Bukti Korespondensi Author

A. PAPER 2:	**CORONARY HEART DISEASE USING SUPPORT VECTOR MACHINE** **Penulis: OKFALISA, LESTARI HANDAYANI, DINDA JUWITA P, MUHAMMAD AFFANDES, S.S.M. FAUZI3, SAKTIOTO. (Penulis 1 dan Corresponding Author), Journal of Engineering Science and Technology, April 2021, Vol.16, Issue 2. Scopus (Q2) SJR: 0.24**

Paper ini diterima oleh pihak jurnal pada tanggal 3 April 2020, Memperoleh review 1 pada 26 Juli 2020, Review ke 2 di 22 Agustus 2020, mendapatkan full accepted pada 17 Oktober 2020, dan full published pada April 2021.

Bukti Korespondensi dapat dilihat pada Gambar berikut dan lengkapnya dapat dilihat pada lampiran:

1. Submission diterima oleh pihak jurnal, tanggal 3 April 2020

6/26/2021 Gmail - Submission of a Manuscript (EE20099) / First Round of the Review Process

Gmail okfalisa saktioto <okfalisa@gmail.com>

Submission of a Manuscript (EE20099) / First Round of the Review Process
3 messages

Jestec <Jestec@taylors.edu.my> Fri, Apr 3, 2020 at 7:50 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

Thank you for submitting your research paper to the Journal of Engineering Science and Technology (JESTEC)

Kindly note that we have received the paper entitled

IDENTIFY THE CLASSIFICATION OF DATASET CORONARY HEART DISEASE: SUPPORT VECTOR MACHINE (SVM) EMPLOYMENT

Your paper ID is **EE20099** (*Please quote the above manuscript ID in all future correspondence with us.*)

Soon we will initiate the first round of the review process.

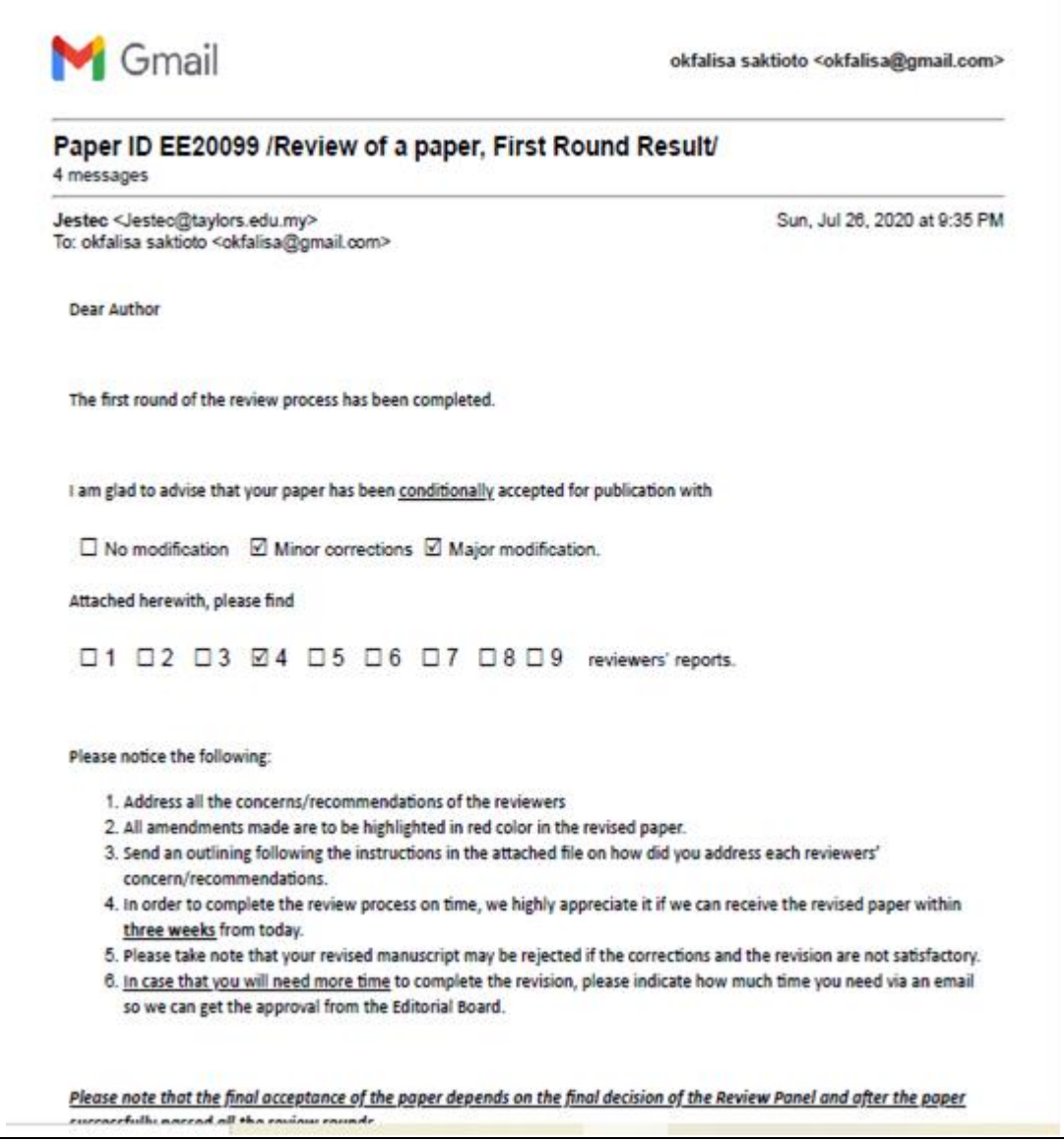
Please be reminded that upon the full acceptance of your paper, publication fee in amount of USD300 must be paid before the article is published in the journal website.

Best regards

JESTEC Editor

http://jestec.taylors.edu.my

2. Memperoleh revisi 1 dari reviewer, 26 Juli 2020



Gmail okfalisa saktioto <okfalisa@gmail.com>

Paper ID EE20099 /Review of a paper, First Round Result/
4 messages

Jestec <Jestec@taylors.edu.my> Sun, Jul 26, 2020 at 9:35 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

The first round of the review process has been completed.

I am glad to advise that your paper has been conditionally accepted for publication with

☐ No modification ☑ Minor corrections ☑ Major modification.

Attached herewith, please find

☐1 ☐2 ☐3 ☑4 ☐5 ☐6 ☐7 ☐8 ☐9 reviewers' reports.

Please notice the following:

1. Address all the concerns/recommendations of the reviewers
2. All amendments made are to be highlighted in red color in the revised paper.
3. Send an outlining following the instructions in the attached file on how did you address each reviewers' concern/recommendations.
4. In order to complete the review process on time, we highly appreciate it if we can receive the revised paper within three weeks from today.
5. Please take note that your revised manuscript may be rejected if the corrections and the revision are not satisfactory.
6. In case that you will need more time to complete the revision, please indicate how much time you need via an email so we can get the approval from the Editorial Board.

Please note that the final acceptance of the paper depends on the final decision of the Review Panel and after the paper successfully passed all the review rounds

so we can get the approval from the Editorial Board.

Please note that the final acceptance of the paper depends on the final decision of the Review Panel and after the paper successfully passed all the review rounds.

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

6 attachments

outlining of Review Report_v3.docx
75K

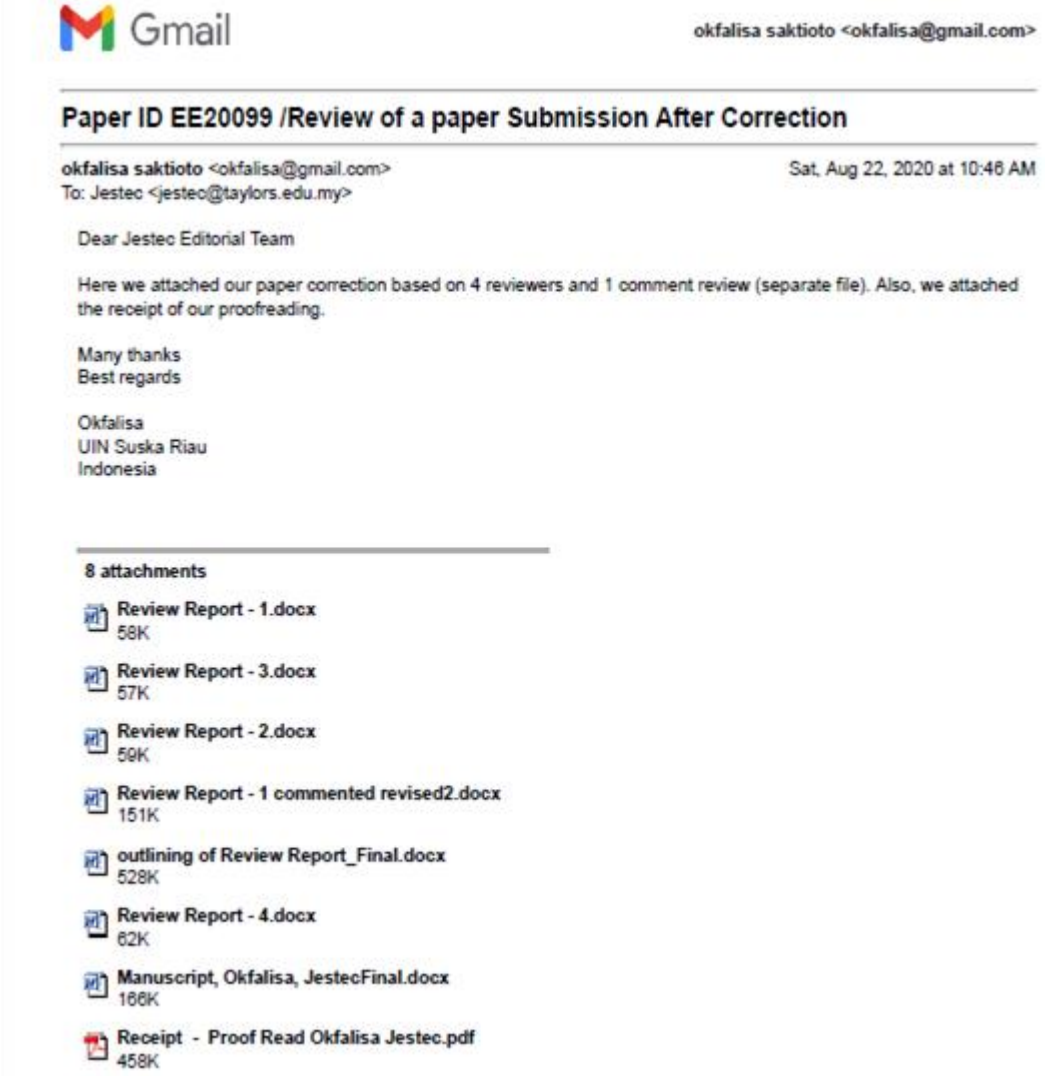
Review Report - 1 commented.docx
208K

Review Report - 1.docx
47K

Review Report - 2.docx
40K

Review Report - 3.docx
48K

3. Menjawab revisi reviewer pada tanggal 22 Agustus 2020



Gmail okfalisa saktioto <okfalisa@gmail.com>

Paper ID EE20099 /Review of a paper Submission After Correction

okfalisa saktioto <okfalisa@gmail.com> Sat, Aug 22, 2020 at 10:46 AM
To: Jestec <jestec@taylors.edu.my>

Dear Jestec Editorial Team

Here we attached our paper correction based on 4 reviewers and 1 comment review (separate file). Also, we attached the receipt of our proofreading.

Many thanks
Best regards

Okfalisa
UIN Suska Riau
Indonesia

8 attachments

Review Report - 1.docx 58K
Review Report - 3.docx 57K
Review Report - 2.docx 59K
Review Report - 1 commented revised2.docx 151K
outlining of Review Report_Final.docx 528K
Review Report - 4.docx 62K
Manuscript, Okfalisa, JestecFinal.docx 166K
Receipt - Proof Read Okfalisa Jestec.pdf 458K

4. Contoh hasil correction sesuai dengan commentar reviewer dapat dilihat pada Gambar dibawah. Lengkapnya dapat dilihat pada Link: https://drive.google.com/file/d/1VaR7RMB5rHtGW5rZxU4sHzlct8Gq5reC/view Dan di **Lampiran.**

OUTLINING HOW THE ISSUES ARE ADDRESSED

Title of paper:

1. Address all the concerns/recommendations of the reviewers.
2. All amendments made are to be highlighted in red color in the revised paper.

Reviewer # 1

Final Recommendation Please tick	Accepted without modification	Accepted with minor corrections	Accepted with major modification	Rejected
	☐	☐	☐	☐

Comments	Addressed (Y/N)	Reply/Action taken
• Some tables can be combined in one table		Done (Combine Table 2-9) into Table 4

Table 4. Attribute Discretization

Age discretization (1)		Systolic TD discretization (Sis) (9)	
Age (years)	Discretization	Systolic BP (mmHg)	Discretization
25 < U <35	0	Sis<120	Optimal (0)
35 ≤ U <45	0.2	120< Sis <130	Normal (0.2)
45 ≤ U <55	0.4	130< Sis <140	Normal Height (0.4)
55 ≤ U <65	0.6	140< Sis <150	Low hypertension (0.6)
65 ≤ U <75	0.8	150< Sis <160	Moderate hypertension (0.8)
U ≥ 85	1	Sis >160	Severe hypertension (1)
Diastolic TD (Dias) discretization (10)		Discretization of LDL (LDL) levels (11)	
Diastolic BP (mmHg)	Discretization	LDL levels (mg / dL)	Discretization
Dias<80	Optimal (0)	LDL<100	Optimal (0)
80≤Dias<85	Normal (0.2)	100< LDL <130	Approaching optimal (0.25)
85≤Dias<90	Normal Height (0.4)	130< LDL <160	Borderline high (0.5)

			145≤ Glu <200 Moderate hypertension (0.8) Glu ≥200 Severe hypertension (1)
• Some tables can be converted to a figures to be more clear such as table 11, 12, and 13		Table 11, 12, and 13 has been converted into Table 6,7, and 8 respectively. Figure 3,4, and 5 explained in more details regarding on the performance. It explained in the text as follows. To investigate the implication of pre-processing against SVM, the analysis is conducted by comparing the accuracy within dataset changes in the original data (without missing values), the reduced (with missing values), k-NN (with distance calculation), and pre-processing (KDD formatted). This was executed through the selection of the best parameters for 10-fold cross-validation in Table 6 and percentage split in Table 7 for four scenarios dataset. The graphical views of performances are shown in Figure 3, 4, and 5.	

