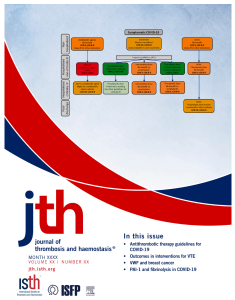
Journal Pre-proof



Artificial Intelligence for Automated Detection of Joint Bleeding via Ultrasound in Hemophilia: Advancing Standardization

Roberta Gualtierotti, Andrea Giachi, Sara Arcudi, Ada Truma, Chiara Suffritti, Marco Colussi, Matteo Manzoni, Dragan Ahmetovic, Salvatore Alessio Angileri, Giampaolo Carrafiello, Sergio Mascetti, Claudio Bettini, Flora Peyvandi

PII: S1538-7836(26)00145-5

DOI: https://doi.org/10.1016/j.jtha.2026.02.025

Reference: JTHA 1442

To appear in: *Journal of Thrombosis and Haemostasis*

Received Date: 1 August 2025

Revised Date: 18 January 2026

Accepted Date: 23 February 2026

Artificial Intelligence for Automated Detection of Joint Bleeding via Ultrasound in Hemophilia: Advancing Standardization

Roberta Gualtierotti[1,2], Andrea Giachi[1], Sara Arcudi[1], Ada Truma[1], Chiara Suffritti[1], Marco Colussi[3], Matteo Manzoni[3], Dragan Ahmetovic[3], Salvatore Alessio Angileri[4], Giampaolo Carrafiello[4,5], Sergio Mascetti[3], Claudio Bettini[3], Flora Peyvandi[1,2]*

1 Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico di Milano, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Milan, Italy
2 Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy
3 Department of Computer Science, Università degli Studi di Milano, Milan, Italy
4 Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico di Milano, Diagnostic and Interventional Radiology Department, Milan, Italy
5 Department of Oncology and Hematology-Oncology, Università degli Studi di Milano, Milan, Italy

Corresponding author:
Flora Peyvandi, MD, PhD
Angelo Bianchi Bonomi Hemophilia and Thrombosis Center,
Fondazione IRCS Ca' Granda Ospedale Maggiore Policlinico,
Department of Pathophysiology and Transplantation,
Università degli Studi di Milano,
Via Pace 9 – 20122 Milano
+390255033228
flora.peyvandi@unimi.it

Running title: AI-based Ultrasound for Joint Bleeds in Hemophilia
Number of words text: 4321
Number of words abstract: 250
figure count: 2
table count: 4
reference count: 35

Abstract

Background. Musculoskeletal ultrasound (US) is a non-invasive tool for joint assessment in persons with hemophilia. Early detection of joint bleeding using a remote US system operated by patients or caregivers and reviewed by Comprehensive Care Centers could improve personalized management. A computer-aided diagnosis (CAD) system for automatic detection of joint effusion may support clinicians in prioritizing interventions.

Objectives. This study aimed to validate a novel CAD system using a deep-learning algorithm to identify joint capsule distension in musculoskeletal US images.

Methods. Longitudinal scans of the subquadricipital recess (SQR) of the knee were collected from people with hemophilia and varying degrees of arthropathy and labeled by an expert. The multi-task learning algorithm was trained to detect the recess and classify images as distended or not.

Results. A total of 8,634 images (2,267 scans) were acquired from 158 adult persons with hemophilia (mean age 44.7 ± 18.6 years) and 66 age-matched healthy controls. After selecting longitudinal SQR images, 814 images were used, of which 711 for training and 103 for testing, ensuring a patient-based split. The model achieved a classification accuracy of 89.2% and a balanced accuracy of 93.9% compared to expert annotations. No significant differences were observed in classification performance between male and female healthy controls, supporting its broader applicability.

Conclusions. The CAD system for automatic detection of joint capsule distension is feasible and reliable. It represents an important step toward telemedicine in hemophilia, enabling early recognition of joint bleeding and supporting personalized, timely therapeutic interventions to prevent further joint damage.

Keywords: hemophilia; joint; ultrasonography; telemedicine; artificial intelligence.

Introduction

Hemophilia A and B are rare, inherited X-linked bleeding disorders caused by a complete or partial deficiency in coagulation factors VIII (FVIII) or IX (FIX). In children and adults with severe hemophilia (i.e. plasma FVIII or FIX levels of <1 U/dL), joint bleeding (hemarthrosis) is the most frequent clinical manifestation (1) and recent data show that it frequently occurs in persons with moderate (plasma factor levels of 1-5 UI/dL) or mild disease (plasma factor levels of >5 UI/dL) as well (2). Repeated episodes of joint bleeding lead to joint remodeling and subsequent hemophilic arthropathy, the targets of which are mainly ankles, knees and elbows (3). It is important that joint bleeding is identified and treated early, as there is evidence that even a single episode may cause cartilage and bone damage (4, 5).

Over the last 30 years, therapeutic advances have progressively improved the life expectancy and quality of life of persons with hemophilia (6). The growing availability of drugs for prophylaxis have changed the scenario of persons with hemophilia as a wider use of prophylactic regimens has led to a better prevention of clinically overt bleeding episodes (7). Therefore, some events may not be recognized by patients as joint bleeding events, thus leading to undertreatment. On the other hand, many painful joint events reported by the patients as joint bleeding are instead acute joint pain recurring in the context of chronic arthropathy (8). Even physical examination by a physician may not be sufficient to diagnose joint capsule distension due to hemorrhagic effusion, since this can be subclinical, or it may occur in an already distended capsule joint due to chronic synovitis (9). In particular, several studies suggest that significant discrepancies exist between point-of-care ultrasonographic (US) findings and patient-reported pain classification as bleeding or other musculoskeletal symptoms (10, 11). Therefore, the current practice of prescribing clotting factor vs. conservative measures based on patient-reported classification of joint pain may lead to an inaccurate management of the acute event with several further complications (8, 12).

