

BUKTI KORESPONDENSI ARTIKEL JURNAL INTERNASIONAL BEREPUTASI

Judul Artikel : Prosodic Spatio-Temporal Feature Fusion with Attention
Mechanisms for Speech Emotion Recognition

Nama Jurnal : Computers

Publhiser : MDBI

Quarttile Scopus : Q1

SJR : 0,805

Penulis Korespondensi : Kristiawan Nugroho

No	Kegiatan	Tanggal	Halaman
1	Submission to Computers	4 Agustus 2025	2
2	Decession to Revised (Major Revisiion)	13 Agustus 2025	3
3	Resubmit Paper Revision	24 Agustus 2025	5
4	Decission Minor Revision	28 Agustus 2025	6
5	Resubmit Paper Revision	29 Agustus 2025	7
6	Decisiion Accepted	29 Agustus 2025	8
7	Proofread	30 Agustus 2025	9
8	Full Published (Volume 14 No 09, 31 Agustus 2025)	31 Agustus 2025	10

1. Submission to COMPUTERS

Computers Editorial Office <computers@mdpi.com>

Mon, Aug 4, 2025 at 8:08 PM

Reply-To: computers@mdpi.com

To: De Rosal Ignatius Moses Setiadi <moses@dsn.dinus.ac.id>, Kristiawan Nugroho <kristiawan@edu.unisbank.ac.id>

Cc: Computers Editorial Office <computers@mdpi.com>

Dear Dr. Nugroho, Dr. Setiadi,

Thank you for submitting the following manuscript for publication in
Computers:

Manuscript ID: computers-3829370

Type of manuscript: Article

Title: Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for
Speech Emotion Recognition

Authors: Kristiawan Nugroho *, Imam Husni Al Amin, Nina Anggraeni Noviasari,
De Rosal Ignatius Moses Setiadi *

Received: 4 Aug 2025

E-mails: kristiawan@edu.unisbank.ac.id, imam@edu.unisbank.ac.id,
ninaanggraeni@unimus.ac.id, moses@dsn.dinus.ac.id

We have identified that you have available discount vouchers that can be
applied in the submission. You are able to add the vouchers via the following
link:

<https://susy.mdpi.com/user/manuscripts/upload/c4c9b73a062ca3458506de174d552e4b/5#choose-discount-section>

Note that vouchers issued to specific individuals are not transferable. Only
vouchers that were issued to one of the authors of the current paper are
valid and must be mentioned during the submission process.

Please note that reviewer vouchers must be applied before acceptance for
publication. Vouchers can no longer be applied once an APC invoice has been
issued.

If you have any question, please contact Computers Editorial Office.

Kind regards,

Computers Editorial Office

Postfach, CH-4020 Basel, Switzerland

Office: Grosspeteranlage 5, CH-4052 Basel

Tel. +41 61 683 77 34 (office)

E-mail: computers@mdpi.com

<https://www.mdpi.com/journal/computers/>

2. Decection to Revised (Major Revisiion)

Computers Editorial Office <computers@mdpi.com>

Wed, Aug 13, 2025 at 8:18 AM

Reply-To: rena.zhu@mdpi.com

To: Kristiawan Nugroho <kristiawan@edu.unisbank.ac.id>, De Rosal Ignatius Moses Setiadi <moses@dsn.dinus.ac.id>

Cc: Imam Husni Al Amin <imam@edu.unisbank.ac.id>, Nina Anggraeni Noviasari <ninaanggraeni@unimus.ac.id>, Computers Editorial Office <computers@mdpi.com>

Dear Dr. Nugroho, Dr. Setiadi,

Thank you again for your manuscript submission:

Manuscript ID: computers-3829370

Type of manuscript: Article

Title: Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition

Authors: Kristiawan Nugroho *, Imam Husni Al Amin, Nina Anggraeni Noviasari, De Rosal Ignatius Moses Setiadi *

Received: 4 Aug 2025

E-mails: kristiawan@edu.unisbank.ac.id, imam@edu.unisbank.ac.id,

ninaanggraeni@unimus.ac.id, moses@dsn.dinus.ac.id

Section: AI-Driven Innovations

https://www.mdpi.com/journal/computers/sections/AI_innovations_in_driving

Special issue: Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition)

https://www.mdpi.com/journal/computers/special_issues/CPA7UOWOH2

Your manuscript has now been reviewed by experts in the field and can be found with the review reports at:

<https://susy.mdpi.com/user/manuscripts/resubmit/c4c9b73a062ca3458506de174d552e4b>

Please revise the manuscript found at the above link according to the reviewers' comments and upload the revised file within 10 days. Note the following check-list:

- (I) Ensure all references are relevant to the content of the manuscript.
- (II) Highlight any revisions to the manuscript, so editors and reviewers can see any changes made.
- (III) Provide a cover letter to respond to the reviewers' comments and explain, point by point, the details of the manuscript revisions.
- (IV) If the reviewer(s) recommended references, critically analyze them to ensure that their inclusion would enhance your manuscript. If you believe these references are unnecessary, you should not include them.
- (V) If you found it impossible to address certain comments in the review reports, include an explanation in your appeal.

