



Delft University of Technology

Document Version

Final published version

Licence

CC BY-NC

Citation (APA)

Hofman, D. E., Zhang, R., Oltmans, R., Hendriks, J. M., Moor, C. C., Wijsenbeek-Lourens, M., & Jung, J. (2026). A machine learning approach to identify perspectives of people living with pulmonary fibrosis. *ERJ Open Research*, 12(5), Article 01836-2025. <https://doi.org/10.1183/23120541.01836-2025>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

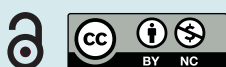


A machine learning approach to identify perspectives of people living with pulmonary fibrosis

Copyright ©The authors 2026

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 21 Dec 2025
Accepted: 30 March 2026



To the Editor:

Pulmonary fibrosis is a fibrotic interstitial lung disease that is caused by a heterogeneous group of underlying aetiologies [1]. People with pulmonary fibrosis often experience debilitating symptoms, including dry cough, dyspnoea and fatigue, that negatively influence the quality of life of patients [2, 3] and their families [4]. Understanding perspectives of people with pulmonary fibrosis is important to address their unmet care needs.

Social media are used by many patients and their families to search for information and share insights with peers about their disease experiences. Social media can serve a valuable source for assessing patients' perspectives [5], beyond structured interviews and surveys. Analysing these online conversations through social media analysis may provide an opportunity to understand patient perspectives throughout the disease course. A machine learning approach was previously used to support topic modelling of social media posts in other areas and could be useful to identify underlying patterns in large data sets.

The objective of this study is to identify perspectives of people living with pulmonary fibrosis by analysing social media posts with the use of machine learning. We aimed to validate this machine learning approach by applying human qualitative analysis.

Posts (both main threads and replies) from the platform “Living with pulmonary fibrosis” from Inspire.com (<https://www.inspire.com/groups/living-with-pulmonary-fibrosis/>) were used as a data source with permission of the platform owner. Prior to the analysis, the dataset was pseudonymised by removing content including family names, dates and medicine dosages. An unsupervised machine learning technique called BERTopic [6] was used to extract topic clusters from the posts, involving hierarchical density-based clustering [7] to find underlying patterns. The optimal number of clusters identified by BERTopic was 315. Model performance was evaluated through a human-in-the-loop validation process to assess for overlap and to determine the optimal number of topic clusters. To achieve an optimal balance between topic diversity and information loss, the number of topic clusters was iteratively reduced from 315 to 100. The reduction process combined similarity matrices, hierarchical plots [7] and human review of keywords related to each topic cluster. Similarity matrices were used to evaluate the distinctiveness between clusters, while hierarchical plots illustrated their structure, enabling assessment of potential information loss when merging similar clusters. Subsequently, all topic clusters were manually labelled by two researchers (one medical doctor and one computer scientist) through reading the keywords related to the topic cluster and the top 50 posts within that topic cluster. Discrepancies in label names were solved by a third researcher (a pulmonologist). The research team collectively reviewed the list of topic clusters in a research meeting, manually merging clusters where overlap was identified. This resulted in a final set of 80 topic clusters. These topic clusters were subsequently grouped into themes, with input and collaboration from the entire research team.

A sentiment analysis was performed by using a pre-trained BERT model. This creates a sentiment (negative, neutral or positive) per post. Sentiment scores were calculated for topic clusters by subtracting the number of negative posts from the number of positive posts, divided by positive, neutral and negative posts combined. Sentiment scores range from -1 to 1. We considered sentiment scores below -0.5 as negative and above 0.5 as positive.

Shareable abstract (@ERSpublications)

People with PF and their families share disease experiences and information on social media, providing a novel source for insights in their perspectives. While machine learning analysis shows promise, further refinement is needed to enhance accuracy. <https://bit.ly/4cezYmR>

Cite this article as: Hofman DE, Zhang R, Oltmans R, *et al.* A machine learning approach to identify perspectives of people living with pulmonary fibrosis. *ERJ Open Res* 2026; 12: 01836-2025 [DOI: 10.1183/23120541.01836-2025].