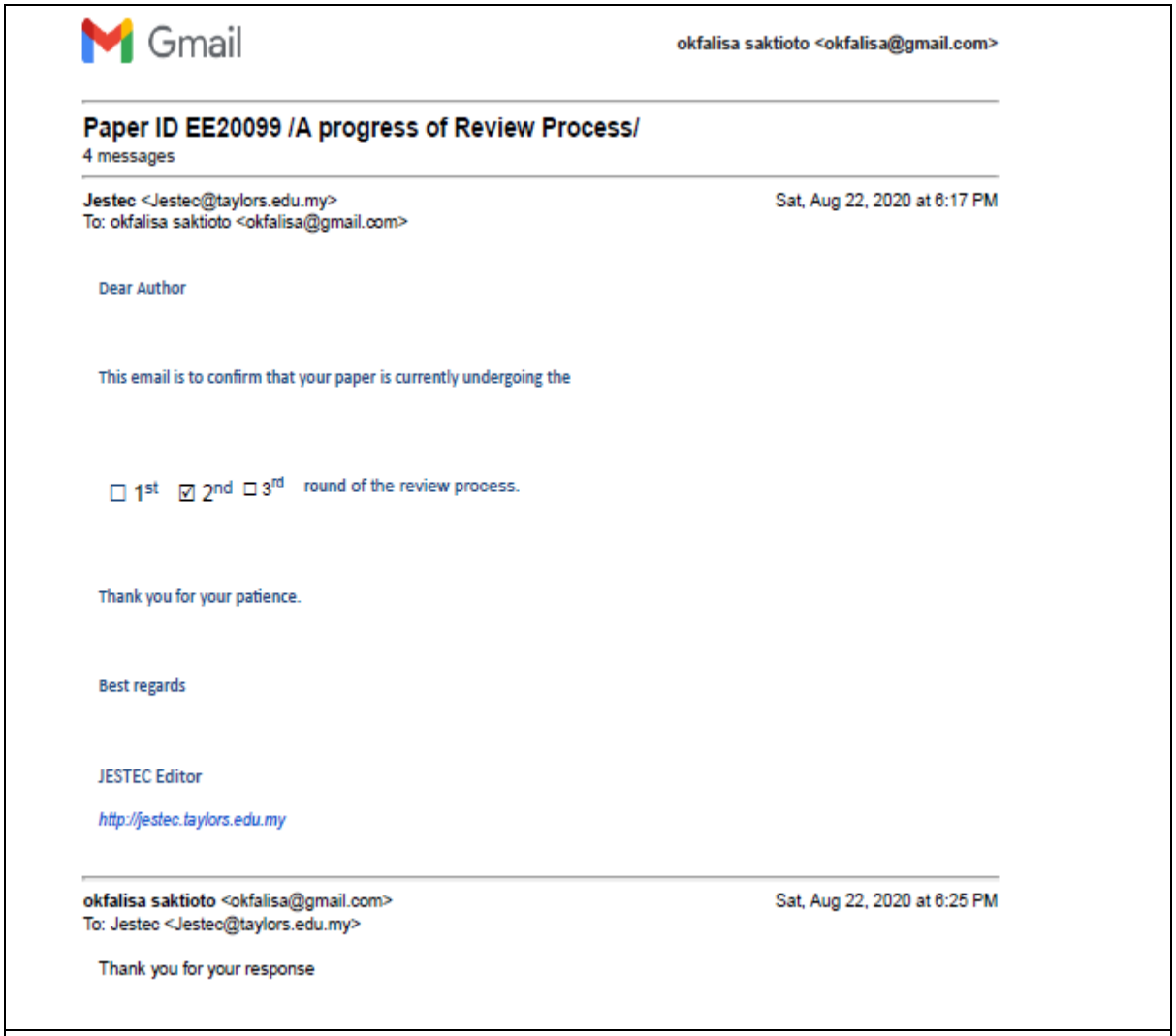
Table 6. The accuracy of the best parameter - 10-fold cross validation.

Kernel	Polynomial			RBF		
Parameters	C	d/σ	Accuracy	C	d/σ	Accuracy
Original Dataset	0.03	1	100%	0.01	1	47.9%
Reduced Dataset	0.03	1	100%	0.01	1	48.1%
k-NN Dataset	0.03	1	100%	0.01	1	47.9%
Pre-processing Dataset	0.02	2	100%	0. 8	1	98.9%

Table 7. The accuracy of the best parameter pairs -percentage split.

Kernel	Polynomial			RBF		
Parameters	DC	T (s)	Accuracy	DC	T(s)	Accuracy
Original Dataset	70:30	27.49	100%	40:60	0.06	49.4%
Reduced Dataset	70:30	21.37	100%	70:30	0.13	53.8%
k-NN Dataset	80:20	24.55	100%	40:60	0.08	49.4%
Pre-processing Dataset	70:30	0.06	100%	80:20	0.08	100%

5. Proses review ke 2 dari jurnal diperoleh tanggal 22 Agustus 2020



Gmail okfalisa saktioto <okfalisa@gmail.com>

Paper ID EE20099 /A progress of Review Process/

4 messages

Jestec <Jestec@taylors.edu.my> Sat, Aug 22, 2020 at 6:17 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

This email is to confirm that your paper is currently undergoing the

☐ 1st ☑ 2nd ☐ 3rd round of the review process.

Thank you for your patience.

Best regards

JESTEC Editor

http://jestec.taylors.edu.my

okfalisa saktioto <okfalisa@gmail.com> Sat, Aug 22, 2020 at 6:25 PM
To: Jestec <Jestec@taylors.edu.my>

Thank you for your response

6. Menerima full accepted dari jurnal 11 Oktober 2020 dan Full Published di April 2021.

Gmail okfalisa saktioto <okfalisa@gmail.com>

Paper ID (EE20099) Review process is completed

2 messages

Jestec <Jestec@taylors.edu.my> Sun, Oct 11, 2020 at 10:08 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

I am glad to advise that your paper has been accepted for publication without modification. The reviewers have no more comments and they are satisfied with the revised paper.

By this the review process is completed and we kindly ask you to check the format of the paper according to the instructions for authors and JESTEC template (attached).

Special attention to be paid for list of symbols used and the references. Please follow strictly the instructions for citation of the references (attached are instructions) and explain each symbol you used and its SI units. Also refer to this link: http://jestec.taylors.edu.my/instructions.html

You are also kindly required to fill in the JESTEC-Copyright transfer form (use this link to download

http://jestec.taylors.edu.my/Copyright%20transfer%20ver%20190818.doc **and send to the journal.**

Kindly note that you have only four weeks to submit the above.

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

Gmail okfalisa saktioto <okfalisa@gmail.com>

Review process is completed paper (EE20099) /formatting, proofreading, payment/

4 messages

Jestec <Jestec@taylors.edu.my> Sat, Oct 17, 2020 at 10:42 AM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author (s)

Thank you for your email and sending your modified paper. We found that the paper still contains some formatting mistakes.

We would like to inform you that your paper has been scheduled to be published in April 2021, Volume 16 Issue 2

Attached please find the acceptance letter.

Please send us up-to-date copyright transfer form. Download from here JESTEC-Copyright transfer form (CRTF)

Payment of the publication is needed before the paper is published online.

Kindly refer to the attached sample of the invoice and amend it (Red text only) according to your up-to-date and accurate information for the purpose of the payment. Once submitted we will send you an official invoice with all details to make safe payment.

We thank you very much for your interest in JESTEC and looking forward for new contribution.

Best regards

Assoc. Prof. Dr. Abdulkareem Sh. Mahdi Al-Obaidi, CEng MIMechE

Executive Editor, Journal of Engineering Science & Technology

http://jestec.taylors.edu.my

2 attachments

177 LoA_16_2_21 OKFALISA et al.pdf
56K

20_177.docx
22K

KESIMPULAN:

Paper 2 telah melampirkan bukti korespondensi pengusul dengan pihak editor jurnal.

LAMPIRAN 2

BUKTI KORESPONDING AUTHOR

PAPER 2. CORONARY HEART DISEASE USING SUPPORT VECTOR MACHINE

Penulis: OKFALISA, LESTARI HANDAYANI, DINDA JUWITA P, MUHAMMAD AFFANDES, S.S.M. FAUZI3, SAKTIOTO. (Penulis 1 dan Corresponding Author), Journal of Engineering Science and Technology, April 2021, Vol.16, Issue 2. Scopus (Q2) SJR: 0.24

Gmail okfalisa saktioto <okfalisa@gmail.com>

Submission of a Manuscript (EE20099) / First Round of the Review Process

3 messages

Jestec <Jestec@taylors.edu.my> Fri, Apr 3, 2020 at 7:50 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

Thank you for submitting your research paper to the Journal of Engineering Science and Technology (JESTEC)

Kindly note that we have received the paper entitled

IDENTIFY THE CLASSIFICATION OF DATASET CORONARY HEART DISEASE: SUPPORT VECTOR MACHINE (SVM) EMPLOYMENT

Your paper ID is **EE20099** (*Please quote the above manuscript ID in all future correspondence with us.)*

Soon we will initiate the first round of the review process.

Please be reminded that upon the full acceptance of your paper, publication fee in amount of USD300 must be paid before the article is published in the journal website.

Best regards

JESTEC Editor

http://jestec.taylors.edu.my

okfalisa saktioto <okfalisa@gmail.com> Fri, Apr 3, 2020 at 8:44 PM
To: Jestec <Jestec@taylors.edu.my>

Dear Editor,

Thank you very much for your information. Let me know when the paper will be published? I will make a payment soon after the paper is fully accepted.
Please advise how the payment is made.

Best regards

Okfalisa

[Quoted text hidden]

Jestec <Jestec@taylors.edu.my> Fri, Apr 3, 2020 at 8:45 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Payment is due once the paper is accepted for publication

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

[Quoted text hidden]
[Quoted text hidden]

Gmail

okfalisa saktioto <okfalisa@gmail.com>

Review Status of a paper (EE20099)

1 message

Jestec <Jestec@taylors.edu.my> Sun, Jun 7, 2020 at 8:29 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

The review of your paper has been not completed yet. Up to this moment we do not have adequate numbers of review reports to share.

Thank you for your patience.

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

Gmail okfalisa saktioto <okfalisa@gmail.com>

Paper ID EE20099 /Review of a paper, First Round Result/

4 messages

Jestec <Jestec@taylors.edu.my> Sun, Jul 26, 2020 at 9:35 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

The first round of the review process has been completed.

I am glad to advise that your paper has been conditionally accepted for publication with

☐ No modification ☑ Minor corrections ☑ Major modification.

Attached herewith, please find

☐ 1 ☐ 2 ☐ 3 ☑ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 reviewers' reports.

Please notice the following:

1. Address all the concerns/recommendations of the reviewers
2. All amendments made are to be highlighted in red color in the revised paper.
3. Send an outlining following the instructions in the attached file on how did you address each reviewers' concern/recommendations.
4. In order to complete the review process on time, we highly appreciate it if we can receive the revised paper within **three weeks** from today.
5. Please take note that your revised manuscript may be rejected if the corrections and the revision are not satisfactory.
6. In case that you will need more time to complete the revision, please indicate how much time you need via an email so we can get the approval from the Editorial Board.

Please note that the final acceptance of the paper depends on the final decision of the Review Panel and after the paper successfully passed all the review rounds.

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

6 attachments

Gmail okfalisa saktioto <okfalisa@gmail.com>

Paper ID EE20099 /Review of a paper, First Round Result/

Jestec <Jestec@taylors.edu.my> Sun, Jul 26, 2020 at 9:35 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

The first round of the review process has been completed.

I am glad to advise that your paper has been conditionally accepted for publication with

☐ No modification ☑ Minor corrections ☑ Major modification.

Attached herewith, please find

☐ 1 ☐ 2 ☐ 3 ☑ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 reviewers' reports.

Please notice the following:

1. Address all the concerns/recommendations of the reviewers
2. All amendments made are to be highlighted in red color in the revised paper.
3. Send an outlining following the instructions in the attached file on how did you address each reviewers' concern/recommendations.
4. In order to complete the review process on time, we highly appreciate it if we can receive the revised paper within **three weeks** from today.
5. Please take note that your revised manuscript may be rejected if the corrections and the revision are not satisfactory.
6. In case that you will need more time to complete the revision, please indicate how much time you need via an email so we can get the approval from the Editorial Board.

Please note that the final acceptance of the paper depends on the final decision of the Review Panel and after the paper successfully passed all the review rounds.

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

6 attachments

outlining of Review Report_v3.docx
75K

Review Report - 1 commented.docx
208K

Review Report - 1.docx
47K

Review Report - 2.docx
40K

Review Report - 3.docx
48K

Review Report - 4.docx
41K

Gmail

okfalisa saktioto <okfalisa@gmail.com>

Paper ID EE20099 /Review of a paper, First Round Result/

okfalisa saktioto <okfalisa@gmail.com> Mon, Jul 27, 2020 at 5:48 AM
To: Jestec <Jestec@taylors.edu.my>

Ok thank you.

I'll revise as requirements.

Thanks & Regards
[Quoted text hidden]

Gmail okfalisa saktioto <okfalisa@gmail.com>

SUBMISSION OUR PAPER: "IDENTIFY THE CLASSIFICATION OF DATASET CORONARY HEART DISEASE: SUPPORT VECTOR MACHINE (SVM) EMPLOYMENT " AUTHOR: OKFALISA

1 message

okfalisa saktioto <okfalisa@gmail.com> Fri, Mar 27, 2020 at 8:42 PM
To: jestec@taylors.edu.my, toto saktioto <saktioto@yahoo.com>, lestari handayani <lestari.handayani@uin-suska.ac.id>

Dear Jestec Editorial Team,

Please find attached our article and supported documents to your journal.
Thank you,

Regards,

Dr.Okfalisa
Faculty Science and Technology
Universitas Islam Negeri Sultan Syarif Kasim Riau
Riau
Indonesia

5 attachments

- **Manuscript, Okfalisa, Jestec ready submit.docx** 183K
- **Copyright Transfer Okfalisa.pdf** 295K
- **CV Okfalisa 2020.pdf** 219K
- **PPR_ Okfalisa.xlsx** 21K
- **Similaity Index, Jestec.pdf** 2142K

Gmail **okfalisa saktioto <okfalisa@gmail.com>**

Paper ID EE20099 /Review of a paper Submission After Correction

okfalisa saktioto <okfalisa@gmail.com> Sat, Aug 22, 2020 at 10:46 AM
To: Jestec <jestec@taylors.edu.my>

Dear Jestec Editorial Team

Here we attached our paper correction based on 4 reviewers and 1 comment review (separate file). Also, we attached the receipt of our proofreading.

Many thanks
Best regards

Okfalisa
UIN Suska Riau
Indonesia

8 attachments

- **Review Report - 1.docx** 58K
- **Review Report - 3.docx** 57K
- **Review Report - 2.docx** 59K
- **Review Report - 1 commented revised2.docx** 151K
- **outlining of Review Report_Final.docx** 528K
- **Review Report - 4.docx** 62K
- **Manuscript, Okfalisa, JestecFinal.docx** 166K
- **Receipt - Proof Read Okfalisa Jestec.pdf** 458K

Gmail okfalisa saktioto <okfalisa@gmail.com>

Paper ID EE20099 /A progress of Review Process/

4 messages

Jestec <Jestec@taylors.edu.my> Sat, Aug 22, 2020 at 6:17 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

This email is to confirm that your paper is currently undergoing the

☐ 1st ☑ 2nd ☐ 3rd round of the review process.

Thank you for your patience.

