to provide an innovative solution for these problems, using 3D technology and 3D printing. In our work, the production of the patented exoskeleton and the validation of its volumetric and surface quality, was presented, by comparing the virtual and 3D printed exoskeletons.

Keywords: cardiac surgery; mechanical circulatory support; 3D technology; 3D printing; personalized medicine

1 Introduction

According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of death worldwide. Among cardiovascular diseases, end-stage heart failure is one of the most severe syndromes, with constantly increasing morbidity- and mortality rates [1]. For patients in end-stage heart failure who are unresponsive to medical treatments, the gold standard therapeutic option is transplantation. However, the access to donor candidates for heart transplantation is limited [2] [3]. The implantation of a left ventricle assist device (LVAD), is a life-saving treatment option for these patients [4]. A LVAD is a mechanical pump with an external battery, that helps pumping blood from the left ventricle (LV) to the systemic circulation, thereby capable of partially or completely replacing the function of a failing heart [5].

Figure 1 shows the four main parts of the LVAD: the **pump**, the **inflow cannula**, the **outflow cannula** and the **driveline**. For further information about the device structure and function, see the paper of Englert et al. [6], and the paper of Worku et al. [7].

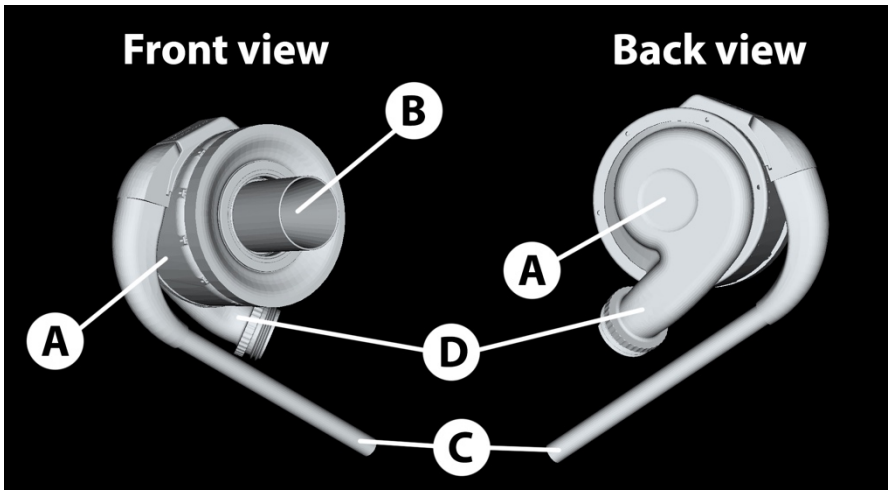


Figure 1
Virtual 3D model of the left ventricle assist device. A: Pump, B: Inflow cannula, C: Driveline, D: Outflow cannula

Success of the LVAD implantation largely depends on the position of the device, however ideal positioning is determined by various factors, that cannot be or are difficult to influence [8] [9]. Ideally, the inflow-cannula (IC) of the LVAD should be parallel with the interventricular septum (IVS) and should point towards the center of the mitral valve (MV). Based on literature data, the ideal degree between the IC and the IVS should be within -7 and +7 degrees of the left ventricular septal

axis [8], otherwise the unfavorable postoperative fluid mechanics of the left ventricle may lead to life-threatening events [10]. Directly after implantation, one of the most severe complications is the suction event, when the pump draws the IVS to the IC, causing the LVAD to stop immediately [11].

The computational fluid dynamical simulation (CFD) is becoming a more and more prominently used mechanical engineering tool in the field of medicine. Using CFD, cardiovascular experts could gain a better knowledge of blood flow dynamics in the left ventricle around the IC. The CFD provides an opportunity for simulation before the clinical usage of new medical devices, facilitating the development of new medical interventions. Despite the CFD simulation advantages, the limitation must be considered [12].

In the field of dentistry and oral surgery, static prefabricated surgical guides are utilized to execute the navigated positioning of dental implants [13] [14]. Dentists and oral surgeons utilize 3D data acquired from cranio-facial computed tomography (CT) scans to determine implant positions and to manufacture surgical guides [15]. Similarly, routine presurgical computed tomography angiography (CTA) scans can be utilized to digitally plan LVAD positions and perform a navigated surgical treatment.

Hence, the aim of this article was to describe a computer-aided design and computer-aided manufacturing process to produce a surgical positioning tool (exoskeleton) for LVADs and to validate the accuracy of the additive manufacturing procedure. The exoskeleton is under the process of international patent protection [16].

2 Materials and Methods

2.1 Image Acquisition

Prior to surgery, spiral electrocardiogram (ECG)-gated thoracic CTA scans are acquired utilizing a multidetector CT machine. Angiographic contrast imaging lasts for one cardiac cycle, during which the machine is capable to take multiple images. 3D planning is carried out relying on images taken at the end-diastolic stage.

Iodine-based intravenous contrast materials are utilized to enhance the visibility of the vessels during radiographic imaging. At the end-diastolic phase, the contrast agent clearly draws out the inner surface and the inner structures of the left ventricle. Therefore, the IVS, the MV, the mitral annulus, and the papillary muscles can be separated from each other. During the virtual model building process, these structures can be reconstructed in 3D. The given flowchart below illustrates our study structure (Figure 2).

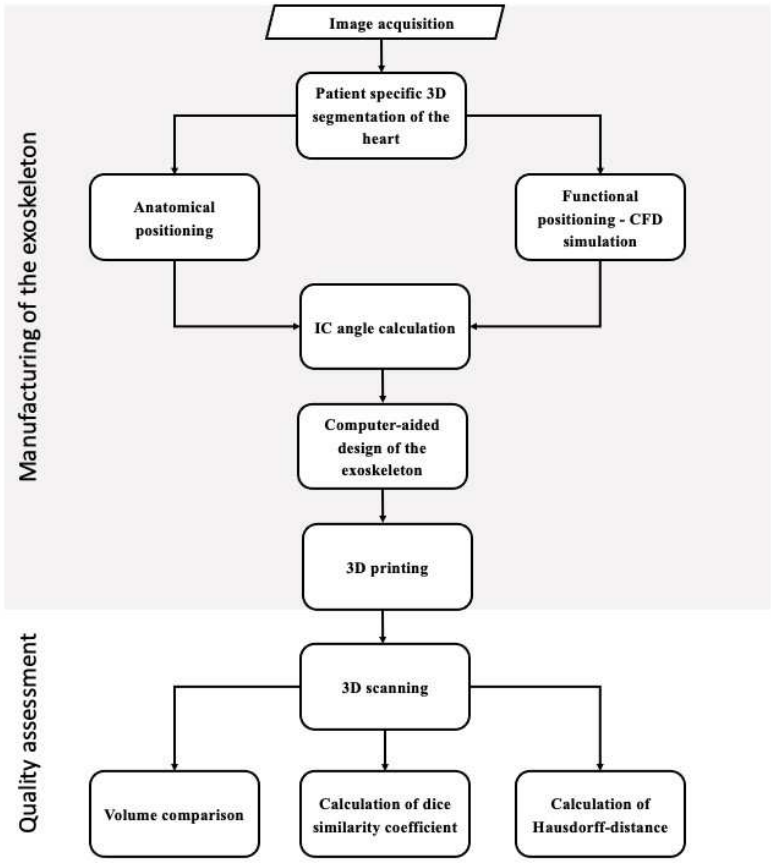


Figure 2
The flowchart illustrates the steps of the exoskeleton manufacturing and quality assessment process

2.2 Radiographic Image Segmentation

DICOM (Digital Imaging and Communications in Medicine) datasets were imported into an open-source medical image processing software (3D Slicer v5.4.0, Boston, MA, USA) [17]. During image segmentation, a binary labelmap for every corresponding anatomic structure is being generated on the multiplanar CTA images (Figure 3). Semi-automatic external heart surface segmentation consists of the following steps: (i) edge detection: outlining the borders of the heart on 5-7 axial slices of the CTA dataset, utilizing a voxel intensity level tracing tool; (ii) morphological contour interpolation: based on the outline of the previously

generated labelmaps, the algorithm calculates the labelmaps on interim slices, interpolating between slices drawn by free hand; (iii) Gaussian method is utilized to smoothen the outline of the binary labelmap. Labelmaps for the internal anatomic structures of the left ventricle (i.e., IVS, MV, papillary muscles) are generated within the area of the previously created labelmap of the heart. Local thresholding segmentation is applied to generate a labelmap for the cavity of the left ventricle. After occasional manual correction and smoothing, the labelmap representing the cavity of the left ventricle is subtracted from the labelmap representing the entire heart. All steps of the segmentation were inspected by a radiologist, a cardiologist and a cardiac surgeon. If necessary, the model was changed with the full agreement of the clinical specialist.

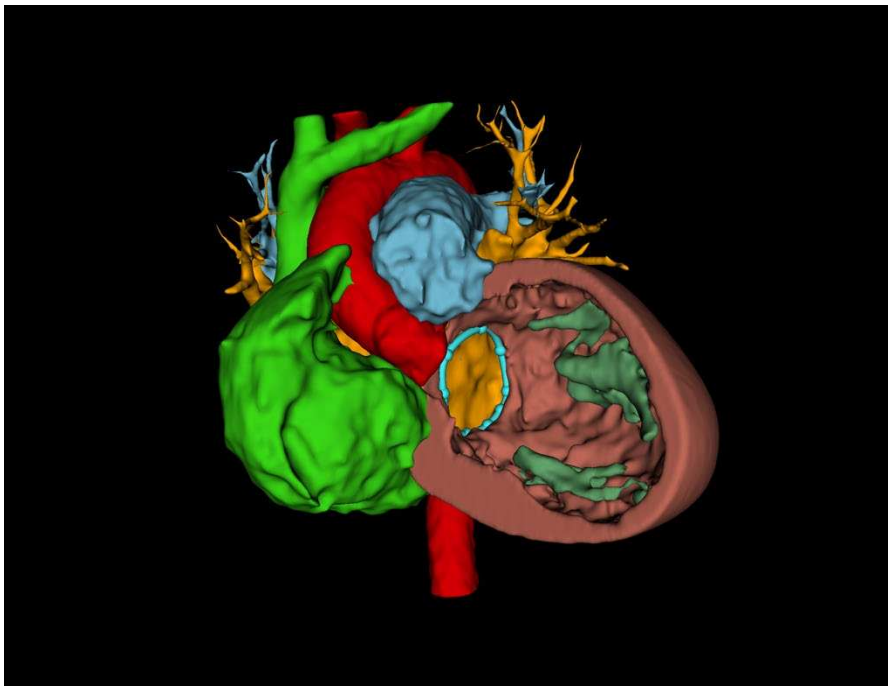


Figure 3
3D model of the segmented heart. A patient-specific computational model provides the possibility to design a personalized exoskeleton for the LVAD implantation. Neon green: right atrium, red: aorta, blue: pulmonary artery, orange: left atrium, brown: heart muscle (myocardium), green: papillary muscles, blue circle: mitral annulus.

With real-time rendering, the software automatically generates a 3D surface representation of the 2D binary label map. 3D surface representation of the segmentation is exported as a Standard Tessellation Language (STL) file.

2.3 Positioning of the IC

2.3.1 Anatomic Positioning

A 3D model of the segmented heart (with the inner structure of the left ventricle), and the 3D model of the LVAD are imported into the Autodesk Fusion 360 (**AutoDesk** Inc. San Rafael, CA, USA) 3D CAD software for the anatomical positioning of the device (Figure 4). The LVAD is placed (i) at 5-15 mm distance from the IVS, (ii) at equal distance from the papillary muscles and (iii) the IC should point towards the center of the MV. The final result of the anatomical positioning step is to orient the LVAD and inflow cannula inside the left ventricle.

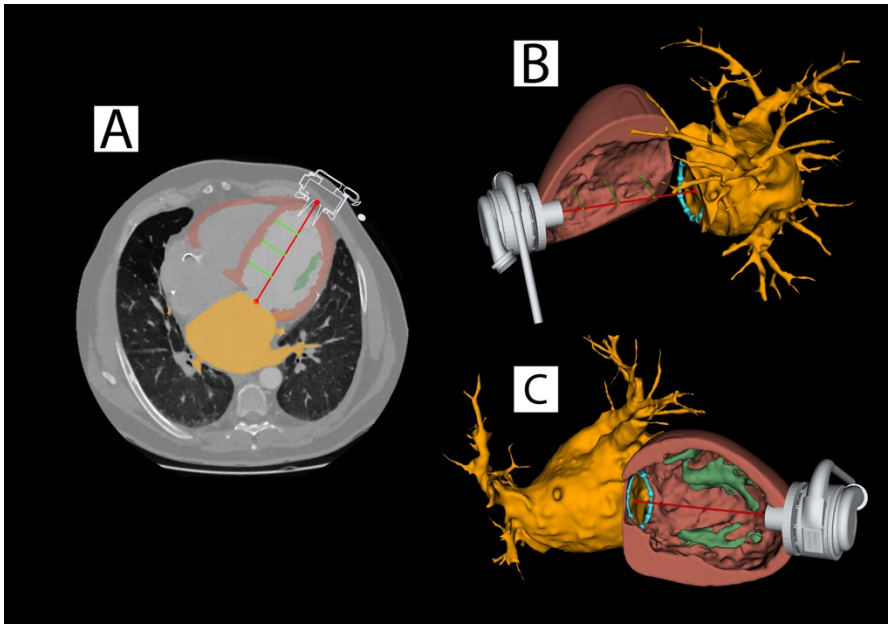


Figure 4

Anatomic positioning of the LVAD **A:** the axial plane of the CT angiography shows the distance (neon green lines) between the inflow cannula's axis (red line) and the interventricular septum (current distance: 13.2 mm); **B-C:** the relation of the IC long axis with the papillary muscles (green) and the mitral annulus (blue circle)

2.3.2 Functional Positioning – Computational Fluid Dynamic Simulation

During the functional positioning step, the most preferable angle was calculated, using CFD simulation, Ansys Fluent 2022 R1 (**ANSYS** Inc. Southpointe, Canonsburg, PA, USA). In our model the left ventricle geometry was patient specific, but other CFD factor was standardized. For fine-tune the IC position for

ideal postoperative ventricular blood flow, a CFD simulation is carried out for all 5 different potential IC angles. A pivot point is placed at the intersection between the IC long axis and the heart muscle, where the LVAD will be implanted. The IC is then rotated into 5 different angles: the center and the four corner points of the MV. Subsequently, fluid mechanical analysis is performed for all 5 potential IC angles.

Standardized fluid characteristics of blood are programmed into the CFD simulation software. Our fluid simulation based on the research of Chivukula et al. [8], and Méndez et al. [18] theoretical flow simulation setup. During the simulation thrombus formation factors were not conducted. Blood is treated as a non-Newtonian fluid with a viscosity of 3.5 mPa·s and a density of 1050 kg/m^3. The inlet part of the heart was the mitral valve, and the inlet boundary condition was specified in flow velocity as 180 m/s, with no pulsatility. Meanwhile, the outlet boundary condition is set on top of the IC inside the left ventricle, with the same conditions as at the inlet point. The inner diameter of the IC was measured as 23.5 mm, while the outer diameter was 25 mm, maintaining a 1:1 ratio with the original device. The insertion depth of the IC was calculated as the difference between the original length of the IC (22 mm) and the thickness of the left ventricle (7.2 mm ± 2.1 mm). The wall (heart muscle) boundary condition is defined as no-slip condition, wherein the normal component of velocity is fixed at zero, and the tangential component is set equal to the velocity of the wall.

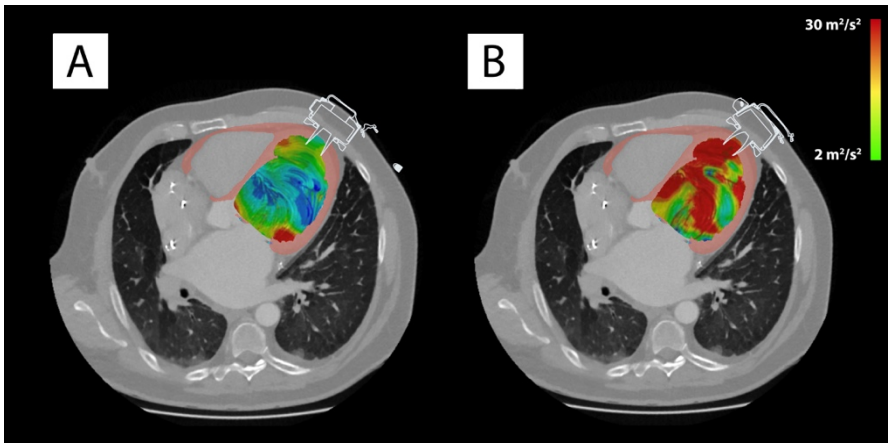


Figure 5

Functional positioning of IC based on CFD simulation. The colormap reveals the turbulence differences between the proper and malpositioned LVAD flow patterns. **A:** turbulence in case of optimal angulation; **B:** turbulence in case of malpositioned IC inside the left ventricle.

During CFD simulation, fluid inflow is set to the MV, whereas fluid outflow is set to the IC of the LVAD. For each of the 5 possible IC positions, turbulence is calculated. Colormap visualizes turbulence values, expressed as: m^2/s^2; the ideal turbulence value for the LV should be between 2 m^2/s^2 and 22 m^2/s^2. The angle with

the most ideal blood flow dynamic is selected as the final planned position of the LVAD (Figure 5).

2.4 Computer-aided Design of the Exoskeleton

The exoskeleton is designed in the Autodesk Fusion 360 (**AutoDesk** Inc., San Rafael, CA, USA) 3D CAD software, capable of free-form surface modelling (Figure 6). The base of the exoskeleton is designed over the outer surface at the apex of the heart, utilizing the heart model generated with the segmentation of the CTA scans. The convex surface of the exoskeleton provides an appropriate fit on the outer surface of the patient's heart during the operation. The extent of the exoskeleton is limited as much as possible, to allow for a minimally invasive incision design and surgical wound, while still providing sufficient support for the guidance of the coring knife. The coring knife is a specific surgical instrument, that is used to cut out a cylindrical shape of the heart muscle for the implantation of the inflow cannula. The guide base spans from the apex of the heart to the IVS with a final leaf-like shape.

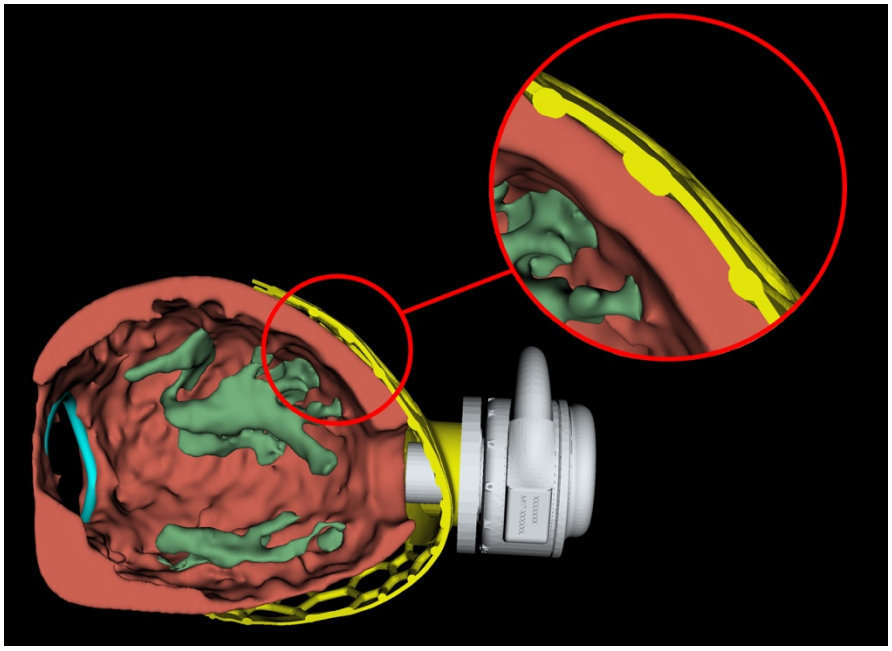


Figure 6
The exoskeleton fits precisely on the surface of the heart. Precise fitting ensures the proper positioning of the IC during the surgical implantation.

At the final position of the IC, a cylindrical guide channel (guiding tube) with a diameter that is capable of housing the coring knife is designed into the exoskeleton base, at the planned anatomical position and preferable angle of the IC. This tube will guide the coring knife during surgery, ensuring that the LVAD will be placed in the planned position. Finally, a Voronoi-pattern is applied to skeletonize the base of the guide, hence the name exoskeleton. The final step of CAD modelling is the smoothing procedure of the exoskeleton. All sharp edges are removed, avoiding any kind of geometry distortion around the guiding tube. At the end of the process, the virtual exoskeleton is suitable for 3D printing and surgical use.

Once the design is finalized, the model of the exoskeleton is exported as a standard tessellation language (STL) file, and is prepared for 3D printing and quality control.

2.5 Additive Manufacturing of the Exoskeleton

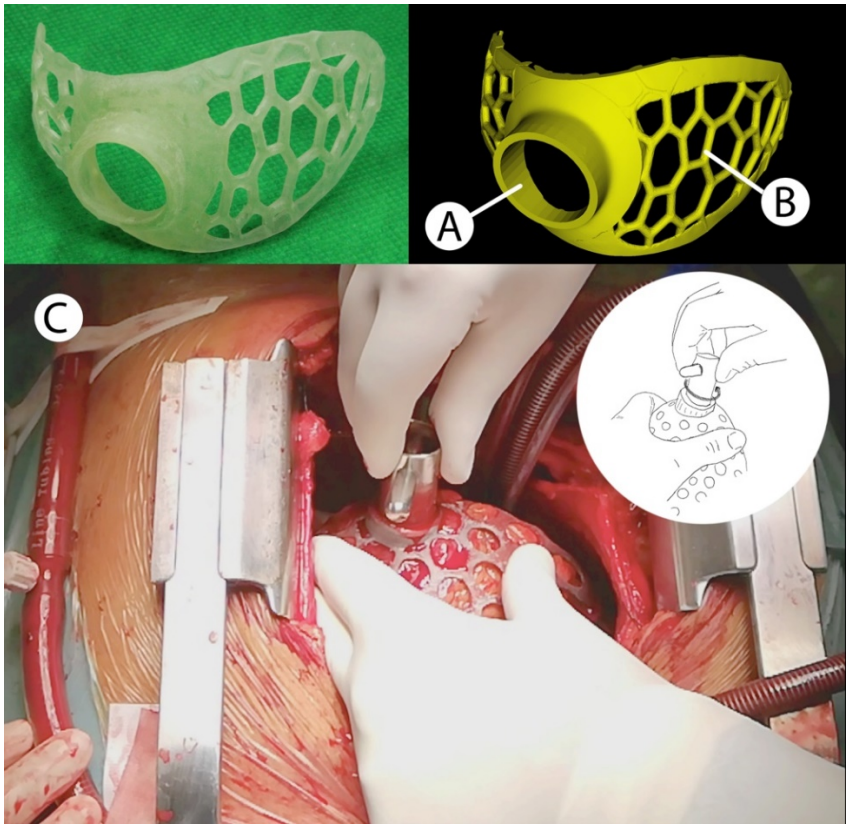


Figure 7
3D digital model, additively manufactured model and surgical implementation of the exoskeleton. **A:** guiding tube for the coring knife, **B:** Voronoi-pattern skirt of the exoskeleton, which ensures a clear view to confirm proper fit of the exoskeleton. **C**: Surgical use of the exoskeleton

The STL file of the exoskeleton is imported to the dedicated software of a 3D printer capable of producing devices suitable for surgical applications. A wide variety of printers are available on the market using stereolithography (SLA) technology, which can be utilized to manufacture the exoskeleton. A sterilizable medical grade photopolymer resin (USP class VI; ISO 10993-5, -10, -23) with adequate mechanical properties can be utilized to manufacture the exoskeleton (Figure 7). Prior to surgery, the exoskeleton is sterilized in an autoclave at 121 °C for 30 minutes [19].

2.6 Outcome Measures - Quality Control of the Printed Exoskeleton

Fifteen patient-specific virtual 3D exoskeleton (VE) masks were extracted, accompanied by the retrieval of their corresponding 3D printed exoskeleton (3DPE) models. The 3D scanning process was carried out using a desktop optical 3D scanner (Einscan-SP V2, SHINING 3D, Hangzhou, China), which ensured high-resolution point cloud data with an exceptional single-shot accuracy of 0.05 mm, conforming to the ISO 12836 standard. These data were then transformed from 'Mask to Surface' utilizing a 3D inspection software, Geomagic Control X (3D Systems, Rock Hill, SC, USA). Later it was used to generate colormaps between the VE STL file, and the re-scanned 3DPE, STL file (Figure 8).

The two datasets (from VE and 3DPE) were registered. The VE was chosen as the fixed reference dataset and the corresponding printed model surface for the corresponding registration dataset. To ensure uniformity, six consistent anatomical landmarks were manually pinpointed on each surface. This selection comprised three landmarks on the convex surface and an additional three on the concave surface, situated within the upper, middle, and lower regions of both the VE and 3DPE surfaces.

The VEs were compared to the 3DPE scans, to validate the accuracy of the manufacturing process. The primary outcome measure was the volumetric difference between the VE models and the 3DPE models.

Secondary outcome measures were: dice similarity coefficient (DSC) and Hausdorff-distance (HD). These metrics assessed the surface quality between the VE and 3DPE models. The DSC uses a reproducibility validation metric and an index of spatial overlap. DSC values range from 0 to 1, which denotes the entire spatial similarity between the two objects. The value of 1 means a total overlap between two 3D models [20].

On the other hand, the Hausdorff-distance [21] measures distance between two-point sets. Thus, this distance can be used to determine the degree of resemblance between two objects that are superimposed on one another. Three different values were measured on the models: (i) maximum, (ii) average distance between two points, and (iii) the 95^{th} percentile of measured distances.

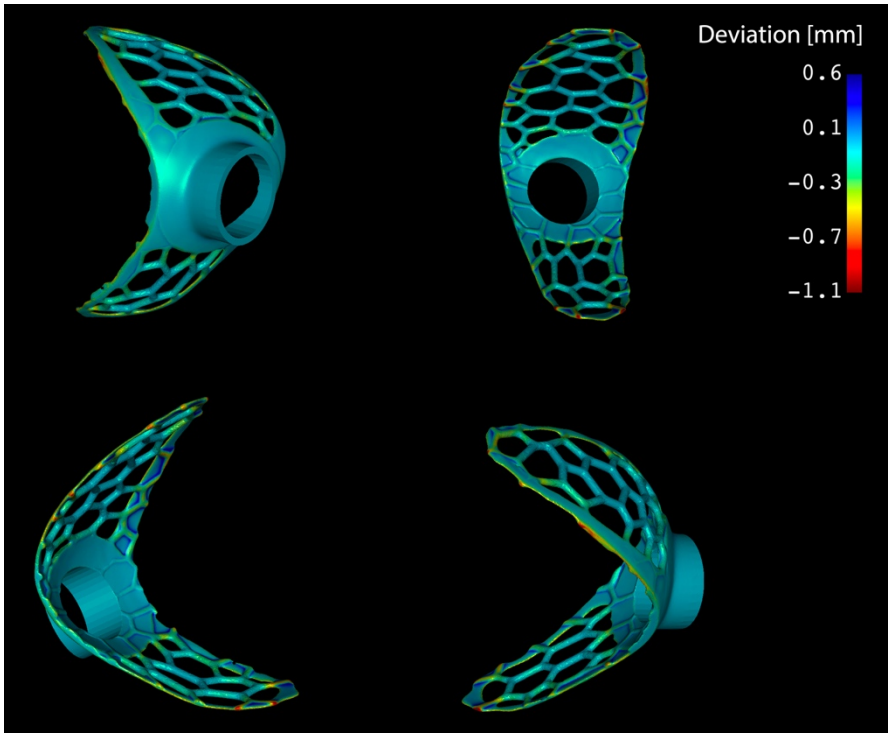


Figure 8
Comparison of the VE and 3DPE models. The surface similarity data were visualized in a deviation map, where the colors red, orange and yellow describe areas with negative material discrepancy, while cyan and blue colors describe areas with surplus material compared to the original STL object.

In all cases, the master and scanned models were considered identical, if the DSC was greater than 0.9 and the HD 95th percentile was less than 1mm.

Statistical significance between the values of each outcome measure was calculated paired t-test with a significance level of 0.05. Statistical evaluation was done with SPSS Statistics (IBM, Armonk, NY, USA).

3 Results

3.1 Primary Outcome – Volumetric Differences

The volume comparison between the VE model and the 3DPE model revealed no statistically significant difference (p= 0.76). Volume of the VE models averaged at

10.81 cm^3 ± 1.69 cm^3, whereas average volume of the 3DPE models were 10.80 cm^3 ± 1.70 cm^3. Volumetric differences are shown in Figure 9.

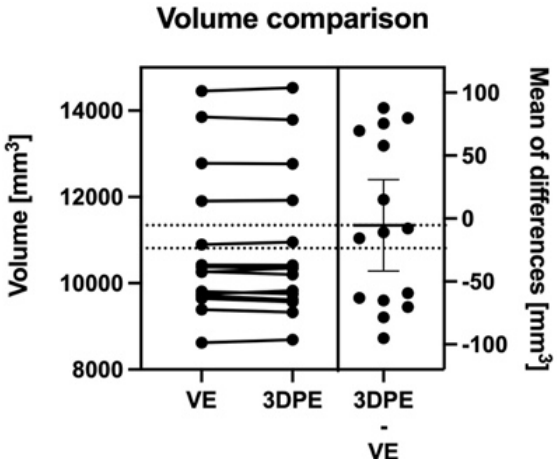


Figure 9
Estimation plot of the virtual exoskeleton (VE) and 3D printed exoskeleton (3DPE).

3.2 Secondary Outcome Measures – Surface and Volumetric Quality Assessment

The surface quality measurements did not show considerable differences. Spatial overlap between the VE and 3DPE models measured by the DSC showed an average 0.95 ± 0.03 value (minimum: 0.91, maximum: 0.98). The maximum HD value averaged at 2.83 mm ± 1.31 mm. Whereas, the mean HD distance between VE and 3DPE averaged at 2.17 mm ± 0.52 mm, with an average 0.16 mm ± 0.06 mm 95th percentile HD value. The 95th percentile of the distance was less than 1mm in every scanned exoskeleton. DSC and HD values are shown in Table 1.

Table 1
Surface quality measurements of the virtual and the 3D printed exoskeleton.

	Dice similarity coefficient	**Hausdorff-distance**		
		Maximum [mm]	**Average [mm]**	**95% [mm]**
Maximum	0.98	2.83	0.26	0.94
Minimum	0.91	1.31	0.08	0.31
Mean	0.95	2.17	0.16	0.60
SD	0.03	0.52	0.06	0.21

4 Discussion

The aim of the current paper was to describe the fundamental aspects of planning a navigated surgical LVAD implantation, and the additive manufacturing of the patented exoskeleton. Re-scanning of the actual 3D printed exoskeletons provides an ideal solution to compare the volume- and the surface stability of two 3D models.

The implantation of an LVAD is a highly technical, sensitive procedure, that mostly relies on previous LVAD implantation experience. Therefore, it is implemented in expert and high-volume centers, as these are associated with a low risk of post-operative bleeding, infection, RV failure and mortality [22]. Yet, even in the hands of an experienced surgeon, results may differ greatly, without the navigated placement of the device. Unfavorable postoperative IC position, caused by inadequate implantation, has negative consequences. The malposition of the device may increase the risk of postoperative complications on long term [10]. The incidence of the complications mentioned in section introduction, arising from device misplacement, could potentially be diminished or entirely prevented through the implementation of the exoskeleton. Important to note that this study did not gather data on long-term postoperative complications. Therefore, additional research is required to prove our device positive effects on clinical outcomes.

A case report demonstrating the clinical application of the intra-surgical application of the exoskeleton has previously been published by our group [19]. In that case, the accuracy of the LVAD position was validated by comparing angulations between the IVS and the IC on the preoperative plan and the postoperative control CT scan. The deviation between the planned position and the actual postoperative position of the IC was 2.6°. In a recently published article by Pearman et al., the authors demonstrated an average of 16.3°–23.2° deviation between the actual IC position and the postoperatively determined hypothetical ideal IC position (the axis connecting the center of the MV and the center of the IC) [23]. Although slightly different reference axes were utilized in the two studies, and it is difficult to compare one single case to a series of 42 cases, the difference in deviation from the ideal values is remarkable. Based on our clinical experience, and so far, unpublished data, high accuracy surgical results can be achieved with high reproducibility. Additionally, with the application of the exoskeleton, postoperative outcomes seem to be influenced less by the experience of the surgeon. An additional benefit of 3D virtual planning is that the surgeon has a more accurate insight of the patient's anatomy. Besides planning, virtual and/or 3D printed anatomical models can be utilized during surgery to visualize anatomical relations of different structures, even in cases with limited direct visualization.

With the re-scanning process it was demonstrated that the 3D printed exoskeleton does not show considerable differences to the original digital model. Only the edges reveal more than 1 mm distance inequality, at a maximum value of 2.83 mm. This occurrence is most likely due to the post-processing of the 3D printed models, namely the removal of the support structure. During the post-process, the edges of the exoskeleton were rounded with a bur. This post-process step is essential to manufacture a safe surgical instrument, which can be used during the operation. On the other hand, the average 95th percentile of two points distance was 0.60 mm, which is noticeably less than the accepted 1mm distance difference. The distance observed on the concave surface of the exoskeleton measures below 0.1 mm (0.04 ± 0.038 mm). The convex surface around the guiding tube determines the accuracy of the implantation process. This indicates that the quality of the exoskeleton was sufficient for surgical application. DSC values marking the spatial overlap between the VE and 3DPE models averaged at 0.95 ± 0.03. Results of the volumetric measurements are in line with the surface assessment, with no statistically significant difference between the volume of the VE and the 3DPE models.

While the medical discipline holds significant promise for the application of CFD simulation, establishing accurate model parameters proves to be a challenging task. This difficulty arises from the fact that the properties of blood undergo changes in varying environments [12] [24]. Consequently, the concept of simulating patient-specific scenarios becomes a topic of discussion in existing literature, particularly when considering essential thrombosis-related parameters such as activated platelets, resting platelets, prothrombin concentration, anti-thrombin III concentration, and other factors. For that reason, standardized parameters were used in our simulation, conducting a theoretical flow simulation.