If your manuscript requires improvement to the language and/or figures, you may consider MDPI Author Services: <https://www.mdpi.com/authors/english>.

Please do not hesitate to contact us if you have any questions regarding the revision of your manuscript or if you need more time. We look forward to hearing from you soon.

Journal	Computers (ISSN 2073-431X)
Manuscript ID	computers-3829370
Type	Article
Title	Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition
Authors	Kristiawan Nugroho *, Imam Husni Al Amin , Nina Anggraeni Noviasari , De Rosal Ignatius Moses Setiadi *
Section	AI-Driven Innovations
Special Issue	Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition)
Abstract	Speech Emotion Recognition (SER) plays a vital role in supporting applications such as healthcare, human-computer interaction, and security. However, many existing approaches still face challenges in achieving robust generalization and maintaining high recall, particularly for emotions related to stress and anxiety. This study proposes a dual-stream hybrid model that combines prosodic features with spatio-temporal representations derived from the Multitaper Mel-Frequency Spectrogram (MTMFS) and the Constant-Q Transform Spectrogram (CQTS). Prosodic cues, including pitch, intensity, jitter, shimmer, HNR, pause rate, and speech rate, were processed using dense layers, while MTMFS and CQTS features were encoded with CNN and BiGRU. A Multi-Head Attention mechanism was then applied to adaptively fuse the two feature streams, allowing the model to focus on the most relevant emotional cues. Evaluations conducted on the RAVDESS dataset with subject-independent 5-fold cross-validation demonstrated an accuracy of 97.64% and a macro F1-score of 0.9745. These results confirm that combining prosodic and advanced spectrogram features with attention-based fusion improves precision, recall, and overall robustness, offering a promising framework for more reliable SER systems.

The cover letter for this review report has been saved in the database. You can safely close this window.

Authors' Responses to Reviewer's Comments (Reviewer 1)

Author's Notes	Comment 1: The present study proposes a model for SER and demonstrate its effect. The research approach is generally fine and the research results are good. However, there are parts of the manuscript that are not logical and need to be revised. It does not take a verification way to directly answer some of the issues discussed in the background. Therefore, it is necessary to state the research gap in line with what was accomplished in the present study. Response 1: Thank you for your helpful feedback. We have revised the end of the Introduction (lines 101–107) to explicitly state the research gaps and clarify how the present study addresses them. Comment 2: Furthermore, although the conclusion emphasizes the increased accuracy, it is unclear how significant a 0.1–0.2% increase would be. If 97.4% of the previous studies are insufficient and 97.6% of the present study is important, then specifics should be given as to what this will enable. The reviewer's understanding is that the significance of the present study seems to lie in the uniqueness of the proposed model rather than in this minor numerical difference. If so, I recommend that the background and conclusions be revised to make it clearer what is unique and significant about the model. Response 2: Thank you for your valuable comment. We have revised the end of Section 4 and the Conclusion to clarify that the significance of our work lies not merely in the small numerical improvement but in the novelty of integrating prosodic features with MTMFS and CQTS under an attention-based fusion scheme, which provides a more balanced trade-off between precision and recall and enhances robustness. Comment 3: The word hypothesis appears in line 353, but the term hypothesis here should be changed to a milder term since the manuscript is not designed to test an explicit hypothesis. Response 3: Thank you for raising this point. Although it is a simple change, we agree it is important for accuracy. We have replaced the term "hypothesis" with "expectation" at line 367 (after revision) to better reflect the nature of the study.
----------------	---

3. Resubmit Paper Revision

Computers Editorial Office <computers@mdpi.com>

Sun, Aug 24, 2025 at 3:48 PM

Reply-To: david.hou@mdpi.com

To: Kristiawan Nugroho <kristiawan@edu.unisbank.ac.id>, De Rosal Ignatius Moses Setiadi <moses@dsn.dinus.ac.id>

Cc: Imam Husni Al Amin <imam@edu.unisbank.ac.id>, Nina Anggraeni Noviasari <ninaanggraeni@unimus.ac.id>, Computers Editorial Office <computers@mdpi.com>