TABLE 1 Results of the topic cluster analysis of posts (Inspire.com) performed with machine learning, along with the associated sentiment

Theme (number of posts, %)	Topic cluster	Number of posts (% of total posts)	Sentiment score
Pharmacological treatment (7312, 15.5)	Use and safety of novel and off-label medication	2671 (5.6)	0.1
	Pirfenidone: usage, effects and side-effects	1605 (3.4)	−0.1
	Nintedanib: usage, effects and side-effects	1127 (2.4)	−0.1
	Prednisone: dosage of use and side-effects	744 (1.6)	−0.2
	Side-effects with any medication	537 (1.1)	0.2
	Mycophenolate mofetil: indication, effects and side-effects	234 (0.5)	0.0
	Side-effects of oral and inhaled corticosteroids: fungal infection and thrush	219 (0.5)	−0.2
	Morphine for relief of breathlessness, cough or pain: dosage of use, effects and side-effects	175 (0.4)	0.1
Supportive care (6868, 14.5)	Management of oxygen therapy	6093 (12.9)	0.0
	End-of-life care: experience and decision-making for hospice referral	524 (1.1)	0.2
	Nocturnal oxygen use	251 (0.5)	0.0
Coping strategies (6378, 13.5)	Climatic condition: influences on patients' symptoms	850 (1.8)	0.0
	Diet: appetite loss, weight loss and weight gain	808 (1.7)	0.0
	Oximetry use: levels, monitoring and taking measurements at home	769 (1.6)	0.0
	Lung function measurements: interpreting and comparing with each other	614 (1.3)	−0.1
	Pain/discomfort: chest pain, back pain, wearing a bra and muscle cramps	598 (1.3)	−0.5
	Labour: quitting jobs, disability approval and social security acknowledgement	583 (1.2)	−0.1
	Physical activity, pulmonary rehabilitation and breathing exercises	430 (0.9)	0.4
	Skin rash and photosensitivity: potential causes and protection strategies	244 (0.5)	−0.3
	Hobbies and activities	239 (0.5)	0.9
	Voice hoarseness and singing	214 (0.5)	0.1
	Giving up on pets due to health concerns	200 (0.4)	0.2
	Daily mobility: mobility scooter, wheelchair and stair lift	199 (0.4)	0.3
	Exposure on smoking and asbestos: quitting smoking and legal matters to asbestos exposure	188 (0.4)	−0.4
	Strategies to improve sleeping	163 (0.3)	0.3
	Clubbing and thinning of finger- and toenails: causes and practical issues	162 (0.3)	−0.4
	COVID-19 pandemic: precautions taken, infectious risk, experience with COVID-19 infection and hospital admission	117 (0.2)	−0.4
Lung transplant (4475, 9.5)	Lung transplant: sharing experience, decision-making and aftercare	4240 (9.0)	0.3
	Lung transplant: concerns about costs and insurance of aftercare	132 (0.3)	0.0
	Lung transplant: waiting list	103 (0.2)	0.5
Peer support (4057, 8.6)	Loss of loved ones	2165 (4.6)	0.2
	Support groups and webinars	564 (1.2)	0.4
	Fear when newly diagnosed and loneliness	412 (0.9)	0.2
	Best wishes for recovery	318 (0.7)	0.9
	Humour and jokes	204 (0.4)	0.8
	Raising awareness	164 (0.3)	0.1
	Sharing patient recovery pathway as caregiver	116 (0.2)	0.5
	Sharing support for surgery	114 (0.2)	0.6