Because of this limitation, CFD was exclusively utilized to refine the positioning of the inflow cannula. The patient specific anatomy of the left ventricle defined the position of the inflow cannula. The CFD approach was aimed at demonstrating the future capabilities of CFD in effectively determining the optimal positioning of LVADs. Another drawback of the technique is the relatively long duration of the planning process and the fact, that the user must have a complex knowledge, in radiographic image processing, CAD modelling and 3D printing. Some of the steps (e. g., segmentation, CAD modelling of the exoskeleton) still require substantial human interaction. The automation of these processes shall be possible in the future with the development of deep learning segmentation algorithms and dedicated software, to reduce the duration of the planning phase and to reduce the rate of occasional human-made errors.

Conclusions

3D modeling and printing have proven to be promising technologies, with impressive potential applications for medicine. Among other benefits, they provide the possibility to plan surgical procedures, even before the first incision. In our ongoing study, all cases were prepared with the help of a virtual, patient-specific

3D model of their heart, with the surrounding vessels and organs. The overall responses from the operating surgeons to this technology were positive. The main opinion is, that 3D technology supports the success of the surgical intervention. However, the leading barrier for the use of 3D technology in clinical practice, is the accessibility of the necessary engineering knowledge (usage of different software, 3D modeling and printing, post processing and availability of a 3D printer). With the enthusiasm for this technology, comes the need for the standardization of the technique, the establishment of a clinical facility and improved accessibility.

It can be concluded, that, with the elaborative planning process, the LVAD can be implanted into an ideal position. Therefore, the complication rate of IC malpositioning could be reduced or completely avoided. The exoskeleton is a surgical instrument which provides the possibility of a personalized, calculation-based implantation of the LVAD. One of the mayor drawbacks of the current planning process is the lack of a patient specific, thrombus formation simulation. Therefore, in the future, individualized haemostaseology processes will be considered, during the planning stage. In order to evaluate the clinical performance of the device, large sample comparative studies must also be conducted.

Acknowledgement

This work was supported by the Semmelweis Science and Innovation Fund as part of the "STIA-KFI-2020" research and development project.

References

[1] Roger, V. L. et al.: Epidemiology of Heart Failure: A Contemporary Perspective, Circulation research, 2021, Vol. 128, pp. 1421-1434

[2] Tatum, R. et al.: Evaluation of donor heart for transplantation, Heart failure reviews, 2022, Vol. 27, pp. 1819-1827

[3] Severino, P. et al.: Advanced Heart Failure and End-Stage Heart Failure: Does a Difference Exist, Diagnostics (Basel, Switzerland), 2019, Vol. 170, pp. 65-71

[4] Varshney, A. S. et al.: Trends and Outcomes of Left Ventricular Assist Device Therapy: JACC Focus Seminar, Journal of the American College of Cardiology, 2022, Vol. 79, pp. 1092-1107

[5] Pietras, C. et al.: Surgical Implantation of Intracorporeal Devices: Perspective and Techniques, Cardiology clinics, 2018, Vol. 36, pp. 465-472

[6] Englert, J. A. R. et al.: Mechanical Circulatory Support for the Failing Heart: Continuous-Flow Left Ventricular Assist Devices, The Ochsner journal, 2016, Vol. 16, pp. 263-9

[7] Worku, B. et al.: Thoracotomy versus sternotomy? The effect of surgical approach on outcomes after left ventricular assist device implantation: A review of the literature and meta-analysis, J Card Surg, 2021, Vol. 7, pp. 2314-2328

[8] Chivukula, V. K. et al.: Left Ventricular Assist Device Inflow Cannula Angle and Thrombosis Risk, Circulation Heart failure, 2018, Vol. 11, pp. 23-41

[9] Chivukula, V. K. et al.: Small Left Ventricular Size Is an Independent Risk Factor for Ventricular Assist Device Thrombosis, ASAIO journal (American Society for Artificial Internal Organs: 1992), 2019, Vol. 65, pp. 152-159

[10] Schlöglhofer, T. et al.: Inflow cannula position as risk factor for stroke in patients with HeartMate 3 left ventricular assist devices, Artificial organs, 2022, Vol. 46, pp. 1149-1157

[11] Kortekaas, K. A. et al.: Left ventricular assist device and pump thrombosis: the importance of the inflow cannula position, The international journal of cardiovascular imaging, 2022, Vol. 38, pp. 2771-2779

[12] Lee, B. K.: Computational fluid dynamics in cardiovascular disease, Korean circulation journal, 2011, Vol. 41, pp. 423-430

[13] Ganz, S. D.: Three-dimensional imaging and guided surgery for dental implants, Dental clinics of North America, 2015, Vol. 59, pp. 265-290

[14] Sass, T. et al.: Esthetic and functional reconstruction of large mandibular defects using free fibula flap and implant-retained prosthetics - a case series with long-term follow-up, Head & face medicine, 2021, Vol. 17, p. 43

[15] Palkovics, D. et al.: Digital Hybrid Model Preparation for Virtual Planning of Reconstructive Dentoalveolar Surgical Procedures, Journal of visualized experiments: JoVE, 2021, Vol. 174, pp. 16-25

[16] Barabás, I. J. et al.: Exoskeleton for assisting surgical positioning, method for producing the exoskeleton and a surgical method applying such an exoskeleton, Patent, 2018, PCT/HU2019/050049

[17] Herz, C. et al.: dcmqi: An OpenSource Library for Standardized Communication of Quantitative Image Analysis Results Using DICOM, Cancer research, 2017, Vol. 77, pp. 87-90

[18] Méndez R. R. et al.: Multi-constituent simulation of thrombus formation at LVAD inlet cannula connection: Importance of Virchow's triad, Artificial organs, 2021, Vol 45, pp. 1014-1023

[19] Barabás, I. J. et al.: A 3D printed exoskeleton facilitates HeartMate III inflow cannula position, Interactive cardiovascular and thoracic surgery, 2019, Vol. 29, pp. 644-646

[20] Krithika Alias AnbuDevi, M. and Suganthi, K.: Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET, Diagnostics (Basel, Switzerland), 2022, Vol. 12, p. 3064

[21] Taha, A. A. and Allan H.: Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, BMC medical imaging, 2015, Vol. 15, p. 12

[22] Loforte, A. et al.: Ventricular assist devices implantation: surgical assessment and technical strategies, Cardiovascular diagnosis and therapy, 2021, Vol. 11, pp. 277-291

[23] Pearman, M. et al.: Comparing left ventricular assist device inflow cannula angle between median sternotomy and thoracotomy using 3D reconstructions, Artificial organs, 2022, Vol. 29 p. 144

[24] Berg, P. et al.: A review on the reliability of hemodynamic modeling in intracranial aneurysms: why computational fluid dynamics alone cannot solve the equation, Neurosurg Focus, 2019, Vol. 47 p. 15

Exploring Robustness in Blood Glucose Control with Unannounced Meal Intake for Type-1 Diabetes Patient

Péter Szalay[1], Dániel András Drexler[2], and Levente Kovács[2]

[1]Elektrobit Automotive Finland Oy, Elektroniikkatie 13, Oulu 90590, Finland, peter.szalay@elektrobit.com
[2]Physiological Controls Research Center, Óbuda University, Bécsi Street 96/B, Budapest 1034, Hungary, kovacs@uni-obuda.hu, drexler.daniel@nik.uni-obuda.hu

Abstract: The importance of efficient diabetes treatment and its' reliable automation is rising with the prevalence of this chronic condition worldwide. Robustness is one of the enablers of the safe automation of plasma glucose control, for it can ensure consistent behavior when the controlled dynamics are changing or partially unknown. Hence, this work focuses on assessing the capabilities of a robust nonlinear controller framework in a simulated environment. Linear Parameter-Varying modeling is combined with robust control techniques, supported by an Unscented Kalman filter. These controllers are then subjected to additional constraints in search of a practical trade-off between disturbance rejection and the severity of transient behavior. Simulations with unannounced meal intakes explore and compare the effect of these additional constraints and configurations using the virtual patients provided by a well-known in silico simulator. The simulation results indicate that this control method can ensure adequate blood glucose control and has the potential to support other control algorithms to realize a safe and reliable artificial pancreas.

Keywords: T1DM; Artificial pancreas; Robust control; LPV; Glycemic controller

1 Introduction

In a healthy individual, a complex endocrine system keeps the glucose concentration in the blood within a narrow range (3.9 - 7.8 [mmol/L]). Diabetes Mellitus (DM) is a collective term referring to several chronic metabolic diseases where this system is impaired, resulting in elevated glucose levels. In particular, Type 1 Diabetes Mellitus (T1DM) is diagnosed when the β-cells in the pancreas cannot produce insulin, a peptide hormone that plays a crucial role in decreasing plasma glucose concentration [1]. The most common treatment for this condition includes regular insulin injections. Due to an increasing prevalence in the population [2, 3], the recent decades saw extensive research in the automation of insulin treatment [4, 5], commonly referred to as Artificial Pancreas (AP). Some

AP implementations have already undergone clinical trials [6–8].

Maintaining normal glucose concentration (normoglycemia) is a challenging control problem due to the following reasons:

- Most models approved for representing a T1DM patient [9–11] are severely nonlinear, even though they focus on only the most relevant aspects of human metabolism.
- Tuning the parameters of these models has practical limitations [12, 13]. Hence, the nominal model can deviate significantly from the actual patient behavior [14, 15].
- The commercially available continuous glucose monitoring (CGM) sensors have significant noise and drift [16–18].
- A single-hormone controller that can only administer insulin has no means to elevate glucose levels. Therefore, the controller must be designed so that the control signal is always non-negative. Controllers with the capability to increase glucose concentration using glucagon [19] is outside the scope of this work due to some practical challenges in glucagon delivery [20].
- Even rapid-acting insulin - if injected subcutaneously - has a significantly slower effect on the plasma glucose concentration than meal intake or physical activity [9, 10].
- The dynamics of the human metabolism concerning glucose is slower when the glucose levels are lower compared to when they are high [21].
- The plasma glucose concentration is affected by various factors that are difficult to measure, detect, or even quantify. These include meal intake, physical activity, insulin administered by the patient (in a semi-automated setting), and various other physiological conditions.
- Having glucose levels below 3.9 mmol/L or 72 mg/dL (hypoglycemia) is a more severe acute complication than high glucose concentration (hyperglycemia). While reducing the latter in severity and frequency is the primary goal of AP, the former can lead to loss of consciousness, seizures, and even death [1] and hence must be strictly avoided [22].

Numerous different approaches were presented for control algorithms [23] that can address the above-listed particularities. These include different incarnations of PID control [24, 25], adaptive controllers [26], machine learning algorithms [27, 28], fractional order controllers [29] among others. One of the most widely accepted approaches is model predictive control (MPC), which showed remarkable results both in simulation and in practice [30, 31]. However, MPC requires an accurate model, which is rarely available in clinical practice. Hence, increasing the robustness of the controllers [32, 33] or making robustness the core feature of the control algorithm [34–36] gained popularity. Despite their limited disturbance rejection capabilities compared to MPC adjusted to individual patients, robust control methods have the potential to achieve a desirable compromise between safety and the efficiency of the T1DM treatment.

This work focuses on a class of robust nonlinear controllers aiming to ensure normoglycemia for T1DM patients. The T1DM model presented in [37] serves as the basis of the controller design, transformed into an LPV model, extended with an output uncertainty model, and weighting functions representing expected tracking performance, reference dynamics, and constraints on the control signal. The LPV controller can ensure robustness by minimizing $\mathscr{H}_\infty$ or $\mathscr{H}_2$ norms between various inputs and outputs of the extended LPV model.

2 Modeling the Human Metabolism

The following differential equations describe the 11th order Cambridge model introduced by Hovorka et al. [9] and later updated by [37]:

$$\begin{aligned}
\dot{C}(t) = \;& -k_{a,int}C(t) + \frac{k_{a,int}}{V_G}Q_1(t) \\
\dot{Q}_1(t) = \;& -\left(\frac{F_{01}}{Q_1(t)+V_G} + x_1(t)\right)Q_1(t) + k_{12}Q_2(t) - \\
& -R_{cl}\max\{0, Q_1(t) - R_{thr}V_G\} - Phy(t) + \\
& +EGP_0\max\{0, 1-x_3(t)\} + \min\left\{U_{G,ceil}, \frac{G_2(t)}{t_{max}}\right\} \\
\dot{Q}_2(t) = \;& x_1(t)Q_1(t) - \Big(k_{12} + x_2(t)\Big)Q_2(t) \\
\dot{x}_1(t) = \;& -k_{b1}x_1(t) + S_{IT}k_{b1}I(t) \\
\dot{x}_2(t) = \;& -k_{b2}x_2(t) + S_{ID}k_{b2}I(t) \\
\dot{x}_3(t) = \;& -k_{b3}x_3(t) + S_{IE}k_{b3}I(t) \\
\dot{I}(t) = \;& \frac{k_a}{V_I}S_2(t) - k_eI(t) \\
\dot{S}_2(t) = \;& -k_aS_2(t) + k_aS_1(t) \\
\dot{S}_1(t) = \;& -k_aS_1(t) + u(t) \\
\dot{G}_2(t) = \;& \frac{G_1(t) - G_2(t)}{t_{abs}(t)} \\
\dot{G}_1(t) = \;& -\frac{G_1(t)}{t_{abs}(t)} + D(t),
\end{aligned} \tag{1}$$

where $C(t)$ is the glucose concentration in the subcutaneous tissue [mmol/L], $Q_1(t)$ and $Q_2(t)$ are the masses of glucose in accessible and non-accessible compartments [mmol], $x_1(t)$ [1/min], $x_2(t)$ [1/min], and $x_3(t)$ [-] are the remote effects of insulin on glucose distribution, disposal and endogenous glucose production, respectively, $I(t)$ is the insulin concentration in plasma [mU/L], $S_1(t)$ and $S_2(t)$ are the insulin masses in the accessible and non-accessible compartments [mU], while $G_1(t)$ and $G_2(t)$ are the masses of ingested glucose in the stomach and gut [mmol/kg]. $u(t)$ is the injected insulin flow of rapid-acting insulin [mU/min], which is the input of the system. $D(t)$ is the amount of ingested carbohydrates [mmol/min], and $Phy(t)$ is the effect of physical activity [mmol/min]. Both are considered as disturbances. Table 1 provides details on model parameters. The glucose absorption time constant $t_{abs(t)}$ is a function of state variables,

Table 1
Cambridge model parameters.

Name	Unit	Description
$k_{a,int}$	1/min	transfer rate constant between the plasma and the subcutaneous compartment
V_G	L	distribution volume of glucose in the accessible compartment
F_{01}	mmol/min	total non-insulin dependent glucose flux
k_{12}	1/min	transfer rate constant from the non-accessible to the accessible compartment
R_{cl}	1/min	renal clearance constant
R_{thr}	mmol/L	glucose threshold for renal clearance
EGP_0	mmol/min	endogenous glucose production extrapolated to the zero insulin concentration
t_{max}	min	time-to-maximum appearance rate of glucose in the accessible compartment
$U_{G,ceil}$	mmol/min/kg	maximum glucose flux from the gut
k_{b1}, k_{b2}	1/min	deactivation rate constants
k_{b3}	1/min	deactivation rate constant for the insulin effect on endogenous glucose production
S_{IT}	L/mU/min	insulin sensitivity for transport
S_{ID}	L/mU/min	insulin sensitivity for distribution
S_{IE}	L/mU	insulin sensitivity for endogenous glucose production
k_a	1/min	insulin absorption rate constant
V_I	L	volume of distribution of rapid-acting insulin
k_e	1/min	fractional elimination rate from plasma

and calculated as follows:

$$t_{abs}(t) = \max\left\{t_{max}, \frac{G_2(t)}{U_{G,ceil}}\right\}. \tag{2}$$

Parameters $k_{a,int}$, F_{01}, k_{12}, EGP_0, k_{b1}, k_{b2}, k_{b3}, S_{IT}, S_{ID}, S_{IE}, k_a and k_e are time-varying with $\pm 5\%$ deviation. In the in silico tests, this is represented by sinusoidal oscillations superimposed on the nominal values with three hour period and a randomly generated phase.

2.1 Model reduction

A high-order model is rather difficult to handle, let it be analysis, identification, observer, or controller design, regardless of the method used. Given the limited measurement capabilities, we can only acquire information from the dominant components of the model dynamics. It would be advantageous to reduce the model to one that retains the most characteristic properties, yet the error resulting from this simplification is minimal. For example, early diabetes models used only three state variables [38], and it is not uncommon in ICU patient models to use only 3-5 state variables [39]. However, the errors resulting from model reduction

must be considered in the controller design.

For the states that belong to meal absorption (which is considered disturbance), it is easy to choose a linear substitution that represents worst-case meal intake. Let us use the notation $W_{meal}(s)$ for the transfer function of this model:

$$W_{meal}(s) = \frac{U_{G,ceil}}{t_{max}s+1}. \tag{3}$$

It is frequent among the most commonly used T1DM models [9, 10, 37, 40, 41] to contain a second-order nonlinear component. In the Cambridge model, these are denoted as $Q_1(t)$ and $Q_2(t)$. Aside from this and the meal absorption, the rest of the dynamics are entirely linear. We can distinguish two main parts: subcutaneous glucose transfer and insulin dynamics. The former can be neglected if an adequate state observer can accurately estimate plasma glucose concentration. The rest incorporates the transfer of the fast-acting insulin from the subcutaneous regions to the plasma, insulin degradation, and insulin effect. The overall model can be simplified by truncating this single input-multiple output linear system.

In the case of (1), the transfer speed between specific compartments is comparable to the sampling time of the CGM. Hence, the states associated with them can be eliminated. The resulting reduced model is as follows:

$$\begin{aligned}
\dot{Q}_1(t) = \; & -\left(\frac{F_{01}}{Q_1(t)+V_G}+x_1(t)\right)Q_1(t)+k_{12}Q_2(t)- \\
& -R_{cl}\max\{0,Q_1(t)-R_{thr}V_G\}-Phy(t)+ \\
& +EGP_0\max\left\{0,1-\frac{k_aS_{IE}}{V_Ik_e}S_2(t)\right\}+\min\left\{U_{G,ceil},\frac{G_2(t)}{t_{max}}\right\} \\
\dot{Q}_2(t) = \; & x_1(t)Q_1(t)-\left(k_{12}+\frac{k_aS_{ID}}{V_Ik_e}S_2(t)\right)Q_2(t) \\
\dot{x}_1(t) = \; & k_{b1}\left(\frac{k_aS_{IT}}{V_Ik_e}S_2(t)-x_1(t)\right) \\
\dot{S}_2(t) = \; & -k_aS_2(t)+k_aS_1(t) \\
\dot{S}_1(t) = \; & -k_aS_1(t)+u(t) \\
\dot{G}_2(t) = \; & \frac{G_1(t)-G_2(t)}{t_{abs}(t)} \\
\dot{G}_1(t) = \; & -\frac{G_1(t)}{t_{abs}(t)}+D(t)
\end{aligned} \tag{4}$$

where the output is $C(t) \approx Q_1(t)/V_G$.

Finally, considering the range of parameter k_{12} as presented in [37], if $1/k_{12}$ is comparable to the CGM sensor sample time, then there is a need to apply reduction in the nonlinear dynamics as well. Section 3.1 provides details on how to perform the necessary changes.

3 Controller Synthesis

3.1 Linear Parameter-Varying representation

Although continuous time T1DM models can have a varying degree of nonlinearity, the vast majority of them can be transformed into a linear parameter-varying model: [42]

$$\begin{array}{c} \dot{\mathrm{x}}(t) = \mathbf{A}(\rho(t))\mathrm{x}(t) + \mathbf{B}(\rho(t))\mathrm{u}(t) \\ \mathrm{y}(t) = \mathbf{C}(\rho(t))\mathrm{x}(t) + \mathbf{D}(\rho(t))\mathrm{u}(t) \\ \mathbf{A}(\rho(t)) = \mathbf{A}_0 + \sum_{i=1}^{m} \rho_i(t)\mathbf{A}_i \quad \mathbf{B}(\rho(t)) = \mathbf{B}_0 + \sum_{i=1}^{m} \rho_i(t)\mathbf{B}_i \\ \mathbf{C}(\rho(t)) = \mathbf{C}_0 + \sum_{i=1}^{m} \rho_i(t)\mathbf{C}_i \quad \mathbf{D}(\rho(t)) = \mathbf{D}_0 + \sum_{i=1}^{m} \rho_i(t)\mathbf{D}_i, \end{array} \tag{5}$$

where the scheduling variables $\rho_i(t)$ are bounded, as well as their time derivatives, and these bounds are known. Furthermore, the scheduling variables should be available for measurement. For a T1DM model, the bound constraint is satisfied, but only estimation of the scheduling variables is available.

The chosen LPV representation of the Cambridge model is the following:

$$\begin{aligned} \dot{C}(t) = \;& -k_{a,int}C(t) + \frac{k_{a,int}}{V_G}Q_1(t) \\ \dot{Q}_1(t) = \;& -\left(F_a\rho_1(t) + F_b\right)Q_1(t) - \rho_1(t)x_1(t) + k_{12}Q_2(t) - dist_{Rcl}(t) - \\ & -Phy(t) + EGP_0\left(1 - x_3(t)\right) + \frac{U_{G,ceil}}{t_{max}}\tilde{G}(t) \\ \dot{Q}_2(t) = \;& \rho_1(t)x_1(t) - k_{12}Q_2(t) - \rho_2(t)x_2(t) \\ \dot{x}_1(t) = \;& -k_{b1}x_1(t) + S_{IT}k_{b1}I(t) \\ \dot{x}_2(t) = \;& -k_{b2}x_2(t) + S_{ID}k_{b2}I(t) \\ \dot{x}_3(t) = \;& -k_{b3}x_3(t) + S_{IE}k_{b3}I(t) \\ \dot{I}(t) = \;& \frac{k_a}{V_I}S_2(t) - k_eI(t) \\ \dot{S}_2(t) = \;& -k_aS_2(t) + k_aS_1(t) \\ \dot{S}_1(t) = \;& -k_aS_1(t) + u(t) \\ \dot{\tilde{G}}(t) = \;& -\frac{1}{t_{max}}\tilde{G}(t) + dist_{meal}(t), \end{aligned} \tag{6}$$

which is constructed with the following considerations. The state variables Q_1 and Q_2 were chosen as scheduling variables: $\rho_1 = Q_1$ and $\rho_2 = Q_2$. The Hill function $F_{01}(Q_1(t)+V_G)^{-1}$ has a linear approximation $F_a\rho_1(t)+F_b$, as presented in [43] and [44]. The worst-case meal model (3) replaces the second-order meal intake. Moreover, to avoid switching control, endogenous glucose production (EGP) has no saturation, and the renal extraction (R_{cl}) is considered a disturbance: $dist_{Rcl}(t)$. In exchange, the controller must ensure that $x_3(t) <= 1$.

If model reduction is necessary, the same method is applicable as presented in Section 2.1. In addition, if $1/k_{12}$ is comparable to the CGM sensor sample time, then further model reduction is possible by replacing $Q_1(t)$ and $Q_2(t)$ with $\tilde{Q}(t) \approx Q_t(t) - x_1(t)/k_{12}$ as follows:

$$\begin{aligned}\dot{C}(t) = &\quad -k_{a,int}C(t) + \frac{k_{a,int}}{V_G}\tilde{Q}(t) - \frac{k_{a,int}}{k_{12}V_G}x_1(t)\\ \dot{\tilde{Q}}(t) = &\quad -(F_a\rho_1(t)+F_b)\tilde{Q}(t) + (F_c\rho_1(t)+F_d)x_1(t) - dist_{Rcl}(t) -\\ &\quad -Phy(t) - \rho_2(t)x_2(t) + EGP_0(1-x_3(t)) + \frac{U_{G,ceil}}{t_{max}}\tilde{G}(t),\end{aligned} \tag{7}$$

where we used the approximation $F_c\rho_1(t)+F_d \approx Q_1(t)F_{01}/(k_{12}(Q_1(t)+V_G))$. Furthermore, the output can be approximated with

$$C(t) \approx \frac{1}{V_G}\left(\tilde{Q}(t) - \frac{1}{k_{12}}x_1(t)\right), \tag{8}$$

if necessary.

Note, that even though $Q_1(t)$ and $Q_2(t)$ are not part of the model (7), the scheduling variables $\rho_1(t)$ and $\rho_2(t)$ are still present, hence the state observer must provide reliable estimation, and cannot use (7) instead of (1).

Considering the nonlinearity of the Cambridge model and the presence of disturbances, sigma-point filters are good candidates for estimating the scheduling variables [45–47]. These filters provide further benefits of reducing the measurement noise of CGMS, estimating the glucose flux from unannounced meal intakes, as well as enabling state feedback control.

The state observer used in this work is the same as presented in [43], i.e., we use unscented square-root filter with lognormal distribution for state variables associated with glucose concentration and meal intake.

3.2 Modeling Uncertainties

Not even one of the most complex T1DM models [48] can fully capture a system as complex as human metabolism. All models used in practice employ some level of simplification, capturing only the most essential aspects of the glucose-insulin interaction. Additionally, human physiology tends to change and adapt over time and is affected by various external factors that are difficult to quantify, let alone measure [49, 50]. No matter the model used, controller design must address deviation from the actual dynamics. This approach uses uncertainty weighting functions to that end. Let P denote a nominal T1DM model, with a single output $y(t)$, a single controlled input $u(t)$, and a set of external disturbances $\mathrm{d}(t)$. Figure 1 presents a simple system that uses output uncertainty weighting functions.

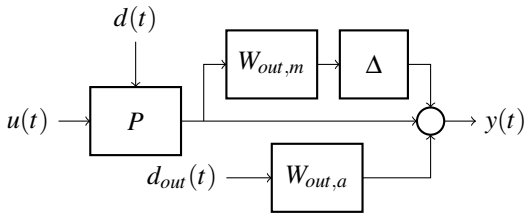


Figure 1
Simple system with output uncertainty weighting functions

$W_{out,a}$ and $W_{out,m}$ are linear minimum-phase systems representing additive and multiplicative output uncertainty, respectively. Δ is an unknown linear system that is of minimum-phase and has an $\mathscr{H}_\infty$-norm that is less or equal to 1. $d_{in}(t)$ and $d_{out}(t)$ are considered stochastic disturbance signals, that are usually modeled with distribution $\mathscr{N}(0,1)$ or $\mathscr{U}(-1,1)$.

There are other ways to model uncertainties, but this work only considers output uncertainties. The parameters of $W_{out,a}$ and $W_{out,m}$ may come from the residual error of fitting T1DM parameters or from a priori knowledge from clinical practice. Here the uncertainty model accounts for the sinusoidal changes in model parameters presented in section 2, and any reduction ((7) and section 2.1) that may take place.

Extending the nominal model with uncertainties is vital for both the controller and state observer, as it can ensure robustness and stability for both control and state estimation. However, it also poses a challenge if the control law is realized with state feedback. In this particular case, the state variables of the uncertainty models shall be estimated as well.

3.3 LPV Model in the Clinical Practice

Before presenting the controller synthesis, it is worth examining the LPV transformed model from a different perspective. In (7), in the subsystem consisting of $Q_1(t)$ and $Q_2(t)$, $\rho_1(t)$ defines the rate of transfer of glucose from $Q_1(t)$ to $Q_2(t)$ compartments, and $\rho_2(t)$ is responsible for how fast glucose in $Q_2(t)$ dissipates from the system. Since $\rho_1(t)$ is identical to $Q_1(t)$, and the output is $\approx Q_1(t)/V_G$, hence the larger the glucose levels are, the faster the dynamics of the system. This is a common observation in clinical practice as well.

Consequently, this poses a challenge to controller design: Suppose the controller injects too much insulin during hyperglycemia. In that case, it is difficult to compensate for its effect once the glucose levels decrease, not only because negative control signal is not an option but also because the system will be less sensitive to any intervention.

3.4 Robust Controller Synthesis

In order to address the need for robustness and stability, $\mathscr{H}_\infty$ and hybrid $\mathscr{H}_\infty/\mathscr{H}_2$ controllers [51] can provide a solid foundation. Let $\mathbf{G}(j\omega,\rho)$ denote the transfer matrix of an LPV (5) closed-loop system with an $\mathscr{H}_\infty$ controller, that was successfully synthesized for a positive γ value. This controller ensures that $\|\mathbf{G}\|_\infty$, the $\mathscr{H}_\infty$ norm of the transfer matrix is smaller than γ:

$$\|\mathbf{G}\|_\infty = \sup_{\rho} \sup_{\omega\in\mathbb{R}} \sigma_{max}(\mathbf{G}(j\omega,\rho)) < \gamma, \tag{9}$$

where σ denotes the singular value. An $\mathscr{H}_2$ controller on the other hand ensures that $\|\mathbf{G}\|_2$, the $\mathscr{H}_\infty$ norm of the transfer matrix is smaller than γ:

$$\|\mathbf{G}\|_2 = \sup_{\rho} \sqrt{\frac{1}{2\pi} trace \int_{-\infty}^{\infty} \mathbf{G}(j\omega,\sup_{\rho})(\mathbf{G}(j\omega,\sup_{\rho}))^H d\omega} < \gamma. \tag{10}$$

$\mathscr{H}_\infty$ and $\mathscr{H}_2$ constraints can be imposed on the controlled system with carefully chosen linear matrix inequalities (LMIs) [52].

Figure 2 and Figure 3 display two different schemes: LPV state feedback, denoted with $K_{fb}(\rho)$, and dynamic LPV controller, denoted with $K_d(\rho)$.

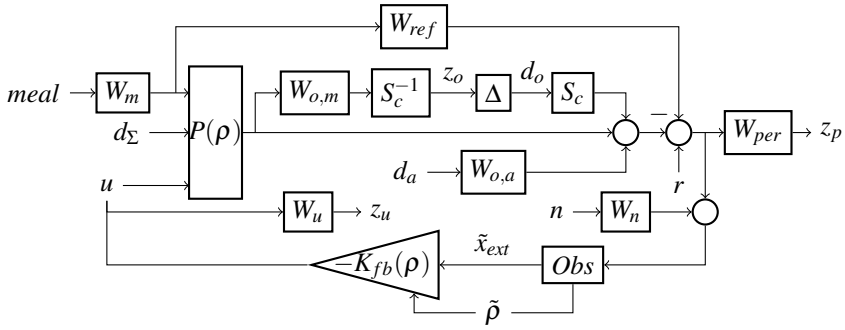


Figure 2
Controller realized as LPV state feedback

$P(\rho)$ represents the LPV transformed T1DM model (6) with reduction applied if necessary, and the meal absorption dynamics separated into W_m. *Obs* denote the sigma point filter that provides state variable ($\tilde{x}_{ext}$) and scheduling variable estimation ($\tilde{\rho}$). $W_{o,m}$ and $W_{a,m}$ are weighting functions for multiplicative and additive output uncertainty models, respectively, with the latter driven by d_a. W_{per} defines expected tracking performance. W_u serves two purposes:

1. it defines the maximum value for the control signal;
2. it ensures that endogenous glucose production does not saturate.

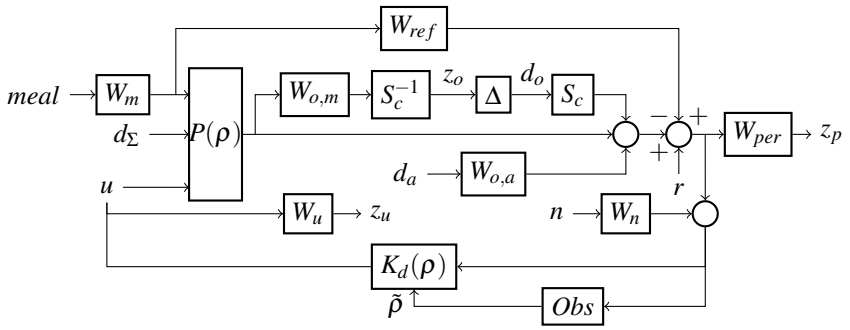


Figure 3
Controller realized as dynamic LPV system

d_Σ represents all disturbance signals affecting $P(\rho)$ directly: *Phy* physical activity, the estimation error of glucose flux from the gut, and quantization error if the control signal has a finite resolution. Since an estimated output of W_m is available, it is possible to define reference tracking dynamics W_{ref} that mimics the behavior of a healthy patient. Using the output of W_{ref} instead of relying on a constant reference signal r for feedback can potentially help avoid hypoglycemia. Finally, W_n noise model is not present for state feedback control.

The state observer works with (1) instead of (6), and will use $W_{o,m}$, $W_{a,m}$, and W_n regardless of the control method used. Section 3.5 details the role of scaling factor S_c.

Pure $\mathscr{H}_\infty$ controller will ensure minimal $\mathscr{H}_\infty$ norm for the transfer to the performance output z_p and smaller than 1 $\mathscr{H}_\infty$ norm to z_o and z_u from the disturbance inputs *meal*, d_a, d_o, d_Σ and n. The latter ensures robust stability for the controlled system. Hybrid $\mathscr{H}_\infty/\mathscr{H}_2$ controller minimizes the $\mathscr{H}_2$ norm for the performance output instead of $\mathscr{H}_\infty$ norm.

Finally, additional LMI constraints limit the poles of the controlled system to a defined region, reducing oscillatory transients and rejecting poles too fast compared to the sampling frequency of the CGM sensor (Figure 4).

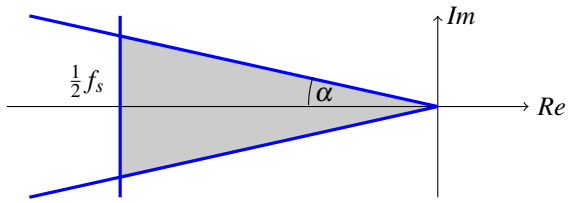


Figure 4
Constraints for the poles of the controlled system: f_s is the sampling frequency of the CGM sensor, α is the maximum angle of the complex conjugate pole pairs

The control law for LPV state feedback is as follows:

$$u(t) = -\mathrm{K}_{fb}(\rho(t))\tilde{x}_e xt(t), \tag{11}$$

where $\tilde{x}_{ext}(t)$ is the estimated state variables of $P(\rho)$, W_m, W_u, $W_{o,m}$, $W_{o,a}$, W_{ref}, and W_{per}.

If the controller is a dynamic LPV system (5) instead, the control signal is the output of the following:

$$\begin{array}{rl} \dot{\mathrm{x}}_d(t) & = \mathbf{A}_d(\rho(t))\mathrm{x}_d(t) + \mathrm{B}_d(\rho(t))e(t) \\ u(t) & = \mathrm{C}_d(\rho(t))\mathrm{x}_d(t) + D_d(\rho(t))e(t). \end{array} \tag{12}$$

3.5 Scaling

In order to achieve robust stability, the $\mathscr{H}_\infty$ norm of the transfer function from the disturbance inputs to the output of W_u and $W_{out,m}$ shall be smaller than 1. On the other hand, the glucose levels can reach up to 17 mmol/L, even with a well-functioning artificial pancreas.