Best regards

JESTEC Editor

http://jestec.taylors.edu.my

---

okfalisa saktioto <okfalisa@gmail.com> Sat, Aug 22, 2020 at 6:25 PM
To: Jestec <Jestec@taylors.edu.my>

Thank you for your response

[Quoted text hidden]

---

okfalisa saktioto <okfalisa@gmail.com> Sun, Aug 23, 2020 at 7:31 AM
To: Jestec <Jestec@taylors.edu.my>

Dear editor

Thank you for your response

On Sat, Aug 22, 2020, 6:17 PM Jestec <Jestec@taylors.edu.my> wrote:
[Quoted text hidden]

---

okfalisa saktioto <okfalisa@gmail.com> Mon, Sep 28, 2020 at 7:53 AM
To: Jestec <Jestec@taylors.edu.my>

Dear editorial team.

Is there any progress regarding the correction of my manuscript?
and when will the paper be published?

Thank you

Best regards

Okfalisa

[Quoted text hidden]

Gmail **okfalisa saktioto <okfalisa@gmail.com>**

Paper ID EE20099: Extended Revision

3 messages

okfalisa saktioto <okfalisa@gmail.com> Mon, Aug 10, 2020 at 11:02 PM
To: Jestec <jestec@taylors.edu.my>

Dear Jestec Editorial Team,

As your email reply and notice in point **6**, please allow me to complete the revision by 25th of August 2020 since I have to do some more corrections both content and grammatical order in English.

I am looking forward to your prompt response. Thank you.

Best regards,

Okfalisa

Jestec <Jestec@taylors.edu.my> Wed, Aug 12, 2020 at 10:51 PM
To: okfalisa saktioto <okfalisa@gmail.com>

We agree

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

[Quoted text hidden]

---

okfalisa saktioto <okfalisa@gmail.com> Wed, Aug 12, 2020 at 11:52 PM
To: Jestec <Jestec@taylors.edu.my>

Many thanks prof..
[Quoted text hidden]

Gmail **okfalisa saktioto <okfalisa@gmail.com>**

Paper ID (EE20099) Review process is completed

2 messages

Jestec <Jestec@taylors.edu.my> Sun, Oct 11, 2020 at 10:08 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

I am glad to advise that your paper has been accepted for publication without modification. The reviewers have no more comments and they are satisfied with the revised paper.

By this the review process is completed and we kindly ask you to check the format of the paper according to the instructions for authors and JESTEC template (attached).

Special attention to be paid for list of symbols used and the references. Please follow strictly the instructions for citation of the references (attached are instructions) and explain each symbol you used and its SI units. Also refer to this link: http://jestec.taylors.edu.my/instructions.html

You are also kindly required to fill in the JESTEC-Copyright transfer form (use this link to download

http://jestec.taylors.edu.my/Copyright%20transfer%20ver%20190818.doc and send to the journal.

Kindly note that you have only four weeks to submit the above.

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

2 attachments

JESTEC template (Camera Ready)_new.docx
219K

about formatting the references.docx
15K

okfalisa saktioto <okfalisa@gmail.com> Sun, Oct 11, 2020 at 10:14 PM
To: Jestec <Jestec@taylors.edu.my>

Thank you, I will do that as soon as possible.

Best regards

Okfalisa
[Quoted text hidden]

Gmail okfalisa saktioto <okfalisa@gmail.com>

Review process is completed paper (EE20099) /formatting, proofreading, payment/

4 messages

Jestec <Jestec@taylors.edu.my> Sat, Oct 17, 2020 at 10:42 AM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author (s)

Thank you for your email and sending your modified paper. We found that the paper still contains some formatting mistakes.

We would like to inform you that your paper has been scheduled to be published in **April 2021, Volume 16 Issue 2**

Attached please find the acceptance letter.

Please send us up-to-date copyright transfer form. Download from here JESTEC-Copyright transfer form (CRTF)

Payment of the publication is needed before the paper is published online.

Kindly refer to the attached sample of the invoice and amend it (Red text only) according to your up-to-date and accurate information for the purpose of the payment. Once submitted we will send you an official invoice with all details to make safe payment.

We thank you very much for your interest in JESTEC and looking forward for new contribution.

Best regards

Assoc. Prof. Dr. Abdulkareem Sh. Mahdi Al-Obaidi, CEng MIMechE

Executive Editor, Journal of Engineering Science & Technology

http://jestec.taylors.edu.my

2 attachments

177 LoA_16_2_21 OKFALISA et al.pdf
56K

20_177.docx
22K

okfalisa saktioto <okfalisa@gmail.com> Sat, Oct 17, 2020 at 11:58 AM
To: Jestec <Jestec@taylors.edu.my>

Ok thank you for your information.
[Quoted text hidden]

okfalisa saktioto <okfalisa@gmail.com> Sat, Oct 17, 2020 at 8:00 PM
To: Jestec <Jestec@taylors.edu.my>

Dear Executive Editor,

Here we attached our copyright transfer and invoice of my paper id EE20099

My modified paper formatted will be sent to you as soon as possible

Best Regards

Okfalisa

[Quoted text hidden]

2 attachments

Copy Right Okfalisa.pdf
1900K

20_177.docx
32K

Jestec <Jestec@taylors.edu.my> Sat, Oct 17, 2020 at 8:05 PM
To: okfalisa saktioto <okfalisa@gmail.com>

Dear Author

Thank you for submitting the information.

Soon, our finance department will send an official invoice containing all details for making safe payment.

Please take note of the following:

- The only payment method is via Telegraphic Transfer (outside Malaysia) or Online Transfer (inside Malaysia).
- Banking details are provided in the invoice that will be sent to you.
- You have option to pay either in USD or RM.
- In either case the net amount to be received is exactly as stated in the invoice.
- The journal will not accept any bank charges associated with the transfer of money or currency exchange charges. Authors should bear all these service charges.

Best Regards

JESTEC Editor

http://jestec.taylors.edu.my

[Quoted text hidden]

OUTLINING HOW THE ISSUES ARE ADDRESSED

Title of paper:

1. Address all the concerns/recommendations of the reviewers.
2. All amendments made are to be highlighted in red color in the revised paper.

Reviewer # 1

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
• Some tables can be combined in one table		Done (Combine Table 2-9) into Table 4

Table 4. Attribute Discretization

Age discretization (1)		Systolic TD discretization (Sis) (9)	
Age (years)	**Discretization**	**Systolic BP (mmHg)**	**Discretization**
25 ≤ U <35	0	Sis<120	Optimal (0)
35 ≤ U <45	0.2	120< Sis <130	Normal (0.2)
45 ≤ U <55	0.4	130< Sis <140	Normal Height (0.4)
55 ≤ U <65	0.6	140< Sis <150	Low hypertension (0.6)
65 ≤ U <75	0.8	150< Sis <160	Moderate hypertension (0.8)
U ≥ 85	1	Sis >160	Severe hypertension (1)
Diastolic TD (Dias) discretization (10)		**Discretization of LDL (LDL) levels (11)**	
Diastolic BP (mmHg)	**Discretization**	**LDL levels (mg / dL)**	**Discretization**
Dias<80	Optimal (0)	LDL<100	Optimal (0)
80≤Dias<85	Normal (0.2)	100< LDL <130	Approaching optimal (0.25)
85≤Dias<90	Normal Height (0.4)	130< LDL <160	Borderline high (0.5)
90≤Dias<100	Low hypertension (0.6)	160< LDL <190	High (0.75)
100≤Dias<110	Moderate hypertension (0.8)	LDL >190	Very high (1)
Dias≥110	Severe hypertension (1)		
Discretization of HDL(HDL) (12)		**Discretization of total cholesterol (Chol)(13)**	
HDL levels (mg / dL)	**Discretization**	**Chol levels (mg / dL)**	**Discretization**
HDL<40	Low (0)	Chol <200	Desirable (expected to be safe) (0)
40≤ HDL <60	Normal (0.5)	200≤ Chol <240	Borderline (must be aware- begin to control) (0.5)

		HDL ≥60 High (1) Chol ≥240 High (1) **Triglyceride discretization (14)** **Glucose Level discretization (Glu) (15)** **Triglyceride levels (mg / dL)** **Discretization** **Glucose Levels (mg/dL)** **Discretization** **trig <150** Normal (0) Glu<40 Optimal (0) **150≤ trig <200** Borderline high (0.33) 40≤ Glu <60 Normal (0.2) **200≤ trig <500** High (0.66) 60≤ Glu <125 Normal Height (0.4) **trig ≥500** Very High (1) 125≤ Glu <145 Low hypertension (0.6) 145≤ Glu <200 Moderate hypertension (0.8) Glu ≥200 Severe hypertension (1)
• Some tables can be converted to a figures to be more clear such as table 11, 12, and 13		Table 11, 12, and 13 has been converted into Table 6,7, and 8 respectively. Figure 3,4, and 5 explained in more details regarding on the performance. It explained in the text as follows. To investigate the implication of pre-processing against SVM, the analysis is conducted by comparing the accuracy within dataset changes in the original data (without missing values), the reduced (with missing values), k-NN (with distance calculation), and pre-processing (KDD formatted). This was executed through the selection of the best parameters for 10-fold cross-validation in Table 6 and percentage split in Table 7 for four scenarios dataset. The graphical views of performances are shown in Figure 3, 4, and 5.

Table 6. The accuracy of the best parameter - 10-fold cross validation.

Kernel	**Polynomial**			**RBF**		
Parameters	**C**	**d/σ**	**Accuracy**	**C**	**d/σ**	**Accuracy**
Original Dataset	0.03	1	100%	0.01	1	47.9%
Reduced Dataset	0.03	1	100%	0.01	1	48.1%
k-NN Dataset	0.03	1	100%	0.01	1	47.9%
Pre-processing Dataset	0.02	2	100%	0. 8	1	98.9%

Table 7. The accuracy of the best parameter pairs -percentage split.

Kernel	**Polynomial**			**RBF**		
Parameters	**DC**	**T (s)**	**Accuracy**	**DC**	**T(s)**	**Accuracy**
Original Dataset	70:30	27.49	100%	40:60	0.06	49.4%
Reduced Dataset	70:30	21.37	100%	70:30	0.13	53.8%
k-NN Dataset	80:20	24.55	100%	40:60	0.08	49.4%
Pre-processing Dataset	70:30	0.06	100%	80:20	0.08	100%

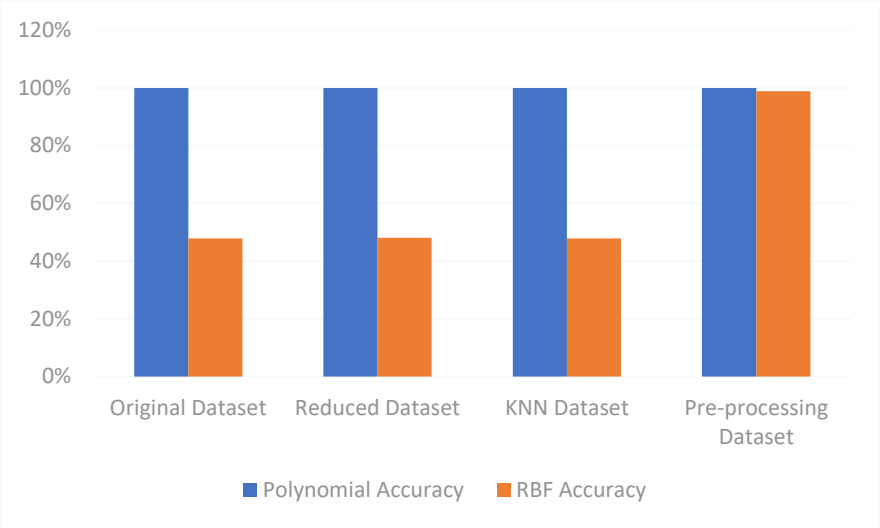


Fig 3. Dataset performance based on accuracy - 10-fold cross validation

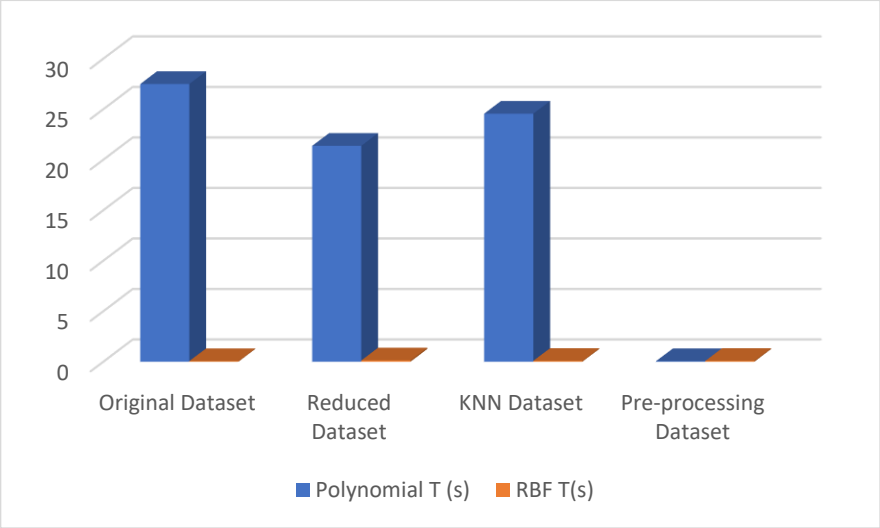


Fig 4. Dataset performance based on time (s) - percentage split

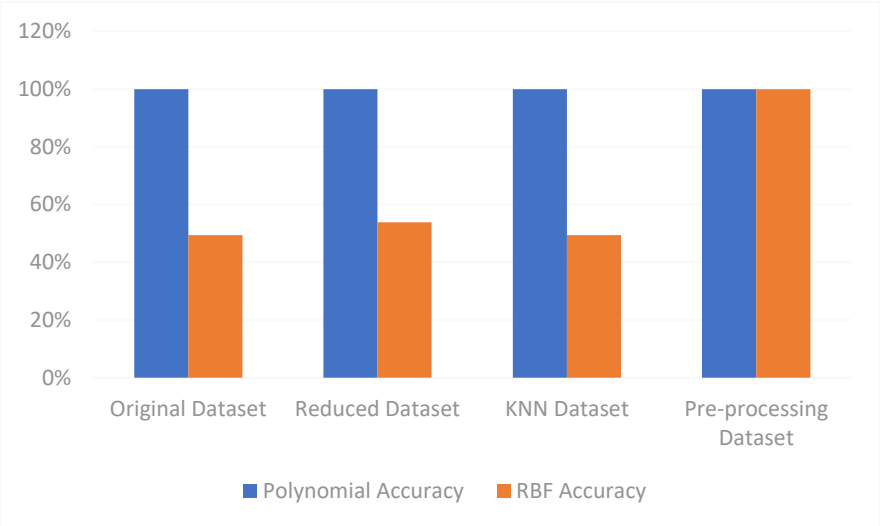


Fig 5. Dataset performance based on accuracy - percentage split

The execution of 10-folds cross-validation in Table 6 explained that the pre-processing dataset improved accuracy level up to 100% and 98.9% in kernel polynomial and RBF, respectively with the superior parameters at C = 0.02 and d = 2, C = 0.8 and σ = 1, respectively. Similarly, Table 7 shows that the pre-processing dataset with the percentage split treatment also provided a significant growth of accuracy in polynomial and RBF kernel. Moreover, the execution time in model development considerably impacts the performance of pre-processing both in Polynomial and RBF kernel at the data composition of 70:30 and 80:20, respectively. Figures 3, 4, and 5 explained that the pre-processing dataset increases its performance in terms of time (s) and accuracy for Polynomial and RBF kernel.

