



Semi-automated image analysis of cellulose nanofibrils using machine learning segmentation and morphological thinning

Carlos Baez · Udit Ringania ·
Saad Bhamla · Robert J. Moon

Received: 17 November 2025 / Accepted: 31 March 2026 / Published online: 15 April 2026

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2026

Abstract Quantitative width distributions of cellulose nanofibrils (CNFs) are difficult to obtain from microscopy when fibrils are highly branched and entangled, and when manual measurements rely on sparse sampling and subjective fibril selection. Here we present FACT (Fibril Analysis for Cellulose Technology), a semi-automated image-analysis framework that extracts length-weighted fibril width distributions from negative-contrast scanning electron microscopy (NegC-SEM) images. FACT uses machine-learning segmentation (either the ImageJ Weka plugin or a U-Net convolutional neural network) to generate binary CNF masks, then applies morphological thinning to obtain a one-pixel-wide skeleton. Local fibril width is calculated from the distance between each skeleton pixel and the segmented fibril boundary. Using idealized simulated geometries, hierarchical branched structures, and

fixed-diameter wire micrographs, we identify practical operating limits for robust width statistics: high-contrast images, fibril aspect ratio greater than 10, and at least ~5 pixels across the fibril width. FACT was then applied to NegC-SEM images representing low- and high-branching CNF morphologies and compared with manual ImageJ measurements. Central tendencies were similar, while distribution shapes differed because FACT measures widths at many points along each fibril (effectively length-weighted), whereas manual analysis typically records one width per fibril (number-weighted); these outputs are therefore complementary rather than directly interchangeable. Once a trained image segmentation model is available, FACT analyzes each image in under 5 min, enabling higher-throughput morphology reporting.

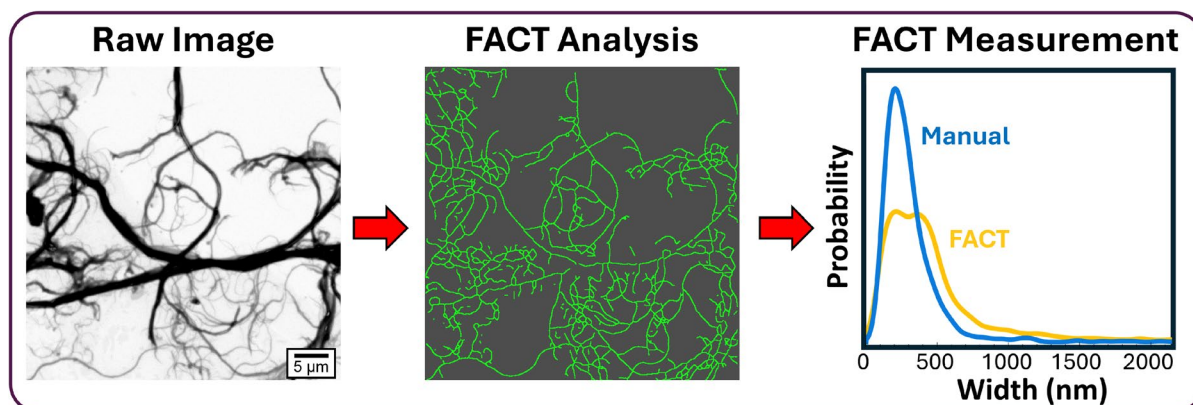
Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10570-026-07023-1>.

C. Baez (✉) · R. J. Moon (✉)
The Forest Products Laboratory, USDA Forest Service,
Madison, WI 53726, USA
e-mail: carlos.l.baez@usda.gov

R. J. Moon
e-mail: Robert.j.moon@usda.gov

U. Ringania · S. Bhamla
Chemical and Biomolecular Engineering, Georgia Institute
of Technology, Atlanta, GA 30332, USA

Graphical Abstract



Keywords Cellulose nanofibrils · Negative-contrast scanning electron microscopy · Semi-automatic image analysis · Particle size measurements · Morphological skeletonization · Convolutional neural network · Image segmentation

Introduction

Cellulose nanomaterials (CNMs) possess a unique set of characteristics that give them utility across a wide range of applications, including composite materials, biodegradable and renewable packaging, biomedical, and rheology modification (Moon et al. 2011; Zambrano et al. 2020; Chen et al. 2021; Li et al. 2021a, 2021b; Eichhorn et al. 2022). The International Organization for Standardization (ISO) has defined three categories of woody plant-based CNMs: cellulose nanocrystals (CNCs), individual cellulose nanofibrils (iCNFs), and cellulose nanofibrils (CNFs) (ISO 2023). CNCs primarily display a spindle-like, non-branching morphology, while iCNFs primarily display a high aspect ratio, non-branching fibril morphology. Approaches for measuring CNC and iCNF particle length, width, and height have been developed (ISO 2021a, 2021b). In contrast, CNFs are complex, fibril-like objects with extensive branching that entangle into networks; as a result, there are no

technical specifications for general CNF particle size measurement. (Moon et al. 2023, 2025).

Reliable and rapid measurement of CNF particle morphology (object size, degree of branching, and fibril width and length) is crucial for quality control, grade definition, and investigation of morphology-performance relationships for various applications. This need has been identified as a priority within the CNF research community (Kangas et al. 2014; Desmaisons et al. 2017; Moon et al. 2023). The perspective paper by Moon et al. (2023) presents a summary of responses from a 2020 survey and outcomes from a 2021 international workshop on CNF measurement and standardization needs that was organized and hosted by ISO/TC 6/TG 1 Cellulosic nanomaterials and the technical association of the pulp and paper industry (TAPPI) NanoDivision. Participants were from CNF manufacturers, commercial users, and researchers. CNF suspensions exhibit a broad range of particle morphologies and dimensions, spanning from micrometers to nanometers. As a result, a broad range of complementary analysis approaches (direct imaging, indirect scattering, sedimentation, rheology, image skeletonization, machine learning, automated assessment, etc.) are being explored to quantify the morphology and dimensions of fibril systems across different length scales (Kangas et al. 2014; Huang et al. 2016; Griebner et al. 2023; Moon et al. 2023; Nakayama et al. 2025; Verho et al. 2025). Macro- and

micro-sized objects can be directly measured using fiber analyzers designed for pulp fiber measurement, which can typically detect object widths in the tens of microns and object lengths ranging from 100 μm to several millimeters. To directly image the nanoscale fibril features within CNF objects, techniques such as Transmission Electron Microscopy (TEM), Atomic Force Microscopy (AFM), or Scanning Electron Microscopy (SEM) are typically used. SEM has a much larger field of view and a broader magnification range than TEM or AFM. With SEM, it is possible to capture entire CNF objects while still providing sufficient resolution to capture the broad size-scale range of widths and lengths of hierarchical fibril branching within a single CNF object or across a tangled fibril network. SEM imaging offers the potential for enhanced characterization of CNF particle morphology (Moon et al. 2025).

However, using SEM to measure micro- to nanoscale features of CNFs remains challenging. Firstly, improved sample preparation is needed to minimize agglomeration and overlap of branching fibrils, which negatively affect fibril identification and measurements during image analysis. Highly diluted suspensions can offset some of these problems (Ringania et al. 2022). Secondly, improved SEM imaging techniques are necessary for non-conducting objects to achieve higher image contrast and sharpness, thereby making fibril edges more distinct and facilitating the identification, measurement, and analysis of fibrils. Conventionally, non-conductive or beam-sensitive materials, such as CNFs, are imaged after they are coated with a thin conductive layer, typically 1–2 nm thick (Ang et al. 2020), which at higher imaging magnifications may obscure small features, fill gaps, artificially broaden objects, and typically results in lower contrast images. An alternative approach is to use the negative-contrast SEM (NegC-SEM) methodology, in which non-conductive objects are deposited onto an atomically smooth and highly conductive substrate (Mattos et al. 2019; Beaumont et al. 2021; Ringania et al. 2022; Moon et al. 2025). The secondary electron response for CNF is low, whereas it is high for the substrate, resulting in a high-contrast image that facilitates the analysis of CNF fibril length and width across various length scales. Thirdly, improved image analysis is necessary to measure a sufficient quantity of fibril features that are statistically relevant while minimizing analyst

bias. This study addresses these challenges by utilizing dilute suspensions, negative-contrast scanning electron microscopy (NegC SEM), and a novel image analysis approach.

Current image analysis of CNFs often requires manual measurements and calculations, which are labor-intensive and time-consuming (Ang et al. 2020). Various image analysis tools are available, offering a wide range of capabilities for measuring particle dimensions. Manual image analysis with Gwyddion (Nečas and Klapetek 2012; Mattos et al. 2019) and ImageJ (Schneider et al. 2012; Ang et al. 2020) has been widely used for CNF image analysis. However, object selection and measurement are susceptible to user bias and fatigue, which contribute to discrepancies in the dimensional data of CNFs reported in the literature. Semi-automated image analysis programs designed for the analysis of high aspect ratio fiber particles without branching morphologies, such as FibrilJ (Sokolov et al. 2017), FiberApp (Usov and Mezzenga 2015; Persson et al. 2017), DiameterJ (Hotaling et al. 2015), and a custom Python package (Willhammar et al. 2021), have had some success when applied to CNC or CNFs with minimal or low levels of branching. However, when applied to CNFs with a high level of branching, these programs struggle with the complexity of overlapping fibrils, varying degrees of fibrillation or branching, and fibrils of varying size scales. The MATLAB code UNDFiber, which was effective to some extent for highly dispersed optical images (Ringania et al. 2022), failed to capture any meaningful data when encountering overlapping fibrils in the SEM images. This study addresses this gap by developing a semi-automatic image analysis approach, Fibril Analysis for Cellulose Technology (FACT), that measures the width distribution of hierarchical, branched, and entangled CNF structures from NegC-SEM images.

This paper first describes the development of FACT, which uses machine learning (ML) for image segmentation, followed by morphological thinning to measure fibril width distributions. Next, validation experiments using simulated fibril-like geometries are used to help interpret the width distribution results. FACT was then used to analyze NegC-SEM images of CNFs from two published studies: one with low branching (Mattos et al. 2019; Beaumont et al. 2021) and one with high branching and networking (Ringania et al. 2022). FACT width results were compared

with manual measurements. Where possible, this study followed the recommendations for sample preparation, imaging, analysis, and reporting as outlined in the paper "Perspectives on Cellulose Nanofibril Size Measurement Using Scanning Electron Microscopy" by an International Organization for Standardization (ISO) task group working on standards for CNF characterization (Moon et al. 2025).

Materials and methods

Cellulose nanofibrils (CNFs)

FACT was used to analyze NegC-SEM images of two CNF variants (low-level and high-level branching) from published studies. A brief description of the reported CNF material preparation from these studies is given here. The low-level branching CNFs used in the Mattos et al. (2019) and Beaumont et al. (2021) studies were prepared from a never-dried, fully bleached, and fines-free sulfite birch pulp, without any chemical or enzymatic pretreatment, diluted to 0.5 wt.% solids in water, and fibrillated via six passes through a high-pressure microfluidizer (Microfluidics M110P). The resulting CNF suspension (0.5 wt.% solids in water) consists of a low-branched fibril structure, with fibril dimensions ranging between 5 and 30 nm in width and 50–5000 nm in length (Beaumont et al. 2021). The CNFs with high-level branching, as used in the Ringania et al. (2022) study, were produced by the Process Development Center at the University of Maine [lot # U-103, 90% fines] (UMaine 2024). The 'fines' percentage, defined by the ISO specification (ISO 2014), is the percentage of fibrils with a length less than 200 μm . These CNFs were prepared by mechanical fibrillation of wood pulp fibers without prior chemical or enzymatic pretreatment. A 5 wt% solids slurry with water was passed through mechanical refiners (Valmet, DD6700, commercial scale), using a two-stage approach. This process takes about 5 h in stage one and 3 h in stage two. The resulting CNF suspension (3 wt.% solids in water) consists of a hierarchical branched fibril structure, with fibril branching element dimensions ranging between 20 and 300 nm in width and overall branched particle size of several tens of micrometers (Ringania et al. 2022).

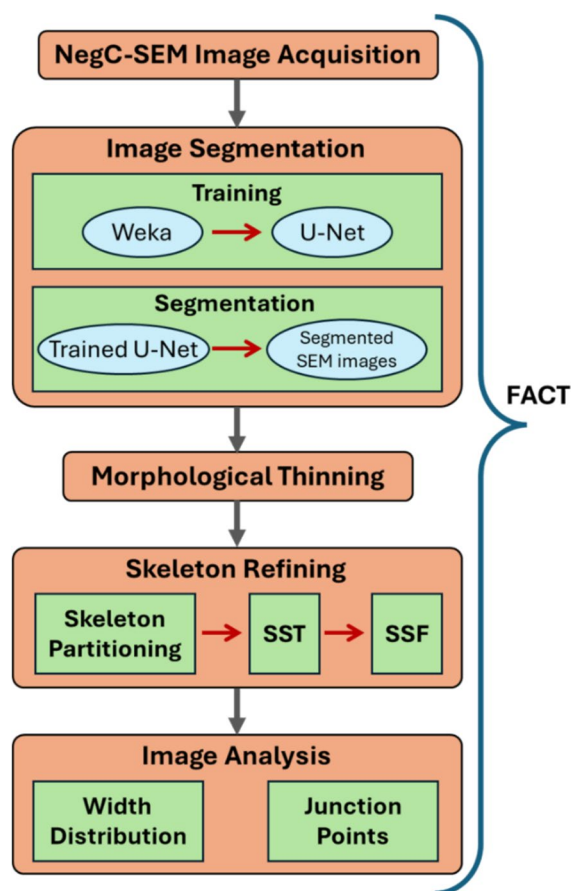


Fig. 1 Flowchart of the FACT automated image analysis process. Where SST = skeleton segment trimming, and SSF = skeleton segment filtering

Semi-automated image analysis

A semi-automated image analysis framework, Fibril Analysis for Cellulose Technology (FACT), was developed to measure the fibril width distributions for the entire CNF hierarchical branching structure. A flowchart of this process is illustrated in Fig. 1. The FACT framework consists of five parts: 1) image acquisition, 2) image segmentation, 3) morphological thinning, 4) skeleton refining, and 5) image analysis.