Point-of-care US imaging is a useful tool that is growingly gaining interest in modern hemophilia care (13). A remote patient-based MSK-US imaging system may support immediate and accurate diagnosis of acute joint bleeding and prompt effective treatment at home or an evaluation at the Center in unclear cases, thus reducing out-patient visits and hospitalization. In addition, patient empowerment would lead to an improved awareness and adherence to treatment (14). Implementing such a telemedicine system for patients requiring continuous follow-up necessitates the transmission of numerous musculoskeletal ultrasound (MSK-US) images to the physicians at the Care Center for evaluation. This influx of a big number of images can complicate the process of prioritizing interventions, making efficient management and timely decision-making more challenging in this context. Therefore, an artificial intelligence (AI)-based tool for the automatic recognition of joint capsule distension could be the first step to help prioritizing intervention. Recently, we developed in collaboration with the Computer Science Department of University of Milan a computer-aided diagnosis (CAD) tool based on a deep-learning algorithm to automatically recognize joint capsule distension in MSK-US images collected during clinical practice starting with the knee as first joint (15, 16).

The aim of the present study is to evaluate the feasibility and accuracy of the CAD tool in a different dataset of images to validate it as the first step for the development of a telemedicine system in persons with hemophilia.

Methods

Patients

Adult persons with hemophilia A and B of any severity referring to the Angelo Bianchi Bonomi Hemophilia and Thrombosis Center from January 2021 to October 2024 were asked to participate in this observational study providing written informed consent to use pseudonymized MSK-US images collected during the routine visit of the patient at the Center or in the case of an acute event suspected as joint bleeding. Healthy controls of both sexes were enrolled as healthy control group. Approval from the Ethics Committee for the “Joint ultrasound evaluation in persons with hemophilia - JOINEM study” was obtained by the Milan Area 2 Ethics Committee (199_2021bis and subsequent amendment by the CET Lombardy 3, ID 2086 – n° 2086_SA_22.03.2024_P).

Ultrasound protocol

We here describe the US study protocol and scoring system that we followed in the present study. The protocol is based on the European Society of MusculoSkeletal Radiology (ESSR) recommendations for US imaging of the knee for the procedure of acquisition and collection of images (17) and the suggested knee scans of the Haemophilia Early Arthropathy Detection with Ultrasound (HEAD-US) protocol (18). As a scoring system, we used a semi-quantitative 4-level classification of SQR distension (absent, mild, moderate, severe), namely: capsular distension is considered absent when the prefemoral and suprapatellar fat pads are in contact, resulting in non-visualization of the suprapatellar recess (<1 mm); mild when the recess extends in width between 1 and 5 mm; moderate when the recess extends between 6 mm and 5 cm (i.e. it does not exceed the limit of the frame in which the image is included, which in our case corresponds to the width of the Philips Affiniti 50 5-12 MHz ultrasound probe with a setting of 5 cm depth); severe when the capsule distension exceeds the frame in which the image corresponding to the width of the ultrasound probe is included 5-12 MHz Philips Affiniti 50 with a setting of 5 cm depth).

The US examination was performed with a linear array 12-5 MHz US transducer (Philips Affiniti 50, www.usa.philips.com/healthcare/product/HC795208/affiniti-50-ultrasound-system). The patient was asked to lie down on the examination bed keeping the knee at 30° flexion (17). This position allows operators to collect the maximum amount of intra-articular fluid. The US probe was then placed longitudinally along the long axis of the quadriceps tendon in a midsagittal plane with its inferior edge just proximally to the base of the patella. The subquadricipital recess (SQR) is localized cranial to the patella and underneath the quadriceps tendon. In the HEAD-US scoring system, synovitis is systematically assessed by evaluation of the main joint recesses. In fact, other minor recesses rarely distend in isolated form and are usually involved as an extension of the larger ones; therefore, they were excluded from the scanning procedures (18).

Data acquisition

Images were acquired by a single specialized practitioner during routine visits of hemophilic patients (RG). Each image had a resolution of 1024 × 780 pixels and contained acquisition parameters and the actual US scan.

Image acquisition and annotation

The US images were collected over the study on a server dedicated to the task. US images were annotated by two expert US reader (RG, AA) through a web application specifically developed for this purpose (16). In order to preserve patients' privacy, US images were pseudonymized before being stored on the server by removing explicit identifiers (i.e., patients' names and date of birth) and associating the data with a random unique code. The association between the random codes and the patients' names was stored on a separate database, with increased security requirements (*i.e.*, restricted access).

During the annotation process, the practitioners reported for each image the level of distension of the SQR on four different levels: absent, mild, moderate and severe. Also, the Minimum Bounding Rectangle (MBR) of the SQR was annotated by one expert operator (RG). To standardize annotations, we set up criteria for absent, mild, moderate and severe capsule distension (figure 1A-D).