Dear Dr. Nugroho, Dr. Setiadi,

Thank you very much for providing the revised version of your paper:

Manuscript ID: computers-3829370

Type of manuscript: Article

Title: Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition

Authors: Kristiawan Nugroho *, Imam Husni Al Amin, Nina Anggraeni Noviasari, De Rosal Ignatius Moses Setiadi *

Received: 4 Aug 2025

E-mails: kristiawan@edu.unisbank.ac.id, imam@edu.unisbank.ac.id, ninaanggraeni@unimus.ac.id, moses@dsn.dinus.ac.id

AI-Driven Innovations

https://www.mdpi.com/journal/computers/sections/AI_innovations_in_driving

Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition)

https://www.mdpi.com/journal/computers/special_issues/CPA7UOWOH2

https://susy.mdpi.com/user/manuscripts/review_info/c4c9b73a062ca3458506de174d552e4b

We will continue processing your paper and will keep you informed about the status of your submission.

Kind regards,

Mr. David Hou

E-Mail: david.hou@mdpi.com

Assistant Editor

/Fractal and Fractional/

2024 Impact Factor: 3.3 (Q1 in category Mathematics, Interdisciplinary Applications)

2024 CiteScore: 6.0 (Q1 in category "Analysis" & "Statistics and Probability" & "Statistical and Nonlinear Physics")

News: Highly Cited/Downloaded Paper (Free to Read/Download)

Multi-Machine Power System Transient Stability [...]

<http://www.mdpi.com/2504-3110/7/11/808>

Fractional-Order Impulsive Delayed Reaction-Diffusion [...]

<http://www.mdpi.com/2504-3110/7/5/384>

Modified FOPID Controller for Frequency Regulation [...]

<http://www.mdpi.com/2504-3110/7/1/89>

4. Decision Minor Revision

Computers Editorial Office <computers@mdpi.com>

Thu, Aug 28, 2025 at 9:16 PM

Reply-To: Rena Zhu <rena.zhu@mdpi.com>, Computers Editorial Office <computers@mdpi.com>

To: Kristiawan Nugroho <kristiawan@edu.unisbank.ac.id>, De Rosal Ignatius Moses Setiadi <moses@dsn.dinus.ac.id>

Cc: Imam Husni Al Amin <imam@edu.unisbank.ac.id>, Nina Anggraeni Noviasari <ninaanggraeni@unimus.ac.id>, Computers Editorial Office <computers@mdpi.com>, Rena Zhu <rena.zhu@mdpi.com>

Dear Dr. Nugroho, Dr. Setiadi,

Thank you again for your manuscript submission:

Manuscript ID: computers-3829370

Type of manuscript: Article

Title: Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition

Authors: Kristiawan Nugroho *, Imam Husni Al Amin, Nina Anggraeni Noviasari, De Rosal Ignatius Moses Setiadi *

Received: 4 Aug 2025

E-mails: kristiawan@edu.unisbank.ac.id, imam@edu.unisbank.ac.id, ninaanggraeni@unimus.ac.id, moses@dsn.dinus.ac.id

Section: AI-Driven Innovations

https://www.mdpi.com/journal/computers/sections/AI_innovations_in_driving

Special issue: Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition)

https://www.mdpi.com/journal/computers/special_issues/CPA7UOWOH2

Your manuscript has now been reviewed by the Academic Editor and can be found with the review reports at:

<https://susy.mdpi.com/user/manuscripts/resubmit/c4c9b73a062ca3458506de174d552e4b>

Please revise the manuscript found at the above link according to the comments and upload the revised file within 3 days. Note the following check-list:

- (I) Ensure all references are relevant to the content of the manuscript.
- (II) Highlight any revisions to the manuscript, so editors and reviewers can see any changes made.
- (III) Provide a cover letter to respond to the reviewers' comments and explain, point by point, the details of the manuscript revisions.
- (IV) If the reviewer(s) recommended references, critically analyze them to ensure that their inclusion would enhance your manuscript. If you believe these references are unnecessary, you should not include them.
- (V) If you found it impossible to address certain comments in the review reports, include an explanation in your appeal.

Our expert editors offer professional Language and Figure Editing, each completed within 1 day. We ensure the high standard required to publish your work with clarity and boost its impact. You can submit here:

https://susy.mdpi.com/user/manuscripts/review_info/c4c9b73a062ca3458506de174d552e4b#author_services_section

Please do not hesitate to contact us if you have any questions regarding the revision of your manuscript or if you need more time. We look forward to hearing from you soon.

