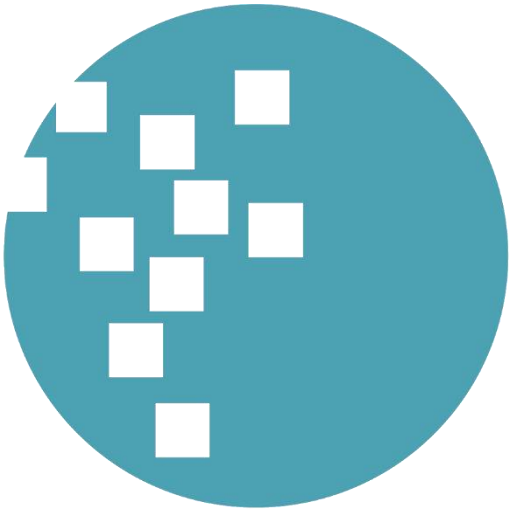
ISSN 2355-0082

IJNMT

International Journal of New Media Technology

VOLUME 11 No. 2



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2355-0082

EDITORIAL BOARD

EDITORIAL ADDRESS

Universitas Multimedia Nusantara (UMN)
Jl. Scientia Boulevard
Gading Serpong
Tangerang, Banten - 15811
Indonesia
Phone. (021) 5422 0808
Fax. (021) 5422 0800
Email : ultimacomputing@umn.ac.id

IJNMT (International Journal of New Media Technology) is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published regularly twice a year (June and December) by Faculty of Engineering and Informatics, Universitas Multimedia Nusantara in cooperation with UMN Press.

Call for Papers

International Journal of New Media Technology (IJNMT) is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published twice a year by Faculty of Engineering and Informatics of Universitas Multimedia Nusantara in cooperation with UMN Press.

Ultimatics : Jurnal Teknik Informatika is the Journal of the Informatics Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the fields of Analysis and Design of Algorithm, Software Engineering, System and Network security, as well as the latest theoretical and practical issues, including Ubiquitous and Mobile Computing, Artificial Intelligence and Machine Learning, Algorithm Theory, World Wide Web, Cryptography, as well as other topics in the field of Informatics.

Ultima Computing : Jurnal Sistem Komputer is a Journal of Computer Engineering Study Program, Universitas Multimedia Nusantara which presents scientific research articles in the field of Computer Engineering and Electrical Engineering as well as current theoretical and practical issues, including Edge Computing, Internet-of-Things, Embedded Systems, Robotics, Control System, Network and Communication, System Integration, as well as other topics in the field of Computer Engineering and Electrical Engineering.

Ultima InfoSys : Jurnal Ilmu Sistem Informasi is a Journal of Information Systems Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the field of Information Systems, as well as the latest theoretical and practical issues, including database systems, management information systems, system analysis and development, system project management information, programming, mobile information system, and other topics related to Information Systems.

ISSN 2355-0082

FOREWORD

Greetings!

IJNMT (International Journal of New Media Technology) is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published regularly twice a year (June and December) by Faculty of Engineering and Informatics, Universitas Multimedia Nusantara in cooperation with UMN Press.

In this June 2024 edition, IJNMT enters the 1st Edition of Volume 11 No 1. In this edition there are ten scientific papers from researchers, academics and practitioners in the fields covered by IJNMT. Some of the topics raised in this journal are: The Effect of Using the Al-Mumtaz Application on Student Learning Outcomes UIN Mahmud Yunus Batusangkar, The Effect of Video Games Towards the Students' Academic Performance, Approach Convolutional Neural Network LeNet-5 for Interactive Learning of Korean Syllables (Hangul), Ensemble Learning - Random Forest Algorithm to Classify Obesity Level, Implementation Of Heuristic Evaluation Method For Evaluation And Recommendations UI/UX Design Improvements On The Cinepolis Website, Data Quality Issues : Case Study of Claim and Insured in Indonesia Insurance Company, Evaluating the Impact of Particle Swarm Optimization Based Feature Selection on Support Vector Machine Performance in Coral Reef Health Classification, Enhancing Support Vector Machine Classification of Nutrient Deficiency in Rice Plants Through Particle Swarm Optimization-Based Feature Selection, Cross-Platform Mobile Based Crowdsourcing Application for Sentiment Labeling Using Gamification Method, Cost Estimation for Software Development Using Function Point Analysis Method.

On this occasion we would also like to invite the participation of our dear readers, researchers, academics, and practitioners, in the field of Engineering and Informatics, to submit quality scientific papers to: International Journal of New Media Technology (IJNMT), Ultimatics : Jurnal Teknik Informatics, Ultima Infosys: Journal of Information Systems and Ultima Computing: Journal of Computer Systems. Information regarding writing guidelines and templates, as well as other related information can be obtained through the email address ultimaijnmt@umn.ac.id and the web page of our Journal here.

Finally, we would like to thank all contributors to this June 2024 Edition of IJNMT. We hope that scientific articles from research in this journal can be useful and contribute to the development of research and science in Indonesia.

December 2024,

Alexander Waworuntu, S.Kom., M.T.I.
Editor-in-Chief

ISSN 2355-0082

TABLE OF CONTENT

ISSN 2355-0082

The Effect of Using the Al-Mumtaz Application on Student Learning Outcomes UIN Mahmud Yunus Batusangkar

Rahmat Hidayat[1], Munirul Abidin[2], Danial Hilmi[3]
[1,2,3] State Islamic University of Maulana Malik Ibrahim Malang, Indonesia
[1] Rrahmad667@gmail.com, [2] munirul@bio.uin-malang.ac.id, [3] Danialhilmi@uin-malang.ac.id

Accepted 18 May 2024
Approved 15 November 2024

***Abstract*— In today's digital era, the utilization of technology in the learning process is becoming increasingly important. Online learning applications such as Al-Mumtaz have the potential to overcome the limitations of conventional learning methods and increase student interest and motivation to learn. However, the effect of using the Al-Mumtaz application on student learning outcomes at Mahmud Yunus State Islamic University Batusangkar has not been widely studied. This study aims to identify and analyze the effect of using the Al-Mumtaz application on student learning outcomes at UIN Mahmud Yunus Batusangkar and provide evidence-based recommendations regarding the effectiveness of digital learning applications. This study used a quantitative pre-experiment design with a sample of Arabic language education students who used the Al-Mumtaz application. Data were collected through initial and final test, and analyzed using validity, reliability, normality, homogeneity, and paired t tests. The analysis showed that 10 out of 15 questions were valid and reliable. The data were normally distributed and met the assumption of homogeneity of variance. The paired t-test revealed a significant mean difference between the pre-test (62.50) and post-test (89.00) scores after the use of Al-Mumtaz application, with an increase of 23.50.The use of Al-Mumtaz application in Arabic language learning at UIN Mahmud Yunus Batusangkar provides significant results.This study confirms the positive effect of Al-Mumtaz application on student learning outcomes and highlights the importance of technology in improving the effectiveness of learning in the academic environment.**

***Index Terms*— Al-Mumtaz App; Learning Outcomes; Students of UIN Mahmud Yunus Batusangkar**

I. INTRODUCTION

In the rapidly growing digital era, information and communication technology (ICT) has become an inseparable part of human life. One form of ICT implementation in education is the use of digital learning applications [1], [2]. Al-Mumtaz application is one of the online learning platforms developed by Mahmud Yunus State Islamic University (UIN) Batusangkar. This application is designed using Kodular web to facilitate the learning process and assist students in improving their learning outcomes [3].

A problem often faced in the learning process is the lack of effectiveness of conventional learning methods. Traditional learning methods that rely on face-to-face meetings and lectures are often less attractive to students, especially for millennials who are accustomed to digital technology [4]. This can lead to a lack of interest and motivation to learn, as well as decreased student learning outcomes.

In addition, the limited time and resources owned by lecturers and educational institutions are also a problem that must be faced. Lecturers often find it difficult to provide learning materials that are up-to-date and relevant to student needs, while educational institutions are faced with the challenge of providing adequate facilities and infrastructure to support an effective learning process.

This study aims to analyze the effect of using the Al-Mumtaz application on the learning outcomes of UIN Mahmud Yunus Batusangkar students. By using this online learning application, it is expected to overcome the problems faced in the conventional learning process. Al-Mumtaz application can be an effective solution in increasing student interest and motivation to learn, as well as facilitating more interactive and interesting learning.

The discussion of the effect of using the Al-Mumtaz application on the learning outcomes of UIN Mahmud Yunus Batusangkar students is very important to do. This is because the results of this study can provide an overview of the effectiveness of online learning applications in improving student learning outcomes. In addition, this research can also be a reference for other educational institutions in developing and implementing online learning applications that are in accordance with student needs and the latest technological trends.

By utilizing the Al-Mumtaz application, it is expected to overcome the problems faced in the conventional learning process. This application can provide up-to-date, interactive, and interesting learning materials for students. In addition, this application can also facilitate flexible learning and is not limited by

time and place, so that students can learn anytime and anywhere according to their needs.

The rapid development of information and communication technology (ICT) has brought significant changes in various aspects of life, including in education [5]. Educational institutions are required to be able to follow this development and utilize technology to improve the quality of learning. One of the efforts made by Mahmud Yunus State Islamic University (UIN) Batusangkar is to develop an online learning application called Al-Mumtaz.

Research on the effect of using the Al-Mumtaz application on student learning outcomes at UIN Mahmud Yunus Batusangkar is important to do. The main reason underlying this research is the gap between the needs of students in accessing more interactive and interesting learning resources with conventional learning methods that are still widely used in educational institutions [6]. Traditional learning methods are often considered ineffective in increasing students' interest and motivation to learn, especially for millennials who are accustomed to digital technology.

This study contributes to filling this gap by evaluating the effectiveness of Al-Mumtaz in improving student learning outcomes. Online learning applications such as Al-Mumtaz have the potential to be a solution to overcome the limitations of conventional learning methods. By utilizing digital technology, this application can provide learning materials that are more interactive, interesting, and easily accessible to students.

Thus, research on the effect of using the Al-Mumtaz application on the learning outcomes of UIN Mahmud Yunus Batusangkar students is important to do. This research is expected to contribute to filling existing gaps and providing solutions in improving the quality of learning through the use of digital technology. Seeing from the introduction and the problems above, the researcher wants to ask questions about this research, namely whether there is an effect of using the Al-Mumtaz application on the learning outcomes of Mahmud Yunus Batusangkar State Islamic University students.

The main objective of this study is to identify and analyze the effect of using Al-Mumtaz application on the learning outcomes of UIN Mahmud Yunus Batusangkar students. This study aims to assess whether the implementation of educational technology such as the Al-Mumtaz app can effectively improve students' concept understanding, learning motivation, and ultimately learning outcomes. In addition, this study also aims to provide evidence-based recommendations to the university regarding the effectiveness of digital learning applications and learning strategies that can be integrated to maximize the educational potential of these digital tools in the academic environment.

II. METHOD

This study uses a quantitative Pre-x periment design to measure the effect of using the Al-Mumtaz application on student learning outcomes at Mahmud Yunus Batusangkar State Islamic University [7]. This design was chosen because it allows researchers to make objective and systematic measurements of the variables studied, namely the use of the Al-Mumtaz application (independent variable) and student learning outcomes (dependent variable) [8]. The sample of this study consists of selected Arabic language education students who use Al-Mumtaz application in their learning process. Sampling was done using Pre-x periment technique to ensure equal representation of each student, taking into account demographic factors such as semester.

The necessary data were collected using two types of instruments: first, tests in the form of multiple choice exams using google from. second, Field documentation on students of Mahmud Yunus Islamic University Batusangkar. To determine whether there is a significant effect of using the Al-Mumtaz application on student learning outcomes, inferential analysis such as the independent t test is used. The independent t test is used to compare the average learning outcomes between students who use the Al-Mumtaz application and those who do not. All data analysis will be assisted by using statistical software such as SPSS, which allows researchers to conduct more in-depth analysis and produce reliable findingS [9], [10].

III. RESULT AND DISCUSSION

A. Validity Test

This study uses descriptive quantitative methods, the sample used in this study was Arabic language education students in the second semester of Mahmud Yunus State Islamic University Batusangkar. Through the Post test research instrument, namely the Al-Mumtaz application to improve the learning outcomes of Arabic language education students.

The last job is statistical calculation and reporting of results. The test data that has been obtained from respondents is then tabulated into a table that can be filled in all values and a number of data from the response data. Tabulation of this data is made to facilitate subsequent statistical calculations, namely to determine the trend value. Furthermore, the results of the calculations that have been analyzed are outlined in the results of the research discussion.

The basis for decision making in the validity test is: First, if Rhitung> Rtabel, then the question items in the question are declared valid. Second, if the value of Rhitung < Rtabel, then the question item in the question is declared invalid. The following data is obtained from the results of filling out the test questions that have been answered by respondents:

TABLE I. VALIDATION

Question	R-Table	R-Calculate	Description
1.	0,514	0,662	Valid
2.	0,514	0,525	Valid
3.	0,514	0,876	Valid
4.	0,514	0,525	Valid
5.	0,514	0,876	Valid
6.	0,514	0,344	Invalid
7.	0,514	0,025	Invalid
8.	0,514	0,306	Invalid
9.	0,514	0,123	Invalid
10.	0,514	0,876	Valid
11.	0,514	0,662	Valid
12.	0,514	0,525	Valid
13.	0,514	0,876	Valid
14.	0,514	0,525	Valid
15.	0,514	0,306	Invalid

From the results of the validity of the questions that have been tested using SPSS to 20 respondents, it turns out that 10 out of 15 questions are declared valid.

The data presented consists of 15 lines of information regarding the validity of the questions based on the comparison of the R table value with the R count. The R table value given is 0.514, which is the threshold for determining the validity of a question. A question is considered valid if its calculated R value is higher than the R table, and invalid if it is lower. Of the 15 questions analyzed, 10 of them had R values that exceeded R table, so they were categorized as valid. This indicates that the majority of items have a strong correlation with what the test is measuring and are considered suitable for further use.

On the other hand, there were 5 questions that did not meet the validity criteria as their R calculated values were below the R table value. These questions had much lower R values, such as in questions 6 (0.344), 7 (0.025), 8 (0.306), and 9 (0.123), as well as question 15 (0.306) which showed an R value that was quite close to the R table but still below it. These questions require further evaluation or improvement as their low correlation may indicate that they are not effective in measuring the ability or knowledge that the test is supposed to measure. This requires adjustments to the questions to improve their validity, or possibly replacing the questions with ones that have a higher correlation to the overall material being tested.

B. Reliability Test

The reliability test is used to determine the consistency of the measuring instrument, whether the measuring instrument used is reliable and remains consistent if the measurement is repeated. The test used in this study used the Cronbach alpha technique. Instrument reliability is considered reliable if it has a reliability coefficient> Rtabel. This means that the measurement is relatively consistent if the measurement is repeated.

The basics of decision making: First, if Cronbach Alpha is greater than Rtabel = real. Second, if Cronbach Alpha is smaller than Rtabel = not real. The reliability test of the test questions in this study is using the SPSS version 26 Cronbach's alpha program, as shown in the table below:

TABLE II. B. RELIABILITY

Reliability Statistics

Cronbach's Alpha	*N of Items*
.803	15

From the above calculations, it is known that the reliability index value is 0.803> 0.514 so that the research instrument is declared reabel. Thus the test question can be used as a research instrument.

The reliability analysis of the pre-test and post-test questions used to measure the effectiveness of the Al-Mumtaz application in an educational context. In this study, the resulting Cronbach's Alpha was 0.803 out of 15 items tested. This value indicates that the questionnaire has a high level of reliability, so it can be considered consistent in measuring what it is intended to measure.

This high reliability of the questionnaire is important in the context of educational research, as it ensures that the variability in the measured learning outcomes is due more to the effect of using the educational app than to measurement error. The internal consistency shown through Cronbach's Alpha values of more than 0.8 supports the overall validity of the instruments used in the study. Thus, the results obtained can be used as a strong basis for making conclusions about the effect of the Al-Mumtaz application on

student learning outcomes at UIN Mahmud Yunus Batusangkar.

C. Normality Test

The normality test is carried out to find out whether the sample under study is normally or abnormally distributed using the shapiro wilk test. In this normality test, researchers used the SPSS version 25 program with the following data:

TABLE III. NORMALITY

Tests of Normality

Kelompok		Kolmogorov-Smirnova			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Hasil	PreTest	.165	20	.158	.951	20	.379
	PostTest	.178	20	.098	.961	20	.572

a. Lilliefors Significance Correction

The basis for decision making in the normality test is: First if the significant value <0.05 then the data is not normally distributed. Second if the significant value > 0.05 then the research data is normally distributed.

Based on the normality test results presented, we see that both groups of data, PreTest and PostTest, show a normal distribution. This analysis uses two methods, namely Kolmogorov-Smirnov and Shapiro-Wilk. For the PreTest data, the Shapiro-Wilk test results gave a significance value of 0.379, which is much greater than the alpha significance level of 0.05. This indicates that the PreTest data does not deviate significantly from the normal distribution. Something similar is seen in the PostTest data, where the Shapiro-Wilk test gives a significance value of 0.572, which again confirms that the data follows a normal distribution. The fact that both sets of data are normal is very important in the context of advanced statistical analysis, allowing the use of various parametric techniques that require the assumption of normality.

With this normality assumption confirmed, researchers can be more confident in applying statistical analyses that assume normality of data distribution, such as ANOVA or linear regression. The reliability of the conclusions resulting from such analyses also increases because the basic assumptions have been met. Furthermore, researchers may be interested in exploring the differences between PreTest and PostTest scores using parametric t-tests that compare the means of two independent or paired groups, depending on the study design. This analysis may provide useful insights into the effectiveness or impact of the intervention implemented between the two testing times.

D. Homogeneity Test

The homogeneity test is a test conducted to determine that two or more groups of sample data come from populations that have the same or homogeneous variants in this homogeneous test the researcher uses the SPSS 26 program, with the following data:

TABLE IV. HOMOGENITY

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Hasil	Based on Mean	3.545	1	38	.067
	Based on Median	2.317	1	38	.136
	Based on Median and with adjusted df	2.317	1	35.935	.137
	Based on trimmed mean	3.309	1	38	.077

The basics of decision making in the homogeneity test are: First, when the significant value <0.05, it can be decided that the variance in the two groups is not homogeneous. Second, when the significant value> 0.05, it can be decided that the variance in the two groups is homogeneous. Based on the table of research results above, the significant value obtained is 0.077, which is more than 0.05 and it can be concluded that the research data is homogeneous.

The results of the Test of Homogeneity of Variance conducted using Levene's statistics showed that the

variances between the groups were not significantly different at the 0.05 level of significance. In the test based on the mean, the significance value was 0.067, indicating that there was no significant difference in variance between the tested groups. Similar results were found in the tests based on the median and median with adjusted degrees of freedom, with significance values of 0.136 and 0.137 respectively, as well as in the test based on the trimmed mean with a significance value of 0.077. All this confirms that the assumption of homogeneity of variance is met, which is important for further statistical analyses that require this assumption, such as analysis of variance (ANOVA).

Ensuring homogeneity of variance is important because it allows the use of statistical techniques that assume similarity of variance between groups. Since there is no evidence to suggest a significant difference in variance, researchers can proceed with analyses that combine data from different groups without the need to adjust for unequal variances. This eases interpretation of the results and reduces the need to apply more complex statistical techniques that would be required if the variances between groups were significantly different. Conclusions drawn from further analysis using ANOVA or similar techniques will, therefore, be grounded in strong statistical preconditions, resulting in more reliable findings.

E. Paired sample t-test

The paired t-test is a test of paired sample data, to determine whether there is a difference in the mean or average of the two paired groups.

TABLE V. PAIRED SAMPLE T-TEST

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	*Std. Deviation*	*Std. Error Mean*	*Lower*	*Upper*	*t*	*df*	*Sig. (2-tailed)*
Pair 1	*Pre-test -post-test*	-26.500	8.751	1.957	-30.596	-22.404	-13.543	19	.000

The basics of decision making in the paired sample t-test are: First, if the significant value <0.05 then H0 is rejected and Ha is accepted. Second, if the significant value> 0.05 then H0 is accepted and Ha is rejected.

Research hypothesis: if the hypothesis is H0, then: There is no average difference between the use of Al-Mumtaz application and student learning outcomes. If the hypothesis shows Ha, then: there is an average difference between the use of the Al-Mumtaz application and student learning outcomes.

Based on the table above, a significant value of 0.000 <0.05 is obtained, it can be concluded that there is an average difference between the use of the Al-Mumtaz application and student learning outcomes.

From the explanation above, the researcher concludes that there is an effect of the Al-Mumtaz application on student learning outcomes as seen from the significant value above in Arabic language learning in the Arabic language education department of Mahmud Yunus Batusangkar State Islamic University in the second semester.

TABLE VI. PAIRED SAMPLES STATISTICS

Paired Samples Statistics

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	*Pretes*	62.50	20	11.642	2.603
	Postes	89.00	20	8.522	1.906

The table above is to see more clearly the average learning outcomes before and after using the Al-Mumtaz application. The average pre-test results 62.50 and post-test 89.00 increased by 26.50. It can be stated that there is an increase after the use of the Al-Mumtaz application for Arabic language learning for second semester students specifically in the mufradat learning section.

IV. CONCLUSION

The use of Al-Mumtaz application in Arabic language learning at Mahmud Yunus State Islamic University Batusangkar provides significant results.

The validity and reliability of the research instruments were confirmed, with 10 out of 15 questions declared valid and high Cronbach's Alpha values. PreTest and PostTest data distribution showed normality, while the homogeneity of variance test met the assumptions of statistical analysis. The results of the analysis showed significant mean differences, confirming the positive influence of the Al-Mumtaz application on student learning outcomes. This conclusion highlights the importance of technology in improving the effectiveness of Arabic language learning in such academic environments.

ACKNOWLEDGMENT

This work was accompanied by prof munirul abidin in the field of quantitative research at Maulana Malik Ibrahim State Islamic University Malang.

REFERENCES

[1] I. Adnyana, *PENGEMBANGAN MEDIA PEMBELAJARAN BERBASIS APLIKASI KODULAR MENDUKUNG MODEL PEMBELAJARAN MATEMATIKA BERBASIS KOLABORASI …*. repo.undiksha.ac.id, 2023. [Daring]. Tersedia pada: https://repo.undiksha.ac.id/id/eprint/14169

[2] F. Andriawan, *Pengembangan mobile learning berbasis kodular pada mata pelajaran teknologi perkantoran untuk meningkatkan kemandirian belajar (studi pada peserta didik …*. mulok.lib.um.ac.id, 2021. [Daring]. Tersedia pada: http://mulok.lib.um.ac.id/index.php?p=show_detail&id=110545

[3] N. Atika, *Pengembangan media pembelajaran berbasis Android dengan Kodular untuk meningkatkan hasil belajar siswa kelas VIII materi statistika*. etheses.uingusdur.ac.id, 2023. [Daring]. Tersedia pada: http://etheses.uingusdur.ac.id/id/eprint/6515

[4] N. M. Sueni, "Metode, Model dan Bentuk Model Pembelajaran (Tinjauan Pustaka)," *Wacana Maj. Ilm. Tentang Bhs. Sastra Dan Pembelajarannya*, vol. 19, no. 1, hlm. 3–3, 2019.

[5] A. Awaluddin, "Evaluasi Pembelajaran Bahasa Arab Berbasis Informasi, Komunikasi, Dan Teknologi," no. Query date: 2024-03-12 12:33:48, 2023, [Daring]. Tersedia pada: http://repositori.iain-bone.ac.id/id/eprint/1542

[6] S. Suhardi, N. Nevrita, dan D. Sarkity, *Pengembangan E-Modul Praktikum Berbasis Kodular Sebagai Sumber Belajar Serta Panduan Dalam Pelaksanaan Praktikum Mikrobiologi SMK*. repositori.umrah.ac.id, 2023. [Daring]. Tersedia pada: http://repositori.umrah.ac.id/id/eprint/5842

[7] M. S. Priadana dan D. Sunarsi, *Metode Penelitian Kuantitatif*. Pascal Books, 2021.

[8] D. D. Unaradjan, *Metode penelitian kuantitatif*. Penerbit Unika Atma Jaya Jakarta, 2019.

[9] A. England, "Quantitative and qualitative research methods," *Res. Med. Imaging Radiat. Sci.*, no. Query date: 2023-11-30 23:13:48, hlm. 71–96, 2022, doi: 10.1007/978-3-030-79956-4_5.

[10] R. Ewing, *Basic quantitative research methods for urban planners*. 2020, hlm. 328. doi: 10.4324/9780429325021.

ISSN 2355-0082

The Effect of Video Games Towards the Students' Academic Performance

Yohanes Brian Caesaryano Lala[1], Raymond Sunardi Oetama[2], Kimberly Lvina[3]
[1,2,3] Department of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
raymond@umn.ac.id

Accepted 28 May 2024
Approved 19 November 2024

***Abstract*— Video games, originating in the mid-20th century, have become a major cultural phenomenon. Recently, about so many people, mostly young and mobile gamers, play video games. Despite their popularity, video games face controversy over potential negative impacts, such as promoting violence and harming academic performance. However, research on the correlation between gaming and academic performance is inconclusive. Some studies suggest a negative impact, while others find no significant effect. This study re-examines this relationship among university students in Tangerang using various statistical methods. The research process includes data collection via a questionnaire, followed by normality tests (Shapiro-Wilk Test and QQ Plot), the Kruskal-Wallis Test, and concluding with the Spearman Correlation Test. The results showed a p-value of 0.6376 and a rho value of -0.04767789, indicating no significant correlation and weak negative correlation. The scatterplot also supports this with widely dispersed data points. Therefore, we conclude that no significant correlation exists between time spent playing video games and students' GPA in the Tangerang area.**

***Index Terms*— *correlation; GPA;* video games; statistics.**

I. INTRODUCTION

Video games are an exciting byproduct that emerged from the rapid development of technology in the 20th Century. The first video game was made in 1958, a simple tennis game displayed via an oscilloscope as a screen. The game was known as tennis for two, where two players would be playing tennis on the screen [1]. Video games would then continue developing as time progressed, with Spacewar! Emerging in 1962 [2]. However, video games at that time had not yet entered popular culture the way it has now. It wasn't until the late 1970s and early 1980s that video games became the cultural phenomenon it is. It can be attributed to the fact that home systems became more widely available to the general populace by that time, allowing people of all ages to play in the comfort of their own homes [3]. Thus, that period became known as the golden age of video games, where video games began to see massive commercial success and influence within popular culture. Even today, video games remain massively popular worldwide, and the number of players continues to grow as time passes. As of 2023, it is estimated that the number of active video game players worldwide is around 3 billion. Most video game players are young, making up more than half of the world's active players. The smartphone is the most popular platform on which people play video games, with an estimated 1.93 billion mobile video game players worldwide [4].

However, despite its popularity, video games have been subject to societal controversies. Public perception of video games is mixed, with many holding positive or negative views. Most of the negative perceptions towards video games are based on the beliefs that video games are a waste of time and energy, are the cause of violent behavior, and can lead to addictions [5]. In addition, video games have also been viewed as negatively impacting academic performance. Numerous studies have claimed that the time spent playing games negatively affected students' grades. These studies found that the time spent playing video games negatively correlated with student's academic performance and thus concluded that video games negatively impact grades. Interestingly, however, video games have also been found to bring many positive benefits despite people's negative perceptions towards them. A study conducted in 2022 by IEEE found that playing video games, specifically in the strategy genre, can improve problem-solving and communication skills [6]. Video games have also been found to help improve mental health, and they are used to help patients as a tool for therapy [7]. Video games have also been found to help children improve their social skills, as they encourage collaboration in problem-solving through in-game challenges. It creates an opportunity for people, especially children, to socialize and develop their social skills while playing [8].

There is mix opinion between researchers in this area. Some believes that there is negative correlation between the time spent on video games and academic performance, while others believes there is no significant correlation. A survey conducted in 2022 estimated the Correlation between the time spent on video games and the Grade Point Average (GPA) of students in Universiti Teknologi Mara and found an association between playing video games and lower academic achievement to some degree [9]. A similar

research conducted in 2023 also came with a similar conclusion, where the findings indicate a negative correlation between the time spent on video games and the student's academic achievement measured in GPA one Eastern European country [10]. On the other hand, one research study concluded that there was no significant correlation between playing video games and lower student grades. Other study also found that non-gamers in Saudi Arabia still achieved slightly better grades than their gamer counterparts [11].

So, at this point, different area shows different results whether the correlation is significantly negative or not significant. Therefore, this study tries to retest this condition with varying subjects of research. The new subjects are university students in Tangerang area. Thus, this study will employ several statistical methods to determine whether video games correlate with student academic performance in Tangerang.

II. LITERATURE REVIEW

A. *The Shapiro-Wilk Test*

The Shapiro-Wilk test is a statistical test used to evaluate whether a sample of data comes from a normally distributed population [12]. In this test, the data are sorted, and the test statistic W is calculated to measure how well the data conform to a normal distribution using Equation (1).

$$W = \frac{\left(\sum_{i-1}^{n} a_i X_{(i)}\right)^2}{\sum_{i-1}^{n}(X_i - \bar{X})^2} \quad (1)$$

X(i) is the ordered sample values and the sample mean. $\bar{X}$ is the sample mean, and a_i is the weights that depend on the sample size n and the expected values of the order statistics of a standard normal distribution.