Hence, when using the multiplicative output uncertainty model, the output of the corresponding weighting function (z_o in Figure 2 and Figure 3) should be scaled (S_c) to satisfy robust stability constraints. If S_c is too small, the controller synthesis is infeasible. Conversely, too large scaling factor will lead to reduced disturbance rejection performance. In this work, we used an iterative algorithm to set a patient-specific scaling factor.

4 Results

All simulations were performed on the original eight patient parameter sets introduced in [37]. The weighting functions presented in Figure 2 and Figure 3 are as follows:

- $P(\rho)$ is the LPV-transformed model (6), with patient-specific reduction applied only to remove state variables that would otherwise introduce time constants less than 20 minutes.
- Transfer function (3) is chosen as W_m in Figure 2 and Figure 3, using patient-specific parameters.
- $W_{o,m}$ is a low pass filter that represents 10% to 25% multiplicative uncertainty on frequencies below $\frac{2\pi}{120}$ [rad/min]:

  $$W_{o,m}(s) = \frac{0.1}{120s+1} \tag{13}$$

  The higher uncertainty value is necessary when working with a linear model and controller instead of LPV.

- $W_{o,a}$ assume ± 0.5 [mmol/L] additive uncertainty on top of the multiplicative uncertainty.
- W_u is patient-specific, ensuring that endogenous glucose production does not reach zero and that u is smaller than 4500 [mU/min].
- The sensor noise is assumed to be Gaussian white noise with 0.1 [mmol/L] standard deviation, hence $n \approx \mathcal{N}(0,1)$ and $W_n = 0.1$.
- W_{per} specifies that the controller should only minimize tracking error below $\frac{2\pi}{180}$ [rad/min]:

$$W_{per}(s) = \frac{1}{180s+1} \tag{14}$$

- If applicable, the reference dynamics W_{ref} is:

$$W_{ref}(s) = \frac{11}{U_{G,ceil}(60s+1)} \tag{15}$$

  The constant reference signal is 4.5 [mmol/L] and 4.9 [mmol/L] with and without W_{ref} respectively.

We used MATLAB and SIMULINK for controller synthesis and simulation, including the CVX toolbox [53, 54]. For each controller, $\alpha = 45^\circ$ and 5 minutes sampling time constrain the closed-loop poles as shown in (Fig. 4).

For easier comparison with other methods, the simulations were done using two commonly used meal intake scenarios spanning 24 hours. The controller administered the insulin without manual intervention and any meal intake announcement. The meal intake scenarios were as follows:

1. 150 g of carbohydrate (CHO) intake per day. The meal intake consists of a 35 g CHO breakfast at 8:30, a 65 g CHO lunch at 13:00, and a 50 g CHO dinner at 19:00.
2. The meal intake protocol is presented in [55]. It consists of a 45 g CHO breakfast at 9:30, a 75 g CHO lunch at 13:30, and an 85 g CHO dinner at 19:30.

Control variability grid analysis (CVGA) [55] visualizes and compares the capabilities of different types of controller configurations on Figure 5, Figure 6, Figure 7, and Figure 8. The x and y axis is the minimum and maximum glucose levels throughout the simulations, respectively, in [mg/dL]. The aim is to stay in the A and B zones for both meal scenarios for all virtual patients. In each figure, two sets of simulation results are presented: black circles represent simulations using a state feedback controller. In contrast, the simulations result when a dynamic LPV controller was applied are represented with white circles.

Figure 5 displays the CVGA of simulations results for linear $\mathscr{H}_\infty$ controllers, without reference dynamics W_{ref}. The output multiplicative uncertainty is 25%, as ensured by (13). Both the state feedback and the dynamic controller performed poorly, reaching only C and D zones with the majority of virtual patients.

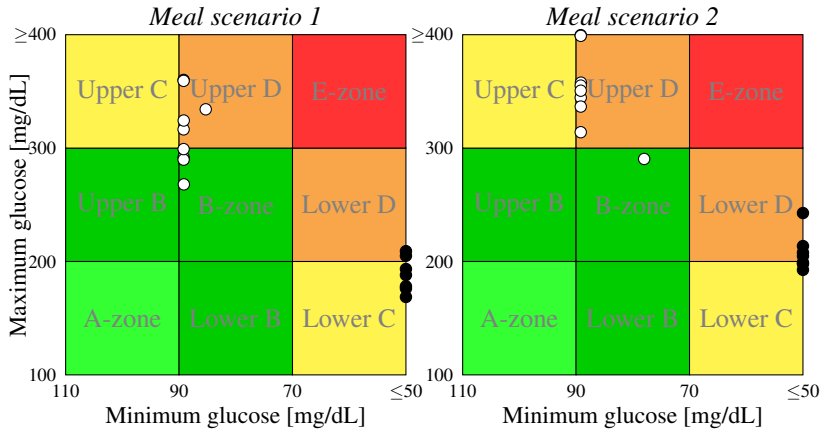


Figure 5
CVGA for linear $\mathscr{H}_\infty$ controllers using (**a**) meal scenario 1 and (**b**) meal scenario 2.

Since the linearization of the model was at 4.9 [mmol/L] ($\approx$ 90 [mg/dL]), both controllers assume slow dynamics, which does not hold once the glucose levels elevate after each meal intake. However, the behavior of the two controller configurations is different. The state feedback controller administers more insulin than necessary, leading to lower maximum values but severe hypoglycemia. In contrast, the dynamic controller avoids hypoglycemia at the cost of higher maximum glucose levels.

Choosing a different working point with a higher glucose level for linearization could improve the capabilities of both controllers. Furthermore, using hybrid $\mathscr{H}_\infty/\mathscr{H}_2$ norm or reference tracking dynamics can bring some minor improvement. However, using a purely linear controller has its limits, and hence a nonlinear approach is necessary.

Using LPV model and LPV controllers reduce the occurrence of both hypo- and hyperglycemia, as shown in Figure 6. Since the controllers directly address the changing dynamics of the model, more than half of the virtual patients are kept in the B-zone for both controller types.

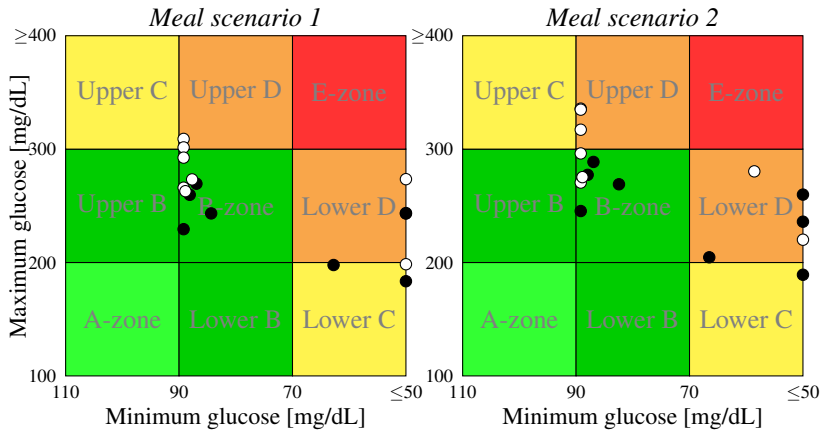


Figure 6
CVGA for LPV $\mathscr{H}_\infty$ controllers using (**a**) meal scenario 1 and (**b**) meal scenario 2.

However, there are cases of severely low glucose concentrations for both controller types. Introducing reference tracking dynamics can mitigate these hypoglycemic episodes, as shown in Figure 7. Instead of a constant reference signal at 4.9 [mmol/L], the additional weighting function $W_{ref}(s)$ defines the desired disturbance rejection. The input of $W_{ref}(s)$ is the estimated glucose flux resulting from meal intake. Both the occurrence and severity of hypoglycemia decreased for the two types of controllers compared to Figure 6, although they are not eliminated completely. The maximum glucose concentration is below 300 [md/dL] for the state feedback controller but increased for the dynamic controller.

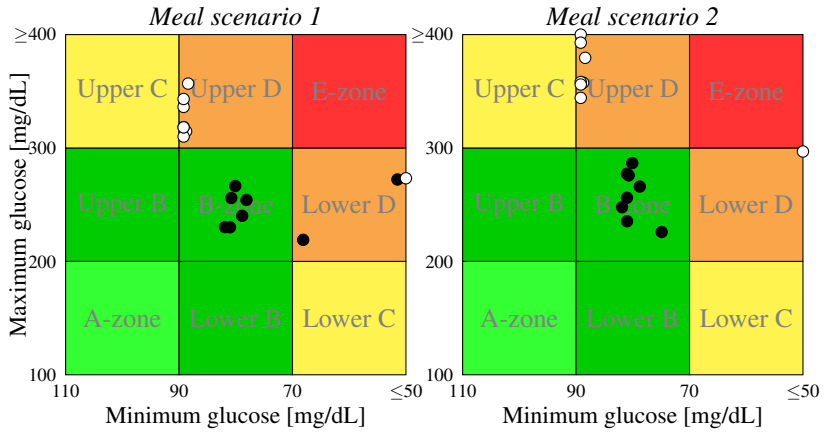


Figure 7
CVGA for state-feedback and dynamic controllers using reference tracking dynamics W_{ref} for (**a**) meal scenario 1 and (**b**) meal scenario 2.

Finally, the results can be further improved by using $\mathscr{H}_2$ norm for the performance output instead of $\mathscr{H}_\infty$. The results are displayed in Figure 8. As a result, the minimum glucose values in both meal scenarios and controller types have less variance and hence lessen the chance of hypoglycemia.

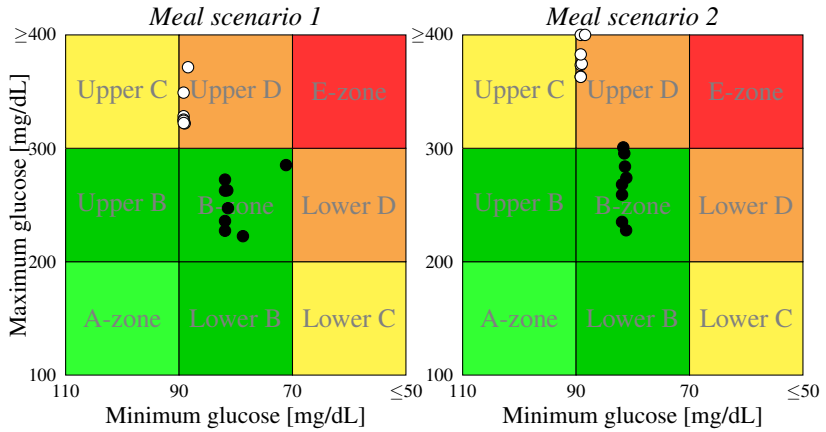


Figure 8
CVGA for LPV $\mathscr{H}_\infty/\mathscr{H}_2$ controllers with W_{ref} using (**a**) meal scenario 1 and (**b**) meal scenario 2.

Conclusions

Based on the simulation results presented in section 4, a linear controller is insufficient for glycemic control. LPV controllers are more reliable than linear ones, with less variability in both maximum and minimum glucose levels for the used meal intake protocols. However, they are still prone to hypoglycemia. Using reference tracking dynamics instead of a constant reference signal can considerably lessen both the occurrence and the severity of these events without compromising the severity of hyperglycemia. Moreover, using a hybrid norm for controller synthesis: $\mathscr{H}_\infty$ for stability and $\mathscr{H}_2$ for performance can provide further benefits. Generalized $\mathscr{H}_2$ and $\mathscr{L}_1$ norm [52] was not in the scope of this work. However, there is a significant difference between using state feedback or a dynamic controller, even though the combination of a sigma point filter and an LPV state feedback is technically a dynamic controller as well. Using the same constraints and extended model, the resulting dynamic controller will lead to higher minimum and maximum glucose levels in the CVGA compared to a state feedback controller. The reason is that the transfer function of the former across all scheduling variables is akin to a high pass filter. Therefore, it will compensate low frequency tracking errors poorly. Introducing an integrator [51] to eliminate this error will result in an infeasible convex synthesis problem.

Even though the presented results may be satisfactory, there are still limitations that must be addressed in future works. LPV model can represent most of the nonlinearity present in the Cambridge model, but it does not capture the positive nature of the system. All the measures to avoid hypoglycemia only indirectly

addressed the constraint that the control signal is non-negative. Despite the parameter variability in the Cambridge model, changes in insulin sensitivity or using different types of insulin are not part of it. Furthermore, a robust controller combined with an accurate meal and fault detection [56] can increase both performance and safety [49, 50]. Finally, a crucial next step is to incorporate the detection of unannounced physical activity. One potential approach is to use the model presented by Resalat et al. [57] is an extension of the Cambridge model that includes a dynamic physical activity subsystem. Early prediction of potential hypoglycemic episodes is necessary since a single hormone system cannot raise the plasma glucose concentration.

Further work shall address the limitations mentioned above and perform validation with the UVA/PADOVA simulator, which is approved by the U.S. Food and Drug Administration (FDA) as an alternative to animal testing of Type 1 diabetes control strategies [11].

Acknowledgment

Project no. 2019-1.3.1-KK-2019-00007. has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2019-1.3.1-KK funding scheme. This project has been supported by the Hungarian National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NKTA-36 funding scheme.

References

[1] A. Fonyo and E. Ligeti. *Physiology*. Medicina, Budapest, third edition, 2008.

[2] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King. Global prevalence of diabetes. Estimates for the year 2000 and projections for 2030. *Diabetes Care*, 27(5):1047–1053, 2004.

[3] R. Buzzetti, S. Zampetti, and P. Pozzilli. Impact of obesity on the increasing incidence of type 1 diabetes. *Diabetes, obesity & metabolism*, 22(7):1009–1013, 2020.

[4] H. Thabit and R. Hovorka. Coming of age: the artificial pancreas for type 1 diabetes. *Diabetologia*, 59(9):1795–1805, 2016.

[5] J. Tašić, M. Takács, and L. Kovács. Control engineering methods for blood glucose levels regulation. *Acta Polytechnica Hungarica*, 19:127–152, 08 2022.

[6] J. Kropff and J. H. DeVries. Continuous glucose monitoring, future products, and update on worldwide artificial pancreas projects. *Diabetes Technology & Therapeutics*, 18(S2), 2016.

[7] G. Forlenza, S. Deshpande, T. Ly, D. Howsmon, F. Cameron, N. Baysal, E. Mauritzen, T. Marcal, L. Towers, B. Bequette, L. Huyett, J. Pinsker, R. Gondhalekar, F. Doyle, D. Maahs, B. Buckingham, and E. Dassau. Application of zone model predictive control artificial pancreas during extended use of infusion set and sensor: A randomized crossover-controlled home-use trial. *Diabetes Care*, 40:dc170500, 06 2017.

[8] P. Colmegna, F. Garelli, H. De Battista, F. Bianchi, and R. S. Sánchez-Peña. The arg algorithm: clinical trials in Argentina. In R. S. Sánchez-Peña and D. R. Cherñavvsky, editors, *The Artificial Pancreas*, pages 79–104. Academic Press, 2019.

[9] R. Hovorka, V. Canonico, L. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. Pieber, H. Schaller, L. Schaupp, T. Vering, and M. Wilinska. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiological measurement*, 25:905–920, 2004.

[10] L. Magni, D. M. Raimondo, C. D. Man, G. D. Nicolao, B. Kovatchev, and C. Cobelli. Model predictive control of glucose concentration in type 1 diabetic patients: An in silico trial. *Biomedical Signal Processing and Control*, pages 338–346, 2009.

[11] C. D. Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, and C. Cobelli. The UVA/PADOVA type 1 diabetes simulator: New features. *Journal of Diabetes Science and Technology*, 8(1):26–34, 2014.

[12] H. Kirchsteiger, S. Pölzer, R. Johansson, E. Renard, and L. del Re. Direct continuous time system identification of MISO transfer function models applied to type 1 diabetes. *In Proc. of 50th IEEE CDC-ECC Conference, Orlando, USA*, pages 5176–5181, 2011.

[13] G. Eigner, K. Koppány, P. Pausits, and L. Kovács. Nonlinear identification of glucose absorption related to diabetes mellitus. In *2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES)*, pages 265–270, 2017.

[14] L. Kovács and P. Szalay. Uncertainties and modeling errors of type 1 diabetes models. *Prediction Methods for Blood Glucose Concentration: Design, Use and Evaluation*, pages 211–225, 2016.

[15] R. Blanc, H. M. R. Ugalde, P.-Y. Benhamou, G. Charpentier, S. Franc, E. Huneker, E. Villeneuve, and M. Doron. Modeling the variability of insulin sensitivity for people with type 1 diabetes based on clinical data from an artificial pancreas study. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5465–5468, 2019.

[16] T. Battelino and J. Bolinder. Clinical use of real-time continuous glucose monitoring. *Current Diabetes Reviews*, 4:218–222, 2008.

[17] A. Facchinetti, S. Del Favero, G. Sparacino, J. R. Castle, W. K. Ward, and C. Cobelli. Modeling the glucose sensor error. *IEEE Transactions on Biomedical Engineering*, 61(3):620–629, 2014.

[18] M. Vettoretti, G. Cappon, G. Acciaroli, A. Facchinetti, and G. Sparacino. Continuous glucose monitoring: Current use in diabetes management and possible future applications. *Journal of Diabetes Science and Technology*, 12(5):1064–1071, 2018.

[19] M. Infante, D. A. Baidal, M. R. Rickels, A. Fabbri, J. S. Skyler, R. Alejandro, and C. Ricordi. Dual-hormone artificial pancreas for management of type 1 diabetes: Recent progress and future directions. *Artificial Organs*, 45(9):968–986, 2021.

[20] L. M. Wilson and J. R. Castle. Stable liquid glucagon: Beyond emergency hypoglycemia rescue. *Journal of Diabetes Science and Technology*, 12(4):847–853, 2018. PMID: 29415555.

[21] M. Somogyi and M. Kirstein. Insulin as a cause of extreme hyperglycemia and instability. *Weekly Bulletin of the St Louis Medical Society*, 32:498–510, 1938.

[22] T. Mohammadridha, P. Rivadeneira, N. Magdelaine, M. Cardelli, and C. Moog. Positively invariant sets of a t1dm model: Hypoglycemia prediction and avoidance. *Journal of the Franklin Institute*, 356, 05 2019.

[23] S. Mehmood, I. Ahmad, H. Arif, U. Ammara, and A. Majeed. Artificial pancreas control strategies used for type 1 diabetes control and treatment: A comprehensive analysis. *Applied System Innovation*, 3:31, 07 2020.

[24] F. Shmarov, N. Paoletti, E. Bartocci, S. Lin, S. A. Smolka, and P. Zuliani. Smt-based synthesis of safe and robust pid controllers for stochastic hybrid systems. In O. Strichman and R. Tzoref-Brill, editors, *Hardware and Software: Verification and Testing*, pages 131–146, Cham, 2017. Springer International Publishing.

[25] D. Calupiña, A. García, O. Camacho, A. Rosales, and P. Rivadeneira. Nonlinear PID and dynamic SMC for the artificial pancreas control in the treatment of type 1 diabetes. In *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, pages 1–6, 2018.

[26] K. Turksoy and A. Cinar. Adaptive control of artificial pancreas systems - a review. *Journal of Healthcare Engineering*, 5, 2014.

[27] T. Zhu, K. Li, P. Herrero, and P. Georgiou. Basal glucose control in type 1 diabetes using deep reinforcement learning: An in silico validation. *IEEE Journal of Biomedical and Health Informatics*, 25(4):1223–1232, 04 2021.

[28] J. C. Peiró. Comparing artificial pancreas controlled by hybrid "closed-loop" machine learning (ML) trained algorithm to multi-daily injection (MDI), insulin pump without CGM and "sensor assisted" insulin pump therapies for diabetes type 1 (DT1) treatment. In *Proceedings of the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, pages 1–6, 2020.

[29] A. Mohammadzadeh and T. Kumbasar. A new fractional-order general type-2 fuzzy predictive control system and its application for glucose level regulation. *Applied Soft Computing*, 91:106241, 2020.

[30] B. Kovatchev, C. Cobelli, and E. Renard. Multi-national study of subcutaneous model-predictive closed-loop control in type 1 diabetes: summary of the results. *Journal of Diabetes Science and Technology*, 4:1374–1381, 2010.

[31] D. Shi, E. Dassau, and F. J. Doyle III. Multivariate learning framework for long-term adaptation in the artificial pancreas. *Bioengineering & Translational Medicine*, 4(1):61–74, 2019.

[32] D. Boiroux, M. Hagdrup, Z. Mahmoudi, K. Poulsen, H. Madsen, and J. B. Jørgensen. An ensemble nonlinear model predictive control algorithm in an artificial pancreas for people with type 1 diabetes. In *Proceedings of the 2016 European Control Conference (ECC)*, pages 2115–2120, 2016.

[33] N. Paoletti, K. S. Liu, H. Chen, S. Smolka, and S. Lin. Data-driven robust control for a closed-loop artificial pancreas. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6):1981–1993, 11 2020.

[34] R. Parker, F. Doyle, J. Ward, and N. Peppas. Robust $\mathscr{H}_\infty$ glucose control in diabetes using a physiological model. *American Institute of Chemical Engineers*, 46(12):2537–2549, 2000.

[35] G. Rigatos, P. Siano, and A. Melkikh. A nonlinear optimal control approach of insulin infusion for blood glucose levels regulation. *Intellectual Industrial Systems*, 3(2):91–102, 11 2017.

[36] A. Mirzaee, M. Dehghani, and M. Mohammadi. Robust LPV control design for blood glucose regulation considering daily life factors. *Biomedical Signal Processing and Control*, 57:101830, 2020.

[37] M. Wilinska, L. Chassin, C. Acerini, J. Allen, D. Dunger, and R. Hovorka. Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes. *Journal of diabetes science and technology*, 4(1):132–144, 2010.

[38] R. Bergman, L. Phillips, and C. Cobelli. Physiological evaluation of factors controlling glucose tolerance in man. *Journal of Clinical Investigation*, 68:1456–1467, 1981.

[39] J. G. Chase, B. Benyo, and T. Desaive. Glycemic control in the intensive care unit: A control systems perspective. *Annual Reviews in Control*, 48:359–368, 2019.

[40] C. Dalla Man, R. Rizza, and C. Cobelli. Meal simulation model of the glucose-insulin system. *IEEE Transactions on Biomedical Engineering*, 54(10):1740–1749, 2007.

[41] P. Colmegna and R. Sánchez-Peña. Analysis of three T1DM simulation models for evaluating robust closed-loop controllers. *Computer methods and programs in biomedicine*, 113, 10 2013.

[42] F. Bianchi, M. Moscoso-Vásquez, P. Colmegna, and R. Sánchez-Peña. Invalidation and low-order model set for artificial pancreas robust control design. *Journal of Process Control*, 76:133–140, 02 2019.

[43] P. Szalay, Z. Benyó, and L. Kovács. Long-term prediction for T1DM model during state-feedback control. In *Proceedings of the 2016 12th IEEE International Conference on Control and Automation (ICCA)*, pages 311–316, 2016.

[44] P. Szalay, G. Eigner, and L. A. Kovács. Linear matrix inequality-based robust controller design for type-1 diabetes model. *Proceedings of the 19th IFAC World Congress*, 47(3):9247–9252, 2014. 19th IFAC World Congress.

[45] D. Boiroux, T. B. Aradóttir, M. Hagdrup, N. K. Poulsen, H. Madsen, and J. B. Jørgensen. A bolus calculator based on continuous-discrete unscented kalman filtering for type 1 diabetics. *IFAC Papers Online*, 48(20):159–164, 2015. 9th IFAC Symposium on Biological and Medical Systems BMS 2015.

[46] Z. Mahmoudi, K. Nørgaard, N. K. Poulsen, H. Madsen, and J. B. Jørgensen. Fault and meal detection by redundant continuous glucose monitors and the

unscented Kalman filter. *Biomedical Signal Processing and Control*, 38:86–99, 2017.

[47] E. Aguirre-Zapata, J. Cardenas-Cartagena, and J. Garcia-Tirado. Glycemic monitoring in critical care using nonlinear state estimators. *Proceedings of the 20th IFAC World Congress*, 50(1):4430–4435, 2017.

[48] J. Sorensen. *A Physiologic Model of Glucose Metabolism in Man and its Use to Design and Assess Improved Insulin Therapies for Diabetes*. PhD thesis, Massachusetts Institute of Technology, 1985.

[49] K. Turksoy, S. Samadi, J. Feng, E. Littlejohn, L. Quinn, and A. Cinar. Meal detection in patients with type 1 diabetes: A new module for the multivariable adaptive artificial pancreas control system. *IEEE Journal of Biomedical and Health Informatics*, 20(1):47–54, 2016.

[50] L. Meneghetti, A. Facchinetti, and S. D. Favero. Model-based detection and classification of insulin pump faults and missed meal announcements in artificial pancreas systems for type 1 diabetes therapy. *IEEE Transactions on Biomedical Engineering*, 68(1):170–180, 2021.

[51] K. Zhou. *Robust and Optimal Control*. Prentice Hall, New Jersey, 1996.

[52] C. Scherer and S. Weiland. *Linear Matrix Inequalities in Control*. Lecture Notes, Delft Center for Systems and Control, Delft, The Netherlands, 2004.

[53] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 03 2014.

[54] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[55] L. Magni, D. Raimondo, C. D. Man, M. Breton, S. Patek, G. D. Nicolao, C. Cobelli, and B. Kovatchev. Evaluating the efficacy of closed-loop glucose regulation via control-variability grid analysis. *Journal of diabetes science and technology*, 2(4):630–635, 2008.

[56] L. Pokorádi, S. Koçak, and E. T'oth-Laufer. Fuzzy failure modes and effects analysis using summative defuzzification methods. *Acta Polytechnica Hungarica*, 18(9):111–126, 2021.

[57] N. Resalat, J. Youssef, R. Reddy, and P. Jacobs. Design of a dual-hormone model predictive control for artificial pancreas with exercise model. In *Proceedings of 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016*, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 2270–2273. Institute of Electrical and Electronics Engineers Inc., October 2016.

Staining Independent Nonrigid Iterative Registration Method, for Microscopic Samples

Róbert Paulik[1], Viktor Jónás[1], Miklós Vincze[2], Miklós Kozlovszky[2] and Béla Molnár[1]

[1] Image Analysis Department, 3DHISTECH Ltd., 1141 Budapest, Hungary; robert.paulik@3dhistech.com

[2] Department of BioTech Research Center, Óbuda University, 1034 Budapest, Hungary; kozlovszky.miklos@nik.uni-obuda.hu

Abstract: Using digital microscope scanners, gigapixel-scale images for tissue samples are scanned in a minute, which provides an opportunity for quantitative evaluation at the cellular or gene level. However, to make an accurate diagnosis for clinical or research cases, it is necessary to make serial sections and stain them using different reagents. Since digital scanning and processing are preceded by manual workflows, the orientations between the images are lost. In the absence of adjustment, we cannot compare them to each other, for colocalization or correlation analysis. A registration method is needed that organizes the samples in the same orientation. The proposed method is inspired by the traditional and deep-learning based registration methods (SURF, SIFT, ORB, SuperPoint, SuperGlue) and further developed to manage the tearing, creasing and other deformations between the samples. Based on the validation results, the basic methods give moderate results, however, by utilizing a grid-based approach and by choosing the appropriate number of recursive iterations and resolution, the methods can be improved. The proposed stain-independent, iterative, non-rigid registration method can manage not only tears, creases and deformations, but also correct structural changes between series sections.

Keywords: digital pathology; digital microscope; stain-independent; image registration; iterative; recursive; non-rigid; elastic; deep-learning; convolutional neural network

1 Introduction

The integration of digital imaging in medical diagnostics first began in radiology, and due to the benefits of the digital file, such as the ability to share, integrate, and archive, the same request has emerged in routine histopathology. The digital revolution began with the introduction of whole slide imaging (WSI) technology in pathology. Various scanner devices creating large files have been introduced, presenting tissue structures in an appropriate resolution with a high color fidelity [1] [2].

In digital pathology, sections stained with different reagents are used. To examine these sections together, it is necessary to register them with each other. Since the sectioning and floating up are manual and mechanical processes, the tissue sample can be deformed to a great extent by the time it is placed on the slide (Figure 1). Furthermore, if the physical distance between the sections was too large, they may not have the same structure visually. Similarly, different reagents may stain different tissue structures. Because of all this, we need a method that is flexible enough to register visually slightly different and deformed tissue sections.

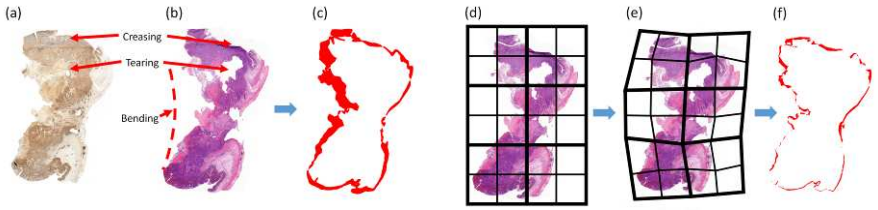


Figure 1

Illustration of deformations resulting from sample preparation and the required registration for correction: (a) source sample (MSH2 stained); (b) target sample (H&E stained); (c) difference of input samples; (d) Subdivided target sample; (e) registered target sample; (f) difference of registered samples

A fundamental problem with many computer vision tasks is finding visual correspondence between similar images. Stereo vision [3], object recognition [4], image stitching [5], visual odometry [6] are tasks which need a method for the registration problem. The feature extraction, description, matching and correspondence estimator methods were highly researched areas during the last three decades, many methods have been proposed that have sought a solution to these. Basically, the registration methods can be divided into three main categories: traditional feature-based approaches, deep-learning feature-based methods and homography learning.

Traditional feature-based algorithms, like Scale Invariant Feature Transform (SIFT) [7] and Speeded Up Robust Features (SURF) [8] methods are excellent feature detection, description, and matching algorithms, but even when an advanced matching process is applied, a considerable number of incorrect matches remains and needs to be eliminated. Random Sample Consensus (RANSAC) [9] is a widely used algorithm for removing false matches.

Nowadays, state-of-the-art methods in the field of image registration concern the use of deep learning. The Generic Feature Learning method [10] and SuperPoint approach [11] are using a training step of a convolutional neural network using only unlabeled public image datasets. SuperGlue network [12] is based on the SuperPoint “keypoint” detector and descriptor, which matches two pointset with a combination of Graph Neural Network and Optimal Matching layer. These methods can outperform the traditional methods; by resulting less outliers during the feature matching, the homography estimation can be more accurate.

Homography learning methods tries to solve the tasks of the previous methods in one step. They use a neural network to directly learn the transformation between an image pair. Deep Image Homography Estimation method [13] describes a regression homography network, a Visual Geometry Group (VGG) style model which is able to learn the homography by two images.

The previous registration methods are generally formulated as an optimization problem that satisfies constraints, such as coordinate displacements that are affine or volume preserving. Non-rigid and adaptive regularization methods [14] often outperform the traditional approaches in cases when the image pair contain non-linear and elastic deformations.

Some solutions found in the current literature try to perform the registration with a multi-modal approach [15] [16], with the help of which the samples with different colors can be registered to each other, but most of them use rigid registration and transformation.

We propose a novel approach, based on these traditional, deep-learning based and non-rigid solutions, enhanced with iterative and adaptive enhancements specialized for the registration of creased and torn tissue sections. The proposed method can also be useful in other disciplines, where it is needed to register samples that contain missing image parts to each other and their correct registration can only be solved by applying elastic transformation.

2 Materials and Methods

There are methods that perform registration on images, but it is exceedingly difficult to find a correspondence based on macrostructures in one step, which also gives satisfactory results in terms of microstructures. One reason for this is the tissue destruction [17] that occurs during sectioning, which can result in some regions of the sample being distorted, and another is due to the difference in staining [18], each reagent amplifies the tissue structures visually and differently.

Since there are many tissue samples that occur in real life that contain a large amount of tissue deficiency, according to our experience, homography learning approaches are not or only limitedly applicable, since these solutions can work well primarily when all the relevant details of the images are similar to each other. When there are too many torn and missing parts, they give an unsuccessful registration result. For this reason, the proposed approach relies more on the classical three-step approach, which is based on feature detection, correspondence and homography estimation, supplemented by a grid-based iterative method that ensures the flexible and non-rigid transformation (Figure 2).

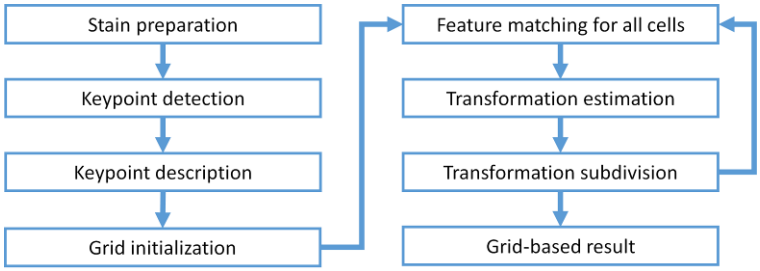


Figure 2
Algorithm overview of proposed method

2.1 Preparation Method

Since serial sections are often made to stain with different reagents, it is necessary to use a staining-independent preparation method to register them with each other. The proposed method uses an operation that tries to find visually similar-looking components between staining and then rely on them when registering. One of the most common such components is the nuclear-type structures [19] [20], which are present in both IHC [21] [22] (Immunohistochemistry) staining, and FISH [23] [24] (Fluorescence in situ hybridization) staining.

The proposed method uses a combination of two solutions for preparation of brightfield slides that have been extensively researched in other applications: staining unmixing [25-27] and staining normalization [28-31]. Staining unmixing or color deconvolution is a method used in brightfield microscopy to transform color images of multiple stained biological samples into images representing the stain concentrations. Staining normalization methods are designed to compensate the differences in intensity, saturation, and hue between samples, using a template image as a target image. These methods utilize color and spatial information to classify the image pixels into different stain components to reduce the effect of the variations of color and intensities which are caused by the sample preparation. Because the digitized image may contain significant amounts of camera noise, which may impair the efficiency of the recording, noise filtering, such as a median filter, may be required for some samples [32]. The chromatic and density distributions for the stain components in the hue-saturation-intensity color model are normalized to match with the distributions of a template image. In the case of a fluorescent sample, the proposed method uses the nuclei channel.

2.2 Keypoint Detectors and Descriptors

One of the basic ideas is to perform the registration in an iterative way in several steps: first detect an approximate transformation based on the larger

macrostructures, then further refine the registration considering the microstructures, trying to improve the result of the previous iteration in each iteration until it is possible.

The other basic idea is not to determine a single transformation value for the sample, but to assign different transformations to various parts of the sample. In this way, not only a rigid transformation with 6 or 8 degrees of freedom is performed, but an elastic deformation is used.