Artificial intelligence algorithm for the detection of subquadricipital recess distension

Artificial intelligence is a term that defines the intelligence demonstrated by machines. Usually, researchers create rules for computers to make decisions. In machine learning, computers learn from themselves and design new rules through practice and repetition. Deep learning in particular is an application of machine learning that imitates the functioning of the human brain. Deep learning networks interpret big data and recognize patterns. The more data they can learn from, the more accurate and informed decisions can be made (19-22). The AI-based CAD was developed in collaboration with the Computer Science Department of Università degli Studi di Milano. Different algorithms were experimented with the aim of identifying the location of the SQR and, at the same time, classify the images as distended or not distended. The algorithm that yields the best results is called multi-task approach and is based on a multi-task framework (15) that is able to address both problems with a single model. Briefly, the multi-task approach has a shared feature extractor followed by two separate neural network branches: one for the detection task and one for the classification, which was pre-trained on COCO (23) dataset to obtain more robust low-level features; this form of transfer-learning is a common practice with small datasets (24). The first branch generates various bounding box candidates. As a post processing step, the algorithm returns only the bounding box with the highest confidence. The branch architecture is based on YOLOv5(25) with a single-class object detection task. The second branch generates a binary label that represents the two different possible classes for the entire image (*i.e.* distended and not distended). The model has been tuned with an evolutionary algorithm to obtain the best hyperparameters, namely, the different augmentation probabilities, the dropout probability and the weights of the multi-task loss.

The multi-task approach was trained with images sized 256x256, using the YOLOv5 augmentation pipeline. The images division into train and test followed a patient-based split approach: all images of each patient are either in the train or test set. Also, the test set contains a single image for each patient. Considering the specific domain, some of the augmentations (like rotation and flip) were not used because they would not generate realistic images. Finally, the model has been trained with an early stop criterion based on balanced accuracy. The evaluation was performed on a set of images from patients excluded from the training set, ensuring the results reflect the model's generalization capability. The test data included two sets of healthy controls: one comprising males and the other females. The measured metrics were balanced accuracy, sensitivity, specificity, and average intersection over union (26, 27).

Statistical Analysis

To avoid data leakage and result overfitting, images from patients used for training were not used also for testing. Consequently, a standard split was applied, with approximately 80% of patients allocated to the training set and 20% to the test set. Only one image per patient was included in the test set. For classification performance, standard metrics were computed, including accuracy, balanced accuracy, sensitivity, and specificity. The classification task was binary (distended vs non-distended) based on the conversion of the original four-level grading of capsule distension. The performance of two different dichotomization strategies was evaluated, and the one combining “absent” and “mild” as non-distended versus “moderate” and “severe” as distended yielded the best results. The weighted Cohen’s κ coefficient was calculated to assess inter-rater agreement between expert annotators. For validation on healthy controls, the Chi-squared goodness-of-fit test was used to estimate appropriate sample size ($\alpha = 0.05$, power = 0.8, effect size = 0.5), and Fisher’s exact test was used to compare classification performance across sex groups. All statistical analyses were performed using Python (version 3.12.6), primarily with the SciPy library (version 1.15) (https://www.python.org/doc/versions/).

Results

Patients

One hundred fifty-eight consecutive male persons with severe hemophilia A and B (71% severe A, 13% moderate A, 6% mild A, 7% severe B, 2% moderate B, 1% mild B), aged 44 ± 15 years, were recruited during annual check-up visit or examination for a suspected intra-articular knee bleeding, subsequently confirmed by response to treatment with factor replacement and ultrasound follow-up. Sixty-six age-matched healthy subjects served as controls (age 48 ± 16, ns).

Ultrasound imaging dataset

From this cohort, a total of 8,634 musculoskeletal US images were collected, of which 2,267 knee scans (multiple frames of SQR, medial and lateral parapatellar recess, medial collateral ligament and femoral trochlea following the HEAD-US scoring system). Of these, 14 scans involved knee prostheses and were excluded. Only SQR longitudinal scans were included, resulting in 814 valid images that were annotated on four different levels: absent, mild, moderate and severe (Table 2). Examples of images with different annotations are shown in Figure 1A-D. The weighted Cohen's κ between the two operators was 0.47, indicating a moderate agreement, while re-mapping the annotations on a binary task (distended vs non-distended), the Cohen's κ reached 0.62, showing a substantial agreement.

Among the knee scans, joint capsule distension was present in 340 out of 814 (41.2%), of these, 42 due to intra-articular bleeding (12.4%), 80 due to synovial thickening (23.5%) and 218 due to the presence of simple synovial fluid effusion (64.2%).

The number of images used to train and to test the system are reported in Table 1.

Thirty-seven of the test images were used to validate the system on persons with hemophilia and the distribution of labels for capsule distension was the following: absent in 24, mild in 9, moderate in 3, severe in one case. The moderate and severe cases were intra-articular bleeding.

Subquadricipital recess detection

Consistently with the previous evaluation proposed in the literature (15), the detection ability of the multi-task model shows that the average IoU is 0.64. The SQR is detected with an IoU greater than 0.5, in 80% of the images. Intuitively, in this case the SQR is correctly detected, as shown in Figure 2A. In some cases, also when the IoU is less than 0.5, the recess is still correctly identified but with a minor precision in size and position (see the example in Figure 2B). In a few cases, when the IoU is less than 0.5, the technique fails to identify the SQR (see the example in Figure 2C).

Classification results

To reduce the complexity of the distension detection task ("distended" vs "non-distended"), we transformed the original four-level annotation system into binary annotations, evaluating two distinct approaches. In the first one, we grouped the images annotated as absent capsule distension in the "non-distended" category *vs* all other levels of distension. The second approach combined those images annotated as absent and mild capsule distension into the "non-distended" category, and moderate and severe capsule distension into the "distended" category.