B. *QQplot*

Q-Q Plot (Quantile-Quantile Plot) is a graphical tool used to compare the distribution of sample data with a particular theoretical distribution, such as the normal distribution, by displaying the quantiles of the sample data against the quantiles of the theoretical distribution [13]. The primary purpose of the Q-Q Plot is to visualize how well the sample data fits the theoretical distribution, helping in checking the normality or fit of the data to another distribution. To read a Q-Q Plot, look at the X-axis, which shows the quantiles of the theoretical distribution, and the Y-axis, which shows the quantiles of the sample data. The diagonal reference line represents the perfect theoretical distribution. If the sample data follows the theoretical distribution, the points on the plot will be close to the reference line. Deviations from this line indicate differences in distribution; for example, points that form a concave curve indicate a thicker-tailed sample data distribution, while a convex curve indicates a thinner-tailed one. Thus, the Q-Q Plot helps identify the fit and deviation of the sample data distribution to the theoretical distribution.

C. *Kruskal-Wallis Test*

The Kruskal-Wallis test is a non-parametric statistical test used to determine whether there are significant differences between three or more groups of data that are not necessarily normally distributed [14]. Unlike ANOVA, which requires the assumption of normality, this test uses the ranks of the data to identify differences in medians between the groups. This test is beneficial when the data are ordinal or interval data do not meet the parametric assumptions. To read the results of the Kruskal-Wallis test, the test statistic H is computed and the resulting using Equation (2).

$$H = \left(\frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i}\right) - 3(N+1) \quad (2)$$

Where N= is the total number of observations.Ri is the sum of ranks for the i-th sample, n is the number of observations for the i-th sample, and k is the number of samples.

D. *Spearman's Correlation Test*

Spearman's Correlation, or Spearman's rank correlation coefficient, is a non-parametric statistical measure used to assess the strength and direction of the relationship between two variables based on their ranks [15]. It is an alternative to Pearson's Correlation and is useful when the data is not normally distributed or is ordinal. Spearman's Correlation does not require the assumption of normality. It can handle data that does not meet the parametric assumptions, making it very suitable for analyzing ordinal data or outliers. To read the Spearman Correlation results, consider the Spearman correlation coefficient value ρ, computed using Equation (3).

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (3)$$

Where ρ which ranges from -1 to 1. A value of ρ close to 1 indicates a solid positive monotonic relationship, while a value of ρ close to -1 indicates a strong negative one. A value of ρ close to 0 indicates the absence of a monotonic relationship. In addition, the p-value is used to test the statistical significance of the Correlation.

III. METHODS

The step by step research processes are shown in Fig. 1. The research is started with data collection through a Questionairre. Afterwards normality test consisit of The Saphiro Wilk Test and QQ Plot. Afterwards Kruskal Wallis Test. Ended with Spearman Correlation Test.

This study combines two distinct research methods: qualitative research and quantitative research. According to Creswell, qualitative research explores and understands the individuals or groups attributed to a social or human problem [16]. Qualitative research aims to answer 'what,' 'how,' or 'why' something happens instead of 'how many' or 'how much' [17]. Quantitative research, however, values breadth,

statistical descriptions, and generalizability. In other words, it values the use of more numbered data that can be measured by statistics and processed using mathematical procedures. Thus, using a mixed research method that utilizes both qualitative and quantitative aspects allows for a more in-depth analysis of a particular problem by measuring the quantitative data and validating it with qualitative data [18].

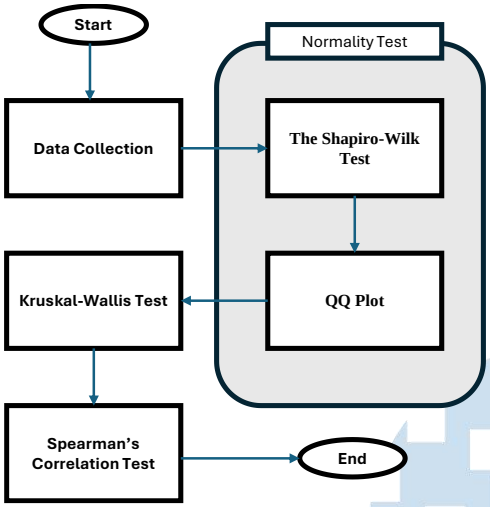


Fig. 1. Research Processes

A. *Data Collection*

To thoroughly analyze whether video games impact university students' academic performance, we first need to obtain data from them. To do this, we created a questionnaire containing questions that ask for general information, information about the games the respondents play, and information regarding the students' grades.

The population that is the subject of this research is university students in Tangerang. Sampling is selecting representative elements from a given population [19]. In other words, we collect data from a small sample – which may include a few members – that represents a larger population. In this research, the data is collected from a sample of 100 respondents of university students. This study's sampling methods include purposive random sampling to active students only. The respondents are randomly selected from the population, guaranteeing minimal sampling bias. Google Forms is a free online form tool provided by Google that can be used to create a wide range of forms. This tool allows users to customize the contents of their form, as well as determine the theme, and even the ability to set the form as a graded quiz. These forms are also easy to share, requiring only a link to access and fill them. The information gathered from these forms can then be placed neatly into a spreadsheet, which makes it easy to retrieve and process.

The R language is a programming language primarily used for statistics. R enables users to conduct statistical analysis and import, export, and visualize the data from the analysis results [20]. The R programming language is modified and run within its dedicated programming environment known as R Studio. Using R in R Studio, we can run several different statistical tests on the data we collected. Furthermore, we can create visualizations of the statistical analysis done in R.

B. *Normality Test*

The Shapiro and Wilk test determines if the numerical variables within this study are distributed normally. It is chosen because its popularity and frequently used in statistical research [21]. The resulting p-value decides whether to reject or accept the null hypothesis that the data are normally distributed. As shown in Table I, a low p-value (0.05) indicates that the data are not normally distributed, so the null hypothesis is rejected.

TABLE I. SHAPIRO WILK TEST DECISION TABLE

p-value	Hypothesis	Decision
> 0.05	Ho is not rejected	a normal distribution
≤ 0.05	Ho is rejected	not a normal distribution

A quantile-to-quantile plot, or a Q-Q plot, is a graph used to test between the empirical or observed distribution and a variable's theoretical or expected distribution [18]. In other words, the Q-Q plot is used to visualize the observed and predicted data distribution. A Q-Q plot is helpful in visually analyzing if a piece of data is distributed normally.

C. *Kruskal-Wallis Test*

The Kruskal-Wallis test is practical when trying to compare the mean of two variables. The Kruskal-Wallis test was chosen over other statistical tests for several reasons. The Kruskal-Wallis test can be applied to data that is not normally distributed [22]. This test is suitable for small sample sizes [23]. The null hypothesis states that the medians of all groups are the same, while the alternative hypothesis states that at least one group has a significantly different median. As shown in Table II, if the p-value is less than the significance level of 0.05, then the null hypothesis is rejected, indicating a significant difference in medians between the groups.

TABLE II. KRUSKAL-WALLIS TEST DECISION TABLE

p-value	Hypothesis	Decision
> 0.05	Ho is not rejected	There is no significant difference in medians between the groups.
≤ 0.05	Ho is rejected	a significant difference in medians between the groups.

D. *Spearman's Correlation Test*

Spearman's Correlation Test helps determine if there is an association between categorical and

numerical variables [24]. This Correlation can be measured between two numerical variables in which the distribution of both is unknown or is not distributed normally. It is chosen beuase its popularity in correlation testing [25]. As shown in Table III, if the p-value is less than the significance level of 0.05, then the relationship between the two variables is considered statistically significant.

TABLE III. SPEARMAN'S CORRELATION TEST DECISION

p-value	Hypothesis	Decision
> 0.05	Ho is not rejected	the relationship between the two variables is not considered significant.
≤ 0.05	Ho is rejected	the relationship between the two variables is considered significant.

IV. RESULTS AND DISCUSSION

A. Questionnaire Result

Table IV shows the categorical variables, while Table V figures numerical variables. For the preferred game platform, among the respondents, 50 preferred mobile, 49 preferred PC/laptop, and only one preferred console, which shows a nearly equal preference split between mobile and PC/laptop among the respondents, with the console being less popular, suggesting that most respondents prefer gaming on more accessible platforms. For the preferred Genre, 33 respondents play shooters, 27 play Multiplayer online battle arena (MOBA) games, 8 play role-playing games (RPG), 7 play racing games, 6 play action-adventure games, 5 each play battle royale and puzzle games, 3 play strategy games, and 1 respondent each plays arcade games, gacha games, fighting games, rhythm games, sports games, and tower defense games. It means that shooters (33%) and MOBA games (27%) are the most popular genres among the respondents, followed by RPGs and racing games. It reflects the various interests of the respondents. For the time spent playing video games (TimeGame), the average (mean) amount of time spent playing video games daily is 4.05 hours, with the median time spent on gaming being 3 hours each day and the most frequently stated time spent playing video games (mode) is 2 hours per day. It means there's a range in video game playtime, with the majority spending around 2-3 hours per day gaming. For the time spent studying (TimeStudy), the average (mean) time spent on studying per day is 4.03 hours, with the median study time being 3 hours each day, and the most frequently stated time spent studying (mode) is 2 hours daily. It shows a similar pattern to time spent on gaming (TimeGame), with most respondents studying for about 2-3 hours daily. For the current Grade Point Average (GPA), the respondents' average (mean) GPA is 3.4923, while the median GPA is 3.56. The most frequently stated GPA (mode) is 3.0, which shows that most respondents have a GPA around 3.0, with the average being somewhat higher at 3.49.

TABLE IV. CATEGORICAL VARIABLES

Data	Category	Frequency
Game Platform	Mobile	50
	PC / Laptop	49
	Console	1
Genre	Shooters	33
	MOBA	27
	RPG	8
	Racing	7
	Action-Adventure	6
	Battle Royale	5
	Puzzle	5
	Strategy	3
	Arcade	1
	Gacha	1
	Fighting	1
	Rhythm Game	1
	Sports	1
	Tower Defense	1

TABLE V. NUMERICAL VARIABLES

Data	Mean	Median	Mode
The Time Spent for Playing Video Games (TimeGame)	4.05	3	2
The Time Spent for Studying (TimeStudy)	4.03	3	2
Current Grade Point Average (GPA)	3.4923	3.56	3

B. Normality Test

The Shapiro-Wilk Test for Normality is used to determine if the numerical variables of a sample are distributed normally, and the Q-Q plot is used to visualize the empirical and theoretical distribution of a given data. In this study, the numerical variables include the time spent playing video games (TimeGame), the time spent studying (TimeStudy), and the GPA of each respondent. The following is the format for the hypotheses for these three data sets.

Ho_1: TimeGame data has a normal distribution. Ha_1: TimeGame data does not have a normal distribution

Ho_2: TimeGame data has a normal distribution. Ha_1: TimeStudy data does not have a normal distribution

Ho_3: GPA data has a normal distribution. Ha_1: GPA data does not have a normal distribution

The Shapiro-Wilk test result is in Table VI. All p-values are far below the significance level of 0.05, meaning that all variables significantly deviate from a normal distribution. Normality Test Result

From QQ plots in Fig. 2 to 4, it is also confirmed that those variables are not normally distributed as too many residues occur outside the area. Their points on the plots are not close to the reference line.

TABLE VI. SHAPIRO-WILK TEST RESULT

Variables	p-values	Status
TimeGame	1.39e-07	not a normal distribution
TimeStudy	1.273e-05	not a normal distribution
GPA	3.807e-05	not a normal distribution

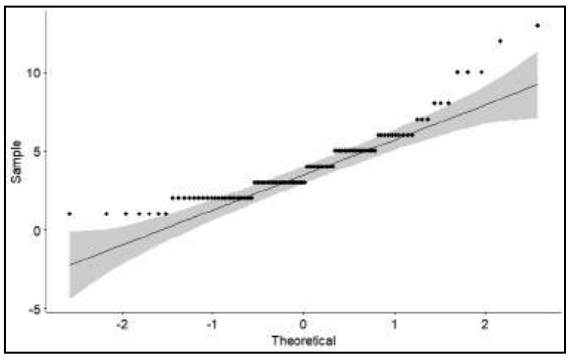


Fig. 2. Q-Q Plot for TimeGame

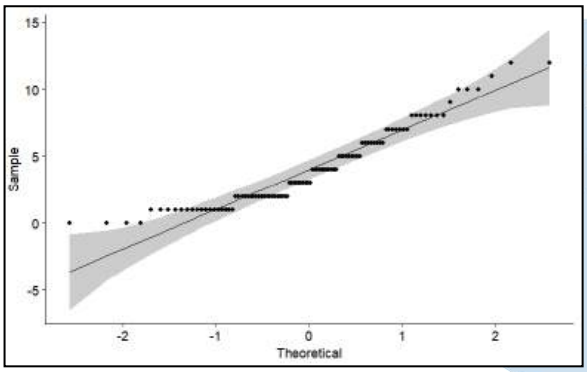


Fig. 3. Q-Q Plot for TimeStudy

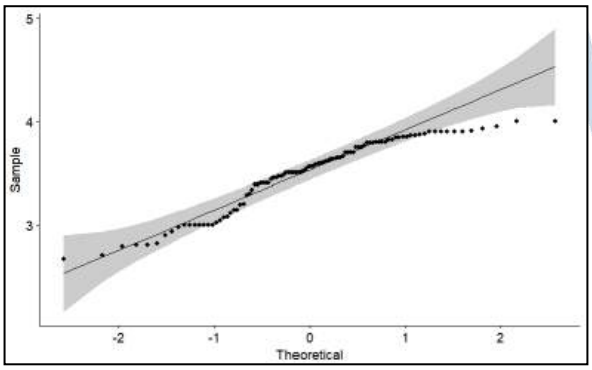


Fig. 4. Q-Q Plot for GPA

C. *Kruskal-Wallis Test*

Since the variables are not normally distributed, we apply The Kurskal-Wallis test as a non-parametric alternative to the one-way analysis of variance (ANOVA) test. The test measures the association of an independent categorical variable with a dependent numerical variable by comparing the mean values of both. In this study, we compare the categorical variables of 'Genre' and 'Platform' with the numerical variable 'TimeGame. The hypothesis formats are:

Ho_1: There is no significant difference in medians between Genre and TimeGame. Ha_1: There is a significant difference between Genre and TimeGame.

Ho_2: There is no significant difference in medians between Platform and TimeGame. Ha_2: There is a significant difference between Platform and TimeGame.

The results of the Kruskal-Wallis test towards the categorical variable of Genre with the numerical variable of TimeGame in Table VII show a p-value of 0.01672. The p-value is less than the significance level of 0.05, suggesting a significant association between the two variables. Thus, based on the test results, we can conclude that there is a significant association between the video game genre and the amount of time spent playing video games. For the variable Platform with TimeGame, we obtain the p-value of 0.0007408, suggesting a significant association between the two variables. Therefore, based on the test results, we can conclude that there is a significant association between the platforms used for playing video games and the time spent playing them.

TABLE VII. KRUSKAL-WALLIS TEST RESULT

Variables	p-values	Status
Genre and TimeGame.	0.01672	a significant difference
Platform and TimeGame.	0.0007408	a significant difference

D. *Spearman's Correlation Test*

In this study, we conducted Spearman's Correlation towards two numerical variables: TimeGame and GPA. Followings are the hypothesis format:

Ho: the relationship between TimeGame and GPA is not considered statistically significant.
Ha: the relationship between TimeGame and GPA is considered statistically significant.

Based on Spearman's Correlation in Fig. 5, we obtain the p-value of 0.6376 and the rho value of -0.04767789. The p-value is more significant than the significance level of 0.05, indicating no significant correlation between TimeGame and GPA. Additionally, the rho value supports the statement as it means a weak negative correlation between the two variables.

```
        Spearman's rank correlation rho

data:  data$TimeGame and data$GPA
S = 174596, p-value = 0.6376
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.04767789
```

Fig. 5. Spearman's Correlation Results

Furthermore, we can observe from the Scatterplot in Fig. 6 that the data points are significantly scattered, deviating from the theoretical distribution line. It further supports the argument that the two variables are not significantly correlated. Therefore, based on Spearman's Correlation and the scatterplot, we can conclude that no significant correlation exists between the Time Spent for Video Games and students' GPA.

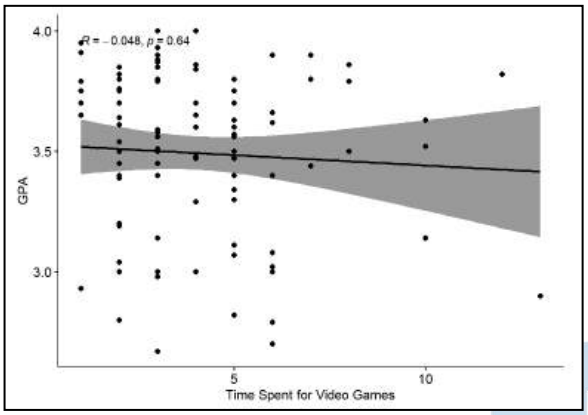


Fig. 6. Spearman's Correlation Scatterplot

E. Discussion

The Shapiro-Wilk tests conducted for the variables 'Time Spent for Video Games,' 'Time Spent Studying,' and GPA yielded p-values below the significance level of 0.05, indicating that these variables are not normally distributed. This result suggests that the assumption of normality, which is crucial for many parametric statistical methods, is violated. Consequently, it is appropriate to use non-parametric statistical methods when analyzing these variables. Non-parametric methods, such as the Mann-Whitney U test, Kruskal-Wallis test, or Spearman's rank correlation, are advantageous in this scenario because they do not assume a specific distribution for the data. They are robust and reliable for analyzing skewed data or data with outliers. Therefore, given the non-normal distribution of the variables, non-parametric methods are recommended to ensure the validity and accuracy of the statistical analysis.

The Kruskal-Wallis tests on the independent categorical variables 'Genre' and 'Platform' concerning the dependent numerical variable 'Time Spent for Playing Video Games' indicate significant correlations. This finding suggests that the type of video game genre and the gaming platform play crucial roles in determining the amount of time students dedicate to gaming. For example, specific genres like role-playing games (RPGs) or massively multiplayer online games (MMOs) might engage students longer than casual or puzzle games. Similarly, different platforms (e.g., consoles, PCs, mobile devices) might offer varying levels of accessibility and immersive experiences, influencing game time.

Based on the findings that video game genre and platform significantly influence the time spent playing video games, students should monitor and balance their gaming time, set specific time limits, prioritize tasks, and choose less time-consuming games. Mindful of their gaming habits can help them make informed choices and improve self-regulation. If managing gaming time becomes challenging, seeking support from peers, family, or counselors can provide additional strategies and accountability.

Spearman's correlation analysis for the variables 'Time Spent for Video Games' and 'Grade Point Average (GPA)' yielded a p-value more significant than the significance level of 0.05. It indicates that there is no statistically significant correlation between the two variables. However, the rho (ρ) value of -0.04767789 suggests a negative but weak correlation between time spent on video games and GPA. While the negative rho value implies that as time spent on video games increases, GPA tends to decrease slightly, the Correlation is weak and not statistically significant.

This lack of significance means we cannot confidently assert that video game time has a meaningful impact on GPA based on this data alone. Several factors could contribute to this weak Correlation. It's possible that moderate gaming does not significantly detract from academic performance for most students or that students who manage their time well can balance both activities effectively. Additionally, other variables not accounted for in this analysis, such as study habits, time management skills, and individual differences in cognitive abilities, could play a more significant role in determining GPA. The weak and non-significant Correlation highlights the complexity of the relationship between recreational activities and academic performance. It suggests that while excessive gaming might have some negative impact, it is not a predominant factor affecting GPA for most students. Further research with a more comprehensive set of variables could provide deeper insights into the dynamics between gaming habits and academic achievement.

V. Conclusion

The Shapiro-Wilk Test for Normality assesses whether numerical variables in a sample are typically distributed. This study's variables include time spent playing video games, studying, and GPA. The test results indicate that all variables deviate substantially from a normal distribution. The Kruskal-Wallis test results show significant associations between the categorical variable Genre and the numerical variable TimeGame, with a p-value of 0.01672, and between Platform and TimeGame, with a p-value of 0.0007408. Thus, video game genre and platform significantly influence the time spent playing video games. In this study, we analyzed the Correlation between TimeGame and GPA using Spearman's Correlation. The results showed a p-value of 0.6376 and a rho value of -0.04767789, indicating no significant correlation, as the

p-value is above 0.05 and the rho value suggests a weak negative correlation. The scatterplot also supports this with widely dispersed data points. Therefore, we conclude that no significant correlation exists between time spent playing video games and students' GPA in Tangerang area.

The finding that there is no significant correlation between time spent playing video games and students' GPA implies that playing video games does not directly affect academic performance. It suggests that factors such as study habits, time management, and individual differences may play a more critical role in influencing GPA. Therefore, concerns that video game time alone might negatively impact academic success may be unfounded, and a more holistic approach should be taken when addressing students' academic performance and extracurricular activities. Besides, considering to analyse respondents' GPA data from several subsequent semesters.

ACKNOWLEDGMENT

We thank Universitas Multimedia Nusantara for their assistance with facilities and funding during our research.

REFERENCES

[1] J. Hasanov, "Advertising with Digital Technologies in Video Games Arena," 2023.

[2] A. J. D. Mirowski, *Making (Outer) Space to Play: A Cosmic History of the American Video Game*. Indiana University, 2021.

[3] G. Dale, F. Kattner, D. Bavelier, and C. S. Green, "Cognitive abilities of action video game and role-playing video game players: Data from a massive open online course.," *Psychol. Pop. Media*, vol. 9, no. 3, p. 347, 2020.

[4] M. L. Shuwandy *et al.*, "Sensor-Based Authentication in Smartphone; a Systematic Review," *J. Eng. Res.*, 2024.

[5] S. Ha and S. Kim, "Barriers to playing digital games: Why do some people choose not to play digital games?," *Telemat. Informatics*, p. 102161, 2024.

[6] A. Lie, A. Stephen, L. R. Supit, S. Achmad, and R. Sutoyo, "Using strategy video games to improve problem solving and communication skills: A systematic literature review," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, 2022, pp. 1–5.

[7] F. Bocci, A. Ferrari, and M. Sarini, "Putting the gaming experience at the center of the therapy—the Video Game Therapy® approach," in *Healthcare*, 2023, vol. 11, no. 12, p. 1767.

[8] B. K. Wiederhold, "Kids will find a way: The benefits of social video games," *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 4. Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New …, pp. 213–214, 2021.

[9] T. Savić Tot, S. Adžić, V. Tot, M. Aleksić, and N. Zakić, "The impact of time devoted to video games on student achievement," *Educ. Inf. Technol.*, vol. 28, no. 5, pp. 5921–5944, 2023.

[10] N. A. A. Wahid and N. N. A. Hassan, "Video Game Engagement and Academic Performance: A Correlation Analysis Approach," *Randwick Int. Soc. Sci. J.*, vol. 4, no. 4, pp. 770–784, 2023.

[11] S. Adžić, J. Al-Mansour, H. Naqvi, and S. Stambolić, "The impact of video games on Students' educational outcomes," *Entertain. Comput.*, vol. 38, p. 100412, 2021.

[12] E. González-Estrada and W. Cosmes, "Shapiro–Wilk test for skew normal distributions based on data transformations," *J. Stat. Comput. Simul.*, vol. 89, no. 17, pp. 3258–3272, 2019.

[13] M. C. Korkmaz, C. Chesneau, and Z. S. Korkmaz, "The unit folded normal distribution: A new unit probability distribution with the estimation procedures, quantile regression modeling and educational attainment applications," *J. Reliab. Stat. Stud.*, pp. 261–298, 2022.

[14] F. N. Nwobi and F. C. Akanno, "Power comparison of ANOVA and Kruskal–Wallis tests when error assumptions are violated," *Metod. Zv.*, vol. 18, no. 2, pp. 53–71, 2021.

[15] R. H. A. Shiekh and E. F. El-Hashash, "A comparison of the pearson, spearman rank and kendall tau correlation coefficients using quantitative variables," *Asian J. Probab. Stat.*, pp. 36–48, 2022.

[16] S. T. Akyıldız and K. H. Ahmed, "An overview of qualitative research and focus group discussion," *Int. J. Acad. Res. Educ.*, vol. 7, no. 1, pp. 1–15, 2021.

[17] R. Maietta, P. Mihas, K. Swartout, J. Petruzzelli, and A. B. Hamilton, "Sort and sift, think and shift: let the data be your guide an applied approach to working with, learning from, and privileging qualitative data.," *Qual. Rep.*, vol. 26, no. 6, 2021.

[18] H. Taherdoost, "What are different research approaches? Comprehensive Review of Qualitative, quantitative, and mixed method research, their applications, types, and limitations," *J. Manag. Sci. Eng. Res.*, vol. 5, no. 1, pp. 53–63, 2022.

[19] M. M. Rahman, M. I. Tabash, A. Salamzadeh, S. Abduli, and M. S. Rahaman, "Sampling Techniques (Probability) for Quantitative Social Science Researchers: A Conceptual Guidelines with Examples," *SEEU Rev.*, vol. 17, no. 1, pp. 42–51, 2022, doi: 10.2478/seeur-2022-0023.

[20] M. Mehmetoglu, "Applied statistics using R: A guide for the social sciences," 2021.

[21] N. Khatun, "Applications of normality test in statistical analysis," *Open J. Stat.*, vol. 11, no. 01, p. 113, 2021.

[22] S. W. Lee, "Regression analysis for continuous independent variables in medical research: statistical standard and guideline of Life Cycle Committee," *Life Cycle*, vol. 2, 2022, doi: 10.54724/lc.2022.e3.

[23] A. Anaya-Isaza and L. Mera-Jimenez, "Data Augmentation and Transfer Learning for Brain Tumor Detection in Magnetic Resonance Imaging," *IEEE Access*, vol. 10, pp. 23217–23233, 2022, doi: 10.1109/ACCESS.2022.3154061.

[24] A. M. Alsaqr, "Remarks on the use of Pearson's and Spearman's correlation coefficients in assessing relationships in ophthalmic data," *African Vis. Eye Heal.*, vol. 80, no. 1, p. 10, 2021.

[25] S. Chatterjee, "A new coefficient of correlation," *J. Am. Stat. Assoc.*, vol. 116, no. 536, pp. 2009–2022, 2021.

ISSN 2355-0082

Approach Convolutional Neural Network LeNet-5 for Interactive Learning of Korean Syllables (Hangul)

Vasyilla Kautsar Al Fitra Yudha[1], Arvita Agus Kurniasari[2],
Aji Seto Arifianto[3], Faisal Lutfi Afriansyah[4]
Departement of Informatics Technology, State Polytechnic of Jember, Jember, Indonesia
[1]vasyillakafy@gmail.com, [2]arvita@polije.ac.id, [3]ajiseto@polije.ac.id, [4]faisal.lutfi@polije.ac.id

Accepted 2 July 2024
Approved 19 November 2024

***Abstract*— The increasing popularity of South Korean culture among Indonesian society has led to a growing interest in gaining a deeper understanding of the country, including a desire to master the Korean language. However, learning the Korean alphabet (hangul) often presents challenges due to its characters being unfamiliar to the Indonesian people. Therefore, engaging and interactive learning media are needed to assist in the learning process. Within this endeavor, a learning website called Learn Hangul was developed, focusing on two main features: learning hangul characters and their arrangement, as well as practicing writing syllables using Korean letters. This website was developed using the Convolutional Neural Network (CNN) LeNet-5 to facilitate learning, with black box testing results indicating good functionality. Model performance evaluation yielded satisfactory values, with model accuracy at 89.2%, precision at 89.7%, recall at 88.8%, and an F1-score of 89.2%. Direct testing with users also showed a high success rate, with 80% of respondents experiencing an increase in their knowledge of Korean characters (Hangul) after trying to learn them on the Learn Hangul website. Thus, the Learn Hangul website serves as a useful learning tool for those interested in studying the Korean alphabet (hangul).**

***Index Terms*— Convolutional Neural Network; LeNet-5; Korean Language; Hangul; Website; Learning Media**

I. INTRODUCTION

Hallyu (한류) is a Korean term where "Han" (한) refers to "Hankuk" meaning Korea, and "Lyu" (류) means wave or flow (Valenciana and Pudjibudojo, 2022). Hallyu refers to the growing public interest in Korean pop and traditional arts around the world. The influence of Hallyu has made Indonesia one of the Asian countries affected by the Korean Wave. With the regular consumption of Korean cultural content, this leads to a desire to learn the Korean language, fostering an interest in studying Korean [1].

Research by Mutiara shows a strong correlation between watching Korean dramas and the interest in learning the Korean language among students at Mercu Buana University's Faculty of Communication Sciences. Their study found that exposure to Korean dramas influences Korean language learning interest by 43.8% [2]. Another research by Hasanah, based on a survey of students in UGM's Korean Language D3 Program revealed that 92.6% were initially drawn to Korean culture before being motivated to learn the language. Moreover, 96.3% confirmed using Korean culture as a learning tool, underscoring its significant role in attracting students to study Korean [1].

Furthermore, Many fans of Korean entertainment are drawn to South Korea not only for its cultural appeal but also for its highly regarded education system, recognized globally for its excellence [3]. The availability of scholarships, including the KGSP (Korea Global Scholarship Program), offered by both local and Indonesian governments, further encourages students to pursue studies at prestigious South Korean universities. As of November 2019, approximately 1,500 Indonesian students were enrolled in universities across South Korea, as reported by the Embassy of the Republic of Indonesia in Seoul.

The impact of the popularity of Korean culture in Indonesia and the growing interest in studying in South Korea has led to an increased interest among Indonesians in learning the Korean language. This ranges from fans who want to know more about South Korea to students who may live alongside Korean society. Before learning the Korean language directly, students first need to familiarize themselves with the official script of South Korea, which is Hangul.