Table 8. Confusion Matrix for SVM and NN-Polynomial and RBF.

SVM: Dataset Pre-processing						
	Polynomial			RBF		
Class	Prediction Class			Prediction Class		
	UAP	NSTEMI	STEMI	UAP	NSTEMI	STEMI
UAP	29	0	0	29	0	0
NSTEMI	0	13	0	0	13	0
STEMI	0	0	14	0	0	14
Accuracy	100%			100%		
Error rate	0			0		
Precision	1			1		
Recall	1			1		

NN-Multilayer Perceptron			
Class	Prediction Class		
	UAP	NSTEMI	STEMI
UAP	18	0	1
NSTEMI	2	18	1
STEMI	0	2	14
Accuracy	89%		
Error rate	0.11		
Precision	0.89		
Recall	0.89		

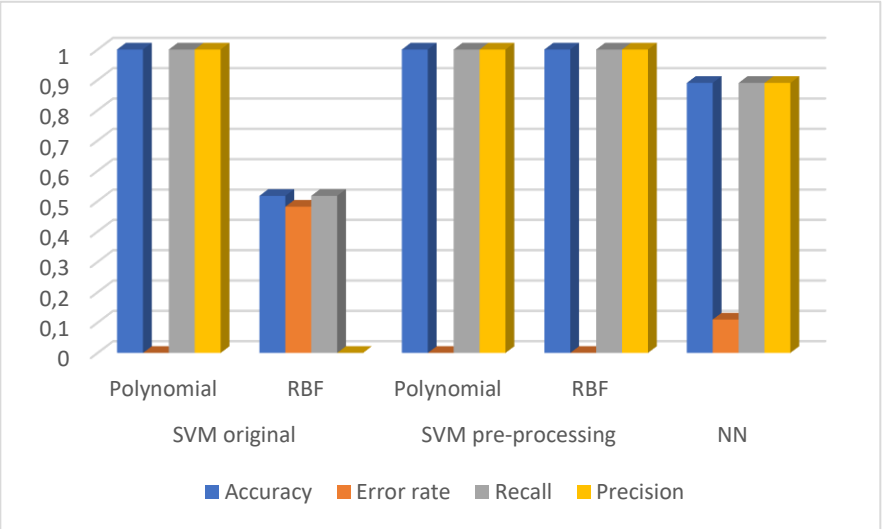


Fig. 6. Performance Polynomial and RBF kernel

(Please add more rows if needed)

Reviewer # 2

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
• How to select the value of C, d in polynomial kernel and c, σ for RBF.		The variable *d* is specified as the degree of the polynomial, the value of *C* is a constant that allows to trade off the influence of the higher and lower-order terms and this is a consideration for varying *C* values between 0.01 and 1. The selection values of *d*, and σ impact the performance accuracy, while *C* is selected based on the *C* function as a constraint, therefore, a greater value of *C* implies more penalty for classification errors. Meanwhile, the values of σ provide a good fit or an overfit to the data, when σ is large compared to the distance between the classes, it results in an overly flat discriminant surface. However, a smaller σ value compared to the distance between classes result in an over-fit [36]. A good choice for σ will be comparable to the distance between the closest members of the two classes. Furthermore, the highest accuracy of parameter pairs during the training session was found at *C and* σ for kernel RBF as well as *C and d* for the polynomial kernel.
• How to measure the testing accuracy and what parameters used to measure the testing accuracy.		the success rate of classification, the determination of accuracy, error rate, precision, and recall values are performed based on the confusion matrix as depicted in Eq. (2)-(5) [40] given by, Accuracy = $\frac{TP+TN}{P+N} \times 100\%$ (2) Error-rate = $\frac{FP+FN}{P+N} \times 100\%$ (3) Precision = $\frac{TP}{TP+FP}$ (4) Recall = $\frac{TP}{TP+FN}$ (5) TP *(True Positive)* = The amount of correctly classified data (*Actual class (yes), Predicted class (yes)).* TN *(True Negative)* = The amount of correctly classified data (*Actual class (no), Predicted class (no)).* FN *(False Negative)* = The amount of incorrectly classified data (*Actual class (yes), Predicted class (no)).* FP *(False Positive)* = The amount of incorrectly classified data (*Actual class (no), Predicted class (yes)).* P = Total of TP and FN N = Total of FP and TN
• Detailed discussion is needed for result		We have explained in more detail for the result and discussion part with additional tables, graphics, and comparison analysis with other classifier, namely Neural Network (NN). We also added one section for discussion in chapter 3.3. **1. The Research Result and Discussion**

and Discussion part

3.1. The Result of KDD analysis

3.1.1. Pre-processing data analysis

The data were manually selected from the medical record of 280 CHD patients at Central Hospital by paying special attention to the feature related to attributes and missing value treatments. The diversity of data based on the feature is shown in Table 1 and missing value consideration in Fig. 1. Table 3 explains that the increasing numbers of training data from 40%, 50%, 60%, 70%, and 80% are directly proportional to the diversity of data in accordance with seventeen attributes and three classes (UAP=1, NSTEMI=2, and STEMI=3). Consequently, the new pattern tested data is recognized easily. Also, Figure 1 describes the transformation of pre-processing activity before and after manipulating the missing values by referring to k-NN distance calculation in Eq. (1). The missing values in the dataset at number 28 column 11, 12, and 14 is replaced by 93, 57, and 84 respectively as well as the missing values at dataset number 69, and 71.

Table 3. Data Diversity according to The Feature.

	Training Data Composition								
Feature	**40%**				…	**80%**			
	Area	**Classes**				**Area**	**Classes**		
		1	**2**	**3**	…		**1**	**2**	**3**
Age (1)	37-44	2	3	2	…	25-31	0	1	0
	45-51	14	5	4	…	32-37	2	1	3
	52-58	16	10	9	…	38-43	1	3	3
	59-65	13	6	5	…	44-49	19	11	10
	66-72	5	5	6	…	50-55	32	22	13
	73-79	3	2	0	…	56-61	20	10	8
	80-86	2	0	0	…	62-67	15	8	8
					…	68-73	6	7	6
					…	74-79	6	5	0
					…	80-86	4	0	0
Gender (2)	M	36	23	21	…	M	67	54	36
	F	19	8	5	…	F	38	14	15
…	…	…	…	…	…	…	…	…	…
Cardiac Enzymes (17)	Norm	55	0	0	…	Norm	105	0	0
	High	0	31	26	…	High	0	68	51

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Case
27	68	F	Yes	No	Yes	Yes	No	No	144	93	160.5	31.1	207	77	381	Yes	High	STEMI
28	63	M	No	Yes	Yes	Yes	No	No	140	90	?	?	212	?	262	Yes	High	STEMI
31	56	M	No	No	No	Yes	No	No	150	90	138.6	38.4	198	105	83	Yes	High	STEMI
69	68	M	No	No	No	No	No	No	100	70	?	?	164	241	76	Yes	High	STEMI
71	54	M	No	No	No	Yes	No	No	120	90	139.1	70	226	?	88	Yes	High	STEMI

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Case
27	68	F	Yes	No	Yes	Yes	No	No	144	93	160.5	31.1	207	77	381	Yes	High	STEMI
28	63	M	No	Yes	Yes	Yes	No	No	140	90	93	57	212	84	262	Yes	High	STEMI
31	56	M	No	No	No	Yes	No	No	150	90	138.6	38.4	198	105	83	Yes	High	STEMI
69	68	M	No	No	No	No	No	No	100	70	90.3	20.1	164	241	76	Yes	High	STEMI
71	54	M	No	No	No	Yes	No	No	120	90	139.1	70	226	140	88	Yes	High	STEMI

Fig. 1. Pre-processing with missing value.

3.1.2. Transformation data analysis

The medical records of CHD patients were collected in a variety of formats. Consequently, the discretization with the equal width approach was applied in expressing the standard range values from 0 to 1 as in Eq. (6).

$$\text{Series of range} = \frac{\text{the highest area} - \text{the lowest area}}{\text{The number of categories}} \quad (6)$$

The discretization of attributes is depicted in Table 4 and Table 5. Table 4 defines the values of attribute 1 for age discretization, attribute 9 for systolic blood pressure (BP), attribute 10 for diastolic blood pressure, attribute 11 for LDL, attribute 12 for HDL, attribute 13 for Total cholesterol, attribute 14 for Triglyceride, and attribute 15 for a glucose level. The rest of the attributes (2,3,4,5,6,7,8,16, and 17) were categorized into two series and discretized into 0 value for "No" and 1 for "Yes" as shown in Table 5. This discretization value will be the format for SVM input. The sample of format SVM input is described in Figure 2.

Table 4. Attribute Discretization

Age discretization (1)		**Systolic TD discretization (Sis) (9)**	
Age (years)	**Discretization**	**Systolic BP (mmHg)**	**Discretization**
25 ≤ U <35	0	Sis<120	Optimal (0)
35 ≤ U <45	0.2	120< Sis <130	Normal (0.2)
45 ≤ U <55	0.4	130< Sis <140	Normal Height (0.4)
55 ≤ U <65	0.6	140< Sis <150	Low hypertension (0.6)
65 ≤ U <75	0.8	150< Sis <160	Moderate hypertension (0.8)
U ≥ 85	1	Sis >160	Severe hypertension (1)
Diastolic TD (Dias) discretization (10)		**Discretization of LDL (LDL) levels (11)**	
Diastolic BP (mmHg)	**Discretization**	**LDL levels (mg / dL)**	**Discretization**
Dias<80	Optimal (0)	LDL<100	Optimal (0)
80≤Dias<85	Normal (0.2)	100< LDL <130	Approaching optimal (0.25)
85≤Dias<90	Normal Height (0.4)	130< LDL <160	Borderline high (0.5)
90≤Dias<100	Low hypertension (0.6)	160< LDL <190	High (0.75)
100≤Dias<110	Moderate hypertension (0.8)	LDL >190	Very high (1)
Dias≥110	Severe hypertension (1)		
Discretization of HDL(HDL) (12)		**Discretization of total cholesterol (Chol)(13)**	
HDL levels (mg / dL)	**Discretization**	**Chol levels (mg / dL)**	**Discretization**
HDL<40	Low (0)	Chol <200	Desirable (expected to be safe) (0)
40≤ HDL <60	Normal (0.5)	200≤ Chol <240	Borderline (must be aware- begin to control) (0.5)
HDL ≥60	High (1)	Chol ≥240	High (1)
Triglyceride discretization (14)		**Glucose Level discretization (Glu) (15)**	

Triglyceride levels (mg / dL)	Discretization	Glucose Levels (mg/dL)	Discretization
trig <150	Normal (0)	Glu<40	Optimal (0)
150≤ trig <200	Borderline high (0.33)	40≤ Glu <60	Normal (0.2)
200≤ trig <500	High (0.66)	60≤ Glu <125	Normal Height (0.4)
trig ≥500	Very High (1)	125≤ Glu <145	Low hypertension (0.6)
		145≤ Glu <200	Moderate hypertension (0.8)
		Glu ≥200	Severe hypertension (1)

Table 5. Attributes with two series discretization

Attributes	Discretization	
Gender (2)	Male	1
	Female	0
Family History (3)	None	0
	Yes	1
Heart History (4)	None	0
	Yes	1
DM History (5)	None	0
	Yes	1
Hypertension History (6)	None	0
	Yes	1
Cholesterol History (7)	None	0
	Yes	1
Obesity (8)	None	0
	Yes	1
Elevation (16)	None	0
	Yes	1
Cardiac Enzymes (17)	None	0
	Yes	1

	Attributes discretization																	
No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Cases
1	0,4	1	0	0	0	0	0	0	0,4	0,2	0	0	0	0	0,4	0	0	UAP
2	0,4	1	0	0	0	0	0	0	0,8	0,8	0	0	0	0	0,4	0	0	UAP
3	0,4	0	0	1	0	0	0	0	0,2	0,2	0	0,5	0	0	0,4	0	0	UAP
4	0,6	1	0	1	0	0	0	0	0,2	0,2	0,25	0,5	0	0	0,8	0	0	UAP
5	0,6	1	0	0	1	0	0	0	1	0,2	0,25	0,5	0	0	0,4	0	0	UAP
6	0,6	1	0	1	1	1	1	0	0,2	0,2	0,5	1	0,5	0	0,4	0	1	NSTEMI
7	0,6	1	0	0	0	0	0	0	0,6	0,2	0	0	0	0,33	0,8	0	0	UAP
8	0,4	0	0	1	0	0	0	0	0,6	0,8	0,25	1	0,5	0	0,4	0	0	UAP
9	0,6	1	0	0	1	1	0	0	0,6	0,2	0,75	0	0,5	0	0,4	0	0	UAP
10	0,4	0	0	1	1	1	0	0	0,4	0,2	0,25	0,5	0,5	0,66	0,4	0	0	UAP
11	0,2	1	0	1	1	1	0	0	0,2	0	0,25	0	0	0	0,4	0	0	UAP
12	0,6	1	0	1	1	1	0	0	1	0,8	0,25	0	0	0,33	0,4	0	0	UAP
13	0,4	0	0	0	0	0	0	0	0	0	1	0	1	0	0,4	0	0	UAP
14	0,4	1	0	1	1	1	0	0	0,6	0,6	0	0	0	0	0,4	0	0	UAP
15	0,4	1	0	1	1	1	0	0	0,6	0,8	0,5	0	0,5	0,66	0,4	0	0	UAP
16	0,4	0	0	0	1	1	0	0	0,6	0,8	0,25	0	0	0	0,8	0	1	NSTEMI
17	0,8	0	0	0	0	0	0	0	1	0,8	0,25	0	0,5	0	0,8	1	1	STEMI
18	0,4	0	1	0	1	1	0	0	0,6	0,6	0	0	0	0	0,4	1	1	STEMI
19	0,6	1	0	0	0	0	0	0	0	0	1	0	1	0	0,4	1	1	STEMI
20	0,4	1	0	1	1	1	0	0	0,4	0,8	0,5	0	0	0	0,4	1	1	STEMI

Fig. 2. The sample of SVM input

3.1.3. SVM mining analysis

To investigate the implication of pre-processing against SVM, the analysis is conducted by comparing the accuracy within dataset changes in the original data (without missing

values), the reduced (with missing values), k-NN (with distance calculation), and pre-processing (KDD formatted). This was executed through the selection of the best parameters for 10-fold cross-validation in Table 6 and percentage split in Table 7 for four scenarios dataset. The graphical views of performances are shown in Figure 3, 4, and 5.

Table 6. The accuracy of the best parameter - 10-fold cross validation.