Results show that the second approach provides more accurate results, the balanced accuracy in this case is 93.9% for the male patients (Table 4), compared to 74.5% of the first approach (Table 3). Compared to other models we experimented, the multi-task approach has a higher sensitivity (15). This is particularly relevant because, in the given domain, the algorithm can be used as a filtering technique and hence it is important to generate less false negative cases.

System validation

The system was validated in a dataset of healthy controls as well: 33 healthy male subjects and 33 healthy female subjects, to ensure validity in both sexes. Data numerosity was established based on the Chi squared goodness of fit test, assuming the following parameters: $\alpha = 0.05$, power (1-ß) = 0.8, effect size = 0.5 (large). The test indicated that a sample size of at least 32 samples per category is needed. The distribution of labels for capsule distension in male healthy subjects was the following: absent in 23, mild in six, moderate in three, severe in one case. In female healthy subjects the distribution was the following: absent in 16, moderate in four, severe in none of the patients.

We found no statistically significant difference in the classification ability (Fisher's Exact Test for Count Data: p-value = 0.18) when separating absent capsule distension vs. mild/moderate/severe capsule distension. Similarly, we found no statistically significant difference in classification ability (Fisher's Exact Test for Count Data: p-value = 0.45) when the classification separated absent or mild capsule distension *vs* moderate or severe capsule distension. Results are reported in Table3 and Table 4.

Discussion

Point-of-care US imaging is gaining growing interest among experts of hemophilia as a simple and widespread tool to evaluate joint health status and to diagnose joint bleeding.

The present study is the first step in a larger project to build a telemedicine system to ensure early diagnosis and personalized treatment of acute joint bleeding with the aim of preventing joint damage occurrence or progression (https://ecare.unimi.it/pilots/practice) (16). In the complete version of our telemedicine system, patients, caregivers, or general practitioners would perform the US scan remotely, with a portable probe connected to a tablet or smartphone device and transmit images to the Comprehensive Care Center. To prioritize intervention when a huge number of images is sent to the Center, a CAD tool was designed in collaboration with the Computer Science Department of Università degli Studi di Milano (15). If joint capsule distension is identified and joint bleeding is suspected - also based on the patient's history - physicians from the Center recommend immediate self-administration of clotting factor. The patient is then invited to attend the Center, where clinicians trained in ultrasound can confirm or rule out intra-articular bleeding and guide further management. If bleeding is excluded, rest and non-steroidal anti-inflammatory drugs selective for cyclooxygenase-2 may be suggested instead, particularly in cases of acute arthropathic pain or active synovitis (28).

The results of our study show that the proposed CAD tool is feasible. The possibility to avoid undertreatment of subclinical bleeding and overtreatment of acute joint pain due to arthropathy represents an important aspect in providing the most accurate personalized care to all patients managed by the hemophilia comprehensive care center.

The present study has both limitations and strengths. Despite showing good accuracy, in the future, we aim to improve the sensitivity of the CAD tool for mild and moderate bleeding, as our goal is to use it as a screening tool for persons with hemophilia with subclinical bleeding to avoid undertreatment. Training deep learning models typically requires large imaging datasets. Despite the rarity of hemophilia and the reduced frequency of bleeding events in contemporary cohorts, likely due to the widespread adoption of novel therapeutic agents and improved treatment availability (6,7), we could collect a large dataset of images to train and test the algorithm. To prevent data leakage and overfitting, images from the same patient were not shared between training and testing phases, and a standard split was applied, with approximately 80% of patients allocated to the training set and 20% to the test set. Despite this constrained framework, the achievement of statistically significant results supports the adequacy of the algorithm's training and its ability to extract meaningful patterns from the available data. A larger imaging dataset is currently being collected as part of routine clinical practice at our Center, and its inclusion in future analyses is expected to further improve the accuracy and sensitivity of the system and allow validation in larger, independent cohorts. In addition, images from examinations performed during routine follow-up visits and those acquired in the setting of suspected acute joint bleeding were not distinguished, as acute bleeding events were infrequent. Future studies will be conducted to incorporate structured, prospective data collection to clearly distinguish routine assessments from acute bleeding-related evaluations, thereby enabling appropriately stratified analyses that more accurately reflect the underlying clinical contexts.

We selected joint capsule distension as a surrogate marker for joint bleeding. Indeed, apart from joint bleeding, capsule distension can be due to many different conditions: simple synovial fluid effusion in the context of chronic arthropathy, which appears as an anechoic compressible effusion; or synovial hyperplasia, which appears as an isoechoic or hypoechoic capsule thickening incompressible during dynamic maneuvers, with or without power-Doppler signal in its context. However, joint capsule distension has the advantage of being easily identifiable even in still US images acquired by non-expert US operators.

The study by Tyrrell et al (29) also uses synovial recess distension detected by point-of-care US as an indicator of potential joint bleeding, across various joints in both adult and pediatric subjects, achieving high accuracy metrics (97%). To initiate the development of our system, we studied only adult subjects because in children, the subchondral bone is not completely formed, and anechoic epiphyseal cartilage may be difficult to distinguish from simple effusion without dynamic maneuvers. In fact, Tyrrell et al observed a decline in performance when their model was applied to pediatric datasets, possibly due to the lower representation of children in the overall data pool but also due to the difficulty in distinguishing simple fluid from growing cartilage.

