

# Hybrid approach redefinition-multi class with resampling and feature selection for multi-class imbalance with overlapping and noise

**Erianto Ongko**

Akademi Teknologi Industri Immanuel

**Hartono**

Universitas Potensi Utama

Bulletin of Electrical Engineering and Informatics Vol. 10 No. 3, Juni 2021

1. First Submit: 24 Juli 2020
2. Accepted with Major Revision: 26 Agustus 2020
3. Submit Hasil Revisi : 29 Desember 2020
4. Last Review: 17 Maret 2021
5. Final Manuscript: 19 April 2021
6. Terbit: Juni 2021

# Bukti Korespondensi Pengajuan Guru Besar Hartono

Judul Karya Ilmiah : Hybrid approach redefinition-multi class with resampling and feature selection for multi-class imbalance with overlapping and noise  
Jurnal : Bulletin of Electrical Engineering and Informatics  
Penulis : Erianto Ongko, Hartono  
Edisi : 2021  
Volume : Vol. 10 No. 3  
Penerbit : Institute of Advanced Engineering and Science (IAES)

The screenshot shows the journal's header with the IAES logo, ISSN information, and various accreditation metrics (3.0, Q3, SJR 2022 0.299, SNIP 2022 0.639). The navigation menu includes Home, About, Login, Register, Search, Current, Archives, and Announcements. The article title is prominently displayed, followed by the authors' names. The abstract section describes the research methodology and findings. A sidebar on the right contains a user login form, citation analysis options (Dimensions, Google Scholar, etc.), quick links to editorial boards and submission guidelines, and a 'How to Submit' section.

iaes  
Bulletin of Electrical Engineering and Informatics  
http://beei.org | ISSN 2089-3191, e-ISSN 2302-9285  
3.0  
Q3  
SJR 2022 0.299  
SNIP 2022 0.639  
Intelektual Pustaka  
Media Utama

HOME ABOUT LOGIN REGISTER SEARCH CURRENT ARCHIVES ANNOUNCEMENTS

Home > Vol 10, No 3 > Ongko

## Hybrid approach redefinition-multi class with resampling and feature selection for multi-class imbalance with overlapping and noise

Erianto Ongko, Hartono Hartono

### Abstract

Class imbalance and overlapping on multi-class can reduce the performance and accuracy of the classification. Noise must also be considered because it can reduce the performance of classification. With a resampling algorithm and feature selection, this paper proposes a method for improving the performance of hybrid approach redefinition-multi class (HAR-MI). Resampling algorithm can overcome the problem of noise but cannot handle overlapping well. Feature selection is good at dealing with overlapping but can experience a decrease in quality if there is a noise. The HAR-MI approach is a way to deal with multi-class imbalance issues, but it has some drawbacks when dealing with overlapping. The contribution of this paper is to suggest a new approach for dealing with class imbalance, overlapping, and noise in multi-class. This is accomplished by employing minimizing overlapping selection (MOSS) as an ensemble learning algorithm and a preprocessing technique in HAR-MI, as well as employing multi-class combination cleaning and resampling (MC-CCR) as a resampling algorithm at the processing stage. When subjected to overlapping and classifier performance, it is discovered that the proposed method produces good results, as evidenced by higher augmented  $r$ -value, class average accuracy, class balance accuracy, multi class  $g$ -mean, and confusion entropy.

### Keywords

Class imbalance; Classifier performance; Multi-class; Noise; Overlapping;

Full Text:  
PDF

DOI: <https://doi.org/10.11591/eei.v10i3.3057>

USER  
Username  
Password  
 Remember me  
Login

CITATION ANALYSIS

- Dimensions
- Google Scholar
- Scholar Metrics
- Scinapse
- Scopus

QUICK LINKS

- Editorial Boards
- Reviewers
- Abstracting and Indexing
- Guide of Authors
- Online Papers Submission
- Peer Review Process
- Publication Fee
- Publication Ethics
- Visitor Statistics
- DOI Deposit Report
- Contact Us

HOW TO SUBMIT

How to submit ...

**First Submit**

---

**Status in EDAS for ICW TELKOMNIKA 2020 : Pending (No Manuscript) for Paper ID: 147 (1570662772)**

3 pesan

---

**Hartono Ibbi** <hartonoibbi@gmail.com>  
Kepada: icwt1 telkomika <icwt1@ee.uad.ac.id>  
Cc: lppi@uad.ac.id

27 Juli 2020 pukul 10.50

Dear Sir/Madam

On July 24, 2020, I have submitted my paper through EDAS for ICW TELKOMNIKA 2020. When I submitted, the upload manuscript button could not be clicked, so then I saved and after that, I couldn't click the upload manuscript button again. So I chose upload via Review Manuscript. However, at the moment my paper status at EDAS is Pending (No Manuscript). My Paper ID is # 147 (1570662772).

I want to ask for information on how to solve this problem. Thank you for the attention and assistance provided.

Best Regards,

Hartono

---

**ICWT1 TELKOMNIKA** <icwt1@ee.uad.ac.id>  
Kepada: Hartono Ibbi <hartonoibbi@gmail.com>  
Cc: TH Sutikno <lppi@uad.ac.id>

27 Juli 2020 pukul 12.28

Dear Sir/Madam,

Status of paper ID 1570662772 entitled HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise is **Active (has manuscript)**. Please check your EDAS account. If the status is different in your account, please inform us. Thank you

Best Regards,  
ICW-TELKOMNIKA Committee

[Kutipan teks disembunyikan]

---

**Hartono Ibbi** <hartonoibbi@gmail.com>  
Kepada: ICWT1 TELKOMNIKA <icwt1@ee.uad.ac.id>

27 Juli 2020 pukul 13.07

Dear Sir/Madam,

Thanks for the assistance provided. Pleased to have the opportunity to participate in the ICW-TELKOMNIKA 2020, after last year had the opportunity to participate in the ICW-TELKOMNIKA 2019. Hopefully this year the ICW-TELKOMNIKA Committee will give an opportunity for me to participate again in this event. Good luck and always be healthy for All of the ICW-TELKOMNIKA committee.

Best Regards,

Hartono

[Kutipan teks disembunyikan]

## **Hasil Review**

**[TELKOMNIKA] Decision for paper 'HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise'**

3 pesan

telkomnika@uad.ac.id &lt;telkomnika=uad.ac.id@edas.info&gt;

25 Agustus 2020 pukul 23.09

Balas Ke: telkomnika@uad.ac.id

Kepada: Hartono Hartono &lt;hartonoibbi@gmail.com&gt;, Erianto Ongko &lt;eriantoongko@gmail.com&gt;

- Please Strictly use and follow to the template Manuscripts
- (Word Format): <http://iaescore.com/gfa/telkomnika.docx>
- Please upload the revised paper within 8 weeks
- Similarity score of your final manuscript must be less than 25%
- by iThenticate or Turnitin software

Dear Mr. Hartono Hartono,

After careful review, your paper #1570662772 "HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise" for TELKOMNIKA Telecommunication Computing Electronics and Control requires MAJOR REVISIONS. You are asked to submit a revised full manuscript, according to the comment from reviewers. The Technical Program Committee (TPC) will check whether the revision already address the reviewers' comments. Failing to do proper revision may lead to the rejection of your paper.

For your information, TELKOMNIKA ISSN: 1693-6930, e-ISSN: 2302-9293 (<http://journal.uad.ac.id/index.php/TELKOMNIKA>) is open to submission from scholars and experts in the wide areas of electrical, electronics, instrumentation, control, robotics, telecommunication, computer engineering, computer science, information system, information technology and informatics from the global world. The aim of this journal is to publish high-quality articles dedicated to all aspects of the latest outstanding developments in the field of electrical engineering. Its scope encompasses the applications of Telecommunication and Information Technology, Applied Computing and Computer, Instrumentation and Control, Electrical (Power), and Electronics Engineering. It was first published in 2003. Beginning with issue 1 of volume 16 (2018), TELKOMNIKA will be published as a bimonthly journal (6 issues/year). The journal registered in the CrossRef system with Digital Object Identifier (DOI) prefix 10.12928. The Journal has been indexed by SCOPUS, Google Scholar, Scholar Metrics etc; accredited 'A' Grade by DGHE (Ministry of Research, Technology and Higher Education, Republic of Indonesia); registered BASE - Bielefeld Academic Search Engine and CORE KMI, etc. The Journal also have a license agreement with ProQuest LLC and EBSCO Publishing.

The reviews are below or can be found at <https://edas.info/showPaper.php?m=1570662772>, using your EDAS login name [hartonoibbi@gmail.com](mailto:hartonoibbi@gmail.com).

Please upload the revised paper using EDAS on the "Revision" upload button within 8 weeks.

Thank you for your cooperation.

Best Regards,  
Assoc. Prof. Dr. Tole Sutikno  
Editor-in-Chief,  
TELKOMNIKA (Telecommunication, Computing, Electronics and Control)  
website: <http://journal.uad.ac.id/index.php/TELKOMNIKA>  
email: [telkomnika@uad.ac.id](mailto:telkomnika@uad.ac.id)

Submit paper: <https://edas.info/N27504>

Below is the reviews on your papers:

===== Review 1 =====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?

Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?

Accept (9)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The paper is written very well and the stream of the paper is clear

> \*\*\* Recommendation: Your overall rating.  
Accept (9)

=====  
Review 2  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Accept (9)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The authors present well written technical paper with relevant topic, proper research methodology and good contribution to the field of studies.

> \*\*\* Recommendation: Your overall rating.  
Accept (9)

=====  
Review 3  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Borderline Accept (5)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

- Please revise the academic writing style and grammars throughout, for examples (but not limited to) the first 3 sentences in the abstract, some broken sentences, casual phrases, etc.