Kernel	Polynomial			RBF		
Parameters	**C**	**d/σ**	**Accuracy**	**C**	**d/σ**	**Accuracy**
Original Dataset	0.03	1	100%	0.01	1	47.9%
Reduced Dataset	0.03	1	100%	0.01	1	48.1%
k-NN Dataset	0.03	1	100%	0.01	1	47.9%
Pre-processing Dataset	0.02	2	100%	0. 8	1	98.9%

Table 7. The accuracy of the best parameter pairs -percentage split.

Kernel	Polynomial			RBF		
Parameters	**DC**	**T (s)**	**Accuracy**	**DC**	**T(s)**	**Accuracy**
Original Dataset	70:30	27.49	100%	40:60	0.06	49.4%
Reduced Dataset	70:30	21.37	100%	70:30	0.13	53.8%
k-NN Dataset	80:20	24.55	100%	40:60	0.08	49.4%
Pre-processing Dataset	70:30	0.06	100%	80:20	0.08	100%

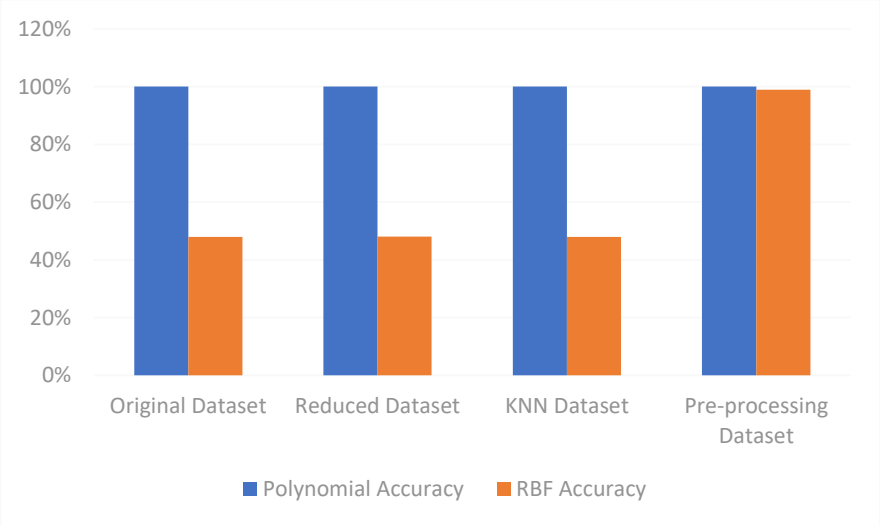


Fig 3. Dataset performance based on accuracy - 10-fold cross validation

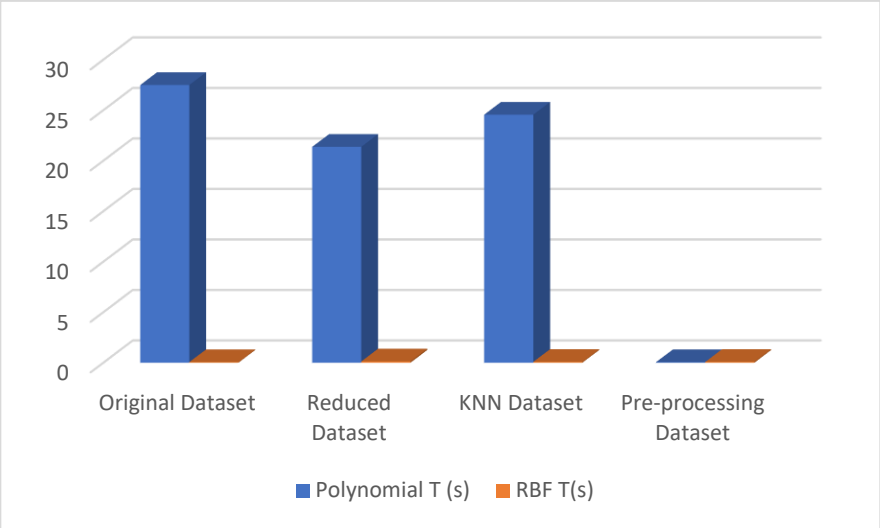


Fig 4. Dataset performance based on time (s) - percentage split

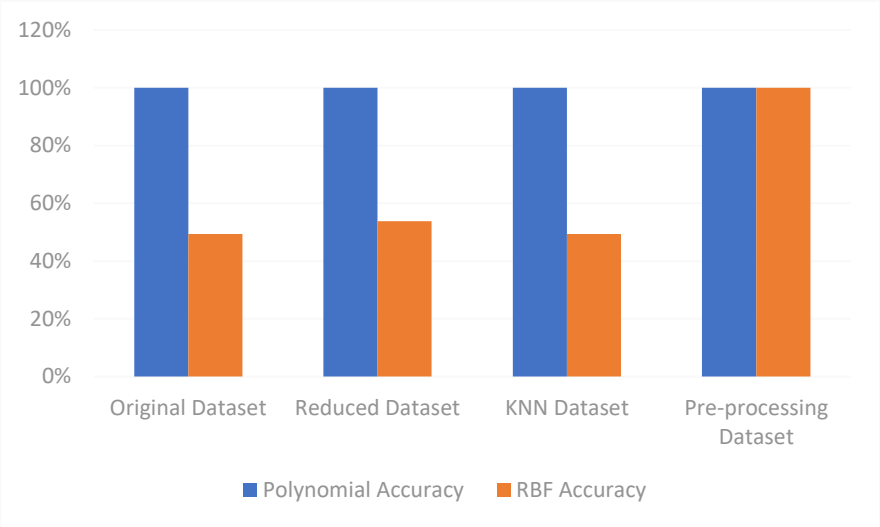


Fig 5. Dataset performance based on accuracy - percentage split

The execution of 10-folds cross-validation in Table 6 explained that the pre-processing dataset improved accuracy level up to 100% and 98.9% in kernel polynomial and RBF, respectively with the superior parameters at C = 0.02 and d = 2, C = 0.8 and σ = 1, respectively. Similarly, Table 7 shows that the pre-processing dataset with the percentage split treatment also provided a significant growth of accuracy in polynomial and RBF kernel. Moreover, the execution time in model development considerably impacts the performance of pre-processing both in Polynomial and RBF kernel at the data composition of 70:30 and 80:20, respectively. Figures 3, 4, and 5 explained that the pre-processing dataset increases its performance in terms of time (s) and accuracy for Polynomial and RBF kernel.

3.2 Testing

To evaluate the classification of CHD patient's dataset in SVM, the testing procedure was undertaken according to the Test Option Supplied on the Confusion Matrix formula [39]. The pre-processing dataset was put in place on 20% of tested data at C = 0.02 and d = 2 in the polynomial kernel and the values of C and σ are 0.8 and 1 respectively, in the RBF. In addition, the resemblance of SVM with another classifier, namely Multilayer perceptron Neural Network (NN) is operated to deeply observe the effectiveness of SVM. The confusion matrix for the above dataset of SVM and NN was explained in Table 8. This table showed that the classification in the pre-processing dataset for SVM is more

accurate compared to NN, especially for RBF kernel. By comparing the values for error rate, precision and recall between polynomial kernel and RBF based on the confusion matrix computation as a side of SVM and NN, Figure 6 is obtained. The figure showed that SVM for Polynomial kernel has 100% accuracy, "0" for error rate, and "1" for precision, and recall. Meanwhile, RBF kernel discharged from 51.79% into 100% accuracy, 0.48 into 1 for error rate, undefined into 1 for precision, and 0.52 into 1 for recall. Also, NN for polynomial kernel achieved 89% accuracy, "0.11" for error rate, and "0.89" for precision and recall.

Table 8. Confusion Matrix for SVM and NN-Polynomial and RBF.

SVM: Dataset Pre-processing						
	Polynomial			RBF		
Class	Prediction Class			Prediction Class		
	UAP	NSTEMI	STEMI	UAP	NSTEMI	STEMI
UAP	29	0	0	29	0	0
NSTEMI	0	13	0	0	13	0
STEMI	0	0	14	0	0	14
Accuracy	100%			100%		
Error rate	0			0		
Precision	1			1		
Recall	1			1		

NN-Multilayer Perceptron			
Class	Prediction Class		
	UAP	NSTEMI	STEMI
UAP	18	0	1
NSTEMI	2	18	1
STEMI	0	2	14
Accuracy	89%		
Error rate	0.11		
Precision	0.89		
Recall	0.89		

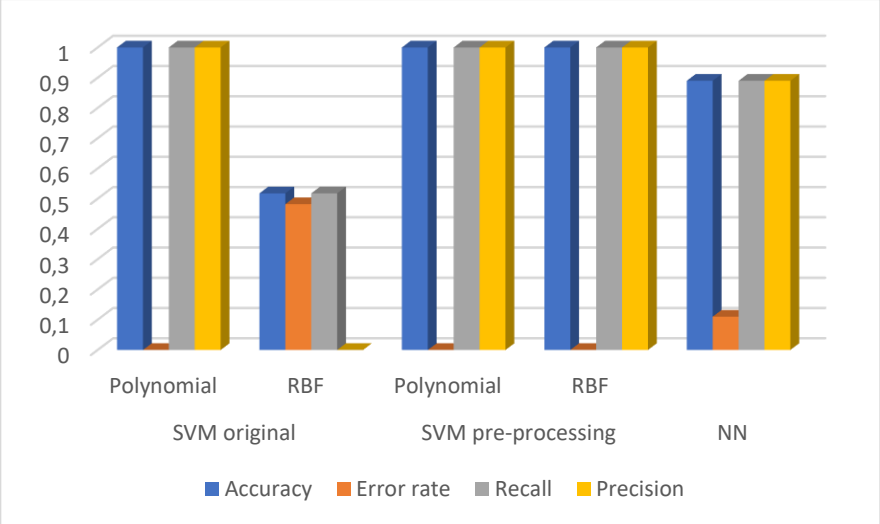


Fig. 6. Performance Polynomial and RBF kernel

3.3 Discussion

This result reveals that the pre-processing dataset in SVM provides significant values on the accuracy, error rate, precision, and recall, even though it exceeds NN capacity. As studied by [42], the SVM approach gives better predictive capability than other models, including NN. This, of course, has far-reaching implications in the medical context that

<table>
<tr><td></td><td></td><td>require increasing sensitivity, specificity (the ability to predict the absence of the condition when it is not present) as well as discriminatory power of the classifier as key features to consider when comparing classifiers and diagnostic methods [45]. In the reviews on kernel type, the simulation presented that SVM polynomial is more reliable on the dataset changes compare to RBF. Consequently, the pre-processing prescription on SVM-RBF will undoubtedly boost RBF performance. Furthermore, selecting the specific kernel is an important research issue for kernel-based learning in the data mining area and the problem of SVM kernels is found in fitting the appropriate parameter values [46]. This investigation revealed that the SVM polynomial kernel mediates the accuracy and efficiency of the diagnostic results based on the parameters defined in CHD.</td></tr>
<tr><td>• Why you are particularly select the SVM ? what are the features compared to other AI techniques.</td><td></td><td>SVM is a classification method that produces a fairly high degree of accuracy and is commonly used compared with the conventional decision tree, ANN [16, 17] and other classifiers [18]. Furthermore, Sivagami [19] compared SVM, Multilayer Perception (MLP), One R, and Decision Tree J48 methods in the classification of breast cancer. The results showed that SVM with kernel type RBF provided the highest accuracy rate of 95%, 91% in polynomial type, and 90% in linear type. One R exhibited 83%, 80% in J48 and 74.1% in MLP, which is the lowest performance. The comparison of SVM and Left Anterior Descending (LDA) for the classification of Coronary heart showed accuracy at 96.86% and 78.18% respectively [20]. Furthermore, Mo and Xu [21] attempted to improve the performance of SVM based on the hybrid kernel function using the optimization of the Particle Swarm Optimization (PSO) algorithm in heart disease diagnosis. Meanwhile, the accuracy of SVM in the early diagnosis of a heart condition by modifying the kernel width using trial and error approach significantly increase by 18.2% [22]. This showed that the kernel function on SVM provides the opportunities in enhancing the accuracy. Unfortunately, some difficulties in choosing the SVM kernel function were encountered [23], as well as flexibility in dataset changing [24], selecting optimal features, and time-consumption [25].</td></tr>
<tr><td>• Why the % split are different? 70:30, 80:20.</td><td></td><td>A common strategy is to take all available labeled data, and split it into training and evaluation subsets, usually with a ratio of 70-80 percent for training and 20-30 percent for evaluation [39]. To make a deep investigation, we use several percentage splits as comparison, namely 40:60, 50:50, 60:40, 70:30, 80:20.

Explanation in the paper:
Also, the 10-folds validation and confusion matrix with percentage splits on the portion of training data compare to test data in 40:60, 50:50, 60:40, 70:30, and 80:20 is applied to support the assessment process. However, there are no specific rules in the distribution of training-data and test-data, therefore, a large number of the former will represent the diversity of the data [39].

39. Kemal Polat, Bayram Akdemir, Salih Gunes. (2008). Computer aided diagnosis of ECG data on the least square support vector machine. Digital Signal Processing, 18(1), 25-32.</td></tr>
<tr><td>• How to select the input parameters? Any analysis is doing for selection of input parameters.</td><td></td><td>In selecting data that limits the patient's age beyond 25 years, 17 attributes were exploited and they were defined based on the reviews of previous researches [27-33] as presented in Table 1.

Table 1. Numbers of Attributes
<table><tr><th>Code</th><th>Attributes</th></tr><tr><td>1</td><td>Age</td></tr><tr><td>2</td><td>gender</td></tr><tr><td>3</td><td>family history</td></tr><tr><td>4</td><td>heart history</td></tr><tr><td>5</td><td>history of diabetes mellitus</td></tr><tr><td>6</td><td>history of hypertension</td></tr><tr><td>7</td><td>history of cholesterol</td></tr><tr><td>8</td><td>obesity</td></tr></table></td></tr>
</table>

		9 systolic blood pressure **10** diastolic blood pressure **11** LDL levels **12** HDL levels **13** total cholesterol levels **14** triglyceride levels **15** blood levels glucose **16** elevation **17** cardiac enzymes 27. Dirjen Bina Kefarmasian dan AlKes DepKes RI. (2006). Pharmaceutical Care Untuk Pasien Penyakit jantung Koroner : Fokus Sindrom Koroner Akut. Jakarta: Departmen Kesehatan RI. 28. Magesh, G.; and Swarnalatha, P. (2020). Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction. Evolutionary Intelligence. Special Issue, 1-11. 29. Arad, Y.; Goodman, K.J.; Roth, M.; Newstein, D.; and Guerci, A.D. (2005). Coronary calcification, coronary disease risk factors, c-reactive protein, and atherosclerotic cardiovascular disease events. Journal of the American College of Cardiology, 46(1), 158–165. 30. Hand, D.; Mannila, H.; and Smyth, P. (2001). Principles of data mining, Massachusetts London: The MIT Press. 31. Nauta, S. T.; Deckers, J.W.; Boon, R.M. Van Der; Akkerhuis, K.M.; and Domburg, R.T. Van. (2014). Risk factors for coronary heart disease and survival after myocardial infarction. European Journal of Prevetive Cardiology, 21(5), 576–583. 32. Mannsverk, J.; Wilsgaard, T.; Mathiesen, E.B.; Løchen, M.; Rasmussen, K.; Thelle, D.S.; and Bonna, K.H. (2015). Trends in modifiable risk factors are associated with declining incidence of hospitalized and nonhospitalized acute coronary heart disease in a population. Circulation, 133(1),74–81. 33. Sharma, P.; Choudhary, K.; Gupta, K.; Chawla, R.; Gupta, D.; and Sharma, A. (2019). Artificial plant optimization algorithm to detect heart rate and presence of heart disease using machine learning. Artificial Intelligence in Medicine, 102, 101752.
• What is the necessity of preprocessing and what method used for the preprocessing?		Following the employment of SVM in KDD [34]. Step 1: Pre-processing This step is to reduce data, therefore there is no missing value. The activity begins with data selection from CHD and then performed as an effort to feature subset selection by ignoring the irrelevant attributes CHD risk factors and missing values. In view of this, k-NN with the Euclidian distance calculation is performed in Eq. (1) $dist = \sqrt{\sum_{k=1}^{n}(pk - qk)^2}$ (1) where *n* is number of attributes, *pk* and *qk* values are the -*k* attribute. 34. Ivezic, Z. (2011). Data Mining and Machine Learning in Astronomy: A Practical Guide. Princeton: Princeton University Press.