We selected the knee - a large, easily accessible joint, even in patient-based settings - focusing on a specific view, the longitudinal scan of the SQR, in which capsule distension of the main recess is easily recognizable in a population of adult persons with hemophilia (18). In this phase, to train the CAD tool, we used high-quality US images from persons with hemophilia acquired by an expert US operator with a high-end machine. In the future, we plan to use images from all the main hemophilia index joints (ankles, knees, and elbows) and images acquired both by expert operators with high-end devices and by patients or caregivers using portable ultrasound systems, in order to further train the tool to recognize capsule distension under these real-world conditions.

In this dataset, the number of images showing joint capsule distension due to bleeding is low, reflecting the effective bleeding control achieved with current treatment strategies (6,7). In addition, when present, the joint capsule distension is not massive, but mild or moderate in the majority of cases, for the same reasons. However, compared with Tyrrel et al (29), we did not include patients with other joint conditions, such as rheumatoid arthritis, so the validity of our recognition system is good for persons with hemophilia.

The binary classification (distended/non-distended) we tested bears a higher inter-rater concordance compared with the four-class classification of capsule distention, based on the Cohen's kappa. However, a binary classification may lead to a limitation in the sensitivity of this approach in detecting small amounts of effusion, which is important in the era of increased availability of prophylaxis and effectiveness of currently available drugs. Larger studies will clarify the validity of the two approaches.

All images were acquired at a single center using identical equipment and were obtained by a single operator, although image annotation was performed independently by two expert operators. While this approach ensured technical and acquisition consistency, it represents a limitation in terms of generalizability. Further efforts toward standardized annotation criteria are warranted to facilitate future validation studies involving multiple centers, different ultrasound systems, and operators with varying levels of expertise. An International Society on Thrombosis and Haemostasis Scientific and Standardization Committee Subcommittee on Factor VIII, Factor IX and Rare Coagulation Disorder project on this topic is ongoing (https://cdn.ymaws.com/isth.site-ym.com/resource/resmgr/ssc/ssc_myisth/factor8&9_musculoskeletal_ul.pdf). Such validation will be essential to confirm the robustness of our findings and to support their applicability in telemedicine settings.

Another important limitation of the presented approach is the differential diagnosis between complex joint effusion due to joint bleeding vs synovial thickening, which is still a challenge in some cases, even for expert US operators, and requires dynamic maneuvers. In our system, we use still images. In the future, videos incorporating compressive maneuvers may be used to improve the detection of small volumes of blood and to differentiate them from synovial thickening.

Another limitation of this study is the lack of techniques to support the interpretability of automatic classification results. To address this limitation, we plan to experiment XAI techniques like Grad-CAM (30) and SHAP (31). Another possible direction to improve the interpretability of the classification results is to display the confidence value computed by the model, in order to support clinicians in identifying borderline cases, where greater attention is required. This would require a dedicated study to assess the correlation between confidence values and model accuracy.

The next step of our project will be to add elbow and ankle evaluation to our CAD system. When a complete CAD tool is available, we will conduct a pilot study to demonstrate the validity of images collected in a home-based setting using portable ultrasound systems.

From a health system perspective, the clinical feasibility of telemedicine-based ultrasound assessment should ultimately be complemented by an evaluation of its economic and organizational impact. A formal Health Technology Assessment would allow comparison of this telemedicine approach with standard care pathways in terms of cost effectiveness, resource utilization, and potential effects on patient access and continuity of care. Such analyses are particularly relevant in chronic conditions requiring longitudinal monitoring, where telemedicine may reduce travel burden, optimize specialist time, and improve equity of access. Incorporating economic evaluation into future studies would therefore provide critical evidence to inform decision-makers and support the sustainable integration of telemedicine solutions into routine clinical practice.

One strength of our study is the inclusion of both male and female healthy controls to ensure greater validity of findings and relevance within the framework of precision medicine. This approach allows the future application of our study on musculoskeletal complications in the rare subgroup of women and girls with hemophilia, and other rare bleeding disorders affecting both sexes, such as Von Willebrand disease (32-35).

Another strength of the present study is its focus on the detection of capsule distension even of mild entity, reflecting the current clinical reality of hemophilia, in which severe overt joint bleeding have become increasingly rare due to advances in treatment. In the future, sensitivity to minimal deviations from individual baseline rather than cross-sectional discrimination between patients and controls, will be particularly relevant for telemedicine-based follow-up strategies.

Conclusion

The present study demonstrates the feasibility and good accuracy of our CAD tool, which can be improved to become more sensitive even in the case of subclinical joint bleeding. This is the first step of our proposed telemedicine system to achieve an early and accurate diagnosis of joint bleeding in persons with hemophilia, with the ultimate goal to avoid undertreatment, thus preventing the occurrence and delaying progression of arthropathy while also avoiding overtreatment due to unnecessary concentrate administration in the absence of bleeding.

In a novel era of great progress and availability of replacement and non-replacement drugs for persons with hemophilia, our telemedicine system will allow an accurate personalized management.

Author Contribution: F. Peyvandi, R. Gualtierotti, C. Bettini, S. Mascetti concept and design, R. Gualtierotti, A. Giachi, S. Mascetti, S. Arcudi, A. Truma, C. Suffritti, M. Colussi, M. Manzoni, D. Ahmetovic, S.A. Angileri, G. Carrafiello, analysis and/or interpretation of data; R. Gualtierotti, S. Mascetti, M. Colussi, D. Ahmetovic, C. Bettini, F. Peyvandi critical writing or revising the intellectual content, all authors provided final approval of the version to be published.

Acknowledgments

The Hemostasis & Thrombosis Unit of the Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico is member of the European Reference Network on Rare Haematological Diseases EuroBloodNet-Project ID No 101157011. ERN-EuroBloodNet is partly co-funded by the European Union within the framework of the Fourth EU Health Programme.