One of the most widely used traditional keypoint detector and descriptor algorithm is the scale-invariant feature transform (SIFT) [7] method. Although it is almost two decades old, has proven remarkably successful in several applications using visual features, including image stitching, object recognition, or stereo vision. However, the usage of it means a large computational requirement, which can be critical for systems which are real-time or has limited computational power.

This limitation was the cause of intense research in the direction of replacing it with a computationally more favorable alternative. One of the best alternatives are the Speeded Up Robust Features (SURF) [8], Features from Accelerated Segment Test (FAST) [33], Binary Robust Independent Elementary Features (BRIEF) [34] and Oriented FAST and Rotated BRIEF (ORB) [35] methods.

The ORB method has similar matching performance (Table 1) as the SIFT and SURF method but computationally it is more efficient. It utilizes the FAST method as keypoint detector and uses the BRIEF descriptor for feature description. These methods have superior performance and low computational cost. The ORB method adds a fast and accurate orientation component to the FAST method, optimizes the BRIEF feature computation, analyzes the variance and correlation of oriented BRIEF features, and to improve the performance of nearest-neighbor applications it has a learning method for de-correlating BRIEF features under rotational invariance.

Regarding the feature extraction and description, SuperPoint is one of the most state-of-the-art methods, which offers a fully convolutional model operates on full-sized images and computes pixel-level feature point positions together with their descriptions in one pass. In this study, we examine both the classical ORB method and the more modern SuperPoint approach as the basis of our method.

Table 1
Performance of keypoint detectors and descriptors

Method	Matching Performance	Computational Performance
SIFT	●●●○○	●●○○○
SURF	●●●○○	●●●○○
ORB	●●●●○	●●●●●
SuperPoint	●●●●●	●●●●○

2.3 Feature Matching and Transformation Estimation

After we detected features on the image pair and calculated descriptions for all of them, it is needed to match them. One of the simplest feature matching algorithms is the brute-force method [36], which takes descriptor of a feature on the first image and matches with all other features on the second image using distance calculation. It returns the closest pairs. When we matched the feature points on the image pair with each other, it is needed to estimate an optimal affine transformation (T) for them.

The widely used random sample consensus (RANSAC) [9] [37] method is a simple but effective iterative algorithm; it can estimate parameters of a mathematical model from a data set that contains outliers, when outliers are to be accorded no influence on the values of the estimates. Therefore, it also can be interpreted as an outlier detection method. In our case, we choose three keypoint matches at random and solve the T affine transformation as a system of equations. We count the number of matches which are inliers according to T transformation and distance limit. Repeat these steps for N rounds and return the T transformation which provided the highest inlier count.

2.4 Simple Grid-based Registration

The computed transformation matrix (T) can provide a good registration estimation for image pairs that include strictly only translational, rotational, and scaling differences. However, if the samples contain other types of deformations, the result may be inaccurate, or the result of RANSAC method may fail because choosing any of the three points will result too many outliers.

The proposed method utilizes grid-based registration, where we divide the image into smaller regions (patches), register these patches on their own, and then apply the separate transformations to the whole image like a grid (Algorithm 1).

Algorithm 1: Simple grid-based registration

Input: *matches, patches*

Output: *G* grid with affine transformation matrices

1. *G* = *patches* × *patches* empty grid
2. Forall *i* in 0..*patches*:
   2.1. Forall *j* in 0..*patches*:
      2.1.1. *roi* = Calculate ROI(*i, j, patches*)
      2.1.2. *matches'* = Collect matches in ROI(*roi*)
      2.1.3. *G*[*i, j*] = Estimate transformation(*matches'*)
3. Return *G*

2.5 Iterative Registration

One of the weaknesses of simple grid-based registration is that it can manage the non-linearity characteristics of the sample, but if we choose too large a grid size, there will be too little information available in a cell to perform the registration. In this case, we can only transform the cells by considering the local environment, so there may be many cells that you may not match properly or at all. If we choose a grid size that is too small, we will lose the ability to register nonlinear. In these cases, an iterative approach can help: the sample is first registered without a grid, in a rigid manner, considering all key points and the whole image (Figure 3).

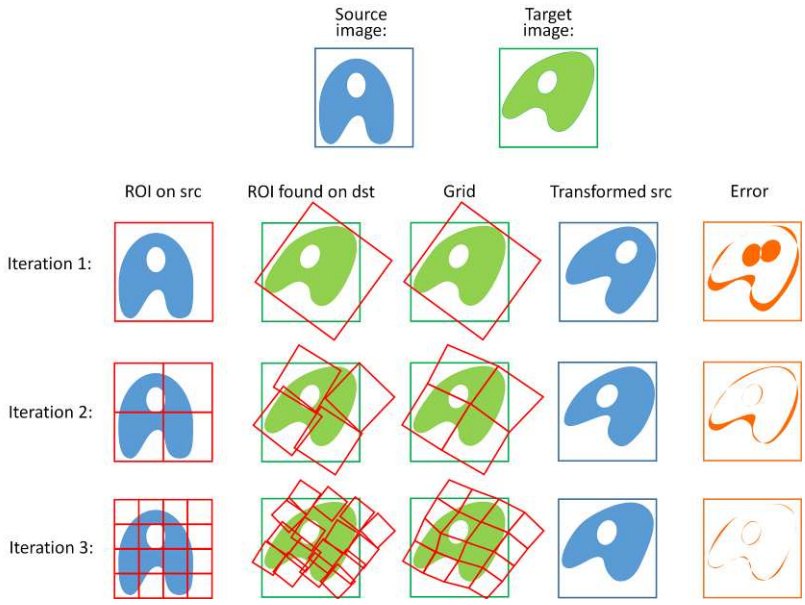


Figure 3
Illustration of iterative grid-based method

We assume that we can calculate globally the best possible rigid affine transformation. In the next iteration, the plane is divided into four equal parts, and the refined transformation matrices are calculated based only on the key points there but using the result of the previous iteration: only key pairs are considered whose transformed distance is less than a certain threshold. In the second iteration, we redistribute the previous four subdivided regions, so we will have a total of 16 regions, and we will also reduce the threshold used for the distance of the key points (Algorithm 2).

Algorithm 2: Recursive registration sub-method

Input: *matches*, *region, depth*

Output: *G* grid with affine transformation matrices

4. Filter the keypoints in the cell by their positions (*region*)
5. Filter the keypoints in the cell by their distances
6. T = Estimate transformations
7. If T is not valid:
   7.1. Return *Unit Matrix*
8. Decrease the distance limit
9. Apply *T* for *matches*
10. *G* = 2 × 2 empty subgrid
11. Forall *i, j*, *region'* in subdivide(*region*):
    11.1. *G*[*i, j*] = Recursive call with *depth'* and *depth+1*
12. Return *G*

Utilizing the recursive sub-method, we can calculate the elastic, non-rigid, grid-based registration (Algorithm 3).

Algorithm 3: Iterative grid-based registration method

Input: *img1*, *img2, depth*

Output: *G* grid with affine transformation matrices

1. *kp1* and *kp2* = Keypoint detection on *img1* and *img2*
2. *km* = Keypoint matching on *kp1* and *kp2*
3. G = Call the first iterative step

Return *G*

3 Validation

Routine anonymized slides were used from the archive of the 1st Department of Pathology and Experimental Cancer Research of the Semmelweis University, Budapest, Hungary. Digital slides in digital pathology are primarily identified by two key data: the inscription/barcode/QR code on the label area of the glass slide

and the name of the slide, often serving as a unique identifier. Using either of these, the patient can be traced back using the hospital information system. In case of research samples, both of these data are always removed right after (or before) scanning the samples. The samples providing the basis of this study were also de-identified and did not contain any details of the patient. The samples were used for a retrospective study without impacting patient care.

The study was conducted in accordance with the Declaration of Helsinki, approved by the Ethics Committee of the Institutional Review Board and Regional Ethics Committee of Semmelweis University (permit no. 14383-2/2017/EKU Semmelweis University, Budapest, Hungary).

Among the available samples, we collected cases in which there were at least 3 serial sections. It was important that these serial sections were not too far from each other in terms of distance, because that would make registration impossible (they cannot even be reconciled by humans in this case). Another aspect was that they should be stained with different reagents and that they should also have fluorescent sections.

A total of 268 slide pairs (53 slides from 12 cases) were selected for the validation and for each slide it was extracted only the scanned area in a size of 4000×4000 pixels (16 megapixels). This image size corresponds to a resolution of 4.185 µm/pixel on average. Images of this size clearly show larger and smaller tissue structures, but this resolution is too small to see cell compartments. The slides in the validation set were stained with 29 types of reagents.

Experts have placed five keypoints on each slide, in the middle of relevant regions or a well-defined part of slides. In the images belonging to the same cases, the same regions were marked. These regions were defined in such a way that both colleagues with biological and with bioengineering experience working in the field could mark them. Such areas were, for example, the center of a tumor area, the center of glands or border of a larger tissue region. The marking was performed at native resolution, and compared to the accuracy of the algorithm, the accuracy of the marking was an order of magnitude higher, so the consistency of the ground truth marking was not examined separately. Altogether it means 268×5 =1340 keypoint-pairs, they were used as a ground through during the validation.

Since the registration of slide pairs in the sample set is strongly influenced by whether they are pairs with the same staining and how they are prepared, we divided the image pairs into four categories (Figure 4):

a) Same staining and well prepared: Slide-pairs prepared with the same reagent, and they do not contain major creasing or tearing. It is usually easiest to register these samples together.

b) Same staining and creased/torn: The stainings are the same but the slide-pairs contain major tearing or creasing. With these samples, the need for an iterative approach can emerge.

c) Different staining and well prepared: The samples prepared correctly but the stainings are different.

d) Different staining and creased/torn: Samples with different stainings and contain major creases or tears. In general, these pairs are the most difficult to register with each other.

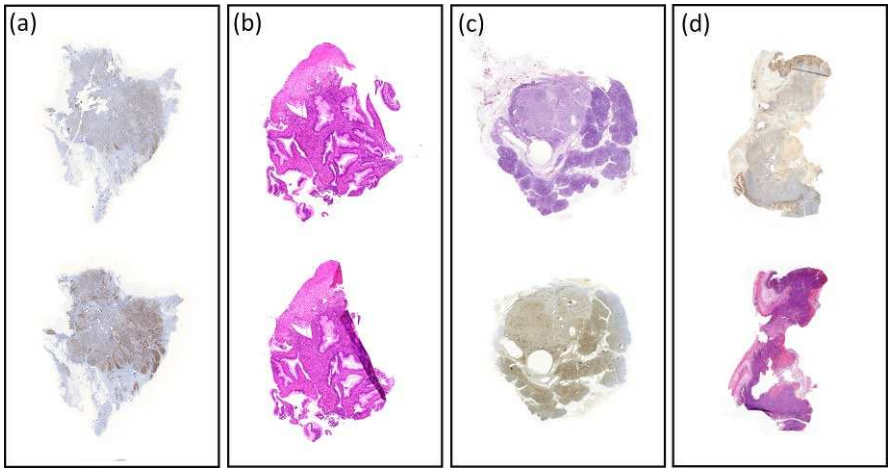


Figure 4

Illustration of the four sample categories: (a) same staining and well prepared; (b) same staining and creased/torn; (c) different staining and well prepared; (d) different staining and creased/torn

During validation, we examined the ideal resolution value and how the quality of the registration improves during the iterations. For scaling values, 1:1 resolution means the resolution of the input image (4.185 µm/pixel), 1:2 means half resolution (8.37 µm/pixel), until the 1:20 resolution (83.7 µm/pixel). Patch sizes 1, 2, 4, 8, 16, and 32 were examined for number of iterations. For optimal iteration number, 1, 2, 3, 4, and 5 iterations were examined, corresponding to 1×1, 2×2, 4×4, 8×8, 16×16, and 32×32 patches.

By validating the proposed method with each parameter, we compared the position of the keypoints marked by the experts with the positions transformed by the registration algorithm. From the distances, an error was calculated for each slide-pair using root mean squared error *(RMSE(c))* method, and then an average was calculated from them *(ARMSE)*.

Once the mean error *(ARMSE)* has been calculated for each number of scaling and iteration, we can determine their optimal values for our sample set. Both the mean error and standard deviation values are calculated back to µm/pixel.

Knowing the ideal parameters for the sample set, the proposed method is compared using the classic ORB keypoint detector and descriptor, the more modern SuperPoint algorithm and executing the SuperGlue approach.

4 Results

4.1 Optimal Number of Patches and Scale Factors

Table 2 shows the average root mean squared errors *(ARMSE)* of the registration using different patch numbers and scale factors. The lower value means the more precise registration. We can see that minimal value (137.12 µm) is reached using the parameters patches = 16 and scale = 1:6. Without iterative method (patches = 1) the error numbers are everywhere worse than with utilizing iterations. The average error *(ARMSE)* is the highest at 1:1 magnification without iterations (284.35 µm).

Table 2

Heatmap colorized table about the average root mean squared errors (ARMSE) using different scale factors and patch sizes. Lower and greener value means the more accurate registration. Values are in µm.

		Patches						
		1	2	4	8	16	32	Avg.:
1	1:1	284.35	244.05	237.55	233.11	233.34	235.34	**236.68**
2	1:2	233.75	195.87	187.04	184.26	181.38	169.59	**183.63**
3	1:3	215.32	176.27	165.27	163.02	162.99	154.67	**164.44**
4	1:4	196.24	159.99	150.11	148.05	147.49	142.09	**149.55**
6	1:6	179.83	149.22	139.19	138.70	137.12	147.50	**142.35**
8	1:8	179.47	152.84	141.25	141.07	139.81	147.89	**144.57**
10	1:10	177.67	151.22	144.31	143.05	141.41	149.37	**145.87**
12	1:12	180.46	157.09	145.39	144.66	146.65	144.89	**147.74**
14	1:14	179.80	152.63	143.38	140.37	140.19	156.43	**146.60**
20	1:20	182.92	161.73	152.43	153.24	154.12	153.68	**155.04**
	Avg.:	**200.98**	**170.09**	**160.59**	**158.95**	**158.45**	**160.14**	

Scale

As can be seen, the proposed method provides the lowest error value using patch-size 16×16 and scale factor 1:6. It corresponds to 5 iterations and 25.11 µm/pixel resolution.

A distance of 137.12 µm in terms of the error value means that there is an average difference of 137.12 µm between the ground truth manual alignment and the algorithmic result. This is a distance of approximately 10-15 nuclei. At the level of the cell compartment, it is therefore not possible to register the different sections using the method, but this is not the purpose of the registration; since there is a depth distance of 10-50 µm between the serial sections anyway, they will not have common nuclei anyway. In terms of macrostructures, however, it may be appropriate; their size is above this error value (e.g., glands), so they can be matched and examined together.

4.2 Optimal Number of Keypoints

Since the parameters scale = 1:6 and patches = 16 parameters proved to be the best, we examined the ideal keypoints number (Figure 5) with these parameters. The lowest number of keypoints examined was 200, while approaching a keypoint number of 10,000, the algorithm achieved the lowest error value: 137.12 µm. Using a higher keypoint count, the efficiency deteriorated again. There was no notable improvement in error value above 50,000 keypoints, suggesting that the sample contains an average of this number of usable key points.

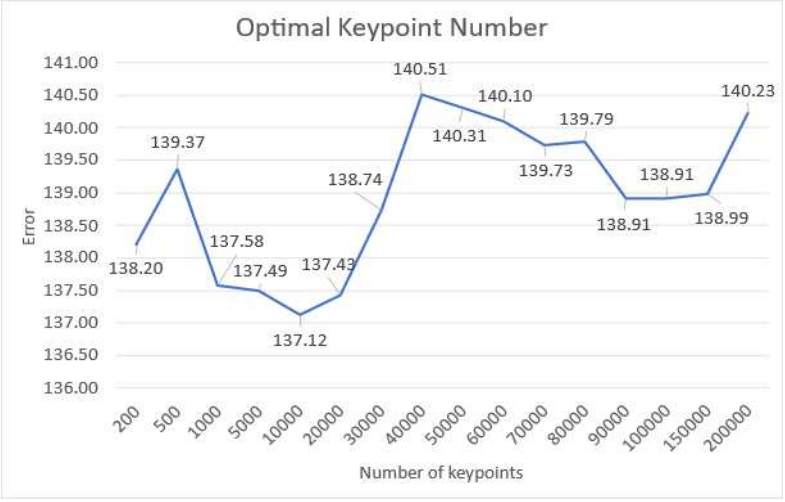


Figure 5
Average root mean squared error values *(ARMSE)* of using different number of keypoints. Values are in µm.

4.3 Average Error Values

Figure 6 shows the comparison of the average error values in each sample category applying the ORB based keypoint detection and description, the SuperGlue method, and the proposed method.

Analyzing the average error values, it can be concluded that the SuperGlue solution proves to be more effective in the case of well-prepared samples. However, in the case of samples that are not well prepared, i.e., contain creases or tears, the iterative nature of the proposed method can provide better results. The classic ORB key point detector always falls short of the results of the proposed method, it can only give similar results in the case of samples with same staining and well prepared.

Figure 6
Average root mean squared error values *(ARMSE)* applying different methods on sample categories. Values are in µm.

Looking at the results, it can be said that the proposed solution can bring better results mostly in the case of poorly prepared samples, under ideal conditions the SuperGlue solution may be sufficient.

4.4 Results Illustrations

To verify the robustness of image registration methods we compared the proposed method with the ORB keypoint detection using RANSAC matching algorithm, and with the SuperGlue registration method. Since the efficiency of the registration methods depends greatly on the quality of the sample preparation and staining, we separately analyzed the algorithms for identically stained and tear/crease-free samples and for differently stained and poorly prepared image-pairs. For demonstration purposes, the algorithm was also executed on high-resolution image pairs cut from whole slides.

Figure 7 shows the matching results in the case of five well-prepared and identically stained sections. It can be observed that usually the ORB+RANSAC based method provides 10-20 percent worse results than the proposed and SuperGlue approaches. Since these samples do not contain major tears or creases and visually look the same, the iterative nature of our method could not help, so the proposed method and SuperGlue approach provide similar results.

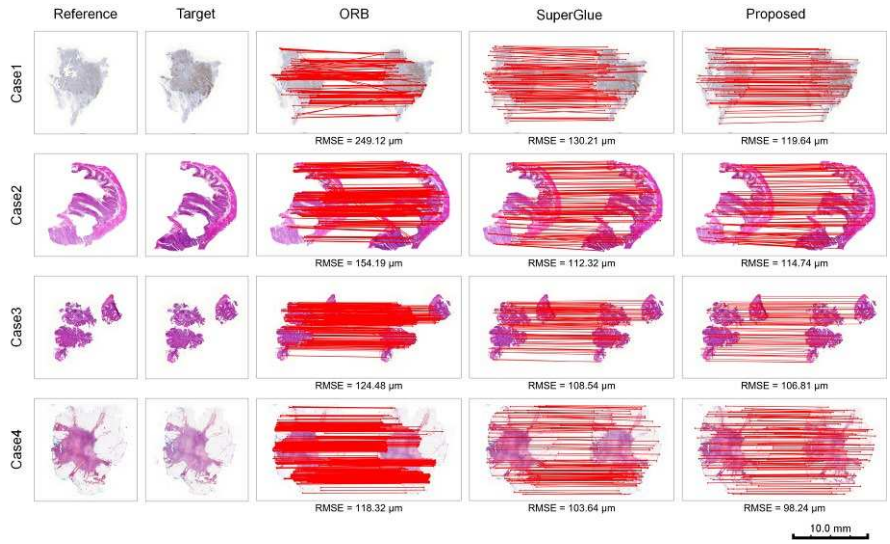


Figure 7
Matching results of five **well-prepared** and **identically stained** samples using the ORB+RANSAC, SuperGlue and proposed approaches

Figure 8 shows five examples, where the methods are executed on smaller parts of a whole slide image. The magnifications of these images are ten times larger than in Figure 7. These samples are also well-prepared and have the same staining. It can be observed that all three models provide similar results, this is due to the fact that the reference and target images are visually very similar.

In Figure 9, we can see the real differences between the analyzed models. In this picture we can see five samples, which are poorly prepared, containing tears and creases, and they are stained using different reagents. Since the image-pairs contain different tissue structures or have different colorizations, the classical ORB based method finds common feature points extremely hard and it matches them incorrectly. Compared to the proposed approach, the ORB method can achieve two-three times worse results on these samples than the proposed one. The iterative nature of the proposed approach can help these samples to a great extent, as in the first iteration it tries to find and match features on a low-resolution image that can be found on both images regardless of staining and tissue structure, and then this found similarity is further refined by further iterations. Using this approach, for image pairs that are difficult to register, the proposed approach mostly achieved better results than the SuperGlue solution.

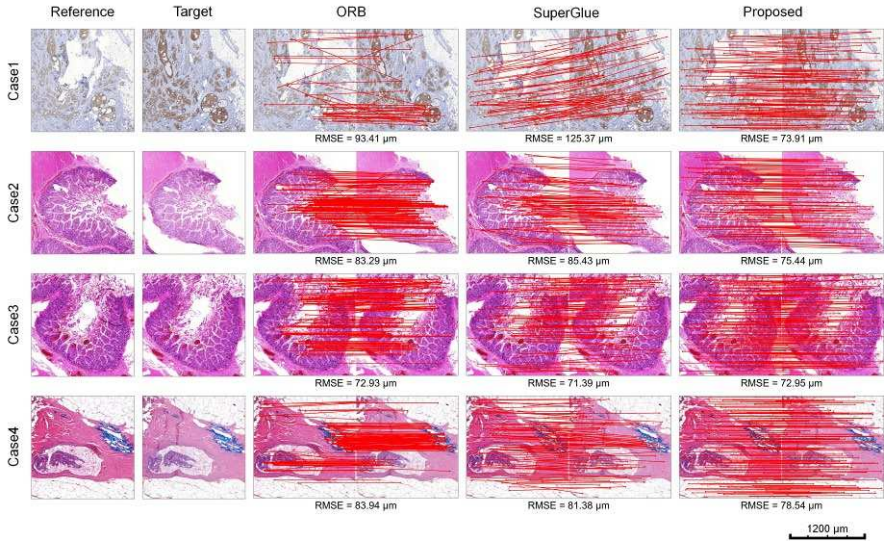


Figure 8
Matching results of five **well-prepared** and **identically stained high-resolution** samples using the ORB+RANSAC, SuperGlue and proposed approaches

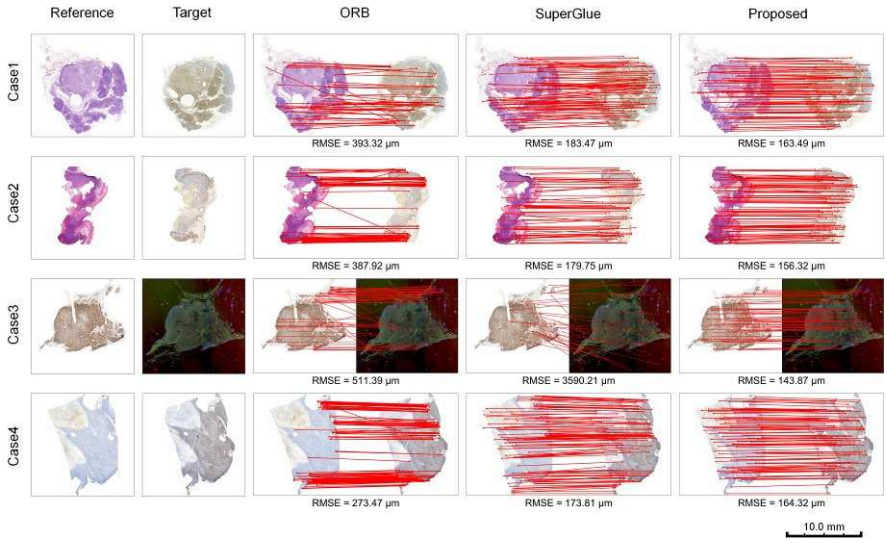


Figure 9
Matching results of five **differently stained** or **poorly prepared** samples using the ORB+RANSAC, SuperGlue and proposed approaches

From the illustration of results in Figure 10, it can be observed that the registration of high-resolution whole-slide parts are often unsuccessful, while the proposed method mostly can provide acceptable results. The reason for this is that ORB and SuperGlue methods do not have prior information about the approximate orientation of image parts and its key points, considering the whole-slide. In contrast to that the proposed method can know the rough position from the previous iterations, and it can consider the positions during the matching of keypoints.

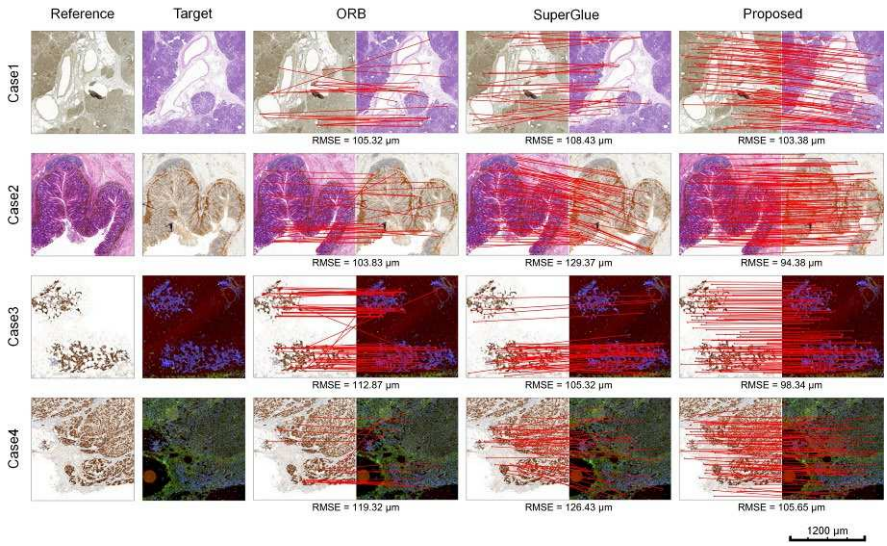


Figure 10

Matching results of five **differently stained** or **poorly prepared high-resolution** samples using the ORB+RANSAC, SuperGlue and proposed approaches

5 Discussion

The purpose of this study was to investigate how different registration methods perform in the task of aligning tissue samples onto each other in a whole slide imaging system and what parameters are used to achieve the best results. The proposed method was examined based on three main aspects and compared with competing solutions.

The first parameter examined was the value of the optimal patch number and scale factor, which was characterized by the mean squared error (ARMSE) metric, which gave the average accuracy of the registration expressed in microns. Based on our validation data set, the proposed solution achieved the best result using

16×16 patches (5 iterations) and a scale-factor value of 1:6, which resulted an error value of 137.12 μm.

The second examined parameter was the optimal number of keypoints, which was determined using the previously established patch and scale factor values. We analyzed key-points between 200 and 200,000, and the best performing key-point number was 10,000, where the proposed method achieved the best error value of 137.12 μm. It performed worse on fewer and more key-points than this.

The third evaluation method was a comparison with competitive solutions, where we examined the performance of the algorithm separately for the four types of samples, in comparison with the ORB-based and SuperGlue solutions. The results showed that under good conditions (with well-prepared samples) the SuperGlue solution gave the best results, however, for various types of stains and in the overall comparison, the proposed solution outperformed the others. The ORB-based solution fell short of the other two solutions in all categories. Overall, the recommended approach performs best in situations where an iterative approach is needed.

Registration is essential for making diagnosis or conducting research in serial sections. Because we used ground truth keypoints for validation, which experts laid down by marking the middle of the relevant regions, examining the results, we found that the appropriate scaling factor and grid-based iterative approach helped to perform the correct registration, thus helping the doctoral work.

However, it is important to mention that the parameters are strongly sample dependent. There are tissue samples that contain only a small amount of nonlinear distortion, either due to the thickness of the sample or due to better or more automated sample preparation. In this case, using a higher number of iterations is unnecessary. It is also important that some samples do not contain enough visual information to use the iterative method, and if you choose a patch number that is too high, there is a chance that they will be falsely registered, so the sample may be incorrectly deformed in those areas.

Analyzing the results of the validation set, several ideas have been put forward to improve the proposed methods. One such enhancement may be when the sample does not visually carry enough information or when completely different structural elements are painted between stains. Recognition of different tissue structures is a well-studied research field in the literature, with many approaches and methods already available from traditional pattern recognition algorithms [38] to sophisticated, deep learning based convolutional neural network methods [39-41]. By applying such methods on samples, we can detect features in the images that were not present locally at the pixel level and amplify similar areas of the different stained samples. By performing the proposed registration method on images prepared in this way, you can get better results.

Another improvement method is to try to use different magnifications between iterations. Initial iterations are often best recorded on low-resolution images, relying on macrostructures. However, this level of magnification no longer carries enough information in later iterations to calculate a more accurate transformation. To overcome this, we can try to use an increasingly high-resolution image as the basis between the iterations. A similar result can be obtained by using a key-point metric that carries information about a multi-resolution image, i.e., a multiscale feature.

Conclusions

The use of registration methods is critical for whole slide imaging (WSI), in the case of serial sections. In this study we collected a set of digital slides which required registration and after the implementation of registration method, we completed a validation study, to determine which model and which parameters provide the best results. According to the characteristics of the collected validation set, executing the methods, the following conclusions were drawn:

(1) For the registration of digital microscopic samples, the SuperPoint method is excellent for keypoint detection and description, supplemented with RANSAC-based keypoint pairing and affine transformation calculation. However, since not all parts of the sample are deformed in the same way during sample preparation, the transformation of the whole sample is not linear. In this case, the grid-based approach can be used to compute transformations for different regions of the sample separately.

(2) By choosing a high patch size, we can manage non-linearity, but there will be many patches that contain only microstructures, and without knowledge of the environment with macrostructures, we cannot accurately transform them into each other. The iterative approach can help the problem by initially performing the initial registration by choosing a low patch size and then further refining the regions by recursively redistributing them. The iterative approach performs non-rigid registration with more confidence and a lower error rate.

Acknowledgement

The authors would like to thank the expert colleagues who collected the samples representing the validation dataset and marked the ground-truth dataset: T.M. (pathologist, Semmelweis University) and G.K. (biologist, 3DHISTECH Ltd.).

This research was supported by the ÚNKP-22-3 New National Excellence Program of the Ministry for Culture and Innovation from the Source of the National Research, Development and Innovation Fund. The study received external funding from the national project 2019-1-3-1-KK-2019-00007.