- Please make the abstract more concise, containing only the key contributions and findings (leaving out detailed discussion to the main text).

- Please restructure the motivation parts, regarding the imbalance and overlapping and make them more focus. For example (but not limited to), several statements are dedicated to problems caused by overlapping, but none of them is more specific than others. This pattern repeats for most parts of the introduction. Please revise. Moreover, feature selection algorithms are mentioned but unfortunately not so specific nor much discussed. Please elaborate.

- Majority of the section 2 are performance metrics and not the related works but more the material components in the proposed method. Please consider.

- Please use more appropriate font/ format for pseudocode presented in section 2.

- Please add structural outline of how components presented in section 2 would be used in section 4.

- Similarly, please refer each block in Figure to respective topic ID in section 2.

- The results discussion are well written. However, please add in section 4.4., comparative discussion with similar works in the areas (impalance, noise handling).

- Please compare the proposed technique with that without MOSS and MC-CCR processing.

> \*\*\* Recommendation: Your overall rating.  
Weak Accept (7)

=====  
Revision review 4  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Borderline Accept (5)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Borderline Accept (5)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The paper presentation is sufficient. Enough theoretical background to support the proposed method.

The authors should state the contribution and novelty of the proposed method compared to the current researches.

The discussions should be more elaborated to support the effectiveness of the proposed method.

> \*\*\* Recommendation: Your overall rating.  
Borderline Accept (5)

We would like your cooperation with the double check of your revised paper:

(1) TEMPLATE, Please Strictly use and follow to the template Manuscripts (Word Format): <http://iaescore.com/gfa/telkomnika.docx>

(2) Authors are suggested to present their articles with IMRaD sections structure (outline): Introduction - The Proposed Method/Algorithm/Procedure specifically designed (optional) - Research Method - Results and Discussion – Conclusion. Authors may present complex proofs of theorems or non-obvious proofs of correctness of algorithms after introduction section (obvious theorems & straightforward proofs of existing theorems are NOT needed).

(3) Introduction section: explain the context of the study and state the precise objective. An Introduction should contain the following three parts:

- Background: Authors have to make clear what the context is. Ideally, authors should give an idea of the state-of-the art of the field the report is about.

- The Problem: If there was no problem, there would be no reason for writing a manuscript, and definitely no reason for reading it. So, please tell readers why they should proceed reading. Experience shows that for this part a few lines are often sufficient.

- The Proposed Solution: Now and only now! - authors may outline the contribution of the manuscript. Here authors have to make sure readers point out what are the novel aspects of authors work.

Authors should place the paper in proper context by citing relevant papers. At least, 15 references (recently journal articles) are referred in this section.

(4) Method section: the presentation of the experimental methods should be clear and complete in every detail facilitating reproducibility by other scientists.

(5) Results and discussion section: The presentation of results should be simple and straightforward in style. This section report the most important findings, including results of statistical analyses as appropriate and comparisons to other research results. Results given in figures should not be repeated in tables. This is where the author(s) should explain in words what he/she/they discovered in the research. It should be clearly laid out and in a logical sequence. This section should be supported suitable references.

(6) (URGENT)!!! About Figures & Tables in your manuscript:

- Because tables and figures supplement the text, all tables and figures should be REFERENCED in the text. Authors MUST EXPLAIN what the reader should look for when using the table or figure. Focus only on the important point the reader should draw from them, and leave the details for the reader to examine on her own.

- Tables are to be presented with single horizontal line under: the table caption, the column headings and at the end of the table. All tables are produced by creating tables in MS Word. Captured tables are NOT allowed.

- All figures MUST in high quality images

(7) Conclusion section: Summarize sentences the primary outcomes of the study in a paragraph. Are the claims in this section supported by the results, do they seem reasonable? Have the authors indicated how the results relate to expectations and to earlier research? Does the article support or contradict previous theories? Does the conclusion explain how the research has moved the body of scientific knowledge forward?

(8) Most importantly, please ensure the similarity score is less than 25%. You can refer to EDAS to see the similarity score of your paper. Any paper with a similarity score of more than 25% will be dropped. Please make sure your revised paper follow this rule. If the similarity score of final version is more than 25%, the TPC has the right to cancel the paper to be presented at TELKOMNIKA.

(9) Please ensure the maximum page of your final paper is 8-page, but still allowed up to 12 pages (required to pay an extra fee).

I will immediately revise my paper according to the suggestions from you and also from the reviewers. Thank you for providing this invaluable advice.

Best Regards,

Hartono

[Kutipan teks disembunyikan]

---

**TELKOMNIKA Telecom Comp Elect and Ctrl** <telkomnika@uad.ac.id>

27 Agustus 2020 pukul 10.29

Kepada: hartonoibbi@gmail.com

Dear Sir/Madam,

Your email has reached my queues safely, so i will get back in touch with you there as soon as i can. Just a heads up i might not get back to you straight away, as i answer all emails in order to keep things fair.

**\*How to submit your manuscript\***

To make a submission, you must have a user account and be enrolled as an Author. User accounts can either be created by the Journal Manager or you can register yourself (this journal policy allow you create user account by your self as a Reader, an Author and/or a Reviewer). All fields with an asterisk beside them (Username; Password; Repeat Password; First Name; Last Name; Email) are mandatory. Your username and your email address must be unique; furthermore, while you can change your email address at a later date, you will be unable to change your username. If you want to register in another role within the same journal (for example, if you are already a Reader, but also want to become an Author) you can log in; go to Edit My Profile (under My Account on your User Home page); and check off the check-boxes next to any available roles, near the bottom of the page. Once you have an account, log in to the journal site and select the role of Author. The Author is asked to upload a submission file and to provide metadata or indexing information. (The metadata improves the search capacity for research online and for the journal.) The Author can upload Supplementary Files, in the form of data sets, research instruments, or source texts that will enrich the item, as well as contribute to more open and robust forms of research and scholarship.

**\*Your paper ID number?\***

Please ALWAYS put your paper ID number in the subject line of email communication when making inquiries. For example, if your submitted manuscript URL:  
<http://journal.uad.ac.id/index.php/TELKOMNIKA/author/submission/10234>, your paper ID number is 10234.

Your cooperation is very appreciated.

Thank you

Best Regards,

Tole Sutikno

Editor:

TELKOMNIKA TCEC, <http://journal.uad.ac.id/index.php/TELKOMNIKA>

---

-----  
Our events in 2020  
-----

2020 1st Conference on Internet of Things and Embedded Intelligence (CITEI 2020)

Venue: Yogyakarta, Indonesia

Date: July 29-30, 2020

Website: <http://citei.intconference.org>

Link of Paper submission: <https://edas.info/N27031>

Paper submission Deadline: February 20, 2020

2020 2nd ADICS Int Conference on Engineering, Science and Information Technology

Venue: Yogyakarta, Indonesia

Date: August 26-27, 2020  
Website: <http://esit.intconference.org>  
Link of Paper submission: <https://edas.info/N27031>  
Paper submission Deadline: March 20, 2020

2020 7th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2020)  
Venue: Yogyakarta, Indonesia  
Date: September 23-25, 2020  
Website: <http://eecs.org/2020/>  
Link of Paper submission: <https://edas.info/N27101>  
Paper submission Deadline: April 15, 2020

2020 3rd International Conference and Workshop on Telecommunication, Computing, Electrical, Electronics and Control (ICW-TELKOMNIKA 2020)  
Venue: Yogyakarta, Indonesia  
Date: November 19-21, 2020, 2020  
Website: <http://icw.telkomnika.com>  
Link of Paper submission:  
<http://journal.uad.ac.id/index.php/TELKOMNIKA> (section ICW....)  
Paper submission Deadline: June 15, 2020

Kindly forward this email to other interested parties.

--

Best Regards,  
Tole Sutikno  
\*Editor-in-Chief, TELKOMNIKA (Telecommunication, Computing, Electronics and Control) \*  
ISSN: 1693-6930, e-ISSN: 2302-9293  
<http://journal.uad.ac.id/index.php/TELKOMNIKA>

--

\*-----\*

\*\*UNIVERSITAS AHMAD DAHLAN <<http://uad.ac.id>>\*

\*Kampus 1:\* Jln. Kapas No. 9  
Yogyakarta

\*Kampus 2:\* Jl. Pramuka 42, Sidikan, Umbulharjo, Yogyakarta  
55161

\*Kampus 3:\* Jl. Prof. Dr. Soepomo, S.H., Janturan, Warungboto,  
Umbulharjo, Yogyakarta 55164

\*Kampus 4:\* Jl. Ringroad Selatan, Yogyakarta

\*Kampus 5: \*Jl. Ki Ageng Pemanahan 19, Yogyakarta

\*\*

\*

\*Kontak\*

\*Email:  
[info@uad.ac.id](mailto:info@uad.ac.id) <<mailto:info@uad.ac.id>>  
Telp. : (0274) 563515, 511830,  
379418, 371120  
Fax. : (0274) 564604\*

## **Revisi Atas Masukan yang Diberikan Reviewer**

# Hybrid Approach Redefinition-Multi Class with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise

Erianto Ongko<sup>1</sup>, Hartono<sup>2,3</sup>

<sup>1</sup>Department of Informatics, Akademi Teknologi Industri Immanuel, Indonesia