Hangul is the official name of the Korean alphabet used by the Korean people, created by the Great King Sejong. Hangul consists of 40 characters, including 21 vowels and 19 consonants (Seon Jung et al., 2015). Hangul is written differently from the usual alphabet, with each letter forming a specific character, and the characters are written in square blocks similar to Chinese characters, which is called Gulja [4].

According to the International Standard Curriculum of The Korean Language for level 1 in the writing field, there are several achievement standards that must be

met. One of them is that students should be able to form words by combining consonant and vowel letters according to orthographic rules. To form words, syllables are needed as the building blocks. When writing syllables using Hangul, there are specific writing rules. The unfamiliarity of the character shapes and writing methods for Indonesians may present some difficulties at the beginning of the learning process.

Currently, there are many free platforms that provide learning materials for the Korean language, such as YouTube, websites, and even on social media like Instagram. However, the learning resources often focus only on reading and listening skills or theoretical learning without direct practice. Meanwhile, writing practice or direct application can hone skills and improve proficiency in writing Hangul, whose characters are unfamiliar to Indonesians. Therefore, engaging and interactive learning media are needed to achieve optimal learning outcomes.

Current technological advancements can be maximized to make the learning process easier and more flexible, by creating a "Learn Hangul" website for introducing the writing patterns of Hangul syllables. In addition to theoretical learning, students can also practice writing directly. This way, students will not only become proficient in reading but also be able to write and form Hangul syllables correctly.

Website development can be done using one of the methods from neural networks. Neural networks can be described as functioning similarly to the human brain, by training the system to recognize patterns in training data to achieve a good level of accuracy. In the process, each pixel is analyzed and matched with the training data that has undergone the neural network process, making it very suitable for solving classification or pattern recognition problems in objects [5]. Convolutional Neural Network (CNN) is a type of neural network commonly used to process image data. CNNs are typically employed to detect and recognize objects in images [6]. CNNs have a deep network architecture, allowing them to achieve high accuracy and produce good results [7].

Based on existing issues and previous research conducted on pattern recognition in objects, this study employs the Convolutional Neural Network (CNN) method to recognize the writing of Korean syllabic characters (Hangul).

II. LITERATURE REVIEW

A. Hangul

Hangul (한글) is the alphabet used by the Korean people for everyday, consists of 40 letters, comprising 21 vowel letters and 19 consonant letters. Among the 21 vowel letters, 10 are basic vowels and 11 are expanded vowels derived from the basic forms. As for the 19 consonant letters, 14 are basic consonants, and 5 more letters are double consonants [8].

B. Korean Syllables

Each Hangul letter must have a pair (consonant + vowel) to be pronounced and form a syllable, with each letter placed according to specific positional rules. To form a Hangul syllable, several patterns are used: CV, CVC, CVVC, CVCC, and CVVCC. C represents a consonant and V represents a vowel. The arrangement of Korean syllabic characters shown in Fig 1.

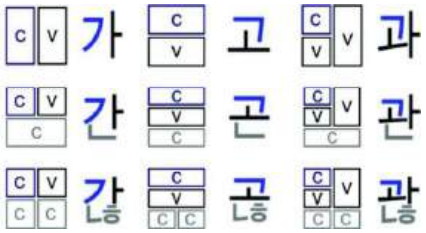


Fig 1. Arrangement of Korean syllables [9]

Vowel letters are arranged vertically or horizontally. ㅏ, ㅑ, ㅓ, ㅕ, ㅣ are vertical vowels, while ㅗ, ㅛ, ㅜ, ㅠ, ㅡ are horizontal vowels. When forming Hangul letters into syllables, consonants are written to the left of vertical vowels and above horizontal vowels [8]. The combination of one consonant with one vowel shown in Fig 2.

	ㅏ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ
ㄱ	가	갸	거	겨	고	교	구	규	그	기
ㄴ	나	냐	너	녀	노	뇨	누	뉴	느	니
ㄷ	다	댜	더	뎌	도	됴	두	듀	드	디
ㄹ	라	랴	러	려	로	료	루	류	르	리
ㅁ	마	먀	머	며	모	묘	무	뮤	므	미
ㅂ	바	뱌	버	벼	보	뵤	부	뷰	브	비
ㅅ	사	샤	서	셔	소	쇼	수	슈	스	시
ㅇ	아	야	어	여	오	요	우	유	으	이
ㅈ	자	쟈	저	져	조	죠	주	쥬	즈	지
ㅊ	차	챠	처	쳐	초	쵸	추	츄	츠	치
ㅋ	카	캬	커	켜	코	쿄	쿠	큐	크	키
ㅌ	타	탸	터	텨	토	툐	투	튜	트	티
ㅍ	파	퍄	퍼	펴	포	표	푸	퓨	프	피
ㅎ	하	햐	허	혀	호	효	후	휴	흐	히

Fig 2. Combination 1 consonant + 1 vowel

C. Convolutional Neural Network (CNN)

Convolutional Neural Network is a type of artificial neural network inspired by the functioning of the human brain's visual cortex [10]. CNN is developed as an evolution of the Multilayer Perceptron (MLP), specializing in processing two-dimensional grid-like data, such as images and videos [11]. CNN has many uses, especially in the fields of image and video processing, such as face recognition, object detection, image segmentation, and others [12]. This algorithm is highly popular due to its effective and efficient capability in processing data with large and complex grid structures.

CNNs leverage key image processing techniques such as convolution and pooling to enhance feature extraction and reduce computational complexity. Convolutional layers apply filters to input data to create feature maps, while pooling layers perform downsampling operations to reduce dimensionality, making the network more manageable and less prone to overfitting [13].

The stages of how CNN works are divided into two stages: feature learning (which consists of convolution, pooling, activation) and image classification (which consists of flatten layer, fully connected layer), as shown in Fig 3.

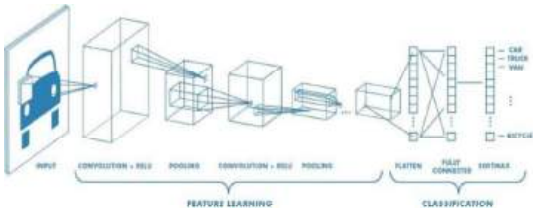


Fig 3. Convolutional Neural Network

D. Convolution Layer

Convolutional layer is a key block in feature learning and the CNN algorithm. In this layer, filters are applied to the incoming image to extract information and values from the previous layer. Filters come in various sizes depending on the type of CNN used, typically 3x3 in size, but there are also those sized 5x5 and 7x7 [10]. Filters or kernels are also commonly referred to as feature detectors, as indicated by the blue color in Fig 4.

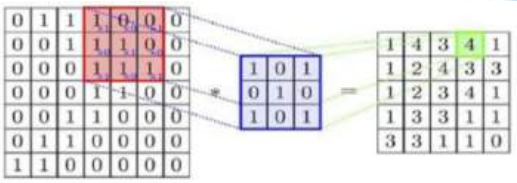


Fig 4. Multiplication with Feature Detector

E. Pooling Layer

Pooling is a layer in CNNs that serves to reduce spatial size so that the subsequent layers do not require excessive computation. This layer is also useful for addressing overfitting issues. There are many types of pooling, such as max pooling, min pooling, average pooling, stochastic pooling, spatial pyramid pooling [10]. The most commonly used pooling methods are max pooling and average pooling. Max pooling uses the highest value, while average pooling computes the average pixel value, as illustrated in Fig 5. However, this study utilizes max pooling.

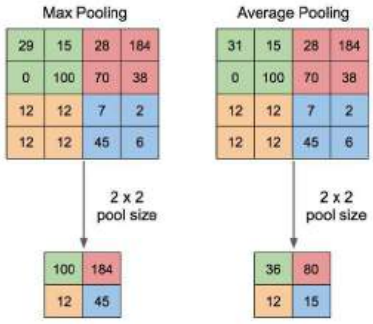


Fig 5. Max Pooling and Average Pooling

F. ReLu Activation

ReLU (Rectified Linear Unit) is used to convert negative values to zero. If the input to a neuron is negative, it will be converted to 0. For positive inputs, the value remains unchanged, meaning the output of the neuron will be the same as the input value itself. The formulation of ReLU shown in Equation 1.

$$f(x) = max\,(0, x) \quad (1)$$

G. Flatten Layer

The feature map generated in the previous process is in the form of a multidimensional array. However, in the fully connected layer, the input needed must be in the form of a vector. Therefore, a flatten layer is required. The Flatten Layer functions to reshape the matrix from the pooling layer into a single column (a single vector). The output of the flatten layer is a vector. Thus, these values can be used as input in the fully connected layer [14]. An illustration of the flatten layer shown in Fig 6.

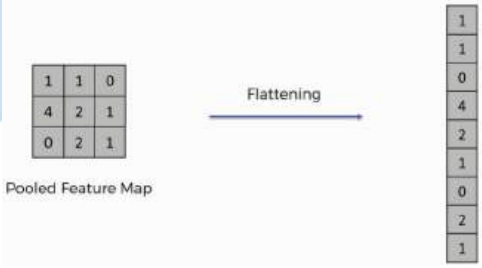


Fig 6. Flatten Layer

H. LeNet-5

LeNet-5 is one of the Convolutional Neural Network (CNN) architectures developed by Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner in 1998, primarily used for handwritten digit recognition tasks [15]. LeNet-5 consists of 7 layers, including convolutional layers, subsampling layers, followed by several fully connected layers. Despite its simplicity compared to contemporary architectures like VGG and ResNet, LeNet-5 laid the groundwork for subsequent developments in deep learning, influencing the design of neural networks for various image processing and classification tasks.

I. Confusion Matrix

Confusion Matrix is a performance measurement for classification problems in machine learning, presented in the form of a 4-table matrix that displays various combinations of predicted and actual values. Essentially, this matrix contains information that compares the system's classification results with the actual classification outcomes [16].

TABLE I. CONFUSION MATRIX

CLASS	*Classified as Positive*	*Classified as Negative*
Positive	TP	FN
Negative	FP	TN

There are 4 terms that represent the results of the classification process:

- TP (True Positive), which represents correctly detected positive data.
- FN (False Negative), which represents positive data incorrectly detected as negative.
- TN (True Negative), which represents correctly detected negative data.
- FP (False Positive), which represents negative data incorrectly detected as positive.

Using the values of TP, TN, FP, and FN, various performance evaluation metrics of the model such as precision, recall (sensitivity) can be calculated [17].

III. METHODOLOGY

A. Object of Research

This research focuses on detecting the writing of Hangul syllables on 'Learn Hangul' website to help students who aim to become proficient in write Korean letters into readable syllables. The output of 'Learn Hangul' website is the romanization of hangul syllables written by users.

The limitation of this research is that it only classifies Hangul syllables formed by one consonant letter combined with one vowel letter. Thus, each syllable is made up of two letters, as shown in Fig 2 (Chapter II, Subchapter B: Korean Syllables). The data is divided into 182 categories, corresponding to the number of syllables composed of two Hangul letters. Syllables used in this research are the basic forms of other Hangul syllables.

B. CNN Model

In this stage, the architecture of the model is designed based on the LeNet-5 network architecture, which is used to train Korean syllable image data.

TABLE II. OUTPUT SHAPE LENET-5

Layer	*Output Shape*	*Parameter*
Input image	**32,32,3**	-
conv2d	(*None*, 28, 28, 6)	456
max_pooling2d	(*None*, 14, 14, 6)	0
conv2d_1	(*None*, 10, 10, 16)	2416
max_pooling2d_1	(*None*, 5, 5 16)	0
conv2d_2	(*None,* 1, 1, 120)	48120
flatten	(*None*, 120)	0
dense	(*None, 84*)	10164
dense_1	(*None*, 182)	15470
Total params	**:**	**76626**
Trainable params	**:**	**76626**
Non-trainable params	**:**	**0**

The first layer is a convolutional layer with ReLU activation, using a 5x5 kernel size and 6 filters. It takes input images sized 32x32 and produces feature maps of size 28x28 with a depth of 6. The second layer is a max pooling layer with a 2x2 kernel, resulting in 6 feature maps sized 14x14. The third layer is another convolutional layer with ReLU activation, using a 5x5 kernel to generate 16 feature maps sized 10x10. The fourth layer is a max pooling layer with a 2x2 kernel, producing 16 feature maps sized 5x5. The fifth layer is the final convolutional layer with ReLU activation, using a 5x5 kernel to create 120 feature maps each of size 1x1. This is followed by a flatten layer to reshape the output into a vector. The sixth layer is a fully connected dense layer with 120 outputs. The seventh and final layer is also a fully connected layer, producing probabilities across 182 classes.

C. Design System

In this stage, two designs are created: system design and design for user flow. The system design shown in Fig 7.

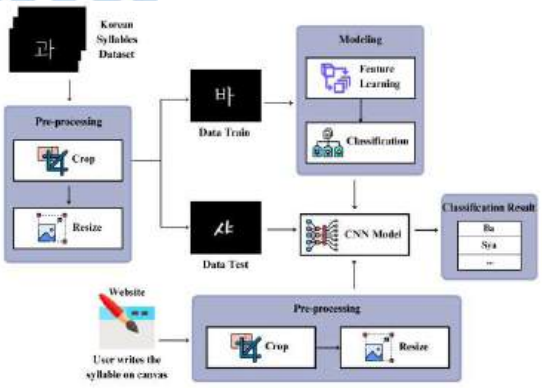


Fig 7. Design System

The first step is data pre-processing, where the data is cropped to remove empty spaces, leaving only the objects, and then resized to 32x32 to fit the CNN input. The dataset is then divided into training and testing data. Training data is used to train the system to recognize patterns and create a model, while testing data is used to evaluate the model's performance. Next,

on the website, users write Korean syllables on the provided canvas. The system then processes the writing, starting with preprocessing steps including cropping and resizing the image to 32x32 pixels to fit the CNN model's input size. After preprocessing, the image is ready to be processed and classified using the CNN model.

User flow or steps that need to be taken by users in the Learn Hangul website shown in Fig 8.

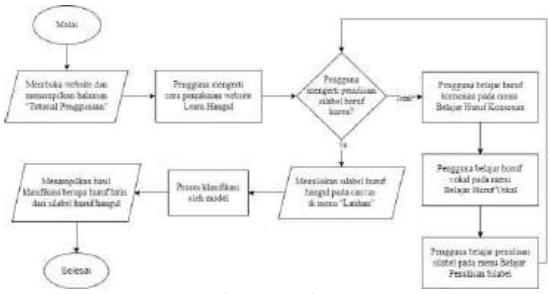


Fig 8. User flow

D. Data Preparation

To develop a model, a dataset is required. The data used is secondary data obtained from Kaggle.com. The data consists of grayscale images of Korean syllables sized 224 x 224 pixels, consists of 20,202 images of Korean Hangul syllables that composed of 2 characters (1 consonant + 1 vowel), divided into 182 labels with 111 images per label. The naming of these 182 classes is based on syllables consisting of 2 characters, such as “da”, “dae”, “de”, “deo”, “deu”, “dya”, “du”, “do”, and so on, following the conventions of writing Korean syllables.

In this study, the dataset is split into data train and data test with a ratio of 80:20. Details of the dataset split shown in Table III. Out of the total 20,202 images, 80% are used as training data, amounting to 16,198 images, and 20% are used as test data, totaling 4,004 images. Each label contains 89 train images and 22 test images.

TABLE III. DATASET DETAIL

	Qty	*%*	*Class*	*Data in 1 class*
Data Train	16.198 images	80%	182 class	89 images
Data Test	4004 images	20%	182 class	22 images

The data, which has been divided into folders according to their classes is ready to be used for developing a CNN model, and then uploaded it to Google Drive to be accessible through Google Colaboratory.

E. Pre-processing

The first step in this preprocessing is cropping. The dataset has a size of 224x224 pixels, as shown in Fig 9. This size is too large and contains too much meaningless black space.

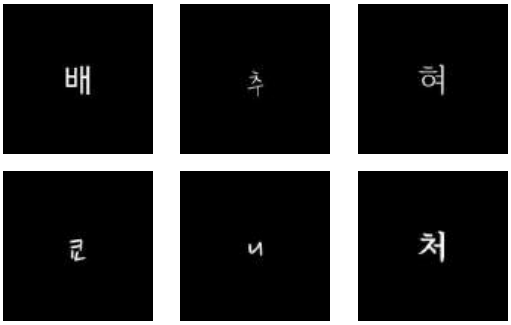


Fig 9. Dataset 224x224

This involves removing the black outer parts of the image, leaving only the object in the center. After removing the black space, the image is resized to 32x32 using the cv.resize() function from the OpenCV library. The results of the cropping and resizing shown in Fig 10.

Fig 10. Cropping data to 32x32px

The first step in feature extraction is Sobel edge detection. Feature extraction using edge detection is chosen to recognize the edges of objects in each image and highlight detailed parts of the image. Research by Widiawati found Sobel superior to Robert and Prewitt methods in detecting facial shapes and edge sharpness [18]. Thus, Sobel is used for edge detection in this study.

Fig 11 displays the original data before edge detection at the top and the data after applying Sobel edge detection at the bottom.

Fig 11. Result of Sobel Edge Detection

The second operation is morphological dilation, which expands or thickens objects to make them clearer. This process is necessary because the original dataset contains small and thin objects. Morphological dilation helps make these objects more visible and easier to analyze. Research by Anwariyah found that this operation effectively improves image quality and simplifies the segmentation of vehicle license plate characters [19]. Morphological dilation is performed by adding pixels to the edges of objects in the image.

The results shown in Fig 12, with the top section showing the data before dilation and the bottom section showing the data after morphological dilation.

Fig 12. Result of Morphological Dilation

F. Model Train Scenario

In this study, there are several scenarios for training the model by varying the batch size across three different datasets: the original dataset, the dataset with Sobel edge detection, and the dataset with morphological dilation operations. Batch size significantly impacts model train in machine learning. It refers to the number of data samples fed into the model during each training iteration. The training scenarios used in this study shown in Table IV.

TABLE IV. MODEL TRAIN SCENARIO

Name	*Data Pre-processing*	*Batch Size*
Dataset A	Crop + Resize	16
		32
		64
		128
		256
Dataset B	Crop + Resize + Sobel Edge Detection	16
		32
		64
		128
		256
Dataset C	Crop + Resize + Morphological Dilation	16
		32
		64
		128
		256

G. Testing

In this stage, testing of the system and application is conducted. Testing is crucial to ensure that the application meets requirements and functions properly without any defects. In this study, testing is conducted using 5 methods outlined as follows:

1) Manual testing: This involves evaluating the CNN model's ability to classify Korean syllables using data that not included in the training or test sets. The images used for this testing are handwritten Korean syllables by the researcher.

2) Confusion Matrix: Used to evaluate the performance of the CNN model.

3) Expert validation: Testing conducted with Korean language experts to ensure that the information conveyed in the website is accurate and does not mislead users.

4) Evaluating the usefulness of the Learn Hangul learning website by administering quizzes to users directly via Google Form, aimed at assessing the website's utility for users interested in learning Korean letters.

IV. RESULT AND DISSCUSSION

A. Model Train Result

In the first scenario, model training was conducted using dataset A, which underwent pre-processing with cropping and resizing only. The highest validation accuracy of 0.8923 was achieved with a batch size of 32, while the highest validation loss of 0.5066 was obtained with a batch size of 128.

Among the five model training sessions with dataset A (pre-processing Crop + Resize), larger batch sizes affected the accuracy results. The highest accuracy was achieved with a batch size of 32, but accuracy declined with batch sizes of 64 and larger. Although accuracy improved at a batch size of 256 compared to 128, it remained lower than the accuracy achieved with a batch size of 32. The results of the first scenario training are summarized in Table V.

TABLE V. RESULT OF TRAIN USING DATASET A

Dataset	*Batch Size*	*Loss*	*Accuracy*	*Val loss*	*Val accuracy*
A	16	0.2496	0.9179	0.3659	0.8911
	32	0.1269	0.9565	0.4074	0.8923
	64	0.2193	0.9265	0.3859	0.8867
	128	0.1983	0.9346	0.5066	0.8670
	256	0.2120	0.9318	0.4243	0.8798

In the second scenario, the dataset B underwent preprocessing steps including cropping, resizing, and Sobel edge detection. The highest validation accuracy, 0.8619, was achieved with a batch size of 16. Meanwhile, the highest validation loss, 0.6598, was reached with a batch size of 64.

Based on the training results with dataset B, using a larger batch size during model training affects accuracy. Similar to the first training scenario, the table shows that batch sizes from 32 to 256 have lower validation accuracy than batch size 16. Although accuracy improves at batch size 128, it still does not surpass the accuracy achieved with batch size 16. The training results for scenario B are summarized in Table VI.

TABLE VI. RESULT OF TRAIN USING DATASET B

Dataset	Batch Size	Loss	Accuracy	Val loss	Val accuracy
B	16	0.2866	0.9009	0.4882	0.8619
	32	0.2728	0.9054	0.4918	0.8553
	64	0.3955	0.8691	0.6598	0.8145
	128	0.2614	0.9138	0.4931	0.8565
	256	0.3166	0.8979	0.5647	0.8380

In the third training scenario, using dataset C involved preprocessing with cropping, resizing, and dilation morphological operations. The model achieved the highest validation accuracy of 0.8920 with a batch size of 32. Meanwhile, the lowest validation loss of 0.4103 was achieved with a batch size of 64.

Based on the training results with dataset C, the validation accuracy increased from batch size 16 to 32. However, it decreased from batch size 64 to 256. Although the accuracy increased again with batch size 256, it remained lower than the accuracy achieved in training with batch size 32. The outcomes of the third training scenario are summarized in Table VII.

TABLE VII. RESULT OF TRAIN USING DATASET C

Dataset	Batch Size	Loss	Accuracy	Val loss	Val accuracy
C	16	0.2554	0.9140	0.4736	0.8704
	32	0.1513	0.9472	0.4312	0.8920
	64	0.2470	0.9178	0.4103	0.8832
	128	0.2673	0.9106	0.5511	0.8447
	256	0.1587	0.9499	0.5058	0.8657

Next, three models with the highest accuracy were obtained from each training scenario. The highest accuracy was achieved with the original dataset, followed by the dataset with morphology dilation, and the lowest accuracy was observed in the model trained with the Sobel edge detection dataset. These results are summarized in Table VIII.

TABLE VIII. MODELS WITH HIGHEST ACCURACY

Dataset	Batch Size	Val Loss	Val accuracy
A	32	0.4074	0.8923
B	16	0.4882	0.8619
C	32	0.4312	0.8920

Among the scenarios, the lowest accuracy was observed with the Sobel edge detection dataset due to Hangul characters' smooth curves and strokes, which Sobel struggles to detect accurately, leading to disconnected or incorrect edges. Additionally, closely spaced or complex strokes in Hangul characters may cause Sobel to detect double edges, complicating character shape interpretation.

Similarly, models trained with the dilation morphology dataset showed lower accuracy than those trained with the original dataset, possibly because dilation could merge closely spaced or complex strokes in Hangul characters, reducing readability and causing recognition errors.

B. User Interface Website Learn Hangul

In this stage, which occurs after the model development phase, the website is built using the Flask framework, incorporating HTML, CSS, and JavaScript for front-end management, and Python for back-end operations.

The Practice Page or “Latihan” is a main feature of this website, designed for users to practice writing Hangul syllables consisting of 2 letters (1 consonant 1 vowel). This page appears when users click the "Latihan" button in the navigation bar. On this practice page, there is a black canvas where users can write Hangul syllables. Then there is a "Process Gambar" button, which is clicked after the user writes a Korean syllable on the canvas. After clicking the "Process Gambar" button, the romanization or alphabet letters of the syllable previously written by the user on the canvas will appear in the "Romanization" section. The layout of the practice page shown in Fig 13.

Fig 13. Practice Page of Website Learn Hangul

The learning page or “Belajar” on this website is designed for users who are not familiar with Hangul at all. This page features 3 menus: learning consonants, learning vowels, and learning how to write syllables. In addition to displaying the characters or forms of the letters, these menus also show the name of each letter, pronunciation guide, and instructions on how to write the letters.

The consonant learning page is a feature that users can utilize to recognize and learn Hangul consonants. Each letter on this page can be clicked to display a modal popup containing the letter's name, pronunciation rules, and writing method. As explained in the previous chapter, there are 14 consonant letters in Hangul, all of which are displayed on the entire consonant learning page in Fig 14.

Fig 14. Learning Page of Consonant Letters

The next learning page is the vowel letters page, where users can learn and familiarize themselves with Hangul vowel letters. Similar to the previous menu for consonant letters, each letter on this menu can be clicked to display a pop-up modal with information on pronunciation and writing instructions. This menu showcases all 10 basic vowel letters and several expanded vowel letters (4), as shown in Fig 15.

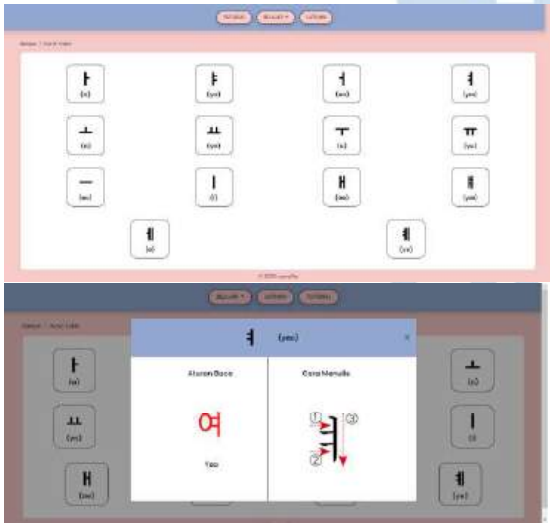


Fig 15. Learning Page of Vowel Letters

The third learning page is the syllable writing page. This page is useful for users to learn about syllable writing rules and the placement of each Hangul letter comprising 2 to 4 letters. This page can be seen in Fig 16.

Fig 16. Learning Page of Syllable Structure

C. *Testing*

In this testing phase, the researcher conducted 5 tests: manual testing, confusion matrix, expert validation, black box, and direct user testing.

1) *Manual Testing*

In this manual testing stage, the researcher tested three models from Table VII. Only the models with the highest accuracy from each training scenario were subjected to manual testing. Out of 182 categories, the researcher selected 10 classes: Hya, Jyeo, Kyu, Tae, Rye, Dyo, Pi, Chu, Ga, and Se.

As shown in Figure 17, to test the model trained with Sobel edge detection data, images processed with Sobel edge detection were used. Similarly, for the model trained with dilation morphology data, images with dilation morphology were used for testing.

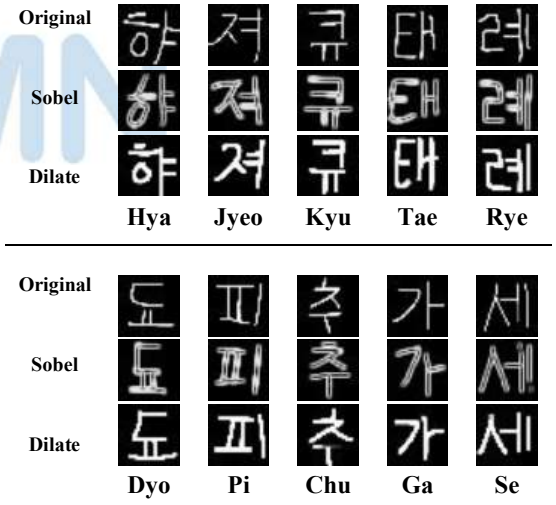


Fig 17. Image for Manual Testing

Each class consists of 10 images handwritten by the researcher and distinct from the dataset, resulting in a total of 100 images for this manual testing. The results of the manual testing shown in Table IX.

TABLE IX. RESULT OF MANUAL TESTING

Model	Class Syllable										Total
	Hya	Jyeo	Kyu	Tae	Rye	Dyo	Pi	Chu	Ga	Se	
A (Original)	9	6	8	6	9	8	6	10	9	6	77
B (Sobel)	6	5	9	5	4	10	7	10	7	7	70
C (Dilate)	9	8	9	6	9	9	6	3	5	8	72

Table IX shows how many times each model correctly predicted images according to their class. As mentioned earlier, each class contains 10 images. Therefore, a number 10 in the table indicates the model correctly predicted all images in that class. Out of 100 images used for testing, Model A, trained with the original dataset, correctly predicted 77 images. Thus, Model A, with an accuracy of 89.2%, is the best model for classifying Korean syllables and is used in the Learn Hangul website..

2) Confusion Matrix

The second test is the Confusion Matrix. Values from the confusion matrix can be used to calculate performance evaluation metrics such as precision, recall, and F1-score. This test aims to determine the percentage of correct predictions made by the model and is performed on the test data. Table X shows the results of the confusion matrix test for each class, with only a portion displayed.

TABLE X. CONFUSION MATRIX RESULT OF EACH CLASS

index	class	precission	recall	F1-score	support
0	A	0.81	0.77	0.79	22
1	Ae	0.91	0.91	0.91	22
2	Ba	0.90	0.86	0.88	22
3	Bae	1.00	0.91	0.95	22
4	Be	1.00	0.91	0.95	22
5	Beo	0.88	1.00	0.94	22
6	Beu	0.92	1.00	0.96	22
7	Bi	0.96	1.00	0.98	22
8	Bo	0.91	0.91	0.91	22
9	Bu	0.77	0.91	0.83	22
10	Bya	1.00	0.91	0.95	22
…	…	…	…	...	…
181	Yu	0.76	0.86	0.81	22

From the confusion matrix values in Table X, the final or average performance scores of Model A can be calculated. Below are the average calculations for precision, recall, and F1-score for the model.

$$Avg\ Precision = \frac{\Sigma\ precision\ seluruh\ class}{Total\ class} \quad (2)$$

$$Avg\ Precision = \frac{163.35}{182} = 0.897$$

Precision is an evaluation metric used to measure how accurate a model is in identifying positive data from all data classified as positive. The results above indicate that Model A has a precision score of 89.7%. A high precision value indicates that the model is effective in avoiding errors in classifying negative data as positive. However, there is still a 10.3% error rate, which can occur because the data used may contain noise, causing the model to learn incorrect patterns and leading to misclassifications.

$$Avg\ Recall = \frac{\Sigma\ recall\ seluruh\ class}{Total\ class} \quad (3)$$

$$Avg\ Recall = \frac{161.54}{182} = 0.8875$$

Recall, also known as sensitivity, is an evaluation metric used to indicate how well a model can accurately find all data belonging to the positive class. From the results above, it shows that Model A has a recall value of 88.8%. A high recall value indicates that the model is quite effective in detecting positive data. However, there is still a 11.2% error rate, which may occur due to variations in features distinguishing between positive and negative data that are not sufficiently understood by the model, or due to insufficient representation of positive data in the training set, which leads to the model being less trained in recognizing that class.