5. Resubmit Paper Revision

Computers Editorial Office <computers@mdpi.com>

Fri, Aug 29, 2025 at 8:25 AM

Reply-To: Rena Zhu <rena.zhu@mdpi.com>, Computers Editorial Office <computers@mdpi.com>

To: Kristiawan Nugroho <kristiawan@edu.unisbank.ac.id>, De Rosal Ignatius Moses Setiadi <moses@dsn.dinus.ac.id>

Cc: Imam Husni Al Amin <imam@edu.unisbank.ac.id>, Nina Anggraeni Noviasari <ninaanggraeni@unimus.ac.id>

Dear Dr. Nugroho, Dr. Setiadi,

Thank you very much for resubmitting the modified version of the following manuscript:

Manuscript ID: computers-3829370

Type of manuscript: Article

Title: Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition

Authors: Kristiawan Nugroho *, Imam Husni Al Amin, Nina Anggraeni Noviasari, De Rosal Ignatius Moses Setiadi *

Received: 4 Aug 2025

E-mails: kristiawan@edu.unisbank.ac.id, imam@edu.unisbank.ac.id, ninaanggraeni@unimus.ac.id, moses@dsn.dinus.ac.id

AI-Driven Innovations

https://www.mdpi.com/journal/computers/sections/AI_innovations_in_driving

Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition)

https://www.mdpi.com/journal/computers/special_issues/CPA7UOWOH2

https://susy.mdpi.com/user/manuscripts/review_info/c4c9b73a062ca3458506de174d552e4b

A member of the editorial office will be in touch with you soon regarding progress of the manuscript.

Kind regards,

Computers Editorial Office

Postfach, CH-4020 Basel, Switzerland

Office: Grosspeteranlage 5, CH-4052 Basel

Tel. +41 61 683 77 34 (office)

E-mail: computers@mdpi.com

<https://www.mdpi.com/journal/computers/>

6. Decision Accepted

Computers Editorial Office <computers@mdpi.com>

Fri, Aug 29, 2025 at 9:57 AM

Reply-To: Rena Zhu <rena.zhu@mdpi.com>, Computers Editorial Office <computers@mdpi.com>

To: Kristiawan Nugroho <kristiawan@edu.unisbank.ac.id>, De Rosal Ignatius Moses Setiadi <moses@dsn.dinus.ac.id>

Cc: Imam Husni Al Amin <imam@edu.unisbank.ac.id>, Nina Anggraeni Noviasari <ninaanggraeni@unimus.ac.id>, Computers Editorial Office <computers@mdpi.com>, Rena Zhu <rena.zhu@mdpi.com>

Dear Dr. Nugroho, Dr. Setiadi,

Congratulations on the acceptance of your manuscript, and thank you for submitting your work to Computers. Please see the relevant manuscript details below:

Manuscript ID: computers-3829370

Type of manuscript: Article

Title: Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition

Authors: Kristiawan Nugroho *, Imam Husni Al Amin, Nina Anggraeni Noviasari, De Rosal Ignatius Moses Setiadi *

Received: 4 Aug 2025

E-mails: kristiawan@edu.unisbank.ac.id, imam@edu.unisbank.ac.id, ninaanggraeni@unimus.ac.id, moses@dsn.dinus.ac.id

AI-Driven Innovations

https://www.mdpi.com/journal/computers/sections/AI_innovations_in_driving

Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition)

https://www.mdpi.com/journal/computers/special_issues/CPA7UOWOH2

https://susy.mdpi.com/user/manuscripts/review_info/c4c9b73a062ca3458506de174d552e4b

We will now edit your paper; once complete, we will return it to you for your review and approval. The standard article processing charge (APC) for Computers is 1800 CHF. Within the next few days, an invoice for a 0 CHF article processing charge (APC) will be sent by email from the Editorial Office in Basel, Switzerland, for your records.

If, however, your manuscript requires extensive English language and grammar revisions, we will need to return the paper and request that such changes are made.

We encourage you to set up a profile on SciProfiles.com, MDPI's researcher network platform. Articles you publish with MDPI will be linked to your SciProfiles page, where your colleagues and peers will be able to see all of your publications, citations, and other academic contributions.