NegC-SEM image acquisition

Image acquisition is critically important for the FACT approach to identify the CNF and provide accurate width measurements. Images should have high

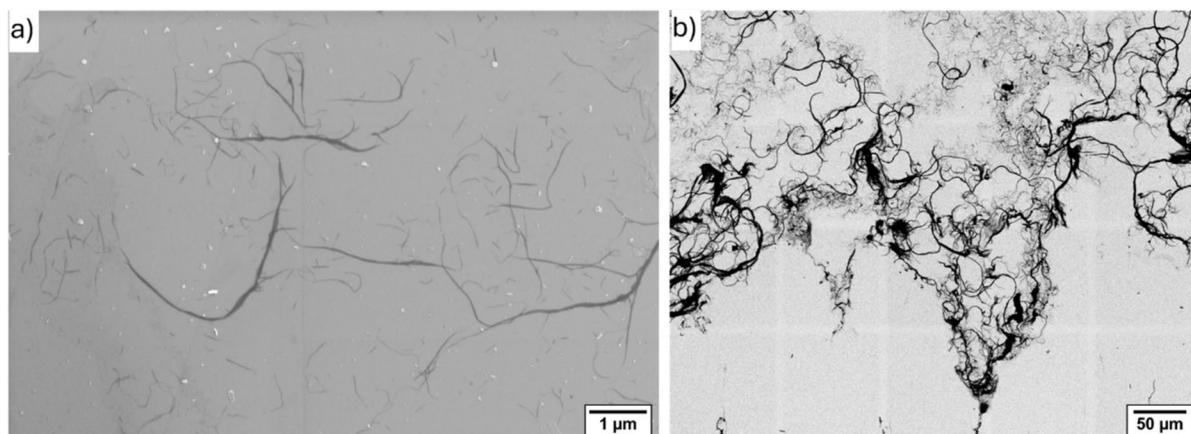


Fig. 2 NegC-SEM images. **a** low-level branching CNFs by Beaumont et al. (2021), and **b** an automated image map of the high-level branching CNFs by Ringania et al. (2022), showing

the notable differences in overall object scale, morphology, level of branching, and network structure

contrast, sharply defined CNF-substrate edges or boundaries, and a low area coverage (<20%) of well-dispersed CNFs on the substrate surface. NegC-SEM imaging of CNFs exhibits greater CNF-substrate contrast differences than conventional SEM imaging (Mattos et al. 2019; Beaumont et al. 2021; Ringania et al. 2022; Moon et al. 2025), enabling more accurate image segmentation.

In this study, FACT is used to analyze NegC-SEM images of two CNF variants (low-level branching and high-level branching) from published studies. The two CNF case studies showed differences in particle morphology. Additionally, the different imaging conditions also resulted in fundamentally different image contrast values and spatial resolutions. A brief description of the reported SEM sample preparation and imaging parameters from these studies is given here. The low-level branching CNF samples used in the Mattos et al. (2019) and Beaumont et al. (2021) studies were prepared by first diluting the starting 0.5 wt.% solid aqueous suspension to 0.01 wt.% using Milli-Q water and tip ultrasonicated (3 min with pulse on/off of 5/1 s at 10% amplitude). Electrically conductive substrates were produced using freshly cleaved mica discs that were spun-coated with a 4 nm thick layer of either gold, platinum/palladium alloy, or iridium. This substrate was then dipped in a 0.33% w/v Poly(ethylene imine) solution for 1 min, rinsed with Milli-Q water, and subsequently dipped into the diluted CNF suspension for 1 min. It was dried

under ambient conditions (Mattos et al. 2019; Beaumont et al. 2021). NegC-SEM images were obtained using an SEM [FE-SEM, Zeiss Sigma VP] at imaging conditions [1–1.5 kV, working distance of 1.8–6 mm] with an in-lens secondary electron detector. Images used in the current study (Fig. 2a) had magnifications of 10,000 $\times$ or 30,000 $\times$, with an image size of 2048 $\times$ 1347 pixels, corresponding to pixel sizes of 5.43 nm/pixel and 1.79 nm/pixel, respectively.

High-level branching CNF samples used in the Ringania et al. (2022) study were prepared by first diluting the starting 3 wt.% solid aqueous suspension to 0.001 wt.% using DI water and subsequently mixing using a vortex mixer (VWR Analog Vortex Mixer No. 10153–838, speed 7, 30 s). A micro-pipette was used to deposit four 2 μ L droplets of the diluted suspension onto a 1 cm $\times$ 1 cm silicon wafer (resistivity: 1–30 Ohm-cm, P-type with no SiO₂ top coating, 460–530 μ m thickness, 2 nm roughness, ordered from Tedpella). These drops remained isolated even after mild spreading and dried overnight in ambient conditions (at least 12 h) before SEM imaging. NegC-SEM images were obtained using a benchtop SEM [Phenom Pure] at imaging conditions [5 kV, working distance of 8.8–8.9 mm] with a backscattering detector. High-level branching CNF images used in the current study had magnifications ranging from 1000 to 6000 $\times$, an image size of 2048 $\times$ 2048 pixels, and a pixel size of 22–130 nm/pixel. The SEM's inbuilt automated image mapping (AIM) feature was used to

capture the majority of a given droplet (Fig. 2b) and subsequently to guide the location of higher-magnification images, taken at various positions within individual droplets.

Image segmentation

The FACT image analysis program relies on binary segmented images, such as black-and-white contrast images that separate the foreground (e.g., CNFs) from the background (e.g., the substrate), for subsequent measurements. The quality of the starting image, particularly the contrast differences between the foreground and background regions and the contrast variation within each image, dictates the approach and the amount of effort required to produce segmented images that accurately represent the fibril structure and dimensions. No subsequent segmentation is needed for graphically simulated fibril structures, as the images are already binary. Traditional grayscale thresholding may be adequate for images of high contrast and low variation across the CNFs and the substrate. As contrast decreases and/or variability increases across individual CNF particles, the CNF network, and the substrate, more sophisticated segmentation tools are required (e.g., Weka and U-Net).

For the grayscale NegC-SEM images used in the current study, low-level branching CNFs by Beaumont et al. (2021) and high-level branching CNFs by Ringania et al. (2022), despite their high contrast, standard grayscale image thresholding resulted in suboptimal segmentation (Figs. S1 and S2). Visual inspection found that fibrils were either not correctly identified or that substrate regions were identified as CNFs. Consequently, machine learning algorithms were employed to enhance the segmentation, utilizing the Weka segmentation plugin in ImageJ and the U-Net Convolutional Neural Network (CNN) implemented in Mathematica V12. Weka uses a library of classical machine learning algorithms to perform pixel classification. Weka is simple to use and fast to train, suitable for individual images. U-Net training was more involved and time-consuming, but the segmentation was more robust at handling image-to-image contrast variability within and between the CNF and substrate regions. The trained U-Net model can be applied to multiple images (training and non-training), significantly expanding the set of images

that can be segmented, and its training can be continuously expanded upon.

For the low-level branching CNF material, five NegC-SEM images were received from the Beaumont et al. (2021) study. The Weka segmentation adequately segmented these images (Fig. S1). Since this data set contained only five images and the Weka segmentation was adequate, a U-Net CNN was not trained on these images. The Weka plugin operates by having the user manually delineate regions as foreground or substrate using ImageJ drawing tools (Fig. S3), which are then used to train the Weka classifier. In this case, the two classes of interest are CNF and substrate. The Weka classifier then outputs a probability map (one image per class), where each pixel in each image has a value equal to the probability of belonging to a particular class. The class probability values are bounded between 0 and 1. Pixel classification assigns each pixel a value representing a class the user is interested in spatially identifying in an image (e.g., CNF or substrate). The binary image is produced by thresholding one of the probability maps at 0.5.

If the quality of the final segmented image is sub-optimal, then Weka can be retrained by selecting additional features within the given image. In this study, a single analyst trained the Weka model. To ensure consistency, the freehand line tool in ImageJ with a 1-pixel line width was used to maximize the analyst's control when selecting CNF examples from thin fibrils. For each training image, at least 20 training examples were selected from both the CNF and the background regions.

For the high-level branching CNF material, 22 NegC-SEM images were received from the Ringania (2023) study. It was found that the Weka segmentation was too noisy, leading to jagged edges around the finer fibrils. Consequently, a multi-step process was developed to improve the segmentation. First, Weka segmentation was completed on eight randomly selected grayscale NegC-SEM images to create binary training images. Second, the noise in these binary images was minimized using a morphological operator to remove isolated foreground and background pixels. The resulting segmented images were thresholded at the 0.5 level to create "ground truth images" considered the "correct" segmentation. Third, the eight initial grayscale NegC-SEM images and their corresponding ground truth images were used to train

a U-Net CNN implemented in Mathematica V12 to build a single machine-learning segmentation model. Finally, the trained U-Net model was used to segment all grayscale NegC-SEM images in this dataset. Note that the quality of the ground truth images is critical for U-Net training. The approach used in this study was consistent with practices in the ML community. Based on our observations, we believe the quality and quantity of the ground-truth images were sufficient for this study (see Fig. S2).

The U-Net CNN used in this study was designed by Ronneberger et al. (2015) for pixel classification of biomedical images of HeLa cells. U-Net takes a single-channel grayscale image as input and outputs a multi-channel image, where each pixel contains the probability of belonging to a particular class (e.g., CNF or substrate). The original U-Net architecture was modified in this study to reduce the total network size by reducing the number of feature maps per layer. This modification decreased the training parameters from 31 million (124 MB) in the original U-Net to 7.7 million (31 MB) in our modified version. This modified network requires less GPU memory for training and also reduces both network training and image segmentation time.

In practice, the network input image size is bounded by the available GPU memory. One approach to address this computational limitation is to reduce the network's input image size by partitioning larger images into smaller subimages. An overlap-tile strategy, as described by Ronneberger et al. (2015), was used to eliminate information loss when partitioning larger images. In the current study, the NegC-SEM images of the high-level branching CNF material were 2048×2048 pixels. Because these images were too large, they were subdivided using the overlap-tile strategy into 700×700 pixel subimages. For this, the original image is first padded by mirroring the periphery, resulting in a final image dimension of 2736×2736 pixels. Then, the padded image was partitioned into 64 sub-images of 700×700 pixels with an overlap of 342 pixels.

The U-Net CNN was trained on eight randomly selected grayscale NegC-SEM images and their corresponding ground-truth segmentations (produced by Weka). One advantage of U-Net training is that very few acquired images are needed to yield precise segmentation. The current study partitions each image into 64 sub-images of 700×700 pixels.

Additional training images were produced by rotating each sub-image by 0, 90, 180, and 270 degrees, and extracting the mirror image at 0 degrees. These five symmetry operations increased the total number of training images from 8 (2048×2048 pixels) to 2,560 (700×700 pixels), of which 62% were used for training, 33% for validation, and 5% for testing. This 62/33/5% split was a design choice as it is essential to keep a fraction of the images for validation. These ratios could be slightly adjusted without significantly altering the segmentation performance of the resulting trained U-Net. Note that having more unique images in the training set is usually better than augmenting the training set with fewer unique images. However, data augmentation is a common practice in ML model training when training data is limited. U-Net was trained on these images using stochastic gradient descent for 60 rounds. This training took approximately 3 h on the computer used in this study (Intel Xeon Gold 5218R CPU running at 2.1 GHz, 192 GB of RAM, and an NVIDIA RTX A5000 (24 GB) graphics card).

After training, the U-Net segmentation of a given grayscale high-level branching CNF image consists of partitioning the original image into sub-images using the overlap-tile strategy (as described above), classifying every pixel into either CNF or substrate, and reassembling the sub-images back together into a segmented image having the original starting dimensions, in this case, 2048×2048 pixels. The U-Net segmentation took approximately 5 s for each image.

Morphological thinning

Morphological thinning (Wolfram Research 2010a) is applied to skeletonize the segmented NegC-SEM images. This morphological thinning approach removes pixels from all sides of an object. Thus, the same number of pixels are removed along the fibril edges/sides as they are from the fibril ends. The ramifications of this are that, for each pixel within a given skeleton segment, the width measurement obtained is based on the shortest distance from that pixel to its nearest background pixel.