Continued

TABLE 1 Continued

Theme (number of posts, %)	Topic cluster	Number of posts (% of total posts)	Sentiment score
Aetiology, diagnosis and comorbidities (4165, 8.8)	GORD: pharmaceutical, nonpharmaceutical and surgical options	704 (1.5)	−0.1
	Hypersensitivity pneumonitis: exposures and removal	698 (1.5)	−0.3
	Hypersensitivity pneumonitis: drug-induced pneumonitis	685 (1.4)	−0.2
	Cardiac comorbidities and complications	443 (0.9)	−0.1
	Agent Orange and other chemical use in veterinary settings	351 (0.7)	−0.2
	Idiopathic pulmonary fibrosis: information	319 (0.7)	−0.2
	CTD-ILDs: information and experience with misdiagnosis	273 (0.6)	−0.2
	Familial pulmonary fibrosis: diagnosis and (dis)advantages of genomic testing	256 (0.5)	−0.3
	Nonspecific interstitial pneumonitis: information and diagnostic trajectory	158 (0.3)	0.1
	Pulmonary fibrosis: information and diagnostic trajectory	148 (0.3)	−0.3
Decision making (3354, 7.1)	Scleroderma-associated interstitial lung disease: diagnostic trajectory and eligibility for lung transplant	130 (0.3)	−0.1
	Centres of excellence and recommendations	1177 (2.5)	0.4
	Choosing healthcare insurance	1037 (2.2)	0.0
	Lung biopsy: sharing experience and decision-making	916 (1.9)	−0.1
	Dissatisfaction with doctors and centres	224 (0.5)	−0.2
Symptoms (2336, 4.9)	Cough: coping strategies	1866 (3.9)	0.1
	Respiratory tract infection and exacerbation	220 (0.5)	−0.2
	Fatigue: experience with tiredness, causes, remedies	129 (0.3)	−0.3
	Anxiety and depression: experience, breathing exercises and antidepressive therapy	121 (0.3)	−0.1
Nontraditional treatment (2210, 4.7)	Stem cell therapy: ongoing research, research outcomes and controversy	699 (1.5)	−0.2
	Enzyme supplements: efficacy, research results, way of administering, interactions	683 (1.4)	0.1
	Antioxidants, NAC and curcumin: efficacy, dosage of use, safety	545 (1.2)	0.1
	Laser therapy: efficacy and controversiality	158 (0.3)	−0.1
	Vitamin D: immunomodulating effects, prevent decalcification, dosage of use	125 (0.3)	0.0
Preventative measures (605, 1.3)	Vaccines: types of vaccinations and timing of administration	492 (1.0)	−0.1
	Face mask: selection, when to use and breathing challenges	113 (0.2)	0.1
Not relevant (4114, 8.7)			
Miscellaneous (1408, 3.0)			
Total		47 282	
Sentiment scores range from −1 to 1, where negative values indicate negative sentiment and positive values indicate positive sentiment. GORD: gastro-oesophageal reflux disease; CTD-ILD: connective tissue disease-associated interstitial lung disease; NAC: <i>N</i> -acetylcysteine.			

47 282 posts created from 5 May 2012 to 3 January 2023 were extracted from Inspire.com. Inspire is a patient platform initiated by the American Lung Association where patients from the USA, along with their families, share personal experiences. Posts were mainly created by patients, but also by their families, and can be short stories or replies to other posts. Specific details of the topic clusters, along with their associated sentiments, are presented in table 1. The most reported theme was pharmacological treatment (15.5%, n=7312), with discussions on the use and safety of novel and off-label treatments (5.6%, n=2671) and antifibrotic therapies, including pirfenidone (3.4%, n=1605) and nintedanib (2.4%, n=1127).

Furthermore, supportive care (14.5%, n=6868) was a prominent theme, characterised by many discussions on the use and management of oxygen therapy (12.9%, n=6093). Coping strategies, mostly related to managing pulmonary fibrosis at home, were also widely discussed (13.5%, n=6378). Among symptom-related discussions (4.9%, n=2336), cough was the most frequently discussed (3.9%, n=1866), followed by discussions on respiratory tract infections and exacerbations (0.5%, n=220), fatigue (0.3%, n=129), and anxiety and depression (0.3%, n=121). 8.7% (n=4114) of the posts did not contain relevant information (*e.g.* instructions regarding platform usage).

Commonly reported topic clusters showed an overall neutral sentiment. The topic cluster “pain and discomfort” (−0.5) scored negatively. Most topic clusters that regarded peer support had a positive sentiment score. Furthermore, the topic clusters “lung transplant: waiting list” (0.5) and “hobbies and activities” (0.5) scored positively.

In this social media analysis, we found that patients and their families discuss a wide variety of topics on social media to seek information, and share experiences with their disease and beyond. Among reported symptoms, cough was most frequently discussed, while fatigue and anxiety were also mentioned, highlighting the psychological impact of pulmonary fibrosis. Beside discussions on disease experiences, patients and their families appear to support one another remotely by sharing their fears and uplifting each other with humour. This is reflected in positive sentiment scores observed within the topic clusters under the theme of peer support. Topic clusters concerning lung transplant waiting lists were more positive in tone, possibly reflecting hope, while clusters related to pain showed more negativity. Overall, the machine learning classification of sentiment scores aligned with intuitive interpretations.