(Please add more rows if needed)

Reviewer # 3				
Final Recommendation	**Accepted without modification**	**Accepted with minor corrections**	**Accepted with major modification**	**Rejected**
Please tick	☐	☐	☐	☐

Comments	Addressed (Y/N)	Reply/Action taken
• For preprocessing what algorithm is employed or a clear explanation about preprocessing is required		Following the employment of SVM in KDD [34]. Step 1: Pre-processing This step is to reduce data, therefore there is no missing value. The activity begins with data selection from CHD and then performed as an effort to feature subset selection by ignoring the irrelevant attributes CHD risk factors and missing values. In view of this, k-NN with the Euclidian distance calculation is performed in Eq. (1) $dist = \sqrt{\sum_{k=1}^{n}(pk - qk)^2}$ (1) where *n* is number of attributes, *pk* and *qk* values are the -*k* attribute. 34. Ivezic, Z. (2011). Data Mining and Machine Learning in Astronomy: A Practical Guide. Princeton: Princeton University Press.
• Explanation about KNN data set is required.		Also, Figure 1 describes the transformation of pre-processing activity before and after manipulating the missing values by referring to k-NN distance calculation in Eq. (1). The missing values in the dataset at number 28 column 11, 12, and 14 is replaced by 93, 57, and 84 respectively as well as the missing values at dataset number 69, and 71. Eq. (1) $dist = \sqrt{\sum_{k=1}^{n}(pk - qk)^2}$ (1) where *n* is number of attributes, *pk* and *qk* values are the -*k* attribute. **Fig. 1. Pre-processing with missing value.**
• Grammar corrections have to done		We have sent this paper to proof read. Here we attached the receipt

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Case
27	68	F	Yes	No	Yes	Yes	No	No	144	93	160.5	31.1	207	77	381	Yes	High	STEMI
28	63	M	No	Yes	Yes	Yes	No	No	140	90	?	?	212	?	262	Yes	High	STEMI
31	56	M	No	No	No	Yes	No	No	150	90	138.6	38.4	198	105	83	Yes	High	STEMI
69	68	M	No	No	No	No	No	No	100	70	?	?	164	241	76	Yes	High	STEMI
71	54	M	No	No	No	Yes	No	No	120	90	139.1	70	226	?	88	Yes	High	STEMI

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Case
27	68	F	Yes	No	Yes	Yes	No	No	144	93	160.5	31.1	207	77	381	Yes	High	STEMI
28	63	M	No	Yes	Yes	Yes	No	No	140	90	93	57	212	84	262	Yes	High	STEMI
31	56	M	No	No	No	Yes	No	No	150	90	138.6	38.4	198	105	83	Yes	High	STEMI
69	68	M	No	No	No	No	No	No	100	70	90.3	20.1	164	241	76	Yes	High	STEMI
71	54	M	No	No	No	Yes	No	No	120	90	139.1	70	226	140	88	Yes	High	STEMI


billing@native-proofreading.com

NATIVE PROOFREADING SERVICE (NPS)
WWW.NATIVE-PROOFEADING.COM

RECEIPT

AUGUST 15, 2020

BILL TO

Okfalisa
Universitas Islam Negeri Suska
Pekanbaru

Receipt No.: NPS/2020/2985

ITEM DESCRIPTION	PRICE
Coronary Heart Disease using Support Vector Machine	Rp 927,500
TOTAL	Rp 927,500

PAID

NATIVE PROOFREADING
SINCE 2013
www.native-proofreading.com

THIS DOCUMENT CERTIFIES THAT THE MANUSCRIPT LISTED ABOVE WAS EDITED FOR PROPER ENGLISH LANGUAGE, GRAMMAR, PUNCTUATION, SPELLING, AND OVERALL STYLE BY ONE OR MORE OF THE HIGHLY QUALIFIED NATIVE ENGLISH SPEAKING EDITORS AT NATIVE PROOFREADING SERVICE (NPS)

- Whether it is possible to compare SVM classifier with any other classifiers

We tried to compare SVM with Neural Network in testing.

2 Testing

To evaluate the classification of CHD patient's dataset in SVM, the testing procedure was undertaken according to the Test Option Supplied on the Confusion Matrix formula [39]. The pre-processing dataset was put in place on 20% of tested data at $C = 0.02$ and $d = 2$ in the polynomial kernel and the values of C and σ are 0.8 and 1 respectively, in the RBF. In addition, the resemblance of SVM with another classifier, namely Multilayer perceptron Neural Network (NN) is operated to deeply observe the effectiveness of SVM. The confusion matrix for the above dataset of SVM and NN was explained in Table 8. This table showed that the classification in the pre-processing dataset for SVM is more accurate compared to NN, especially for RBF kernel. By comparing the values for error rate, precision and recall between polynomial kernel and RBF based on the confusion matrix computation as a side of SVM and NN, Figure 6 is obtained. The figure showed that SVM for Polynomial kernel has 100% accuracy, "0" for error rate, and "1" for precision, and recall. Meanwhile, RBF kernel discharged from 51.79% into 100% accuracy, 0.48 into 1 for error rate, undefined into 1 for precision, and 0.52 into 1 for recall. Also, NN for polynomial kernel achieved 89% accuracy, "0.11" for error rate, and "0.89" for precision and recall.

Table 8. Confusion Matrix for SVM and NN-Polynomial and RBF.

SVM: Dataset Pre-processing

	Polynomial			RBF		
Class	**Prediction Class**			**Prediction Class**		
	UAP	**NSTEMI**	**STEMI**	**UAP**	**NSTEMI**	**STEMI**
UAP	29	0	0	29	0	0
NSTEMI	0	13	0	0	13	0
STEMI	0	0	14	0	0	14
Accuracy	100%			100%		
Error rate	0			0		
Precision	1			1		
Recall	1			1		

NN-Multilayer Perceptron			
Class	**Prediction Class**		
	UAP	**NSTEMI**	**STEMI**
UAP	**18**	**0**	**1**
NSTEMI	**2**	**18**	**1**
STEMI	**0**	**2**	**14**
Accuracy	**89%**		
Error rate	**0.11**		
Precision	**0.89**		
Recall	**0.89**		

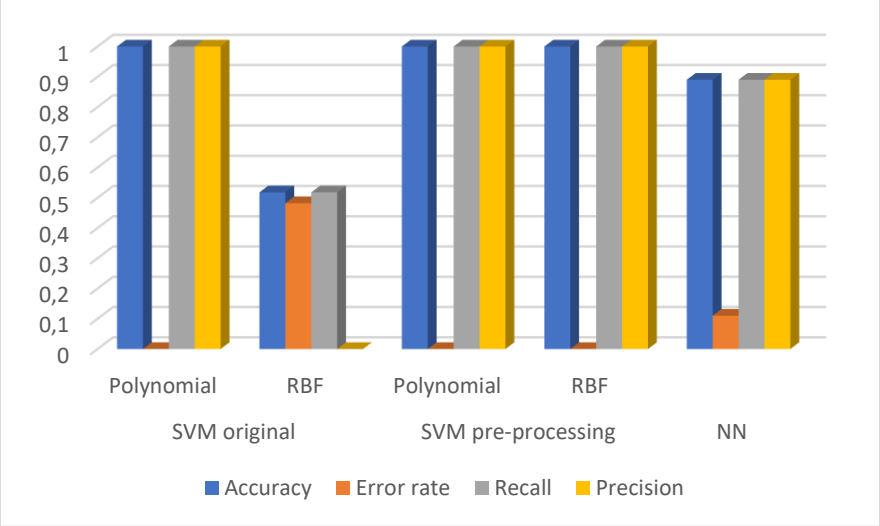


Fig. 6. Performance Polynomial and RBF kernel

(Please add more rows if needed)

Reviewer # 4

Final Recommendation Please tick	**Accepted without modification**	**Accepted with minor corrections**	**Accepted with major modification**	**Rejected**
	☐	☐	☐	☐

Comments	**Addressed (Y/N)**	**Reply/Action taken**
• **Title:** Should be deleted abbreviation		**CORONARY HEART DISEASE USING SUPPORT VECTOR MACHINE**

SVM and my suggestion is your title should be "Coronary Heart Disease using Support Vector Machine" or if you use other classifier you can mention it.		
• Abstract: Need to explain problem background, why you are interested in making research, what is your motivation.		**Abstract** The preference of SVM kernel function with optimal features that flexibly applied for dynamic dataset is a new challenge. The restriction of technology and infrastructure support for diagnosing the bioinformatics at rural area is a major concern for developing countries towards excellent health services. Therefore, this study aimed at evaluating the utilization of Support Vector Machine (SVM) in classifying patients of coronary heart disease with Unstable Angina Pectoris (UAP), Non-Segment (ST) Elevation Myocardial Infarction (NSTEMI) and ST-Elevation Myocardial Infarction (STEMI) classes. So far, 280 samples were experimented with 17 attributes by considering four types of dataset, which include the original, reduced, pre-processing and K-Nearest Neighbours (k-NN). To evaluate the optimal parameter pairs in terms of accuracy and processing time for the above dataset types, 10-folds cross-validation and percentage split were carried out on Polynomial and Radial Basis Function (RBF) kernels. Waikato Environment for Knowledge Analysis (WEKA) tool for 10-folds reveals the optimum accuracy of 100% for polynomial kernel and 98.9% for RBF. Also, the percentage split of 70:30 affirms 100% accuracy with 0.06 seconds of processing time as the ideal values of Polynomial kernel test. Meanwhile, RBF exhibits 80:20 split for 100% accuracy with 0.08 seconds in dataset pre-processing. In a nutshell, SVM enhances the data precision and recall as well as minimizes the error possibility for the greatest classification of coronary heart disease patients in Polynomial and RBF kernel than other classifier such as Neural Network (NN). Therefore, the application of SVM improves the accuracy of coronary heart disease diagnostics. **Keywords**: Neural Network, K-Nearest Neighbours, Data Mining, Support Vector Machine, Coronary Heart Diseases.
• Keywords: Should be not more than five keywords.		Neural Network, K-Nearest Neighbours, Data Mining, Support Vector Machine, Coronary Heart Diseases.
Introduction, Research Method, Results and Discussion: (1) • In paragraph one, separate this paragraph become two paragraphs.		**1. Introduction** Data mining provides various manipulation services to achieve the prediction, classification, clustering, mapping, and anomalous detection of data. The utilization of this technique in various disciplines has evolved and shown a significant contribution to the field of knowledge, including medicine, finance, industry, technology, and even molecular biology as well as bioinformatics. With an emphasis on classification, the advent of methods in disaggregation data improves its usefulness and maneuverability in interpreting information, for examples Nijssen and Fromont [1] studied the optimal constraint of Decision tree method induction in pattern mining; Network and Tree-based methods were applied for data mining modeling in the corrosion of concrete sewer [2]; k-NN for scholarship recipient cases [3]; Multilayer Perceptron (MPL) for data mining in healthcare operations [4]; Naive Bayes approach in classifying the analysis of students' performance [5], Artificial Neural Network as a validation tool of Loud Haul Dump (LHD) machine performance characteristics [6], Neural Network in designating the water cycle problems [7] and the utilization of SVM in data mining [8].

		Recently, the enforcement of the above methods in analysing the complex bioinformatics data was put into practice. Big data opportunities bring unprecedented potential and challenges in data mining and biological analysis systems in a cost-efficient manner [9]. Also, big data technology ensures that the biologist generates large amount of facts and measurement of genomic sequences, images of physiological structures, measuring the messenger Ribonucleic Acid (mRNA) and protein expression, transcription factor binding, and metabolite concentration with limitation of programming skills [10]. In addition, Majhi et al [11] utilized bioinformatics techniques to identify the early stages of diseases such as metabolic and urea cycle disorders, inborn errors and path-aligners through genetics analysing processes and proteomics reports, which are therefore compared with health care data. Furthermore, Dashtban and Balafar [12] found the significance of data mining as artificial intelligent tools in classifying the microarray cancer data. The adoption of machine learning algorithms in bioinformatics accomplished the reduction of complex data and allocated the feature selection of biomarkers in raw data. Serra et al [13] verified the successful employment of machine learning techniques as well as clustering, classification, embedding techniques and network-based approaches in addressing bioinformatics problems which include gene expression clustering, patient classification, brain network analysis, and identification of biomarkers. In addition, this technology's ability to capture biomedical data has reformed machine learning into a sophisticated way to solve the complexities of big data. The number of heterogeneity modalities in biological and neurobiological phenomena insists on the multi-view of intelligent data integration from several resources. In addition, multi-view learning and data integration offers greater statistical power analysis [14]. In the process of improvising classification parameters, especially in predicting bioinformatics data, a high level of precision is required to produce the best and most effective classifier tool. The classification techniques that involve data mining, as well as machine learning, reduce computational time and improve categorization precision in determining the optimum values as clarifying in the case of unknown protein sequence classification [15].
• (2) • Need to use standard stages in your research. Step 1 is pre-processing. This step is to reduce data, so there is no missing value. Step 2 is Feature Extraction. This step is to produce seventeen attributes and using k-NN (formula Equation 1 is		Following the employment of SVM in KDD [34]. Step 1: Pre-processing This step is to reduce data, therefore there is no missing value. The activity begins with data selection from CHD and then performed as an effort to feature subset selection by ignoring the irrelevant attributes CHD risk factors and missing values. In view of this, k-NN with the Euclidian distance calculation is performed in Eq. (1) $dist = \sqrt{\sum_{k=1}^{n}(pk - qk)^2}$ (1) where *n* is number of attributes, *pk* and *qk* values are the *-k* attribute. Step 2: Transformation This step is to produce seventeen attributes and using k-NN and it is driven by discretizing the attributes with an equal width approach. The Equal width is one of the unsupervised discretizations of continuous features to obtain a better precision rate in dealing with data manipulation with high cardinality attributes [35] and its outputs become an input to the classification. Step 3: Classification using SVM Subsequently, the core process of data mining, which is the one-against-one SVM multiclass method is defined with a value of *d*, sigma *σ*, and *C* as explained in Table 2. **Table 2. The Define of SVM Value** **d** | **sigma (σ)** | **C** **1** | 1 | 0.01 **2** | 2 | 0.02

Euclidean Distance). Step 3 is Classification. Outputs of the feature extraction become input to the classification. This step using Support Vector Machines (SVM) and two kernels uses namely, Radial Basis Function (RBF) and Polynomial. Step 1 in your article is called Pre-processing, Step 2 is called transformation data and Step 3 is called testing using SVM.		