References

[1] Krenacs, T.; Zsakovics, I.; Micsik, T.; Fonyad, L.; Varga, S. V.; Ficsor, L.; Kiszler, G.; Molnar, B. Digital microscopy – the upcoming revolution in histopathology teaching, diagnostics, research and quality assurance. *Microscopy: Science, Technology, Applications and Education Book Series,* **2010**, *Number 4*, 965-977

[2] Paulik, R.; Kozlovszky, M.; Molnár, B. Regression Based Iterative Illumination Compensation Method for Multi-Focal Whole Slide Imaging System. *Sensors,* **2021**, *21*, 7085, doi:10.3390/s21217085

[3] Song, H.; Xiao, H.; He, W.; Wen, F.; Yuan, K. A fast stereovision measurement algorithm based on SIFT keypoints for mobile robot. *2013 IEEE International Conference on Mechatronics and Automation,* **2013**, 1743-1748, doi:10.1109/ICMA.2013.6618179

[4] Lowe, D. G. Object recognition from local scale-invariant features. *Proceedings of the IEEE International Conference on Computer Vision,* **1999**, *2*, 1150-1157, doi:10.1109/ICCV.1999.790410

[5] Zhang, W.; Li, X.; Yu, J.; Kumar, M.; Mao, Y. Remote sensing image mosaic technology based on SURF algorithm in agriculture. *EURASIP Journal on Image and Video Processing* **2018**, *2018*, 85, doi:10.1186/s13640-018-0323-5

[6] Boulekchour, M.; Aouf, N.; Richardson, M. Robust L∞ convex optimisation for monocular visual odometry trajectory estimation. *Robotica,* **2014**, 1-20, doi:10.1017/S0263574714001829

[7] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision,* **2004**, 91-110, doi:10.1023/B:VISI.0000029664.99615.94

[8] Bay, H.; Tuytelaars, T.; Gool, L. Surf: Speeded up ro-bust features. *European Conference on Computer Vision,* **2006**, doi:10.1007/11744023_32

[9] Fischler, M. A.; Bolles, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM,* **1981**, *24*, 381-395, doi:10.1145/358669.358692

[10] Dosovitskiy, A.; Fischer, P.; Springenberg, J.; Riedmiller, M.; Brox, T. Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence,* **2016**, *38*, 1734-1747, doi:10.1109/TPAMI.2015.2496141

[11] DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-Supervised Interest Point Detection and Description. *Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, **2018**, 224-236, doi:10.1109/CVPRW.2018.00060

[12] Sarlin, P.-E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching with Graph Neural Networks. *arXiv,* **2019**, doi:10.1109/CVPR42600.2020.00499

[13] DeTone, D.; Malisiewicz, T.; Rabinovich, A. Deep image homography estimation. *arXiv,* **2016**, doi:10.48550/arXiv.1606.03798

[14] Myronenko, A.; Song, X. Adaptive Regularization of Ill-Posed Problems: Application to Non-rigid Image Registration. *Comput. Res. Repos,* **2009**, *6*, doi:10.48550/arXiv.0906.3323

[15] Hoque, M. Z.; Keskinarkaus, A.; Nyberg, P.; Mattila, T.; Seppänen, T. Whole slide image registration via multi-stained feature matching. *Computers in Biology and Medicine,* **2022**, *144*, 105301, doi:10.1016/j.compbiomed.2022.105301

[16] Cooper, L.; Sertel, O.; Kong, J.; Lozanski, G.; Huang, K.; Gurcan, M. Feature-based registration of histopathology images with different stains: an application for computerized follicular lymphoma prognosis. *Comput Methods Programs Biomed,* **2009**, *96*, 182-192, doi:10.1016/j.cmpb.2009.04.012

[17] Taqi, S. A.; Sami, S. A.; Sami, L. B.; Zaki, S. A. A review of artifacts in histopathology. *J Oral Maxillofac Pathol,* **2018**, 22(2):279., doi:10.4103/jomfp.JOMFP_125_15

[18] Meyerholz, D. K.; Beck, A. P. Principles and approaches for reproducible scoring of tissue stains in research. *Laboratory Investigation,* **2018**, 98(7):844-855, doi:10.1038/s41374-018-0057-0

[19] Szenasi, S.; Vamossy, Z. Implementation of a Distributed Genetic Algorithm for Parameter Optimization in a Cell Nuclei Detection Project. *Acta Polytechnica Hungarica,* **2013**, *10*, 59-86, doi:10.12700/APH.10.04.2013.4.4

[20] Szenasi, S.; Vamossy, Z. Evolutionary Algorithm for Optimizing Parameters of GPGPU-based Image Segmentation. *Acta Polytechnica Hungarica,* **2013**, *10*, 7-28, doi:10.12700/APH.10.05.2013.5.2

[21] Gobbo, F.; Sarli, G.; Silva, M.; Galiazzo, G.; Chiocchetti, R.; Morini, M. A Double Histochemical/Immunohistochemical Staining for the Identification of Canine Mast Cells in Light Microscopy. *Vet. Sci,* **2021**, *8*, 229, doi:10.3390/vetsci8100229

[22] Smith, N. R.; Womack, C. A matrix approach to guide IHC-based tissue biomarker development in oncology drug discovery. *J Pathol,* **2014**, *Jan;232(2):190-8*, doi:10.1002/path.4262

[23] Frickmann, H.; Zautner, A. E.; Moter, A.; Kikhney, J.; Hagen, R. M.; Stender, H.; Poppert, S. Fluorescence in situ hybridization (FISH) in the microbiological diagnostic routine laboratory: A review. *Crit. Rev. Microbiol,* **2017**, *43*, 263-293, doi:10.3109/1040841X.2016.1169990

[24] Kuo, J.-T.; Chang, L.-L.; Yen, C.-Y.; Tsai, T.-H.; Chang, Y.-C.; Huang, Y.-T.; Chung, Y.-C. Development of Fluorescence In Situ Hybridization as a Rapid, Accurate Method for Detecting Coliforms in Water Samples. *Biosensors,* **2021**, *11*, 8, doi:10.3390/bios11010008

[25] Bianconi, F.; Kather, J. N.; Reyes-Aldasoro, C. C. Experimental Assessment of Color Deconvolution and Color Normalization for Automated Classification of Histology Images Stained with Hematoxylin and Eosin. *Cancers,* **2020**, *12*, 3337, doi:10.3390/cancers12113337

[26] Khan, A.; Rajpoot, N.; Treanor, D.; Magee, D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans. Biomed. Eng,* **2014**, *61*, 1729-1738, doi:10.1109/TBME.2014.2303294

[27] Ruifrok, A.; Johnston, D. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol,* **2001**, *23*, 291-299, pmid:11531144

[28] Bejnordi, E.; Babak; Litjens, G.; Timofeeva, N.; Otte-Höller, I.; Homeyer, A.; Karssemeijer, N.; Laak; Jeroen Stain Specific Standardization of Whole-Slide Histopathological Images. *IEEE transactions on medical imaging,* **2015**, *35*, doi:10.1109/TMI.2015.2476509

[29] Zheng, Y. Stain Standardization Capsule for Application-Driven Histopathological Image Normalization. *IEEE Journal of Biomedical and Health Informatics,* **2021**, *25*, 337-347, doi:10.1109/JBHI.2020.2983206

[30] Anghel, A.; Stanisavljevic, M.; Andani, S. A High-Performance System for Robust Stain Normalization of Whole-Slide Images in Histopathology. *Front Med Lausanne,* **2019**, *6*, doi:10.3389/fmed.2019.00193

[31] Balakrishnan, L.; Anand, S.; Jenitha, T. Stain Removal Through Color Normalization of Haematoxylin and Eosin Images: A Review. *Journal of Physics: Conference Series,* **2019**, *12108*, doi:10.1088/1742-6596/1362/1/012108

[32] Szántó, P.; Fehér, B. Hierarchical Histogram-based Median Filter for GPUs. *Acta Polytechnica Hungarica,* **2018**, *15*, 49-68, doi:10.12700/APH.15.1.2018.2.3

[33] Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. *European Conference on Com-puter Vision,* **2006**, *1*, doi:10.1007/11744023_34

[34] Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. Brief: Binary robust independent elementary features. *In EuropeanConference on Computer Vision,* **2010**, doi:10.1007/978-3-642-15561-1_56

[35] Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: an efficient alternative to SIFT or SURF. *Proceedings of the IEEE International Conference on Computer Vision,* **2011**, 2564-2571, doi:10.1109/ICCV.2011.6126544

[36] Jakubović, A.; Velagić, J. Image Feature Matching and Object Detection Using Brute-Force Matchers. *2018 International Symposium ELMAR,* **2018**, 83-86, doi:10.23919/ELMAR.2018.8534641

[37] Mac, T.; Lin, C.-Y.; Huan, N.; Luong, D.-N.; Hoang, P.; Hai, H. Hybrid SLAM-based Exploration of a Mobile Robot for 3D Scenario Reconstruction and Autonomous Navigation. *Acta Polytechnica Hungarica,* **2021**, *18*, 197-212, doi:10.12700/APH.18.6.2021.6.11

[38] Burges, C. J. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov,* **1998**, *2*, 121-167, doi:10.1023/a:1009715923555

[39] Liu, X.; Song, L.; Liu, S.; Zhang, Y. A Review of Deep-Learning-Based Medical Image Segmentation Methods. *Sustainability* **2021**, *13*, 1224, doi:10.3390/su13031224

[40] Vu, Q. D.; Graham, S.; Kurc, T.; To, M. N. N.; Shaban, M.; Qaiser, T.; Koohbanani, N. A.; Khurram, S. A.; Kalpathy-Cramer, J.; Zhao, T. Methods for Segmentation and Classification of Digital Microscopy Tissue Images. *Front. Bioeng. Biotechnol,* **2019**, *7*, 53, doi:10.3389/fbioe.2019.00053

[41] Wetteland, R.; Engan, K.; Eftestøl, T.; Kvikstad, V.; Janssen, E. A. M. A Multiscale Approach for Whole-Slide Image Segmentation of five Tissue Classes in Urothelial Carcinoma Slides. *Technol. Cancer Res. Treat,* **2020**, 19, doi:10.1177/1533033820946787

Lung Ultrasound Imaging and Image Processing with Artificial Intelligence Methods for Bedside Diagnostic Examinations

Gábor Orosz[1,2,*], Róbert Zsolt Szabó[1,3*], Tamás Ungi[4], Colton Barr[4], Chris Yeung[4], Gábor Fichtinger[4], János Gál[2], and Tamás Haidegger[1,5]

[1]Antal Bejczy Center for Intelligent Robotics, Óbuda University, EKIK; Becsi ut 96/B, Budapest, 1034, Hungary, gabor.orosz@irob.uni-obuda.hu
[2]Semmelweis University, Department of Anesthesiology and Intensive Therapy; Ulloi ut 78, Budapest, 1083, Hungary, gal.janos@med.semmelweis-univ.hu
[3]John von Neumann Faculty of Informatics, Óbuda University, Budapest, Hungary, robert.szabo@irob.uni-obuda.hu
[4]Laboratory for Percutaneous Surgery, Queen's University; 2K7L 2N8, ON, Canada, {ungi,c.barr,chris.yeung,gabor}@queensu.ca
[5]Austrian Center for Medical Innovation and Technology; Viktor-Kalpan-str. 2., Wiener Neustadt, 2700, Austria, haidegger@acmit.at
[*]*These authors contributed equally*

Abstract: Artificial Intelligence-assisted radiology has shown to offer significant benefits in clinical care. Physicians often face challenges in identifying the underlying causes of acute respiratory failure. One method employed by experts is the utilization of bedside lung ultrasound, although it has a significant learning curve. In our study, we explore the potential of a Machine Learning-based automated decision-support system to assist inexperienced practitioners in interpreting lung ultrasound scans. This system incorporates medical ultrasound, advanced data processing techniques, and a neural network implementation to achieve its objective. The article provides a comprehensive overview of the steps involved in data preparation and the implementation of the neural network. The accuracy and error rate of the most effective model are presented, accompanied by illustrative examples of their predictions. Furthermore, the paper concludes with an evaluation of the results, identification of limitations, and recommendations for future enhancements.

Keywords: AI-based image processing; Surgical Data Science; Applied medical imaging; Deep neural networks; Lung ultrasound

1 Introduction

Over the past decade, lung ultrasound has emerged as a widely utilized diagnostic tool [1–6], particularly in bedside examinations [7, 8], owing to the advent

of portable and handheld devices, and even supporting robotic surgery applications [9–12]. Critical care [13], anesthesia [14], and emergency medicine [15] are among the primary medical specialties that have embraced this technique. Proficiency in recognizing lung ultrasound artifacts is essential for practitioners in these fields [16]. The critical care perspective played a pivotal role in demonstrating the utility of this technique to the medical community [17–19]. Unlike general ultrasonography, lung ultrasound examinations focus on identifying and analyzing patterns of artifacts [20]. Therefore, lung ultrasonography has become a distinct clinical modality within acute care specialties.

1.1 Sustainable Radiology

Ultrasound is also seen as an affordable, wide-spread diagnostic tool, subject to continuous innovation, able to bring sustainability to the imaging domain of modern medicine [21, 22].

Emergency ultrasonography has emerged as a widely researched application, supporting euqal access to care. This is aligned with the United Nations' Sustainable Development Goals (SDG) – especially with SDG 3 for promoting well-being across all age groups [23].

Notably, lung emergency ultrasound offers an eco-friendly and non-ionizing imaging option, particularly valuable in resource-limited regions and during pandemics like COVID-19.

In low- and middle-income countries (LMICs), diagnostic imaging is often inadequate [24], but clinician-performed, hand-carried, bedside ultrasound has gained popularity globally. Affordable, portable, and user-friendly machines have expanded its reach, bolstering diagnostic capabilities in rural hospitals.

The use of ultrasound in LMICs has gained recognition from health ministries, non-governmental organizations, and the World Health Organization (WHO). It significantly improves patient diagnosis and management, particularly in remote regions [25].

1.2 Automating Ultrasound Diagnostics

Our research aims to develop an automated ultrasound system, conserving resources and aiding in training and diagnostics [26]. This innovation holds the potential to transform healthcare delivery in resource-limited settings, empowering local practitioners and enhancing patient care. Emergency lung ultrasonography represents a powerful tool in achieving sustainable healthcare goals in underserved regions. Through our efforts to create an automated solution, we hope to strengthen healthcare infrastructure and reduce disparities, fostering improved health outcomes for all.

The utilization of ultrasound imaging in lung examinations presents distinct challenges. These difficulties arise from the presence of the rigid chest structure and the unsuitability of air as an ultrasound medium. As a result, the emergence of acoustic

shadows, a common occurrence due to the transducer's orientation perpendicular to the ribs, becomes noteworthy. Nevertheless, adept manipulation and positioning of the transducer can mitigate this artifact.

In the context of air-filled lungs, the ultrasound beam experiences complete reflection at the boundary between soft tissue and air or fluid. This interaction yields a tangible image of the chest's soft tissues and pleural layers, accompanied by distinctive artifact patterns. The crux of lung ultrasound examinations revolves around the detection and analysis of these artifacts. Among the array of artifacts, reverberation artifacts are prominent. In physiological conditions, they are referred to as A-lines, while in pathological contexts, they are predominantly known as B-lines. [14]. The nomenclature of these artifacts originates from the work of Lichtenstein et al. [19, 27], and consensus has been reached regarding their application [28, 29]. An atelectatic lung appears as a true tissue image, and the description of each significant pathological phenomenon is provided within its respective field of application.

1.3 Technology Against a Pandemic

During the COVID-19 pandemic, there was a rapid development and deployment of supporting technologies for diagnosis and treatment [30–33]. Lung ultrasound emerged as a reliable alternative to chest X-rays, exhibiting comparable accuracy to CT scans, which are considered the gold standard for lung imaging. According to a Cochrane review on thoracic imaging tests for diagnosing COVID-19, chest CT and lung ultrasound were found to be sensitive and moderately specific. Therefore, these modalities may be more valuable in ruling out COVID-19 than in differentiating SARS-CoV-2 infection from other respiratory illnesses [34].

1.4 Medical Background

Ultrasound stands as a preferred imaging modality in various scenarios owing to its array of advantages over other methods. Foremost, its cost-effectiveness renders it more accessible to patients. Furthermore, ultrasound furnishes real-time imaging capabilities, enabling instantaneous observation of anatomical structures and physiological dynamics. Distinct from alternative techniques like X-rays, ultrasound circumvents the use of ionizing radiation, thereby ensuring patient safety during recurrent examinations without the potential for adverse repercussions. [35].

The portability of ultrasound devices is another significant advantage, enabling their use at the bedside. This eliminates the need to transport patients within the hospital, reducing the risk of spreading infectious diseases and minimizing the challenges associated with handling critically ill and unstable patients [36]. In situations where other scanning methods may be impractical or unsafe, bedside lung ultrasound has emerged as a viable solution. Moreover, its diagnostic accuracy is comparable to or even superior to conventional radiographic measures, further validating its utility. However, a notable challenge with lung ultrasound lies in the interpretation of images, which demands experienced medical professionals. Diagnosis through ultrasound is inherently subjective and heavily reliant on the competency, experience, and mental state of the performing physicians. Factors like stress or exhaustion can

impact the accuracy of interpretations [20]. Given that lung ultrasound is frequently performed on critically ill patients, prompt and accurate diagnosis is of utmost importance. As not all physicians performing the examination possess expertise in interpreting lung ultrasound, the provision of a decision support system becomes crucial in expediting accurate diagnoses.

Moreover, despite the generally acknowledged safety and reproducibility of ultrasound assessments, the fundamental "ALARA" (As Low As Reasonably Achievable) principle remains pivotal in curtailing ultrasonography exposure. Hence, expediting the diagnostic process assumes significance from a safety standpoint. Regrettably, the absence of uniform standards hinders the seamless comparison of ultrasound investigations and their outcomes. Acknowledging this concern, authorities within the realm of lung ultrasound have advocated for the establishment of standardization guidelines encompassing machine configurations and scanned regions. [5, 31].

In addition to machine settings, the selection of the optimal scanning technique (such as longitudinal or transverse) remains a topic of ongoing discourse. This debate can give rise to potential misinterpretations or misdiagnoses in routine clinical settings. Such ambiguity may lead to uncertainty among medical personnel and elongate the duration of lung ultrasound assessments, ultimately extending patient contact time. It becomes imperative to address these challenges pertaining to standardization and scanning techniques, as doing so holds the key to enhancing the efficiency and dependability of lung ultrasound procedures. [37].

1.5 The BLUE protocol

The BLUE (Bedside Lung Ultrasound in Emergency) protocol, developed by Daniel Liechtenstein [19], is a clinical procedure commonly used in intensive care medicine to rapidly diagnose the underlying causes of acute respiratory failure (Fig. **??**). When administered by skilled practitioners, the BLUE protocol has demonstrated its efficacy and potency as a tool for assessing lung conditions in critically ill individuals. Furthermore, it is noteworthy to emphasize the exceptional adaptability of the BLUE protocol, which readily lends itself to a seamless evolution into the advanced semi-quantitative framework named BLUE-LUSS (Bedside Lung Ultrasound in Emergency-Lung UltraSound Score). This innovative concept, previously expounded upon by the authors in a prior publication, signifies a testament to the protocol's progressive potential.n [37]. This remarkable evolution empowers medical professionals to leverage the capabilities of numerical scoring, ushering in a paradigm shift towards enhanced precision and objectivity within the domain of diagnostic assessment.

Previously, the protocol relied on still frames of ultrasound images captured from standardized locations. However, recent advancements in the field have demonstrated the superiority of utilizing short ultrasound loops, typically ranging from 3 to 10 seconds in duration [31]. These dynamic ultrasound loops provide a more comprehensive view of the lung, allowing for a more accurate assessment of lung artifacts and abnormalities.

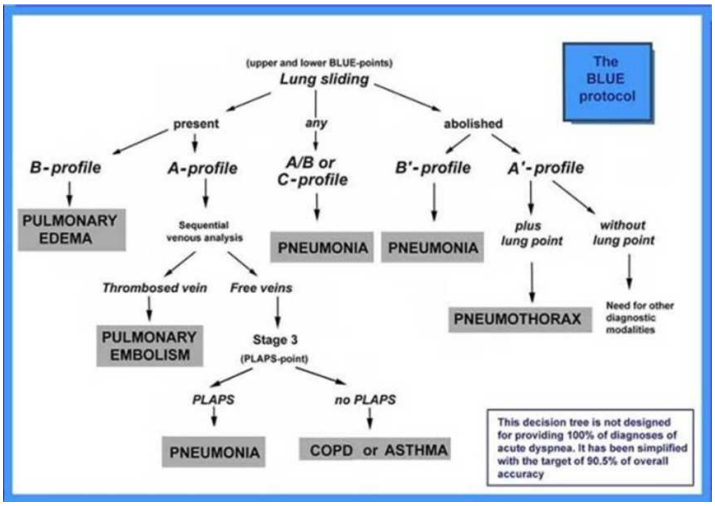


Figure 1
The core objective of the BLUE protocol revolves around streamlining the diagnostic timeline. This is achieved through the utilization of predefined points of analysis for ultrasound loop recording, coupled with a straightforward decision tree catering to the primary causes of acute respiratory failure. The protocol's design focuses on expediting the diagnostic process and enhancing its efficiency. Adopted from Lichtenstein et al.

The main objective of the BLUE protocol is to categorize lung artifacts into a distinct "lung profile" based on anatomical signs and visual artifacts observed on the ultrasound images. By following a predefined decision tree outlined in the protocol, physicians can make on-the-spot diagnoses and identify the underlying causes of acute respiratory failure. The protocol has shown remarkable diagnostic accuracy, successfully identifying the six most common diseases associated with acute respiratory failure in 97 percent of cases, with an overall accuracy of 90.5 percent [35].

The profiling of the lung involves detecting specific signs in the ultrasound loops. The protocol defines ten basic signs, which can be further categorized into primary and secondary signs. Primary signs, such as the pleural line, A-line, quad sign, fractal sign, tissue-like sign, and B-lines or lung rockets, can be observed on single still frames without considering the entire ultrasound loop [19]. These signs serve as important indicators of lung pathology and assist in the diagnostic process.

In contrast, secondary signs, including lung sliding, sinusoid sign, and lung point, require a comprehensive analysis of the entire video loop. These signs often exhibit movement or dynamic patterns that are not discernible on still frames alone. Detecting these secondary signs necessitates a thorough examination of the ultrasound loop as a whole, considering the temporal progression of lung artifacts.

To streamline the identification of initial indicators, we've examined different methods, such as the utilization of 2D multiclass semantic segmentation methodologies. Through the segmentation of static frames, it becomes feasible to train deep learning

algorithms in precisely recognizing and categorizing various pulmonary anomalies. This automatic identification of primary indicators shows potential in accelerating and establishing a uniform diagnostic procedure, especially for medical practitioners with limited experience or within scenarios requiring prompt decisions.

Concurrently, we've undertaken the development of a classification framework that is specifically concentrated on discerning the presence or absence of lung sliding. This secondary indicator holds pivotal importance in distinguishing among diverse pulmonary conditions. Leveraging an extensive, custom-designed clinical dataset, we've educated convolutional neural networks (CNNs) to scrutinize B-mode (2D mode) image sequences and transform them into M-mode ("motion" mode) images. These M-mode images facilitate the extraction of sample lines by proficient annotators, thus facilitating the training and assessment of the CNNs. The most proficient CNN achieved an impressive precision rate of 93.0

In summary, the BLUE protocol has risen as a valuable instrument for swiftly diagnosing acute respiratory failure, especially within intensive care environments. The transition from static images to brief ultrasound sequences has notably elevated the precision and effectiveness of lung assessments. The continuous endeavors in automating the identification of initial indicators using deep learning methodologies demonstrate encouraging outcomes, laying the groundwork for more consistent and easily accessible diagnostic procedures. Furthermore, the creation of a classification network dedicated to identifying lung sliding enhances the comprehensive diagnostic prowess of the BLUE protocol. These innovations hold the capacity to transform the landscape of lung ultrasound analysis, empowering timelier and more precise diagnoses for critically unwell patients.

2 Methods

2.1 Deep Neural Networks

In the past few years, there has been a growing trend in employing deep neural networks for enhancing the precision and effectiveness of medical procedures within the realm of medical imaging. These sophisticated neural networks have found application across a range of medical imaging assignments, encompassing tasks like object identification, segmentation, and the reconstruction of images [38, 39].

Illustratively, convolutional neural networks have found utility in discerning and categorizing diverse irregularities present in radiological images. Noteworthy instances encompass the detection of tumors within MRI scans or the identification of lesions in dermatological images. As an illustration, a study conducted in 2020 focused on utilizing a convolutional neural network to classify chest X-ray images. The network was primed for binary categorization and successfully achieved a commendable accuracy rate of 94.6

In addition to its advantages, it's crucial to acknowledge the challenges associated with the application of deep learning networks in medical image analysis. Among

the foremost hurdles is the establishment of a dataset containing a requisite quantity and caliber of annotated medical data. The task of annotation involves the expertise of medical professionals, who furnish the reference truth necessary for training and appraising the network's performance. Given that annotation demands the involvement of these specialized medical experts, procuring a substantial volume of annotated medical images often proves to be arduous and complex.

Deep neural networks have demonstrated their effectiveness in the scrutiny of lung ultrasound images as well. Over the past years, a plethora of research findings have emerged in this domain, examining the efficacy of deep neural networks in identifying various irregularities within lung ultrasound images.

An additional instance is found in a study authored by Cheng and Lam in 2021. In this research, a customized U-Net architecture was employed for the binary segmentation of lung ultrasound images. Notably, the encoder component of the U-Net was substituted with a VGG16 network that had been pre-trained on the ImageNet database. This hybrid architecture was subsequently trained on a dataset comprising 400 ultrasound images tailored to the specific challenge. The outcome was a Dice score of 0.86, underscoring the success achieved in accurate binary segmentation.6 [40–42].

In a separate investigation documented in 2020, Roy et al. explored the segmentation of COVID biomarkers within lung ultrasound images. Their approach hinged on a U-Net-based model, which underwent training using a comprehensive set of 277 annotated ultrasound loops. This concerted effort yielded a noteworthy 96

These studies unequivocally showcase the practicality and relevance of employing deep neural networks for the realm of medical image processing.

2.2 U-Net

U-Net constitutes a fully convolutional neural network architecture that was initially devised specifically to address medical image segmentation assignments [42]. Since its original introduction in 2015, U-Net has persisted as the foundational framework for numerous cutting-edge neural network architectures, particularly within the sphere of medical image analysis, up until the present day. [43, 44]. The architecture of the network employs an encoder-decoder design, enriched by skip connections. The network's initial segment is the encoder, often referred to as the "contracting path," where each stage progressively diminishes the dimensions of the input image in terms of width and height, while simultaneously amplifying the extracted features. Subsequently, the decoder component (termed the "expanding path") methodically upscales the input and carries out additional convolutions.

A significant aspect of this architecture is the incorporation of skip connections. These connections link the output activation map of convolutional layers in the encoder segment to the input of convolutional layers in the decoder segment, preserving a match in the width and height dimensions between the encoder's output and the decoder's input. This strategic linkage allows the network to retain and integrate more information from preceding layers. This encompasses retaining the original

spatial characteristics of the input image, a pivotal attribute for segmentation undertakings.

3 Implementation

Although the present implementation is a considerable distance from direct clinical utilization, the chosen methodologies for its development and prototyping hold the potential to pave the way for a more robust and diagnosis-focused decision support system.

3.1 Approach

The central objective addressed within this paper revolves around the identification of fundamental indicators present in lung ultrasound images, bearing significance to the BLUE protocol. To address this objective, we employ a deep neural network tailored for multi-class semantic segmentation. This process is executed on every annotated frame contained within the captured ultrasound sequences. Essentially, this undertaking can be deconstructed into two primary components, each of which can be further subdivided into smaller constituent subtasks:

- Data Conditioning – in this step, we prepare and arrange our raw data into a format that is suitable for network training;
- Network Implementation – the actual implementation, training, and fine tuning of the neural network.

In the following sections we discuss both stages in detail.

3.2 Imaging protocol

Following the method developed by Daniel Lichtenstein, the creator of the BLUE and PINK protocols, we adhered to his specified approach of using three designated points on each side of the chest to gather imaging data. The imaging was integrated into the daily medical routine. Informed by prior expert analyses and considering the uneven and sporadic nature of COVID-19 pneumonia, we captured video loops in both longitudinal and transverse orientations. The specific locations and orientations were systematically labeled using predetermined codes. Our protocol enabled a single operator to conduct the assessments without the need for additional assistance. Given that all the examinations were carried out by lung ultrasound (LUS) specialists with over seven years of experience, we didn't document the precise duration of each examination, as it wasn't deemed significant. The assessment of the loops was generally done off-line. However, if the operator identified an urgent situation during the analysis, such as an emergency equivalent, they immediately communicated this to the clinical team. An example of this occurred during one examination when an acute pneumothorax on the right side was unexpectedly detected, leading to a timely intervention that ultimately saved the patient's life.

Basic Settings	Obligatory value/range
Depth	Pleural line + 3–8 cm
Focus	Multifocus: OFF; Single focus: at pleural line
Gain	Optimized for main artifacts (A-lines, B-lines, consolidation)
Image-processing features	THI: OFF, XRes: OFF, CrossXBeam: OFF, SRI: OFF
MI	< 0.7
TIs	< 0.1

Table 1
Abbreviations: MI: mechanical index; TI: soft tissue thermal index; THI: tissue harmonic imaging; XRes: speckle noise reduction; CrossXBeam: spatial compounding; SRI: speckle reduction

3.3 Ultrasound settings, data collection protocol

Our objective was to establish a consistent and comprehensive US examination procedure. Utilizing specific settings on our equipment, we successfully gathered a collection of images that effectively depicted the pleura and underlying artifacts, ensuring no critical observations were overlooked. Prioritizing patient safety, we carefully regulated the thermal index (TI) and mechanical index (MI) to adhere to global safety norms. The devices employed included a Philips CX50 (Philips Healthcare, The Netherlands) with a Philips C5-1 convex probe (1-5 MHz) and a GE Venue GO (GE Healthcare, IL, USA) with a C1-5-RS convex probe (1.4-5.7 MHz). Essential default settings are detailed in Table 1. Drawing from our clinical experience and aligning with global practices in lung ultrasound (LUS), we opted to capture video loops lasting 4–6 seconds instead of single images, for enhanced accuracy. The operator of the equipment was not informed about the patient's clinical progress, and did not participate in their care. These loops were recorded under a pseudo-anonymized system (using a unique patient code) and stored on the hard drive in DICOM (Digital Imaging and Communication in Medicine) format, fully compliant with GDPR (EU General Data Protection Regulation) guidelines. Subsequently, patient information was transferred to an encrypted, Microsoft Excel-based database, through a triple-layered security protocol.

3.4 Imaging protocol

Following the method developed by Lichtenstein, the creator of the BLUE and PINK protocols, we adhered to his specified approach of using three designated points on each side of the chest to gather imaging data. The imaging was integrated into the daily medical routine. Informed by prior expert analyses, and considering the uneven and sporadic nature of COVID-19 pneumonia, we captured video loops in both longitudinal and transverse orientations. The specific locations and orientations were systematically labeled using predetermined codes.

Our protocol enabled a single operator to conduct the assessments without the need

for additional assistance. Given that all the examinations were carried out by LUS specialists with over seven years of experience, we did not document the precise duration of each examination, as it was not deemed significant. The assessment of the loops was generally done off-line. However, if the operator identified an urgent situation during the analysis, such as an emergency equivalent, they immediately communicated this to the clinical team. An example of this occurred during one examination when an acute pneumothorax on the right side was unexpectedly detected, leading to a timely intervention that ultimately saved the patient's life.

3.5 Data Conditioning

The dataset harnessed for this endeavor has been meticulously curated by medical professionals affiliated with Semmelweis University, Budapest. In addition to supplying the requisite data, these experts contributed their invaluable medical acumen to the project. This encompassed the meticulous annotation of raw data and ensuring that the project's trajectory was aligned with a clinically coherent perspective.

As of the time of composing this article, our repository comprises data sourced from 22 RT-PCR confirmed SARS-CoV-2 infected patients, in addition to data extracted from 18 RT-PCR confirmed SARS-CoV-2 negative ("non-COVID") patients. This data has been meticulously gathered by Semmelweis University's Department of Anaesthesiology and Intensive Therapy. Notably, all the data acquisition procedures adhere rigorously to the BLUE protocol, a universally recognized and adopted standardized methodology. This meticulous adherence ensures the replicability of the procedure and the coherence of our data.

Each patient's dataset comprises multiple ultrasound loops, with each loop spanning 5 to 6 seconds in duration. These loops are captured from distinct points as delineated by the BLUE protocol. Given the aggregate of 40 patients, the dataset encompasses a cumulative count of approximately 630 loops, encompassing over 200,000 individual frames. Our efforts to augment this dataset are ongoing, involving the collaboration of additional hospitals and research establishments. To uphold the ethical considerations, all our data is pseudonymized, and we have procured explicit authorization from Semmelweis University's Research Ethics Committee to utilize this dataset. In general, most recent ethically aligned engineering design principles (IEEE 7000, IEEE 70007 standards) were followed during the project [45].

In the process of data preparation, we extensively leverage the capabilities of the 3D Slicer tool. [46]. Certainly, 3D Slicer stands out as an open-source, research-centric image processing tool, primarily geared towards medical imaging tasks. While its scope is not limited solely to medical imaging, it excels in this domain. A notable attribute of 3D Slicer is its robust segmentation capabilities. Moreover, the tool offers programmable extension capabilities and a conveniently embedded Python command line, which significantly facilitates data transformation and extraction procedures. These inherent functionalities align seamlessly with our objectives. We adeptly harness the built-in segmentation module to annotate our raw ultrasound loops, and the Python interface expedites the automated extraction of this annotated data, rendering 3D Slicer an ideal fit for our requirements.

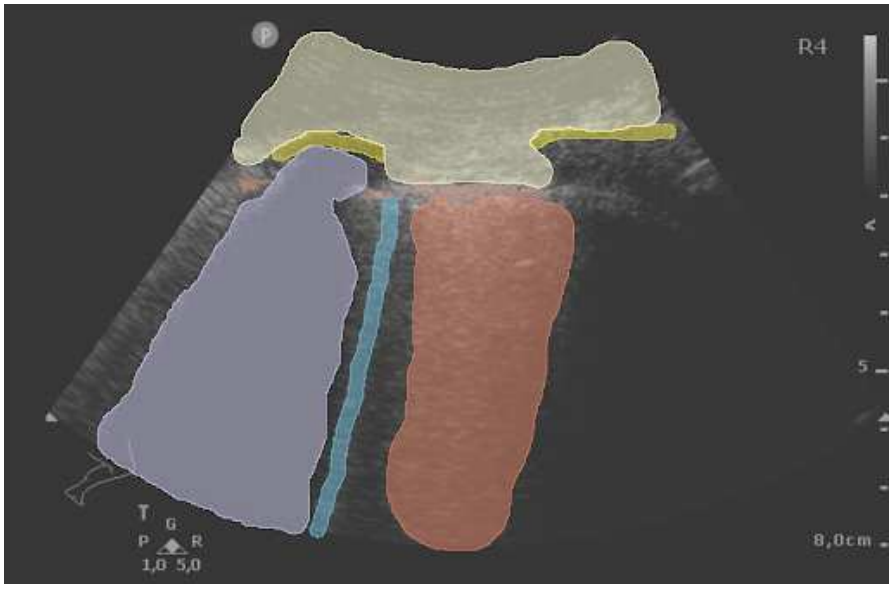


Figure 2
A manually segmented ultrasound frame. Visible segmented classes are: thoracic wall (pale yellow), rib (yellow), rib shadow (purple), B-line (blue), septal rockets (red), unhealthy pleural line (pale orange)

In the initial step, the raw data undergoes annotation under the supervision of medical experts within the 3D Slicer platform. This process involves the manual delineation of segments pertinent to our particular scenario. These encompass the primary indicators outlined earlier, accompanied by a select few supplementary landmark segments. The segmentation occurs on a frame-by-frame basis, with every fifth frame within each loop being subjected to this procedure. It's noteworthy that the entirety of the loop doesn't necessarily require annotation, as capturing a complete breathing cycle suffices for our purposes.

Upon completion of annotation, the subsequent phase involves the extraction of data from 3D Slicer. As not every individual frame undergoes segmentation, identification of frames with attached segmentation data becomes imperative. Moreover, a series of standardization and cleanup steps are necessary for each relevant frame. Certain extraneous details, originating from the ultrasound machine itself (such as machine settings and brand), are irrelevant to our analysis. To address this, a mask is applied to retain solely the pertinent ultrasound image, effectively eliminating the extraneous information. Additionally, the images are transformed into a uniform quadratic format, specifically 256x256 pixels.

To achieve this quadratic format, an initial resizing is performed, such that the larger dimension of the original image matches the desired 256 pixels, while maintaining the original aspect ratio. During this resizing, linear interpolation is employed for the ultrasound images, while nearest neighbor interpolation is used for the segmentation masks. Subsequently, the smaller dimension is extended to reach 256 pixels

using zero values, while preserving the central segment of the image.

Post-cleanup and extraction, these relevant frames are integrated to constitute a comprehensive dataset. This dataset is structured on a patient-specific basis, ensuring that data from a given patient remains exclusive either to the training or validation set. This segregation mimics real-world application scenarios, where the trained model is applied to new, unknown patients' BLUE ultrasound loops. This setup ensures more realistic outcomes in model evaluation.

At present, our dataset encompasses a total of 1097 manually segmented ultrasound frames, derived from the data of four patients.

3.6 Network Training

Our approach involved implementing a standard U-Net model, enhanced by the incorporation of zero-padding within the convolutional layers. This modification was strategically employed to preserve the original dimensions of the image. In the testing phase, our network was configured with an input size of 256 by 256 pixels, and each convolutional layer was equipped with 16 filters, contributing to the model's feature extraction capabilities.

During the training process, we adopted a leave-one-out cross-validation technique rooted in patient-centric segmentation. This methodology entails setting aside the data from a single patient as the validation set in each round, while the model is simultaneously trained on the remaining dataset. This cross-validation scheme ensures comprehensive validation and helps mitigate potential overfitting.

In evaluating the performance of our model, we focused on two key metrics: validation accuracy and validation loss. To quantify the model's predictive accuracy and assess the fidelity of its output, we computed the validation accuracy. Meanwhile, the validation loss, measured using the sparse categorical cross-entropy loss function, provides insight into the dissimilarity between predicted and actual segmentation outcomes. This comprehensive approach offers a robust framework for effectively gauging the model's proficiency in the context of medical image analysis.

4 Results and Discussion

By employing the method outlined in the preceding section, we attained an average validation accuracy of 87.53

At this juncture, the most pronounced challenge lies in the scarcity of available annotated data. With the present dataset size, the model exhibits signs of overfitting within only a few training epochs, necessitating premature training termination. Our forthcoming efforts will be directed towards addressing this limitation. We intend to bolster our annotated data reservoir and explore diverse data augmentation techniques to alleviate this challenge effectively.