<sup>2</sup>Department of Computer Science, Universitas IBBI, Indonesia

<sup>3</sup>Department of Computer Science, Universitas Potensi Utama, Indonesia

## Article Info

### Article history:

Received Jun x, 20xx

Revised Nov x, 20xx

Accepted Dec x, 20xx

### Keywords:

Class Imbalance

Multi-Class

Overlapping

Noise

Classifier Performance

## ABSTRACT

The handling of class imbalance and overlapping on multi-class problems is one of the most important issues in data classification. Both of these problems can reduce the performance and accuracy of the classification. Another problem that must also be considered is handling noise, because noise can also reduce the performance of classification. With the aid of a resampling algorithm and feature selection, this paper proposes a method for improving the performance of Hybrid Approach Redefinition-Multi Class (HAR-MI). Resampling algorithm can overcome the problem of noise but cannot handle overlapping well. On the other hand, feature selection is good at dealing with overlapping but can experience a decrease in quality if there is a noise. The HAR-MI Approach is a great way to deal with Multi-Class Imbalance issues, but it has some drawbacks when dealing with overlapping. The key contribution of this paper is to suggest a new approach for dealing with class imbalance, overlapping, and noise in multi-class. In terms of overlapping handling and Classifier Performance, this approach must be capable of producing good results. This is accomplished by employing Minimizing Overlapping Selection (MOSS) as an ensemble learning algorithm and a preprocessing technique in HAR-MI, as well as employing Multi-Class Combination Cleaning and Resampling (MC-CCR) as a resampling algorithm at the processing stage to substitute Random Under Sampling. When subjected to overlapping and classifier performance, it is discovered that the proposed method produces good results, as evidenced by higher Augmented R-Value, Class Average Accuracy, Class Balance Accuracy, Multi Class G-Mean, and Confusion Entropy.

*This is an open access article under the [CC BY-SA](#) license.*



**Comment [Office1]:** Please revise the academic writing style and grammars throughout, for examples (but not limited to) the first 3 sentences in the abstract, some broken sentences, casual phrases, etc

## Corresponding Author:

Erianto Ongko,  
Department of Informatics,  
Akademi Teknologi Industri Immanuel,  
Jalan Jend. Gatot Subroto No. 325 Medan.  
Email: eriantoongko@gmail.com

## 1. INTRODUCTION

The class imbalance occurs when a class has a significantly smaller number of instances than other classes, as determined by the Imbalance Ratio (IR), which is the ratio of a class with a significantly smaller number of instances (minority class) to a class with a significantly larger number of instances (majority class) (majority class)[1] and basically machine learning algorithms work optimally if each class has a number of instances that are not much different[2]. This problem is one of the causes of the low accuracy of

classification problems and also causes important information contained in the minority class can not be obtained due to better coverage on the majority class[3]. Handling of multi-class imbalance has greater difficulty compared to two-class problems, especially when it comes to accuracy and difficulty of training data on large datasets with high imbalance ratios[4]. Another thing that escapes attention is the overlapping problem, where several classes overlap with other classes. [The overlapping problem has far more impact on accuracy compared to class imbalance[5]. Overlapping conditions can increase the accuracy of one class by decreasing the accuracy of another class. For example, although the overlapping regions have a high concentration of minority classes, the classification results can also provide low accuracy because some instances associated with majority classes are eliminated[6].

The overlapping problem in multi-class can be overcome by using feature selection method which is very effective in dealing with overlapping problems[7]. [The reason why feature selection is effective in dealing with overlapping is because of its ability to eliminate uninformative predictors and reduce dimensionality of feature space[8]. However, on the other hand with the noise, the performance given by Feature Selection can decrease[9] and noise basically has an influence on classification performance[10]. Noise handling in general uses the method of resampling, but often encounters obstacles if there is a state of overlapping[6]. This is an obstacle when handling multi-class imbalance and at the same time also faces other obstacles in the form of overlapping and noise. A number of studies have discussed handling class imbalance accompanied by overlapping or noise. [11] has proposed the MC-CCR method which has the ability to overcome the noise problem but has obstacles in dealing with overlapping. [The Feature Selection method on the other hand has been used by a number of researchers in dealing with overlapping problems, such as: [12] that has proposed Density Based Feature Selection and [13] that has proposed RSFSAID Algorithm.

The Ensemble Learning approach, especially the Hybrid Ensembles, is very commonly used in overcoming multi-class imbalance problems[14]. The Hybrid Ensembles approach has a good ability in handling multi-class imbalances accompanied by overlapping, but what needs to be considered is in situations where the imbalance ratio is high and also conditions that contain noise and overlapping[15]. [16] has stated a number of things that need to be considered in improving performance on Hybrid Ensembles, such as the use of the right selection method on the noise label and also the right sampling at the processing stage. Research conducted by [17] and [18] shows that using the appropriate feature selection method at the preprocessing stage in the hybrid ensembles can provide a good result in handling class imbalance and overlapping. Noise handling in Hybrid Ensembles can be overcome by choosing the right sampling method at the processing stage[19][20].

[The Hybrid Ensembles approach that combines the application of preprocessing by using feature selection and sampling at the processing stage is Hybrid Approach Redefinition-Multiclass Imbalance[21]. The HAR-MI approach will be combined with the Resampling Algorithm in the processing stage and Feature Selection in the preprocessing stage in this analysis. Selection Under No Sampling[22][23] and Selection Under SMOTE[24] are two feature selection approaches that can be used to overcome overlapping problems. According to research conducted by [5] the Minimizing Overlapping Selection under No-Sampling (MOSNS) and Minimizing Overlapping Selection under SMOTE (MOSS) methods provide very satisfactory results in dealing with overlapping. Both methods have a similar level of efficiency.

Noise handling by using resampling at the processing stage, there have been a number of studies that have been done. The research conducted by [25] uses the SMOTE Sampling method in handling noise but this research has problems with overlapping classes and also has problems in accuracy. The same thing is also found in using the SMOTE Oversampling with Edited Nearest Neighbors (ENN) method[26]. The Multi-Class Combination Cleaning and Resampling (MC-CCR) system is One Sampling method that produces excellent results.

HAR-MI Method which has good ability in overcoming multi-class imbalance problems but the result would be worse if there are overlapping between class and noise. In the preprocessing step, the HAR-MI system employs the Random Balance Ensemble method, which will be paired with the MOSS method for the preprocessing stage, as well as the MC-CCR method for processing step. The findings will be compared to those obtained using the Neighbourhood-Based Undersampling process, which is one of the best techniques for dealing with class imbalance and overlapping in multi-class imbalanced conditions[18]. Augmented R-Value, Class Average Accuracy, Class Balance Accuracy, Multi Class G-Mean, and Uncertainty Entropy were used to compare these results.

**Comment [Office2]:** Please restructure the motivation parts, regarding the imbalance and overlapping and make them more focus. For example (but not limited to), several statements are dedicated to problems caused by overlapping, but none of them is more specific than others. This pattern repeats for most parts of the introduction. Please revise. Moreover, feature selection algorithms are mentioned but unfortunately not so specific nor much discussed. Please elaborate.

**Comment [Office3]:** Please restructure the motivation parts, regarding the imbalance and overlapping and make them more focus. For example (but not limited to), several statements are dedicated to problems caused by overlapping, but none of them is more specific than others. This pattern repeats for most parts of the introduction. Please revise. Moreover, feature selection algorithms are mentioned but unfortunately not so specific nor much discussed. Please elaborate.

**Comment [Office4]:** Introduction section: explain the context of the study and state the precise objective. An Introduction should contain the following three parts:  
 - Background: Authors have to make clear what the context is. Ideally, authors should give an idea of the state-of-the art of the field the report is about.  
 - The Problem: If there was no problem, there would be no reason for writing a manuscript, and definitely no reason for reading it. So, please tell readers why they should proceed reading. Experience shows that for this part a few lines are often sufficient.  
 - The Proposed Solution: Now and only now! - authors may outline the contribution of the manuscript. Here authors have to make sure readers point out what are the novel aspects of authors work.  
 Authors should place the paper in proper context by citing relevant papers. At least, 15 references (recently journal articles) are referred in this section.

**2. RELATED WORKS**

**2.1. Augmented R-Value for Multi-Class**

Each class's R-Value shows how much of an instance overlaps the area. R-Value has a strong correspondence with classifier performance, according to research conducted by [27]. As can be seen in Equation 1, [23] has suggested a method for calculating this.

$$R_{aug}(D[V]) = \frac{\sum_{i=0}^{k-1} |C_{k-1-i}| R(C_i)}{\sum_{i=0}^{k-1} |C_i|} \tag{1}$$

Where  $C_0, C_1, \dots, C_{k-1}$  are  $k$  class labels with  $|C_0| \geq |C_1| \geq \dots \geq |C_{k-1}|$  and  $D[V]$ : Dataset  $D$  restraining predictors in set  $V$ . A higher  $R_{Aug}$  indicates a higher overlap degree.

**2.2. Confusion Matrix**

For the general classification results the classification results can be grouped into 4 (four) groups, namely: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) and can be presented in the Confusion Matrix as can be seen in Table 1[28].

Table 1. Confusion Matrix For A Classification Problem

		Predicted (Classified) as	
		Positive	Class Negative
Actually (Really) is	Positive Samples	TP	FN
	Negative Samples	FP	TN

**2.3. Classifier Performance**

The following parameters are used to determine the classifier performance.

- Class Average Accuracy with  $C$  classes can be calculated using Equation 2[29].

$$AvAcc = \frac{1}{C} \sum_{i=1}^C \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i} \tag{2}$$

Where  $C$  is number of class with  $TP, TN, FP,$  and  $FN$  are the result of predicted (classified) that was obtained from confusion matrix.

- Class Balance Accuracy for any  $C^k$ , confusion matrix, Class Balance Accuracy is defined as[30].

$$CBA = \frac{\sum_{i=1}^k \frac{c_{ii}}{\max(c_{i,j}, c_{j,i})}}{k} \tag{3}$$

According to Confusion Matrix for Class Balance Accuracy as can be seen in Table 2.