The thinning operator is defined as the intersection between the original image and the complement of the hit-and-miss operator. The hit-and-miss operator (a Mathematica built-in function) performs template matching, using selected structuring elements as

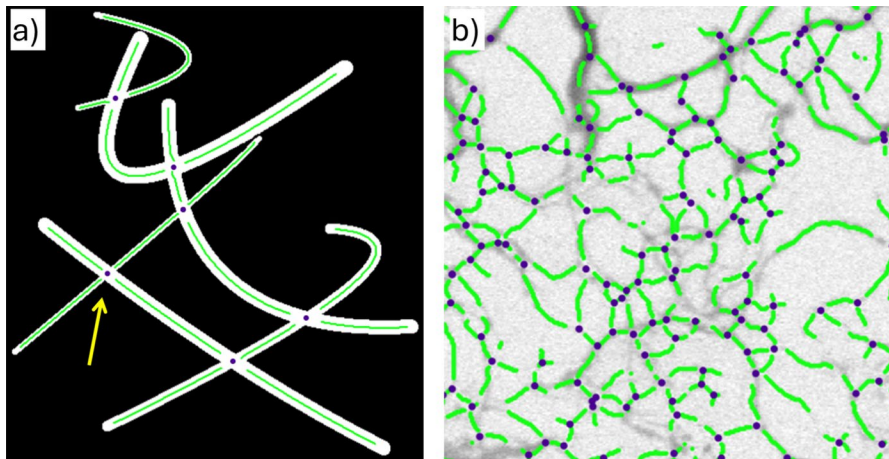


Fig. 3 Examples of skeleton refinement of fibril networks analyzed by FACT, resulting in skeleton junction points (purple dots) and segments (green lines). **a** Simulated fibril structure (SST=20%, SSF=1°), and **b** NegC-SEM image of highly branched CNFs (SST=25%, SSF=25°). Skeleton segment

trimming (SST) removes pixels from the ends of skeleton segments, thus increasing the gap between the green skeleton segments and the purple skeleton junction points (yellow arrow in part a)

input and rastering them across the image to find a match (Wolfram Research 2008). The structuring element must correspond with the current image subsection's foreground and background pixels. If a match is found, the pixel beneath the origin of the structuring element is set to 1; otherwise, it is set to 0. Each iteration of the thinning operation effectively erodes the foreground objects in the binary image by one pixel. The thinning operator is iteratively applied until convergence, yielding a 1-pixel-wide image skeleton.

Subsequently, the distance transform of the unfiltered segmented images is computed. In this process, the value of each foreground pixel is substituted by its Euclidean distance to the nearest background pixel. In cases where the particles in the binary image have jagged edges, it is recommended to apply a Gaussian blur filter to smooth the edges. FACT allows the user to select a kernel size for this filter. Recommended kernel size values are 0 (no filter) or 2. The thinning operator is applied to the filtered binary image. Then, the coordinates of the resulting skeleton pixels are used to extract their corresponding values from the unfiltered distance transform image.

Skeleton refining

The sensitivity of the thinning operator to sharp foreground edges sometimes results in undesirable

skeleton segments, which subsequently degrade the quality of the CNF width information extracted from the image skeleton. A skeleton-refinement method was developed that maintains the integrity of the original binary image while improving fibril width analysis. The skeleton is refined in three steps: 1) skeleton partitioning, 2) skeleton segment trimming (SST), and 3) skeleton segment filtering (SSF). Skeleton partitioning splits the interconnected skeleton into separate segments by identifying all the skeleton junction points using the hit-and-miss operator described earlier. This results in an image containing only the junction points. The pixels of the junction point image are then dilated from single-pixel points into clusters of 9 pixels using a dilation operator, forming a 3×3 matrix with the original junction point in the center. Then, the skeleton image is subtracted by the edited junction point image, effectively partitioning the continuous skeleton network (Fig. 3). Finally, pixel connectivity-based methodologies (Wolfram Research 2010b) are used to extract data from each skeleton segment, such as pixel count, pixel positions, pixel values, and centroids. In addition to SST and SSF, FACT allows the user to select and remove skeleton segments based on a minimum length threshold. This filter was not used in the current study, as a fibril width measurement is independent of its skeleton segment length.

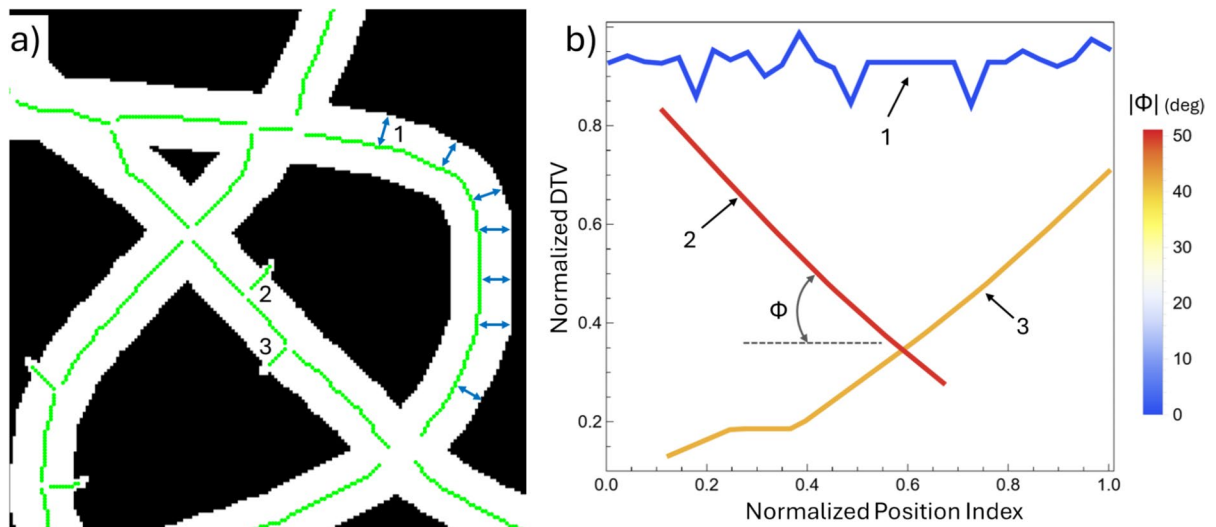


Fig. 4 Skeleton segment filtering (SSF) process. **a** Simulated fibril structure showing how the skeleton segments follow the fibril profile. The blue double-ended arrows (labeled 1) represent the distance from the skeleton to the fiber edge. Simulated edge defects in the periphery of the fibril produce spurious skeleton segments (labeled 2 and 3). **b** Normalized distance transform value (DTV) versus normalized position index. The absolute value of the mean profile slope of a given segment, $|\Phi|$, is color-coded. $|\Phi|$ is what we refer to as the SSF angle in

the rest of the document. For low $|\Phi|$ (i.e., < 10 degrees), the DTV shows low variability (blue profile) that is indicative of skeleton segments following the fibril centerline (labeled 1 in part a). In contrast, for the spurious segments (labeled 2 and 3 in part a), the $|\Phi|$ is high, varying between 30 and 50 degrees (orange/red profiles). FACT removes segments with high $|\Phi|$ from the analysis. SST = 5%, and SSF refinement was not applied in this example, so that the unwanted skeleton segments would be shown in part a

The SST is used to eliminate pixels along the skeleton segment near the skeleton junction points, as they can incorrectly represent the width of the corresponding CNFs in this region. In SST, the points of a skeleton segment are first sequentially ordered, and then a fraction of the skeleton segment points are discarded from the skeleton segment ends, resulting in shorter, trimmed skeleton segments. This process is replicated across all skeleton segments within an image, using an identical skeleton segment endpoint discard fraction. One aspect of this approach is that the number of pixels removed per segment is not constant but instead scales directly with the length of the skeleton segment.

The SSF is used to eliminate unwanted small skeleton segments that do not track along the center line of a fibril and likely originate from segmentation edge defects (e.g., steps, discontinuities) along a given fibril (see segments labeled 2 and 3 in Fig. 4). For a given skeleton segment, its constituent points are plotted as $\{x, y\}$ in two-dimensional (2D) space, where 'x' represents the skeleton segment position index, and 'y' is the corresponding distance transform

value (DTV) at that position. Then, for each segment, its points are standardized to zero mean and unit variance, and the best-fit line and its slope (Φ) are computed. This slope (Φ) represents the average rate of change of the distance transform values along a specific skeleton segment. SSF uses the absolute value of this slope $|\Phi|$ to differentiate between wanted and unwanted skeleton segments.

For skeleton segments that follow the "fibril" centerline (labeled 1 in Fig. 4a), the $|\Phi|$ is low (< 10 degrees), having a low DTV variability (blue profile, labeled 1 in Fig. 4b). The subsequent width measurements do not vary much along the skeleton segment and are representative of that fibril. In contrast, for the spurious segments (labeled 2 and 3), $|\Phi|$ is high, varying between 30 and 50 degrees. Here, the DTV changes considerably along the length of the skeleton segment. In most cases, such skeleton segments do not track the center line of a given fibril and thus are not representative of any fibril. FACT removes segments with high $|\Phi|$ from the analysis and their corresponding junction point from the analysis.

A general approach for selecting SST and SSF values is as follows. For SST, it is recommended to initially use a value between 10 and 30 percent. Going below 10% increases the risk of having a higher fraction of skeleton counts near junction points, while going above 30% increases the risk of removing too many skeleton counts. For SSF, it is recommended to use a value between 20 and 40 degrees. Going below 20 degrees increases the risk of removing skeleton segments of fibrils, fibril branching, and fibril network structures, while above 40 degrees increases the risk of missing spurious skeleton segments produced from edge defects. To optimize, complete a sensitivity analysis on a representative image using the following ranges of SST (0 to 40%) and SSF (7.5 to 60 degrees), and assess the extent of changes to measured properties (mean width, standard deviation, histogram profile, counts, etc.). The optimal SST and SSF values will depend on image resolution, segmentation quality, and the fibril distribution density on the substrate (level of branching and extent of the fibril network structure).

For the current study, a sensitivity analysis was conducted, varying SST (0, 10, 20, 40%) and SSF (7.5, 15, 30, 60 degrees) for a simulated CNF structure (Fig. S4), a low-branching CNF (Fig. S5), and a high-branching CNF (Fig. S6). For the simulated structure the only influence was a decrease in skeleton counts with increasing SST. For the low-branched CNF, there was minimal influence on mean width, but a decrease in standard deviation and skeleton counts with increasing SST. For the high-branched CNF, the effects of SST and SSF were more noticeable. The mean width decreased with increasing SST and decreasing SSF, but the overall change was less than 10%. The following SST values were used: 10%, 30%, or 50% for simulated fibril structures, 10% for low-branched CNF images, and 25% for highly branched CNF images. The following SSF values were used: 10 degrees for simulated fibril structures, 20 degrees for low-branched CNF images, and 25 degrees for highly branched CNF images.

Measurement & analysis

FACT measures several features within each SEM image: the pixel size (nm/pixel), the number of junction points, the number of skeleton segments, the number of pixels within skeleton segments, the

Euclidean length of each skeleton segment, and the corresponding fibril width value for every pixel within all skeleton segments. In FACT, the width of a fibril is defined as twice the radial distance from a skeleton pixel to the nearest background pixel (e.g., the outer edge of the fibril). Fibril radial values are extracted from the distance transform image of the unfiltered binary using the skeleton pixel coordinates. The limit of FACT width measurement resolution is 2 pixels, a consequence of defining width as doubling the radial distance from the skeleton segment to its outer edge. Therefore, a fibril must be two pixels wider than another fibril for FACT to detect a width difference between them. For this reason, the histograms presented in this report will have a bin width of 2 times the image pixel size (rounded to the nearest integer or multiple of 10). The effective width resolution is dependent on pixel size (nm/pixel). Higher-resolution imaging or imaging at higher magnifications decreases pixel size, which subsequently results in higher resolution fibril width measurements. In contrast, digital zooming (post-processing) of captured images does not influence pixel size and will not influence the resolution of fibril width measurements completed by FACT.

Three approaches to analyzing FACT data are considered: skeleton segments, skeleton junction points, and individual pixels. However, relating these measurements to specific features that describe fibril and network structures can be problematic. The issue stems from variability in skeleton junction formation, which depends on the quality of image segmentation, skeletonization, and skeleton refinement, as well as the complexity of the fibrils and network structures. In general, as the fraction of skeleton junction locations deviates from fibril branching or overlapping points, the less relevant the skeleton junction and segment analysis are in describing the fibril material analyzed. For simulated fibril structures, skeleton junctions are predominantly at branching and overlapping points, and minimal SST and SSF are needed. Thus, it is probable that skeleton junction analysis can be strongly related to branching and/or network density, while the skeleton segment analysis can be related to fibril length and width. In contrast, the grayscale NegC-SEM images of CNFs will exhibit segmentation errors in regions of ambiguity, either between fibrils or at the fibril/substrate interface, and the CNF

network can be very complex, requiring more aggressive SST and SSF. These errors significantly increase the number of skeleton junctions produced at locations unrelated to fibril branching or crossover points. Additionally, fibrils will be sectioned into multiple skeleton segments, and thus, skeleton segments no longer represent the length or number of fibrils within the system.