(a)
Ground truth

(b)
Model prediction

Figure 3
Illustrative Prediction Outcome of the Model: Noteworthy Segmentation Classes include the Thoracic Wall (highlighted in yellow), Ribs (displayed in orange), Rib Shadows (rendered in pale orange), and B-lines (depicted in purple).

Once a substantial annotated dataset is amassed, our focus will shift towards conducting a comprehensive hyperparameter sweep to determine the optimal configuration. While we've engaged in some level of hyperparameter optimization, the relatively modest dataset size has somewhat limited its impact. Once segmentation attains a satisfactory level of performance, our endeavors will extend towards identifying the secondary signs elucidated in the second chapter. Subsequently, we will be poised to embark on the comprehensive implementation of the entire decision tree integral to the BLUE protocol.

Another avenue for potential enhancement lies in revisiting the roster of segmentation classes currently utilized for model training. Streamlining the model by omitting or consolidating specific classes that do not contribute to relevant diagnostic insight could potentially simplify training and expedite the annotation process. An associated adjustment could involve modifying the loss function to facilitate the assignment of varying weights to individual segmentation classes. Notably, even in the existing segmentation schema, the prominence of the background class outweighs others. This concern could intensify with a reduced number of segments, and recalibrating the significance of the background class might yield improved outcomes.

Fine-tuning the model architecture and experimenting with diverse model types to facilitate performance comparison is another avenue for potential improvement. However, the dearth of sufficiently annotated data has thus far hindered the pursuit of this optimization. Given the present stage, a meaningful model architecture comparison might be elusive.

In addition to enhancing the current model, we intend to reevaluate our use case scenario to explore alternative approaches beyond segmentation. Our foremost challenge is the time-consuming process of manual annotation, particularly the meticulous contouring required for each frame. We're actively considering the prospect of leveraging object detection methods as an alternative. By employing this approach, annotating would involve simply drawing bounding boxes around objects rather than intricate contouring, thereby significantly expediting the annotation process. This transition alone could prove advantageous, even if the new approach doesn't yield immediate tangible benefits.

Another avenue involves the integration of the segmentation model with a medical ultrasound device, enabling real-time ultrasound image segmentation at the patient's bedside. While not directly leading to diagnosis, this real-time segmentation could still furnish valuable information to the attending physician during the examination, potentially aiding in timely decision-making.

These potential directions underscore our commitment to refining and innovating upon our current framework to overcome existing challenges and enhance the clinical utility of our approach.

Future research endeavors encompass the pursuit of full automation in image acquisition by integrating a collaborative robotic arm. This innovative approach holds the promise of enhancing data consistency, a crucial facet for accurate analysis. Moreover, such automation could prove particularly advantageous in pandemic scenarios, mitigating the risk of human exposure and facilitating safe and efficient medical

procedures. [33, 47].

5 Conclusion

Our research is dedicated to advancing computer-assisted support for the BLUE protocol. In the initial phase, we have successfully implemented a U-Net based deep neural network tailored for segmenting and identifying vital anatomical elements within lung ultrasound images.

The outcomes thus far reveal a highly promising conceptual foundation. Notwithstanding the modest count of annotated images and the brief training duration, the model's predictions exhibit discernible alignment with the ground truth. This alignment is particularly pronounced for segments that are more prevalent in nature, such as the thoracic wall or rib shadows. This initial progress underscores the potential of our approach to significantly enhance the diagnostic capabilities of the BLUE protocol through computer-assisted assistance.

Future work

Our current project is an essential component of a modular decision support system, incorporating artificial intelligence solutions. As per our vision, this system can be immensely beneficial to our clinical colleagues. The foundation of the system relies on well-validated and widely used emergency lung ultrasound protocols.

Our primary objective is to implement and enhance a framework that offers real-time support to physicians during bedside examinations. Additionally, this framework aims to facilitate the acquisition of lung ultrasound skills as part of structured medical training.

We plan to further advance the system by gathering more clinical data and expanding our annotated data bank. This expansion is expected to improve the system's metrics, making it suitable for real testing in clinical conditions.

Looking ahead, we also consider the possibility of automating and robotizing the system in various directions. Such advancements could lead to additional cost savings in terms of human resources for healthcare systems that operate within limited frameworks.

Acknowledgment

The authors thank the support of Roland Incze in the research.

C. Barr is supported by an NSERC CGS-D. G. Fichtinger is supported as an NSERC Canada Research Chair.

R. Szabó was supported by the MITACS Globalink Research Award.

T. Haidegger's work is partially supported by project no. 2019-1.3.1-KK-2019-00007, provided by the National Research, Development and Innovation Fund of

Hungary. T. Haidegger is a "Consolidator Researcher", supported by the Distinguished Researcher program of Óbuda University.

References

[1] F. Mojoli, B. Bouhemad, S. Mongodi, and D. Lichtenstein. Lung ultrasound for critically ill patients. *American journal of respiratory and critical care medicine*, 199(6):701–714, 2019.

[2] S. Mongodi, D. De Luca, A. Colombo, A. Stella, E. Santangelo, F. Corradi, L. Gargani, S. Rovida, G. Volpicelli, B. Bouhemad, et al. Quantitative lung ultrasound: technical aspects and clinical applications. *Anesthesiology*, 134(6):949–965, 2021.

[3] R. Brat, N. Yousef, R. Klifa, S. Reynaud, S. S. Aguilera, and D. De Luca. Lung ultrasonography score to evaluate oxygenation and surfactant need in neonates treated with continuous positive airway pressure. *JAMA pediatrics*, 169(8):e151797–e151797, 2015.

[4] B. Bouhemad, S. Mongodi, G. Via, and I. Rouquette. Ultrasound for "lung monitoring" of ventilated patients. *Anesthesiology*, 122(2):437–447, 2015.

[5] G. Volpicelli, M. Elbarbary, M. Blaivas, D. A. Lichtenstein, G. Mathis, A. W. Kirkpatrick, L. Melniker, L. Gargani, V. E. Noble, G. Via, et al. International evidence-based recommendations for point-of-care lung ultrasound. *Intensive care medicine*, 38:577–591, 2012.

[6] S. Mongodi, B. Bouhemad, A. Orlando, A. Stella, G. Tavazzi, G. Via, G. A. Iotti, A. Braschi, and F. Mojoli. Modified lung ultrasound score for assessing and monitoring pulmonary aeration. *Ultraschall in der Medizin-European Journal of Ultrasound*, 38(05):530–537, 2017.

[7] M. Szabó, A. Bozó, K. Darvas, S. Soós, M. Őzse, and Z. D. Iványi. The role of ultrasonographic lung aeration score in the prediction of postoperative pulmonary complications: an observational study. *BMC anesthesiology*, 21:1–10, 2021.

[8] A. Dargent, E. Chatelain, L. Kreitmann, J.-P. Quenot, M. Cour, L. Argaud, and C.-L. study group. Lung ultrasound score to monitor covid-19 pneumonia progression in patients with ards. *PLoS One*, 15(7):e0236312, 2020.

[9] R. Elek, T. D. Nagy, D. A. Nagy, B. Takács, P. Galambos, I. Rudas, and T. Haidegger. Robotic platforms for ultrasound diagnostics and treatment. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 1752–1757. IEEE, 2017.

[10] G. Fichtinger, J. Troccaz, and T. Haidegger. Image-guided interventional robotics: Lost in translation? *Proceedings of the IEEE*, 110, 2022.

[11] T. Haidegger, S. Speidel, D. Stoyanov, and R. Satava. Robot-assisted minimally invasive surgery—surgical robotics in the data age. *Proceedings of the IEEE*, 110:835–846, 2022.

[12] T. D. Nagy and T. Haidegger. Performance and capability assessment in surgical subtask automation. *Sensors*, 22:2501, 2022.

[13] R. Raheja, M. Brahmavar, D. Joshi, and D. Raman. Application of lung ultrasound in critical care setting: a review. *Cureus*, 11(7), 2019.

[14] M. Szabó, G. Orosz, Z. D. Iványi, and K. Darvas. Use of thoracic and lung ultrasound in general anesthesia. *Orvosi Hetilap*, 164(22):864–870, 2023.

[15] P. S. Pang, F. M. Russell, R. Ehrman, R. Ferre, L. Gargani, P. D. Levy, V. Noble, K. A. Lane, X. Li, and S. P. Collins. Lung ultrasound–guided emergency department management of acute heart failure (blushed-ahf) a randomized controlled pilot trial. *Heart Failure*, 9(9):638–648, 2021.

[16] C. Urbán, P. Galambos, G. Györök, and T. Haidegger. Simulated medical ultrasound trainers a review of solutions and applications. *Acta Polytechnica Hungarica*, 15(7):111–131, 2018.

[17] D. A. Lichtenstein and G. A. Meziere. Relevance of lung ultrasound in the diagnosis of acute respiratory failure*: the blue protocol. *Chest*, 134(1):117–125, 2008.

[18] D. A. Lichtenstein. Lung ultrasound in the critically ill. *Annals of intensive care*, 4(1):1–12, 2014.

[19] D. Lichtenstein. *Lung Ultrasound in the Critically Ill: The BLUE Protocol*. Springer International Publishing, 2015.

[20] G. Soldati, A. Smargiassi, L. Demi, and R. Inchingolo. Artifactual lung ultrasonography: It is a matter of traps, order, and disorder. *Applied sciences*, 10(5):1570, 2020.

[21] A. Khamis, H. Li, E. Prestes, and T. Haidegger. Ai: a key enabler of sustainable development goals, part 1. *IEEE Robotics & Automation Magazine*, 26(3):95–102, 2019.

[22] A. Khamis, H. Li, E. Prestes, and T. Haidegger. Ai: A key enabler for sustainable development goals: Part 2. *IEEE Robotics & Automation Magazine*, 26(4):122–127, 2019.

[23] T. Haidegger, V. Mai, C. Mörch, D. Boesl, A. Jacobs, B. Rao R, A. Khamis, L. Lach, and B. Vanderborght. Robotics: Enabler and inhibitor of the sustainable development goals. *Sustainable Production and Consumption*, 43:422–434, 2023.

[24] H. Ostensen. Developing countries. *Ultrasound in Medicine and Biology*, 26:S159–S161, 2000.

[25] S. Sippel, K. Muruganandan, A. Levine, and S. Shah. Use of ultrasound in the developing world. *International journal of emergency medicine*, 4(1):1–11, 2011.

[26] R. Z. Szabó, G. Orosz, T. Ungi, C. Barr, C. Yeung, R. Incze, G. Fichtinger, J. Gál, and T. Haidegger. Automation of lung ultrasound imaging and image processing for bedside diagnostic examinations. In *2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 779–784. IEEE, 2023.

[27] D. A. Lichtenstein. Current misconceptions in lung ultrasound: a short guide for experts. *Chest*, 156(1):21–25, 2019.

[28] S. Mongodi, E. Santangelo, D. De Luca, S. Rovida, F. Corradi, G. Volpicelli, L. Gargani, B. Bouhemad, and F. Mojoli. Quantitative lung ultrasound: time for a consensus? *Chest*, 158(2):469–470, 2020.

[29] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, T. Perrone, D. F. Briganti, S. Perlini, E. Torri, A. Mariani, E. E. Mossolani, et al. Time for

a new international evidence-based recommendations for point-of-care lung ultrasound. *Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine*, 40(2):433–434, 2021.

[30] F. Mento, T. Perrone, V. N. Macioce, F. Tursi, D. Buonsenso, E. Torri, A. Smargiassi, R. Inchingolo, G. Soldati, and L. Demi. On the impact of different lung ultrasound imaging protocols in the evaluation of patients affected by coronavirus disease 2019: how many acquisitions are needed? *Journal of Ultrasound in Medicine*, 40(10):2235–2238, 2021.

[31] G. Soldati, A. Smargiassi, R. Inchingolo, D. Buonsenso, T. Perrone, D. F. Briganti, S. Perlini, E. Torri, A. Mariani, E. E. Mossolani, et al. Proposal for international standardization of the use of lung ultrasound for patients with covid-19: a simple, quantitative, reproducible method. *Journal of Ultrasound in Medicine*, 39(7):1413–1419, 2020.

[32] L. Demi. Lung ultrasound: The future ahead and the lessons learned from covid-19. *The Journal of the Acoustical Society of America*, 148(4):2146–2150, 2020.

[33] A. Khamis, J. Meng, J. Wang, A. T. Azar, E. Prestes, Á. Takács, I. J. Rudas, and T. Haidegger. Robotics and intelligent systems against a pandemic. *Acta Polytechnica Hungarica*, 18(5):13–35, 2021.

[34] N. Islam, S. Ebrahimzadeh, J.-P. Salameh, S. Kazi, N. Fabiano, L. Treanor, M. Absi, Z. Hallgrimson, M. M. Leeflang, L. Hooft, et al. Thoracic imaging tests for the diagnosis of covid-19. *Cochrane Database of Systematic Reviews*, (3), 2021.

[35] D. A. Lichtenstein and D. A. Lichtenstein. The luci-flr project: Lung ultrasound in the critically ill–a bedside alternative to irradiating techniques, radiographs and ct. *Lung Ultrasound in the Critically Ill: The BLUE Protocol*, pages 217–225, 2016.

[36] B. Yousuf, K. S. Sujatha, H. Alfoudri, and V. Mansurov. Transport of critically ill covid-19 patients. *Intensive Care Medicine*, 46:1663–1664, 2020.

[37] G. Orosz, P. Gyombolai, J. T. Tóth, and M. Szabó. Reliability and clinical correlations of semi-quantitative lung ultrasound on blue points in covid-19 mechanically ventilated patients: The 'blue-luss'—a feasibility clinical study. *PLoS One*, 17(10):e0276213, 2022.

[38] S. K. Zhou, D. Rueckert, and G. Fichtinger. *Handbook of medical image computing and computer assisted intervention*. Academic Press, 2019.

[39] S. R. Chandaran, G. Muthusamy, L. R. Sevalaiappan, and N. Senthilkumaran. Deep learning-based transfer learning model in diagnosis of diseases with brain magnetic resonance imaging. *Acta Polytechnica Hungarica*, 19(5):127–147, 2022.

[40] D. Cheng and E. Lam. Transfer learning u-net deep learning for lung ultrasound segmentation. 10 2021.

[41] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. 2014.

[42] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[43] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Köhler, T. Norajitra, S. J. Wirkert, and K. H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *CoRR*, abs/1809.10486, 2018.

[44] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.

[45] M. A. Houghtaling, S. R. Fiorini, N. Fabiano, P. J. Gonçalves, O. Ulgen, T. Haidegger, J. L. Carbonera, J. I. Olszewska, B. Page, Z. Murahwi, et al. Standardizing an ontology for ethically aligned robotic and autonomous systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023.

[46] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. M. Fennessy, M. Sonka, J. M. Buatti, S. R. Aylward, J. V. Miller, S. Pieper, and R. Kikinis. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341, 2012.

[47] T. Haidegger. Autonomy for Surgical Robots: Concepts and Paradigms. *IEEE Transactions on Medical Robotics and Bionics*, 1(2):65–76, 2019.

Effect of Single-Slice CT Segmentation Methods on Fat Volume and Body Shape Estimation

Lilla Szatmáriné Egeresi

Division of Radiology and Imaging Science, Department of Medical Imaging, Faculty of Medicine, University of Debrecen, Nagyerdei krt. 98, 4032 Debrecen, Hungary, E-mail: egeresi.lilla@med.unideb.hu

András Székely

Kenézy Gyula Hospital, Department of Radiology, Bartók Béla út 2-26, 4031 Debrecen, Hungary, E-mail: andras.szekely@med.unideb.hu

Piroska Kallos-Balogh

Division of Nuclear Medicine and Translational Imaging, Department of Medical Imaging, Faculty of Medicine, University of Debrecen, Nagyerdei krt. 98, 4032 Debrecen, Hungary, E-mail: balogh.piroska@med.unideb.hu

Lajos Trón

Division of Nuclear Medicine and Translational Imaging, Department of Medical Imaging, Faculty of Medicine, University of Debrecen, Nagyerdei krt. 98, 4032 Debrecen, Hungary, E-mail: tron.lajos@med.unideb.hu

Ildikó Garai

Scanomed Nuclear Medicine Center, Nagyerdei krt. 98, 4032 Debrecen, Hungary, E-mail: garai.ildiko@scanomed.hu

László Balkay

Division of Nuclear Medicine and Translational Imaging, Department of Medical Imaging, Faculty of Medicine, University of Debrecen, Nagyerdei krt. 98, 4032 Debrecen, Hungary, E-mail: balkay.laszlo@med.unideb.hu

Abstract: Several automatic and semi-automatic algorithms for adipose tissue (AT) segmentation in CT have been proposed. Our study aimed to determine the effect of the preselected HU range, on the resulting AT volumes and establish whether there is a relationship between body shape and AT values determined, from a single slice. Scans of 98 patients acquired using two CT protocols, were used. Three axial slices were selected from each subject's CT data. Subcutaneous and visceral adipose tissues (SAT, VAT) were manually segmented and analyzed using three different HU ranges. In addition, a simple BMI calculation model was created by the segmented data. The areas segmented with the three different HU ranges, correlated well with each other in the case of SAT (r^2=0.99, and r^2=0.99) and VAT (r^2=0.99, and r^2=0.998). The preselected slice position had no significant effect on correlation; however, the absolute values of ATs were statistically different. CT image data acquired with higher tube current yielded a better correlation between SAT, VAT, and BMI. We also found that the correlation of VAT area to mass and BMI was weaker than the corresponding SAT correlations. The simple model-based BMI estimation is in line with real BMI data (males: r^2=0.78, females: r^2=0.841). The segmentation threshold does not substantially affect correlation, among the segmented AT values; however, their absolute values are significantly different. In addition, and interestingly, the body shape can be accurately described from the segmented AT data from a single CT slice.

Keywords: CT; fat; segmentation

1 Introduction

It is well known that obesity is closely associated with diabetes, fatty liver, cardiovascular disease and various cancers [2, 17, 24, 33]. Visceral fat resides between organs, while subcutaneous fat is located beneath the skin. Another substantial difference between these two types of fat is that visceral fatty tissue functions as an endocrine organ, contributing to the pathogenesis of several diseases [5]. Particularly, VAT has a major impact on these diseases, and several articles underline the importance of accurately measuring visceral tissue volume [11, 28, 35]. On the other hand, there is also increased attention on body composition in which SAT and VAT determination is a critical element in addition to muscle volume measurement. Neither X-ray nor any of the nuclear medicine techniques are unable to help in precisely determining the amount of fatty tissue, only MRI and CT can provide means to segment the adipose regions in the body [9, 12, 13, 15, 18, 25]. Reviewing the literature, we found some articles that tried to estimate adipose tissue with other modalities, such as ultrasound. For example, Stolk et al. measured the distance between the peritoneum and the lumbar spine using ultrasound with a strict protocol to calculate adipose tissue volume [29] [30].

The numerical characterisation of adipose tissue volume calculated from CT data can be regarded as the most frequently used method [16, 34]. Measurement of SAT and VAT values is usually accomplished on a single CT slice at the level of the umbilicus [24] or the level of T12-L1, L2-L3, L3-L4 or L4-L5 vertebral bodies [2,

11, 13, 15, 16, 27, 35]. In theory, total volume measurement of VAT or SAT would be more advantageous than estimation from a single CT slice, however proper segmentation may not be automatically performed at the diaphragm or near the pelvis, moreover limited number of slices available at CT. Several investigators have suggested and have developed sophisticated automatic algorithms for the segmentation of adipose tissue in a large body volume [23] [31], however it has been shown recently that the single slice measurement can accurately predict the changes in VAT volume and body weight.

Today, CT is the only clinically applied imaging modality that provides quantification of the different tissues based on pixels and Hounsfield unit values. Actually, the HU scale is relative because different X-ray beam energies result in different CT values, thus, it is essential to take into consideration the applied tube voltage. Furthermore, the mentioned scale depends on the radiologists. While CT is one of the most important modalities to assess adipose volume, in the literature, several independent segmentation algorithms have been introduced applying different Hounsfield Unit (HU) ranges: −190/–30 HU, −200/–10 HU, −250/–50 HU, −195/–45 HU [5, 11, 19, 23, 24, 28, 34]. To the best of our knowledge, it has not been extensive investigation pertaining to how the HU range selection might influence the segmented SAT/VAT volumes, so we aimed to:

i) Compare three commonly used ranges

ii) Study how the selected slice position affects the segmented adipose tissue area. In addition, one of the research groups has developed a linear model to estimate the patient's body shape and BMI index along with the fat-tissue segmentation technique [10]. However, this model is rather complex, and it depends on at least six different parameters, and the actual formula varies with gender. The model also includes several arbitrary constants, depending on the CT scanner.

iii) Examine these models based on our CT data and to simplify the model.

2 Materials and Methods

Ninety-eight human CT examinations were randomly selected from March 2012 to September 2013. Since all patient related data were retrospectively analyzed, informed consent was not obtained. All CT examinations were scanned by Philips Brilliance 64 CT scanner with two different protocols. However, of the 98 participants, only 14 were tested with both protocols, where the axial length scan was shorter.

The patient population included outpatients who were being assessed for several diseases. Participants' anthropometric data: 51 males mean age of 59 years (range

31-83 years), mean weight of 87 kg (range 54-150 kg), mean height of 1.7 m (range 1.52- 1.9 m), mean body mass index 34 kg/m^2 (range 23-58 kg/m$^{2)}$) and 47 females mean age of 62 (range 32-94 years), mean weight of 75 kg (range 44-110 kg), mean height of 1.62 m (range 1.5-1.77) and mean body mass index 33 kg/m^2 (range 23-55 kg/m$^{2)}$). The local medical ethics committee approved the study and waived the requirement to obtain informed consent (DE RKEB/IKEB 6593-2023). To ensure a constant signal-to-noise ratio, a standard dose optimization algorithm was applied, in which increasing X-ray exposures were used with increasing body weight in both CT protocols. In Protocol-I, a larger mAs range was set (100-200 mAs) for better image quality, while in Protocol-II, X-ray exposures were about half of those used in Protocol-I. The tube voltage was set to 120 kV in all cases.

Three axial slices (L1 vertebra and the hilar region of the right and left kidney, respectively) [25] [26] were selected from the sagittal reconstruction for image processing. Two ROIs (region of interest) were drawn on each image manually (Figure 1) in the MATLAB program on a diagnostic monitor by a radiographer with five years of professional experience in CT imaging; the larger one was defined by the body contour (green line), and the smaller one encompassed the abdominal cavity (red line).

Figure 1
The amount of subcutaneous fat tissue (SAT) is defined by the total number of segmented pixels within the space between the green and red ROIs, while the visceral fat tissue (VAT) is defined by those within the region in the red ROI

Although several automatic adipose tissue segmentation algorithms are proposed in the literature [14, 23, 28, 32], they are not entirely accurate, confirmation and correction are usually necessary by radiologists. In our study, fat tissue segmentation was performed based on three different ranges of attenuation [window level/window width in HU]: -190/-30 HU, -150/-40 HU, -195/-45 [8] [31]. SAT and VAT were determined as the number of segmented pixels within the red ROI and between the red and green ROIs, respectively. Thus, SAT and VAT values were

calculated by using three different attenuation ranges. In addition, the three different SAT (SATi, i=1,2,3) and three different VAT (VATi, i=1,2,3) estimations were obtained for three different sections (Figure 2).

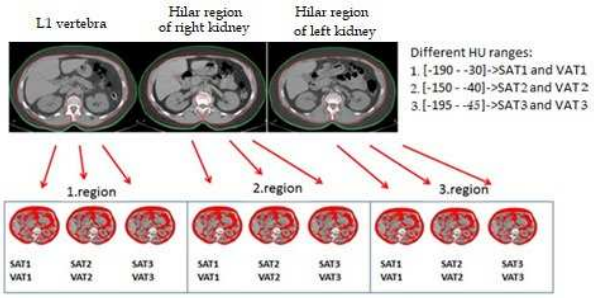


Figure 2

Segmentations were prepared with three different HU ranges using slices acquired at the level of the L.1. vertebra, and at the level of the hilum of the right and the left kidney

In each case, we calculated the BMI as follows:

$$BMI = \frac{bodyweight}{height^2}\left[\frac{kg}{m^2}\right] \tag{1}$$

The weight and height data were collected by the patients' self-declaration. Data obtained from the reconstructed images allowed us to estimate the BMI by the following formulas [8] [10]:

$$BMIest\ male: 2.069 + (0.037 \cdot SQA) - (0.05 \cdot age) + (0.984 \cdot BTD) - (2.647 \cdot L1APD) \tag{2}$$

$$BMIest\ female: -9.163 + (0.252 \cdot BC) + (10.621 \cdot \frac{SQA}{BA}) - (0.08 \cdot age) + (0.597 \cdot BAPD). \tag{3}$$

These formulas require the horizontal and anteroposterior (AP) diameters of the body (BTD, BAPD), the diameter of the vertebral body (L1APD), the body circumference (BC), the total body area (BA) of the axial slice, the subcutaneous fatty area (SQA) and the patient's age (see Figure 3).

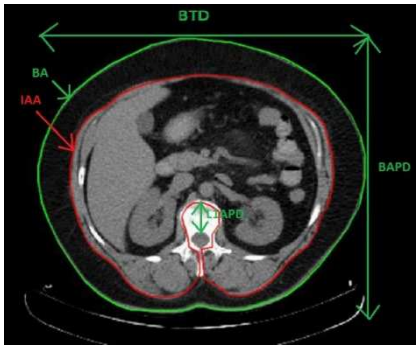


Figure 3

Different parameters for BMI estimation on an axial CT slice. The following parameters are drawn on the image: horizontal and anteroposterior diameters of the body (BTD, BAPD), the diameter of the vertebral body (L1APD), intraabdominal adipose tissue (IAA) and total body area (BA).

We thought the original model was too complex therefore, we tried to simplify it. The BMIest depends on 4-5 data, and the dependence differs for both genders. Furthermore, it is questionable whether constants depend on the actual CT setting. In the literature, we found several additional BMI models [7], and we proposed nine new models based on the original and the current models and labelled them from I to IX (see Table 1).

Table 1

Nine different models were created based on the original model for both sexes

Model numbers	**Female models**
I	BMI_{model}=-9.163+(0.252 ×BC)+ $\left(10.621\times \frac{SQA}{BA}\right)$-(0.081 ×age)+(0.597 ×BAPD)
II	BMI_{model}=a+(b ×BC)+ $\left(c\times \frac{SQA}{BA}\right)$-(d ×age)+(e ×BAPD)
III	BMI_{model}=2.069+(0.037 ×SQA)- (0.051× age)+(0.985 ×BTD)-(2.648 ×L1APD)
IV	BMI_{model}=a+(b ×SQA)- (c× age)+(d ×BTD)-(e ×L1APD)
V	BMI_{model}=a+(b ×BC)
VI	BMI_{model}=a+(b ×BTD)+ (c× BAPD)
VII	BMI_{model}=a+(b ×BC)+c ×age
VIII	BMI_{model}=a+(b ×BTD)+ (c× BAPD)+ (d ×BC)
IX	BMI_{model}=a+(b ×BTD)+ (c× BAPD)+ (c ×BC)+ (e ×age)
Model numbers	**Male models**
I	BMI_{model}=2.069+(0.037 ×SQA)- (0.051× age)+(0.985 ×BTD)-(2.648 ×L1APD)
II	BMI_{model}=a+(b ×SQA)- (c× age)+(d ×BTD)-(e ×L1APD)
III	BMI_{model}=-9.163+(0.252 ×BC)+ $\left(10.621\times \frac{SQA}{BA}\right)$-(0.081 ×age)+(0.597 ×BAPD)
IV	BMI_{model}=a+(b ×BC)+ (c×SQR)-(d ×age)+(e ×BAPD)

V	BMI_{model}=a+(b ×BC)
VI	BMI_{model}=a+(b ×BTD)+ (c× BAPD)
VII	BMI_{model}=a+(b ×BC)+c ×age
VIII	BMI_{model}=a+(b ×BTD)+ (c× BAPD)+ (d ×BC)
IX	BMI_{model}=a+(b ×BTD)+ (c× BAPD)+ (c ×BC)+ (e ×age)

We applied non-linear regression to calculate the coefficients of the nine equations where the following weight function was minimized for each case:

$$h(a,b,c,d,e) = \sum_{k=1}^{n}(BMI_k - BMImodel((a,b,c,d,e))_k)^2 \quad (4)$$

In Eq. 4, the BMI and BMImodel stand for the real and the model equation-based (presented in Table 1) BMIs, respectively. In addition, the sum runs over the whole male or female population (n_{female}= 47, n_{male}= 51).

Hypothesis tests for paired datasets were performed to compare the values obtained from different axial slices and different ranges. The distribution of the data was evaluated using the Anderson-Darling normality test. After that, the corresponding hypothesis test was chosen: paired t-test for normally distributed data and Wilcoxon signed-rank test for the rest. The datasets are considered significantly different if the p-value is smaller than 5%.

All data evaluation and processing were performed using Microsoft Office EXCEL and MATLAB, commercially available programs.

3 Results

3.1 Influence of Segmentation on SAT and VAT Values

AT data of all patients were separated according to gender, slice, X-ray exposure, type of adipose tissue, and applied range of HU during segmentation. First, we analyzed how the AT data obtained by the three segmentation methods are interrelated for gender and X-ray exposure. Figure 4 shows that the correlation coefficients are somewhat higher in the case of Protocol-I compared to Protocol-II for both subcutaneous fat and visceral fat AT data, furthermore, the correlations were independent of HU ranges.

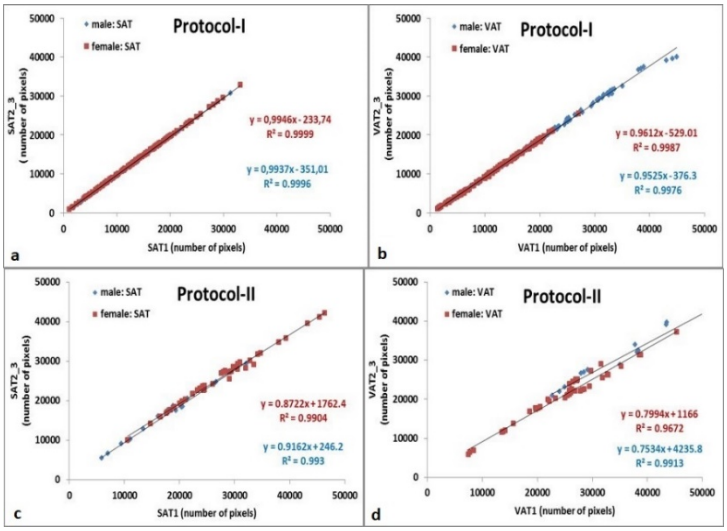


Figure 4

Correlation plots of SAT1-SAT2, SAT1-SAT3(a, c) and VAT1-VAT2, VAT1-VAT3 (b, d) for male and female subjects. Protocol-I was used in a, b and Protocol-II in c, d

The role of the protocol is understandable since Protocol-I generates more photons that may be involved in imaging. Interestingly, the correlation coefficients and the regression equations are very similar for male and female patients. Although the correlations are close, all y axis intersections differ significantly from zero. Generally, the correlation between SATi values is closer as composed to those between VATi data.

As a next step, hypothesis tests were performed to analyze how the predefined HU segmentation ranges influence the AT values obtained at different anatomical regions. Summary statistics for the hypothesis test are provided in Table 2.

Table 2

The resulting p values of the hypothesis test for paired subcutaneous and visceral adipose tissue data from different axial slices

	VAT1, VAT2	**VAT1, VAT3**	**VAT2, VAT3**
male, vertebra	5.14E-10	5.14E-10	4.17E-08
male, left kidney	5.14E-10	5.15E-10	2.66E-07
male, right kidney	5.14E-10	1.19E-28	1.31E-07
female vertebra	4.09E-28	1.46E-26	5.77E-18
female left kidney	3.31E-25	8.21E-25	7.24E-09
female right kidney	7.98E-23	1.12E-22	2.67E-15
	SAT1, SAT2	**SAT1, SAT3**	**SAT2, SAT3**

male, vertebra	5.14E-10	2.77E-30	1.45E-05
male, left kidney	5.14E-10	1.93E-30	5.24E-05
male, right kidney	5.14E-10	1.41E-29	2.41E-05
female vertebra	3.47E-26	2.33E-25	1.45E-05
female left kidney	9.13E-28	4.95E-26	5.24E-05
female right kidney	1.85E-29	8.54E-28	7.23E-13

3.2 Relation of SAT and VAT Data to Patient's Mass

The interrelationship between the number of AT pixels and the patients' weight, or BMIs was also studied. Figure 5a and 5b presents data on the correlation between AT data and patients' weight.

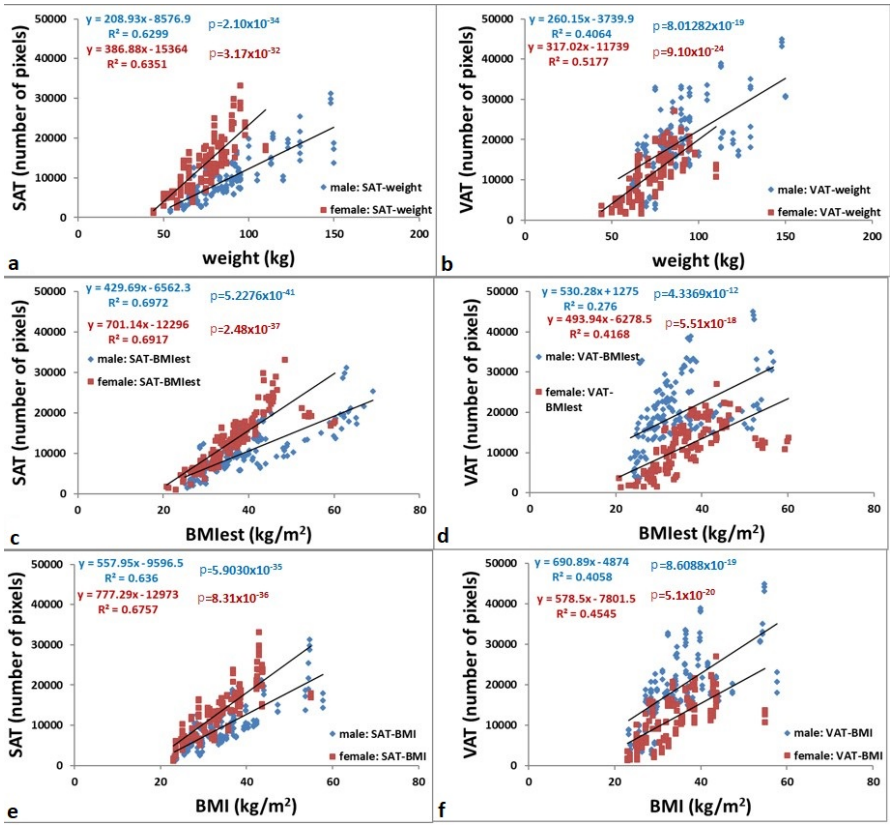


Figure 5

The interrelationship between SAT or VAT and the weight, BMIest and BMI. Adipose tissue data determined with all three HU ranges are plotted. The regression equation, the R2 and p values are shown in each plot.