3	3	0.03
4	4	0.04
5	5	0.05
		0.06
		0.07
		0.08
		0.09
		0.1
		0.2
		0.3
		0.4
		0.5
		0.6
		0.7
		0.6
		0.9
		1

The variable *d* is specified as the degree of the polynomial, the value of *C* is a constant that allows to trade off the influence of the higher and lower-order terms and this is a consideration for varying *C* values between 0.01 and 1. The selection values of *d*, and σ impact the performance accuracy, while *C* is selected based on the *C* function as a constraint, therefore, a greater value of *C* implies more penalty for classification errors. Meanwhile, the values of σ provide a good fit or an overfit to the data, when σ is large compared to the distance between the classes, it results in an overly flat discriminant surface. However, a smaller σ value compared to the distance between classes result in an over-fit [36]. A good choice for σ will be comparable to the distance between the closest members of the two classes. Furthermore, the highest accuracy of parameter pairs during the training session was found at *C and* σ for kernel RBF as well as *C and d* for the polynomial kernel. To process the data, WEKA 3.7.10, which is a powerful tool in data mining [37] and machine learning [38] was adopted.

Step 4: Evaluation using SVM

The evaluation process was carried out to ensure the performance of the classification methods in the SVM with two kernel trick types on polynomial and RBF. The value of accuracy and time in the building model is thoroughly investigated to achieve the superlative one. Also, the 10-folds validation and confusion matrix with percentage splits on the portion of training data compare to test data in 40:60, 50:50, 60:40, 70:30, and 80:20 is applied to support the assessment process. However, there are no specific rules in the distribution of training-data and test-data, therefore, a large number of the former will represent the diversity of the data [39]. Furthermore, to calibrate the testing procedure and the overcoming of various issues related to percentage splits in defining the best *C* and parameter values, 10-folds validation was exploited. Also, the test simulation took place in four stages, viz the original dataset (with missing values), the reduced (no missing values), the k-NN (with Euclidian distance calculation), and the Pre-processing (with KDD formation). Therefore, the success rate of classification, the determination of accuracy, error rate, precision, and recall values are performed based on the confusion matrix as depicted in Eq. (2)-(5) [40] given by,

$$\text{Accuracy} = \frac{TP+TN}{P+N} \times 100\% \quad (2)$$

$$\text{Error-rate} = \frac{FP+FN}{P+N} \times 100\% \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

		TP *(True Positive)* = The amount of correctly classified data (*Actual class (yes), Predicted class (yes)).* TN *(True Negative)* = The amount of correctly classified data (*Actual class (no), Predicted class (no)).* FN *(False Negative)* = The amount of incorrectly classified data (*Actual class (yes), Predicted class (no)).* FP *(False Positive)* = The amount of incorrectly classified data (*Actual class (no), Predicted class (yes)).* P = Total of TP and FN N = Total of FP and TN
• (3) • Organization of the paper need to explain at the end of paragraph in part of introduction.		The organization of this study begins with an introduction that explains the background, previous reviews on the SVM method, the objectives, the research work, and implications. Furthermore, detailed data, instruments, and step processes are elucidated in the research method. The output of Knowledge Discovery and Data mining (KDD) and SVM analysis as well as and SVM evaluation are deliberated in the research result and discussion. Finally, the conclusion is given as a resume and suggestion is made for future studies.
• (4) • Check format of citation, especially how to cite research papers. For example: [16 and 17] or [16-17] or [16, 17]?		For two citations: [16, 17], more than two citation [27-33]. Have been checked. SVM is a classification method that produces a fairly high degree of accuracy and is commonly used compared with the conventional decision tree, ANN [16, 17] To scrutinize the performance of SVM with other classifiers, the calculation of confusion matrix in NN Multilayer perceptron is measured by considering the values of accuracy, error rate, precision, and recall. This was adopted because research has steadily built on the accuracy and efficiency of data mining using NN and SVM for medical prediction and classification tasks [41, 42]. NN methods were extensively adopted in classifying problems and as one of the most active research and application areas. Furthermore, SVM and NN have been used with high accuracy in classification with relatively small sample data [43, 44].
• (5) • Every equation, make the label of number and mention Equation 1, Equation 2 and etc.		Have been done the correction. Its refers to Jestec format template. In view of this, k-NN with the Euclidian distance calculation is performed in Eq. (1) $dist = \sqrt{\sum_{k=1}^{n}(pk - qk)^2}$ (1) where *n* is number of attributes, *pk* and *qk* values are the *-k* attribute. ---- Therefore, the success rate of classification, the determination of accuracy, error rate, precision, and recall values are performed based on the confusion matrix as depicted in Eq. (2)-(5) [40] given by, $\text{Accuracy} = \frac{TP+TN}{P+N} \times 100\%$ (2) $\text{Error-rate} = \frac{FP+FN}{P+N} \times 100\%$ (3) $\text{Precision} = \frac{TP}{TP+FP}$ (4) $\text{Recall} = \frac{TP}{TP+FN}$ (5)

		----- Consequently, the discretization with the equal width approach was applied in expressing the standard range values from 0 to 1 as in Eq. (6). Series of range= the highest area – the lowest area / The number of categories (6)
• (6) • Need to explain format input of SVM.		The medical records of CHD patients were collected in a variety of formats. Consequently, the discretization with the equal width approach was applied in expressing the standard range values from 0 to 1 as in Eq. (6). Series of range= the highest area – the lowest area / The number of categories (6) The discretization of attributes is depicted in Table 4 and Table 5. Table 4 defines the values of attribute 1 for age discretization, attribute 9 for systolic blood pressure (BP), attribute 10 for diastolic blood pressure, attribute 11 for LDL, attribute 12 for HDL, attribute 13 for Total cholesterol, attribute 14 for Triglyceride, and attribute 15 for a glucose level. The rest of the attributes (2,3,4,5,6,7,8,16, and 17) were categorized into two series and discretized into 0 value for "No" and 1 for "Yes" as shown in Table 5. This discretization value will be the format for SVM input. The sample of format SVM input is described in Figure 2.

Table 4. Attribute Discretization

Age discretization (1)		**Systolic TD discretization (Sis) (9)**	
Age (years)	**Discretization**	**Systolic BP (mmHg)**	**Discretization**
25 ≤ U <35	0	Sis<120	Optimal (0)
35 ≤ U <45	0.2	120< Sis <130	Normal (0.2)
45 ≤ U <55	0.4	130< Sis <140	Normal Height (0.4)
55 ≤ U <65	0.6	140< Sis <150	Low hypertension (0.6)
65 ≤ U <75	0.8	150< Sis <160	Moderate hypertension (0.8)
U ≥ 85	1	Sis >160	Severe hypertension (1)
Diastolic TD (Dias) discretization (10)		**Discretization of LDL (LDL) levels (11)**	
Diastolic BP (mmHg)	**Discretization**	**LDL levels (mg / dL)**	**Discretization**
Dias<80	Optimal (0)	LDL<100	Optimal (0)
80≤Dias<85	Normal (0.2)	100< LDL <130	Approaching optimal (0.25)
85≤Dias<90	Normal Height (0.4)	130< LDL <160	Borderline high (0.5)
90≤Dias<100	Low hypertension (0.6)	160< LDL <190	High (0.75)
100≤Dias<110	Moderate hypertension (0.8)	LDL >190	Very high (1)
Dias≥110	Severe hypertension (1)		
Discretization of HDL(HDL) (12)		**Discretization of total cholesterol (Chol)(13)**	
HDL levels (mg / dL)	**Discretization**	**Chol levels (mg / dL)**	**Discretization**

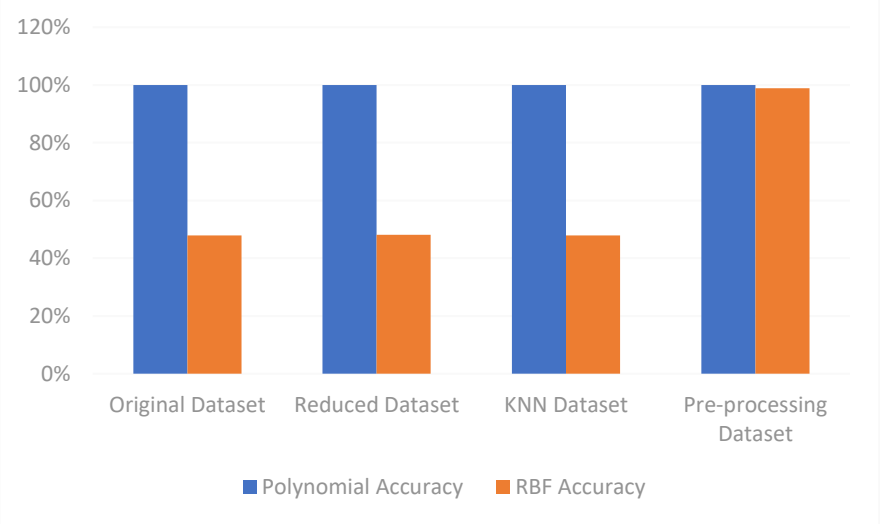
HDL<40	Low (0)	Chol <200	Desirable (expected to be safe) (0)
40≤ HDL <60	Normal (0.5)	200≤ Chol <240	Borderline (must be aware- begin to control) (0.5)
HDL ≥60	High (1)	Chol ≥240	High (1)

Triglyceride discretization (14)		**Glucose Level discretization (Glu) (15)**	
Triglyceride levels (mg / dL)	**Discretization**	**Glucose Levels (mg/dL)**	**Discretization**
trig <150	Normal (0)	Glu<40	Optimal (0)
150≤ trig <200	Borderline high (0.33)	40≤ Glu <60	Normal (0.2)
200≤ trig <500	High (0.66)	60≤ Glu <125	Normal Height (0.4)
trig ≥500	Very High (1)	125≤ Glu <145	Low hypertension (0.6)
		145≤ Glu <200	Moderate hypertension (0.8)
		Glu ≥200	Severe hypertension (1)

Table 5. Attributes with two series discretization

Attributes	**Discretization**	
Gender (2)	Male	1
	Female	0
Family History (3)	None	0
	Yes	1
Heart History (4)	None	0
	Yes	1
DM History (5)	None	0
	Yes	1
Hypertension History (6)	None	0
	Yes	1
Cholesterol History (7)	None	0
	Yes	1
Obesity (8)	None	0
	Yes	1
Elevation (16)	None	0
	Yes	1
Cardiac Enzymes (17)	None	0
	Yes	1

	Attributes discretization																
No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Cases
1	0,4	1	0	0	0	0	0	0	0,4	0,2	0	0	0	0	0,4	0	0	UAP
2	0,4	1	0	0	0	0	0	0	0,8	0,8	0	0	0	0	0,4	0	0	UAP
3	0,4	0	0	1	0	0	0	0	0,2	0,2	0	0,5	0	0	0,4	0	0	UAP
4	0,6	1	0	1	0	0	0	0	0,2	0,2	0,25	0,5	0	0	0,8	0	0	UAP
5	0,6	1	0	0	1	0	0	0	1	0,2	0,25	0,5	0	0	0,4	0	0	UAP
6	0,6	1	0	1	1	1	1	0	0,2	0,2	0,5	1	0,5	0	0,4	0	1	NSTEMI
7	0,6	1	0	0	0	0	0	0	0,6	0,2	0	0	0	0,33	0,8	0	0	UAP
8	0,4	0	0	1	0	0	0	0	0,6	0,8	0,25	1	0,5	0	0,4	0	0	UAP
9	0,6	1	0	0	1	1	0	0	0,6	0,2	0,75	0	0,5	0	0,4	0	0	UAP
10	0,4	0	0	1	1	1	0	0	0,4	0,2	0,25	0,5	0,5	0,66	0,4	0	0	UAP
11	0,2	1	0	1	1	1	0	0	0,2	0	0,25	0	0	0	0,4	0	0	UAP
12	0,6	1	0	1	1	1	0	0	1	0,8	0,25	0	0	0,33	0,4	0	0	UAP
13	0,4	0	0	0	0	0	0	0	0	0	1	0	1	0	0,4	0	0	UAP
14	0,4	1	0	1	1	1	0	0	0,6	0,6	0	0	0	0	0,4	0	0	UAP
15	0,4	1	0	1	1	1	0	0	0,6	0,8	0,5	0	0,5	0,66	0,4	0	0	UAP
16	0,4	0	0	0	1	1	0	0	0,6	0,8	0,25	0	0	0	0,8	0	1	NSTEMI
17	0,8	0	0	0	0	0	0	0	1	0,8	0,25	0	0,5	0	0,8	1	1	STEMI
18	0,4	0	1	0	1	1	0	0	0,6	0,6	0	0	0	0	0,4	1	1	STEMI
19	0,6	1	0	0	0	0	0	0	0	0	1	0	1	0	0,4	1	1	STEMI
20	0,4	1	0	1	1	1	0	0	0,4	0,8	0,5	0	0	0	0,4	1	1	STEMI

		Fig. 2. The sample of SVM input
• (7) • In Table 11 and Table 12, what is your justification to produce accuracy for original dataset, reduce dataset, kNN data set? Is not our focused-on pre-processing with do transformation ?		We have changed Table 11 to Table 6 and Table 12 to Table 7. **Table 6. The accuracy of the best parameter - 10-fold cross validation.** **Kernel** | **Polynomial** | | | **RBF** | | | **Parameters** | **C** | **d/σ** | **Accuracy** | **C** | **d/σ** | **Accuracy** **Original Dataset** | 0.03 | 1 | 100% | 0.01 | 1 | 47.9% **Reduced Dataset** | 0.03 | 1 | 100% | 0.01 | 1 | 48.1% **k-NN Dataset** | 0.03 | 1 | 100% | 0.01 | 1 | 47.9% **Pre-processing Dataset** | 0.02 | 2 | 100% | 0. 8 | 1 | 98.9% **Table 7. The accuracy of the best parameter pairs -percentage split.** **Kernel** | **Polynomial** | | | **RBF** | | | **Parameters** | **DC** | **T (s)** | **Accuracy** | **DC** | **T(s)** | **Accuracy** **Original Dataset** | 70:30 | 27.49 | 100% | 40:60 | 0.06 | 49.4% **Reduced Dataset** | 70:30 | 21.37 | 100% | 70:30 | 0.13 | 53.8% **k-NN Dataset** | 80:20 | 24.55 | 100% | 40:60 | 0.08 | 49.4% **Pre-processing Dataset** | 70:30 | 0.06 | 100% | 80:20 | 0.08 | 100% These above tables are used to investigate the implication of pre-processing against to SVM. Therefore, the evaluation is conducted by comparing the accuracy within dataset changes in the data original (without missing values), the reduced data (with missing values), k-NN (with distance calculation), and pre-processing (KDD formatted). These above tables reveal that pre-processing provided a significant growth of accuracy in SVM for polynomial and RBF kernel. The graphical views of performances are shown in Figure 3, 4, and 5.  **Fig 3. Dataset performance based on accuracy - 10-fold cross validation**