3) Expert Validation

In this testing, the expert (someone who proficient in the Korean language field) checked all content on the website, including verifying the writing of each Hangul letter and the arrangement of syllables displayed on the Learn Hangul website. Additionally, the expert also tested the main menu of the Learn Hangul website, specifically the Practice menu for writing Korean syllables.

In this testing scenario, the expert wrote 20 syllables: 10 in large size and 10 in small size. The expert then checked the classification results to see if they matched the Romanization displayed (classification results) or if there were any errors. The results of the testing with the expert can be seen in Table XI and Table XII

TABLE XI. TESTING RESULT WITH LARGE TEXT SIZE

Syllable written by the expert	*True Label*	*Predicted Label*	*Information*
	A	A	True
	Bya	Ba	False
	Cha	Cha	True
	Do	Do	True
	Ju	Ju	True
	Mi	Mi	True
	Ri	Ri	True
	Sa	Sa	True
	Seo	Seo	True
	Yo	Seo	False

TABLE XII. TESTING RESULT WITH SMALL TEXT SIZE

Syllable written by the expert	*True Label*	*Predicted Label*	*Information*
	Beo	Beo	True
	Chyeo	Ru	False
	Dyu	Chi	False
	Heo	Chi	False
	Ko	Pu	False
	Pi	Ji	False
	Ro	Po	False
	Seo	Seo	True
	U	Ho	False
	Yeo	Ye	False

Based on the results in Table XI, by writing syllables in large and clear sizes, 8 out of 10 syllables were successfully classified correctly into their respective classes. However, the classification of the remaining 2 images was incorrect.

Based on the results from Table XII, where small-sized syllables were written, only 2 out of 10 images were correctly classified, while the other 8 images were classified into incorrect classes. This could happen because the images submitted to the model contained information that differed from what it had learned. When the input differs significantly from what the model has learned, such as images with very small information (Korean syllables) as shown in Table XI, it is likely to affect the classification results.

From this testing, it can be concluded that syllables written in small sizes on the canvas tend to result in misclassification or incorrect classification.

4) Direct User Testing

The last testing involved direct user testing. It was conducted by distributing questionnaires using Google Form to collect user response data. Two questionnaires were given to users: pre-learning and post-learning. The pre-learning questionnaire was completed by users before they tried learning on Learn Hangul website, while the post-learning questionnaire was filled out after they had used Learn Hangul website for learning.

In the pre-learning and post-learning questionnaires on Google Form, there are 5 questions identical to those in Table XIII. This approach aimed to observe changes in users' knowledge of Korean (Hangul) before and after using the Learn Hangul website.

TABLE XIII. QUIZ FOR USER

No.	Question	Point
1	The vowel letter "ㅖ" is placed ... the consonant letter	20
2	The vowel letter "ㅛ" is placed ... the consonant letter	20
3	What is the Latin letter equivalent of "大" ?	20
4	What is the Latin letter equivalent of "ㅍ" ?	20
5	What is the Latin letter of the syllable "져"?	20

Out of 20 participants, 16 showed an improvement in their scores compared to the pre-learning results. However, 4 respondents scored the same as in the pre-learning phase, indicating no improvement for these 4 individuals. The scores of the 20 respondents shown in Table XIV.

TABLE XIV. THE RESPONDENTS' SCORES

No	Res	Score		No	Res	Score	
		Pre	Post			Pre	Post
1.	Res.1	20	80	**11.**	Res.11	80	100
2.	Res.2	0	80	**12.**	Res.12	60	60
3.	Res.3	20	80	**13.**	Res.13	80	100
4.	Res.4	40	60	**14.**	Res.14	40	60
5.	Res.5	40	80	**15.**	Res.15	80	80
6.	Res.6	40	100	**16.**	Res.16	100	100
7.	Res.7	20	80	**17.**	Res.17	0	40
8.	Res.8	40	60	**18.**	Res.18	60	60
9.	Res.9	40	60	**19.**	Res.19	40	80
10.	Res.10	80	100	**20.**	Res.20	40	60

*Res = Respondent

Based on direct user testing, it was found that 80% of respondents experienced an increase in their knowledge of Korean (Hangul) letters after trying to learn them on the Learn Hangul website, while the remaining 20% did not experience any increase or decrease. These figures indicate the success of the Learn Hangul website in benefiting users who wish to learn Korean letters and show that it was fairly well-received by users.

V. CONCLUSION

Based on the results of the conducted research, several conclusions were drawn as follows: The CNN model with LeNet-5 architecture demonstrated the highest accuracy in identifying Korean Hangul syllables at 89.2% and achieved the greatest number of correct classifications in the manual testing conducted by Model A. The performance results of Model A, tested using a confusion matrix, indicated a precision value of 89.7% and a recall of 88.8%. Conversely, the CNN model with the lowest accuracy of 86% was observed in the second training scenario, which utilized dataset B with Sobel edge detection, and this model also recorded the lowest number of correct classifications in the manual testing. In the exercise menu on the Learn Hangul website, it is essential for writing to be clear and in large letters, as determined by expert testing, because it significantly influences the classification results. Lastly, 80% of respondents showed an improvement in their knowledge of Korean Hangul letters, whereas the remaining 20% exhibited no improvement or decline.

ACKNOWLEDGMENT

The author would like to express gratitude to State Polytechnic of Jember and the chief counselor for their support in this research.

REFERENCES

[1] U. Hasanah and M. Kharismawati, "Penggunaan Budaya Pop Korea dalam Proses Pembelajaran Bahasa Korea bagi Mahasiswa dengan Gaya Belajar Campuran," *JLA (Jurnal Ling. Appl.*, vol. 3, no. 1, p. 10, 2020, doi: 10.22146/jla.52060.

[2] Mutiara, N. E. Prasatia, E. Evangeline, and N. H. Marhaeni, "Pengaruh Drama Korea Terhadap Minat Belajar Bahasa Korea Mahasiswa," *Inov. J. Imiah Pengemb. Pendidik.*, vol. 1, no. 2, pp. 33–42, 2023, [Online]. Available: https://ejournal.lpipb.com/index.php/inovasi/article/view/20

[3] N. Leni, "Faktor yang Membuat 7 Negara (Finlandia, Korea Selatan, Hongkong, Jepang, Singapura , Belanda, Kanada) Diakui Memiliki Sistem Pendidikan Terbaik di Dunia dalam Kajian Antropologi dan Matematika," *Pros. Semin. Nas. Mat. dan Pendidik. Mat.*, pp. 219–229, 2019.

[4] J. R. Cho and C. McBride, "Different Cognitive Correlates of Early Learning of Spelling of Different Target Types in Korean Hangul Among First Language Children and Adult Foreign Language Learners," *J. Learn. Disabil.*, vol. 55, no. 2, pp. 138–153, 2022, doi: 10.1177/0022219420978231.

[5] C. Umam and L. B. Handoko, "Convolutional Neural Network (CNN) Untuk Identifkasi Karakter Hiragana," *Pros. Semin. Nas. Lppm Ump*, vol. 0, no. 0, pp. 527–533, 2020, [Online]. Available: https://semnaslppm.ump.ac.id/index.php/semnaslppm/article/view/199

[6] C. A. Lorentius, R. Adipranata, and A. Tjondrowiguno, "Pengenalan Aksara Jawa dengan Menggunakan Metode Convolutional Neural Network," *e-Proceeding Eng.*, vol. 7, no. 1, pp. 2558–2567, 2020.

[7] M. R. Alwanda, R. P. K. Ramadhan, and D. Alamsyah, "Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle," *J. Algoritm.*, vol. 1, no. 1, pp. 45–56, 2020, doi: 10.35957/algoritme.v1i1.434.

[8] K. Seon Jung, M. Min Kyung, P. Sung Tae, and H. Yung, *Buku Standar Bahasa Korea EPS-TOPIK*. 2015.

[9] M. Pastore, "Processing an Alphabetic Syllabary : Investigating the Orthographic Code for Korean Processing an Alphabetic Syllabary : Investigating the Orthographic Code for Korean," 2019.

[10] A. Ajit, K. Acharya, and A. Samanta, "A Review of Convolutional Neural Networks," *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, pp. 1–5, 2020, doi: 10.1109/ic-ETITE47903.2020.049.

[11] T. Tinaliah, "Penerapan Convolutional Neural Network Untuk Klasifikasi Citra Ekspresi Wajah Manusia Pada MMA Facial Expression Dataset," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 4, pp. 2051–2059, 2021, doi: 10.35957/jatisi.v8i4.1437.

[12] Lusiana, A. A. Kurniasari, and I. F. Kusumawati, "Enhancing Accuracy in the Detection of Pneumonia in Adult Patients: An Approach by Using Convolutional Neural Networks BT - Proceedings of the 4th International Conference on Electronics, Biomedical Engineering, and Health Informatics," T. Triwiyanto, A. Rizal, and W. Caesarendra, Eds., Singapore: Springer Nature Singapore, 2024, pp. 593–609.

[13] R. C. Gonzalez and R. E. Woods, "Digital Image Processing," in *A Guide for Machine Vision in Quality Control*, Fourth., Addison Wesley Publishing Company, 2018, pp. 19–44.

[14] P. A. Nugroho, I. Fenriana, and R. Arijanto, "Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia," *Algor*, vol. 2, no. 1, pp. 12–21, 2020.

[15] A. Faizin, M. Lutfi, and Achmyatari, "Perbandingan Arsitektur Lenet Dan Googlenet Dalam Klasifikasi Diabetic Retinopathy Pada Citra Retina Fundus," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 1, pp. 342–347, 2022, doi: 10.36040/jati.v6i1.4581.

[16] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163–170, 2019, doi: 10.33480/pilar.v15i2.658.

[17] M. Sholawati, K. Auliasari, and F. Ariwibisono, "Pengembangan Aplikasi Pengenalan Bahasa Isyarat Abjad Sibi Menggunakan Metode Convolutional Neural Network (Cnn)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 1, pp. 134–144, 2022, doi: 10.36040/jati.v6i1.4507.

[18] L. Widiawati and N. Wulandari, "Akurasi Deteksi Tepi Wajah dengan Metode Robert, Metode Prewitt Dan Metode Sobel," *J. Ilm. MIKA AMIK Al Muslim*, vol. 3, no. 2, pp. 79–87, 2019.

[19] K. Anwariyah, F. Sains, and D. Teknologi, "Deteksi Objek Nomor Kendaraan Pada Citra Kendaraan Bermotor (Detection of Vehicle License Objects in Motorized Vehicle Images)," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 1, no. 4, pp. 311–317, 2020.

ISSN 2355-0082

Ensemble Learning - Random Forest Algorithm to Classify Obesity Level

Tesalonika Abigail[1], Marlinda Vasty Overbeek[2]
[1,2] Dept. of Informatics, Universitas Multimedia Nusantara, Tangerang, Banten
[1]tesalonika.abigail@student.umn.ac.id, [2]marlinda.vasty@umn.ac.id

Accepted 5 July 2024
Approved 15 November 2024

***Abstract*— Obesity is one of the serious global health problems caused by excessive accumulation of body fat. According to the World Health Organization (WHO), the prevalence of obesity has tripled in the last 40 years, with 650 million out of 1.9 billion overweight adults suffering from obesity. Obesity is a non-communicable disease that increases the risks of more dangerous diseases, such as heart disease and cancer. Therefore, early detection of obesity level is crucial. Currently, Body Mass Index (BMI) serves as a measurement indicator, but it tends to overestimate obesity for those with high muscle mass and vice versa, making it ineffective as it only relies on height and weight, without considering body composition and daily activities. To address this limitation, a Random Forest model was developed and selected based on the results of model selection, feature selection, and hyperparameter tuning. This model improved accuracy by 1.4% and was implemented into a web-based system for classifying obesity levels. Evaluation of the model yielded Precision, Recall, F1-Score, and Accuracy of 97%, 97%, 97%, and 96.8%, respectively. Based on these results, it can be concluded that this system is highly effective in classifying obesity levels.**

***Index Terms*— Feature Selection; Hyperparameter Tuning; Model Selection; Obesity Level Classification; Random Forest**

I. INTRODUCTION

Obesity is a serious global health problem caused by excessive accumulation of fat [1]. The World Health Organization (WHO) states that the global obesity rate has tripled in the last 40 years, with over 1.9 billion adults suffering from overweight and 650 million of those classified as obese [2]. Obesity, primarily caused by unhealthy lifestyles, increases the risk of diseases such as type-2 diabetes, cancer, heart disease, hypertension, and stroke [2,3]. Therefore, obesity should not just be seen as a lifestyle issue but as a disease posing significant health risks. Consequently, early detection of obesity levels is crucial [1].

BMI is a common indicator for measuring obesity, but it tends to overestimate obesity in individuals with high muscle mass and underestimate it in those with low muscle mass [4]. This is because BMI relies solely on height and weight, ignoring body composition and daily activities. Alternative approaches by classifying obesity levels based on lifestyle and dietary habits with machine learning which is increasingly popular in medical studies for classification, clustering, and anomaly detection can be used [1,5].

Previous research using the K-Nearest Neighbor (KNN) algorithm to classify obesity levels with the "PIMA Indian Diabetes" dataset achieved a 78.98% accuracy rate, with a recommendation to use other algorithms to improve model accuracy [1]. Another study comparing Random Forest, Decision Tree, Support Vector Machine, and KNN for diabetes classification, found that Random Forest is the best algorithm to classify diabetes with 97.5% accuracy, 97.4% precision, 96.6% recall, and 97% f1-score [6]. These results show that Random Forest performs exceptionally well in classifying diabetes.

To improve model accuracy, this research will use the Random Forest algorithm to classify obesity levels using the "PIMA Indian Diabetes" dataset as the development from the first previous study. This choice is based on its excellent performance in diabetes classification from the second previous study which is closely related to obesity itself. Random Forest itself is widely used in medical classification due to its ability to handle mixed datasets, high efficiency, excellent accuracy, and low error rates [5,7,8]. The expectation is that it will similarly excel in obesity classification just like how it achieved an exceptionally good performance in classifying diabetes.

During development, a model selection by doing feature selection techniques and hyperparameter tuning will be implemented to select the best model with the highest accuracy. The selected model will then be integrated into a system for accurate early detection of obesity levels. Therefore, this research will focus on designing and developing an obesity levels classification system using the Random Forest algorithm to provide high-accuracy classification results.

II. LITTERATURE REVIEW

A. Obesity

Indonesia obesity prevalence has been increasing, with rates reported by the *Badan Penelitian dan Pengembangan Kesehatan* in 2019 rising from 8.6% in

2007 to 11.5% in 2013, and 13.6% in 2018 [9]. Obesity has thus become a significant health issue in Indonesia, contribute to the dual burden of disease [10]. While infectious diseases remain a leading cause of health problems and deaths, non-communicable diseases are also on the rise, necessitating focused prevention and management efforts [10].

B. Machine Learning

Machine learning, a subset of Artificial Intelligence (AI), empowers systems or applications to autonomously learn from data, enhancing their reliability and predictive accuracy without human intervention [11]. One of its main approaches is supervised learning [12]. Supervised learning itself it utilized when datasets include outputs or classes, rely on accurate class assignments for effective model training [12]. The learning process involves dataset partitioning into training and testing subsets, model training to learn feature-class relationship, and evaluation of model performance using testing data [13].

Classification is a common supervised learning method, involve teaching algorithms to categorize data into predefined classes [11,13]. This classification encompasses dual and multiple class distinctions, with binary outcomes in the former and multi-category outcomes in the latter [11]. Random Forest stands out as one of the top algorithms for classification tasks [11].

C. Decision Tree

Decision tree learning is a predictive model approach commonly used in machine learning [14] and is often employed for classification tasks as it doesn't require extensive information [15]. It takes the form of a tree with subdivisions that repeatedly divide data into smaller subsets based on specific criteria [16]. It is constructed using nodes and branches with root as the top node of the tree for tree construction [16]. Branches stemming from the root, which still have child branches, are called internal nodes thats connected to the root, other internal nodes, and leaf nodes through branches [16]. Leaf nodes, on the other hand, are branches with no further branches, representing the decision outcomes of the decision tree. Fig 1 illustrates the structure of a decision tree [16].

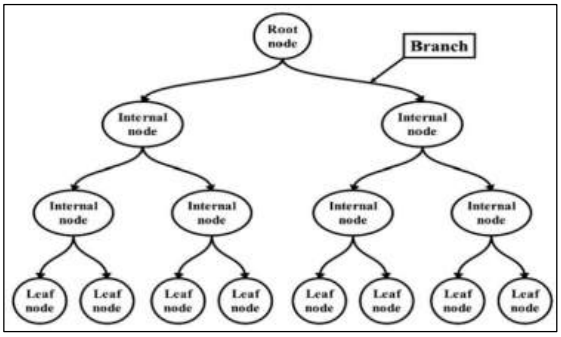


Fig. 1. Decision Tree Structure [16]

D. Random Forest

Random Forest is an ensemble algorithm [17], stemming from the development of decision trees. It operates by constructing multiple decision trees for prediction and making final decisions based on the most voted outcome [18]. Below are the steps in creating a random forest model [19].

1. Selecting n random samples from the training data to be used in the dataset bootstrapping process.
2. Selecting m random features with m < p, where p is the total number of features, to select the best feature in each node as a node splitting separator with optimization criteria, such as "gini", "entropy", or "log_loss".
3. The second process will continue until the minimum number of observations at the node is reached.
4. The entire process above will be repeated until k decision trees are formed. From each decision tree, one class will be obtained and used in the voting process to determine the final classification class

The overview/ illustration of how the Random Forest algorithm works can be seen in Fig 2.

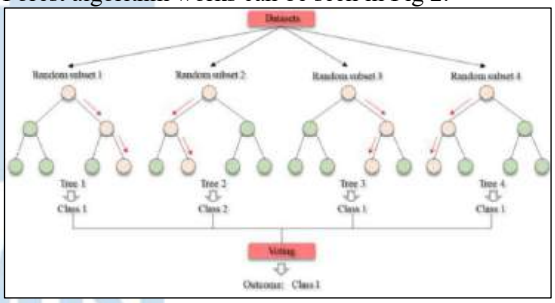


Fig. 2. Flow of the Random Forest Algorithm [20]

Next, here's the formula for calculating the Random Forest algorithm to obtain the final classification result. The formula represents the predicted class of the b-th tree for data x [21], which will be used in the voting process by considering the classification results from the 1st tree to the Bth tree to determine the final classification outcome based on the majority vote.

$$\hat{C}_{rf}^{B} = majority\ vote\{\hat{C}_b(x)\}_1^B \quad (1)$$

However, this algorithm inherently cannot achieve optimal performance if relying solely on one value for each hyperparameter used, as it may not be the optimal value. This is because its performance heavily depends on the hyperparameters used [22]. To address this limitation, hyperparameter tuning can be conducted to find the most optimal hyperparameter set to be used for the model construction through trial and error testing of all hyperparameter combinations [22].

E. Randomized Search

Randomized Search is a hyperparameter tuning technique that utilizes randomness or probability, because it explores the search space randomly using only a few randomly selected hyperparameter settings without following any specific pattern or sequence [23]. Thus, Randomized Search can quickly evaluate several important combinations to achieve optimal results, making computational time more efficient.

The Randomized Search algorithm is evaluated using cross-validation methods [23], which divide the data into k subsets called folds. The evaluation process is repeated k times, using 1 fold as validation data to test the model after being trained on the other k-1 folds as training data. Below is the pseudocode illustrating the workflow of Randomized Search [23].

Algorithm 1: Randomized Search Algorithm

```
Data: Input data
Result: Output Result
Initializing criterion, max_depth, n_estimators, max_features,
 min_samples_split, and max_leaf_nodes hyperparameters;
Initializing estimators, search space, number of iterations, and
 number of k-folds;
while Stop criteria is not fulfilled do
    Randomly select parameters from the search space;
    Split the dataset into K-Folds evenly;
    for each fold k in K-Fold do
        Set k as the validation data, the rest as training data;
        Training the data using the estimator with the
         hyperparameter combination;
        Evaluate model performance;
        Calculate the average score from each fold;
    end
end
Return the best hyperparameters;
```

F. Confusion Matrix

The confusion matrix is a metric used to assess the performance of classification outcomes and is easy to interpret because the model's performance can be directly observed from the distribution of values in the confusion matrix [24]. A good model will only have values along the diagonal line of the matrix and vice versa [24]. Confusion matrix itself is divided into binary and multi-class classification [25].

The binary classification confusion matrix is a 2x2 matrix with positive and negative labels for each actual class and classification outcome [14]. Each cell represents the outcome between the actual and predicted class [24]. There are four terms used to represent the classification outcomes [14]. True Positive (TP) represents the number of classifications predicted as positive by the model that match reality [26]. True Negative (TN) represents the number of classifications predicted as negative by the model that match reality [26]. False Positive (FP) represents the number of classifications predicted as positive that should have been classified as negative [26]. False Negative (FN) represents the number of classifications predicted as negative that should have been classified as positive [26]. Table I shows the representation of the confusion matrix for binary classification.

TABLE I. CONFUSION MATRIX FOR BINARY CLASSIFICATION

	Predicted Positive	**Predicted Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

From the confusion matrix, accuracy and several other evaluation metrics can be calculated, including precision, recall (or sensitivity), and F1-score. Here are the formulas for calculating these metrics [27].

- Accuracy : used to measure how well the model predicts correctly.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

- Precision : used to measure the model's ability to predict positive class correctly from all positive predictions.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

- Recall : used to measure the model's ability to predict true positive cases.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

- Recall : used to measure the model's ability to predict true positive cases.

$$F1 - Score = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recall} \quad (5)$$

Meanwhile, in multi-class classification, the confusion matrix has a dimension of $n \times n$, where n is the number of classes and $n > 2$. Therefore, the characterization of TP, TN, FP, and FN does not apply in multi-class classification confusion matrix [14]. However, analysis for classification results can still be conducted by focusing the analysis on a specific class, as shown in Table II with class C_2 as the main focus of analysis [14].

TABLE II. MULTI-CLASS CONFUSION MATRIX

	Predicted C_1	**Predicted C_2**	**…**	**Predicted C_N**
Actual C_1	$C_{1,1}$	FP	…	$C_{1,N}$
Actual C_2	FP	TP	…	FN
…	…	…	…	…
Actual C_N	$C_{N,1}$	FP	…	$C_{N,N}$

III. METHODS

A. Collecting Data

In this study, the "Estimation of obesity levels UCI" dataset obtained from the Kaggle website is used. The dataset was collected through a 30 day online survey accessible via a website platform with the subject of

people from Mexico, Peru, and Colombia, aged 14-61, regarding various dietary patterns and physical conditions [28]. In addition, the data were also synthetically generated using the SMOTE filter to balance the sample sizes between majority and minority classes [28]. SMOTE is an oversampling technique that generates new synthetic samples for the minority class by interpolation [29]. It first selects a random sample from the minority class and then selects another random sample from its nearest neighbors to perform interpolation by estimating new values from both samples [29].

From this entire process, a dataset with a total of 2111 instances and 17 attributes was created [28]. The features include gender, age, height, weight, family history of overweight, frequent consumption of high-caloric food (FAVC), frequency of vegetable consumption (FCVC), number of main meals (NCP), consumption of food between meals (CAEC), smoking habits (SMOKE), daily water consumption (CH2O), calories consumption monitoring (SCC), physical activity frequency (FAF), time using technology devices (TUE), alcohol consumption (CALC), and transportation used (MTRANS). Additionally, there are ”NObeyesdad” class with values such as ”Insufficient Weight,” ”Normal Weight,” ”Overweight Level I,” ”Overweight Level II,” ”Obesity Type I,” ”Obesity Type II,” and ”Obesity Type III.” The feature data were obtained from questionnaires and the SMOTE filter, while the class data were obtained from BMI calculations [28]. Therefore, this dataset combines BMI values and features that can influence obesity levels with the aim of addressing the limitations of BMI as an indicator of obesity measurement.

B. *Pre-processing*

In this stage, data checking and deletion will be performed to remove all null and duplicated instances, ensuring data consistency, uniqueness, and preventing bias in the analysis. Furthermore, the processed data will be re-represented to be understood by machine learning by performing Label Encoding to convert all categorical data into numerical data.

C. *Model Construction*

This stage marks the initial step in the model selection process, starting with the construction and accuracy check of the initial model using the Random Forest algorithm to classify obesity levels and obtain the initial accuracy level of the model. This process involves dividing the dataset using a holdout scenario (8:2), which splits the dataset into training and testing data [30]. From the splitting process, 80% of the data is used to train the model with the training data, and the remaining 20% is used to evaluate how well the model classifies with the testing data. Subsequently, the model’s accuracy is checked using the “*accuracy_score()*” function to obtain the initial accuracy level of the model.

D. *Feature Selection and Model Construction*

This stage marks the second step in model selection process, starting with feature selection using the ANOVA (Analysis of Variance) F-score method aided by the “*SelectKBest()*” function. The feature selection process iterates 16 times in accordance to the dataset’s total features. In each iteration, data is split using an 8:2 holdout ratio to construct the second model for 16 times, which will be trained with the chosen features. The accuracy of the new model is then assessed using “*accuracy_score()*”. The optimal number and the best-selected features are determined based on the model yielding the highest accuracy and will be used as the question lists in the classification system.

E. *Hyperparameter Tuning and Model Construction*

This stage marks the third stage in the model selection process, starting with hyperparameter tuning to find the best combination of hyperparameter values for optimal model classification performance. Hyperparameter tuning is performed using Randomized Search aided by the “*RandomizedSearchCV()*” function with 10-fold cross-validation, as it offers the most optimal k-fold value [23]. After that, the data is divided again into 2 parts using the 8:2 holdout scenario, which will be trained using the best combination of hyperparameter values obtained from the tuning results. Subsequently, the accuracy of this new model is assessed using the “*accuracy_score()*” function. Since this stage marks the final step in the model selection process, the next step involves comparing all three models to identify the one with the highest accuracy. This step is crucial as the best-performing model is chosen for further evaluation using various metrics and implementation into the classification system.

F. *Model Performance*

In this stage, firstly, the multi-class classification confusion matrix will be used to observe the distribution of predicted class results generated by the model against the actual classes. Then, further evaluation will assess the overall model performance for each class based on precision, recall, and F1-score, along with macro, weighted, and micro averages from the “*classification_report()”* function and micro average calculation. Lastly, log loss metrics will be used to check the model’s error level in predicting class probabilities because it can be used for multi-class classification problems and aligns with the Random Forest algorithm’s operation, which also computes class probabilities at each node for classification.

G. *System Development*

In this stage, firstly, a flowchart will be created to outline the system’s workflow. Then, the system will be designed in prototype form to design the user interface and system flow to achieve the final classification results for each user. The system will be made like an online form with three pages according to each group

of questions, namely personal data, family health and eating habits, and daily routines, with a total of 16 questions corresponding to the original number of features in the dataset and will be made in Bahasa Indonesia. The final classification results will be displayed in a popup, which will either show an error message if an error occurs during the classification process because of the required fields validation error or show the classification results if the process is successful.

IV. Result and Analysis

A. Collecting Data

The first step in building the classification model is to read and store the dataset to be used. This dataset is taken from the Kaggle.com platform entitled "Estimation of Obesity Levels UCI" dataset consists of 2111 data with a total of 17 attributes. The implementation results can be seen in Fig 3.

Fig. 3. Data Collection from Kaggle.com

B. Model Construction

In this stage, the dataset is split into training and testing data using the "*train_test_split()*" function, with 20% of the data used for testing data. Then, the construction of the Random Forest model will start with training using the training data first to find relationships/ patterns between features and classes. After that, the model will use this trained data to predict on the testing data. In the final step, the accuracy of the model is checked by comparing the true classes with the classification results, resulting in an accuracy score of 95.4%. The implementation results can be seen in Fig 4.

Fig. 4. First Model Construction Code

C. Feature Selection and Model Construction

In this stage, the model will be rebuilt for the second time after feature selection to improve the accuracy obtained previously. This process is repeated 16 times to determine the number of features that produce the highest accuracy of the model. Feature selection technique is performed using "*SelectKBest()"* function which retains only the top *k* features in the dataset by involving the F-score analysis of variance (ANOVA) to evaluate the significance of the relationship between each feature and the class using f-score. Then, it will learn the patterns and relationships between features and classes and select only the top *k* features based on the f-score from the learning process.

In this loop, model building is also performed using different datasets in each iteration from *k=1* to *k=16* to select the optimal number and the best features. In its construction, the model will learn and predicts classes from the training and testing data in each iteration to check the model's accuracy score to determine the optimal number of features. The implementation code is shown in Fig 5.

Fig. 5. Feature Selection and Second Model Construction

Next, the *k* value and its corresponding accuracy, which have been stored, will be displayed collectively for easier comparison. From this feature selection result, it is known that the optimal number of features is 12 with an accuracy of 96.4%. The implementation result is shown in Fig 6.