As expected, the AT pixel numbers increase as a function of body weight, but the correlation coefficients are rather low, especially for visceral fat. In addition, VAT and BMIest correlations (see Figure 5d) are definitely the lowest (r^2=0.276, r^2=0.416). The only exception is the correlation between SAT and BMIest for male subjects (r^2=0.697) and female subjects (r^2=0.691, shown in Figure 5c. SAT and VAT data are better correlated with real BMI data (Figure 5e, f) being characterized by r^2 values similar to those in the correlation of AT data and real weight (Figure 5a, b). Visceral data show a significantly poorer correlation in all cases. It can also be noted that the female body contains more subcutaneous fat, whereas men have more visceral fat. This can serve as useful information since visceral fat can function as an endocrine organ and produce various hormones [16]. These hormones may damage the body in different ways.

3.3 Effect of X-Ray Exposure

Correlation analysis was performed with data obtained from the investigation of the 14 patients scanned using both protocols. The most critical difference between data using the two protocols, is the higher pixel number of the segmented adipose tissue regions yielded by Protocol-II (Figure 6).

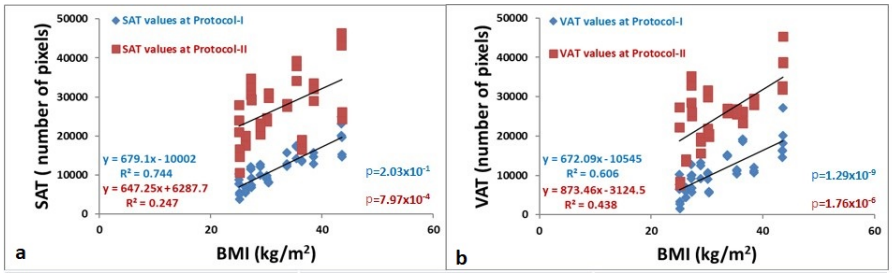


Figure 6

Interrelationship between SAT and BMI based comparison of Protocol-I and Protocol-II.
The numerical result of the regression (equation, the R^2 and p values) are presented in each graph.

The difference can be more than double. Displayed data also show a better correlation of AT data by Protocol-I with real BMI than Protocol-II, suggesting that the segmentation is more accurate if Protocol-I is applied. The correlation is high with the first protocol at real BMI (r^2=0.74 and r^2=0.606). The p value of regression analysis is almost zero (values are shown in Figure 6). Finally, it can be generally stated that the determination of subcutaneous adipose areas is more accurate with higher X-ray exposure. Plots of SAT and VAT versus body mass show an increase of adipose tissue with greater body mass, however, the interrelationship and the related correlations are less pronounced with Protocol-II (Protocol-I vs Protocol-II; SAT: r^2 = 0.745 vs 0.248 and VAT: r^2=0.606 vs 0.439).

3.4 Reliability of BMIest

Additional correlation analysis was conducted to test the reliability of the formula used to calculate BMIest. The correlation between BMIest and real BMI data is displayed in Figure 7 for both male (r^2=0.78 with p<0.001) and female patients (r^2=0.841 with p<0.001) using data obtained by Protocol-I.

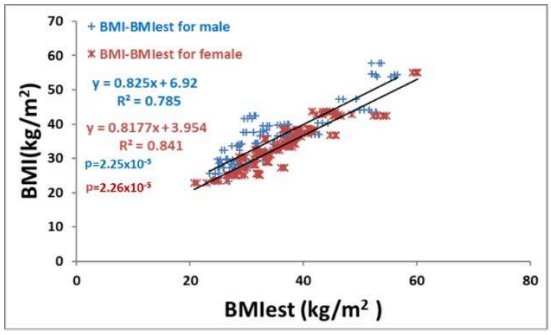
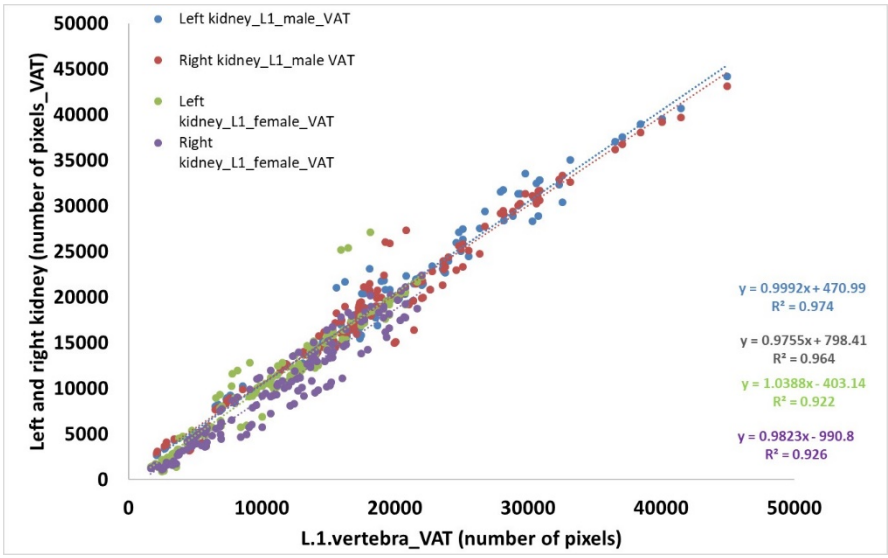


Figure 7
Correlation plot between BMI and BMIest. The resulted correlation parameters (equation, R^2 and p) are also presented.

3.5 Effect of Segmentation Levels

In the study, we also examined the segmentation differences due to the three selected regions (L.1 vertebra, left and right kidney). The results are illustrated in Figures 8 and 9. High correlations are shown at both fat tissues by the large r^2 value (Figs. 8 and 9) with a definite positive intersection at the Y axes.



The interrelationship between the VAT values originated from the L.1vertebra and the left and the right kidney

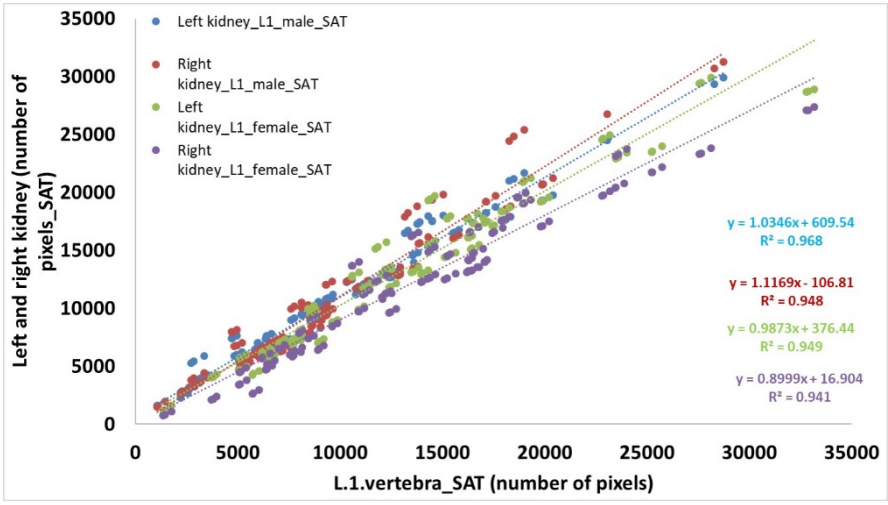


Figure 9
Correlation between the SAT values selected from L1 vertebra and left and right kidney

Hypothesis tests were also performed to compare the AT values obtained from different axial regions (Table 3).

Table 3
Summary of paired hypothesis test for subcutaneous and visceral adipose tissue data from different anatomical ranges. If p is larger than 0.05, the value is presented in red color

	vertebra, left kidney	vertebra, right kidney	left kidney, right kidney
male VAT1	1.82E-02	2.45E-02	2.69E-01
male VAT2	3.18E-02	2.19E-02	4.59E-01
male VAT3	2.83E-02	2.63E-02	3.03E-01
female, VAT1	1.33E-01	2.52E-06	2.30E-06
female, VAT2	1.74E-01	4.66E-06	4.82E-06
female, VAT3	2.05E-01	4.85E-06	3.52E-06
	vertebra, left kidney	**vertebra, right kidney**	**left kidney, right kidney**
male SAT1	7.06E-08	4.57E-06	8.52E-01
male SAT2	5.95E-08	4.18E-06	8.62E-01
male SAT3	3.75E-08	3.82E-06	8.11E-01
female, SAT1	3.66E-01	3.23E-06	2.05E-07
female, SAT2	3.48E-01	3.19E-06	1.93E-07
female, SAT3	3.52E-01	2.72E-06	1.82E-07

3.6 BMI Estimation by Models

In this section, we show the BMIs calculated by the 9 different models and then compare them. Models are shown in Table 1. The constants of models (a, b, c, d, e) are estimated by non-linear regression, and the resulting values are presented in Table 4 for men (n=51) and Table 5 for women (n=47).

Table 4
Model parameters (a, b, c, d, e) are estimated by non-linear regression, and the estimated values are presented. The h is the resulting value of the weight function used in the non-linear regression.

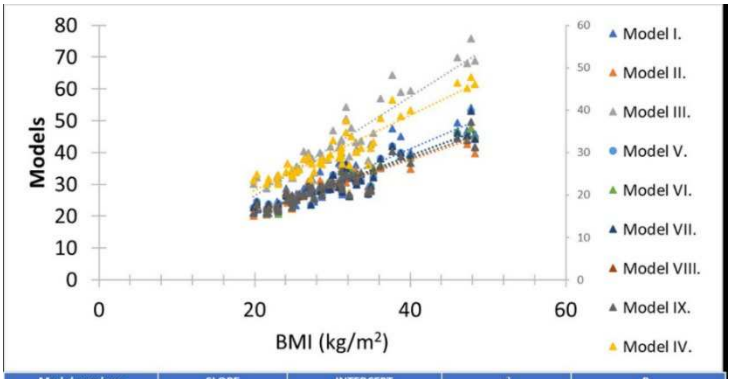
Male constants	*a*	*b*	*c*	*d*	*e*	*h*
Model I	2.069	0.037	0.051	0.985	2.648	125.89
Model II	-18.275	0.010	$3x10^{-5}$	1.116	-1.505	92.72
Model III	-9.163	0.253	10.621	0.081	0.597	616.80
Model IV	-0.26	0.114	$2.9x10^{-4}$	0.021	0.489	97.30
Model V	-3.22	0.249	-	-	-	103.23
Model VI	-15.799	0.911	0.395	-	-	80.46
Model VII	-3.19	0.248	$2.8x10^{-4}$	-	-	103.23
Model VIII	-12.75	0.844	$3.3x10^{-4}$	0.092	-	85.04
Model IX	-13.42	0.900	$3x10^{-4}$	0.085	-0.007	85.04

Table 5
Parameters of female models (a, b, c, d, e) are estimated by non-linear regression, and the estimated values are presented. The h stands for the resulting value of the weight function

Female constants	a	b	c	d	e	h
Model I	-9.163	0.253	10.621	0.081	0.597	390.14
Model II	-5.087	0.187	11.525	0.055	0.390	72.95
Model III	2.069	0.037	0.051	0.985	2.648	91.68
Model IV	0.110	0.014	0.045	0.953	1.114	65.26
Model V	-9.680	0.324	-	-	-	79.93
Model VI	-12.38	1.002	0.262	-	-	72.56
Model VII	-7.500	0.334	-0.060	-	-	71.23
Model VIII	-12.864	1.129	0.469	-0.082	-	71.48
Model IX	0.002	0.234	0.001	0.208	-0.069	83.65

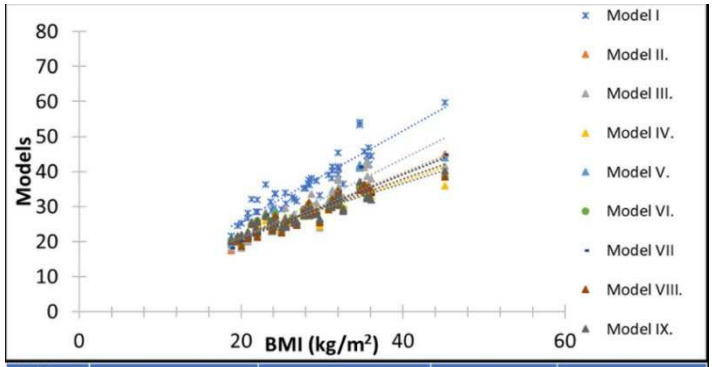
In the last column of both tables, the value of the weight function of the non-linear regression (see Eq. 4) can also be seen. In general, the h value may depend on the number of cases and the goodness of the fitted model, thus we need to consider it when interpreting the results. Fortunately, our two groups contain almost the same number of cases, thus the h value will be comparable between the two gender groups. In the case of Table 4 the highest h value (meaning the worst model) belongs to Model III, which is the original model for the other sex (women). In this case, the h value is 616.8, approximately five times greater than Model I. The best model is Model VI, where only the AP and horizontal diameters are included. There are reduced differences among the remaining models, thus there is no need to use a complicated, multifactorial model for BMI estimation. The h values of women are in a similar range as those of men, which is probably explained by the similar number of patients in both groups (Table 5). The best h value belongs to Model IV (h=65.26) which is based on Model I of man. The worst case is Model I, which is the original model of women, and this conclusion is the same for both sexes. There are minor differences among the resting models, and especially Model VI-VIII are very close to each other because the h value differences are less than 1. Considering all models, the Model VI seems to be the most optimal for both genders because the h values are low and comprise only two distance data (BTD and BAPD).

The results of correlation analysis between the model based and real BMI are presented in Figures 10 and 11, showing the data for men and women.



Model numbers	SLOPE	INTERCEPT	r^2	P
I.	1.043	-0.394	0.838	5.26×10^{-21}
II.	0.821	4.925	0.858	2.2×10^{-22}
III.	1.559	-4.865	0.862	9×10^{-23}
IV.	0.884	3.429	0.863	8.7×10^{-23}
V.	0.914	2.392	0.852	5.42×10^{-22}
VI.	0.923	2.218	0.899	4.51×10^{-26}
VII.	0.913	2.440	0.852	5.42×10^{-22}
VIII.	0.851	4.204	0.882	2.17×10^{-24}
IX.	0.858	3.879	0.882	2.02×10^{-24}

Figure 10
Regression analysis between BMI (kg/m^2) and models in the case of males, the regression data are in the table below the figure



Model numbers	SLOPE	INTERCEPT	r^2	p
I.	1.282	0.341	0.852	2.86×10^{-20}
II.	0.944	1.889	0.853	2.42×10^{-20}
III.	1.147	-2.214	0.846	6.57×10^{-20}
IV.	0.785	5.773	0.867	2.66×10^{-21}
V.	0.901	3.159	0.828	8.6×10^{-19}
VI.	0.813	5.273	0.866	3×10^{-21}
VII.	0.928	2.149	0.853	2.37×10^{-20}
VIII.	0.823	4.963	0.863	4.82×10^{-21}
IX.	0.720	7.852	0.880	2.14×10^{-22}

Figure 11
Regression analysis between BMI (kg/m^2) and models in the case of females; the regression data are in the table below the figure

The points overlap in both figures. Results of regression analysis are similar in all cases because all the regression coefficients are larger than 0.8, and the ranges for women and men are 0.828-0.880 and 0.838-0.899, respectively. In addition, all correlations were significant, which is explained by the small p values (< 0.001) in the hypothesis tests.

Discussion

Determination of the amount and localization of adipose tissue plays an essential role in several diseases [20]. Several study groups have developed methods to assay tissue volume and explore the interrelationship between bariatrics and the development of defined diseases. Davies and co-workers developed an application that can successfully calculate the amount of adipose tissue in polytraumatic patients from routinely used CT scans [6]. These methods may be helpful in intensive care units for COVID-19 patients because obesity is a major risk factor and may result in a worse prognosis, mostly among young patients [4] [26]. Sala and co-workers investigated the correlation between adipose tissue distribution and cardiometabolic diseases using adipose tissue volume and distribution data obtained from CT [22]. They applied an automatic tool that could recognize fat tissue and set the Hounsfield units of fat between -150 and -30 HU and the axial slices were selected from the level of the 5th lumbar vertebra.

After the literature review, we did not find gold standard for the segmentation method. The accuracy of the segmented areas could only be verified quantitatively using phantoms, where the volume of adipose tissue or adipose tissue equivalent material in the phantom is known. Such a study was performed by Yoon et al. where they compared phantom and human CT and MR scans [9], however, in this study the main focus was the effect of varying tube current on segmentation. However, our results definitely show that image quality (influenced by the scan setup parameters), the chosen HU range and the segmentation region affect the final result. We applied the manual delineating technique and three previously published HU ranges to segment the visceral and subcutaneous areas in this study. Usually, part of the lumbal spine is the base for the adipose tissue determination, however some groups choose the level of the umbilicus [24]. In our study, adipose areas were measured on each slice acquired at the level of L1, hilar region of left and right kidney. We chose the hila of the kidneys because these are usually located between L1 and L3. The volume of subcutaneous and visceral adipose tissue was defined by these HU ranges using two different X-ray tube currents. We found that the volume of adipose tissue obtained with the higher X-ray tube current (Protocol-I) by segmentation based on different HU ranges was in good correlation ($r^2 \geq 0.998$) for both male and female subjects. Remarkable differences were detected between images and correlation analysis data applying different protocols. Images obtained with a lower (≈ 50 % lower) tube current were less reliable, and the correlations between the adipose tissue volume data segmented using different segmentation procedures were not as close as between those obtained with a larger tube current ($r^2 > 0.967$). Troschel and co-workers showed a similar tendency [3] changing the

tube current–exposure time product values in their work. In Table 2, we presented the p values for the hypothesis test, which investigated the possible effect between the segmented AT areas and predefined HU ranges at each possible region (L1 vertebra, right and left kidney level). It is visible that all p values are less than 0.05, thus it can be stated that the three different HU ranges of the segmentation affect the amount of adipose tissue segmented. This does not contradict the fact that the different AT values correlate well for each HU range, as the correlation equations' axis intercept is not zero (Fig. 4). Thus, the segmented AT values obtained with the three distinct HU ranges are not proportional to each other, only an excellent linear relationship is between them.

We were also curious about how the extent of segmented AT pixels is affected by the preselected anatomical areas (L1 vertebra, left and right kidney). The results are presented in Figures 8 and 9. Good correlations and linear relationships could be observed among each ATs, with non-zero intersections of the linear equations in every analysis. The results of the associated hypothesis tests are demonstrated in Table 3, where the p value was larger than 0.05 the text was colored in red. In the case of male and both adipose tissues, there were no significant differences between the right and left kidney level (VAT: $p>0.269$, $p>0.4591$, $p>0.303$, SAT: $p>0.852$, 0.862, 0.811), while in the case of female the vertebra and left kidney region gave same result ($p>0.133$, $p>0.174$, $p>0.205$, SAT: 0.3663, 0.348, 0.352. In all other cases, there is a difference between the segmented SAT and VAT values.

In the last step, we examined the relationship between the amount of segmented adipose tissues and the measured and estimated body shape BMI parameters. The highest correlation between segmented volume and BMIest was found with $r^2 = 0.69$ for male and female data. Comparison of the same parameters using a high and a low tube current protocol resulted in significant dissimilarities. The correlation between SAT and BMIest was characterized with $r^2 = 0.7445$ using Protocol-I and $r^2 = 0.2477$ using Protocol-II. Correlation coefficients between segmented adipose tissue and BMI or BMIest are closer if we use images obtained by the higher tube current. Studying CV data of SAT and VAT data revealed that the volume of both kinds of adipose tissue can be estimated with a $CV \leq 10$ % if the axial slice is chosen from a 5 cm wide area of the lumbal region.

CT-based determination of body mass index can be useful in two cases. One is in clinical practice, when the patient is unconscious and height and weight data are unavailable. The other is in research where no BMI or other anthropometric data are available. The latter was the case in an article by O'Leary and colleagues published in 2012. In their research, they performed adipose tissue segmentation in patients with severe acute pancreatitis without anthropometric data, so they used the BMI estimator model created by Geraghty and colleagues in 2003. In Geraghty's study, in addition to BMI, height, weight and body surface area were estimated using a single abdominal CT image.

Analyzing the nine proposed BMI models, we found that the calculated BMIest values agreed with the real BMI data (males: $r^2 = 0.7802$, females: $r^2 = 0.841$), and 7 out of the 9 models gave excellent accuracy to estimate the body shape. In addition, we found that having these seven models, the simpler ones (V, VI and VII) gave similar reliability to the original equations (I and III), so it follows that there is no need to use a complex, multifactorial model for BMI estimation.

There are some limitations to our study. First, based on the previous article, we selected the slices including the L1 [1] [8] and hila of the kidneys for adipose tissue segmentation. At the same time, other authors prefer L2-L3 or the level of umbilicus as reference [21, 24, 32]. The second limitation was the relatively few numbers of patients involved in Protocol-II. Third, we did not utilize an automatic method for segmentation, a better way could be to develop an algorithm to reduce the adipose tissue segmentation time. However, the segmentation time was not a critical aspect of our study. Finally, our study did not investigate the effects of other setting parameters on segmentation, such as the reconstruction kernel or slice thickness.

Conclusions

As three specific objectives have been identified in the introduction, we can state the following:

i) In our study, we found a good correlation between the three HU ranges, without proportionality, thus, the segmented SAT and VAT areas are actually different.

ii) The chosen anatomical level of segmentation significantly affects the VAT and SAT values.

iii) For the 98 patients, model-based BMI and real BMI were determined and compared.

We have simplified the model-based BMI estimation and as a result, we have shown that a complex multivariate model is not necessary, for CT-based BMI calculations.

Funding

Project no. TKP2021-NKTA-34 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme.

References

[1] A. A. Perez, P. J. Pickhardt, D. C.Elton et al. Fully automated CT imaging biomarkers of bone, muscle, and fat: correcting for the effect of intravenous contrast. Abdominal Radiology. 2020; 48:1229-1235

[2] A. M. Sironi, A. Gastaldelli, A. Mari et al. Visceral fat in hypertension: influence on insulin resistance and beta-cell function. Hypertension. 2004;44(2):127-33

[3] A. S. Troschel, F. M. Troschel, G. Fuchs et al. Significance of Acquisition Parameters for Adipose Tissue Segmentation on CT Images. AJR AM J Roentgenol. 2021; 217(1):177-185

[4] A. Tolonen, T. Pakarinen, A. Sassi et al. Methodology, clinical applications, and future directions of body composition analysis using computed tomography (CT) images: A review. European Journal of Radiology. 2021;145, 109943

[5] C. S. Fox, J. M. Massaro, U. Hoffmann et al. Abdominal visceral and subcutaneous adipose tissue compartments: Association with metabolic risk factors in the Framingham heart study. Circulation. 2007;116(1):39-48

[6] D. F. Ferguson, B. J. Busenlehner, M. D. Rahm et al. The Use of Routine Thoracoabdominal CT Scans in the Polytrauma Patient to Estimate Obesity. Obesity. 2013;21(5):997-1003

[7] D. M. Thomas, C. Bredlau, A. Bosy-Westphal et al. Relationships between body roundness with body fat and visceral adipose tissue emerging from a new geometrical model. Obesity. 2013;21(11):2264-71

[8] D. P. O'Leary, D. C'Neill, P. McLaughlin et al. Effects of abdominal fat distribution parameters on severity of acute pancreatitis. World Journal of Surgery. 2012;36(7):1679-85

[9] D. Y. Yoon, J. H. Moon, H. K. Kim et al. Comparison of Low-Dose CT and MR for Measurement of Intra-Abdominal Adipose Tissue: a Phantom and Human Study. Academic Radiology: 2008;62-70

[10] E. M. Geraghty, J. M. Boone. Determination of Height, Weight, Body Mass Index, and Body Surface Area with a Single Abdominal CT image. Radiology. 2003;(228):857-63

[11] H. Fang, E. Berg, X. Cheng, Wei S. How to best assess abdominal obesity. Current Opinion in Clinical Nutrition and Metabolic Care. 2018;21(5):360-5

[12] H. H. Hua, H. E. Kan. Quantitative proton MR techniques for measuring fat. NMR in Biomedicine. 2013;26(12):1609-29

[13] I. Janssen, S. B. Heymsfield, D. B. Allison et al. Body mass index and waist circumference independently contribute to the prediction of nonabdominal, abdominal subcutaneous, and visceral fat. American Journal of Clinical Nutrition. 2002;75(4):683-8

[14] J. Kullberg, A. Hedström, J. Brandberg et al. Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. Scientific reports. 2017;(November 2016):1-11

[15] J. LaForgia, J. Dollman, M. J. Dale et al. Validation of DXA body composition estimates in obese men and women. Obesity (Silver Spring).

2009;17(4):821-6, Available from: http://www.ncbi.nlm.nih.gov/pubmed/19131939

[16] J. L. Kuk, T. S. Church, S. N. Blair, R. Ross. Does measurement site for visceral and abdominal subcutaneous adipose tissue alter associations with the metabolic syndrome? Diabetes Care. 2006;29(3):679-84

[17] J. Machann, A. Horstmann, M. Born et al. Diagnostic imaging in obesity. Best Practice Research Clinical Endocrinology and Metabolism. 2013;27(2):261-77, Available from: http://dx.doi.org/10.1016/j.beem.2013.02.003

[18] J. Sun, B. Xu, J. Freeland-Graves. Automated quantification of abdominal adiposity by magnetic resonance imaging. American Journal of Human Biology. 2016;28(6):757-66

[19] K. J. Rosenquist, A. Pedley, J. M. Massaro et al. Visceral and subcutaneous fat quality and cardiometabolic risk. JACC Cardiovascular Imaging. 2013;6(7):762-71

[20] L. E. Wagenknecht, C. D. Langefeld, A. L. Scherzinger et al. Insulin sensitivity, insulin secretion, and abdominal fat: the Insulin Resistance Atherosclerosis Study (IRAS) Family Study. Diabetes. 2003;52(10):2490-6. Available from: http://diabetes.diabetesjournals.org/content/52/10/2490.full.pdf

[21] M. Hilmi, A. Jouinot, R. Burns et al. Body composition and sarcopenia: The next-generation of personalised oncology and pharmacology? Pharmacology and Therapeutics. 2019; 196:135-59, Available from: https://www.sciencedirect.com/science/article/abs/pii/S0163725818302195?via%3Dihub

[22] M. L. Sala, B. Röell, N. van der Bijl et al. Genetically determined prospect to become long-lived is associated with less abdominal fat and in particular less abdominal visceral fat in men. Age Ageing. 2015;44(4):713-7

[23] M. Nemoto, T. Yeernuer, Y. Masutani et al. Development of automatic visceral fat volume calculation software for CT volume data. Journal of Obesity. 2014: 495084

[24] M. S. Kim, Y.-J. Choi, Y. H. Lee. Visceral fat measured by computed tomography and the risk of breast cancer. Translation Cancer Research. 2019;8(2):1939-49

[25] M. S. Middleton, W. Haufe, J. Hooker et al. Quantifying Abdominal Adipose Tissue and Thigh Muscle Volume and Hepatic Proton Density Fat Fraction: Repeatability and Accuracy of an MR Imaging-based, Semiautomated Analysis Method. Radiology. 2017;283(2):438-49, Available from: http://pubs.rsna.org/doi/10.1148/radiol.2017160606

[26] M. Watanabe, D. Caruso, D. Tuccinardi et al. Visceral fat shows the strongest association with the need of intensive care in patients with COVID-19, Metabolism. 2020; 111:154319

[27] N. Linder, S. Michel, T. Eggebrecht et al. Estimation of abdominal subcutaneous fat volume of obese adults from single-slice MRI data – Regression coefficients and agreement. European Journal of Radiology. 2020;130(December 2019):1-7

[28] P. Decazes, A. Rouquette, A. Chetrit et al. Automatic Measurement of the Total Visceral Adipose Tissue from Compute Tomography Images by Using a Multi-Atlas Segmentation Method. Journal of Computer Assisted Tomography. 2018;42(1):139-45

[29] R. L. R. Novais, A. C. C. Café ,A. A. Morais et al. Intra-abdominal fat measurement by ultrasonography: association with anthropometry and metabolic syndrome in adolescents. Jornal Pediatria (Rio J). 2019;95(3):342-9, Available from: https://doi.org/10.1016/j.jped.2018.03.004

[30] R. P. Stolk, O. Wink, P. M. J. Zelissen et al. Validity and reproducibility of ultrasonography for the measurement of intra-abdominal adipose tissue. International Journal of Obesity. 2001;25(9):1346-51

[31] S. Cecchini, E. Cavazzini, F. Marchesi et al. Computed tomography volumetric fat parameters versus body mass index for predicting short-term outcomes of colon surgery. World Journal of Surgery. 2011;35(2):415-23

[32] S. J Lee, J. Liu, J. Yao al. Fully automated segmentation and quantification of visceral and subcutaneous fat at abdominal CT: application to a longitudinal adult screening cohort. British Journal of Radiology. 2018;(March):4-9

[33] S. M. Grundy., I. J. Neeland, A. T. Turer, G. L. Vega. Waist Circumference as Measure of Abdominal Fat Compartments, J Obesity, J Obes [Internet] 2013; e454285, Available from: http://www.hindawi.com/journals/jobe/2013/454285/abs/

[34] W. O. Grauer, A. A. Moss, C. E. Cann, H. I. Goldberg. Quantification of body fat distribution in the abdomen using computed tomography. The American Journal of Clinical Nutrition. 1984;39(4):631-7

[35] X. Cheng, Y. Zhang, C. Wang et al. The optimal anatomic site for a single slice to estimate the total volume of visceral adipose tissue by using the quantitative computed tomography (QCT) in Chinese population. European Journal Clinical Nutrition. 2018;72(11):1567-75

Examining the Efficacy of Lower Extremity Exoskeletons in the Rehabilitation Process of Spinal Cord Injury Patients – Case Studies

Benjamin Shenker[1], Melinda Fehér[1], Petra Erdősi[1], Leila Papp[1], Kitti Vadai[1], Gergely Hrivnák[1], Luca Tóth[2], Péter Maróti[3], Csaba Horváth[4], Péter Cserháti[1,2]

[1]National Institute of Locomotor Diseases and Disabilities – National Institute of Medical Rehabilitation, Szanatórium utca 19, 1528 Budapest, Hungary, e-mail: shenkerb@revitarendelo.hu; Feher.Melinda@omint.hu; Erdosi.Petra@omint.hu; Papp.Leila@omint.hu; Vadai.Kitti@omint.hu; Hrivnak.Gergely@omint.hu; Cserhati.Peter@omint.hu

[2]University of Pécs, Medical School, Department of Neurosurgery, Rét u. 2, 7623 Pecs, Hungary, e-mail: toth.luca@pte.hu

[3]University of Pécs, Medical School, Medical Skills Education and Innovation Centre, Szigeti u. 12, 7624 Pecs, Hungary; e-mail: peter.maroti@aok.pte.hu

[4]Semmelweis University, Faculty of Medicine, Department of Internal Medicine and Oncology, Korányi Sándor utca 2/a, 1083 Budapest, Hungary, e-mail: horvath.csaba@med.semmelweis-univ.hu

Abstract: Patients suffering from spinal cord injury (SCI) have shown a growing number in recent times. The effects of SCI have severe consequences, not only to the individual but their family and close surroundings too. The secondary medical issues related to SCIs have further negative effects, thus the rehabilitation of locomotion is becoming a key element for patients with SCI. Numerous devices are being tested and used to assist locomotion and gait therapy, one of them being exoskeletons. Lower extremity exoskeletons (LEE) have shown growing interest within health professionals, industrial and military stakeholders and of course SCI patients. This article documents 2 SCI patients using a LEE for gait training. Firstly, a literature review demonstrates the scientific relevance of the topic. Secondly, presentation of the study. Thirdly, the demonstration of the case studies, changes of the 2 SCI patients' functional and physiological parameters during the process. The aim of the study is to show the effects and efficacy of intensive gait training with LEE of SCI patients.

Keywords: Medical rehabilitation; Lower extremity exoskeleton; Spinal cord injuries

1 Introduction

The field of medical rehabilitation is beginning to involve automated technologies at a growing rate, with the social robotics sector seemingly growing faster than other areas [1]. Exoskeletons are revolutionizing medical rehabilitation with biomechanics, robotics, and healthcare. They help individuals with mobility impairments regain independence and functional mobility. This study explores exoskeleton's historical evolution, biomechanical foundations, technological advancements, clinical efficacy, patient outcomes, practical implementation challenges, and knowledge gaps in medical rehabilitation.

1.1 Historical Trajectory: Pioneering Visionaries to Real-World Applications

The concept of exoskeletons dates back to 1890 with Nicholas Yagn's patent "Apparatus for facilitating walking, running, and jumping" [2]. The Hardiman exoskeleton developed by General Electric in the 1960s aimed to enhance industrial labor, but technical limitations such as power supply constraints and complex controls limited practical applications [3]. The 21st Century saw advancements in materials, robotics, and control systems, leading to the development of the Berkeley Lower Extremity Exoskeleton (BLEEX) by Kazerooni et al. BLEEX bridged the gap between concept and reality [4]. Today, exoskeletons are revolutionizing medical rehabilitation and transforming the landscape of mobility challenges.

1.2 Biomechanical Fundamentals: Fusing Human Physiology with Technological Innovation

A nuanced understanding of human biomechanics is crucial for exoskeleton efficacy. Replicating and augmenting natural movements requires a delicate comprehension of joint mechanics, muscle coordination, and gait dynamics. Key biomechanical studies by Winter and Cappellini have shed light on human locomotion [5, 6]. These biomechanical foundations support exoskeletons that seamlessly integrate with the body, mitigate disparities, and enhance comfort.

1.3 Technological Strides: Precision and Adaptation

Exoskeletons have evolved with advancements in materials, actuators, and control systems. Durable yet comfortable materials, such as lightweight alloys and carbon fiber composites have been developed. Various actuation mechanisms are available, including electric motors and hydraulic, and pneumatic systems, which enable exoskeletons to mimic natural joint movements and provide support [7]. Sophisticated control algorithms and sensor integration strategies have been created

for adaptability, allowing exoskeletons to navigate different terrains and respond to user intent [8].