The FACT width measurement is effectively a length-weighted distribution. The resulting width distribution represents the fibril width across all skeleton segments. The resulting data is not relatable to a specific fibril but to the "fibril population" within the image. Every fibril will be measured multiple times along its length. The number of measurements per fibril is based on the number of pixel counts along the skeleton segment of that fibril. An object's skeleton pixel count is proportional to its length, aspect ratio, and inversely proportional to the number of junction points within the object. The FACT analysis can be used in conjunction with manual image analysis, which gives a number-weighted distribution.

FACT can be applied to either individual images or multiple images as a batch. Batch image analysis is a sequential process that employs a consistent approach and parameters for segmentation, morphological thinning, and skeleton refinement. FACT outputs a statistical values table, a width histogram, a box-whisker chart for each image within the batch, and skeleton overlay images (as displayed in this manuscript). Results are exported in Excel (.xlsx) or tab-delimited (.csv) formats. The user can group the statistical results to compare the distribution of results between sample groups from the same batch of images. Additionally, users are welcome to modify the FACT code to measure features beyond those listed above.

Results and discussion

FACT was developed to analyze CNF particles from NegC-SEM images and give a detailed assessment of the fibril width distributions of the hierarchical branching and entangled fibril structures typical of CNF materials. The FACT approach was systematically investigated to validate the effectiveness of fibril identification and width measurement of 1) simulated fibril structures, 2) micrographs of uniform diameter wires, 3) NegC-SEM images of low-level branching

CNF material by Beaumont et al. (2021), and 4) NegC-SEM images of high-level branching CNF material by Ringania et al. (2022). Additionally, the relevance of skeleton segment and junction analysis for describing fibril length, branching, and network density is discussed.

Validation with idealized structures

The effectiveness of the FACT approach on width measurements was assessed and validated using idealized simulated fibril structures. Simulated fibril structures were used to assess the effects of object aspect ratio (Fig. S7), length (Fig. S8), shape (Fig. S9), orientation (Fig. S10, S11), junction points (Fig. S12), hierarchical level branching using rectangular segments (Fig. 5), and hierarchical level branching using curved segments (Fig. 6) on the FACT analysis. The geometry of an object affects its skeletonization, which subsequently alters the number of pixels from which measurements are made, influencing the resulting width statistical analysis. The effect of object aspect ratio on the skeleton counts results from the morphological thinning during skeletonization, in which pixel removal around the periphery of the object is limited by the smaller dimension as the periphery converges into a central line. In general, as an object's aspect ratio increases (Fig. S7), its skeleton segment length increases, which results in a greater fraction of pixel counts contributed by the object. However, for a given fixed aspect ratio, the number of skeleton pixel counts as a fraction of the object length is slightly higher for shorter objects as compared to longer objects (Fig. S8). In addition, objects with rounded ends will contribute a slightly higher fraction of pixel counts than objects with square ends of the same aspect ratio and length (Fig. S9). These effects are reduced when comparing fibril segments of higher aspect ratios (greater than 10). When analyzing CNF materials, the variations in aspect ratio, fibril length, and fibril end shape are not expected to have a significant influence, as fibril aspect ratios are commonly greater than 10, and ends are predominantly tapered.

Object orientation with respect to the image coordinate axis (e.g., pixel axis) can affect object skeleton pixel counts and, to a small extent, the width measurements. The number of pixels comprising a given skeleton segment decreases with increasing object

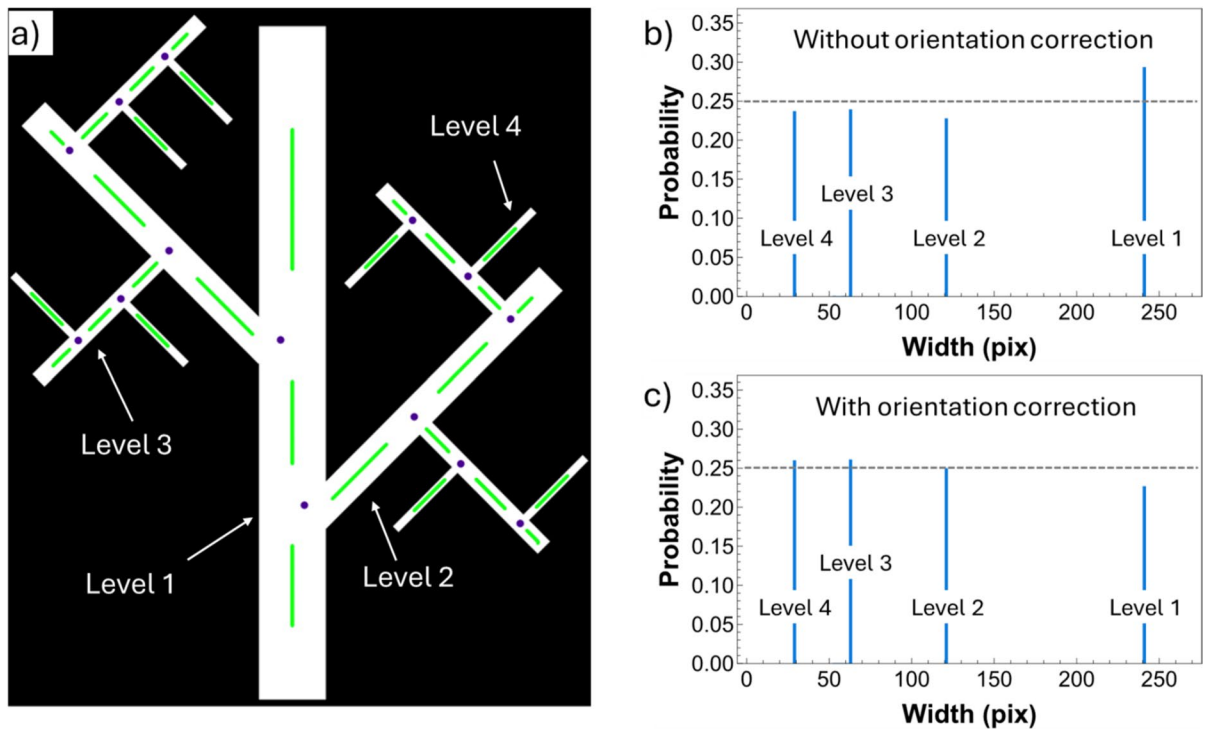


Fig. 5 Simulated four-level hierarchical branched structure using rectangular segments (image size of 2107×2468 pixels). **a** FACT analyzed image with skeleton segment (green lines) and junction (purple dots) overlay. **b** FACT width assessment

without orientation correction. **c** FACT width assessment with orientation correction. The histogram bin width is 2 pixels (limit of FACT width measurement resolution). SST = 50% and SSF = 10°

misalignment from the image coordinate axis and thus will be underrepresented in the final width measurement distribution of the analyzed structure. The minimum occurs at an off-axis angle of ± 45 degrees, resulting in approximately 30% lower pixel counts. A correction factor was calculated by evaluating the change in skeleton pixel counts as a function of orientation and is given in Fig. S10. Additionally, there is an edge effect at the periphery of off-axis objects, as pixels align stepwise, resulting in slight width variations along the object of approximately ± 1.25 pixels (Fig. S11).

Skeleton junction points result from fibril branching, crossover, or overlapping fibrils, and as a result of discontinuities along the fibril length (Fig S12). An object's skeleton pixel count is reduced by 3 pixels for each junction point within after partitioning the skeleton image. For the idealized fibril shown in Fig S12a, the total number of skeleton pixels was 1962, while for a fibril of exact length but containing 4 junction points within (Fig S12b), the total number of pixels

within the 5 skeleton segments was 1950. A more consequential effect of junction points created from discontinuities is that they complicate associations to branching density, network density, and relating skeleton segments to fibril lengths or to fibril counts.

A four-level hierarchical branched structure was analyzed using FACT (Fig. 5). Each branch is a rectangular object with an aspect ratio of 10; all branching levels have the same total length, and the width of the four branch levels is constant at 30, 62, 120, and 240 pixels, respectively. FACT was able to measure the distinct widths for each branch level and adjust for orientation effects on the probability values (Fig. 5b). The effect of branch orientation and connectivity (i.e., isolated versus connected objects) on skeleton counts is shown in Fig. S13. Each branch level is expected to contribute 25% of the total skeleton pixel counts to the width statistics (Fig. S13a). However, levels 2, 3, and 4 are oriented at ± 45 degrees and contribute fewer skeleton pixel counts than expected (Fig. S13b). To correct

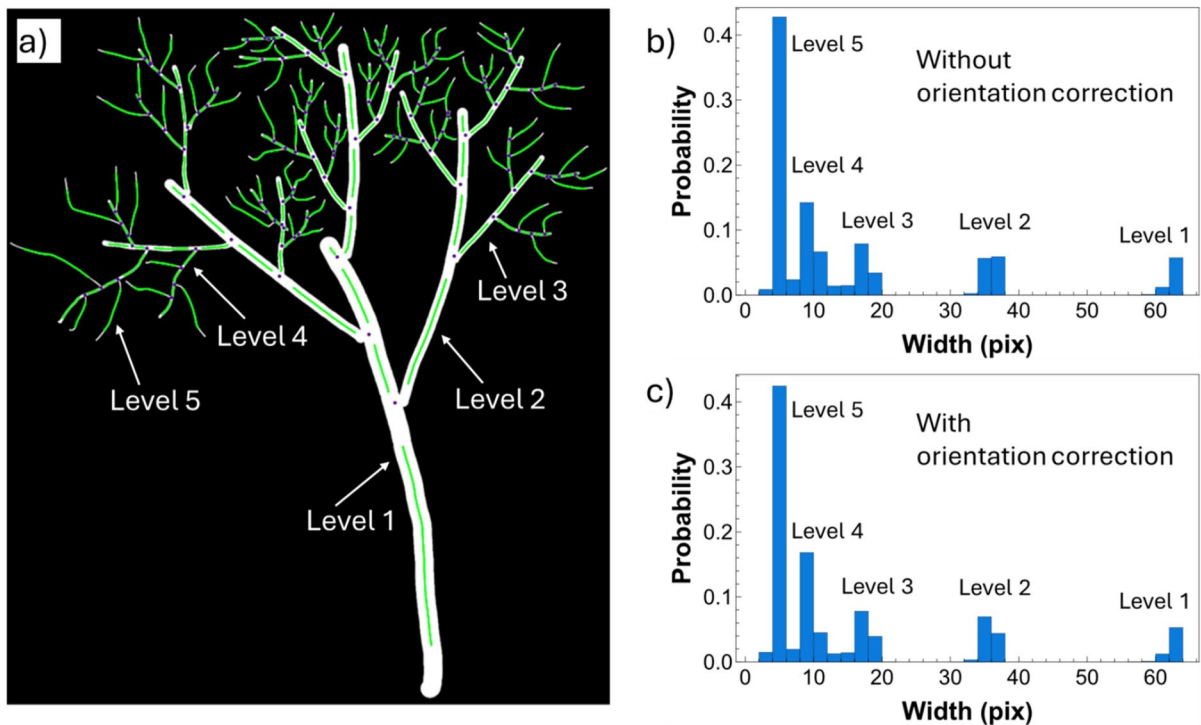


Fig. 6 Simulated five-level hierarchical branched structure using curved segments (image size of 2016×2319 pixels). **a** FACT analyzed image with skeleton segment (green lines) and junction (purple points) overlay. **b** FACT width assessment

without orientation correction. **c** FACT width assessment with orientation correction. The histogram bin width is 2 pixels (limit of FACT width measurement resolution). SST = 30% and SSF = 10°

for this orientation effect, the average angle relative to the vertical axis for each skeleton segment is calculated, and the corrective factor is then applied to resample the skeleton's pixels (Fig. S10). In the case of isolated branches, the orientation correction adjusted the skeleton counts so that each level contributed approximately 25% (Fig. S13b,f). However, for connected branches (Fig. 5b, S13c,g), the corrected skeleton counts do not match the expectation that all width levels have the same 25% probability. This discrepancy is caused by skeleton segment encroachment from higher-level branches through the branch connection boundary into lower-level branches (Fig. S14). Encroachment is problematic because this extension increases the length of the skeleton segment of a given object branch. Additionally, the width measurements from the skeleton pixels within the encroachment area do not accurately reflect the width of either branching level. For this reason, SST is used to remove skeleton pixels

between fibril branch levels by trimming skeleton segment pixels near junction points.

A five-level hierarchical branched structure, with additional structural variability and complexity (e.g., curved branches, orientation distribution, branch length distributions, overlapping fibrils) was analyzed by FACT (Fig. 6). The width of the five branch levels was set at 4, 9, 18, 36, and 62 pixels (Fig. 6). FACT measured distinct width distributions for each level, the mean and standard deviation of which were 4.5(0.6), 9.2(1.0), 17.2(1.3), 35.6(0.7), 62.3(0.8) pixels, respectively. (Fig. 6b). The orientation correction had minimal influence on width 4.5(0.6), 9.2(1.0), 17.1(1.3), 35.6(0.7), 62.3(0.8) pixels, as all the branches were off-axis to the image coordinate axis (Fig. 6c). The broader width distribution for each branching level as compared to the four-level rectangular branched structure in Fig. 5, was a result of variability in widths of the lines drawn, and the curved fibril profile.