Fig. 6. Accuracy Summary at each Iteration

In the previous stage, it was determined that the optimal number of features is 12, so the next step is to

find out what the selected features are and retrain the model with *k=12,* because the model was last trained with 16 features (*k=16*) from the iteration result. The initial step in this process is to redefine the "*SelectKBest()"* object with *k=12*, then perform the feature selection. From the feature selection result, the F-score for each selected feature is displayed using. Then, the model will be rebuilt using the dataset from feature selection results. The selected features include "Gender", "Age", "Height", "Weight", "family history with overweight", "FAVC", "FCVC", "NCP", "CAEC", "SCC", "CALC", and "MTRANS". Meanwhile, "SMOKE", "CH2O", "FAF", and "TUE" are not used. From the model building result, an accuracy of 96.4% is obtained, indicating that the implementation of feature selection can increase the classification accuracy of the model by 1%, from 95.4% to 96.4%. The implementation result can be seen in Fig 7.

Fig. 7. Selected Features and Second Model Reconstruction

D. *Hyperparameter Tuning and Model Construction*

In this stage, the model will be rebuilt for the third time after performing hyperparameter tuning to improve the accuracy obtained previously. The hyperparameters used are "*criterion*", "*max_depth*", "*n_estimators*", "*max_features*", "*min_samples_split*", and "*max_leaf_nodes*". Criterion determines the criteria for measuring the quality of splits in tree building. Max_depth determines the maximum depth of the tree. *n_estimators* determines the number of decision trees to be built. *max_features* determines the maximum number of features to consider for splitting a node. *Min_samples_split* determines the minimum number of samples required to split a node. *Max_leaf_nodes* determines the maximum number of leaf nodes in the tree. Fig 8 represents the hyperparameters used along with their respective values.

```
param_dist = {
    'criterion': ['gini', 'entropy', 'log_loss'],
    'max_depth': [i for i in range(1, 101)], # 101
    'n_estimators': [i for i in range(10, 201)], # 201
    'max_features': ['sqrt', 'log2', None],
    'min_samples_split': [i for i in range(2, 6)], # 6
    'max_leaf_nodes': [i for i in range(10, 501)] # 501
}
```

Fig. 8. Hyperparameter Tuning Values

Next, hyperparameter tuning is performed using the values in Fig 8 using the "*RandomizedSeachCV()"* function to obtain the best combination of values for each hyperparameter. Firstly, the Randomized Search is initialized with several parameters. Then, the initialized result is used to train the model, which simultaneously searches for parameters to obtain the best parameter combination. From the tuning result, the best hyperparameter combination values for *n_estimators, min_samples_split, max_leaf_nodes, max_features, max_depth, and criterion* are 131, 2, 361, "*None*", 68, and "*log_loss*" respectively, as seen in Fig 9.

Fig. 9. Hyperparameter Tuning with Randomized Search

Then, the model is rebuilt for the third time to improve the accuracy from the previously obtained value. This process involves learning and predicting by utilizing the best hyperparameter results obtained earlier. From this model building, a new accuracy value of 96.8% is obtained that successfully increases the model's accuracy by 0.4%. This indicates that the implementation of feature selection and hyperparameter tuning can improve the model's accuracy in classification. The implementation result can be seen in Figure 10.

Fig. 10. Third Model Construction using the Tuning Results

Based on the results of the entire model development process, it can be observed that the use of feature selection and hyperparameter tuning successfully increased the model accuracy by a total of 1.4%. This improvement amounted to 1% and 0.4% in each respective stage. It indicates that implementing feature selection and hyperparameter tuning can enhance the model's accuracy in classifying obesity levels. The summary of accuracy from all model developments can be seen in the Table III.

TABLE III. SUMMARY OF THE ACCURACY LEVEL FROM ALL MODEL

Model	Feature Selection	Hyper-parameter Tuning	Accuracy	Increase
1	-	-	95.4%	-
2	✓	-	96.4%	1%
3	✓	✓	96.8%	0.4%

The third model achieved the highest accuracy of 96.8% from the model selection process. Therefore, it is chosen as the best selected model in the study that will be evaluated and implemented into the obesity level classification system.

E. Model Performance Evaluation

a) Evaluation with Confusion Matrix: This evaluation process begins with the calculation of confusion matrix, which will provide results in the form of a 2D array. This calculation process continues until the calculation for the last class is completed. The implementation result of the TP, TN, FP, and FN values for each class can be seen in Table IV.

TABLE IV. TP, TN, FP, AND FN VALUES FOR EACH CLASS

	TP	FN	FP	TN
Insufficient_Weight	45	1	2	370
Normal_Weight	45	6	1	366
Overweight_Level_I	54	2	5	357
Overweight_Level_II	58	2	3	355
Obesity_Type_I	74	1	2	341
Obesity_Type_II	69	1	0	348
Obesity_Type_III	60	0	0	358

From the summary in Table IV, it's evident that the largest misclassified class involves 4 samples incorrectly classified as "*Overweight_Level_I*" from "*Normal_Weight*" class. To conduct further checks, the 4 instances with its original values to before label encoding was performed from the most misclassified "*Normal_Weight*" class from the testing data will first be obtained.

b) Evaluation with Classification Report: From the result of classification report, several evaluation metrics are obtained, including the precision, recall, and f1-score for each class, as well as the accuracy, macro average, and weighted average values from the overall evaluation of the model. Precision represents how well the model classifies positive classes out of all positive predictions. Recall represents how well the model classifies positive classes out of all actual positive classes. F1-score represented how well the model balances precision and recall in classifying positive classes. Accuracy represents how accurate the model is in overall predictions. Macro avg represents the average evaluation metrics based on the precision, recall, and f1-score values from each class, without considering class distribution. Weighted Avg represents the average evaluation metrics based on the precision, recall, and f1-score values from each class, considering class distribution.

Additionally, there is one evaluation metric not included in classification report, which is the micro average, representing the average evaluation metrics based on the TP, FP, and FN values from all classes without considering class distribution. The implementation result along with the model evaluation can be seen in Fig 11, and the result of the classification report and the calculation of the micro average can be seen in Fig 12.

```
from sklearn.metrics import classification_report

print('Report: ' + classification_report(y_test2, y_pred3))
```

```
Report:                      precision    recall  f1-score   support

Insufficient_Weight       0.96      0.98      0.97        46
      Normal_Weight       0.98      0.88      0.93        51
     Obesity_Type_I       0.97      0.99      0.98        75
    Obesity_Type_II       1.00      0.99      0.99        70
   Obesity_Type_III       1.00      1.00      1.00        60
 Overweight_Level_I       0.92      0.96      0.94        56
Overweight_Level_II       0.95      0.97      0.96        60

           accuracy                           0.97       418
          macro avg       0.97      0.97      0.97       418
       weighted avg       0.97      0.97      0.97       418
```

Fig. 11. Classification Report Result

```
from sklearn.metrics import precision_recall_fscore_support

# Menghitung precision, recall, dan F1-score menggunakan micro averaging
precision, recall, f1_score, _ = precision_recall_fscore_support(y_test2, y_pred3, average='micro')

# Menampilkan nilai metrik evaluasi
print("Micro average metrics:")
print("Precision: {:.2f}".format(precision))
print("Recall: {:.2f}".format(recall))
print("F1-score: {:.2f}".format(f1_score))
```

```
Micro average metrics:
Precision: 0.97
Recall: 0.97
F1-score: 0.97
```

Fig. 12. Micro Average Calculation Result

c) Evaluation with Log Loss: After evaluating the model's prediction results with Confusion Matrix and Classification Report, the next step is evaluating the model's uncertainty/ error level in predicting class probabilities. Log loss is a evaluation metric that measures how well the model predicts the true class probabilities. The smaller the value, the better the model's performance. In other words, log loss measures how accurate and reliable a model is in providing probabilities that a sample/entity belongs to a certain class. The steps calculating the probability values for each entity in each class, then use it to calculate the log loss. From this calculation, a log loss value of 0.09 or 9% is obtained. Fig13 represents the implementation result of log loss.

```
from sklearn.metrics import log_loss

y_proba = rfc3.predict_proba(X_test2)
logloss = log_loss(y_test2, y_proba)

print(f'Log loss {logloss}')
```

```
Log loss 0.09504090778292554
```

Fig. 13. Log Loss Result

F. User Interface Implementation

Overall, the system's interface implementation will resemble the prototyping results, but with only 12 questions as this is the optimal number of features determined by feature selection. The first page, will include questions about gender, age, height, weight, and an additional question for the user name to be displayed with the classification result. The second page will include questions about family histories of overweight, frequency of eating high-calorie foods, frequency of vegetable consumption, and the number of main meals per day. The third page will include questions about snacking frequency, daily calorie

monitoring, alcohol consumption frequency, and usual mode of transportation.

Additionally, there will be an additional feature on pages 2 and 3 to view more information about non-quantifiable answer options like “Sometimes”, “Rarely”, etc. This aims to clarify and standardize these options, which can vary between individuals. Furthermore, the classification result popup will now also include personalized health suggestions based on each user’s classification level. These additions were derived from an online interview with a doctor called dr. Jesslyn Valentina, M.M., conducted on April 16, 2024. Below is a detailed explanation of the three pages and the classification result popup in this classification system.

In the first page, users can fill in the five fields and navigate through the pages. Fig 14 represents the implementation for page 1.

Fig. 14. Implementation Result for Page 1

The second page contains a total of 4 questions as previously described. In page 2, users can also fill in the four fields and navigate through the pages. Fig 115 represents the implementation for page 2.

Fig. 15. Implementation Result for Page 2

The answer options for third question on the second page is unmeasurable, and this question is one of the questions that needs the “view more information” feature that was explained before that was obtained from the online interview with a doctor. Fig16 represents a more detailed explained for the third question.

Fig. 16. Detailed Information for the Third Question on the Second Page

The third page contains a total of 4 questions as previously described. In page, users can also fill in the four fields and navigate through the pages. In this page, there is also the submit button “Cek Hasil” to process all inputs and returns the obesity level classification result. Fig 17 represents the implementation for page 3.

Fig. 17. Implementation Result for Page 3

The answer options for first and third questions on the third page are also unmeasurable and needs the “view more information” feature that was explained before. Fig 18 and 19 represents a more detailed explained for the first and third question respectively.

Fig. 18. Detailed Information for the First Question on the Third Page

Fig. 19. Detailed Information for the Third Question on the Third Page

Then, these are the user interface for the error popup because of the failed validation in required fields and server/ API error respectively in Fig 20 and 21.

Fig. 20. Error Popup Interface for Required Field Validation Error

Fig. 21. Error Popup Interface for Server/ API Error

Then, these are the user interface for the success popup for class "Normal_Weight" and all other 6 classes that outputs the classification result and the personalized suggestions based on the obtained obesity level. The icon/ picture representing insufficient weight, normal weight, overweight, and obesity was obtained from the flaticon.com website. The personalized suggestions for each level are also obtained from the online interview. Here is the implementation respectively in Fig 22 and Fig 23.

Fig. 22. Success Popup Interface for Class "Normal_Weight"

Fig. 23. Success Popup Interface for Six Other Classes

V. CONCLUSION

This research shows that the Random Forest algorithm can effectively classify obesity levels. The model also successfully implemented into a system using HTML, CSS, JS, and Flask with the assistance of the Pickle module, accurately processes user input and provides reliable classification results and health advice. The model itself exhibits excellent performance with 96.8% accuracy, 97% precision, 97% recall, and 97% f1-score. However, it shows a relatively high misclassification of "Normal_Weight" into "Overweight_Level_I" class.

For further development, the previously ignored features from the feature selection process can be combined to enhance the model performance. Additionally, using or collecting a new dataset reflecting Indonesian adults' dietary and daily habits can also be conducted to better match local characteristics. Lastly, a classification history features can also be added to the system to allows users to track and monitor their own health progress.

REFERENCES

[1] A. M. S. I. Dewi and I. B. G. Dwidasmara, "Implementation of the k-nearest neighbor (knn) algorithm for classification of obesity levels," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 9, 2020.

[2] G. Castelnuovo, B. D. Cuevillas, S. Navas-Carretero, and J. A. Mart´ ınez, "Body fat mass assessment and obesity classification: A review of the available methods for adiposity estimation," *Progress in Nutrition*, vol. 23, 2021.

[3] S. Jeon, M. Kim, J. Yoon, S. Lee, and S. Youm, "Machine learningbased obesity classification considering 3d body scanner measurements," *Scientific Reports*, vol. 13, 2023.

[4] K. M. Heinrich, K. G. Gurevich, A. N. Arkhangelskaia, O. P. Karazhelyaskov, and W. S. Poston, "Despite low obesity rates, body mass index under-estimated obesity among russian police officers when compared to body fat percentage," *International Journal of Environmental Research and Public Health*, vol. 17, 2020.

[5] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak smote terhadap kinerja random forest classifier berdasarkan data tidak seimbang," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, 2022.

[6] M. Phongying and S. Hiriote, "Diabetes classification using machine learning techniques," *Computation*, vol. 11, 2023.

[7] B. O. Macaulay, B. S. Aribisala, S. A. Akande, B. A. Akinnuwesi, and O. A. Olabanjo, "Breast cancer risk prediction in african women using random forest classifier," *Cancer Treatment and Research Communications*, vol. 28, 2021.

[8] R. Susetyoko, W. Yuwono, E. Purwantini, and N. Ramadijanti, "Perbandingan metode random forest, regresi logistik, na¨ ıve bayes, dan multilayer perceptron pada klasifikasi uang kuliah tunggal (ukt)," *Jurnal Infomedia*, vol. 7, 2022.

[9] T. Kok, V. Wiriantono, J. Bakhriansyah, and L. Aditama, "The factors affecting the occurrence of obesity in college students," *Unnes Journal of Public Health*, vol. 12, 2023.

[10] D. S. Harbuwono, L. A. Pramono, E. Yunir, and I. Subekti, "Obesity and central obesity in indonesia: Evidence from a national health survey," *Medical Journal of Indonesia*, vol. 27, 2018.

[11] S. Madakam, T. Uchiya, S. Mark, and Y. Lurie, "Artificial intelligence, machine learning and deep learning (literature: Review and metrics)," *Asia-Pacific Journal of Management Research and Innovation*, vol. 18, 2022.

[12] A. Alanazi, "Using machine learning for healthcare challenges and opportunities," 2022.

[13] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Introduction to machine learning, neural networks, and deep learning," *Translational Vision Science and Technology*, vol. 9, 2020.

[14] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass confusion matrix reduction method and its application on net promoter score classification problem," *Technologies*, vol. 9, 2021.

[15] H. Khalid, A. Khan, M. Z. Khan, G. Mehmood, and M. S. Qureshi, "Machine learning hybrid model for the prediction of chronic kidney disease," *Computational Intelligence and Neuroscience*, vol. 2023, 2023.

[16] C. S. Lee, P. Y. S. Cheang, and M. Moslehpour, "Predictive analytics in business analytics: Decision tree," *Advances in Decision Sciences*, vol. 26, 2022.

[17] S. Hamsa, I. Shahin, Y. Iraqi, and N. Werghi, "Emotion recognition from speech using wavelet packet transform cochlear filter bank and random forest classifier," *IEEE Access*, vol. 8, 2020.

[18] C. A. Bugeac, R. Ancuceanu, and M. Dinu, "Qsar models for active substances against pseudomonas aeruginosa using disk-diffusion test data," *Molecules*, vol. 26, 2021.

[19] Sriyanto and A. R. Supriyatna, "Teknika 17 (1): 163-172 prediksi penyakit diabetes menggunakan algoritma random forest," *Jurnal Teknika*, vol. 17, 2023.

[20] J. Yang, J. Gong, W. Tang, Y. Shen, C. Liu, and J. Gao, "Delineation of urban growth boundaries using a patch-based cellular automata model under multiple spatial and socio-economic scenarios," *Sustainability (Switzerland)*, vol. 11, 2019.

[21] T. J. Kiely and N. D. Bastian, "The spatially conscious machine learning model," *Statistical Analysis and Data Mining*, vol. 13, 2020.

[22] I. M. M. Matin, "Hyperparameter tuning menggunakan gridsearchcv pada random forest untuk deteksi malware," *MULTINETICS*, vol. 9, 2023.

[23] F. Rahmadayana and Y. Sibaroni, "Sentiment analysis of work from home activity using svm with randomized search optimization," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, 2021.

[24] H. Yun, "Prediction model of algal blooms using logistic regression and confusion matrix," *International Journal of Electrical and Computer Engineering*, vol. 11, 2021.

[25] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, 2018.

[26] T. F. Monaghan, S. N. Rahman, C. W. Agudelo, A. J. Wein, J. M. Lazar, K. Everaert, and R. R. Dmochowski, "Foundational statistical principles in medical research: Sensitivity, specificity, positive predictive value, and negative predictive value," 2021.

[27] H. Agatha, F. Putri, and A. Suryadibrata, "Sentiment Analysis on Song Lyrics for Song Popularity Prediction Using BERT", Ultimatics : Jurnal Teknik Informatika, vol. 15, no. 2, pp. 99-105, Jan. 2024.

[28] F. M. Palechor and A. de la Hoz Manotas, "Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from colombia, peru and mexico," *Data in Brief*, vol. 25, 2019.

[29] Yusuf Gregorius Neven and Gunawan Dennis,SMOTE and Random Forest Algorithm for Detecting DoS Attacks in Wireless Sensor Networks,"ICIC Express Letters, Part B: Applications",21852766,ICIC International, 2022,13,07,715,https://cir.nii.ac.jp/crid/1390855190129179776,https://doi.org/10.24507/icicelb.13.07.715.

[30] F. Tempola, R. Rosihan, and R. Adawiyah, "Holdout validation for com parison classfication na¨ ıve bayes and knn of recipient kartu indonesia pintar," *IOP Conference Series: Materials Science and Engineering*, vol. 1125, 2

ISSN 2355-0082

Implementation Of Heuristic Evaluation Method For Evaluation And Recommendations UI/UX Design Improvements On The Cinepolis Website

Cindy Aristawati[1], Fenina Adline Twince Tobing[2]*, Eunike Endariahna Surbakti[3], Jimmy Peranginangin[4], Anjar Pinem[5]
[1,2,3]Informatics Department, Universitas Multimedia Nusantara, Tangerang, Indonesia
[3,4]Informatics Engineering, Universitas Mandiri Bina Prestasi, Medan, Indonesia
[1]cindy.aristawati@student.umn.ac.id, [2]fenina.tobing@umn.ac.id*, [3]eunike.endariahna@umn.ac.id
[4]jimmy@umbp.ac.id, [5]anjar@umbp.ac.id

Accepted 30 July 2024
Approved 16 January 2025

***Abstract*— UI/UX is one of the most important elements of a website. One of the tasks of UI/UX is to make it easier to achieve a goal that the user wants. Cinépolis is a cinema that has been established in Indonesia since 2014. Cinépolis then launched its own website to make it easier for users to view movie information and order tickets. Based on the questionnaires that have been distributed and calculated using the System Usability Scale or SUS method, the Cinépolis website gets a score of 54.03 and is below the SUS standard of 68. The predicate obtained from the Cinépolis website is grade D with the predicate Poor. Heuristics are methods for finding interface problems to improve usability and user experience. The joint evaluation of 2 evaluators showed that there were 20 problems on the Cinépolis website based on 10 heuristic principles, while the evaluation of the Cinépolis website improvement prototype with 1 other evaluator found 5 problem findings based on 10 heuristic principles on the Figma prototype. The prototype that has been implemented gets a final score of 88.01 using the SUS calculation based on the questionnaire data that has been distributed. The final predicate obtained from the Cinépolis repair website is grade A with the predicate of Excellent.**

***Index Terms*— Evaluation; Figma; Heuristic; SUS; UI/UX; Website**.

I. BACKGROUND

Movie is one of the most widely enjoyed entertainment media by the public [1]. There are many alternative choices that we can choose to be able to enjoy a movie, one of which is by watching a movie in the cinema. Cinépolis is a cinema from Mexico that has been operating in Indonesia since 2014 [2]. To support its distribution, Cinépolis has created a website for cinemas in Indonesia with the aim that customers can see currently showing and upcoming films, and can order tickets online [2]. A good web-based application is an application that can interact well with its users [3]. Two important elements that determine this are User Interface (UI) and User Experience (UX). A good UI will provide user convenience in seeing the appearance and understanding the functionality of the application. Meanwhile, good UX can provide ease of use of applications for users [3]. Based on daily visitor checks carried out using the https://www.visitorsdetective.com/ website, on 3 February 2022 daily visitors to the Cinépolis website were 732 visitors, on 5 February 2022 visitors increased to 1,751 visitors and on 7 February 2022 decreased again to 1,084 visitors. Based on the drastic change in the number of visitors, the Cinépolis website was chosen as the object of this research. To find out more about the user experience when using the Cinépolis website, a form containing questions from the System Usability Scale (SUS) table was distributed which has been translated into Indonesian [4]. From a total of 31 respondents [5] and with calculations using the SUS method, the score obtained by the Cinépolis website is 54.03. So that the Cinépolis website is below the SUS standard score of 68 and is in grade D, namely the Poor category [6]. This problem can be brought to the attention of Cinépolis to re-evaluate its web appearance. Based on the advantages and disadvantages of previous research, Heuristic Evaluation is the most suitable method to be used in this study, because the Heuristic method can provide fast and accurate results in finding problems found on the website being researched.

A. Formulation of the Problem

Based on the background above, it can be formulated several problems that will be solved in this research, namely:

1) How is the recommendation design of the Cinépolis website using the Heuristic Evaluation method?
2) What is the level of user satisfaction with the Cinépolis cinema repair website using the SUS method?

B. Scope of Problem

The following are some limitations for the scope of the research to be carried out:

1) The website being evaluated is the Cinépolis cinema website in desktop view.
2) Recommend UI/UX improvements on the appearance of the next Cinépolis cinema website.
3) The length of time for observing the problem of the Cinepolis cinema website is within a week, from February 1, 2022 to February 7, 2022.

C. Purpose and Objectives

In accordance with the problems that have been formulated, the objectives of this research are as follows:

1) Designing recommendations for improving the Cinépolis cinema website using the Heuristic Evaluation method.
2) Measuring the level of user satisfaction with the Cinépolis cinema repair website using the SUS method.

D. Advantages of Research

The advantages of this research are:

1) Knowing the shortcomings and display problems of the Cinépolis website.
2) Provide convenience and a high impression of comfort in using the Cinépolis website for users.

I. theoretical basis

The following are theories and definitions that explain every aspect involved in this research.

A. Usability

Usability is the ability of a software to assist users in completing a task. Usability quality is defined in 5 components [12]:

1) Learnability
2) Efficiency
3) Memorability
4) Errors
5) Satisfaction

B. System Usability Scale (SUS)

System Usability Scale or SUS is a method to test the usability of an application by providing a table containing 10 questions to the user. The questions consist of 5 positive questions and 5 negative questions and are rated on a scale of 1 (Strongly disagree) - 5 (Strongly agree). Here are 10 SUS questions [14]:

1) I think that I would like to use this system.
2) I found the system unnecessarily complex.
3) I thought the system was easy to use.
4) I think that I would need the support of a technical person to be able to use this system.
5) I found the various functions in the system were well integrated.
6) I thought there was too much inconsistency in this system.
7) I would imagine that most people would learn to use this system very quickly.
8) I found the system very cumbersome to use.
9) I felt very confident using the system.
10) I needed to learn a lot of things before I could get going with this system.

The following are the steps for calculating SUS:

1) The respondent's answer scale is reduced by 1 for each odd numbered statement (1, 3, 5, 7, 9).
2) Statements with even numbers (2, 4, 6, 8, 10) then 5 minus the scale of respondents' answers.
3) Doing the sum of the respondent's scale then multiplied by 2.5.
4) Summing up all respondents' scales and then averaged.

TABLE I. SUS CLASS CATEGORIES

SUS Score	*Grade*	*Rating*
>80.3	A	*Excellent*
68 - 80.3	B	*Good*
68	C	*Okay*
51 - 68	D	*Poor*
<51	E	*Awful*

C. Heuristic Evaluation

Heuristic Evaluation is a method for evaluating the usability of an application in order to find usability problems contained in the interface design. Heuristic evaluation involves several evaluators to assess whether a design violates the 10 principles of usability. This evaluation aims to find deficiencies and errors in the application so that recommendations for improving the interface design can be given to increase the level of usability [15].

D. Prinsip-Prinsip Heuristik

In 1994, Jakob Nielsen and Rolf Molich developed the heuristic principles used to analyze interface design. 10 Heuristic Principles are as follows [15]:

1) Visibility of System Status
2) Match Between System and The Real World

3) User Control and Freedom
4) Consistency and Standards
5) Error Prevention
6) Recognition Rather Than Recall
7) Flexibility and Efficiency of Use
8) Aesthetic and Minimalist Design
9) Help User Recognize, Diagnose, and Recovers from Errors
10) Help and Documentation

E. Severity Ratings

Severity Ratings are numbers that indicate the severity of a usability error. The scale that Severity Ratings has is 0 - 4.

1) 0: *I Don't Agree* - Bukan suatu permasalahan usabilitas
2) 1: *Cosmetic Problem Only* - No need to repair unless there is additional time to work on the project
3) 2: *Minor Usability Problem* - Low priority repair
4) 3: *Major Usability Problem* - High priority repair
5) 4: *Usability Catastrophe* - Repair with the highest priority. Very important to fix.

F. Prototype

A prototype is an initial model or example that is created to test a pre-existing concept. Prototypes are usually made to conduct trials, such as to find out whether the concepts built are feasible to be implemented and used by the wider community [17].

II. ANALYSIS AND DESIGN

This research consists of 8 stages. The first and second stages are Literature Study and data collection where data from user research, journals and previous research are collected to support this research. In the third stage, the stages are divided into 3 small stages, namely evaluation stage 1, making prototypes using figma and evaluation 2. Evaluation 1 is the stage for evaluating the original Cinépolis website together with 2 evaluators. Then from the results of the recommendations for improvement in stage 1, a prototype for figma improvement is made, and at evaluation stage 2 will be re-evaluated the figma prototype that has been made by 1 evaluator. In the next stage, the figma prototype that has been made will be built in the form of a website and a second stage of user research is carried out using the SUS method. After all stages are completed, the entire research process will be documented to facilitate further research with similar topics.

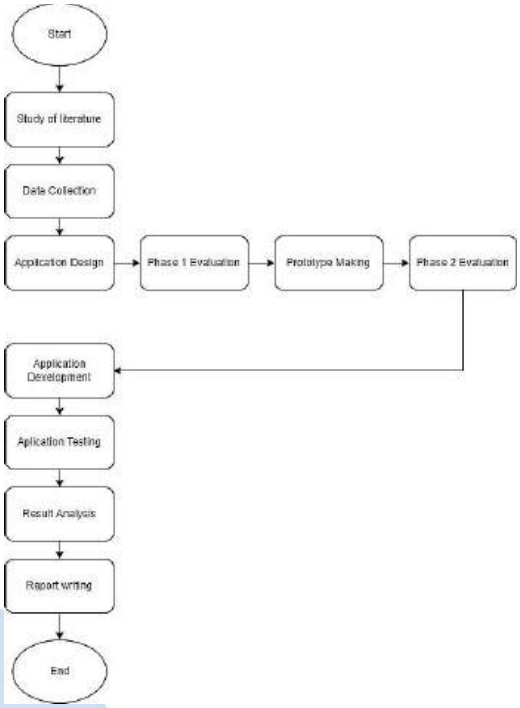


Fig 1. Research methodology

1) Evaluation Object

The main object that will be evaluated from this research is the original Cinépolis website. Here are the pages that are the main focus for analysis:

1. Home
2. Ticket
3. Schedule
4. Schedule with show times of each film
5. Movies
6. Detail Movie
7. Cinemas
8. Cinemas after search filter
9. Cinemas if no filter results were found
10. Detail Cinema
11. F&B
12. News & Promotions
13. Detail News
14. Detail Promotion

2) Evaluator

This research was assisted by 3 evaluators who specialize in the field of UI / UX [8]. The evaluation will be divided into 2 stages, where the first stage consists of Evaluator 1 and Evaluator 2, while the

second stage only consists of Evaluator 3. Each evaluator is given approximately 1 hour to explore and find problems on the Cinépolis website.