Table 2. Confusion Matrix For Class Balance Accuracy

		Predicted			
		Class 1	Class 2	Class 3	
Actual	Class 1	$C_{11}$	$C_{12}$	$C_{13}$	$C_{.1}$
	Class 2	$C_{21}$	$C_{22}$	$C_{23}$	$C_{.2}$
	Class 3	$C_{31}$	$C_{32}$	$C_{33}$	$C_{.3}$
	Total	$C_{.1}$	$C_{.2}$	$C_{.3}$	N

- The G-mean was proposed as the geometric mean of recall (R) values of two groups by Multi Class G-Mean (mGM) [31], [32] To apply this measure to multiple-class situations, define the G-mean as the geometric mean of each class's recall values.

$$R_i = \frac{c_{ii}}{\sum_{j=1}^k c_{ij}} \tag{4}$$

$$mGM = \sqrt[c]{\prod_{i=1}^c R_i} \tag{5}$$

4. Confusion Entropy (CEN). [33] proposed using the Confusion Entropy to determine classifier efficiency. According to the confusion entropy, the misclassification information includes both how the samples with true class label  $c_l$  were misclassified to the other N classes and how the samples from the other N classes were misclassified to class  $c_l$ .

$$CEN = \sum_j P_j CEN_j \quad (6)$$

Where

$$P_j = \frac{\sum_k (C_{j,k} + C_{k,j})}{2 \sum_{k,l} C_{k,l}} \quad (7)$$

$$CEN_j = - \sum_{k=1, k \neq j}^{N+1} (p_{j,k}^i \log_2 p_{j,k}^i + p_{k,j}^i \log_2 p_{k,j}^i) \quad (8)$$

#### 2.4. Minimizing Overlapping Selection under SMOTE (MOSS)

The MOSS algorithm is as follows [5].

- 1:  $X$  – matrix with  $p$  predictors:  $X = [x_1, x_2, \dots, x_p]$ ; class label:  $y$
- 2: Over-sampling the Positive Samples using SMOTE; merging the generated instances with original ones to get updated  $X$ -matrix,  $X_{new}$  and updated class label  $Y_{new}$
- 3:  $X \leftarrow X_{new}$ ;  $Y \leftarrow Y_{new}$
- 4: Establish sparse regularization path  $\hat{\beta}(\lambda, \alpha)$  according to Equation 9
- 5: Compute the optimal  $(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^T$  via the Equation 10
- 6: Select those feature with  $\hat{\beta}_j \neq 0$  for  $j = 1, 2, \dots, p$

Equation 9 shows the sparse collection that would be used to create sparse regularization[34].

$$C_\alpha(\beta) = \frac{1}{2} (1 - \alpha) \|\beta\|_2^2 + \alpha \|\beta\|_1 \quad (9)$$

Equation 10 shows the Loss Penalties that will be used to determine the optimum  $\hat{\beta}_j$ .

$$Loss = - \frac{1}{n} \sum_{i=1}^n (y_i \beta^T x_i - \ln(1 + \beta^T x_i)) \quad (10)$$

The method of handling overlapping has started at the preprocessing stage by adding MOSS, as seen in the previous pseudocode. The MOSS process begins with the provision of  $p$  predictors and class labels  $y$ . The oversampling process in the minority class will be carried out using SMOTE, and then the sparse regularization process will be carried out using sparse collection, which can be measured using Equation 2, and then Loss Penalties will be calculated using Equation 3.

#### 2.5. Multi-Class Combined Cleaning and Resampling (MC-CCR)

The MC-CCR algorithm is as follows [11].

- Input:  $x^{(c)}$  denotes a subcollection of observations belonging to class in the Set of Observations  $x$ .  
Parameters: Each sphere has an energy budget for expansion, and the  $p$ -norm is used to calculate distance  
Output: Observations that have been translated and oversampled  $X$
- 1: Function MC-CCR ( $X$ , energy,  $p$ );
  - 2:  $C \leftarrow$  Collection of all classes; sorted by the number of associated observations in a descending order
  - 3: for  $i \leftarrow 1$  to  $|c|$
  - 4:  $n_{classes} \leftarrow$  number of classes with high number of observations than  $C_i$
  - 5: if  $n_{classes} > 0$  then
  - 6:  $X_{min} \leftarrow X^{(C_i)}$
  - 7:  $X_{maj} \leftarrow \emptyset$
  - 8: for  $j \leftarrow 1$  to  $n_{classes}$
  - 9: add  $\left\lfloor \frac{|X^{(C_j)}|}{n_{classes}} \right\rfloor$  randomly selected from  $x^{(j)}$  to  $X_{maj}$
  - 10: end
  - 11:  $X_{maj}, S \leftarrow$  CCR ( $X_{maj}$ ,  $X_{min}$ , energy,  $P$ )
  - 12:  $X^{(C_i)} \leftarrow X^{(C_i)} \cup S$
  - 13: Substitute observation used to construct  $X_{maj}$  with  $X'_{maj}$
  - 14: end if
  - 15: end

16: *return X*

The MC-CCR method is used to eliminate noise at the processing stage. It should be noted that the impact is limited and basically the combined majority observations.

### 2.6. Hybrid Approach Redefinition for Multi-Class Imbalance (HAR-MI)

The algorithm of Hybrid Approach Redefinition for Multi-Class Imbalance as follows[21].

*Require: Set S of examples (x<sub>i</sub>, y<sub>i</sub>)*

*Ensure: New set S' of examples with Random Balance Ensemble Method*

```

1: totalSize ← |S|
2: Determine k using Dynamic Ensemble Selection
3: Building the candidate ensemble for Safe, Borderline, Rare, and Outlier
4: For all samples in Majority and Minority
5:   Preprocessing Stage using Random Balance Ensemble Method
6:   New Majority and New Minority of Preprocessing Dataset
7: End
8: Determine the Augmented R-Value
9: For all instances in Preprocessing Dataset
10:  Determine Majority and Minority Class
11:  For All Instances in Majority Class
12:    Biased Support Vector Machine for Determine SV Sets and NSV Sets
13:  End
14:  For All Instances in Minority Class
15:    Biased Support Vector Machine for Determine SV Sets and NSV Sets
16:  End
17: End
18: For All Instances in NSV Sets from Majority Class
19:  Process Multiple Random Under Sampling
20: End
21: For All Instances in SV Sets from Minority Class
22: Process SMOTEBoost
23: End

```

Based on the preceding algorithm, it is clear that the HAR-MI method is divided into 2 (two) major stages: preprocessing and processing. The Random Balance Ensemble Method and Dynamic Ensemble Selection are used in the preprocessing stage. It is clear that the preprocessing stages will generate preprocessing datasets, which will then go through processing stages using Different Contribution Sampling. There are Biased Support Vector Machine stages in Different Contribution Sampling that will produce SV Sets and NSV Sets for both the majority and minority classes. NSV Sets from majority classes are then processed using Multiple Random Under Sampling, while SV Sets from minority classes are processed using SMOTEBoost.

### 3. RESEARCH METHOD

Figure 1 shows the stages of this research.

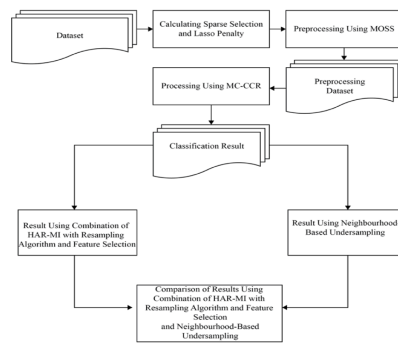


Figure 1. Stages of Research Methods

According to the previous Figure, the process will start with the preprocessing stage, which employs MOSS. The Sparse Selection and Lasso Penalty values are determined first. This stage's output will be a preprocessing dataset, which will then go through processing stages using MC-CCR. The results from HAR-MI with Resampling and Feature Selection will then be compared to the results from Neighborhood-Based Undersampling.

### 3.1. Preprocessing Stage

MOSS will be used to modify the HAR preprocessing stage for multi-class problems. The following algorithm depicts the preprocessing stages.

Require:  $S$  as set of instances  $(X_i, Y_i)$

Ensure:  $S'$  is the new Set with MOSS

- 1:  $totalSize \leftarrow |S|$
- 2: Calculate  $k$  as the number of Nearest Neighbors
- 3: For All Instances of  $S$
- 4: Building the Positive or Minority Class Borderline as  $E_0C_t^+$
- 5: Building the Negative or Majority Class Borderline as  $E_0C_t^-$
- 6: End
- 7: Creating a candidate ensemble based on  $k$  value for safe, borderline, rare, and outlier candidates
- 8: Using Equation 9 to Calculate Sparse Selection
- 9: Using Equation 10 to Calculate the Loss Penalty
- 10: Determine Sparse Regularization
- 11: For All Instances in Positive
- 12: MOSS is used to sample all instances
- 13: Build newMinority
- 14: Build newMajority
- 15: End
- 16: Determine the Augmented R-Value
- 17: return  $S'$

Based on the preceding algorithm, it is clear that the HAR-MI preprocessing stage will be carried out using one of the Feature Selection methods, namely MOSS. MOSS is intended to do overlapping handling before entering the processing stage. This MOSS stage begins with determining the value of Sparse Selection and Loss Penalty. Furthermore, MOSS will be used to sample each instance in the minority class. The result is a preprocessing dataset which will then be measured in Augmented R-Value values.

### 3.2. Processing Stage

The following algorithm depicts the processing steps.