The effect of number versus length-weighted approaches on the width measurement probability can be demonstrated by comparing manual versus FACT analysis of the simulated structures shown in Figs. 5 and 6. For manual measurements, a single measurement is made for each fibril, so the measurement distribution is based on fibril counts. In contrast, FACT makes multiple measurements on each fibril, the number of which is based on the number of skeleton segment pixels of that fibril, and thus the measurement distribution is related to fibril lengths. For example, in the simulated structure in Fig. 5a, the number of fibrils at each level (1, 2, 3, 4) is 1, 2, 4, and 8, respectively. For manual measurement, the resulting number fraction (e.g., measurement probability of 15 fibrils) at each level (1, 2, 3, 4) should be 0.067, 0.13, 0.267, and 0.533, respectively. In contrast, the FACT measurement probability for each level was approximately 0.25 (Fig. 5b and c). This result closely follows the total length distribution of 0.25 for each level (i.e., the total fibril length at each level was set to be equal). Most of the deviation from 0.25 can be attributed to skeleton segment encroachment and variation in fibril junction point count. Likewise, for the simulated structure in Fig. 6a, the number of fibrils at each level (1, 2, 3, 4, 5) is 1, 3, 9, 27, 81, respectively. For manual measurement of the total of 121 fibrils, the resulting number fraction at each level (1, 2, 3, 4, 5) should be 0.008, 0.025, 0.074, 0.223, and 0.670, respectively. The FACT measurement probability was approximately 0.06, 0.11, 0.14, 0.22, and 0.46, respectively. This result followed the total length distribution for the five levels. Most of the deviation stems from the encroachment of skeleton segments from higher to lower levels and from variation in the junction point count between fibrils. These results demonstrate that FACT effectively returns a length-weighted width distribution.

In summary, FACT width distribution measurements perform best for objects with aspect ratios greater than 10 and widths greater than 5 pixels. As fibril orientation becomes more randomized relative to the image coordinate axes, the effects of orientation correction on pixel counts and width statistics are diminished. For the hierarchical branched structures (Fig. 5 and 6), all branching levels were adequately represented by skeleton segments, and each level contributed pixels to the width histogram proportional to its total length and aspect ratio.

However, estimating branch length using pixel counts is not recommended, as discontinuities (e.g., particle edge effects, branch encroachments, and overlapping branches) influence the formation of skeleton junction points. Thus, the length of a skeleton segment, in most real cases, does not represent the length of a given fibril element. The implementation of skeleton refinement (SST and SSF) helps to remove spurious skeleton segments that form from discontinuities and trims unwanted skeleton pixels in the vicinity of junction points, reducing the influence of skeleton segment encroachment.

Validation with wire micrographs

The FACT approach was used to measure the width of small-gauge fixed-diameter wires (37 and 39 ga). The wires were placed on an optical imaging calibration standard (1 mm) and imaged with an I4-infinity optical microscope under backlighting. Image resolution was 2.03 $\mu\text{m}/\text{pixel}$. The Line tool in ImageJ (FIJI) was used to manually measure the width of the raw optical images, with measurements taken at 10 locations along each wire, perpendicular to the wire's longitudinal axis. The manual measurements yielded mean and standard deviation for the 39-ga and 37-ga wires of 92.4 μm (1.6 μm) and 124.0 μm (0.9 μm), respectively.

For the FACT analysis of these wires, images were segmented by grayscale thresholding using ImageJ. FACT was used for morphological thinning, skeleton refining, and width measurement of the wire segments. To refine the skeleton segments, an SST fraction of 0% and an SSF angle of 1 degree were used on these images. As shown in Fig. 7, the wire segments are shaded in white, with the skeleton segments overlaid as green lines. The FACT width measurements were based on 1044 and 1022 points along the skeleton segments, resulting in a mean width of 89.4 (1.9) μm and 118.2 (1.5) μm for the 37 ga and 39 ga wires, respectively. The slight discrepancy between manual and FACT measurements was mainly attributed to the noise floor of these measurements, as the wire images had a pixel size (2.03 $\mu\text{m}/\text{pixel}$), which is about 2% of the total object's width. Nevertheless, the tight width distribution (Fig. 7c and d) indicates measurement consistency.

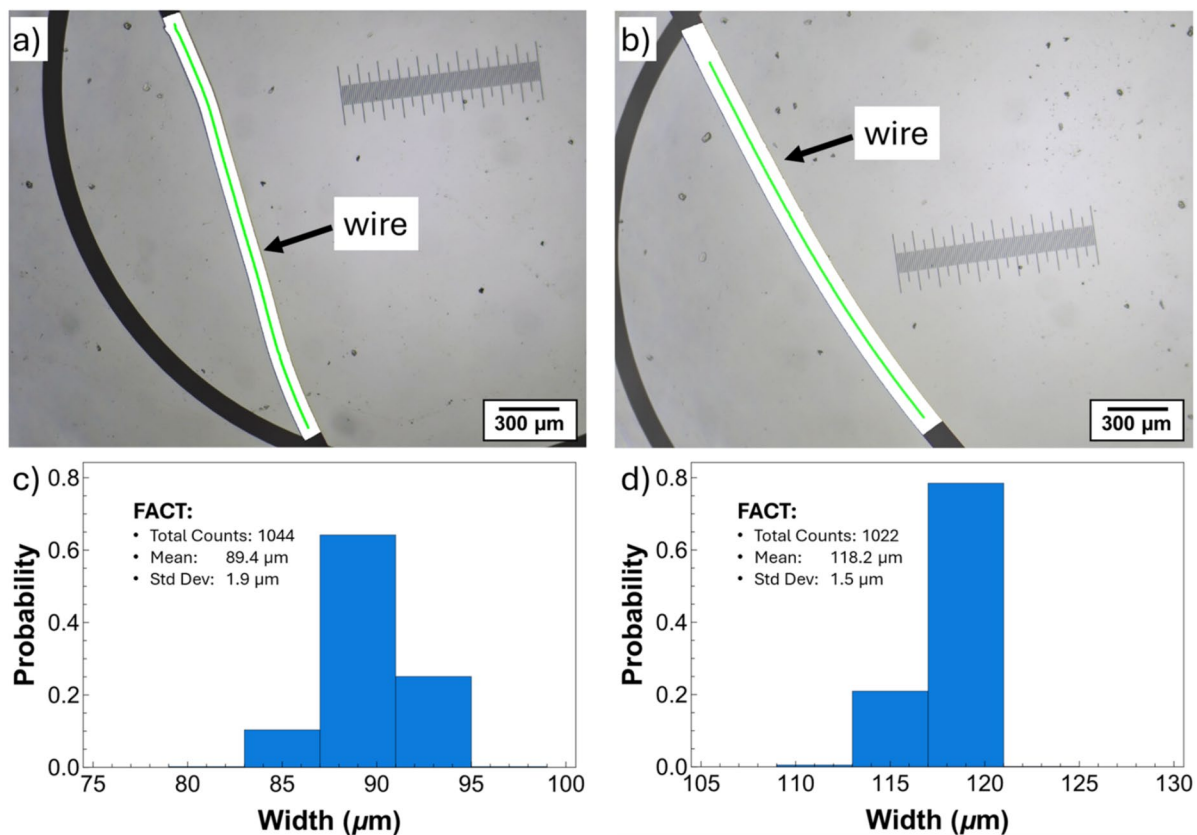


Fig. 7 FACT analyzed optical images of small gauged wires: **a, c** 37 gauge wire, and **b, d** 39 gauge wire. The wire area analyzed is shaded white, with the skeleton segment overlaid as green lines. The 4 μm histogram bin width is approximately 2

times the image pixel resolution (2.03 μm/pixel) to be consistent with the FACT width measurement resolution limit. Fibril width value outliers were not identified or removed. SST=0% and SSF=1°

Low-level branching CNFs

To assess the effectiveness of FACT in measuring the width distributions of isolated, low-level branched CNFs, the FACT approach was applied to five NegC-SEM images captured in the study by Beaumont et al. (2021). The resulting images had sufficient intensity contrast between CNFs (dark) and the substrate (medium gray), a low CNF distribution density, and a low level of CNF entanglement and branching. Such images are considered ideal for FACT analysis (Fig. 8a), enabling unambiguous segmentation of the fibrils using Weka. FACT successfully identified and segmented the CNF objects, as demonstrated by the direct overlap of the segmentation image (purple overlay) shown in Fig. 8b. Careful inspection shows neighboring substrate pixels were captured in the

fibril segmentation, resulting in a 1-to-2-pixel dilation of the fibrils and roughened fibril edge surface. This effect will increase the width measurement and promote the formation of erroneous skeleton segments and skeleton junctions. These issues could be minimized by either imaging at higher magnification to increase pixel density across the fibril width or by improving segmentation using the U-Net algorithm. Since there were insufficient images to train the U-Net, the current analysis used Weka segmentation, followed by applying a Gaussian filter with a kernel size of 2 to reduce edge roughness. The morphological thinning and skeleton refining (SST=10% and SSF=20 degrees) resulted in skeleton segments that predominantly traced the centerlines of the CNF structures, as shown via the green skeleton overlay (Fig. 8c).

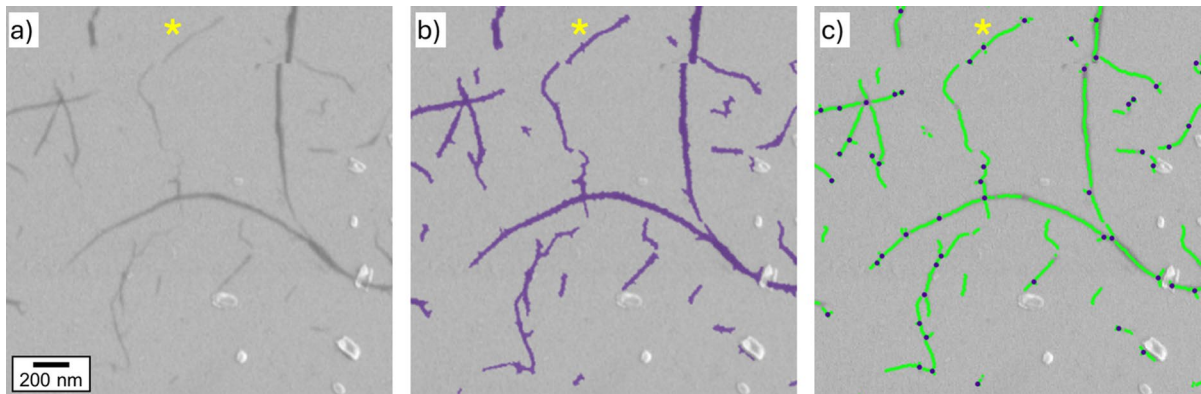


Fig. 8 FACT analysis of NegC-SEM image of low-level branched CNFs from the study by Beaumont et al. (2021). **a** As-received NegC-SEM image, digitally zoomed-in region, showing adequate contrast between CNFs and substrate background. **b** FACT analyzed image segmentation (purple) over-

laid on NegC-SEM image, showing good object identification. **c** FACT skeleton segmentation (green line overlay) and junction points (purple dots) of fibrils (both are dilated for ease of view). SST=10% and SSF=20°. Image pixel resolution: 5.43 nm/pixel

The high aspect ratio of the fibrils resulted in skeleton segments encompassing much of the fibril length. The SST and SSF skeleton-refining techniques were effective at removing many unwanted skeleton pixels. The skeleton junctions (purple dots) were created by fibril branching or overlapping, but also by other features (e.g., roughened fibril edges, kinks in fibril profile, etc.). Note that extra junction points complicate associations to the density of branching or the fibril network structure, as well as relating skeleton segments to the number and/or length of fibrils (Fig. S12), or in the case of network structures, fibril element lengths. For example, examining the unbranched fibril labeled "*" in Fig. 8, it consists of 3 junction points, which are not associated with branching or fibril crossover points, and the fibril is effectively cut into multiple skeleton segments. The ramifications are that the image analysis of this fibril could be misinterpreted. This single, isolated fibril, approximately 300 nm in length, could be misinterpreted as branching or as part of a network structure, and is composed of 6 fibril elements, each approximately 20–200 nm in length. To avoid such errors in image analysis, the authors recommend using the skeleton segment points to measure the widths of the entire fibril structure and using extreme caution when considering any branching and/or fibril element length analysis.

Manual measurements of these images were completed separately by one of their team members

as part of the Beaumont et al. (2021) study. Qualitatively, the FACT and manual image analyses were in reasonable agreement, with similar width distributions, means, and standard deviations, where the differences in means were within the image pixel resolution. Additionally, any differences can primarily be attributed to FACT being a length-weighted measure, while the manual approach is number-weighted. In the manual approach, each fibril is measured once in the mid-section using ImageJ, resulting in significantly lower counts. In contrast, the FACT approach obtains thousands of counts from the total collection of pixels that comprise the entire image skeleton.