TABLE II EVALUATOR TERMS AND CRITERIA

Character	Condition
Usability Knowledge	Have extensive knowledge in terms of UI / UX and understand the 10 principles of Heuristics
Working Experience	Have work experience in the field of UI / UX

3) Evaluation result

TABLE III WEBSITE EVALUATION STAGE

Evaluation Stage	Name	Usability Understanding	Work experience
1	Evaluator 1	Understand	UI/UX Designer at Antikode
1	Evaluator 1	Understand	Senior UX Designer at Tokopedia
1	Evaluator 2	Understand	UI/UX Designer at Antikode

4) Phase1 Evaluation

Stage 1 evaluation is an evaluation carried out on the original Cinépolis website with the assistance of 2 evaluators.

a) Evaluator Results 1

TABEL IV RESULT FINDING PROBLEMS FROM EVALUATOR I

Problem Found	*Severity Rantings*	*Heuristic Number*
Very messy website with no clear CTA for users	4	8 & 6
Main Banner size is not ideal	3	8
The website hierarchy is not clear. Almost the entire website section is larger than the Now Showing section (News, Event, Ad Banner), so users are confused which section they should see first	4	6 & 7
Too much white space on the left and right of the website	3	8
All colors seem to dominate	4	8
The mobile app download banner ad is at the top of the website	2	8
Progress Bar and tab design look the same, can make users very confused	4	8 & 3
The distance between the movie posters is too close, so users are confused about the title of the movie that belongs to the poster above or below	4	6
The function of the "Find a Cinema" bar is not clear	3	4
The purpose of the page is not clear	4	6

b) Evaluator Results 2

TABEL V RESULT FINDING PROBLEMS FROM EVALUATOR II

Problem Found	*Severity Rantings*	*Heuristic Number*
There is no identifier in the header to indicate the active page	3	1
Newsletter is displayed before Now Showing	2	8
The "We Are Hiring" section doesn't fit the context of the web page	3	8
The overall appearance of the website is very stiff	1	8
"Change Showtime" and "Next" positions are reversed	4	4
Step by step ticket booking is displayed from beginning to end, making users not focus on filling out the page that is currently open	3	2
Prices shown are not marked with commas	1	4
The features on the Ticket and Schedule pages are almost exactly the same, so the web becomes redundant	4	4
No message appears when the searched cinema and cinema class has no results (only displays Blank Space)	3	1
Breadcrumbs have no important role on the page	1	8

c) Accumulation Phase 1

$$Accumulated\ Value = \frac{Total\ Severity\ Rating\ Value}{Frequency\ of\ Problems\ Appearing\ in\ Stage\ 1\ Evaluation}$$

The following are the results of the accumulation of the problem findings in the evaluation stage 1:

i) 3 heuristic principles with Usability Catastrophe values (Highest level of repair) on heuristic numbers 3, 6, and 7.

ii) 3 heuristic principles with Major Usability Problem scores on heuristic numbers 1, 2, and 4.

iii) 1 heuristic principle with a value of Minor Usability Problem (Low level improvement) on the heuristic number 8.

iv) 3 heuristic principles with a value of I don't agree (Not a problem) on heuristic numbers 5, 9 and 10.

d) Recommendations for Improvement of Phase 1 Evaluation

TABEL VI RECOMMENDATIONS FOR IMPROVEMENT OF PHASE 1 EVALUATION

Problem Found	*Repair Recommendations*
Very messy website with no clear CTA for users	Revamp entire website pages
Main Banner size is not ideal	Adjust the size of the Main Banner

The website hierarchy is not clear. Almost the entire website section is larger than the Now Showing section (News, Event, Ad Banner), so users are confused which section they should see first	Improve the website hierarchy by determining the order of priority in terms of features, and adjusting the size of features based on their priority
Too much white space on the left and right of the website	The main page of the feature is expanded so that there is less white space on the left and right of the website
All colors seem to dominate	Determine the primary color, second color, and color palette
The mobile app download banner ad is at the top of the website	Create your own special section "Download Mobile App"
Progress Bar and tab design look the same, can make users very confused	The design is made different or simplified
The distance between the movie posters is too close, so users are confused about the title of the movie that belongs to the poster above or below	Give enough space between the posters, the title of the film is placed under the poster
The function of the "Find a Cinema" bar is not clear	Removed "Find a Cinema" bar
The purpose of the page is not clear	F&B menu removed
There is no identifier in the header to indicate the active page	Make an identifier by distinguishing the color of the text on the currently active header menu
Newsletter is displayed before Now Showing	Newsletter feature moved to Footer
The "We Are Hiring" section doesn't fit the context of the web page	Removed from Home
The overall appearance of the website is very stiff	Some elements on the website can be made more modern, for example, a movie poster is given a slight curve at the corners (Rounded edge).
"Change Showtime" and "Next" positions are reversed	Change the position of the "Next" button with "Change Showtime" so that the position of the "Next" button is on the far right (End User Decision). The "Changes Showtime" button can be omitted and use the breadcrumb at the top of the page only if the user wants to return to the previous page
Step by step ticket booking is displayed from beginning to end, making users not focus on filling out the page that is currently open	The tab "Select Seats" et al does not need to be displayed before the step the user is running has not been completed
Prices shown are not marked with commas	Prices are marked with a comma as a separator (For example Rp. 40,000)
The features on the Ticket and Schedule pages are almost exactly the same, so the web becomes redundant	Combined into one tab menu named Schedule
No message appears when the searched cinema and cinema class has no results (only displays Blank Space)	Show a confirmation message to the user, for example: "The cinema you are looking for is not found"
Breadcrumbs have no important role on the page	Breadcrumb removed

5) Phase 2 Evaluation

Stage 2 evaluation is an evaluation carried out on the Cinépolis repair prototype using figma, assisted by 1 evaluator.

a) Evaluator Results 3

TABEL VII RESULT FINDING PROBLEMS FROM EVALUATOR 3

Problem Found	***Severity Rantings***	***Number Heuristic***
Footer integrates with web page	3	8
The font size of the seat number is too small, and each row of seats is too close together	2	8
The font size for ordering details and filling out forms is too small	2	8
The question form is next to the field	2	8
Using previous and next text in pagination	1	2

b) Accumulation Phase 2

$$Accumulated\ Value = \frac{Total\ Severity\ Rating\ Value}{Frequency\ of\ Problems\ Appearing\ in\ Stage\ 2\ Evaluation}$$

The following are the results of the accumulation of the problem findings in the 2nd stage evaluation:

i) 1 heuristic principle with a value of Minor Usability Problem (Low level improvement) on the heuristic number 8.

ii) 1 heuristic principle with a value of Cosmetic Problem Only (lowest level of improvement) on heuristic number 2.

iii) 8 heuristic principles with a value of I don't agree (not a problem) on heuristic numbers 1, 3, 4, 5, 6, 7, 9 and 10.

c) Recommendations for Improvement of Phase 2 Evaluation

TABEL VIII RECOMMENDATIONS FOR IMPROVEMENT OF PHASE 2 EVALUATION

Problem Found	***Repair Recommendations***
Footer integrates with web page	Can be given a different color or given a border between the web page and the footer
The font size of the seat number is too small, and each row of seats is too close together	The font size of the seats is made bigger and gives space between rows of seats
The font size for ordering details and filling out forms is too small	Enlarged font size

The question form is next to the field	Questions can be placed above the field to make it easier for users to read
Using previous and next text in pagination	Can be simplified by using the < icon for previous and > for next

III. IMPLEMENTATION AND TESTING

Based on the questionnaires that have been distributed, from a total of 37 respondents, the results obtained with the SUS calculation method are the Cinépolis repair website which has been built into a website based on recommendations for improvement using the Heuristic Evaluation method, getting a score of 88.10. This means that the Cinépolis repair website is above the SUS standard score of 68 and is in grade A, namely the Excellent category[6].

IV. CONCLUSION

Recommendations for improvements to the Cinépolis repair website using the Heuristic Evaluation method twice, namely the original Cinépolis website in the first stage and the wireframe in the second stage have been successfully implemented. The comparison of the value of the original Cinépolis website with the revamping website of Cinépolis is 54.03 and 88.10. The revamping website Cinépolis got a value of 34.07 greater than the value of the original website. So it can be concluded that the revamping of the Cinépolis website has been successful in improving the appearance of the interface and has succeeded in improving the user experience when using the website.

REFERENCES

[1] P. C. Kartika, "Rasionalisasi perspektif film layar lebar beradaptasi karya sastra," Jurnal Pena Indonesia, vol. 2, 2017.

[2] J. Llamas-Rodriguez, "A global cinematic experience: Cinépolis, film exhibition, and luxury branding," JCMS: Journal of Cinema and Media Studies, vol. 58, 2019.

[3] S. Gupta, "An analysis of ui/ux designing with software prototyping tools," 2019.

[4] "An indonesian adaptation of the system usability scale (sus)," 2017.

[5] R. Stockton and F. N. Kerlinger, "Foundations of behavioral research," American Educational Research Journal, vol. 11, 1974.

[6] "System usability scale vs heuristic evaluation: A review," Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, vol. 10, 2019.

[7] "Uji usability dengan metode cognitive walkthrough pada situs web perpustakaan universitas mercu buana jakarta," Jurnal Pustakawan Indonesia, vol. 15, 2016.

[8] "Uji usability dengan metode cognitive walkthrough pada situs web perpustakaan universitas mercu buana jakarta," Jurnal Pustakawan Indonesia, vol. 15, 2016.

[9] N. Brugger, "Website history and the website as an object of study," New Media and Society, vol. 11, 2009.

[10] H. B. Bekti, "Mahir membuat website dengan adobe dreamweaver cs6, css, dan jquery," AndiPublisher, vol. 1, 2015.

[11] H. Joo, "A study on understanding of ui and ux, and understanding of design according to user interface change," International Journal of Applied Engineering Research, vol. 12, 2017.

[12] J. Nielsen, "Usability engineering," 2004.

[13] S. Reeves, "How ux practitioners produce findings in usability testing," ACM Transactions on Computer-Human Interaction, vol. 26, 2019.

[14] "Sus: a retrospective," Journal of usability studies, vol. 8, 2013.

[15] J. Nielson, "Jakob nielsen ' s ten usability heuristics," Nielsen Norman Group, 1995.

[16] J. Nielsen, "Severity ratings for usability problems: Article by jakob nielsen," 1995.

[17] "The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications," Choice Reviews Online, vol. 50, 2013.

ISSN 2355-0082

Data Quality Issues : Case Study of Claim and Insured in Indonesia Insurance Company

Chris Solontio[1], Achmad Nizar Hidayanto[2]
[1,2] Department of Information Technology, Universitas Indonesia, Jakarta Pusat, Indonesia
[1] chris.solontio21@ui.ac.id, [2] nizar@sci.ui.ac.id

Accepted 24 August 2024
Approved 7 November 2024

***Abstract*— Data has become an asset for insurance companies that have many benefits and management needs to realize the importance of data quality to avoid the impact of poor data quality. In this study, data quality measurement will be carried out by observation to see the total amount of invalid data from data dimensions, namely, accuracy, completeness and consistency of the relationship between claim data and insured, and findings from each data fields in this case study. In addition, researchers conducted interviews to find out the obstacles faced by IT, Customer Retention, Operational, and Actuary teams where they are directly related to data flow and data processing. From the results of the analysis, there is invalid data that will affect the analysis and cause obstacles faced by users according to the interview results. In the conclusion, management needs to form a data governance team to avoid poor data quality that has responsibility for data flow and maintains data quality in order to provide a positive impact such as providing the right data accuracy in data analysis and user time to be more effective in data processing, assisting in making data warehouses, applying AI and digital transformation as a form of improvement in the services provided.**

***Index Terms*— Data Management; Data Quality; Insurance; Data Quality Dimension**

I. INTRODUCTION

According to the results of the populix survey, the majority of Indonesia's population already has BPJS Kesehatan insurance of 83% and private insurance of 38% [1]. One of the most owned private insurances is health insurance [2]. Insurance companies are now carrying out many development initiatives in the IT world such as creating a data warehouse, carrying out digital transformation such as using chatbots or processing claims quickly with the use of artificial intelligence (artificial intelligence), or the use of data analytics and data science to be able to compete between insurance companies and provide the best service for customers. This initiative can be carried out well, inseparable from the importance of a good company data condition where it must start paying attention to data management so that there is no more unreliable data.

The life insurance company that is the place of this research case study is one of the largest life insurance companies in Indonesia by providing health protection for employees of other companies and has a wide network of providers. In addition, this insurance company has two main health insurance products, namely indemnity and managed care. Of the two products, this managed care product will be the source of data used in assessing the quality of the company's data. Managed care is health insurance products that have a different process from indemnity, namely the customer must start the first treatment at a first-level health facility and then get a referral to the hospital if needed for outpatient or inpatient treatment in accordance with the plan and provisions in the policy which is an agreement between the insurance company and other companies.

In facing today's business challenges, the company has the initiative to improve its services to be the best and help management quickly in making decisions through digital transformation. Some of these initiatives have been carried out such as accelerating the claim process, creating dashboards for the company's internal and assisting management in making decisions from cash flow results, claim behavior and others. Meanwhile, initiatives that have not been carried out are the creation of a data warehouse, and the use of AI in detecting fraud. From the initiative that has been running, there are obstacles where the process takes approximately 3 hours because it needs to validate data because there is still incomplete and inconsistent data. Meanwhile, initiatives that have not been implemented cannot be carried out due to the condition of the data. From this, there are obstacles in data management that need to be considered, especially in the quality of the data itself.

From the obstacles faced, the purpose of this study is to evaluate the impact of poor data quality on the company's operations, and customer satisfaction. Besides that, the research question that arises for this research, that is:

1. how to deal with the issue of existing data quality?
2. What quality dimensions are suitable for this insurance industry?.

One of the components of the data management framework that will be discussed is data quality [3]. Data quality can result in efficient operational processes, decision-making, data warehouse creation, and have a positive influence on customer satisfaction [4]. In maintaining data quality, every company needs to define data dimensions with the aim of knowing the impact of poor data quality on costs, reputation, compliance regulations and so on [5]. In the insurance business line, poor data quality in general can lead to losses in operational and strategic costs for hidden costs and direct costs [6]. Hidden fees (hidden cost) and direct costs (direct cost) of the effects of poor quality data on insurers in figures 1.

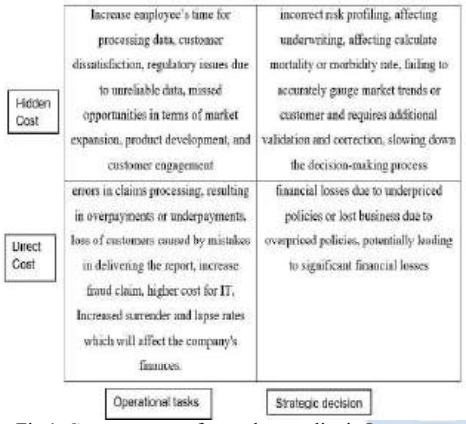


Fig 1. Consequences of poor data quality in Insurance

In Data Quality for the Information Age, Thomas Rednan formulates a set of data quality dimensions that are rooted in data structures. In addition, the dimensions of data cannot be determined the same in every business area but can vary in each company depending on the characteristics of the data or the use of data by the company. The importance of data quality and dimensions as a tool to measure data quality in an organization, many researchers conduct research on this research by conducting literature reviews or with case studies in several business lines such as Table I.

TABLE I. DATA QUALITY RESEARCH AREA

Research Areas of Data Quality	**References**
Government	Government **O**rganization **[7]**, BPS-Statistics Indonesia **[8]**, Malaysian Public **[9]**, State Electricity Company **[10]**.
Financial Industry	PT BPI **[11]**.
Airport Services	PT JAS **[12]**.
Education	Institute of Statistic **[13]**.
Factory	Paper Factory **[14]**.
Medical or Health	Electronic Medical Record **[15]**.

From table I, there is still limited research on quality data in insurance companies. Based on the table of impacts caused by poor quality, poor data quality will affect insurance companies, namely increased employee time to process data, customer dissatisfaction, regulatory problems due to unreliable data, slowing down the decision-making process, resulting in overpayment or underpayment, customer loss caused by errors in submitting reports, and financial losses due to policies that are not too cheap or lose business due to overly expensive policies. Based on research conducted by Chen, assessing the quality of data in the consistency dimension between MIS social insurance data and MIS workforce has the goal of improvising in terms of integrating the two databases [16]. The paper presents an evaluation using conflict classification methods, especially intra-concept conflicts and inter-concept conflicts. Intra-concept conflicts are related to data quality based on variations in data values, such as the appearance of the name "Liu Ming" in social insurance data and MIS Employment data. On the other hand, inter-concept conflict refers to the assessment of data quality exemplified by the inconsistency of the employment status of individuals between the insurance database of non-workers in the company and the database of MIS Workers who are currently employed in the company.

The next research was carried out by Haryadi, an assessment of the data quality of top banks and insurance in Europe with the aim of creating a big data process [17]. In the study, each assessment of the data quality data dimension differed between banks and insurance, where for insurance companies, Allianz data quality was carried out to see the accuracy and timeliness. The next research conducted by Mary discussed the quality of claims database data against Electronic Health Record (EHR) data which was carried out by dividing 9 roles for the relationship between data attributes to see the number of invalid data based on the compatibility dimension, completeness, and the value of the date range for which the results are still found invalid [18]. Thus, the answer to the second research question is that the most suitable data quality dimensions for the insurance industry are accuracy, completeness, and consistency.

The data quality dimension to be used based on the number of quality dimensions that will be used [15] and the results of discussions with the head of the IT division, namely completeness, consistency, and accuracy. In assessing the quality of this data, the study will apply the method used in Mary's paper by forming several roles to assess the quality of the data by seeing if any data is found invalid with several roles that have been determined. Insurance is one of the things that can be categorized in the financial services, it is the same as a bank where data quality is very important so that this research can be an aid to determine the dimensions and assess the quality of data from the characteristics of the data in insurance.

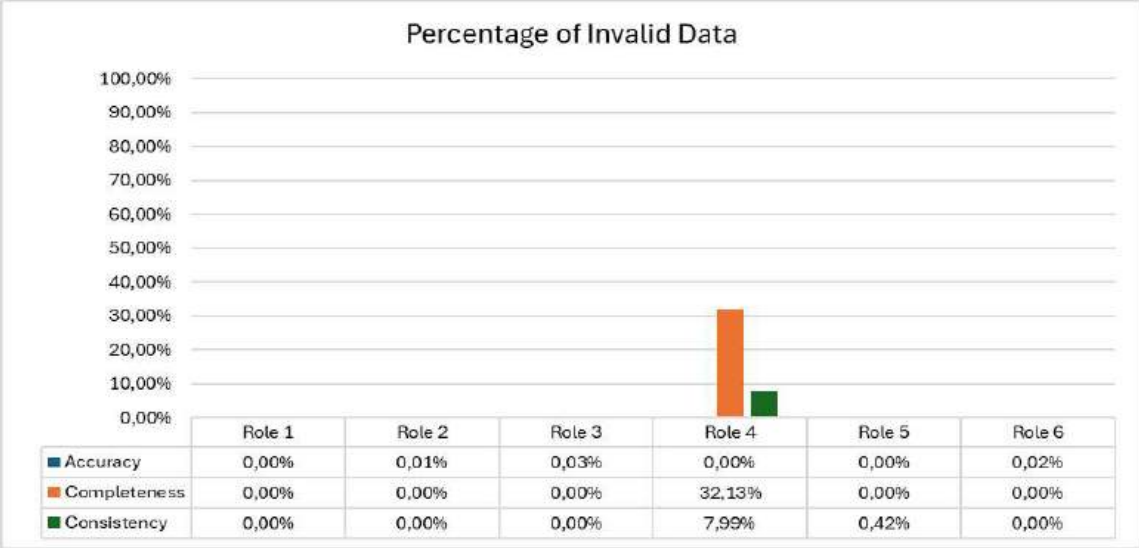


Fig 1. Consequences of poor data quality in Insurance

II. Research Methodology

In this study, qualitative and quantitative research design will be carried out. The quantitative method will be carried out by observation with profiling data in order to find out the results of data assessments that are not valid from the data quality dimension that has been determined by forming 6 roles between two

claim data and participants [18]. Meanwhile, qualitative interviews will be conducted to units that are specifically adjacent to the use of data to add supporting data related to the problems found from the results of the assessment and the consequences generated in the company's activities.

In data processing, it will use excel and python to view invalid data from participant details and claim data. Before conducting a data assessment of the quality data dimension, 6 roles must be formed that are relevant to the interconnectedness of the two data, namely:

TABLE II. Six Roles to Asses Data Quality

Roles	Descriptions
Roles 1	Service Date and Discharge Date for Inpatient in Claim Data.
Roles 2	Participants status with ICD X in claim data.
Roles 3	Gender and ICD X.
Roles 4	Types of Services with ICD X.
Roles 5	Participant Plan and Participant Number.
Roles 6	Participant Status and Age Range.

III. Result and Discussion

A. Result

The limitation of this study is that the claim data analyzed is claim data provided from the case study location, namely the provider history (hospital/clinic) in 2022 and the amount of data provided is 125 thousand data consisting of 38 thousand SJP (letter of guarantee for patients after health services are carried out at a hospital or clinic) data, 9 thousand participant claims and three providers.

Based on the six roles that have been formed from claim and participant data to assess the quality of the data, invalid data is found from each data dimension from roles 2 to 6 that have been formed. From figure 2, it can be concluded in detail as follows:

- Roles 1 does not find invalid data on the receipt date and delivery date data by looking at the length of stay.
- Roles 2 were found to be 2 SJPs invalid with information on the status of child participants but the diagnosis was Z34 which is a normal pregnancy surveillance diagnosis.
- Roles 3 based on a list of diagnoses per gender based on ICD X on ICD10CM website, found that 13 SJPs were invalid in data claiming a link between gender and diagnosis (Table IV).

TABLE III. Invalid Data on ICD X By Gender

Gender	Invalid Diagnosis
Male	O74, N91, N89, N80, dan N93
Female	C61, D17.6, dan N47

- Roles 4 types of services with 32% (12,307 SJP) incompleteness diagnoses where the

service does not have details of the type of medication administered. In addition, there are 8% inconsistent data due to the writing of the type of service that does not match and there are differences in spaces, for example, anasthesi periodal operation actions (caesar / sectio) and anasthesia periodal operation actions (caesar / sectio), dengue ns 1 ag and dengue ns1 antigen.

- Roles 5 there is an inconsistency between the participant plan and the participant data where there are 36 participants who have 2 different plans for example participant number 00001 but have silver and blue insurance plans.
- Roles 6 were found to be 14 participants whose age ranges did not match their membership status (Table V).

TABLE IV. INVALID DATA IN AGE RANGE BASED ON PARTICIPANT STATUS

Status	Age Range	Total
Spouse	0 - 5	1
	6 - 12	2
	13 - 16	6
Child	36 - 45	4
	46 - 55	1

In addition to the results of data analysis, interviews were conducted with departments that work directly or manage data, namely information technology, customer retention, operations and actuary with the aim of finding out the effects and consequences of poor data quality. From the results of the interview, several points were found as follows:

- Department of Information Technology

There is no single department in the IT field that is specifically responsible for maintaining the quality of the data in the system. Although for now the IT service team and business analyst work together in carrying out their duties as MIS and data engineers, but this makes the team overwhelmed because there are dual tasks from the main tasks of the two teams so that in maintaining the existing data, sometimes only relying on findings from users if anomalous data is found when processing data and in terms of validation it also takes a long time. In addition, another problem is the absence of regulations regarding data management and not knowing how to measure data quality.

- Customer Retention Department

Anomalies in the date of birth data of participants and their family members caused by the absence of format locking on the date of birth from a file manually uploaded to the system caused an error in the participant's date of birth.

- Operations Department

The data processing for the analysis material takes a long time because inconsistent and non-uniform data are found, especially in the outpatient or inpatient service detail codes caused by each hospital with different inputs and do not have a uniform code for the service detail code.

- Actuary Department

It takes a long time in the data validation section because there is an inconsistency between participant data and claim data, so it must be validated by the data owner or IT who provides the data. This has an impact in determining premium extensions for customers who are not fast and analyzing Claims for additional audit reports, as well as providing recommendations and information on inflation that occurs to the pricing team in designing a baseline percentage of one of the components of premium calculation and other reports needed by the actuarial team.

From the results of observations and interviews that have been conducted, it is found that there are data quality problems that must be faced due to two things, namely first, the absence of a team responsible for the data entering the database and the absence of regulations and an overview of the upstream to downstream process in the flow of data into the database.

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

B. Discussion

This paper will focus on data quality where companies must start to raise awareness of the importance of avoiding poor data quality where the consequences are very significant impact on both internal and external as explained in figure 1 in the introduction. Despite limitations in the data provided by the case study site (2022 data), connecting data analysis with interview findings reveals that the lack of a dedicated team for data management can result in declining data quality. Furthermore, a data warehouse is essential to resolve inconsistencies in health service classifications, which is a primary responsibility of a specialized data team.

Therefore, this answers the first research question that management should establish a dedicated data team to address current and future data quality issues. To effectively define this team's duties and align them with industry standards, insurance companies can refer to the DAMA-DMBOK (Data Management Body of Knowledge) framework. This framework provides a structured approach for implementing key data governance tasks, ultimately ensuring data quality while leaving data security responsibilities with the IT security team.

Based on this analysis, management can establish a data governance team composed of data governance leads, data stewards, and data engineers. This team will have four primary objectives: developing a data quality framework focused on the dimensions of accuracy, completeness, and consistency; implementing data lifecycle management to oversee data collection, processing, storage, usage, and disposal; defining data ownership by assigning clear accountability for data quality within specific departments and Work with the compliance team to create internal and external data use policies that must comply with personal data protection regulations (PDPs) Personally Identifiable Information) which has been regulated in Law No. 27 of 2022 [19].

IV. CONCLUSION

In conclusion, this study identified significant data quality challenges within the insurance company, primarily due to a lack of management awareness and the absence of a dedicated data governance team. These issues impact both operational efficiency and customer satisfaction, which are critical to the company's competitive advantage in the insurance industry. Given the critical role data quality plays in supporting accurate analysis, timely decision-making, and regulatory compliance, it is essential for the company to adopt structured data governance practices.

To address these issues, the following recommendations are proposed:

- Establishing a Dedicated Data Governance Team:

This team should consist of data engineers, data stewards, and data governance leaders, with each role having clear responsibilities as in the discussion section or can be based on a data governance framework such as DAMA-DMBOK. This structure will ensure that all data management functions, from quality monitoring to data ownership, are systematically overseen.

- Leveraging Data Governance Tools:

Once the governance team is in place, the company can implement tools (e.g., Collibra) to streamline data quality checks, automate data profiling, and ensure consistency across datasets. Such tools will aid in early detection of quality issues, supporting proactive improvements in data accuracy and completeness.

- Building a Data Warehouse for Enhanced Accessibility:

Creating a centralized data warehouse will simplify data access and reporting for all users, reducing time spent on validation and enabling faster, data-driven insights. This infrastructure will also support future advancements, such as dashboards and AI-driven analytics, by providing a reliable data foundation.

By implementing these recommendations, the insurance company can significantly enhance data quality, improve customer satisfaction, and foster more reliable decision-making processes. This approach will position the company for sustainable growth and ensure readiness for future digital transformation initiatives.

REFERENCES

[1] Data Indonesia, "Survei Mayoritas Warga Indonesia Miliki Asuransi Kesehatan.", [Online]. Available: https://dataindonesia.id/varia/detail/survei-mayoritas-warga-indonesia-miliki-asuransi-kesehatan

[2] katadata, "Ini Jenis Asuransi Swasta yang Paling Banyak Dimiliki Penduduk Indonesia," 2023. [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/02/01/ini-jenis-asuransi-swasta-yang-paling-banyak-dimiliki-penduduk-indonesia

[3] I. DAMA, *2017 DAMA-DMBOK Data Management Body of Knowledge, 2nd Edition-1*. Technics Publications, 2017.

[4] L. David, *The Practitioner's Guide to Data Quality Improvement*. Knowledge Integrity, Inc., 2011.

[5] N. Askham, D. Cook, M. Doyle, H. Fereday, and M. Gibson, *The Six Primary Dimensions For Data Quality Assessment*. DAMA International, 2013.

[6] A. Haug, F. Zachariassen, and D. van Liempd, "The costs of poor data quality," *J. Ind. Eng. Manag.*, vol. 4, no. 2, pp. 168–193, 2011, doi: 10.3926/jiem.2011.v4n2.p168-193.

[7] R. Nulhusna, N. F. Taufiq, and Y. Ruldeviyani, "Strategy to Improve Data Quality Management: A Case Study of Master Data at Government Organization in Indonesia," *Proceeding - 2022 Int. Symp. Inf. Technol. Digit. Innov. Technol. Innov. Dur. Pandemic, ISITDI 2022*, pp. 150–155, 2022, doi: 10.1109/ISITDI55734.2022.9944466.

[8] S. D. Rahmawati and Y. Ruldeviyani, "Data Quality Management Strategy to Improve the Quality of Worker's Wage and Income Data: A Case Study in BPS-Statistics Indonesia, 2018," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, 2019, doi: 10.1109/ICIC47613.2019.8985803.

[9] Z. Zarina Abdul Jabar, M. Wook, O. Zakaria, S. Ramli, and N. Afiza Mat Razali, "Establishment of Big Data Analytics Application Model for Malaysian Public Sector: an Expert Validation," *Edpacs*, vol. 65, no. 4, pp. 1–20, 2022, doi: 10.1080/07366981.2021.1958736.

[10] A. Purnomoadi, I. M. Sari, J. S. Anna Maria, D. B. Fiddiansyah, N. E. Saputro, and M. S. Sofan Hadi, "A Method to Quantify Data Quality in Asset Health Indices Model," *Proc. 2023 4th Int. Conf. High Volt. Eng. Power Syst. ICHVEPS 2023*, pp. 16–20, 2023, doi: 10.1109/ICHVEPS58902.2023.10257386.

[11] N. Sunandar and A. Nizar Hidayanto, "Data Quality Management Improvement: Case Studi PT BPI," *2022 IEEE World AI IoT Congr. AIIoT 2022*, pp. 53–58, 2022, doi: 10.1109/AIIoT54504.2022.9817195.

[12] T. Wahyudi and S. M. Isa, "Data Quality Assessment Using Tdqm Framework: a Case Study of Pt Aid," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 9, pp. 3576–3589, 2023.

[13] W. Wijayanti, A. N. Hidayanto, N. Wilantika, I. R. Adawati, and S. B. Yudhoatmojo, "Data quality assessment on higher education: A case study of institute of statistics," *2018 Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2018*, pp. 231–236, 2018, doi: 10.1109/ISRITI.2018.8864476.

[14] D. Xu, Z. Zhang, and J. Shi, "A Data Quality Assessment and Control Method in Multiple Products Manufacturing Process," *2022 5th Int. Conf. Data Sci. Inf. Technol. DSIT 2022 - Proc.*, pp. 1–5, 2022, doi: 10.1109/DSIT55514.2022.9943883.