Conflict of Interest

Roberta Gualtierotti is on the advisory boards of Bayer, Biomarin, Roche, Sanofi, SOBI, and Novo Nordisk, and has participated in speaker bureau/educational meetings for Biomarin, Pfizer, SOBI, Takeda, and Novo Nordisk. Flora Peyvandi is on the advisory board of Biomarin, CSL Behring, Pfizer, Roche, Sanofi, Sobi and Regeneron and has participated in educational meetings sponsored by Takeda and Sanofi. All other authors have no conflicts of interest to disclose.

Funding

The study was partially supported by the nonprofit Angelo Bianchi Bonomi Foundation; Sobi, Bayer and Roche provided unrestricted grants to the Angelo Bianchi Bonomi Foundation (JoinMi project); by the Italian Ministry of Health – Bando Ricerca Corrente; the Multilayered Urban Sustainability Action (MUSA) project, funded by the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (PNRR) Mission 4 Component 2 Investment Line 1.5: Strengthening of research structures and creation of R&D “innovation ecosystems”, set up of “territorial leaders in R&D”; Italian Ministry of Education and Research - MUR (‘Dipartimenti di

Eccellenza' Programme 2023–27 - Dept. of Pathophysiology and Transplantation, Università degli Studi di Milano); PETRA, Italian taxes 5 x 1000 - 2020 devolved to Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico.

Ethics Statement

All procedures were performed in compliance with relevant laws and institutional guidelines. The study was approved by the Milan Area 2 Ethics Committee (199_2021bis and subsequent amendment by the CET Lombardy 3 (ID 2086 – n° 2086_SA_22.03.2024_P)

Table 1. The number of images used to train and to test the system

	Suitable images	Unsuitable Images	Collected images
Train	711	-	-
Test	103	-	-
Total	814	143	957

Table 2. Distribution of the train set images by capsule distension level.

	Absent	**Mild**	**Moderate**	**Severe**
Images	399	185	115	12

Table 3. Classification accuracy absent vs mild or moderate or severe

Absent vs **mild, moderate, severe**	Male persons with hemophilia	Male healthy controls	Female Healthy controls
Absent	24	23	23
mild, moderate, severe	13	10	10
Accuracy	78.4%	84.9%	93.94%
Sensitivity	61.5%	70%	80.00%
Specificity	87.5%	91.4%	100%
Balanced Accuracy	74.5%	80.7%	90.00%
Correct	29	28	31
Wrong	8	5	2

455 **Table 4.** Classification accuracy absent or mild vs moderate or severe

Absent, mild vs **moderate, severe**	Male persons with hemophilia	Male healthy control	Female Healthy control
Absent, mild	33	29	28
moderate, severe	4	4	5
Accuracy	89.2%	87.9%	96.97%
Sensitivity	100%	75%	80%
Specificity	87.9%	89.7%	100%
Balanced Accuracy	93.9%	82.3%	90.00%
Correct	33	29	32
Wrong	4	4	1

456

References

1. Bolton-Maggs PH, Pasi KJ. Haemophilias A and B. Lancet. 2003;361(9371):1801-9.

2. Di Minno MN, Ambrosino P, Franchini M, Coppola A, Di Minno G. Arthropathy in patients with moderate hemophilia a: a systematic review of the literature. Seminars in thrombosis and hemostasis. 2013;39(7):723-31.

3. Gualtierotti R, Solimeno LP, Peyvandi F. Hemophilic arthropathy: current knowledge and future perspectives. J Thromb Haemost. 2021.

4. Hooiveld MJ, Roosendaal G, van den Berg HM, Bijlsma JW, Lafeber FP. Haemoglobin-derived iron-dependent hydroxyl radical formation in blood-induced joint damage: an in vitro study. Rheumatology (Oxford, England). 2003;42(6):784-90.

5. Manco-Johnson MJ, Abshire TC, Shapiro AD, Riske B, Hacker MR, Kilcoyne R, et al. Prophylaxis versus episodic treatment to prevent joint disease in boys with severe hemophilia. N Engl J Med. 2007;357(6):535-44.

6. Mannucci PM. Hemophilia therapy: the future has begun. Haematologica. 2020;105(3):545-53.

7. Manco-Johnson MJ, Soucie JM, Gill JC. Prophylaxis usage, bleeding rates, and joint outcomes of hemophilia, 1999 to 2010: a surveillance project. Blood. 2017;129(17):2368-74.

8. Ceponis A, Wong-Sefidan I, Glass CS, von Drygalski A. Rapid musculoskeletal ultrasound for painful episodes in adult haemophilia patients. Haemophilia. 2013;19(5):790-8.

9. Gualtierotti R, Giachi A, Truma A, Arcudi S, Ciavarella A, Bucciarelli P, et al. Assessing joint health in haemophilia patients: The combined value of physical examination and ultrasound imaging. Haemophilia. 2024;30(4):1018-24.

10. De la Corte-Rodriguez H, Rodriguez-Merchan EC, Alvarez-Roman MT, Martin-Salces M, Martinoli C, Jimenez-Yuste V. The value of HEAD-US system in detecting subclinical abnormalities in joints of patients with hemophilia. Expert Rev Hematol. 2018;11(3):253-61.

11. Melchiorre D, Linari S, Innocenti M, Biscoglio I, Toigo M, Cerinic MM, et al. Ultrasound detects joint damage and bleeding in haemophilic arthropathy: a proposal of a score. Haemophilia. 2011;17(1):112-7.