1.4 Clinical Efficacy: Accelerating Recovery and Fostering Independence

Exoskeletons are making a significant shift in medical rehabilitation, offering new methods for recovery and functional augmentation. In neurorehabilitation, exoskeleton-assisted gait training benefits individuals with spinal cord injuries, stroke survivors, and neuromuscular disorders. Esquenazi's work has demonstrated the potential of exoskeletons to hasten recovery and restore ambulation [9]. A meta-analysis by Liu suggests that exoskeleton robotic training improves ambulation recovery in patients with spinal cord injuries, particularly in those with an injury [10]. Exoskeletons are also making a mark in pediatric rehabilitation, as evidenced by devices such as the PediAnklebot, which holds promise for enhancing motor outcomes in children with cerebral palsy [11]. Beyond physical recovery, exoskeletons foster psychological well-being and a renewed sense of accomplishment.

1.5 Measurement of Patient Outcomes and Compliance: A Pathway to Objective Progress

The effectiveness of exoskeletons depends on measuring patient outcomes and adherence to treatment using metrics such as objective assessment, gait parameters, energy expenditure, and functional independence. Real-time observations can be obtained through wearable sensors, as demonstrated by Kim et al [12]. However, the limited number of high-quality studies on exoskeletons makes it difficult to draw general conclusions, as pointed out in the systematic review by Tamburella et al. [13]. Additionally, user feedback and wearable devices are crucial for monitoring compliance, providing valuable data to refine exoskeleton interventions and optimize rehabilitation protocols.

1.6 Patient Outcomes and Compliance: A Holistic Perspective

The successful use of exoskeletons requires understanding patient outcomes and compliance. Research emphasizes the significance of patient satisfaction and acceptance, such as the systematic review by Cumplido-Trasmonte et al. in neurological pathology [14]. Combining objective measurements with patient-reported outcomes provides a comprehensive understanding of the impact of exoskeleton interventions on patients' lives.

1.7 Practical Implementation Challenges: Navigating Real-world Complexities

Exoskeletons in clinical practice face challenges such as ensuring proper fit, customizing settings, and user training to optimize user experience and rehabilitation outcomes. A systematic review by Charette et al. found that the implementation of exoskeleton locomotor training programs for individuals with spinal cord injuries (SCI) is hindered by gaps in knowledge in the context and implementation process domains [15]. In upper limb rehabilitation, traditional methods can be costly, limiting the amount of time patients can spend on exercises. Bauer et al. introduced a hand rehabilitation module with a direct drive and force measurement for patients with spastic hemiplegia [16]. To maximize the potential of exoskeleton technologies, safety considerations, user education, and integration into rehabilitation settings are crucial.

1.8 Knowledge Gaps and Future Exploration

There is significant potential for exoskeletons, but gaps in knowledge remain, particularly in the areas of physiological changes during exoskeleton-assisted walking, long-term effects of exoskeleton training, and the interaction between exoskeletons and the neuromuscular system [17]. Yip et al. found that underground exoskeletons can help individuals with spinal cord injuries maintain healthy lifestyles and improve their quality of life. Additionally, there are secondary health complications associated with prolonged immobilization, and individuals with spinal cord injuries are interested in the potential to reintegrate into the community [18]. To improve rehabilitation protocols and realize the full potential of exoskeleton interventions, it is crucial to address these knowledge gaps.

1.9 Cost-effectiveness Analysis: Balancing Clinical Gains and Economic Realities

Exoskeletons can improve medical rehabilitation cost-effectiveness. Consider long-term financial implications, insurance coverage, and societal benefits for fair access and sustainable integration into healthcare systems. A study by Pinto et al. found that cost-effectiveness of training strategies varies depending on the completeness of SCI. Conventional training was more cost-effective for incomplete SCI, while overground robotic training was more cost-effective for complete SCI [19].

2 A Lower Extremity Exoskeleton Study with SCI Patients

2.1 Summary

In 2018, the University of Pécs won a European Union tender that enabled the purchase of two ReWalk 6.0 lower extremity exoskeletons (LEE) and the establishment of a research project in collaboration with the National Institute of Medical Rehabilitation. The objective of this study was to evaluate the effectiveness of exoskeletons in the rehabilitation of patients with SCI. The investigation also encompassed the indications for the therapeutic use of the exoskeleton, its physiological effects, and its role in ambulatory rehabilitation.

In December 2019, official training of the ReWalk company took place at the National Institute of Medical Rehabilitation. Nine movement therapists and five medical doctors attended and received certificates to operate and use the ReWalk 6.0 lower extremity exoskeleton.

Six patients from the spinal cord injury patient database at the National Institute of Medical Rehabilitation were selected for the study, which was originally planned to commence in the spring of 2020. However, the COVID-19 pandemic forced the postponement of this study. Due to the continuously changing medical protocols related to the pandemic, the study could not be resumed in autumn 2020.

The University of Pécs had already been using a device for gait training of an SCI patient before the start of the study; however, no patients from this institution were included in the study. Therefore, our work exclusively presents case studies of two individuals with SCI at the National Institute of Medical Rehabilitation.

2.2 Objectives and Method

The purpose of this study was to evaluate the effects of gait therapy on certain functional and physiological parameters in patients who had suffered an injury four months prior. To achieve this, we conducted a prospective, controlled trial over a six-month period, during which we registered and followed these parameters in patients who received traditional rehabilitation. The parameters included DEXA scans for determining bone density and body composition, body impedance measurement for analyzing gastrointestinal and urogenital functions, and questionnaires about general well-being and compliance.

2.3 The Device

To ensure the user's safety, comfort, and functionality, the ReWalk 6.0 Personal Exoskeleton was tailored to their body (Figure 1). The battery-powered system includes a wearable exoskeleton that can support its own weight, with motors at the knee and hip joints, and adjustable ankle joints that allow the user to stand up, sit down, and walk in a natural gait pattern. ReWalkers use buttons on a wristband to control the movement of their exoskeletons and make subtle adjustments to their trunk motion during walking. The body of the device shifts repeatedly, generating a series of steps that simulate a natural and functional gait. Customizable ankle joints and software settings allow clinicians to adjust the gait pattern of each patient for efficiency and comfort. The device is designed for use with crutches, and users can navigate the curbs and stairs using wrist-based controls or hip-mounted buttons [20].

Figure 1
ReWalk 6.0. lower extremity exoskeleton

2.4 Inclusion Criteria

The inclusion criteria for the study were established in accordance with the ReWalk Robotics. These criteria included

- spinal cord injury (SCI) with paraplegia or paraparesis below the 4th thoracic vertebrae,
- a minimum of 4 months since the injury,
- functional use of the hand, shoulder complex and upper extremity to use mandatory crutches,
- a hip T-score ≤ -3.5 on a DEXA test,

- an intact skeleton with only stable or stabilized spine injuries and no other unstable fractures,
- the ability to stand with the device similar to Easystand,
- good general health,
- a height of 160-190 cm,
- an upper leg length (hip axis to knee axis) of 36-48.5 cm and a lower leg length (knee axis to bottom of foot) of 43.5-56 cm,
- a maximum body weight of 100 kg,
- and sufficient physiological lower extremity range of motion without contractures.

27 patients were identified as eligible from the SCI patient database at the National Institute of Medical Rehabilitation, of which 6 agreed to participate in the study.

2.5 Protocol

- Each patient was accompanied by at least two certified therapists during sessions.
- The therapy program was divided into four stages, with the first three stages consisting of intensive training five times a week for, 60-90 minutes each. The last stage offered lower-intensity training to maintain safe and functional gait.
- Patient training was scheduled for a period of 6 months.
- Psychologists were available during the study and the follow-up periods to monitor for any psychological issues.

However, due to practical constraints, we found that it was difficult to maintain the original schedule. Five sessions per week proved to be too much time off work for both patients and physiotherapists, so we adjusted to three sessions per week.

2.6 Therapy, Gait Training

The gait training with the ReWalk LEE was divided into 4 phases:

- 1.: Personalized preparatory therapy improves sitting balance and upper extremity strength if needed. This phase also included learning to use the device and transferring it in and out of the wheelchair.
- 2.: Independent application of the device, stand-up and sitting-down maneuvers, and sufficient use of crutches. The patient should be safe in a standing position and able to shift body weight in all directions.

- 3.: Independent walking, ability to use the controller alone. The patient should be able to safely manage starting and stopping walking as desired.
- 4.: Lower intensity, sustaining gait therapy.

2.7 Limitations

The study had several limitations, including:

- Limited access to transport services for SCI patients makes it difficult for them to attend institutes.
- The lack of financial reimbursement and time off work made the initial phase of intensive gait training (5 times a week, 60-90 minutes/session) unmanageable. This was the main reason why the patients did not participate in the study.
- The COVID-19 pandemic and related restrictions on patients, therapists, and regulatory agencies affected the progress of the study.
- Patients' health issues and lack of continuity also presented challenges.

3 Case Studies

Due to the pandemic, the commencement of this study was delayed. However, the National Institute for Medical Rehabilitation successfully mobilized two hospitalized patients, who attempted to use of the device on three separate occasions. Although we were able to use the device with another SCI patient who was originally enrolled in the study, medical adverse events ultimately led to the cessation of further training.

3.1 Case Study 1

K.Á.

- 2017: fracture of 4th and 5th thoracic vertebrae, complete SCI, underwent rehabilitation at the National Institute of Medical Rehabilitation
- 2019 autumn: Conducted an exoskeleton study.
- 2021 autumn: After 30 successful sessions, reached a maximum of 288 steps per session.
- 2022 February: Experienced left leg spasms that restricted gait training with LEE.
- 2022 March: Infected with COVID, suffered from post-COVID symptoms, underwent pulmonology control. Additionally, the patient had

severe toenail and prolonged skin injuries, which prevented further LEE training.

- 2023 May: Reopened the exoskeleton study (Figure 2).

Figure 2
K.Á. exoskeleton gait training

3.2 Case Study 2

A.Á.

- 2022 June: The patient was diagnosed with polytrauma, including bilateral pneumothorax, lung and liver contusion, fractured ribs, an unstable fracture of the vertebral body at the level of the 12th thoracic vertebrae, a fracture of the transverse process of the vertebral column at the level of the 11th thoracic to 2nd lumbar vertebrae, and a fracture of the pelvic bone. Surgical intervention involved spinal stabilization, and vertebral decompression, and laminectomy.
- 2022 July: the patient underwent complex rehabilitation at the National Institute of Medical Rehabilitation, where they learned to use wheelchairs and perform intermittent catheterization independently.
- 2022 September: continues rehabilitation at the National Institute of Medical Rehabilitation
- Status:
    - Sensorium: anaesthesia below the 12th thoracic vertebrae on the right side and below the 11th thoracic vertebrae on the left side of the trunk.

- Motorium: Complete on the upper extremities (UE), no voluntary movements in the lower extremities (LE).
- Vegetativum: daily 5x clean intermittent self-catheterization (CISC), sufficient bowel management
- Muscle tone: moderate spasticity in lower extremities
- Reflexes: UE: normal, LE: increased patellar and foot reflexes, bilateral positive Babinski reflex
- No contractures, sitting posture maintained, self-sufficient.

- 2023 May: the patient participates in an exoskeleton study (Figure 3)

Figure 3
A.Á. exoskeleton gait training

4 Results

4.1 Bone Density

Bone density data were measured using DEXA. All scans were performed by the team of Prof. Dr. Csaba Horváth at Semmelweis University - Department of Internal Medicine and Oncology. A DEXA is a type of medical imaging test. It uses very low levels of X-rays to measure bone density. DEXA stands for “dual-energy X-ray absorptiometry.” Medical experts consider it to be the most useful, quick, and painless test for diagnosing osteoporosis [21].

K.Á.'s BMD measurements for 2nd to 4th lumbar vertebrae showed a 4.6% decrease, a decrease of 1.6% in the left side femur and a 1.1% increase in the left side radius (Table 1).

A.Á.'s datasets recorded a 1.8% increase in BMD values for 2nd to 4th lumbar vertebrae, a 0.9% decrease at the left side of the femur neck, while a 4.7% increase was noted at the left side radius (Table 2).

Table 1
K.Á. osteodensitometry data

Osteodensitometry		**K.Á.**	
		28.03.2023	**04.08.2023**
2nd-4th lumb.vert.	BMD [g/cm^2]	1,558	1,487
2nd-4th lumb.vert.	Z-Score	2,3	1,3
Left side femur neck	BMD [g/cm^2]	1,079	1,062
Left side femur neck	Z-Score	-0,4	-0,8
Hip	BMD [g/cm^2]	0,954	0,966
Hip	Z-Score	-1,5	-1,6
Left side radius	BMD [g/cm^2]	1,035	1,046
Left side radius	Z-Score	0,5	0,6

Table 2
A.Á. osteodensitometry data

Osteodensitometry		**A.Á.**	
		08.03.2023	**04.08.2023**
2nd-4th lumb.vert.	BMD [g/cm^2]	1,049	1,068
2nd-4th lumb.vert.	Z-Score	-1,2	-1,2
Left side femur neck	BMD [g/cm^2]	1,096	1,086
Left side femur neck	Z-Score	0,0	-0,2
Hip	BMD [g/cm^2]	0,872	0,846
Hip	Z-Score	-1,7	-2,0
Left side radius	BMD [g/cm^2]	0,974	1,020
Left side radius	Z-Score	-	-

4.2 Body Composition

Bioelectrical impedance analysis (BIA) is a technique used to determine body composition by measuring the rate at which a painless low-level electrical current travels through the body. Different tissues in the body allow electrical currents to travel at different speeds, with fat being more resistant than muscle or water. Therefore, a higher resistance (impedance) indicates higher body fat percentage. Most BIA scales estimate the total fat, muscle, water, and bone weight and percentage based on this rate and use other data such as height, sex, and weight

measurements to determine the body's fat percentage. To improve the accuracy of the BIA scale, patients were advised to fast overnight before the test, as suggested by the institute's dietician. Other studies have found that a training program using exoskeletons is effective in preventing continuous muscle loss in patients with SCI and reducing body fat to maintain health [22].

K.Á.'s initial BIA was performed at the start of the original study, and the measurements were repeated after a 3-month period of LEE gait training. Overall, his data showed a decline in weight value (2.1%), fat-free mass index value (0.47%), and visceral adipose tissue value (1.2%). An increase in absolute fat mass (1.14%) was observed, whereas skeletal muscle mass decreased by 1.21%.

The body composition measurements of A.Á. showed a lower absolute fat mass value of 0.14%, and an increase in skeletal muscle mass value (0.45%) and weight value (1.6%). However, the FFMI value dropped by 0.47% and the visceral adipose tissue value decreased by 0.32%.

4.3 Gastrointestinal and Urogenital Changes

Urodynamic testing is a diagnostic procedure that evaluates the functioning of the lower urinary tract, including the bladder, sphincters, and urethra, in relation to the storage and release of urine. The focus of these tests is on the bladder's ability to store and empty urine, as well as to monitor bladder contractions [23]. We observed an improvement in bladder compliance and subsequently prescribed antimuscarinic medications. Interestingly, there was no significant change in the patient's stool type, as measured using the Bristol stool form scale.

4.4 Questionnaires Related to the Patients' General Condition and Compliance

Questionnaires were used for both patients, focusing on well-being, functionality, movement, and gastrointestinal functions. The recording dates of the questionnaires were marked as follows unless marked otherwise.

- K.Á.: T1.1.: 27.09.2021; T1.2.: 03.08.2023
- A.Á.: T2.1.: 08.05.2023; T2.2.: 31.07.2023

4.4.1 Trunk Control Test

The Trunk Control Test (TCT) was used to evaluate the trunk movements in patients with neurological disorders. The test was conducted on a bed and consisted of four tasks: rolling to the weak side, rolling to the strong side, maintaining balance in a sitting position on the edge of the bed with both feet off the ground for at least 30 seconds, and sitting up from a lying down position. The TCT score was determined by adding the scores obtained from each of the four tasks (ranging from 0 to 100).

Both patients displayed equal performance at the beginning and end of the three-month period, with a score of 74.

4.4.2 Spinal Cord Independence Measure

The Spinal Cord Independence Measure (SCIM) was created to evaluate the function of patients with spinal cord injuries in three specific areas: self-care (feeding, grooming, bathing, and dressing), respiration and sphincter management, and mobility (bed and transfer abilities both indoors and outdoors). The scores ranged from 0 to 100, with a score of 0 indicating total dependence and a score of 100 indicating complete independence. SCIM can also aid clinicians in setting treatment goals and objectives for patients with SCI [24]. Both patients showed similar test outcomes and changes.

4.4.3 Barthel Index

The Barthel Index for Activities of Daily Living (ADL) is an ordinal scale that measures a person's ability to complete activities of daily living [25]. The guidelines for interpreting Barthel scores are as follows [26].

- Scores of 0-20 indicate "total" dependency
- Scores of 21-60 indicate "severe" dependency
- Scores of 61-90 indicate "moderate" dependency
- Scores of 91-99 indicate "slight" dependency.

K.Á. performed equally at the beginning and at the end of the study. A.Á. reported on minor changes in toilet use and bathing, improved bladder function, and a total score improvement of 1. Both cases indicate moderate dependency.

4.4.4 Berg Balance Scale

The Berg Balance Scale (BBS) is used to measure a patient's ability to balance safely during a series of predetermined tasks. It is a 14-item list, with each item having a five-point ordinal scale ranging from 0 to 4, with 0 indicating the lowest level of function and 4 indicating the highest level of function [27]. A score of

- 0 to 20 indicates that the person will likely need assistance from a wheelchair to move around safely, while a score of
- 21 to 40 indicates that the person will need some type of walking assistance, such as a cane or walker.

In this case, minor improvements were seen in both patients, with K.Á. improving by 1 point (22-›23) and A.Á. improving by 2 points (20-›22).

4.4.5 36-Item Short Form Survey

The 36-Item Short Form Survey (SF-36) is a widely used self-reported measure of health that was originally developed as a tool for assessing quality of life in the Medical Outcomes Study [28]. It comprises 36 questions that cover eight domains of health [29], and scores for each domain range from 0 to 100, with higher scores indicating better health. Scores for the different domains were converted and pooled using a scoring key, and the total score provided an overall measure of quality of life (QOL), ranging from low to high. The total score can be divided into two parts: a physical component summary and a mental component summary. Research suggests that interpreting SF-36 scores can be difficult and should be done in comparison to the overall score or profile [30] and that SF-36 cannot be used as a single index of overall health-related QOL because it measures two dimensions (physical and mental) [31]. In this study, the SF-36 outcome had a lower score on most dimensions of the questionnaire with a minor reduction particularly in the areas of mental health (Table 3).

Table 3
SF-36 scores for K.Á. and A.Á.

SF-36	**K.Á.**		**A.Á.**	
0-100 scores mean 0-100%	**T.1.1.**	**T1.2.**	**T2.1.**	**T2.2.**
physical functioning	35	10	55	55
bodily pain	100	100	90	90
role limitations due to physical health	75	50	25	25
role limitations due to emotional problems	100	0	100	66,7
emotional well-being	92	72	64	60
social functioning	75	37,5	75	75
energy/fatigue	80	50	55	50
general health perceptions	85	75	90	90

4.4.6 WHOQOL-BREF Questionnaire

The WHOQOL-BREF is a 26-item version of the WHOQOL-100 that assesses the quality of life, health, and well-being of people with and without disease as well as health professionals. Each item on the WHOQOL-BREF was scored from 1 to 5 on a response scale and the scores were then converted to a 0-100 scale. A score of 0 represents the worst possible health state, whereas a score of 100 represents the best possible health state in the respective domain. The physical, psychological, social, and environmental health status of the patients was assessed separately [32]. The WHOQOL-BREF showed that K.Á. experienced a 4-point improvement, while A.Á. indicated a 10-point decrease in the total score (Table 4).

Table 4

WHOQOL-BREF Questionnaire scores for K.Á. and A.Á.

WHOQOL-BREF	**K.Á.**		**A.Á.**	
1: very poor; 2: poor; 3: neither poor nor good; 4: good; 5: very good	**T.1.1.**	**T1.2.**	**T2.1.**	**T2.2.**
How would you rate your quality of life?	4	4	4	5
How satisfied are you with your health?	3	4	4	3
To what extent do you feel that physical pain prevents you from doing what you need to do?	1	1	1	3
How much do you need any medical treatment to function in your daily life?	1	1	1	2
How much do you enjoy life?	4	4	3	1
To what extent do you feel your life to be meaningful?	5	4	3	2
How well are you able to concentrate?	4	4	4	3
How safe do you feel in your daily life?	5	5	4	4
How healthy is your physical environment?	5	4	4	4
Do you have enough energy for everyday life?	4	4	5	4
Are you able to accept your bodily appearance?	5	4	3	3
Have you enough money to meet your needs?	4	5	4	4
How available to you is the information that you need in your day-to-day life?	4	5	3	2
To what extent do you have the opportunity for leisure activities?	2	3	4	4
How well are you able to get around?	2	1	5	4
How satisfied are you with your sleep?	3	4	4	3
How satisfied are you with your ability to perform your daily living activities?	4	4	3	3
How satisfied are you with your capacity for work?	4	4	4	2
How satisfied are you with yourself?	4	4	3	2
How satisfied are you with your personal relationships?	4	4	3	2
How satisfied are you with your sex life?	4	4	2	2
How satisfied are you with the support you get from your friends?	4	5	2	2
How satisfied are you with the conditions of your living place?	5	4	5	5
How satisfied are you with your access to health services?	2	4	3	2
How satisfied are you with your transport?	3	4	4	3
How often do you have negative feelings such as blue mood, despair, anxiety, depression?	1	1	3	4
Total	**91**	**95**	**88**	**78**

4.4.7 Beck's Depression Inventory

The Beck Depression Inventory Short Form (BDI-SF) is a 13-item self-report rating scale that measures the characteristic attitudes and symptoms of depression [33]. The scoring system for the BDI-SF was as follows:

- 1-10: These ups and downs are considered normal
- 11-16: Mild mood disturbance
- 17-20: Borderline clinical depression
- 21-30: Moderate depression
- 31-40: Severe depression
- 40 < : Extreme depression [34]

In the case of K.Á., their BDI-SF score remained within the normal range before and after LEE gait training and increased by 3 points (0-›3).

Meanwhile, A.Á.'s evaluation showed an initial score of 9, which increased to a score of 12 by the end of the 3-month trial period. According to the BDI-SF scoring system, A.Á.'s score increased from the normal range to mild mood disturbances (9-›12).

4.4.8 Beck Anxiety Inventory

The Beck Anxiety Inventory (BAI) was used to measure the severity of anxiety. It is a 21-question multiple-choice self-report inventory that focuses on an individual's feelings in the previous week, primarily on somatic symptoms [35]. The total score was calculated by summing up the 21 items.

- Score of 0 – 21 = low anxiety.
- Score of 22 – 35 = moderate anxiety.
- Score of 36 and above = potentially concerning levels of anxiety [36].

During the trial period, both patient scores changed, but they remained within the range of low anxiety (0-21).

4.5 Gait Training Outcomes

The patient's pulse and blood pressure were measured using the exoskeleton before, during, and after each session of the 3-month gait training period. The number of steps taken with the device's assistance was also registered, and any adverse events were documented. K.Á. had 17 sessions, twice a week, each lasting approximately 60 minutes. A maximum of 640 steps were achieved in session 12, with an average of 301 steps performed per session. The peak heart rate was measured at 127 during session 16. The 10 metre walking test time was 32.44 seconds. A.Á. had 14 sessions,

twice a week. The maximum heart rate observed was 112 in session 6 and 12 in session 12, with high blood pressure recorded in sessions 6 (162/95) and 14 (143/93). The patient experienced vertigo and gait training was not possible at those times. The maximum step count was 1727 in session 12, with an average of 480 steps taken during the training period. The 10 metre walking test time was 26.39 seconds.

Conclusions

In conclusion, the body composition data indicated a decrease in weight and visceral adipose tissue for K.Á. due to the long-lasting COVID-19 and post-COVID syndrome. The patient's immobility contributed to a more sedentary lifestyle. Further BIA tests are planned for this patient. A.Á.'s skeletal muscle mass and weight increased, suggesting the benefits of LEE gait training.

Regarding bone density, the changes were minimal due to the brief trial period.

The first stage of hyperactive detrusor syndrome showed improvement in bladder compliance. Both patients experienced an increase in the volume of involuntary detrusor contractions; however, the administration of antimuscarinic medication was justified.

We found no significant differences in the TCT, SCIM, Barthel index, and BBS scores among the patients. This may have contributed to the successful rehabilitation, as sufficient mobility was achieved. However, decreasing scores on the SF-36, WHOQOL-BREF, BDI-SF, and BAI tests suggest a lower mental health state. A.Á.'s data showed an improvement in physical parameters, but a decrease in mental health measures, which may be related to the documented suicidal ideation at the time of SCI.

The patients showed significant improvements in LEE gait training, and their motivation remained high throughout the trial period. Seeing the patients evolve and mastering the basic skills of LEE training is a motivating experience. The next chapter of LEE training involved stair climbing.

The experiences obtained from our work highlight the need to explore patients' and their caregivers' attitudes towards exoskeletons, to determine the minimum number of weekly treatments needed to achieve clinically significant efficacy on the primary and secondary outcomes, and to investigate whether inpatient versus ambulatory care would be a better form of exoskeleton treatment for SCI patients in a given jurisdiction and infrastructure. The study included multiple outcome measures to track patient parameters, such as physiological indicators, mobility, functioning, health status, and quality of life. The connection between digitally measured parameters and patient-reported outcomes has not been thoroughly explored. An important point to consider is which outcomes should be recorded at which intervals, considering the measurement properties of the measurement tools, the human/technical resource needs for a comprehensive measurement, and patients' tolerance for a comprehensive measurement. Identifying highly relevant

outcomes for patients and caregivers and developing a standardized set of primary and secondary outcomes for clinical trials and practice are crucial. This set can then be customized according to the specific study requirements.

The case studies demonstrated the potential for SCI patients to use robotic rehabilitation devices independently, and the indirect success of the project was the modification of the social security funding regulation for national health care [37]. New therapeutic modalities, including ambulatory gait training with LEE have been accepted, moving Hungary closer to countries that use robotic technology in their therapeutic practice.

Future plans

Our future plans involve exploring the adaptability of LEE gait training in the rehabilitation process of patients with SCI with the aim of including more patients and identifying more effective evaluation methods. This also allows us to examine the financial viability and feasibility of implementing such devices.

Acknowledgement

We extend our gratitude to the staff at the University of Pécs and the National Institute of Locomotor Diseases and Disabilities – National Institute of Medical Rehabilitation for their ongoing support and assistance throughout this study. Their professional expertise, initial contributions, and coordination have been invaluable for the success of this research.

This work was supported by the Project no. TKP2021-NVA-06 with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NVA funding scheme.

References

[1] T. Haidegger, G. S. Virk, C. Herman, R. Bostelman, P. Galambos, Gy. Györök & I. J. Rudas: Industrial and Medical Cyber-Physical Systems: Tackling User Requirements and Challenges in Robotics. Recent Advances in Intelligent Engineering : Volume Dedicated to Imre J. Rudas' Seventieth Birthday Cham, Switzerland : Springer-Verlag (2020) 308 p. pp. 253-277. Paper: Chapter 13 , 25 p.

[2] N. Yagn: "Apparatus for facilitating walking, running, and jumping", U.S. Patents 420 179 and 438 830, 1890

[3] C-J Yang, J-F Zhang,Y Chen,Y-M Dong, & Y Zhang: "A review of exoskeleton-type systems and their key technologies. Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science 1989-1996 (vols 203-210) 222(8):1599-1612. August 2008

[4] Kazerooni H., Racine J. L., & Huang L. (2005) On the control of the Berkeley lower extremity exoskeleton (BLEEX). Proceedings-IEEE International Conference on Robotics and Automation, 2005:4353-4360

[5] Winter D. A. (2009) Biomechanics and motor control of human movement. John Wiley & Sons

[6] Cappellini G., Ivanenko Y. P., Poppele R. E., & Lacquaniti F. (2006) Motor patterns in human walking and running. Journal of Neurophysiology, 95(6), 3426-3437

[7] Gregory S Sawicki, Antoinette Domingo, Daniel P Ferris (2006) The effects of powered ankle-foot orthoses on joint kinematics and muscle activation during walking in individuals with incomplete spinal cord injury. Journal of NeuroEngineering and Rehabilitation, 2006; 3:3

[8] Galle S., Malcolm P., Derave W., & De Clercq D. (2013) Adaptation to walking with an exoskeleton that assists ankle extension. Gait & Posture, 38(3), 495-499

[9] Esquenazi A., Talaty M., Packel A., & Saulino M. (2012) The ReWalk powered exoskeleton to restore ambulatory function to individuals with thoracic-level motor-complete spinal cord injury. American Journal of Physical Medicine & Rehabilitation, 2012 Nov;91(11):911-21

[10] Wentan Liu, Jianer Chen (2023) The efficacy of exoskeleton robotic training on ambulation recovery in patients with spinal cord injury: A meta-analysis. Journal of Spinal Cord Medicine, 3;1-10

[11] Colombo G., Joerg M., Schreier R., & Dietz V. (2000) Treadmill training of paraplegic patients using a robotic orthosis. Journal of Rehabilitation Research & Development, 2000 Nov-Dec;37(6):693-700

[12] Grace J Kim, Avinash Parnandi, Sharon Eva, Heidi Schambra (2022) The use of wearable sensors to assess and treat the upper extremity after stroke: a scoping review. Disability and Rehabilitation. 2022 Oct;44(20):6119-6138

[13] Federica Tamburella, Matteo Lorusso, Marco Tramontano, Silvia Fadlun, Marcella Masciullo, Giorgio Scivoletto (2022) Overground robotic training effects on walking and secondary health conditions in individuals with spinal cord injury: systematic review. Journal of NeuroEngineering and Rehabilitation, 19(1):27

[14] C. Cumplido-Trasmonte, F. Molina-Rueda, G. Puyuelo-Quintana, A. Plaza-Flores, M. Hernández-Melero, E. Barquín-Santos, MA. Destarac-Eguizabal & E. García-Armada (2023) Satisfaction analysis of overground gait exoskeletons in people with neurological pathology: a systematic review. Journal of NeuroEngineering and Rehabilitation, 18(1), 1-12

[15] Caroline Charette, Julien Déry, Andreanne K Blanchette, Céline Faure, François Routhier, Laurent J Bouyer, Marie-Eve Lamontagne (2023) A

systematic review of the determinants of implementation of a locomotor training program using a powered exoskeleton for individuals with a spinal cord injury. Clinical Rehabilitation, 37(8):1119-1138

[16] Márk Ottó Bauer, Máté Benjámin Vizi, Péter Galambos, Tibor Szalay (2021) Direct Drive Hand Exoskeleton for Robot-assisted Post Stroke Rehabilitation. Acta Polytechnica Hungarica 18(5):37-54

[17] Ethier, C., Oby, E. R., Bauman, M. J., & Miller, L. E. (2012) Restoration of grasp following paralysis through brain-controlled stimulation of muscles. Nature, 485(7398), 368-371

[18] Christopher C. H. Yip, Chor-Yin Lam, Kenneth M. C. Cheung, Yat Wa Wong, Paul A. Koljonen (2022) Knowledge Gaps in Biophysical Changes After Powered Robotic Exoskeleton Walking by Individuals With Spinal Cord Injury—A Scoping Review. Frontiers in Neurology, 13: 792295

[19] Daniel Pinto, Allen W Heinemann, Shuo-Hsiu Chang, Susan Charlifue, Edelle C Field-Fote, Catherine L Furbish, Arun Jayaraman, Candace Tefertiller, Heather B Taylor, Dustin D French (2023) Cost-effectiveness analysis of overground robotic training versus conventional locomotor training in people with spinal cord injury. Journal of NeuroEngineering and Rehabilitation, 20(1):10

[20] https://rewalk.com/rewalk-personal-3/(last accessed: 17.08.2023)

[21] https://my.clevelandclinic.org/health/diagnostics/10683-dexa-dxa-scan-bone-density-test (last accessed: 25.08.2023)

[22] Hyuk-Jae Choi, Gyoo-Suk Kim, Jung Hoon Chai, Chang-Yong Ko (2020). Effect of Gait Training Program with Mechanical Exoskeleton on Body Composition of Paraplegics. Journal of Multidisciplinary Healthcare 2020; 13: 1879–1886

[23] https://www.niddk.nih.gov/health-information/diagnostic-tests/urodynamic-testing (last accessed: 25.08.2023)

[24] Catz A, Itzkovich M. Spinal cord independence measure: comprehensive ability to rating scale for the spinal cord lesion. Journal of Rehabilitation Research and Development. 2007;44(1):65-68

[25] Mahoney FI, Barthel DW. Barthel index. Maryland state medical journal. 1965 Feb; 14:61-5

[26] Elite learning. The original Barthel index of ADLs. Available from: https://www.elitecme.com/resource-center/rehabilitation-therapy/the-original-barthel-index-of-adls/ (last accessed: 18.08.2023)

[27] https://www.physio-pedia.com/Berg_Balance_Scale (last accessed: 18.08.2023)

[28] Ware J. SF-36 Health Survey Update. Spine. 2000 Dec 15;25(24): 3130-3139

[29] Ware Jr., J.E. and Sherbourne, C.D. (1992) The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection. Medical Care, 1992 Jun;30(6):473-83

[30] Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res. 2001. 10; 5:395-404

[31] Lins L, Carvalho F. SF-36 total score as a single measure of health-related quality of life: Scoping review. SAGE Open Med. 2016. Published online 2016 Oct 4. Accessed: 16 January 2020

[32] https://www.physio-pedia.com/WHOQOL-BREF (last accessed: 18.08.2023)

[33] Letícia M. Furlanetto, Mauro V. Mendlowicz, J. Romildo Bueno (2005) The validity of the Beck Depression Inventory-Short Form as a screening and diagnostic instrument for moderate and severe depression in medical inpatients. Journal of Affective Disorders. Volume 86, Issue 1, May 2005, Pages 87-91

[34] https://www.ismanet.org/doctoryourspirit/pdfs/Beck-Depression-Inventory-BDI.pdf. (last accessed: 18.08.2023)

[35] Beck, A. T., Epstein, N., Brown, G., & Steer, R. (1988) Beck Anxiety Inventory

[36] Beck, A. T., Epstein, N., Brown, G., Steer, R. A. (1988) An inventory for measuring clinical anxiety: Psychometric properties. Journal of Consulting and Clinical Psychology, 56, 893-897

[37] The decree 33/2023 (VIII. 24.) BM by the Home Secretary regarding the modification of statutes of certain health and social health insurance orders of financial procedures for innovative diagnostic and therapeutic modalities. The modification of decree 9/1993. (IV. 2.) NM regarding questions addressing the financing of healthcare. 1st attachment of decree 33/2023. (VIII. 24.) BM. The part of attachment n.: R1.2 „activity codes for outpatient specialist care" (...) (robotic) gait training with exoskeleton