[15] D. I. Miguel, A. S. Solange N., I. Marcia, and Da Silva Suzana Alves, "Data Quality in health records: A literature review," *Iber. Conf. Inf. Syst. Technol. Cist.*, no. June, pp. 1–6, 2021, doi: 10.23919/CISTI52073.2021.9476536.

[16] W. L. Chen, S. D. Zhang, and X. Gao, "Anchoring the consistency dimension of data quality using ontology in data integration," *2009 6th Web Inf. Syst. Appl. Conf. WISA 2009*, pp. 201–205, 2009, doi: 10.1109/WISA.2009.32.

[17] A. F. Haryadi, J. Hulstijn, A. Wahyudi, H. Van Der Voort, and M. Janssen, "Antecedents of big data quality: An empirical examination in financial service organizations," *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016*, pp. 116–121, 2016, doi: 10.1109/BigData.2016.7840595.

[18] M. A. Gadde, Z. Wang, M. Zozus, J. B. Talburt, and M. L. Greer, "Rules based data quality assessment on claims database," *Stud. Health Technol. Inform.*, vol. 272, pp. 350–353, 2020, doi: 10.3233/SHTI200567.

[19] R. Indonesia, "Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi," *Ditama Binbangkum - BPK RI*, no. 016999, pp. 1–50, 2022, [Online]. Available: https://peraturan.bpk.go.id/Home/Details/229798/uu-no-27-tahun-2022

ISSN 2355-0082

Evaluating the Impact of Particle Swarm Optimization Based Feature Selection on Support Vector Machine Performance in Coral Reef Health Classification

Jessica Carmelita Bastiaans[1], James Hartojo[2], Ricardus Anggi Pramunendar[3], Pulung Nurtantio Andono[4]

[1,2,3,4] Informatics Engineering, Dian Nuswantoro University, Semarang, Indonesia
[1]jessicacarmelita2004@gmail.com, [2]jameshartojo@gmail.com, [3]ricardus.anggi@dsn.dinus.ac.id, [4]pulung@dsn.dinus.ac.id

Accepted 1 September 2024
Approved 19 November 2024

***Abstract*— This research explores improving coral reef image classification accuracy by combining Histogram of Oriented Gradients (HOG) feature extraction, image classification with Support Vector Machine (SVM), and feature selection with Particle Swarm Optimization (PSO). Given the ecological importance of coral reefs and the threats they face, accurate classification of coral reef health is essential for conservation efforts. This study used healthy, whitish, and dead coral reef datasets divided into training, validation, and test data. The proposed approach successfully improved the classification accuracy significantly, reaching 85.44% with the SVM model optimized by PSO, compared to 79.11% in the original SVM model. PSO not only improves accuracy but also reduces running time, demonstrating its effectiveness and computational efficiency. The results of this study highlight the potential of PSO in optimizing machine learning models, especially in complex image classification tasks. While the results obtained are promising, the study acknowledges several limitations, including the need for further validation with larger and more diverse datasets to ensure model robustness and generalizability. This research contributes to the field of marine ecology by providing a more accurate and efficient coral reef classification method, which can be applied to other image classifications.**

***Index Terms*— Coral Reef Classification; Histogram of Oriented Gradients (HOG); Machine Learning; Particle Swarm Optimization (PSO); Support Vector Machine (SVM)**

I. INTRODUCTION

Indonesia as an archipelago that has the second longest coastline reaching more than 95,000 km2 where more than 60% of its territory is the ocean, and its geographical location between the Indian and Pacific Oceans produces a very rich and diverse marine biodiversity. More than 39,500 km2 or as much as 16% of the world's coral reefs are found in Indonesia [1]. Coral reefs are home to marine biodiversity up to more than 6000 species of fish that are very large and unique [2]. Coral reefs play an important role in maintaining biodiversity, preventing coastal erosion, and promoting business trade.

However, coral reefs are experiencing population decline due to overexploitation, ecosystem damage, and climate change causing abrasion [3], [4]. Nearly 25% of Indonesia's 270 million people live and do activities in coastal areas within 30 km of coral reefs, based on this percentage, it affects up to 95% of coral reefs in Indonesia are currently in threatened status where more than 35% are in high or very high threat levels [1]. Meanwhile, about 43% of observations of 324 coral reefs in Indonesia are damaged and even endangered while only 6.48% are still in very good condition and only 5.48% of coral reefs have high status from the results of a survey at 985 stations conducted by the Oceanographic Research Center of LIPI in 2008 [1].

Based on research conducted by several scientists, almost 50% of coral reefs will be destroyed by 2030 [5] Therefore, the destruction of coral reefs affects the health of marine life and also decreases the livelihood of people who depend on it. One of the diseases experienced by coral reefs is bleaching. Coral reefs bleach due to ocean acidification and global warming, which is a serious threat to the earth's ecosystem [6].

With the advancement of technology in image processing, researchers began to conduct research on coral reefs using a variety of machine learning and deep learning algorithms. Such as, the use of CNN (Convolutional Neural Network) Algorithm [7] in the classification of small datasets of coral texture images, data augmentation techniques and transfer learning approaches, using variations of ResNet and ImageNet to improve coral reef image processing results [8], [9]. YOLOv4 (You Only Look Once) was also used in determining coral reef disease for the computer vision algorithm training process. An incremental methodology was developed in three training stages to

evaluate accuracy, by modifying different parameters [2], [10], [11].

But on the other hand, by proposing learning models for species classification in underwater images, as well as many factors such as lighting, color, shape, structure, etc. it was found that the performance of CNN models in image classification decreased significantly [12], [13]. Image classification analysis comparison has also been conducted out with a small sample ImageNet Dataset, obtained SVM accuracy results of 0.85 and CNN 0.82 with runtime for SVM 1.05 minutes and CNN 2.05 minutes [14]. Deep Learning models often require more computational resources for training, and these models are less interpretable due to their black box nature [15], [16]. So, based on the research that has been done, SVM is superior in handling small datasets with the support of a fast runtime process compared to CNN [14], [15], [17], [18].

Previous researchers used three algorithms to model coral reef bleaching areas with 3 models, the results showed that SVM was the most effective classification model with 88.85% accuracy, followed by decision tree and Naïve Bayes with 80.25% and 71.34% accuracy [19]. Other researchers also compared SVM with other machine learning, namely KNN, where SVM is the right choice for the classification of larger datasets [20]. Coral reef researchers in Kapota Atoll (Wakatobi National Park, Indonesia), Harapan Island and Kelapa Island of Kepulauan Seribu Indonesia, Palmyra Atoll also used the SVM (Support Vector Machine) algorithm method to classify and recognize characteristic images of coral reefs [21], [22], [23], [24]. As for the development of other research that aims to classify bleached and healthy coral reefs, this research also uses the SVM (Support Vector Machine) Classifier method and receives input from grouping features based on the similarity of coral reef characteristics [6].

The author also concluded that the Histogram Of Oriented Gradients (HOG) feature provides better results than SIFT and SC in its use with the SVM algorithm model to perform coral reef classification [25]. The SVM algorithm has the advantage of using raw image data as feature vectors, especially for natural random textures where characterization is difficult to obtain. This is relevant to the classification of coral reef images that are rich in patterns, colors, shapes, and textures [26], [27]. Based on previous research, it is an appropriate target to use various HOG feature extraction methods and PSO optimization to determine the best features based on the particle set so as to achieve the best accuracy of the SVM model [28]. The application of this strategy is considered suitable as it aims to obtain a higher level of precision and accuracy. The contribution of this research is validated by comparing the classification results of healthy, whitish and dead coral reef images for: (1). test the use of PSO parameters on the effect of SVM classification performance; (2). investigate the relationship of HOG extraction with PSO optimization; (3). Identify solutions and feature combinations needed to achieve accurate performance of healthy, whitish and dead coral reef image classification.

This document is organized in the following order: Section 1, regarding the background of this proposed research. Section 2, relates the research method and discusses the experimental strategy used by the authors. Section 3 details the evaluation and experimental findings. Section 4 contains the conclusion of the research.

II. METHOD

Contents Based on previous research, it shows that the selection of machine learning algorithms and image preprocessing affects the accuracy and classification of coral reef images. The characteristics of the detected coral reefs will determine the label of each processing dataset collected. The dataset also has an influence for machine learning to learn the uniqueness of the coral reef itself. This research uses a unique approach to recognize the characteristics of healthy, whitish and dead coral reefs.

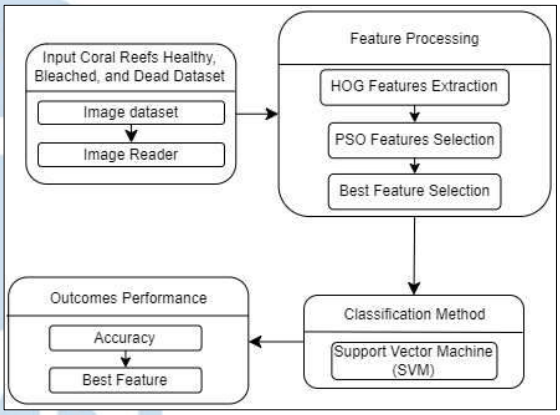


Fig. 1. Research proposed diagram

As shown in Fig. 1, the classification stage begins with the process of collecting coral reef datasets labeled healthy, bleach and dead which will be used as input data for the training process. This classification technique is based on several categories extracted through Histogram Oriented Gradients (HOG), including histograms of color, texture, shape, and standard deviation. HOG itself has superior characteristics in the form of edge structure representation, shape, and adjustable level of variation [29]. The extraction results performed by HOG will then be optimized through the Particle Swarm Optimization (PSO) algorithm feature selection. PSO will take the best value from the optimization results for the classification of coral reef images by PSO-based Support Vector Machine. PSO is a method of finding the best combination of features that will be classified by the SVM model. In the special case of coral reef texture classification, Support Vector Machine allows excellent class separation even when the feature vector size is large and the number of training samples is limited [26], [27].

A. Data Collection

In this study we conducted two initial stages, namely data collection and recognition of coral reef characteristics. Bleached Health Dead Corals Dataset used in this study in the form of images of healthy, whitish, and dead coral reefs (https://www.kaggle.com/datasets/sonainjamil/bhd-corals).

Fig. 2. Bleached coral reef images

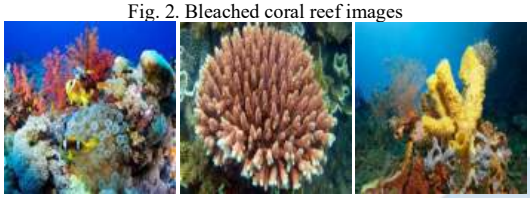


Fig. 3. Healthy coral reef images

Fig. 4. Dead coral reef images

The dataset is divided into three parts, namely training, testing, and validation. Training data includes 576 images of whitish coral reefs, 569 healthy coral reefs, and 120 dead coral reefs, while validation data includes 72 images of whitish coral reefs, 71 healthy coral reefs, and 15 dead coral reefs, then for testing data includes 72 images of whitish coral reefs, 72 healthy coral reefs, and 15 dead coral reefs. The set of image data contained in the training file will be processed so as to obtain information on the unique characteristics of healthy, whitish and dead coral reefs. HOG helps the process of extracting coral reef image recognition and removing noise around the detected image.

B. Histogram of Oriented Gradients (HOG) for Features Extraction

Histogram of Oriented Gradients (HOG) is an extraction feature that helps in coral reef image recognition [30]. In the image processing process, grouping pixel gradient values based on the directional orientation of each part of the local structure and shape characteristics of the image [31]. HOG will convert the input image into a feature vector representation that reflects the gradient orientation [31] in various parts of the coral reef image. To perform the image feature extraction stage, it is necessary to read the dataset from the folder and set parameters for dynamic input.

The feature extraction function in this project will receive an image parameter containing a coral reef image that has been converted to grayscale format using the OpenCV library. This image input is dynamic, where the image dimensions, namely height and width, are extracted automatically. Based on the varying image dimensions, the cell size is calculated by dividing the height and width of the respective image by 9. This parameter will determine how large the size of each cell is where the gradient histogram will be calculated. Then, a fixed value of (2,2) is determined, which means that each block consists of 2x2 cells. The blocks are used as histogram normalization in some cells to increase the robustness against lighting changes. In addition, the gradient orientation in each cell is divided into 9 bins, which is useful for determining the number of gradient orientation intervals calculated in each cell.

After setting the parameters for image processing with HOG, for each pixel in the coral reef image, the gradient is calculated in the x and y directions. This gradient reflects the change in pixel intensity, which can be interpreted as a shape feature.

$$G_x = \frac{\partial I}{\partial x}, \quad G_y = \frac{\partial I}{\partial y} \tag{1}$$

Where I is the intensity of the image, and Gx and Gy represent the change in intensity in the horizontal and vertical directions, respectively (1).

After calculating the image gradient with the formula above, the next step is to determine the magnitude and orientation of the gradient at each pixel.

$$G = \sqrt{G_x^2 + G_y^2} \tag{2}$$

$$\theta = atan2(G_{y,}\, G_x) \tag{3}$$

The magnitude G represents the strength of the intensity change, while the orientation θ determines the direction of the change (2), (3). With the pre-set parameters, HOG will calculate each cell (small region in the image) which will then represent the gradient orientation distribution within a cell.

HOG leverages the use of block normalization to make features more resilient to lighting changes with *L2-Hys* normalization. This normalization is applied to ensure that the magnitude of the gradient vector does not affect feature detection, and helps to improve the quality of the extracted features by reducing sensitivity to lighting and contrast differences, so that the features remain well distributed throughout the image.

$$v' = \frac{v}{\sqrt{||v||_2^2 + \epsilon^2}} \tag{4}$$

Where v is the unnormalized feature vector of the block, and ϵ is a small value to prevent division by zero. v' is the normalized feature vector (4). After all the blocks in the image are normalized, the feature vectors of all the blocks are combined into one large feature vector that represents the image as a whole.

C. Particle Swarm Optimization (PSO) for Features Selection

The feature selection method in the SVM classification model uses Particle Swarm Optimization (PSO) which is an optimization algorithm adopted from the social behavior of animal group movements such as birds flying in flocks and groups of coral reefs [28], [32]. Each particle in PSO will move in the solution space to determine the combination as a potential solution from many available features, so that PSO can choose the best features to be used in SVM model training in order to maximize model accuracy [32], [33].

The process begins with initializing the parameters of the number of epochs and population size. Where the particles will be generated randomly in the solution space, then each particle determines one solution in the form of a binary vector that will be carried out in the feature selection process. Based on the given solution, the selected feature columns are taken from the training data and validation data. PSO will generate a new solution randomly when no features are selected. The selected features are then standardized to ensure that each feature has the same scale [34].

Next, the SVM model is trained based on a subset of features from the retrieved binary vectors to measure the accuracy of the model on the test data. The validation accuracy value obtained is then returned as the "fitness" value of the solution, where the fitness function is a calculation that determines the suitability or objective value of a solution. This research utilizes the classification method to evaluate fitness performance. Then, for the best solution results that have been achieved will be stored by personal best (*pbest*) with the aim of updating the particle position. At each iteration, the particle will compare the fitness value of the current position with the personal best (*pbest*) value. Then an update is made to the global best (*gbest*) value which refers to the best solution among all personal bests (*pbest*) of each particle in a particular iteration population. In each iteration, the global best (*gbest*) is calculated based on the fitness comparison of each personal best (*pbest*). The best solution (a subset of features) found during the optimization process is used to train the final model on training data and tested on test data.

Based on the process previously described, the formula for calculating the position and velocity of particles is generated as below.

$$x_i(t+1) = x_i(t)v_i(t+1) \quad (5)$$

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot rand_1 \cdot (pbest_i - x_i(t)) + c_2 \cdot rand_2 \cdot (gbest - x_i(t)) \quad (6)$$

In PSO, the representation of the solution using a position vector $x_i(t)$ indicates that each particle can have a potential solution in the search space, while the velocity vector $v_i(t)$ determines how fast and in which direction the solution will move in the next iteration. The combination of the two creates a search mechanism that helps the algorithm reach the optimal solution. Then, the personal best (pbset) affects the movement of individual particles, while the global best (gbest) is used to affect the movement of the entire population (5). The following is a description of the use of formulas that affect particle position and velocity updates (5), (6).

- Particle velocity *i* at iteration *t+1* is initialized with $v_i(t)$
- The position of particle *i* at iteration *t* is initialized with $x_i(t)$
- *w* is the inertia factor
- $c_1$dan c_2 are acceleration factors
- Random values between *0* and *1* are assigned to $rand_1$ and $rand_2$

$$fitness = accuracy(SVM_{model}, X_{selected}, Y_{val} \quad (7)$$

It then evaluates potential solutions on each particle representation. Each particle selects which features to use based on its position. $X_{selected}$ is the feature selected based on the particle position, this solution is then tested on the SVM model and evaluated based on accuracy on the validation set (7). By utilizing the feature solution information obtained by personal best (*pbest*) and global best (*gbest*), it can adaptively adjust the movement of particles to approach the most optimal solution.

D. Support Vector Machine (SVM) for Classification Model

Support vector machines (SVM) are binary classifiers that estimate the optimal separating hyperplane that maximizes the margin between two classes [26], [27], [31]. This coral reef image research classification uses the Support Vector Machine method because of the advantages in object- and pixel-based classification methods that have high accuracy [20], [35]. The preprocessing step is followed by training the Support Vector Machine (SVM) classifier [23]. The data splitting used in this study involves 3 categories using the ratio (80% Training: 10% Testing: 10% Validation).

Support Vector Machine receives input from HOG extractions in the form of color, structure, shape, texture, size, and gradient features from coral reef images. The author implemented the SVM model using the *C* parameter to control large margins and misclassification. A larger value of *C* influences the model to classify all samples correctly.

Many mapping functions are available, including linear, polynomial, and radial basis (RBF) kernel functions. Polynomial and RBF kernel functions are commonly used depending on the training dataset [19], it should be noted that the RBF kernel can be considered as a restricted version of the generalized Gaussian version, where the Gaussian matrix is restricted to a unit matrix multiplied by a scale factor [36]. In this study, the RBF kernel is applied to the Support Vector

Machine (8) [35] to handle the non-linearity problem (not linearly separable) by mapping the data to a higher dimension [37]. With gamma set to value scale so that it can automatically calculate the gamma value based on the number of features and data variations (9). there is a formula applied to coral reef classification where :

$$K(X_i, X_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \quad (8)$$

$$\gamma = \frac{1}{2\sigma^2} \quad (9)$$

- x_i and x_j are two input data vectors
- $\left\| x_i - x_j \right\|^2$ is the squared Euclidean distance between the two input data vectors
- γ is a gamma parameter that determines how far the influence of a training sample goes

Then, the SVM model is trained using training data with appropriate labels, and tries to find the most optimal hyperplane that separates the two classes with the largest margin. By maximizing the margin, the SVM model tries to have good generalization ability on data that has never been seen before. Here is the SVM formula to minimize the loss function while still ensuring maximum margin (10),

$$\min_{w,b}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i\right) \quad (10)$$

- $\frac{1}{2}\|w\|^2$ is the regularization part that tries to minimize the norm of the weight vector w, thus ensuring the margin remains large.
- C is a regularization parameter that controls the change between large margins and misclassification.
- ξ_i i is a slack variable that allows the feature vector data to be inside the margin or on the wrong side of the hyperplane, usually occurring when the data cannot be perfectly separated.

After the SVM model is trained, an evaluation is conducted on the training data to see how well the model learns to predict labels and recognize coral reef image patterns. The SVM model will be integrated into the feature selection process with PSO to select the best combination of features that produce the highest accuracy. With the support of PSO will find the optimal feature subset, so that the performance of the SVM model is maximized. The following is the decision formula used as a label prediction from new data (11),

$$y(x) = sign(w \cdot x + b) \quad (11)$$

- *y(x)* is the class prediction for data *x*
- $sign\ (z)$ to return *+1* if $z > 0$ and *-1* if $z < 0$

E. Performance Evaluation

Based on the classification performed by the Support Vector Machine (SVM) algorithm, the accuracy results are measured with a division of 80% training data, 10% test data, and 10% validation data. The trained model is used to predict the test data. Then, comparing the adjustment between the prediction results with the actual labels, namely healthy, whitish and dead reefs (12).

$$Accuracy = \frac{ypred}{ytrue} \times 100 \quad (12)$$

The process of classifying healthy, whitish and dead coral reefs using the Support Vector Machine (SVM) model produces an accuracy value of 85.44%, PSO helps the model in processing and finding the best features from the many combinations of particles available. Where, there will be a combination of the best features selected from all columns. The author decides to set a *feature solution ≥ 0.5* which will then be selected randomly, because the solution has a big influence on the best feature combination to improve model accuracy and speed up the feature selection process and model training (13), (14).

$$Features = Number\ of\ Columns \quad (13)$$

$$solutions \geq 0.5 \quad (14)$$

III. Result and Discussion

A. Result of the Histogram Oriented of Gradients (HOG)

Histogram Oriented of Gradients (HOG) extraction helps the classification process of coral reef images from a total of 2304 extraction features performed as in table 1. There are 158 images processed in the use of the HOG feature extraction method.

TABLE 1 FEATURE EXTRACTION RESULTS WITH HOG

No	X0	X1	X2	X3	X4	X5	X6	...	X2303
1	0.137	0.004	0.014	0.008	0.029	0.014	0.059	...	0.265
2	0.078	0.082	0.053	0.147	0.322	0.132	0.077	...	0.079
3	0.089	0.080	0.051	0.182	0.263	0.085	0.052	...	0.000
4	0.222	0.098	0.206	0.269	0.275	0.275	0.231	...	0.122
5	0.230	0.027	0.015	0.007	0.011	0.004	0.019	...	0.254
6	0.175	0.060	0.083	0.121	0.247	0.162	0.110	...	0.243
7	0.167	0.062	0.099	0.165	0.247	0.247	0.247	...	0.186
...	...	...	...	...	...	...	...	...	...
158	0.263	0.309	0.309	0.309	0.277	0.088	0.083	...	0.182

Then these features will help the PSO method in finding and selecting the best combination of features to find the best fitness value, where each iteration will compare the fitness value of the current position with the personal best (pbest) value obtained.

B. *Feature Selection Results with Particle Swarm Optimization (PSO)*

Experimental iterations were conducted 5 times with a population size parameter range of 5 to 25 iterations. Based on the PSO process that has been carried out, the best combination of solutions is obtained by taking the solution value set by the author, namely the *feature solution* ≥ 0.5. The selected solution represents the best feature columns, which are then used to form the optimal feature combination that will be used in the image classification stage..

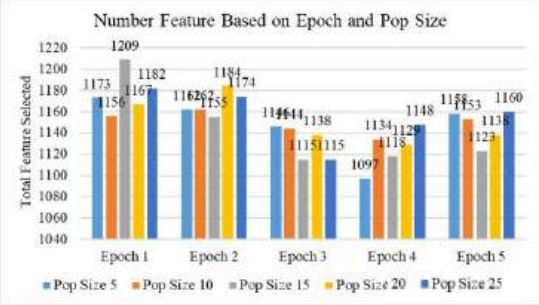


Fig. 5. Comparison total feature selected of each epoch and pop size

Fig. 5 visualizes the effect of epoch and population size on the number of features selected after selection. Looking at the graph above, larger population sizes, such as population size 25 often produce the largest number of features, while smaller populations, such as population sizes 10 and 15, tend to have fewer features. The number of selected features also tends to decrease as epochs increase, although there is movement between populations. Larger population sizes, especially in the final epochs, correlate with the best accuracy, indicating the important role of population size in the optimization process. The two-factor Analysis of Variance (ANOVA) results without replication show that the epoch factor has a significant effect on the number of features selected after selection, with an F-value of 7.8589554 and a P-value of 0.0010557, which is smaller than 0.05. This means that the number of selected features varies significantly between different epochs. In contrast, the population size factor did not show a significant effect on the number of selected features, with an F-value of 0.3041953 and a P-value of 0.8708985, which is much larger than 0.05. Thus, population size does not significantly affect the number of features selected in this selection process.

In Fig. 6, the scatter plot results show the relationship between the number of features selected and accuracy. The X-axis, the total number of features selected after the PSO selection process ranges from 1080 to 1220, while the Y-axis shows the accuracy of the model which reaches between 80% to 86%. It can be seen that the highest accuracy, about 85%, is achieved when the number of features is in the range of 1140 to 1160, although some other points also show high accuracy at different numbers of features. This shows that there is no linear correlation between the number of features and accuracy, so increasing or decreasing the number of features does not always have a consistent impact on accuracy. The use of PSO in feature selection proved to be effective in finding the optimal number of features that can achieve the best accuracy without having to use all the features, thus improving the efficiency of the model and reducing computational complexity.

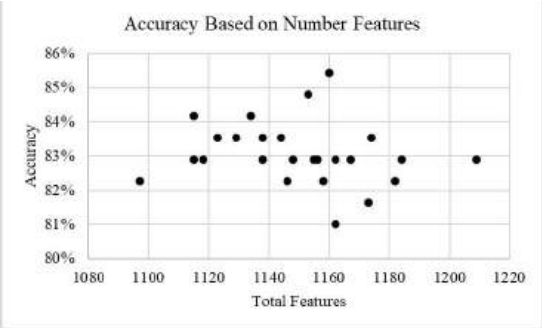


Fig. 6. Comparison accuracy based on total feature selected

The author also conducted an ANOVA analysis to show that the variation between columns, i.e. total features selected and validation accuracy, has a significant effect with an F-value of 50665.6 and a P-value of 2E-41, well below the 0.05 threshold. This indicates that the number of selected features strongly influences accuracy validation. In contrast, the variation between rows represents the iterations with an F-value of 0.99975 and a P-value of 0.50025, which means there is no significant difference between iterations. This result indicates that the main factor affecting accuracy is the number of features selected, while the difference between iterations has no significant impact. Based on the explanation that has been described, the best initial accuracy evaluation results are obtained at epoch 5 population size 25 with a total of 1160 features as in table 2.

TABLE 2 FEATURE SELECTION RESULTS WITH PSO

No	X0	X1	X5	X8	X9	X12	X13	...	X2303
1	0,137	0,004	0,014	0,273	0,169	0,039	0,062	...	0.265
2	0,078	0,082	0,132	0,244	0,039	0,201	0,322	...	0.079
3	0,089	0,080	0,085	0,129	0,085	0,228	0,263	...	0.000
4	0,222	0,098	0,275	0,117	0,247	0,275	0,275	...	0.122
5	0,230	0,027	0,004	0,285	0,110	0,020	0,019	...	0.254
6	0,175	0,060	0,162	0,112	0,093	0,142	0,247	...	0.243
7	0,167	0,062	0,247	0,075	0,176	0,133	0,210	...	0.186
...	...	...	...	...	...	...	...	...	...
158	0,263	0,309	0,088	0,117	0,177	0,309	0,204	...	0.182

From the initial evaluation results, the best accuracy in the classification of healthy, whitish, and dead coral

reefs produced solutions with values {0.817, 0.917, 0.145, 0.077, 0.343, 0.564, 0.324, 0.230, 0.810, 0.670, ... , 0.585, 0.310, 0.813, 0.856, 0.675}. From a total of 2303 features, the selected features that have been selected are 1160 combinations of features, namely {0, 1, 5, 8, 9, 12, 13, 14, 16, 17, 20, 21, 22, 24, 25, 31, 33, 34, 35, 36, ... , 2278, 2279, 2283, 2284, 2287, 2288, 2289, 2290, 2291, 2295, 2298, 2299, 2301, 2302, 2303}, these results are randomly selected by PSO.

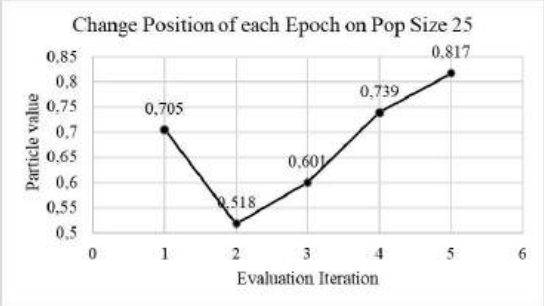


Fig. 7. Position movement of each epoch on pop size 25

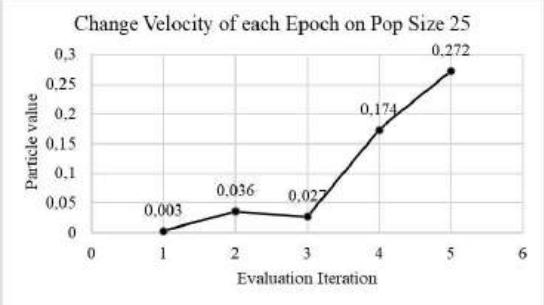


Fig. 8. Movement velocity of each epoch on pop size 25

The following Fig. 7 is a graph of changes in the position of particles in PSO at each epoch with a population size of 25 showing a significant shift at the beginning of the process. In the first epoch, it starts with a relatively high particle position of 0.705, but experiences a fairly significant decrease in epoch 2 with a position of 0.518. This decrease indicates that the particle is exploring the search space to find a more optimal solution. After that, there was a gradual increase in epochs 3 and 4 with positions of 0.601 and 0.739 respectively, until finally reaching the best position at epoch 5 with a value of 0.817. This increase in position indicates that the particles are getting closer to the optimal solution as time goes by.