12. Berro M, Elichiry M, Wasen K, Insagaray J, Rodríguez I. Use of ultrasound for evaluation of painful joint episodes perceived as haemarthrosis in adult patients with severe haemophilia. Haemophilia. 2018;24(3):e124-e5.

13. Gallastegui N, Steiner B, Aguero P, Bailey C, Kruse-Jarres R, Quon DV, et al. The role of point-of-Care Musculoskeletal Ultrasound for Routine Joint evaluation and management in the Hemophilia Clinic - A Real World Experience. BMC Musculoskelet Disord. 2022;23(1):1111.

14. von Drygalski A, Pasta G, de la Corte-Rodriguez H. Ultrasound and patient self-imaging in hemophilia. Haemophilia. 2021;27(2):e298-e301.

15. Colussi M, Civitarese G, Ahmetovic D, Bettini C, Gualtierotti R, Peyvandi F, et al. Ultrasound detection of subquadricipital recess distension. Intelligent Systems with Applications. 2023;17:200183.

16. Ahmetovic D, Angileri A, Arcudi S, Bettini C, Civitarese G, Colussi M, et al. Insights on the Development of PRACTICE, A Research-Oriented Healthcare Platform. 2024 IEEE International Conference on Smart Computing (SMARTCOMP). 2024:2020-385.

17. Beggs I, Bianchi S, Bueno A, Cohen M. ESSR Ultrasound Group Protocols. Musculoskeletal Ultrasound Technical Guidelines: Knee.

18. Martinoli C, Della Casa Alberighi O, Di Minno G, Graziano E, Molinari AC, Pasta G, et al. Development and definition of a simplified scanning procedure and scoring method for Haemophilia Early Arthropathy Detection with Ultrasound (HEAD-US). Thromb Haemost. 2013;109(6):1170-9.

19. Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston AD, Philbrick KA, et al. A Survey of Deep-Learning Applications in Ultrasound: Artificial Intelligence-Powered Ultrasound for Improving Clinical Workflow. Journal of the American College of Radiology : JACR. 2019;16(9 Pt B):1318-28.

20. Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, et al. Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis. Investigative radiology. 1990;25(10):1102-10.

21. Chen M, Shi X, Zhang Y, Wu D, Guizani M. Deep Feature Learning for Medical Image Analysis with Convolutional Autoencoder Neural Network. IEEE Transactions on Big Data. 2021;7:750-8.

22. Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. The British journal of radiology. 2005;78 Spec No 1:S3-s19.

23. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al., editors. Microsoft COCO: Common Objects in Context. Computer Vision – ECCV 2014; 2014 2014//; Cham: Springer International Publishing.

24. Campana MG, Colussi M, Delmastro F, Mascetti S, Pagani E. A Transfer Learning and Explainable Solution to Detect mpox from Smartphones images. Pervasive and Mobile Computing. 2024;98:101874.

25. Jocher G, Chaurasia A, Stoken A, Borovec J, NanoCode012, Kwon Y, et al. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation: Zenodo; 2022 November 22, 2022.

26. Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(07):12993-3000.

27. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision. 2010;88(2):303-38.

28. Srivastava A, Santagostino E, Dougall A, Kitchen S, Sutherland M, Pipe SW, et al. WFH Guidelines for the Management of Hemophilia, 3rd edition. Haemophilia. 2020;26 Suppl 6:1-158.

29. Tyrrell PN, Alvarez-Román MT, Bakeer N, Brand-Staufer B, Jiménez-Yuste V, Kras S, et al. Utilizing artificial intelligence for the detection of hemarthrosis in hemophilia using point-of-care ultrasonography. Research and Practice in Thrombosis and Haemostasis.

30. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 618–626.

31. S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” Advances in neural information processing systems, vol. 30, 2017.

32. Corrales-Medina FF, Fraga K, D'Almeida Bastos M, Rafique A, Kempton CL, Vijayvargia P, et al. Assessment of joint health in females with haemophilia: The carriers ultrasound project (CUP) study. Haemophilia. 2024;30(6):1422-8.