- 1: For all samples in preprocessed dataset
- 2: Preprocessed Dataset should be added to  $S_i$
- 3: Using B-SVM, determine the SV and NSV sets for the majority and minority
- 4: For All Samples in Negative

```

5:   Checking and removing noise from SV and NSV sets with MC – CCR
6:   End
7:   For All Samples in Positive
8:     Check and remove the Noise in SV Sets and NSV Sets using MC – CCR
9:     SMOTEBoost Step for SV Sets and Produce SMOTEsets
10:  End For
11:  For All SV Sets and NSV Sets from Majority Class do
12:    New NegativeSampleSets
13:  End For
14:  For All SV Sets and NSV Sets from Minority Class do
15:    New PositiveSampleSets
16:  End For
17: End For

```

According to the previous algorithm, the processing stage begins with the biased support vector machine process to determine SV Sets and NSV Sets for the majority and minority classes. The next step is for each SV Set and NSV Set in the majority class to go through the process of noise removal and resampling using the MC-CCR. On minority classes, the same thing will be done with SV Sets and NSV Sets. The SMOTEBoost process will be applied to SV Sets in the minority class in particular. This whole process will result in a result dataset.

## 4. RESULTS AND ANALYSIS

### 4.1. Dataset Description

We conducted our experiments using 6 (six) multi-class imbalanced datasets from the KEEL Repository, each with a low, moderate, or high. For datasets with a low IR are New-Thyroid and Balance, datasets with moderate IR are Flare and Car, and dataset with high IR are Red Wine Quality and Yeast. Table 3 contains a description of the dataset[35].

Table 3. Description of Dataset[35]

Dataset	#Ex	#Atts	Distribution of Class	IR
New-Thyroid	215	5	150/35/30	5
Balance	625	4	288/49/288	5.88
Flare	1066	11	331/239/211/147/95/43	7.7
Car	1728	6	65/69/384/1210	18.61
Red Wine Quality	1599	11	10/53/681/638/199/18	68.1
Yeast	1484	8	463/5/35/44/51/163/244/429/20/30	92.6

Following the selection of the dataset, the next step is to assess the presence of noise. This experiment will use a subset of training examples and randomly replace their labels to generate noise. This experiment will use a noise level of 0.1.

### 4.2. Testing Result

The first test compares the Augmented R-Value and Class Average Accuracy obtained by using the HAR-MI with Resampling Algorithm and Feature Selection. Table 4 shows the test results.

Table 4. Augmented R-Value and Class Average Accuracy Testing Results

Dataset	HAR-MI with Resampling Algorithm and Feature Selection		Neighbourhood-Based Undersampling	
	Augmented R-Value	Class Average Accuracy	Augmented R-Value	Class Average Accuracy
New-Thyroid	0.335	0.972	0.341	0.933
Balance	0.327	0.865	0.345	0.891
Flare	0.367	0.713	0.359	0.721
Car	0.361	0.725	0.373	0.693
Red Wine Quality	0.428	0.687	0.415	0.674
Yeast	0.448	0.623	0.452	0.615

*HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise (Erianto Ongko)*

**Comment [Office5]:** Method section: the presentation of the experimental methods should be clear and complete in every detail facilitating reproducibility by other scientists.

According to Table 4, the results obtained by the HAR-MI method with Resampling Algorithm and Feature Selection and Neighborhood-Based Undersampling are not significantly different in terms of overlapping. This is indicated by the value of Augmented R-Value which is not much different. The lower the Augmented R-Value, the lower the overlapping level. There is a strong relationship between overlapping and accuracy. The lower the overlapping, the better the average class accuracy obtained. It should also be noted that Neighborhood-Based Undersampling tends to have a slight advantage in datasets with a large number of attributes such as Flare and Red Wine Quality. The HAR-MI with Resampling Algorithm and Feature Selection has the advantage of 4 other datasets. It should be noted that the imbalance ratio has an impact on the results. The higher the imbalance ratio, the more overlapping there will be, and the accuracy obtained will also be lower.

The second test compares the Class Balance Accuracy, Multi Class G-Mean, and Confusion Entropy obtained by using the HAR-MI method with Resampling Algorithm and Feature Selection, as well as Neighborhood-Based Undersampling. Table 5 shows the test results.

Table 5. Class Balance Accuracy, Multi Class G-Mean, and Confusion Entropy Testing Results for Each Method

Dataset	HAR-MI with Resampling Algorithm and Feature Selection			Neighbourhood-Based Undersampling		
	Class Balance Accuracy	Multi Class G-Mean	Confusion Entropy	Class Balance Accuracy	Multi Class G-Mean	Confusion Entropy
New-Thyroid	0.913	0.898	0.091	0.897	0.875	0.11
Balance	0.927	0.915	0.105	0.911	0.897	0.12
Flare	0.691	0.715	0.292	0.676	0.687	0.282
Car	0.897	0.875	0.311	0.874	0.815	0.327
Red Wine Quality	0.516	0.467	0.472	0.498	0.465	0.493
Yeast	0.495	0.512	0.526	0.476	0.498	0.523

Based on the Table 4, it is obvious that the number of attributes, the number of classes, and the level of IR all have a significant impact on class balance accuracy. The number of attributes and classes will largely determine the results of class balance accuracy for datasets with similar imbalance ratio levels. This can be seen in the Dataset Balance results for improved class balance accuracy when compared to the New-Thyroid dataset. When it comes to Class Balance Accuracy, it can be seen that the HAR-MI With Resampling Algorithm and Feature Selection method produces better results than Neighborhood-Based Undersampling.

Test results for Multi Class G-Mean show that for both HAR-MI with Resampling Algorithm and Feature Selection and Neighborhood-Based Undersampling, the higher the IR, the lower the Multi Class G-Mean value obtained because G- Means stated the equilibrium between positive samples and negative samples.

The test results for Confusion Entropy show that the results obtained depend on the number of classes and imbalance ratios. The number of classes determines the results obtained for imbalance ratios that are not significantly different, such as those in the Flare and Car Datasets. In general, the results obtained by the two methods for Confusion Entropy are not significantly different.

#### 4.3. Statistical Tests

The Wilcoxon signed-rank test is used to perform the statistical test, which is a statistical procedure in order to assess performance on the basis of pairwise comparisons [36]. The Result for statistical tests can be seen in Table 6.

Table 6. Wilcoxon Signed-Rank Test in Order to Assess Performance

Performance Measurement		P-Value	Hypothesis
Augmented R-Value		0.687500	$H_0$ (no significant difference in score between HAR-MI with Resampling Algorithm, Feature Selection, and Neighbourhood-Based Undersampling) is accepted, which means $H_1$ (significant difference in score between HAR-MI with Resampling Algorithm, Feature Selection, and Neighbourhood-Based Undersampling) is rejected because the p-value is greater than 0.05)
Class Accuracy	Average	0.344118	$H_0$ (no significant difference in score between HAR-MI with Resampling Algorithm, Feature Selection, and Neighbourhood-Based Undersampling) is accepted, which means $H_1$ (significant difference in score between HAR-MI with Resampling Algorithm, Feature Selection, and Neighbourhood-Based Undersampling) is rejected because the p-value is greater than 0.05)
Class Accuracy	Balance	0.03552	$H_0$ (no significant score difference between HAR-MI with Resampling Algorithm and Feature Selection and Neighbourhood-Based Undersampling) rejected, which means $H_1$ (there is a significant difference between HAR-MI with Resampling Algorithm and Feature Selection and Neighbourhood-Based Undersampling in score) Accepted because the p-value is less than 0.05
Multi Class G-Mean		0.0312500	$H_0$ (no significant score difference between HAR-MI with Resampling Algorithm and Feature Selection and Neighbourhood-Based Undersampling) rejected, which means $H_1$ (there is a significant difference between HAR-MI with Resampling Algorithm and Feature Selection and Neighbourhood-Based Undersampling in score) Accepted because the p-value is less than 0.05
Confusion Entropy		0.156250	$H_0$ (no significant difference in score between HAR-MI with Resampling Algorithm, Feature Selection, and Neighbourhood-Based Undersampling) is accepted, which means $H_1$ (significant difference in score between HAR-MI with Resampling Algorithm, Feature Selection, and Neighbourhood-Based Undersampling) is rejected because the p-value is greater than 0.05)

#### 4.4. Discussion

According to the results in Tables 4, 5, 6, there is no significant distinction in Augmented R-Value, Class Average Accuracy, and Confusion Entropy between HAR-MI with Resampling Algorithms and Feature Selection and Neighborhood-Based Undersampling. It indicates that both methods have successfully overcome overlapping with positive outcomes. The Confusion Entropy obtained is good and this means that the misclassification is spread evenly for all classes. The test results for Class Balance Accuracy and Multi Class G-Mean show that HAR-MI with Resampling Algorithms and Feature Selection gives better results compared to Neighborhood-Based Undersampling.

It should be noted that overlapping and accuracy are two interrelated things, the higher the overlapping, the lower the accuracy. The imbalance ratio is the main factor that determines how much overlap there is. The higher the imbalance ratio, the higher the overlapping will be. In terms of overlapping, HAR-MI with Resampling Algorithms and Feature Selection has few limitations on datasets with a large number of attributes, in addition to imbalance ratios. The results revealed that, in addition to IR, the number of attributes and the number of classes determined the value of Class Balance Accuracy. The number of classes and the imbalance ratio have a strong influence on Multi Class G-Mean and Confusion Entropy.