At lower magnification (Fig. 9, having a 5.43 nm/pixel resolution), the mean width and standard deviation were: FACT 22.1 (8.9) nm, manual 22.7 (8.2) nm. The counts for FACT at 0 to 10 bin width would result from fibrils having a width of 1 pixel. At higher magnification (Fig. 10, having a 1.79 nm/pixel resolution), the mean width and standard deviation were: FACT 16.9 (5.2) nm, manual 17.1 (2.5) nm. The higher-magnification image resulted in a narrower width distribution, a lower mean width, and a lower standard deviation. This result was likely attributed to three primary factors: 1) Image acquisition in a region with finer fibrils, 2) optimized image contrast, segmentation, and skeletonization of the finer fibril regions, improving the identification of faint fibrils, and 3) improved

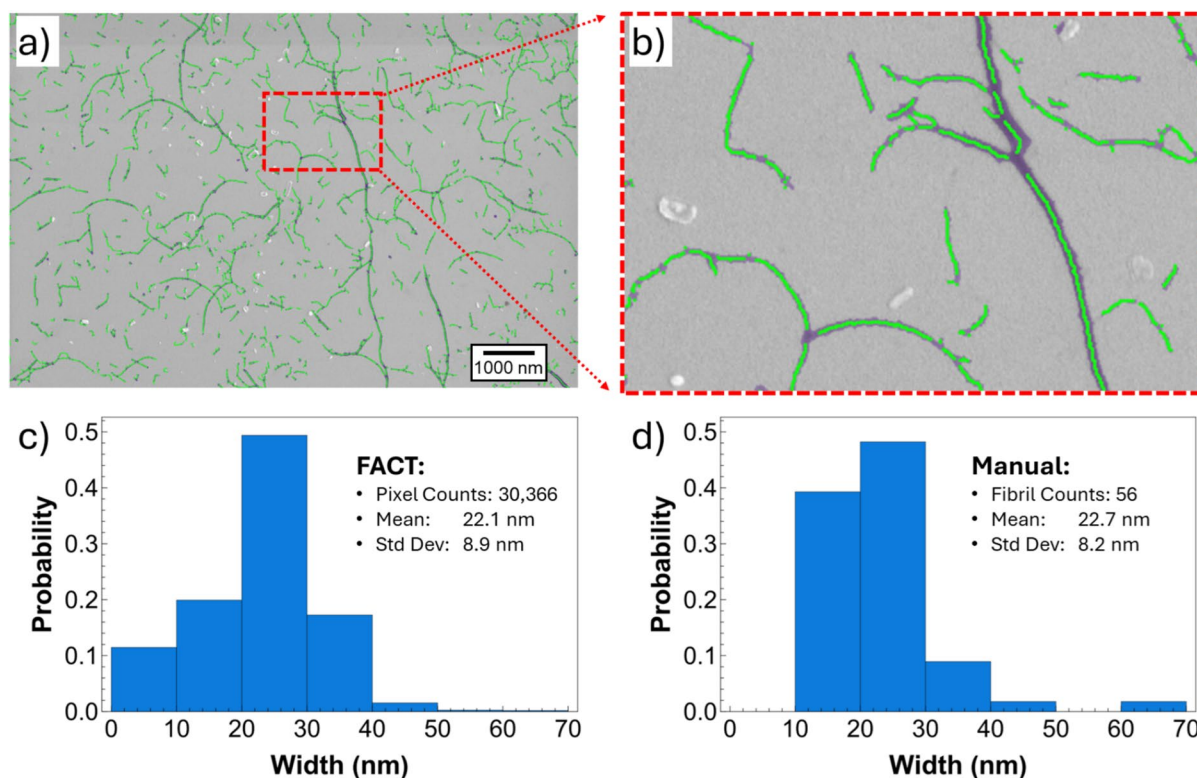


Fig. 9 Image analysis of lower magnification NegC-SEM images of low-level branched CNFs from the study by Beaumont et al. (2021). **a** FACT analyzed image with the FACT segmentation (pale purple) and the skeleton (green lines) overlays. The red rectangular overlay shows the location of the digitally zoomed regions for part b. **b** Digitally zoomed-in region showing good agreement of FACT object identification, segmentation, and skeletonization of CNF fibril structure. **c** FACT

width measurements for all pixels that make up the green skeleton overlay within part a. **d** Manual width measurement of individual fibrils. The 10 nm histogram bin width is approximately 2 times the image pixel resolution (5.43 nm/pixel) to be consistent with the FACT width measurement resolution limit. Fibril width value outliers were not identified or removed. SST = 10% and SSF = 20°

fidelity in fibril width measurement resulting from higher pixel density across the fibril width. Factor 1 pertains to the effect of sampling choice, thus highlighting the potential for analysis bias. Whereas factors 2 and 3 relate to the potential effect of measurement fidelity. The analyst needs to consider the interplay of sampling choice and measurement fidelity when analyzing measurements completed at different imaging magnifications and locations within a given sample.

The "correctness" of image analysis relies on the premise that the images analyzed provide a reasonable representation of the CNF size distribution. For example, two additional images were analyzed at a lower magnification, including larger fibrils. Image analysis of all three images (Fig. S15) resulted in a

broadening of the width distribution and increases in the mean width and standard deviation: FACT 24.8 (15.2) nm, manual 23.1 (15.2) nm. All of which indicate the inclusion of wider fibrils in the analysis, demonstrating the importance of capturing a representative range of CNF particle sizes during imaging. A typical strategy for capturing a representative range of CNF sizes is to analyze multiple images at different magnifications (Moon et al. 2023, 2025).

High-level branching CNFs

To assess the effectiveness of measuring the width distributions of highly branched and networked CNF materials, the FACT approach was applied to NegC-SEM images captured in the study by Ringania et al.

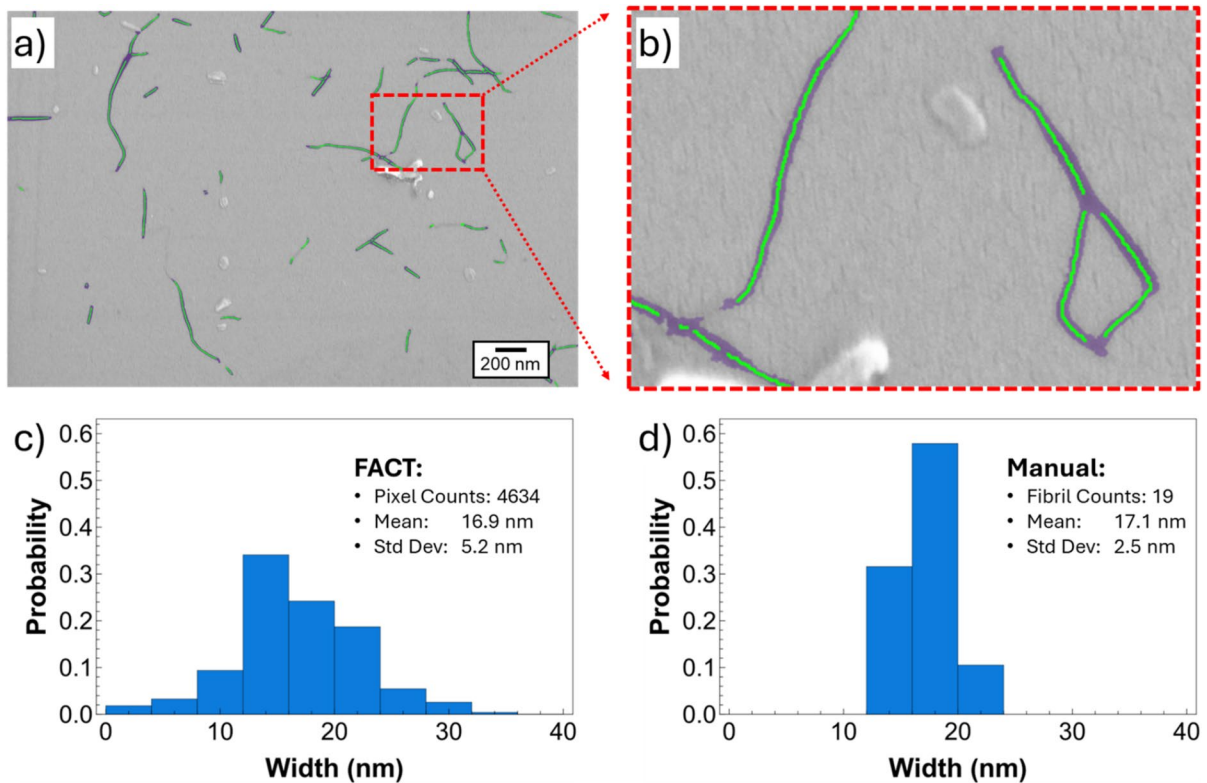


Fig. 10 Image analysis of a higher magnification NegC-SEM image of low-level branched CNFs from the study by Beaumont et al. (2021). **a** FACT analyzed image with the FACT segmentation (pale purple) and the green skeleton overlays. The red rectangular overlay shows the location of the digitally zoomed region for part b. **b** Digitally zoomed-in region showing good agreement of FACT object identification, segmentation (pale purple), and skeletonization (green line) overlays

of CNF fibril structure. **c** FACT width measurements for all pixels that make up the green skeleton overlay within part a. **d** Manual width measurement of individual fibrils. The 4 nm histogram bin width is approximately 2 times the image pixel resolution (1.79 nm/pixel) to be consistent with the FACT width measurement resolution limit. Fibril width value outliers were not identified or removed. SST = 10% and SSF = 20°

(2022), (Ringania 2023). These complex fibril structures are typical of mechanically refined wood pulp fibers (Fig. 11). The coarser CNFs and network structure required imaging at lower magnifications than the low-level branching CNF materials analyzed in the prior section. Additionally, differences in imaging practices compared to the low-level branching CNF material led to differences in image contrast and resolution. Despite these differences, the U-Net segmentation and thinning operations yielded an adequate skeleton trace of the centerline for most fibrils. To further reduce the effect of particle edge roughness, a Gaussian filter with a kernel size of 2 was applied to the resulting U-Net segmented images. For thin fibrils, the skeleton overlay showed good agreement with the CNF network (Fig. 11c and d), indicating

that the subsequent width assessment will be accurate. However, in regions of extensive fibril bundling, branching, or overlapping, as shown in the center of Fig. 11e, there is some discrepancy between the green skeleton segments and the fibril network. This discrepancy is a result of several factors: the encroachment of the higher-level branch segments into the lower-level segments (described in Fig. S14), image segmentation issues in this area due to context ambiguity, and the thinning operator's sensitivity to sharp edges along the fibril boundary, which can result in extraneous skeleton segments (Fig. 4). Such deviations will result in errors in the width assessment of these regions. To reduce measurement errors, the refining steps SST and SSF are used to eliminate pixels from encroachment regions and completely

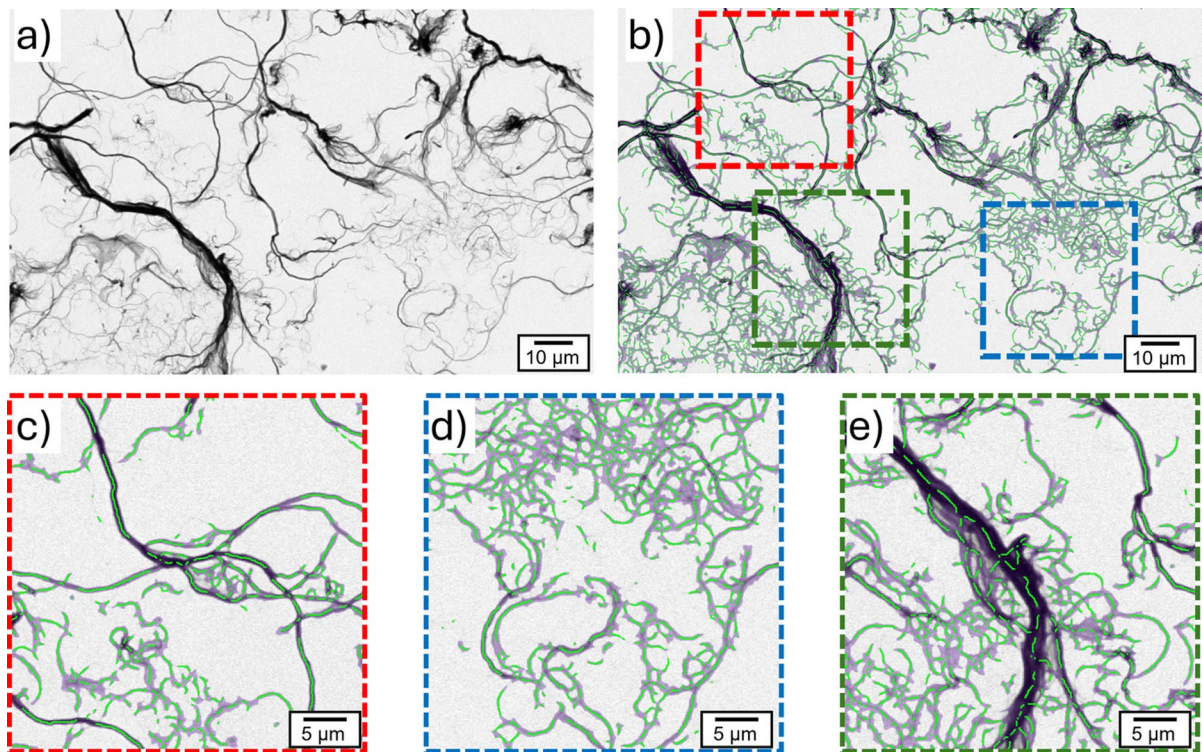


Fig. 11 FACT analysis of NegC-SEM image of high-level branched CNFs from the study by Ringania (2023). **a** Raw NegC-SEM image. **b** FACT analyzed image showing both segmentation (pale purple) and the skeleton segment (green lines) overlays. The three boxed overlays show the locations of digitally zoomed regions for parts c, d, and e. **c** and **d**: Skeleton

segment overlays showing good agreement with the thinner fibril elements. **e** Skeleton segment overlay of thick fibril bundle and network region, showing a moderate level of discontinuity in representing the actual fibril structure. SST = 25% and SSF = 25°. Image pixel resolution: 80 nm/pixel

remove spurious skeleton segments, respectively. Optimizing the SST and SSF parameters to maintain a high fraction of correct segments is imperfect, as some extraneous skeleton segments will inevitably remain, and excessive filtering can remove many correct segments. A robust approach to quantifying the fraction of unwanted skeleton segments remains unclear.