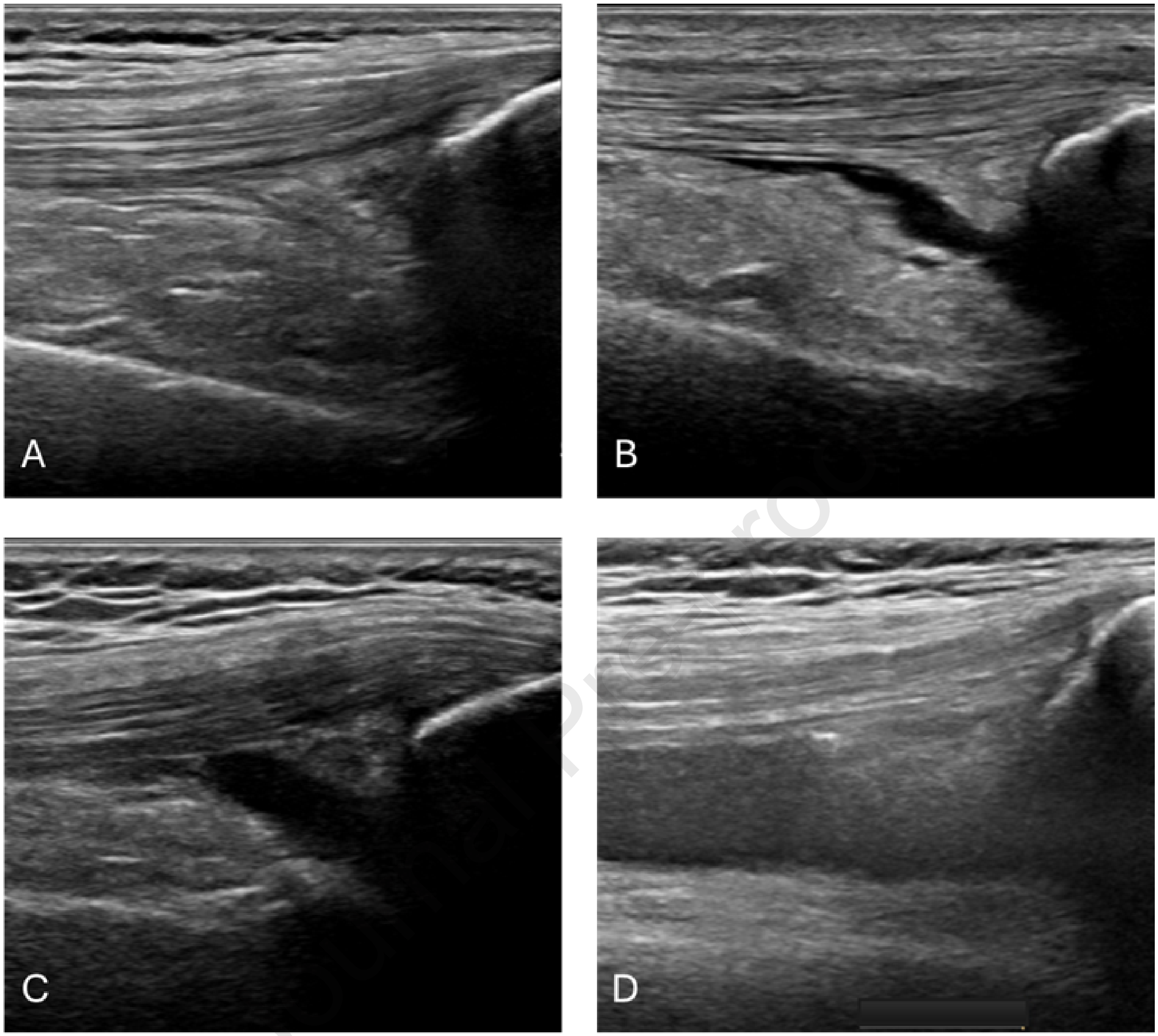
33. Gualtierotti R, Garagiola I, Mortarino M, Spena S, Romero-Lux O, Peyvandi F. Gender equity in hemophilia: need for healthcare, familial, and societal advocacy. Frontiers in Medicine. 2024;11.

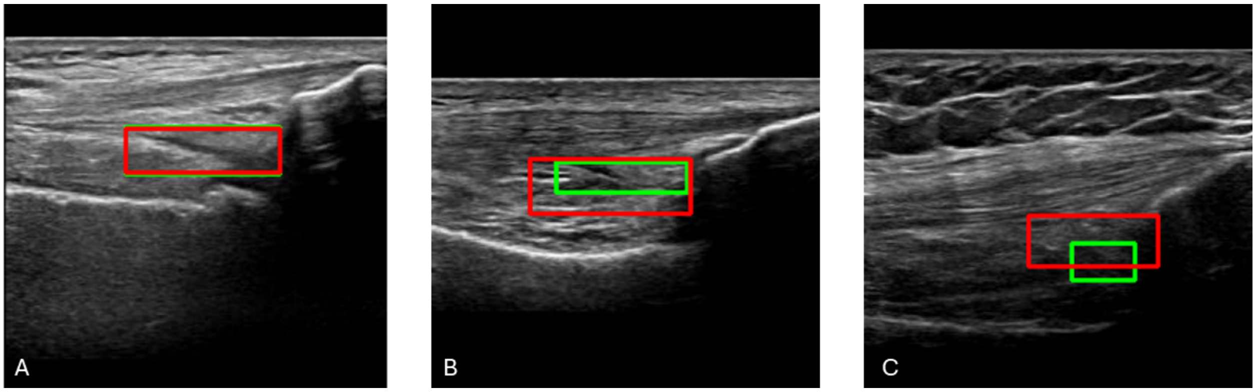
34. Castaman G, Katsarou O, Jansen N, Santos S, Escolar G, Berntorp E. Clinical, economic, and health-related quality of life burden associated with von Willebrand disease in adults and children: Systematic and targeted literature reviews. Haemophilia. 2023;29(2):411-22.

35. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. Res Integr Peer Rev. 2016;1:2.

Figure 1. Criteria for knee capsule distension in a longitudinal scan of the subquadricipital recess. A. Absent: the recess is not visible because the two fat pads collapse (<1 mm); B. Mild: the recess is extended in width between 1 and 5 mm; C. Moderate: the recess is extended between 6 mm and 5 cm (i.e. it does not exceed the limit of the frame in which the image is included, which corresponds to the width of the ultrasound probe with a setting of 5 cm depth); D. Severe: if the distension exceeds the frame in which the image corresponding to the width of the ultrasound probe (in our case we used a Philips Affiniti 50 with a 5-12 MHz probe).

Figure 2. Examples of detected SQRs, the ground-truth is depicted in green, while prediction in red. A. Accurate prediction with IoU > 0.5; B. Slightly inaccurate prediction with IoU < 0.5; C. Wrong prediction with IoU < 0.5.


A
B
C
D


A
B
C