Our findings complement prior qualitative studies exploring patient perspectives regarding symptoms, diagnosis and treatment of pulmonary fibrosis [8–10], adding value by capturing patients’ unsolicited perspectives. This was reflected in discussions on traditional, and often controversial, therapies being discussed, such as stem cell therapy, enzyme supplements and laser therapy, that are often not emphasised in a clinical setting. This could potentially contribute to the spread of misinformation among patients [11–13]. Conventional methods for collecting patient perspectives are primarily initiated by researchers, which may introduce bias in the questions asked and the answers sought. Notably, machine learning allows for efficient (re)sampling of large datasets, an approach that has not been used before in the analysis of social media posts related to pulmonary fibrosis [11, 12, 14]. This could help detect shifts in patient perspectives in response to new developments.

Despite growing benefits of machine learning use, limitations exist for applying it in the analysis of social media posts. Unsupervised machine learning lacks deep contextual understanding and often yields broad or imprecise output. Furthermore, our human qualitative analysis revealed that not all posts within a topic cluster were relevant, underscoring the need to detect noise before drawing conclusions. Additionally, demographics of platform users are unknown and the sample may be skewed toward digitally skilled individuals. This is a recognised limitation in social media research.

In conclusion, our analysis shows that people with pulmonary fibrosis and their families actively share experiences and seek disease- and treatment-related information on social media. This approach may help identify knowledge gaps in care needs and inform patient-centred educational initiatives. While machine learning offers a promising tool for analysing patient discourse at scale, manual review remains essential to ensure accuracy. Future research should focus on refining machine learning models using disease-specific training data to enhance relevance and precision.

Delian E. Hofman¹, Ruixuan Zhang², Rembrandt Oltmans³, Johanna M. Hendriks⁴, Catharina C. Moor¹, Marlies Wijsenbeek-Lourens^{1,5} and Jiwon Jung^{2,4,5}

¹Centre of Excellence for Interstitial Lung Diseases and Sarcoidosis, Department of Respiratory Medicine, Erasmus University Medical Centre, Rotterdam, the Netherlands. ²Faculty of Industrial Design Engineering, Delft University of Technology, Delft, the Netherlands. ³Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, the Netherlands. ⁴Department of Surgery, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁵These authors contributed equally to this article as lead authors.

Corresponding author: Marlies Wijsenbeek-Lourens (m.wijsenbeek-lourens@erasmusmc.nl)

Acknowledgements: The authors thank Inspire.com for granting permission to use platform data for research purposes.

Provenance: Submitted article, peer reviewed.

Ethics statement: Ethical approval was obtained from the Human Research Ethical Committee at Delft University of Technology, the Netherlands.

Author contributions: R. Zhang, R. Oltmans, M. Wijsenbeek-Lourens and J. Jung contributed to the conception and design of the study. D.E. Hofman, R. Zhang, R. Oltmans and J. Jung were involved in data analyses. D.E. Hofman, R. Zhang, R. Oltmans, C.C. Moor, M. Wijsenbeek-Lourens and J. Jung interpreted the data. D.E. Hofman, M. Wijsenbeek-Lourens and J. Jung drafted the manuscript. D.E. Hofman, R. Zhang, R. Oltmans, J.M. Hendriks, C.C. Moor, M. Wijsenbeek-Lourens and J. Jung read and revised the manuscript for important intellectual content. All authors approved the final manuscript.