Coarser CNF fibril structures require imaging over a range of magnifications to capture both the coarse and fine fibril structures, as demonstrated in Fig. 12. Three magnifications were used, with image pixel resolutions of 131, 38, and 22 nm/pixel. The width distribution was non-Normal, being skewed towards higher widths. With higher magnification, the width distributions shifted to lower values, as did their corresponding means and standard deviations (Low: 1234 (1259) nm, medium: 628 (534) nm, and high: 436 (348) nm). The higher magnification image resulted

in a narrower width distribution, a lower mean width, and a lower standard deviation. This result was likely attributed to three primary factors: 1) Image acquisition in a region with finer fibrils, 2) optimized image contrast, segmentation, and skeletonization of the finer fibril regions, improving the identification of faint fibrils, and 3) improved fidelity in fibril width measurement resulting from higher pixel density across the fibril width. Factor 1 pertains to the effect of sampling choice, thus highlighting the potential for analysis bias. Whereas factors 2 and 3 relate to the potential effect of measurement fidelity. The analyst needs to consider the interplay of sampling choice and measurement fidelity when making measurements at different magnifications and locations within a given complex CNF bundle. Digital zooming may improve the analyst's ability to make manual measurements, but it does not help with FACT analysis. At each magnification, there are regions where the green

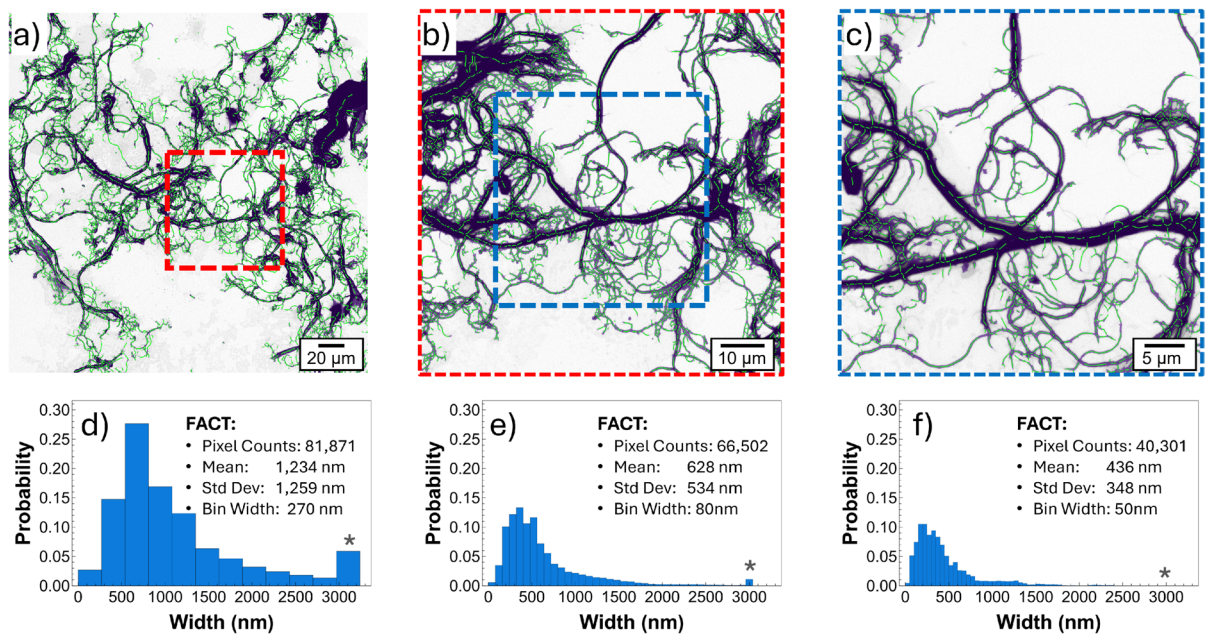


Fig. 12 FACT analysis of NegC-SEM image of high-level branched CNFs from the study by Ringania et al. (2022) at three different magnifications. **a** 131 nm/pixel, **b** 38 nm/pixel, and **c** 22 nm/pixel. FACT analyzed images with binary segmentation (pale purple) and skeleton overlay (green lines), showing good agreement of the skeleton overlay with the CNF fibril structure. **d**, **e**, and **f** FACT fibril width distribution for the corresponding three magnifications, showing a shift

to smaller fibril widths for higher magnification and image positioning to finer fibril regions. * Bin for all width measurements greater than 3000 nm. The histogram bin width is approximately 2 times the image pixel resolution to be consistent with the FACT width measurement resolution limit. For parts d, e, and f, the bin widths are 270 nm, 80 nm, and 50 nm, respectively. Fibril width value outliers were not identified or removed. SST = 25% and SSF = 25°

skeleton segments do not perfectly match the coarse fibril bundling, extensive branching, and overlapping fibrils. The skeleton segments are fragmented, resulting from encroachment of the higher-level skeleton into the lower-level segments and spurious skeleton segments produced by sharp edges along the fibril boundary. These problematic effects were reduced by using SST = 25% and SSF = 25° to trim encroachment regions and remove most spurious skeleton segments, respectively (Fig. 12c).

The FACT width measurements were qualitatively compared to the manual measurements reported by Ringania et al. (2022) for the three images shown in Fig. S16. The contrast between the length-weighted (FACT) and the number-weighted (manual) measurement approaches is evident in comparing Fig. 13a and Fig. 13b. FACT analysis comprises all skeleton pixels. The resulting data is a function of the entire "fibril network" captured within the analyzed images, providing an effective length-weighted distribution (Fig. 13c). The width distribution spans from 22 to

5160 nm, having a broad peak ranging from 50 to 600 nm, and a mean and standard deviation of 510 (532) nm, based on a total of 104,891 pixel counts. In contrast, the manual measurements were of individual fibrils or branching fibril elements with a single measurement using ImageJ, the location of which are shown as red line overlays in Fig. 13b, resulting in a number-weighted distribution (Fig. 13d). The width distribution spans from 64 to 4226 nm, having a broad peak ranging from 100 to 450 nm, and a mean and standard deviation of 320 (310) nm, based on 978 fibril elements measured. The overlap between the two approaches in the peak range (100 nm to 450 nm) indicates consistency. By considering FACT and manual approaches as complementary, additional insights into the CNF particle and network morphology may be gained.

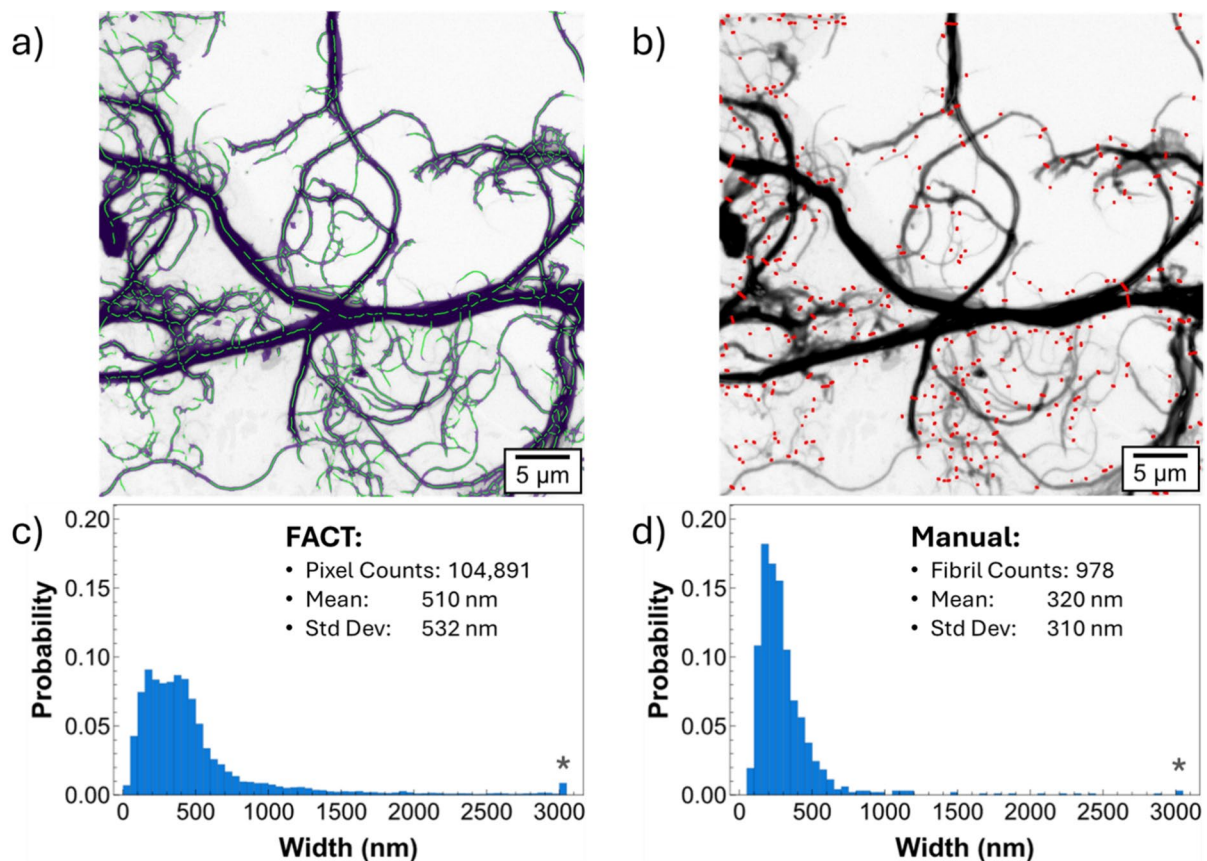


Fig. 13 Comparison between FACT and manual fibril width measurement of high-level branched CNFs from the study by Ringania et al. (2022). **a** FACT analysis showing the skeletonization of the fibril network (green lines). Pixels along these lines are used to measure width. **b** Manual analysis showing the fibril measurement locations (red lines are dilated for easier viewing). The width distributions for **c** FACT analysis and **d** manual analysis. The width histograms are based on all

fibril width counts pooled from the image analysis of the three images in Fig. S16. * Bin for all width measurements greater than 3000 nm. The 50 nm histogram bin width is approximately 2 times the image pixel resolution (22.0 nm/pixel) to be consistent with the FACT width measurement resolution limit. Fibril width value outliers were not identified or removed. SST=25% and SSF=25°

Network assessment

Relating FACT measurements (skeleton segments, skeleton junction points, and individual pixels) to specific features, such as fibril length, branching density, and network density, can be problematic. Skeleton junction point formation dictates the viability of such assessments. The issue is that skeleton junction formation depends on the quality of image segmentation and the complexity of fibril and network structures. In cases where most junction points result from fibril intersections or branching (Fig. 3), it may be possible to correlate the number of junction points with the density of the fibril intersection network or the level

of fibril branching. Likewise, if SST was set to 0, then the skeleton segments could be used to estimate fibril lengths or network segment lengths. However, as the proportion of junction points resulting from edge defects or segmentation imperfections increases, the relevance of the junction-point analysis decreases. For the low-branching CNF material, there were sufficient defects in the segmentation that most junction points were from defects as opposed to fibril branching or intersections (Fig. 8). This was worse for the high-level branched CNF material. Currently, FACT lacks a practical approach for identifying and removing all erroneous junction points. A potential area of future work would be the incorporation of additional

ML step(s) that use supervised learning to differentiate between correct and incorrect junction points.

FACT analysis time

A complete FACT analysis of a new image data set can take between 4 and 10 h, depending on several factors, such as the number of images, image size, number of fibers per image, image segmentation approach (e.g., gray value threshold, Weka, or U-Net CNN), and the user's computer specifications. Analysis time can be categorized into several subtasks that follow the FACT flow chart in Fig. 1. The initial pre-analysis of incoming raw images takes ~1 min/image, in which FACT obtains pixel size using the scale marker within the image, trims the image to remove any boundary or banners, and saves the edited image as a new image file.