#### 5. CONCLUSION

Based on the test results that for handling multi-class imbalances accompanied by overlapping and noise, the results obtained by HAR-MI with Resampling Algorithms and Feature Selection and Neighborhood-Based Undersampling are good. The results obtained by HAR-MI with Resampling Algorithms and Feature Selection are generally better than Neighborhood-Based Undersampling. This is indicated by better Augmented R-Value, Class Average Accuracy, Class Balance Accuracy, Multi-Class G-Mean, and Confusion Entropy. Although statistically for Augmented R-Value, Class Average Accuracy, and Confusion Entropy based on the test results statistically it does not have too significant differences. It should be noted that HAR-MI with Resampling Algorithms and Feature Selection and Neighborhood-Based Undersampling has limitations in handling overlapping, where there is a slight decrease in performance in datasets with large numbers of attributes. Imbalance ratio also has a direct relationship with the performance classifier. Future research is expected to be able to handle the decrease in performance in datasets with large number of attributes and also a high IR.

**Comment [Office6]:** Results and discussion section: The presentation of results should be simple and straightforward in style. This section report the most important findings, including results of statistical analyses as appropriate and comparisons to other research results. Results given in figures should not be repeated in tables. This is where the author(s) should explain in words what he/she/they discovered in the research. It should be clearly laid out and in a logical sequence. This section should be supported suitable references.

## ACKNOWLEDGEMENTS



This study was funded by a fund of the Republic of Indonesia's Ministries of Education and Culture, as well as the Ministry of Research and Technology.

## REFERENCES

- [1] B. Liu and G. Tsoumakas, "Dealing with class imbalance in classifier chains via random undersampling," *Knowledge-Based Systems*, vol. 192, p. 105292, Mar. 2020, doi: 10.1016/j.knsys.2019.105292.
- [2] N. Santoso, W. Wibowo, and H. Hikmawati, "Integration of synthetic minority oversampling technique for imbalanced class," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 1, Art. no. 1, Jan. 2019, doi: 10.11591/ijeecs.v13.i1.pp102-108.
- [3] J. Hamidzadeh, N. Kashefi, and M. Moradi, "Combined weighted multi-objective optimizer for instance reduction in two-class imbalanced data problem," *Engineering Applications of Artificial Intelligence*, vol. 90, p. 103500, Apr. 2020, doi: 10.1016/j.engappai.2020.103500.
- [4] C.-M. Vong and J. Du, "Accurate and efficient sequential ensemble learning for highly imbalanced multi-class data," *Neural Networks*, vol. 128, pp. 268–278, Aug. 2020, doi: 10.1016/j.neunet.2020.05.010.
- [5] G.-H. Fu, Y.-J. Wu, M.-J. Zong, and L.-Z. Yi, "Feature selection and classification by minimizing overlap degree for class-imbalanced data in metabolomics," *Chemometrics and Intelligent Laboratory Systems*, vol. 196, p. 103906, Jan. 2020, doi: 10.1016/j.chemolab.2019.103906.
- [6] E. R. Q. Fernandes and A. C. P. L. F. de Carvalho, "Evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning," *Information Sciences*, vol. 494, pp. 141–154, Aug. 2019, doi: 10.1016/j.ins.2019.04.052.
- [7] A. Fernández, M. J. del Jesus, and F. Herrera, "Addressing Overlapping in Classification with Imbalanced Datasets: A First Multi-objective Approach for Feature and Instance Selection," in *Intelligent Data Engineering and Automated Learning – IDEAL 2015*, Cham, 2015, pp. 36–44, doi: 10.1007/978-3-319-24834-9\_5.
- [8] A. Wahid *et al.*, "Feature selection and classification for gene expression data using novel correlation based overlapping score method via Chou's 5-steps rule," *Chemometrics and Intelligent Laboratory Systems*, vol. 199, p. 103958, Apr. 2020, doi: 10.1016/j.chemolab.2020.103958.
- [9] B. Frénay, G. Doquire, and M. Verleysen, "Estimating mutual information for feature selection in the presence of label noise," *Computational Statistics & Data Analysis*, vol. 71, pp. 832–848, Mar. 2014, doi: 10.1016/j.csda.2013.05.001.
- [10] S. Uyun and E. Sulistyowati, "Feature selection for multiple water quality status: Integrated bootstrapping and SMOTE approach in imbalance classes," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, Art. no. 4, Aug. 2020, doi: 10.11591/ijece.v10i4.pp4331-4339.
- [11] M. Koziarski, M. Woźniak, and B. Krawczyk, "Combined Cleaning and Resampling algorithm for multi-class imbalanced data with label noise," *Knowledge-Based Systems*, vol. 204, p. 106223, Sep. 2020, doi: 10.1016/j.knsys.2020.106223.
- [12] M. Alibeigi, S. Hashemi, and A. Hamzeh, "DBFS: An effective Density Based Feature Selection scheme for small sample size and high dimensional imbalanced data sets," *Data & Knowledge Engineering*, vol. 81–82, pp. 67–103, Nov. 2012, doi: 10.1016/j.datak.2012.08.001.
- [13] H. Chen, T. Li, X. Fan, and C. Luo, "Feature selection for imbalanced data based on neighborhood rough sets," *Information Sciences*, vol. 483, pp. 1–20, May 2019, doi: 10.1016/j.ins.2019.01.041.
- [14] H. Ali, M. N. M. Salleh, R. Saedudin, K. Hussain, and M. F. Mushtaq, "Imbalance class problems in data mining: a review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, Art. no. 3, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1552-1563.
- [15] R. Alejo, R. M. Valdivinos, V. Garcia, and J. H. Pacheco-Sanchez, "A hybrid method to face class overlap and class imbalance on neural networks and multi-class scenarios," *Pattern Recognition Letters*, vol. 34, no. 4, pp. 380–388, Mar. 2013, doi: 10.1016/j.patrec.2012.09.003.
- [16] W. Xie, G. Liang, Z. Dong, B. Tan, and B. Zhang, "An Improved Oversampling Algorithm Based on the Samples' Selection Strategy for Classifying Imbalanced Data," *Mathematical Problems in Engineering*, pp. 1–13, 2019, doi: <https://doi.org/10.1155/2019/3526539>.
- [17] F. N. Koutanaei, H. Sajedi, and M. Khanbabaei, "A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring," *Journal of Retailing and Consumer Services*, vol. 27, pp. 11–23, Nov. 2015, doi: 10.1016/j.jretconser.2015.07.003.
- [18] P. Vuttipittayamongkol and E. Elyan, "Neighbourhood-based undersampling approach for handling imbalanced and overlapped data," *Information Sciences*, vol. 509, pp. 47–70, Jan. 2020, doi: 10.1016/j.ins.2019.08.062.
- [19] W. A. Rivera, "Noise Reduction A Priori Synthetic Over-Sampling for class imbalanced data sets," *Information Sciences*, vol. 408, pp. 146–161, Oct. 2017, doi: 10.1016/j.ins.2017.04.046.
- [20] M. Koziarski, B. Krawczyk, and M. Woźniak, "Radial-Based oversampling for noisy imbalanced data classification," *Neurocomputing*, vol. 343, pp. 19–33, May 2019, doi: 10.1016/j.neucom.2018.04.089.
- [21] H. Hartono, Y. Risyani, E. Ongko, and D. Abdullah, "HAR-MI method for multi-class imbalanced datasets," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, Art. no. 2, Apr. 2020, doi: 10.12928/telkommika.v18i2.14818.
- [22] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

- [23] J. Fan and R. Li, "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [25] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.
- [26] D. Angluin and P. Laird, "Learning From Noisy Examples," *Machine Learning*, vol. 2, no. 4, pp. 343–370, Apr. 1988, doi: 10.1023/A:1022873112823.
- [27] S. Oh, "A new dataset evaluation method based on category overlap," *Comput. Biol. Med.*, vol. 41, no. 2, pp. 115–122, Feb. 2011, doi: 10.1016/j.compbiomed.2010.12.006.
- [28] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A Review," *International Journal of Advances in Soft Computing and Its Application*, vol. 7, no. 3, pp. 176–204, 2015.
- [29] P. Branco, L. Torgo, and R. P. Ribeiro, "Relevance-Based Evaluation Metrics for Multi-class Imbalanced Domains," in *Advances in Knowledge Discovery and Data Mining*, Cham, 2017, pp. 698–710, doi: 10.1007/978-3-319-57454-7\_54.
- [30] L. Mosley, "A balanced approach to the multi-class imbalance problem," *Graduate Theses and Dissertations*, Jan. 2013, doi: <https://doi.org/10.31274/etd-180810-3375>.
- [31] M. Kubat, R. C. Holte, and S. Matwin, "Machine Learning for the Detection of Oil Spills in Satellite Radar Images," *Machine Learning*, vol. 30, no. 2, pp. 195–215, Feb. 1998, doi: 10.1023/A:1007452223027.
- [32] Y. Sun, M. S. Kamel, and Y. Wang, "Boosting for Learning Multiple Classes with Imbalanced Class Distribution," in *Sixth International Conference on Data Mining (ICDM'06)*, Dec. 2006, pp. 592–602, doi: 10.1109/ICDM.2006.29.
- [33] J.-M. Wei, X.-J. Yuan, Q.-H. Hu, and S.-Q. Wang, "A novel measure for evaluating classifiers," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3799–3809, May 2010, doi: 10.1016/j.eswa.2009.11.040.
- [34] H. Zou and T. Hastie, "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [35] J. Alcalá-Fdez *et al.*, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," *Soft Comput.*, vol. 13, no. 3, pp. 307–318, Feb. 2009, doi: 10.1007/s00500-008-0323-y.
- [36] F. Wilcoxon, "Individual Comparisons by Ranking Methods on JSTOR," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