Kind regards,
Paolo Bellavista
Editor-in-Chief

7. Proofread Paper

Computers Editorial Office <computers@mdpi.com>

Sat, Aug 30, 2025 at 9:18 AM

Reply-To: Rena Zhu <rena.zhu@mdpi.com>, Computers Editorial Office <computers@mdpi.com>

To: Kristiawan Nugroho <kristiawan@edu.unisbank.ac.id>, De Rosal Ignatius Moses Setiadi <moses@dsn.dinus.ac.id>

Cc: Imam Husni Al Amin <imam@edu.unisbank.ac.id>, Nina Anggraeni Noviasari <ninaanggraeni@unimus.ac.id>, Computers Editorial Office <computers@mdpi.com>, aviva.yu@mdpi.com, Rena Zhu <rena.zhu@mdpi.com>

Dear Dr. Nugroho, Dr. Setiadi,

We invite you to proofread your manuscript to ensure that this is the final version for publication and to confirm that you require no further changes:

At MDPI, we believe in the fast dissemination of sound, valid scientific knowledge. Once accepted for publication, we aim to ensure that research is published as soon as possible.

Please upload the final proofread version of your manuscript within 24 hours, and please remember that we can be flexible with this timeframe if necessary. If you need more time, please inform the Assistant Editor of the date by which you will be able to return the proofread version.

Please also ensure that you check that your manuscript has been properly categorized by our AI tool during proofreading here:
<https://category-connect.mdpi.com/computers-3829370/f1b7f4d9a8b9713d>

Our new AI tool (Category Connect), once launched, will assist with MDPI's mission to foster open scientific exchange by enhancing the discoverability of related papers on the MDPI website. We hope that you will assist us in the development of this important tool by validating the categories identified for your paper.

Manuscript ID: computers-3829370

Type of manuscript: Article

Title: Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition

Authors: Kristiawan Nugroho *, Imam Husni Al Amin, Nina Anggraeni Noviasari, De Rosal Ignatius Moses Setiadi *

Received: 4 Aug 2025

E-mails: kristiawan@edu.unisbank.ac.id, imam@edu.unisbank.ac.id, ninaanggraeni@unimus.ac.id, moses@dsn.dinus.ac.id

AI-Driven Innovations

https://www.mdpi.com/journal/computers/sections/AI_innovations_in_driving

Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition)

https://www.mdpi.com/journal/computers/special_issues/CPA7UOWOH2

Please read the following instructions carefully before proofreading:

8. Full Published (Volume 14 No 09, 31 Agustus 2025)

Open Access Article

Prosodic Spatio-Temporal Feature Fusion with Attention Mechanisms for Speech Emotion Recognition

by Kristiawan Nugroho ^{1,*} , Imam Husni Al Amin ² , Nina Anggraeni Noviasari ³  and De Rosal Ignatius Moses Setiadi ^{4,*} 

¹ Department of Information Technology, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang 50241, Indonesia

² Department of Industrial Engineering, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang 50241, Indonesia

³ Faculty of Medicine, Universitas Muhammadiyah Semarang, Semarang 50273, Indonesia

⁴ Research Centre for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

* Authors to whom correspondence should be addressed.

Computers **2025**, *14*(9), 361; <https://doi.org/10.3390/computers14090361>

Submission received: 4 August 2025 / Revised: 29 August 2025 / Accepted: 29 August 2025 /

Published: 31 August 2025

(This article belongs to the Special Issue Multimodal Pattern Recognition of Social Signals in HCI (2nd Edition))

Download 

Browse Figures

Versions Notes

Abstract

Speech Emotion Recognition (SER) plays a vital role in supporting applications such as healthcare, human–computer interaction, and security. However, many existing approaches still face challenges in achieving robust generalization and maintaining high recall, particularly for emotions related to stress and anxiety. This study proposes a dual-stream hybrid model that combines prosodic features with spatio-temporal representations derived from the Multitaper Mel-Frequency Spectrogram (MTMFS) and the Constant-Q Transform Spectrogram (CQTS). Prosodic cues, including pitch, intensity, jitter, shimmer, HNR, pause rate, and speech rate, were processed using dense layers, while MTMFS and CQTS features were encoded with CNN and BiGRU. A Multi-Head Attention mechanism was then applied to adaptively fuse the two feature streams, allowing the model to focus on the most relevant emotional cues. Evaluations conducted on the RAVDESS dataset with subject-independent 5-fold cross-validation demonstrated an accuracy of 97.64% and a macro F1-score of 0.9745. These results confirm that combining prosodic and advanced spectrogram features with attention-based fusion improves precision, recall, and overall robustness, offering a promising framework for more reliable SER systems.

Keywords: speech emotion recognition; prosodic features; multitaper mel-frequency spectrogram; constant-Q transform; attention mechanism

Link paper : <https://www.mdpi.com/2073-431X/14/9/361>