The edited gray images can be segmented using Weka, an ImageJ plugin, which can take ~2 h for an initial round of 4 images. If the user has only a small number of images to analyze and determines that the Weka segmentation results were adequate, they can continue to use FACT for width statistical assessment. However, if the user has a larger data set and determines that the segmentations could benefit from training a convolutional neural network, the Weka segmentation can be used as ground truth for training a U-Net CNN. The time required to train a U-Net CNN depends on the computer's specifications, the number of ground truth images, and the number of selected training rounds. This part could take between 2 and 8 h. Finally, the FACT width analysis is fast, taking less than 2 min per image, which includes the computing of various measurements for each skeleton segment and junction points (e.g., counts, point coordinates, distance transform values, centroid, and index), producing image overlays (e.g., segmentation, skeleton segment, junction points), and exporting the width data.

Although a complete FACT analysis for a new dataset can take a reasonable amount of time (e.g., 4 to 10 h), if the U-Net CNN is already trained, FACT can be used without additional training, significantly reducing analysis time to less than 5 min per image. For the high-branched example, the total FACT analysis time for 3 images shown in Fig. S16 was less than 15 min, while manual measurements took over 8 h to measure 978 fibrils. A pre-trained CNN

can be used if the incoming images have features, contrast values, and noise levels similar to the original training images. Out of an abundance of caution, it is recommended to run a test using a pre-trained network to verify that FACT performs segmentation and skeletonization adequately. Otherwise, it is possible to continue training the U-Net CNN by adding additional ground-truth images to the existing training dataset. Training with images of similar contrast, intensity, and noise is recommended.

Conclusion

A semi-automated image analysis framework, Fibril Analysis for Cellulose Technology (FACT), was developed to measure the width distribution of fibril elements that make up cellulose nanofibril (CNF) particles. The FACT framework was validated using idealized geometries of hierarchical branched structures, and simple wire micrographs, and was then applied to negative-contrast scanning electron microscopy (NegC-SEM) images of two contrasting CNF morphologies. The two CNF case studies not only showed differences in particle morphology (low-level versus high-level branching) but also in imaging conditions that led to fundamentally different contrast mechanisms and spatial resolutions. The machine learning (ML) toolsets used (Weka and U-Net) provided greatly improved segmentation compared to traditional grayscale thresholding. The high-contrast NegC-SEM images, combined with ML segmentation tools, enabled FACT to capture and skeletonize the majority of the CNF hierarchical branching structure, spanning length scales from the micron-sized CNF object to the nano-sized fibril features. FACT image analysis was most effective for high-contrast images containing fibrils with aspect ratios greater than 10 and widths of at least 5 pixels.

FACT image analysis provides a length-weighted measure that complements the number-weighted measure of manual image analysis. The key difference is that FACT measures multiple points along each fibril, whereas manual analysis measures a single point per fibril. Thus, FACT can potentially enable a more detailed assessment of the entire CNF branching and network structure. The number of measurements per fibril is based on the number

of pixel counts along the skeleton segment of that fibril, which is proportional to the fibril length, aspect ratio, and is slightly reduced by the number of junction points within the fibril. Skeleton refinement using skeleton segment trimming (SST) and skeleton segment filtering (SSF) was used to minimize inconsistencies when acquiring fibril width information from centerline tracing. Despite this, using pixel counts to estimate fibril or fibril branch length is not recommended due to the inconsistent nature of how discontinuities affect junction point formation and skeleton segment encroachments. Additionally, great caution is needed when using junction-point analysis to assess fibril network or branching density metrics, as junction-point misidentification is often too high.

This study demonstrated that FACT could be successfully applied to two contrasting CNF morphologies (i.e., low- and high-level branching) and yielded effective segmentation, skeletonization, and fibril width measurement. The FACT width measurement results were generally consistent with the manual measurements. A complete FACT training and analysis can take 4 to 10 h. However, if the U-Net CNN is already trained, FACT can be used without additional training, significantly reducing analysis time to less than 5 min per image. The FACT code is publicly available in Zenodo.

Acknowledgments The authors recognize Dr Ishant Tiwari from Chemical and Biomolecular Engineering at Georgia Institute of Technology for optical micrographs used in Fig. 7. The authors also recognize Dr. Bruno Mattos from the Department of Bioproducts and Biosystems at Aalto University for providing NegC-SEM images of their low-level branched CNF material and for conducting the subsequent manual measurements. We also gratefully acknowledge Dr Nayomi Plaza-Rodrigues (Forest Products Laboratory), Dr Nathan Bechle (Forest Products Laboratory), Dr Warren Batchler (Monash University), and Dr Linda Johnston (National Research Council Canada) for critical reading of the document.

Author contributions U.R., C.B., R.M., and S.B developed the study conception. R.M., C.B., and U.R developed the study design. C.B developed coding design. C.B., R.M., and U.R. completed code testing and verification. R.M., C.B., and U.R. completed the results analysis. U.R completed NegC-SEM imaging of high-level branched CNF material and subsequent manual width measurements. R.M., C.B., and U.R. wrote the main manuscript text. All authors critically reviewed draft manuscript(s), which resulted in revisions to the document. All authors have read and approved the final manuscript.

Funding This work was supported by the USDA Forest Service—Forest Products Laboratory, and by both the Renewable Bioproducts Institute and the Brook Byers Institute for Sustainable Systems at Georgia Institute of Technology.

Data availability The FACT code is publicly available in Zenodo (<https://doi.org/10.5281/zenodo.18176747>). A trained U-Net CNN is available in (<https://doi.org/10.5281/zenodo.18176828>)

Declarations

Conflicts of interests The authors declare no conflict of interest.

Consent for publication The authors consent to publication by the journal.

References

- Ang S, Narayanan JR, Kargupta W, Haritos V, Batchelor W (2020) Cellulose nanofiber diameter distributions from microscopy image analysis: effect of measurement statistics and operator. *Cellulose* 27:4189–4208. <https://doi.org/10.1007/s10570-020-03058-0>
- Beaumont M, Otoni CG, Mattos BD et al (2021) Regioselective and water-assisted surface esterification of never-dried cellulose: nanofibers with adjustable surface energy. *Green Chem* 23:6966–6974. <https://doi.org/10.1039/d1gc02292j>
- Chen Y, Zhang L, Yang Y, Pang B, Xu W, Duan G, Jiang S, Zhang K (2021) Recent progress on nanocellulose aerogels: preparation, modification, composite fabrication, applications. *Adv Mater* 33(11):2005569
- Desmaisons J, Boutonnet E, Rueff M et al (2017) A new quality index for benchmarking of different cellulose nanofibrils. *Carbohydr Polym* 174:318–329. <https://doi.org/10.1016/j.carbpol.2017.06.032>
- Eichhorn SJ, Etale A, Wang J et al (2022) Current international research into cellulose as a functional nanomaterial for advanced applications. *J Mater Sci* 57(10):5697–5767. <https://doi.org/10.1007/s10853-022-06903-8>
- Grießer A, Westerteiger R, Glatt E, Hagen H, Wiegmann A (2023) Identification and analysis of fibers in ultra-large micro-CT scans of nonwoven textiles using deep learning. *J Text Inst* 114(11):1647–1657. <https://doi.org/10.1080/00405000.2022.2145429>
- Hotaling NA, Bharti K, Kriel H, Simon CG (2015) DiameterJ: a validated open source nanofiber diameter measurement tool. *Biomaterials* 61:327–338. <https://doi.org/10.1016/j.biomaterials.2015.05.015>
- Huang X, Wen D, Zhao Y, Wang Q, Zhou W, Deng D (2016) Skeleton-based tracing of curved fibers from 3D X-ray microtomographic imaging. *Results Phys* 6:170–177. <https://doi.org/10.1016/j.rinp.2016.03.008>
- ISO (2014) ISO 16065–2 Pulps—Determination of fiber length by automated optical analysis—Part 2: unpolarized light method

- ISO (2021a) Nanotechnologies—Characterization of individualized CNF samples ISO/TS 21346:2021
- ISO (2021b) Nanotechnologies—Particle size distribution for cellulose nanocrystals ISO/TS 23151:2021
- ISO (2023) Nanotechnologies—Vocabulary for cellulose nanomaterial ISO/TS 20477:2023
- Kangas H, Lahtinen P, Sneek A, Saariaho AM, Laitinen O, Hellén E (2014) Characterization of fibrillated celluloses. A short review and evaluation of characteristics with a combination of methods. *Nord Pulp Pap Res J* 29(1):129–143. <https://doi.org/10.3183/NPPRJ-2014-29-01-p129-143>
- Li MC, Wu Q, Moon RJ, Hubbe MA, Bortner MJ (2021a) Rheological aspects of cellulose nanomaterials: governing factors and emerging applications. *Adv Mater* 33(21):2006052. <https://doi.org/10.1002/adma.202006052>
- Li T, Chen C, Brozena AH et al (2021b) Developing fibrillated cellulose as a sustainable technological material. *Nature* 590(7844):47–56. <https://doi.org/10.1038/s41586-020-03167-7>
- Mattos BD, Tardy BL, Rojas OJ (2019) Accounting for substrate interactions in the measurement of the dimensions of cellulose nanofibrils. *Biomacromol* 20(7):2657–2665. <https://doi.org/10.1021/acs.biomac.9b00432>
- Moon RJ, Martini A, Nairn J et al (2011) Cellulose nanomaterials review: structure, properties and nanocomposites. *Chem Soc Rev* 40:3941. <https://doi.org/10.1039/c0cs00108b>
- Moon RJ, Hensdal CL, Beck S et al (2023) Setting priorities in CNF particle size measurement: What is needed vs. what is feasible. *Tappi J* 22(2):116–137. <https://doi.org/10.32964/TJ22.2.116>
- Moon R, Johnston L, Land-Hensdal C, Batchelor W (2025) Perspectives on cellulose nanofibril size measurement using scanning electron microscopy. *Cellulose* 32:2793–2810. <https://doi.org/10.1007/s10570-025-06458-2>
- Nakayama K, Kumagai A, Sakakibara K (2025) Machine learning-assisted sedimentation analysis of cellulose nanofibers to predict the specific surface area. *Carbohydr Polym Technol Appl* 9:100697. <https://doi.org/10.1016/j.carpta.2025.100697>
- Nečas D, Klapetek P (2012) Gwyddion: an open-source software for SPM data analysis. *Open Phys* 10(1):181–188. <https://doi.org/10.2478/s11534-011-0096-2>
- Persson NE, Rafshoon J, Naghshpour K et al (2017) High-throughput image analysis of fibrillar materials: a case study on polymer nanofiber packing, alignment, and defects in organic field effect transistors. *ACS Appl Mater Interfaces* 9:36090–36102. <https://doi.org/10.1021/acsami.7b10510>
- Ringania U, Harrison J, Moon RJ, Bhamla MS (2022) Dewatering of cellulose nanofibrils using ultrasound. *Cellulose* 29:5575–5591. <https://doi.org/10.1007/s10570-022-04626-2>
- Ringania U (2023) Dewatering of cellulose nanofibrils using ultrasound. PhD dissertation, Georgia Institute of Technology. <https://hdl.handle.net/1853/72753>
- Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) *Medical image computing and computer-assisted intervention—MICCAI 2015 lecture notes in computer science*. Springer, Cham
- Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675. <https://doi.org/10.1038/nmeth.2089>
- Sokolov PA, Belousov MV, Bondarev SA, Zhouravleva GA, Kasyanenko NA (2017) FibrilJ: ImageJ plugin for fibrils' diameter and persistence length determination. *Comput Phys Commun* 214:199–206. <https://doi.org/10.17632/ndxb93h4vc.1>
- UMaine The Process development Center. <https://umaine.edu/pdc/nanocellulose/>. Accessed November 2024
- Usov I, Mezzenga R (2015) FiberApp: an open-source software for tracking and analyzing polymers, filaments, biomacromolecules, and fibrous objects. *Macromolecules* 48:1269–1280. <https://doi.org/10.1021/acs.chemmater.6b01825>
- Verho T, Turpeinen T, Asad F, Immonen K (2025) A skeletonization-based approach for individual fiber separation in tomography images of biocomposites. *Comput Mater Sci* 246:113372. <https://doi.org/10.1016/j.commatsci.2024.113372>
- Willhammar T, Daicho K, Johnstone DN et al (2021) Local crystallinity in twisted cellulose nanofibers. *ACS Nano* 15:2730–2737. <https://doi.org/10.1021/acsnano.0c08295>
- Wolfram Research (2008), HitMissTransform, Wolfram Language function, <https://reference.wolfram.com/language/ref/HitMissTransform.html> (updated 2012)
- Wolfram Research (2010a), Thinning, Wolfram Language function, <https://reference.wolfram.com/language/ref/Thinning.html>
- Wolfram Research (2010b), ComponentMeasurements, Wolfram Language function, <https://reference.wolfram.com/language/ref/ComponentMeasurements.html> (updated 2017).
- Zambrano F, Starkey H, Wang Y et al (2020) Using micro- and nanofibrillated cellulose as a means to reduce weight of paper products: a review. *BioResources* 15(2):4553–4590. <https://doi.org/10.15376/BIORES.15.2.ZAMBRANO>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.