#### BIOGRAPHIES OF AUTHORS

	<p>Erianto Ongko earned his Master's degree (2015) from Universitas Sumatera Utara and Bachelor's degree (2012) from STMIK IBBI. Both of them are in Computer Science. He serve as a lecturer in Akademi Teknologi Industri Immanuel in Medan, Indonesia. His research topic are in Machine Learning, AI, and Operational Research.</p>
	<p>Hartono earned his Doctoral's degree (2018) from Universitas Sumatera Utara, Master's degree (2010) from Universitas Putra Indonesia YPTK Padang, and Bachelor's degree (2008) from STMIK IBBI. All of them are in Computer Science. He serve as a lecturer in Universitas IBBI and Universitas Potensi Utama in Medan, Indonesia. His research topic are in Machine Learning, AI, and Operational Research.</p>



**Last Review**

**[BEEI 2021] Your paper #1570662772 ('HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise')**

2 pesan

tole@iaescore.com &lt;tole=iaescore.com@edas.info&gt;

17 Maret 2021 pukul 22.01

Balas Ke: tole@iaescore.com

Kepada: Erianto Ongko &lt;eriantoongko@gmail.com&gt;, Hartono Hartono &lt;hartonoibbi@gmail.com&gt;

- Please Strictly use & follow to the template Manuscripts: <http://iaescore.com/gfa/beeI.docx>
  - Number of minimum references for original research paper is 25 references (and minimum 20 recently journal articles).
  - Number of minimum references for review paper is 50 references (and minimum 40 recently journal articles).
  - Authors MUST upload their Final manuscript through EDAS online system!!
- =====

Dear Mr. Erianto Ongko,

Congratulations - your paper #1570662772 ('HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise') for the 2021 Bulletin of Electrical Engineering and Informatics has been ACCEPTED with minor revisions. The Bulletin of Electrical Engineering and Informatics (BEEI), ISSN: 2089-3191, e-ISSN: 2302-9285, <http://beeI.org> is open to submission from scholars and experts in the wide areas of electrical, electronics, instrumentation, control, robotics, telecommunication, computer engineering, computer science, information system, information technology and informatics from the global world. This international journal is ACCREDITED (recognised) by the Ministry of Research, Technology and Higher Education, Republic of Indonesia (RISTEKDIKTI, Decree No: 60/E/KPT/2016) and indexed by Scopus (Elsevier)/ScimagoJR (SJR), SNIP 2019: 1.358 (SNIP=Source Normalized Impact per Paper measures actual citations received relative to citations expected for the serial's subject field), CiteScore 2019: 1.3, SJR 2019: 0.231; and ACCREDITED (recognised) by the Ministry of Research and Technology/National Agency for Research and Innovation, Republic of Indonesia (Decree No: 60/E/KPT/2016). Beginning with issue 1 of volume 9 (2020), BEEI is published as a bimonthly journal (6 issues/year).

Please make the necessary changes based on reviewers' comments and suggestions. The reviews are below or can be found at <https://edas.info/showPaper.php?m=1570662772>. Please prepare your final camera ready paper (in MS Word file format) adheres every detail of the guide of authors (<http://iaescore.com/gfa/beeI.docx>), and check it for spelling/grammatical mistakes. The goal of this camera ready paper is to describe NOVEL TECHNICAL RESULTS.

>>> For original research paper, there are four (4) types of novel technical results: 1) An algorithm; 2) A system construct: such as hardware design, software system, protocol, etc.; The main goal of your revised paper is to ensure that the next person who designs a system like yours doesn't make the same mistakes and takes advantage of some of your best solutions. So make sure that the hard problems (and their solutions) are discussed and the non-obvious mistakes (and how to avoid them) are discussed; 3) A performance evaluation: obtained through analyses, simulation or measurements; or 4) A theory: consisting of a collection of theorems. Your final camera ready paper should focus on: 1) Describing the results in sufficient details to establish their validity; 2) Identifying the novel aspects of the results, i.e., what new knowledge is reported and what makes it non-obvious; and 3) Identifying the significance of the results: what improvements and impact do they suggest. Number of minimum references for original research paper is 25 references (and minimum 20 recently journal articles).

>>> For review paper, the paper should present a critical, constructive analysis of the literature in a specific field through summary, classification, analysis and comparison. The function and goal of the review paper is: 1) to organize literature; 2) to evaluate literature; 3) to identify patterns and trends in the literature; 4) to synthesize literature; or 5) to identify research gaps and recommend new research areas. The structure includes:

1. Title – in this case does not indicate that it is a review article.
2. Abstract – includes a description of subjects covered.
3. Introduction includes a description of context (paragraph 1 – 3), motivation for review (paragraph 4, sentence 1) and defines the focus (paragraph 4, sentences 2 – 3)
4. Body – structured by headings and subheadings
5. Conclusion – states the implications of the findings and an identifies possible new research fields
6. References ("Literature Review") – organised by number in the order they were cited in the text.

Number of minimum references for review paper is 50 references (and minimum 40 recently journal articles).

For your information, according to international regulations, similarity score of camera-ready paper should be less than 25%. Single author is NOT allowed. The Editor(s) will check whether the final version has been performed and already address the reviewers' comments or not. Failing to do proper revision may lead to the rejection of your paper.

Authors MUST upload their Final manuscript through EDAS online system for similarity checking by EDAS. You also must submit your

final camera ready paper, similarity checking report (by iThenticate or Turnitin software), along with your payment receipt to: [tole@iaescore.com](mailto:tole@iaescore.com) within 5 weeks.

I look forward for your response

Sincerely yours,  
Assoc. Prof. Dr. Tole Sutikno  
Editor-in-Chief,  
Bulletin of Electrical Engineering and Informatics  
website: <http://beei.org>  
email: [tole@iaescore.com](mailto:tole@iaescore.com), [beei.org@gmail.com](mailto:beei.org@gmail.com)

-----  
Submit paper: <https://edas.info/N27818>

-----  
Below is the reviews on your papers:

=====  
Review 1 =====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?

Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?

Accept (9)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The paper is written very well and the stream of the paper is clear

> \*\*\* Recommendation: Your overall rating.

Accept (9)

=====  
Review 2 =====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?

Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?

Accept (9)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The authors present well written technical paper with relevant topic, proper research methodology and good contribution to the field of studies.

> \*\*\* Recommendation: Your overall rating.

Accept (9)

=====  
Review 3 =====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?

Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?

Accept (9)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

This work proposes a methodology for overcoming the problem of class imbalance, overlapping, and noise in the multi-class classification.

The manuscript is quite well written, the theory seems correct, and the reported results and findings are good and interesting. Further comparisons with similar SoA techniques and/or an application to a real world example could definitively improve this paper

strenght.

> \*\*\* Recommendation: Your overall rating.  
Accept (9)

=====  
Review 4  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Borderline Accept (5)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

- Please revise the academic writing style and grammars throughout, for examples (but not limited to) the first 3 sentences in the abstract, some broken sentences, casual phrases, etc.

- Please make the abstract more concise, containing only the key contributions and findings (leaving out detailed discussion to the main text).

- Please restructure the motivation parts, regarding the imbalance and overlapping and make them more focus. For example (but not limited to), several statements are dedicated to problems caused by overlapping, but none of them is more specific than others. This pattern repeats for most parts of the introduction. Please revise. Moreover, feature selection algorithms are mentioned but unfortunately not so specific nor much discussed. Please elaborate.

- Majority of the section 2 are performance metrics and not the related works but more the material components in the proposed method. Please consider.

- Please use more appropriate font/ format for pseudocode presented in section 2.

- Please add structural outline of how components presented in section 2 would be used in section 4.

- Similarly, please refer each block in Figure to respective topic ID in section 2.

- The results discussion are well written. However, please add in section 4.4., comparative discussion with similar works in the areas (impalance, noise handling).

- Please compare the proposed technique with that without MOSS and MC-CCR processing.

> \*\*\* Recommendation: Your overall rating.  
Weak Accept (7)

=====  
Review 5  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Strong Accept (Very good) (10)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Strong Accept (Excellent) (10)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

Very strong technical analysis, lots of details. Very good results.

> \*\*\* Recommendation: Your overall rating.  
Strong Accept (10)

=====  
Review 6  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Borderline Accept (5)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Borderline Accept (5)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

1. Accuracy and Reliable issue.

2. How did the author validate the results?
3. Please perform an experiment related to the work and compare the results with simulation or calculated results.
4. What are the system efficiency?

> \*\*\* Recommendation: Your overall rating.  
Borderline Accept (5)

=====  
Review 7  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Borderline Accept (5)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Borderline Accept (5)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

Needs revision and language check

> \*\*\* Recommendation: Your overall rating.  
Borderline Accept (5)

=====  
Review 8  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Borderline Accept (5)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Borderline Accept (5)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

A more comprehensive literature survey can be included in the paper. The proposed methodology in this work and its performance can be compared in detail with some of the previous research to highlight the novelty and impact of the work.

A more comprehensive analysis of the results, inferences and findings can be included and specific examples from the dataset can be highlighted in the paper. The advantages, limitations and specific applications of the proposed approach should be clearly stated in the paper.

The paper should be proofread in a thorough manner and revised to correct grammatical errors.

> \*\*\* Recommendation: Your overall rating.  
Borderline Accept (5)

=====  
Review 9  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Accept (9)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

This manuscript proposes a method which can improve the performance of Hybrid Approach Redefinition Multi Class (HAR-MI) with resampling algorithm and feature selection. Generally this paper has enough contribution. To improve the quality of this paper please conduct point-point of revision below:

1. Please re-write Algorithm for MOSS. Make sure the content is not written exceed outside the margin of paper.
2. Fourth Paragraph in Section 4.2 : "Based on the test results in Table 4" . I think it should refer to Table 5.
3. How much undersampling ratio was applied in that experiment? Are there 1:1. Please state in the manuscript.