Conflict of interest: D.E. Hofman reports support outside the submitted work from ZonMw (the Netherlands Organisation for Health Research and Development) paid to their institution. C.C. Moor reports grants outside the submitted work from ZonMw, Boehringer Ingelheim and AstraZeneca-Daiichi Sankyo paid to their institution; payments for presentations/lectures from Boehringer Ingelheim and GSK paid to their institution; support for travel from Boehringer Ingelheim paid to their institution; and leadership roles with the ERS Early Career Member Committee (Early Career Member Representative of the Interstitial Lung Disease (ILD) Assembly), Professional Advisory Committee European Lung Foundation, and Advocacy Council ERS. M.S. Wijsenbeek-Lourens reports grants outside the submitted work from ZonMw, AstraZeneca-Daiichi Sankyo, Boehringer Ingelheim, Erasmus MC, Hoffmann-La Roche, Sarcoidosis.nl, The Dutch Lung Foundation, The Dutch Pulmonary Fibrosis Patients Association and The Thorax Foundation paid to their institution; consulting fees from AstraZeneca-Daiichi Sankyo, Avalyn Pharma, Boehringer Ingelheim, Bristol-Myers Squibb, CSL Behring, Galapagos, Galecto, GSK, Hoffmann-La Roche, Horizon Therapeutics, Kinevant Sciences, Molecure, NeRRe Therapeutics, Novartis, PureTech Health, Trevi and Vicore outside submitted work paid to their institution; payment or honoraria for presentations from Avalyn Pharma, Boehringer Ingelheim, Sanofi, CSL Behring and Novartis; support for travel from Boehringer Ingelheim, GSK and Hoffmann-La Roche; and leadership roles as secretary of ILD assembly for the European Respiratory Society, board member of the Netherlands Respiratory Society, scientific advisory board of European Idiopathic Pulmonary Fibrosis and Related Disorders Federation, chair of educational committee of European Reference Network for Rare Lung Diseases, and advisory board of Dutch lung fibrosis and sarcoidosis patient associations. J. Jung reports support for the present study from Convergence, Consultation Room 2030 Flagship (<https://convergence.nl/flagship-consultation-room-2030/>) paid to their institution, is paid a salary by the Dutch Government sector plan (<https://www.government.nl/topics/science/sector-plans>), and received social media forum data from Inspire.com. The data source used in our study was approved for release by the Inspire platform authority, based on a request we formally initiated (<https://www.inspire.com/groups/living-with-pulmonary-fibrosis/>). All other authors report no conflicts of interest.

Support statement: This analysis is sponsored by Convergence, Flagship Consultation Room 2030, a collaboration between TU Delft, Erasmus MC and Erasmus University Rotterdam. Convergence had no role in the design of the study, data collection or analysis or decision to submit the research letter for publication. Funding information for this article has been deposited with the Open Funder Registry.

References

- 1 Wijsenbeek M, Cottin V. Spectrum of fibrotic lung diseases. *N Engl J Med* 32020, 83: 958–968.
- 2 Cottin V, Teague R, Nicholson L, et al. The burden of progressive-fibrosing interstitial lung diseases. *Front Med (Lausanne)* 2022; 9: 799912.

- 3 Ma X, Zhu L, Kurche JS, *et al.* Global and regional burden of interstitial lung disease and pulmonary sarcoidosis from 1990 to 2019: results from the Global Burden of Disease study 2019. *Thorax* 2022; 77: 596–605.
- 4 Wells M, Harding S, Dixon G, *et al.* Patient and caregiver shared experiences of pulmonary fibrosis (PF): a systematic literature review. *Respir Med* 2024; 227: 107659.
- 5 Grewal JS, Kawano-Dourado L, Ryerson CJ. The opportunities and challenges of social media in interstitial lung disease: a viewpoint. *Respir Res* 2021; 22: 247.
- 6 Grootendorst MR. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv* 2022; pre-print [https://doi.org/10.48550/arXiv.2203.05794].
- 7 Campello R, Moulavi D, Zimek A, *et al.* Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Trans Knowl Discov Data* 2015; 10: 5.
- 8 Bonella F, Wijsenbeek M, Molina-Molina M, *et al.* European IPF Patient Charter: unmet needs and a call to action for healthcare policymakers. *Eur Respir J* 2016; 47: 597–606.
- 9 van Manen MJ, Kreuter M, van den Blink B, *et al.* What patients with pulmonary fibrosis and their partners think: a live, educative survey in the Netherlands and Germany. *ERJ Open Res* 2017; 3: 00065-2016.
- 10 Russell A-M, Ripamonti E, Vancheri C. Qualitative European survey of patients with idiopathic pulmonary fibrosis: patients' perspectives of the disease and treatment. *BMC Pulm Med* 2016; 16: 10.
- 11 Kochan A, Ong S, Guler S, *et al.* Social media content of idiopathic pulmonary fibrosis groups and pages on facebook: cross-sectional analysis. *JMIR Public Health Surveill* 2021; 7: e24199.
- 12 Ong S, Koo J, Johannson KA, *et al.* Content analysis of idiopathic pulmonary fibrosis-related information on twitter. *ATS Sch* 2022; 3: 576–587.
- 13 Goobie GC, Guler SA, Johannson KA, *et al.* YouTube videos as a source of misinformation on idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* 2019; 16: 572–579.
- 14 Albright K, Walker T, Baird S, *et al.* Seeking and sharing: why the pulmonary fibrosis community engages the web 2.0 environment. *BMC Pulm Med* 2016; 16: 4.