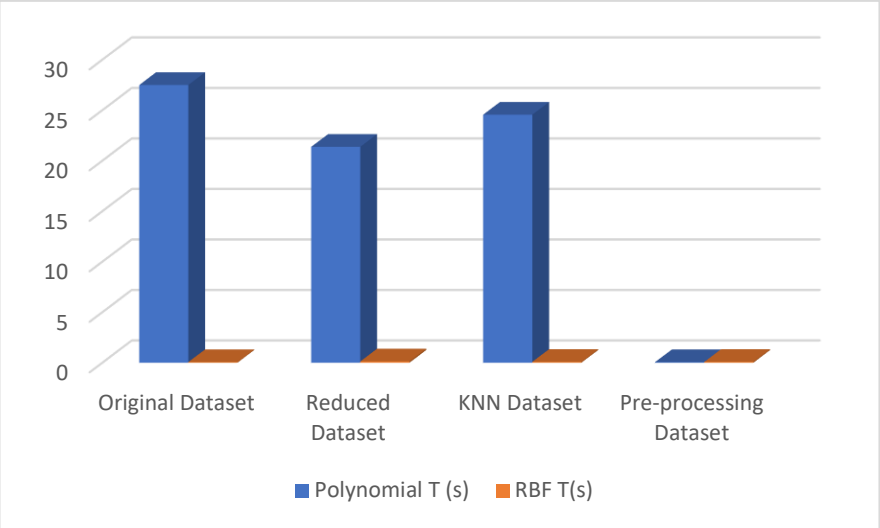


Fig 4. Dataset performance based on time (s) - percentage split

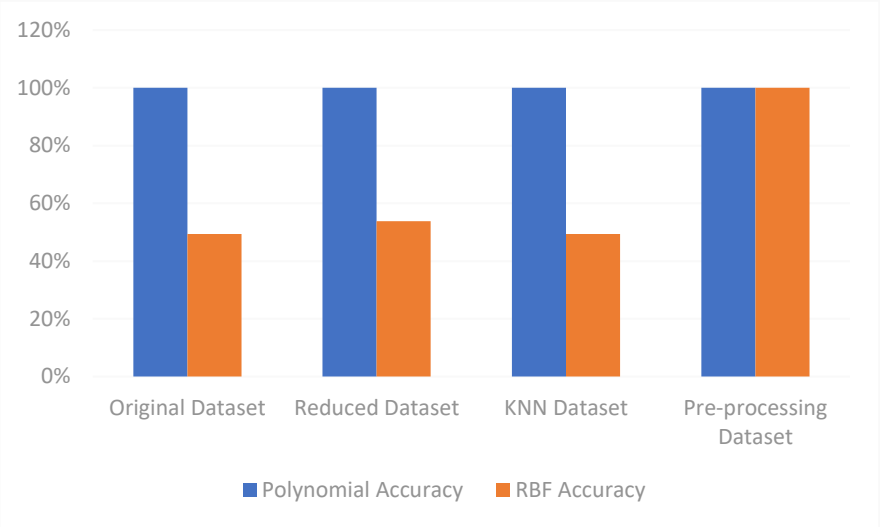


Fig 5. Dataset performance based on accuracy - percentage split

- (8)
- Table 13, why you still use confusion matrix for original dataset again?

Table 13 turned into Table 8. We have eliminated the original dataset in Table 8 and focusing on the pre-processing dataset. Nevertheless, to show the performance of before and after SVM pre-processing, Figure 6 is obtained. The comparison testing analysis between SVM and other classifier, such as NN is also defined.

To evaluate the classification of CHD patient's dataset in SVM, the testing procedure was undertaken according to the Test Option Supplied on the Confusion Matrix formula [39]. The pre-processing dataset was put in place on 20% of tested data at $C = 0.02$ and $d = 2$ in the polynomial kernel and the values of C and σ are 0.8 and 1 respectively, in the RBF. In addition, the resemblance of SVM with another classifier, namely Multilayer perceptron Neural Network (NN) is operated to deeply observe the effectiveness of SVM. The confusion matrix for the above dataset of SVM and NN was explained in Table 8. This table showed that the classification in the pre-processing dataset for SVM is more accurate compared to NN, especially for RBF kernel. By comparing the values for error rate, precision and recall between polynomial kernel and RBF based on the confusion matrix computation as a side of SVM and NN, Figure 6 is obtained. The figure showed that SVM for Polynomial kernel has 100% accuracy, "0" for error rate, and "1" for precision, and recall. Meanwhile, RBF kernel discharged from

51.79% into 100% accuracy, 0.48 into 1 for error rate, undefined into 1 for precision, and 0.52 into 1 for recall. Also, NN for polynomial kernel achieved 89% accuracy, "0.11" for error rate, and "0.89" for precision and recall.

Table 8. Confusion Matrix for SVM and NN-Polynomial and RBF.

SVM: Dataset Pre-processing						
	Polynomial			RBF		
Class	Prediction Class			Prediction Class		
	UAP	NSTEMI	STEMI	UAP	NSTEMI	STEMI
UAP	29	0	0	29	0	0
NSTEMI	0	13	0	0	13	0
STEMI	0	0	14	0	0	14
Accuracy	100%			100%		
Error rate	0			0		
Precision	1			1		
Recall	1			1		

NN-Multilayer Perceptron			
Class	Prediction Class		
	UAP	NSTEMI	STEMI
UAP	18	0	1
NSTEMI	2	18	1
STEMI	0	2	14
Accuracy	89%		
Error rate	0.11		
Precision	0.89		
Recall	0.89		

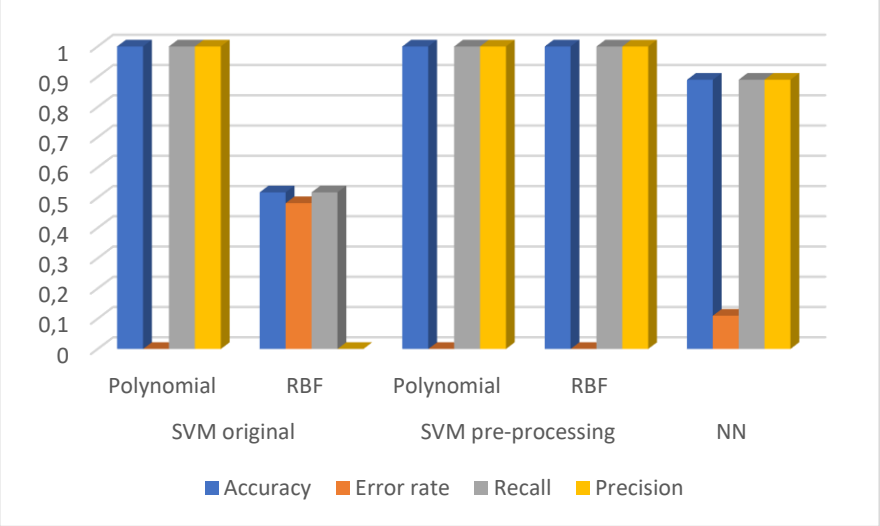


Fig. 6. Performance Polynomial and RBF kernel

- (9)
- Performance measurements are accuracy and processing time. Currently,

We have added error rate, precision, and recall for other performance measurement. The explanation in the text as follows.

To evaluate the classification of CHD patient's dataset in SVM, the testing procedure was undertaken according to the Test Option Supplied on the Confusion Matrix formula [39]. The pre-processing dataset was put in place on 20% of tested data at $C = 0.02$ and $d = 2$ in the polynomial kernel and the values of C and σ are 0.8 and 1 respectively, in the RBF. In addition, the resemblance of SVM with another classifier, namely Multilayer perceptron Neural Network (NN) is operated to deeply observe the effectiveness of SVM. The confusion matrix for the above dataset of SVM and NN was

processing time is not employed for the criteria measurement. If you use high processing, automatically will give better result compared to low processing. In addition, you can add other performance measurements such as error rate, sensitivity and specificity.

explained in Table 8. This table showed that the classification in the pre-processing dataset for SVM is more accurate compared to NN, especially for RBF kernel. By comparing the values for error rate, precision and recall between polynomial kernel and RBF based on the confusion matrix computation as a side of SVM and NN, Figure 6 is obtained. The figure showed that SVM for Polynomial kernel has 100% accuracy, "0" for error rate, and "1" for precision, and recall. Meanwhile, RBF kernel discharged from 51.79% into 100% accuracy, 0.48 into 1 for error rate, undefined into 1 for precision, and 0.52 into 1 for recall. Also, NN for polynomial kernel achieved 89% accuracy, "0.11" for error rate, and "0.89" for precision and recall.

Table 8. Confusion Matrix for SVM and NN-Polynomial and RBF.

SVM: Dataset Pre-processing						
	Polynomial			RBF		
Class	Prediction Class			Prediction Class		
	UAP	NSTEMI	STEMI	UAP	NSTEMI	STEMI
UAP	29	0	0	29	0	0
NSTEMI	0	13	0	0	13	0
STEMI	0	0	14	0	0	14
Accuracy	100%			100%		
Error rate	0			0		
Precision	1			1		
Recall	1			1		

NN-Multilayer Perceptron			
Class	Prediction Class		
	UAP	NSTEMI	STEMI
UAP	18	0	1
NSTEMI	2	18	1
STEMI	0	2	14
Accuracy	89%		
Error rate	0.11		
Precision	0.89		
Recall	0.89		

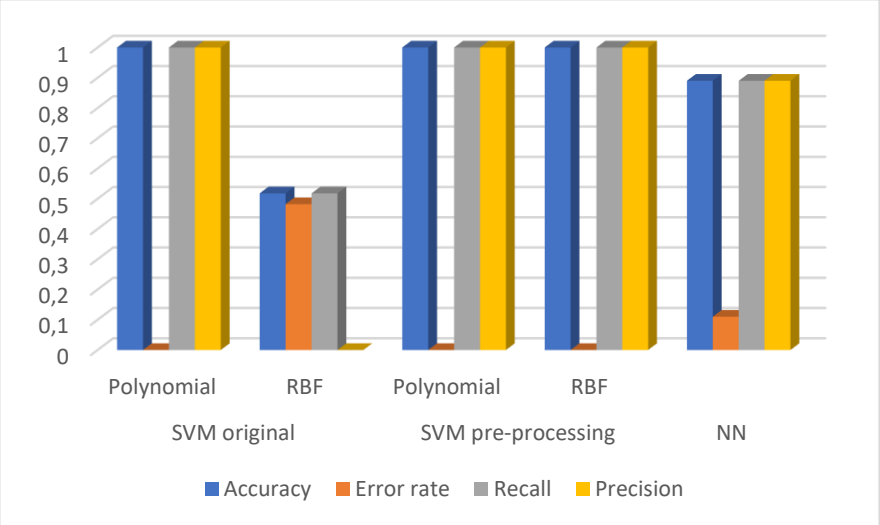


Fig. 6. Performance Polynomial and RBF kernel

- (10)

We did not use a standard database. It is good advice for the future work

• If you use standard databases, you can compare your result with other researchers and you can make the analysis. The output of results, is better use grant chart.		
• (11) • The contribution of paper is very low, because only using one classifier which is SVM. Use another classifier such as Artificial Neural Network, Fuzzy Logic, Ant Colony Optimization and etc.		This research reveals the optimum accuracy of 100% for polynomial kernel and 98.9% for RBF. Also, SVM provides significant values on the accuracy, error rate, precision, and recall, even though it exceeds NN capacity. If we compare to previous researches (which provide the accuracy of SVM just in 95% [19], and 96.86% [20]), the values of SVM in this case has been increase and reach better performance. Even SVM provides the opportunities in enhancing the accuracy by combining with another optimization algorithm such as ANN, Fuzzy logic, Ant Colony, and etc, the increasing is not too significant. Nevertheless, the trial of SVM hybrid can be suggested for the future work. It is explained in conclusion. The explanation in the text can be seen in Discussion and Conclusion. **3.3. Discussion** This result reveals that the pre-processing dataset in SVM provides significant values on the accuracy, error rate, precision, and recall, even though it exceeds NN capacity. As studied by [42], the SVM approach gives better predictive capability than other models, including NN. This, of course, has far-reaching implications in the medical context that require increasing sensitivity, specificity (the ability to predict the absence of the condition when it is not present) as well as discriminatory power of the classifier as key features to consider when comparing classifiers and diagnostic methods [45]. In the reviews on kernel type, the simulation presented that SVM polynomial is more reliable on the dataset changes compare to RBF. Consequently, the pre-processing prescription on SVM-RBF will undoubtedly boost RBF performance. Furthermore, selecting the specific kernel is an important research issue for kernel-based learning in the data mining area and the problem of SVM kernels is found in fitting the appropriate parameter values [46]. This investigation revealed that the SVM polynomial kernel mediates the accuracy and efficiency of the diagnostic results based on the parameters defined in CHD. **4. Conclusion** This study successfully employed the SVM method in classifying the CHD patient's dataset. The simulation of the original, reduced, k-NN, and pre-processing datasets have shown the potential differences between polynomial and RBF kernel in terms of accuracy and processing time. The analysis of 10-folds cross-validation and percentage splits revealed the optimal pairs of parameters and data composition for polynomial and RBF kernel. Furthermore, the confusion matrix presented evidence that the pre-processing dataset delivered greater values in accuracy, precision, error rate, recall, and time model consumption than others. A comparative analysis between SVM and NN has shown the efficiency and accuracy of SVM in accelerating the unsurpassed

		classification of the CHD dataset with minimal errors. Therefore, this classification practically aids the doctors in suggesting medical assistance and taking a curative action. This result methodically answered the difficulties in choosing the SVM kernel function, which is flexible in changing the data set, optimal functionality, and time-consumption with high performance. Nevertheless, integrating SVM with other methods is a new solution to increase SVM performance for future work.
• (12) • References: ok		OK

(Please add more rows if needed)

Reviewer # 5

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
•		
•		
•		
•		

(Please add more rows if needed)

Reviewer # 6

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
•		
•		
•		
•		

(Please add more rows if needed)

Reviewer # 7

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
•		
•		
•		
•		

(Please add more rows if needed)

Reviewer # 8

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
•		
•		
•		
•		

(Please add more rows if needed)

Reviewer # 9

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
•		
•		
•		
•		

(Please add more rows if needed)

Reviewer # 10

Final Recommendation Please tick	Accepted without modification ☐	Accepted with minor corrections ☐	Accepted with major modification ☐	Rejected ☐

Comments	Addressed (Y/N)	Reply/Action taken
•		
•		
•		
•		

(Please add more rows if needed)