4. For noise identification? What method was used in this manuscript? Because MC-CCR was claimed as method to overcome noise, please give additional description about the number of noise for each data set in manuscript. For overlap I things the R-Value has successfully describe the overlap number in those data set.

5. Please state what is accuracy which was been obtained for proposed method in Abstract. Please give quantitative number in abstract.

> \*\*\* Recommendation: Your overall rating.  
Accept (9)

=====  
Revision review 10  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Borderline Accept (5)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Borderline Accept (5)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The paper presentation is sufficient. Enough theoretical background to support the proposed method.

The authors should state the contribution and novelty of the proposed method compared to the current researches.

The discussions should be more elaborated to support the effectiveness of the proposed method.

> \*\*\* Recommendation: Your overall rating.  
Borderline Accept (5)

=====  
Revision review 11  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Reject (1)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Reject (1)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The paper deserves a careful revision. Acronyms are not defined, terms as well. Section 1 does not clearly introduce the problem, it is much more a "related works" than Section 2, while Section 2 reviews several concepts of classifiers and lists some algorithms that are used by the authors. That is, the paper presentation does not allow this reviewer to make an assertive opinion regarding the authors' contribution.

> \*\*\* Recommendation: Your overall rating.  
Reject (1)

=====  
Revision review 12  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Weak Accept (7)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Accept (9)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

\* There are some grammatical and punctuation errors. Please recheck the written language of the paper!

\* Please recheck Eq.(3) on page 3! What are the starting and ending values for index 'i'?

\* Similarly, please recheck Eq.(6) - Eq.(8)!

\* Please use consistent Mathematical notations, for example, non-capital p is different from capital p!

\* In the last paragraph of sub-section 2.4, the referred equations should be Eq.(9) and Eq.(10)!

\* Why don't the authors compare the result of the original HAR-MI method with the combination of HAR-MI and Resampling + Feature Selection method?

> \*\*\* Recommendation: Your overall rating.  
Accept (9)

=====  
Revision review 13  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Weak Accept (7)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Weak Accept (7)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The main contribution of this work is to produce a new method that can overcome the problem of class imbalance, overlapping, and noise in the multi-class. The talked issue is interesting, but the writing should be improved.

> \*\*\* Recommendation: Your overall rating.  
Weak Accept (7)

=====  
Revision review 14  
=====

> \*\*\* Novelty and Contribution: Rate the degree of scientific contribution provided by this paper. Do the authors offer new findings? Do they give proper explanation and detailed analysis?  
Accept (9)

> \*\*\* Paper Presentation: What is your evaluation on the quality of presentation from this paper (e.g. figures, tables, formats, etc.)?  
Weak Accept (7)

> \*\*\* Detailed Comments: Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.

The paper needs improvement in the body of knowledge, English language grammar, and references.

> \*\*\* Recommendation: Your overall rating.  
Weak Accept (7)

Please pay an attention to double check your final camera ready paper!!  
-----

(1) TEMPLATE, Please Strictly use and follow to the template Manuscripts (Word Format): <http://iaescore.com/gfa/bee1.docx> or <http://iaescore.com/gfa/bee1.rar> (LaTeX format). If you use LaTeX, submit your LaTeX source files to [tole@iaescore.com](mailto:tole@iaescore.com)

For original research paper:

(2) Authors are suggested to present their articles with IMRaD sections structure (outline): Introduction - The Proposed Method/Algorithm/Procedure specifically designed (optional) - Research Method - Results and Discussion – Conclusion. Authors may present complex proofs of theorems or non-obvious proofs of correctness of algorithms after introduction section (obvious theorems & straightforward proofs of existing theorems are NOT needed).

(3) Introduction section: explain the context of the study and state the precise objective. An Introduction should contain the following three (3) elements:

- Background: Authors have to make clear what the context is. Ideally, authors should give an idea of the state-of-the art of the field the report is about.
- The Problem: If there was no problem, there would be no reason for writing a manuscript, and definitely no reason for reading it. So, please tell readers why they should proceed reading. Experience shows that for this part a few lines are often sufficient.
- The Proposed Solution: Now and only now! - authors may outline the contribution of the manuscript. Here authors have to make sure readers point out what are the novel aspects of authors work. Authors should place the paper in proper context by citing relevant papers. At least, 5 references (recently journal articles) are cited to support this section.

(4) Method section: the presentation of the experimental methods should be clear and complete in every detail facilitating reproducibility by other scientists.

(5) Results and discussion section: The presentation of results should be simple and straightforward in style. This section report the most important findings, including results of statistical analyses as appropriate and comparisons to other research results. This is where the author(s) should explain in words what he/she/they discovered in the research.

(6) (URGENT)!!! About Figures & Tables in your manuscript:

- Because tables and figures supplement the text, all tables and figures should be REFERENCED in the text. Authors MUST EXPLAIN what the reader should look for when using the table or figure. Focus only on the important point the reader should draw from them, and leave the details for the reader to examine on her own.
- Tables are to be presented with single horizontal line under: the table caption, the column headings and at the end of the table. All tables are produced by creating tables in MS Word. Captured tables are NOT allowed.
- All figures MUST in high quality images

(7) Conclusion section: Summarize sentences the primary outcomes of the study in a paragraph. Are the claims in this section supported by the results, do they seem reasonable? Have the authors indicated how the results relate to expectations and to earlier research? Does the article support or contradict previous theories? Does the conclusion explain how the research has moved the body of scientific knowledge forward?

(8) Each citation should be written in the order of appearance in the text in square brackets. For example, the first citation [1], the second citation [2], and the third and fourth citations [3,4]. When citing multiple sources at once, the preferred method is to list each number separately, in its own brackets, using a comma or dash between numbers, as such: [1], [3], [5] or [4-8]. It is not necessary to mention an author's name, pages used, or date of publication in the in-text citation. Instead, refer to the source with a number in a square bracket, e.g. [9], that will then correspond to the full citation in your reference list. Examples of in-text citations:

This theory was first put forward in 1970 [9]."

Zadeh [10] has argued that...

Several recent studies [7], [9], [11-15] have suggested that....

...end of the line for my research [16].

(9) Please ensure the maximum page of your final paper is 8-page, but still allowed up to 12 pages (required to pay an extra fee).

---

In order to cover part of the event cost, each accepted paper will be charged: 265.00 (USD)

This article publication is to support the cost of wide open access dissemination of research results, to manage the various costs associated with handling and editing of the submitted manuscripts, and the Journal management and publication in general, the authors or the author's institution is requested to pay a publication fee for each article accepted. The 265 USD fee covers the standard eight (8) pages manuscript. For every additional page an extra fee of 50 USD per page will be charged.

The payment should be made by bank transfer (T/T):

---

Bank Account name (please be exact)/Beneficiary: TOLE SUTIKNO

Bank Name: Bank Negara Indonesia (BNI)

Branch Office: UGM Yogya, KCP SARJANA WIYATA

City: Yogyakarta

Country: Indonesia

Bank Account # : 0469076067

SWIFT Code: BNINIDJAXXX

or as an alternative you can pay it by using PayPal to email: [tole@ee.uad.ac.id](mailto:tole@ee.uad.ac.id)

---

Bank's detailed address:

BNI SARJANA WIYATA

Jl. Kusumanegara No.121, Muja Muju, Umbulharjo

City: Yogyakarta

Province: D.I. Yogyakarta (DIY)

Post Code: 55167

Country :Indonesia

The Beneficiary's address:

---

D2, Griya Ngoto Asri, Bangunharjo, Sewon

City: Bantul  
Province: D.I. Yogyakarta (DIY)  
Post Code: 55187  
Country: Indonesia

---

**Hartono Ibbi** <hartonoibbi@gmail.com>  
Kepada: tole=iaescore.com@edas.info

18 Maret 2021 pukul 14.55

Dear Assoc. Prof. Dr. Tole Sutikno,


Well received and we will complete all the requirements needed As Soon As Possible. Thank you for the opportunity given to us and if there is anything that needs to be completed, do not hesitate to let us know.

Best Regards

Hartono




[Kutipan teks disembunyikan]

# Final Manuscript



http://iaes.org | ISSN 2089-3191 e-ISSN 2502-5195

Bulletin of Electrical Engineering and Informatics




Q3

0.251

0.954

1.7



Indonesian Journal of Electrical Engineering and Informatics

Home Papers Register Travel grants My... Help

My... » My papers » #147 (1570662772): HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise

## #147 (1570662772): HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise

[Hide details](#)

**Authors** [Erianto Ongko](#) (Akademi Teknologi Industri Immanuel, Indonesia); [Hartono Hartono](#) (Universitas Sumatera Utara, Indonesia)

**Paper title** HAR-MI with Resampling and Feature Selection for Multi-Class Imbalance with Overlapping and Noise Only the chairs can edit

**Journal and track** 2021 Bulletin of Electrical Engineering and Informatics - Soft Computing and Intelligent System


**Abstract** Only the chairs can edit Handling Class Imbalance on multi-class problems is the crucial problems in data classification....

**Keywords** Class Imbalance; Multi-Class; Overlapping; Noise; Classifier Performance Only the chairs can edit

**Topics** (Only the chairs can edit)

**Personal notes** [+](#)

**Roles** You are the creator and an author for this paper.  
You have authored an accepted paper in this journal.

**Status** Final Camera Ready 

**Review manuscript** **Revision** **Final manuscript**

**Final manuscript**

Created	Apr 19, 2021
Checked	Apr 19, 2021