Meanwhile, the particle velocity in Fig. 8 change graph shows that the particles start with a very low velocity at the first epoch, which is 0.003. This speed increased slightly at epoch 2 to 0.036, but again dropped at epoch 3 with a value of 0.027, indicating that the particles were exploiting deeper around the temporary solution. A significant change occurred at epoch 4, where the velocity jumped to 0.174 and then peaked at epoch 5 with a value of 0.272. Where the acceleration of particle movement that occurs is useful for optimizing its position. This increase shows that the particle moves faster and more intensively to find a better solution at the final stage of the optimization process.

The relationship between position change and velocity shows that higher velocities in the last epochs encourage particles to find more optimal positions. When the particle speed reaches its peak at epoch 5, the particle position also reaches the highest value, which coincides with the highest accuracy of 85.44%. This shows that more intensive particle movement speed helps to find better solutions in coral reef classification. Thus, the combination of increased particle speed and position significantly contributed to the optimal result at the last epoch.

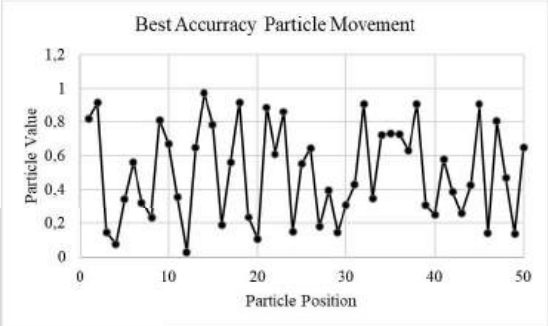


Fig. 9. Particle position movement based on best accuracy

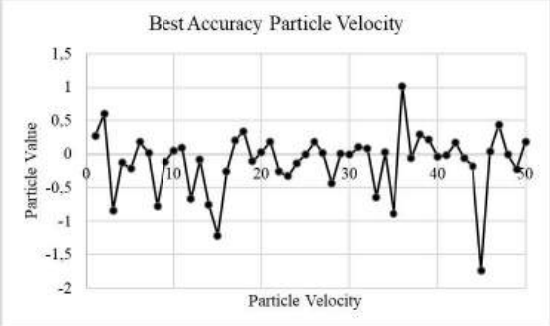


Fig. 10. Particle velocity movement based on best accuracy

Then, the author also looks at how the movement of the velocity and position displacement when it is in the best accuracy, namely epoch 5 population size 25 with 23 particles. In Fig. 9, the graph of particle position changes experiences a significant shift, but remains in a more stable range than the velocity graph, with position values ranging from 0.2 to 1.2. At some points, such as the 5th and 46th positions, the position peaks around the value of 1.2, but also drops to lower values, such as at the 10th and 35th positions. This shifting pattern indicates that particle 23 remains in the process of optimizing the solution space, with more controlled changes in position than in velocity.

And in Fig. 10 the velocity change graph shows a fairly dynamic velocity shift. Where, the particle velocity is around the zero value with some significant peaks, such as at the 30th velocity which reaches a value above 1.5, and a drastic decrease at the 44th velocity which reaches a value close to -2. This pattern

explains that the particles experience rapid changes in acceleration, both in terms of increases and decreases. Such rapid changes in velocity usually reflect intensive exploration of the solution space, where particles move quickly to various points in an attempt to find an optimal solution.

Rapidly moving velocities indicate that the particle is conducting an intensive search, while more stable positions indicate that the particle is focusing on an area that the algorithm considers optimal. The peaks in the velocity graph go hand in hand with larger position changes, indicating that as the particle moves quickly, the position will continue to move to a new, more significant position. At epoch 5 of population size 25, the combination of dynamic speed and stable position helped the particle find the optimal solution in the coral reef classification process.

C. Accuracy Evaluation

PSO is a feature selection method that researchers use to find the best combination of features so as to find the most optimal solution. The use of PSO feature selection improves the performance of the SVM algorithm model by producing the best accuracy of 85.44% at epoch 5 population size 25 as in Fig. 11, while the lowest accuracy is 81.01% at epoch 2 population size 10 and for the best testing accuracy obtained is 80% as in Fig. 12. Researchers used Analysis of Variance (ANOVA) to show statistical analysis conducted on validation accuracy data and testing accuracy data from the classification of healthy, whitish, and dead coral reefs. Two main factors that affect the results of the analysis in this ANOVA are the number of epochs (experiments) and population size in the SVM model used.

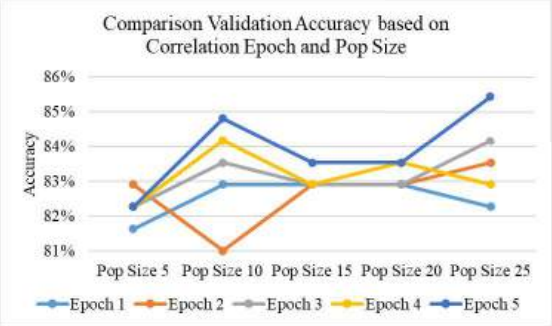


Fig. 11. Comparion validation accuracy based on correlation epoch and pop size

In the ANOVA results of the validation accuracy data, it was found that the F-value for the epoch factor was 2.41026, followed by a P-value of 0.09213. It is also explained that the F-crit value is 3.00692, where this value is relevant at a certain significance level of 0.05. From the results of the analysis it can be seen that, the F-value is smaller than the F crit and the P-value is greater than the general significance level (0.05). It can be concluded that the variation in the number of epochs has no significant effect on the validation accuracy. Then for the population size F-value of 2.08974, with a P-value of 0.12977 and an F crit value of 3.00692. Similar to the epoch factor, where the F-value is smaller than the F crit and the P-value is greater than the general significance level (0.05), indicating that variations in population size have no significant effect on validation accuracy. Based on the results of the ANOVA analysis of the validation accuracy data, it shows that changes in these two factors do not have a significant impact on the validation accuracy results of the coral reef classification model. In other words, there was no significant difference in validation accuracy based on variations in the number of epochs or population size used in the classification model.

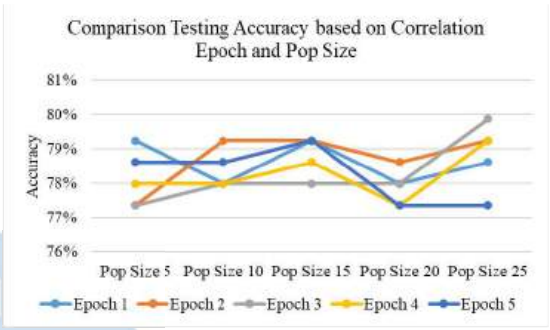


Fig. 12. Comparion testing accuracy based on correlation epoch and pop size

Furthermore, in the ANOVA results of testing accuracy data, the F value for the epoch factor is 0.54874, with a P-value of 0.70264 and a relevant F crit value of 3.00692. Then in the population size section, the F value is 1.84838, with a P-value of 0.16896, and an F crit value of 3.00692. Similarly, the results of the ANOVA analysis of the validation accuracy data show that changes in both factors do not have a large enough impact to be considered significant on the test accuracy results of the coral reef classification model used. In other words, there was no significant difference in testing accuracy based on variations in the number of epochs or population size used. This result is consistent with the results of the validation accuracy data, which also showed that both components did not significantly impact the performance of the model.

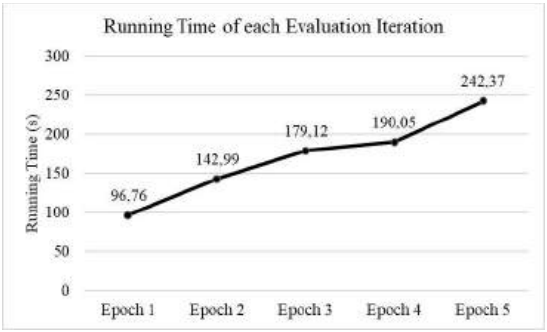


Fig. 13. Running time of each evaluation iteration

In Fig. 13 the graph above shows an increase in the running time of the SVM model with PSO optimization for coral reef classification as the epoch increases, from 96.76 seconds at Epoch 1 to 242.37 seconds at Epoch 5. This increase is due to the increasing complexity of the model and deeper exploration of the solution space by PSO to find the optimal solution. This shows that the optimization process gets more complex as time goes by.

D. Comparison with Ordinary Models

By using the same dataset and data division of 80:10:10, researchers conducted a comparison between the original SVM model and the SVM model optimized using PSO (Particle Swarm Optimization) to see how much influence PSO has on accuracy and running time speed in the classification of healthy, whitish, and dead coral reefs. In the original SVM model, the accuracy result is 79.11% with a running time speed of 916.37 seconds. However, after optimization with PSO, the accuracy of the model increased by 6.33%, reaching 85.44%. In fact, the lowest accuracy of the SVM model results that have been optimized with PSO is still higher than the accuracy of the original SVM model, which is 81.01%. The difference in running time also provides a significant difference of 674 seconds.

This experiment shows that the PSO optimization algorithm has a significant effect on the performance of the SVM model, both in terms of accuracy improvement and running time efficiency. This result confirms that PSO not only improves the model's ability to recognize and classify patterns in coral reef datasets more accurately, but also speeds up the computational process, which is very important in the context of real-time or large-scale applications..

IV. CONCLUSION

This study successfully demonstrated that the combination of feature extraction method with Histogram of Oriented Gradients (HOG), image classification with Support Vector Machine (SVM), and feature selection with Particle Swarm Optimization (PSO) significantly improved the accuracy in the classification of healthy, whitish and dead coral reef images. Experimental results show that the use of PSO successfully increases the accuracy of the SVM model to 85.44%, which is a substantial improvement compared to the original SVM model without optimization, which only achieves an accuracy of 79.11%. In addition, PSO also reduced the running time required for the classification process making it computationally efficient. PSO also showed excellent performance in performing feature selection, by effectively reducing the number of features used in training the model, which originally had a total of 2303 features, at best accuracy only used a total of 1160 features. This not only speeds up the training process but also increases the accuracy of the model in predicting coral reef classification. Although the results achieved are quite satisfactory, this study has some limitations, especially in terms of generalizing the model to new data. Further research using larger and varied datasets is recommended to ensure the robustness and generalization of the resulting model.

ACKNOWLEDGMENT

All praise to God Almighty for all His mercy and grace so that this research can be completed properly. We would also like to thank Dian Nuswantoro University for the support that has been given during this research process. Without continuous help and support, this research could not be completed perfectly. We hope that the results of this research can provide benefits and have a positive impact on the wider community, as well as being a meaningful contribution in the field of science and technology.

REFERENCES

[1] A. Triwibowo, "STRATEGI PENGELOLAAN EKOSISTEM TERUMBU KARANG DI WILAYAH PESISIR," *Jurnal Kelautan dan Perikanan Terapan (JKPT)*, vol. 1, p. 61, Jan. 2023, doi: 10.15578/jkpt.v1i0.12048.

[2] S. Villon *et al.*, "A Deep learning method for accurate and fast identification of coral reef fishes in underwater images," *Ecol Inform*, vol. 48, pp. 238–244, Nov. 2018, doi: 10.1016/j.ecoinf.2018.09.007.

[3] E. Bollati, C. D'Angelo, D. I. Kline, B. G. Mitchell, and J. Wiedenmann, "Development of a multi-excitation fluorescence (MEF) imaging method to improve the information content of benthic coral reef surveys," *Coral Reefs*, vol. 40, no. 6, pp. 1831–1847, Dec. 2021, doi: 10.1007/s00338-021-02101-1.

[4] S. A. Harahap, N. A. Shabrina, N. P. Purba, and M. L. Syamsuddin, "The patterns of changes in coral reef coverage in the Seribu Islands National Park, Jakarta, Indonesia," 1994. [Online]. Available: www.worldnewsnaturalsciences.com

[5] J. Borbon, J. Javier, J. Llamado, E. Dadios, and R. K. Billones, "Coral Health Identification using Image Classification and Convolutional Neural Networks," in *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/HNICEM54116.2021.9731905.

[6] S. Jamil, M. Rahman, and A. Haider, "Bag of features (Bof) based deep learning framework for bleached corals detection," *Big Data and Cognitive Computing*, vol. 5, no. 4, Dec. 2021, doi: 10.3390/bdcc5040053.

[7] C. Jackett *et al.*, "A benthic substrate classification method for seabed images using deep learning: Application to management of deep-sea coral reefs," *Journal of Applied Ecology*, vol. 60, no. 7, pp. 1254–1273, Jul. 2023, doi: 10.1111/1365-2664.14408.

[8] A. Gómez-Ríos, S. Tabik, J. Luengo, A. S. M. Shihavuddin, and F. Herrera, "Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks," *Knowl Based Syst*, vol. 184, Nov. 2019, doi: 10.1016/j.knosys.2019.104891.

[9] A. Gómez-Ríos, S. Tabik, J. Luengo, A. S. M. Shihavuddin, B. Krawczyk, and F. Herrera, "Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation," *Expert Syst Appl*, vol. 118, pp. 315–328, Mar. 2019, doi: 10.1016/j.eswa.2018.10.010.

[10] G. A. Bautista-Hernández, D. A. Jimenez-Nixon, and A. M. Reyes-Duke, "Coral Reef Disease and Bleaching Indentification through Computational Vision Algorithm," in *Proceedings of the 2022 39th IEEE Central America and*

Panama Student Convention Conference, CONESCAPAN 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/CONESCAPAN56456.2022.9959584.

[11] S. Mittal, S. Srivastava, and J. P. Jayanth, "A Survey of Deep Learning Techniques for Underwater Image Classification," *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 10, pp. 6968–6982, Oct. 2023, doi: 10.1109/TNNLS.2022.3143887.

[12] Y. Pei, Y. Huang, Q. Zou, H. Zang, X. Zhang, and S. Wang, "Effects of Image Degradations to CNN-based Image Classification," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.05552

[13] S. Sood, H. Singh, M. Malarvel, and R. Ahuja, "Significance and Limitations of Deep Neural Networks for Image Classification and Object Detection," in *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1453–1460. doi: 10.1109/ICOSEC51865.2021.9591759.

[14] M. Madanan, S. S. Gunasekaran, and M. A. Mahmoud, "A Comparative Analysis of Machine Learning and Deep Learning Algorithms for Image Classification," in *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 2436–2439. doi: 10.1109/IC3I59117.2023.10398030.

[15] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit Lett*, vol. 141, pp. 61–67, Jan. 2021, doi: 10.1016/j.patrec.2020.07.042.

[16] M. Burman, R. Singh, R. Das, S. Gauraha, and S. Kumar, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *International Journal of Scientific Research in Engineering and Management*, 2023, doi: 10.55041/IJSREM23473.

[17] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. Gnrsn Prudhvith, and N. Kumar, "Image Classification using SVM and CNN," in *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*, Institute of Electrical and Electronics Engineers Inc., Mar. 2020. doi: 10.1109/ICCSEA49143.2020.9132851.

[18] H. Hasan, H. Z. M. Shafri, and M. Habshi, "A Comparison between Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Models for Hyperspectral Image Classification," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1755-1315/357/1/012035.

[19] N. Boonnam, T. Udomchaipitak, S. Puttinaovarat, T. Chaichana, V. Boonjing, and J. Muangprathub, "Coral Reef Bleaching under Climate Change: Prediction Modeling and Machine Learning," *Sustainability (Switzerland)*, vol. 14, no. 10, May 2022, doi: 10.3390/su14106161.

[20] N. Ani Brown Mary and D. Dharma, "Coral reef image classification employing Improved LDP for feature extraction," Nov. 01, 2017, *Academic Press Inc.* doi: 10.1016/j.jvcir.2017.09.008.

[21] B. Hamuna, S. Pujiyati, J. L. Gaol, and T. Hestirianoto, "Spatial distribution of benthic habitats in Kapota Atoll (Wakatobi National Park, Indonesia) using remote sensing imagery," *Biodiversitas*, vol. 24, no. 7, pp. 3700–3707, 2023, doi: 10.13057/biodiv/d240706.

[22] Hartoni, V. P. Siregar, S. Wouthuyzen, and S. B. Agus, "Object based classification of benthic habitat using Sentinel 2 imagery by applying with support vector machine and random forest algorithms in shallow waters of Kepulauan Seribu, Indonesia," *Biodiversitas*, vol. 23, no. 1, pp. 514–520, Jan. 2022, doi: 10.13057/biodiv/d230155.

[23] J. J. Gapper, H. El-Askary, E. Linstead, and T. Piechota, "Coral reef change detection in remote Pacific Islands using support vector machine classifiers," *Remote Sens (Basel)*, vol. 11, no. 13, Jul. 2019, doi: 10.3390/rs11131525.

[24] N. W. Prabowo, V. P. Siregar, and S. B. Agus, "KLASIFIKASI HABITAT BENTIK BERBASIS OBJEK DENGAN ALGORITMA SUPPORT VECTOR MACHINES DAN DECISION TREE MENGGUNAKAN CITRA MULTISPEKTRAL SPOT-7 DI PULAU HARAPAN DAN PULAU KELAPA," *Jurnal Ilmu dan Teknologi Kelautan Tropis*, vol. 10, no. 1, pp. 123–134, Apr. 2018, doi: 10.29244/jitkt.v10i1.21670.

[25] S. Villon, M. Chaumont, G. Subsol, S. Villéger, T. Claverie, and D. Mouillot, "Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between deep learning and HOG+SVM methods," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2016, pp. 160–171. doi: 10.1007/978-3-319-48680-2_15.

[26] A. Mehta, E. Ribeiro, J. Gilner, and R. Van Woesik, "CORAL REEF TEXTURE CLASSIFICATION USING SUPPORT VECTOR MACHINES," 2007.

[27] E. Ribeiro, A. Mehta, J. Gilner, and R. Van Woesik, "CORAL REEF TEXTURE CLASSIFICATION USING SUPPORT VECTOR MACHINES," 2007. [Online]. Available: https://www.researchgate.net/publication/221416032

[28] Z. Lu, "Enhanced Accuracy Enabled by Particle Swarm Optimization in Classification Application," in *Proceedings - 2020 International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 146–149. doi: 10.1109/ICAICE51518.2020.00034.

[29] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection." [Online]. Available: http://lear.inrialpes.fr

[30] N. Ani Brown Mary and D. Dharma, "A novel framework for real-time diseased coral reef image classification," *Multimed Tools Appl*, vol. 78, no. 9, pp. 11387–11425, May 2019, doi: 10.1007/s11042-018-6673-2.

[31] F. T. Anggraeny, B. Rahmat, and S. P. Pratama, "Deteksi Ikan Dengan Menggunakan Algoritma Histogram of Oriented Gradients," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 15, no. 2, p. 114, Sep. 2020, doi: 10.30872/jim.v15i2.4648.

[32] C.-J. Tu, L.-Y. Chuang, and C.-H. Yang, "Feature Selection Using PSO-SVM," Feb. 2007.

[33] G. Jin, P. Jin-Ye, and L. Zhan, *Application of Improved PSO-SVM Approach in Image Classification*. IEEE, 2010.

[34] P. N. Andono, G. F. Shidik, D. P. Prabowo, D. H. Yanuarsari, Y. Sari, and R. A. Pramunendar, "Feature Selection on Gammatone Cepstral Coefficients for Bird Voice Classification Using Particle Swarm Optimization," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, pp. 254–264, 2023, doi: 10.22266/ijies2023.0228.23.

[35] O. Beijbom, P. J. Edmunds, D. I. Kline, B. G. Mitchell, and D. Kriegman, "Automated annotation of coral reef survey images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1170–1177. doi: 10.1109/CVPR.2012.6247798.

[36] K. Thurnhofer-Hemsi, E. López-Rubio, M. A. Molina-Cabello, and K. Najarian, "Radial basis function kernel optimization for Support Vector Machine classifiers," Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.08233

[37] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology (Singapore)*, vol. 13, no. 5, Oct. 2021, doi: 10.1007/s41870-017-0080-1.

ISSN 2355-0082

Enhancing Support Vector Machine Classification of Nutrient Deficiency in Rice Plants Through Particle Swarm Optimization-Based Feature Selection

James Hartojo[1], Jessica Carmelita Bastiaans[2], Ricardus Anggi Pramunendar[3], Pulung Nurtantio Andono[4]

[1,2,3,4] Informatics Engineering, Dian Nuswantoro University, Semarang, Indonesia

[1] jameshartojo@gmail.com, [2] jessicacarmelita2004@gmail.com, [3] ricardus.anggi@dsn.dinus.ac.id, [4] pulung@dsn.dinus.ac.id

Accepted 2 September 2024
Approved 6 January 2025

***Abstract*— The research focuses on the classification of nutrient deficiencies in rice plant leaves using a combination of Support Vector Machine (SVM) and Particle Swarm Optimization (PSO) methods for feature selection. Image features are extracted using Histogram of Oriented Gradients (HOG), which is then optimized with PSO to select the most relevant features in the classification process. Indonesia is one of the largest rice producers in the world, with food security as a major issue that requires sustainable solutions, especially in the agricultural sector. The growth and yield of rice plants are highly dependent on the availability of nutrients such as Nitrogen (N), Phosphorus (P), and Potassium (K). However, traditional observation methods to detect nutrient deficiencies in plants become inefficient as the scale of production increases. The dataset used includes images of rice leaves showing nitrogen (N), phosphorus (P), and potassium (K) deficiencies. Experiments show that the SVM model optimized with PSO provides a classification accuracy of 83.19% and a runtime of 129.63 seconds with 1150 best feature combinations out of 2303 extracted features, which is higher accuracy and faster runtime than the model that does not use PSO. These results show that the integration of PSO in the feature selection process not only improves the accuracy of the model, but also reduces the required computation time. This research makes an important contribution to the development of an automated system for the classification of nutrient deficiencies in crops, which can be implemented in large farms or other agricultural fields.**

***Index Terms*— Histogram of Oriented Gradients (HOG); Nutrient Deficiency; Particle Swarm Optimization (PSO); Rice plants; Support Vector Machine (SVM).**

I. INTRODUCTION

Rice is the staple food for more than half of the world's population, so it is very important for food security [1]. According to USDA (U.S Department of Agriculture) Foreign Agricultural Service on Rice-USDA-Foreign-Agricultural-Service (fas.usda.gov), Indonesia is the 4th largest rice producer in the world with a total production of 33 million tons in 2023. With an area of only 7% of Indonesia's land area, Java Island is still the largest contributor to rice production, reaching 53% of the total national rice production [1].

Food security implies that people can consume the ingredients they need for an active and healthy lifestyle at any time, in any quantity or variety. In this context, "food security" is usually considered, meaning when it is physically and economically feasible to use. In terms of physical availability of food, it is natural to consider the amount and variety of food products that meet effective demand in their area of demand. This refers to the food supply of a region or the food supply of a country [2].

To meet the needs of a growing population along with the depletion of food supplies, the world faces a number of serious problems related to the availability of sufficient food in a sustainable way [3]. Food security, quality food production and increasing crop yields are major global challenges in the agricultural sector. Plant growth or yield is highly dependent on several nutrients. Nutrients are essential for overall plant growth and development [4].

Since long after primitive agriculture emerged, field observation has been a common method for monitoring crop growth and identifying plant diseases. However, with the tremendous increase in crop yields, traditional observation methods are inefficient for systematic management and assessment. Meanwhile, aberrations arise when the eyes can hardly distinguish a series of similar crop symptoms [5]. Moreover, inspecting long fields or crops can be tiring and demand a lot of effort. Therefore, an automated mechanism is needed to detect nutrient-deficient vascular plants in a precise and timely manner. Image-based plant nutrient deficiency identification is a promising and efficient solution as it

is a non-invasive, efficient, and accurate method that can be applied in large fields [4].

Given the rapid development of artificial intelligence (AI) such as Machine Learning that can provide innovative solutions to increase rice production, the main source of food for more than half of the world's population. By utilizing it to detect nutrient deficiencies in rice plants quickly and accurately, food security challenges can be addressed more effectively, supporting the sustainability of rice production amidst limited resources and increasing global demand. This research aims to classify nutrient deficiencies in rice plant leaves using image processing and machine learning techniques. Machine learning in agriculture is not new, and several machine learning approaches have been implemented to support agricultural processes, control, or monitoring [6]. Our research will focus on nutrient deficiencies of Nitrogen (N), Phosphorus (P), and Kalium or the same as Potassium (K). The most important nutrients for plants are Nitrogen (N), Phosphorus (P), and Potassium (K). Some specific changes appear in rice leaves due to loss of mineral balance. The scarcity of macronutrients, especially nitrogen, phosphorus and potassium in rice, can result in obvious symptoms [7].

There are several previous studies related to similar problems or related to the classification of nutrient deficiencies in crops or harvests using various machine learning and deep learning. The Support Vector Machine (SVM) method is able to efficiently find the ideal combination of classification bands while maintaining excellent classification accuracy, as shown by Juan W's experiments [8], [9]. Especially in image classification, SVM has the highest performance compared to other machine learning models such as, K-Nearest Neighbor (KNN), Naïve Bayes (NB), Binary Decision Tree (BDT) and Discriminant Analysis (DA) [10]. Research on the classification of nutritional deficiencies in corn plant leaves using the SVM model and achieved 80% accuracy [11]. Classification of nutritional deficiencies in chili plants using the SVM model without feature extraction and achieved 84% accuracy [12].

Some research related methods researchers use Deep Learning, namely, Convolutional Neural Network (CNN). CNN is one of the best image processing approaches in Artificial Intelligence that implements general and detailed tasks [13]. The accuracy of related researches get a high accuracy score in the range of 85% - 99 [4], [5], [6], [14], [15], [16]. Unfortunately, one of the disadvantages of CNN models is that they usually require large datasets to train a decent model [8], [9], [17]. SVM has higher accuracy when testing small datasets. Small data will result in a high difference in training and testing accuracy so that the model will experience overfitting problems [18]. In addition, the time required for testing is also faster than CNN [8], [9], [19].

Some studies use similar Machine Learning models, namely, Support Vector Machine (SVM) assisted with feature extraction such as Histogram Oriented Gradients (HOG) but unfortunately do not use optimization methods such as Particle Swarm Optimization (PSO) which can select features to improve classification performance.

Research by Zhe Xu et al. on a similar problem used the SVM model and HOG extraction features and got a classification accuracy of 56.86%. The accuracy is relatively low due to weaknesses, among others, not using the PSO optimization method [20]. Research by Prabira Kumar et al. on the lack of Nitrogen elements in rice plants that use SVM models and HOG extraction features and get classification accuracy at 55.30% [14]. Just like previous studies, this study did not use PSO optimization methods that could improve model performance. Investigation of machine learning and deep learning models to classify potassium and healthy leaves from grapevine. Researchers used HOG and Principal Component Analysis (PCA) to reduce features and then input to SVM and achieved 66.70% classification accuracy [21]. The weakness of the model could be due to the use of PCA which is less optimal than the PSO method.

Compared to other feature selection methods, the PSO-SVM method simplifies feature selection and effectively reduces the required parameters, resulting in higher classification accuracy most of the time [22]. PSO method is proven to be more effective than other optimization methods such as PCA, especially in improving classification performance in the SVM algorithm. The advantage of PSO lies in its ability to perform better global exploration in the solution space, resulting in more optimal SVM parameters than other optimization methods. Therefore, we conducted research using SVM model with HOG extraction features complemented by PSO method.

In previous research, image classification models using Support Vector Machine (SVM) with Histogram Oriented Gradients (HOG) extraction and Particle Swarm Optimization (PSO) methods were rarely used. Being the right target to achieve a goal by using the HOG feature extraction method and PSO feature selection to select the best features based on a collection of particles. This method is considered suitable to be applied because it aims to achieve a higher level of accuracy more efficiently. Comparing the classification results of nutrient deficiency images on rice plants shows the contributions of this research: (1) testing the effect of PSO method on SVM classification model; (2) investigating the relationship between PSO method and HOG feature extraction; and (3) finding solutions and combinations of selected features to achieve accurate performance of nutrient deficiency image classification in rice plants.

This document is organized in 4 chapters: chapter 1 discusses the background of the research, chapter 2

discusses the model and method of our research, chapter 3 describes the evaluation details and experimental results, and chapter 4 describes the conclusion of the research.

II. METHOD

A. Research Design

The study aims to measure the success rate of the combination of the Support Vector Machine (SVM) model with Histogram of Oriented Gradients (HOG) feature extraction and the Particle Swarm Optimizations (PSO) feature optimization method in classifying rice plants that are deficient in Nitrogen, Phosphorus, and Potassium nutrients based on the accuracy rate while getting the best combination of features. HOG extracts features from each image in the dataset and then selects the best features by PSO to be input to the SVM model as classification training material.

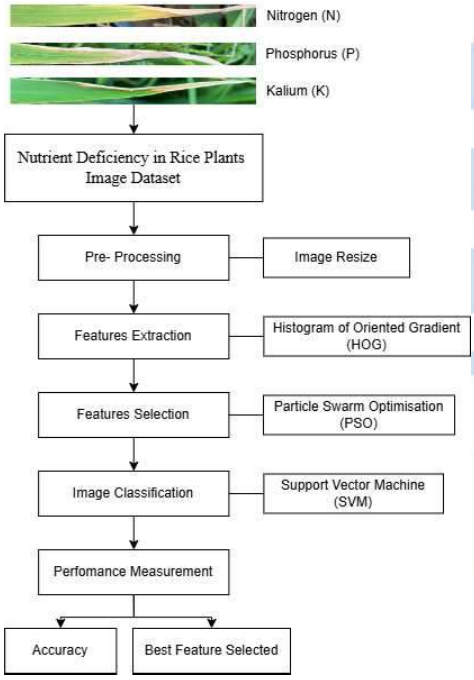


Fig. 1. Research design diagram

Based on Figure 1, the classification process begins by dividing the Nutrient Deficiency in Rice Plants dataset into 70% training data, 20% validation data, and 10% testing data. The images from the dataset were resized to 1024x256 pixels using cv2 to save resources. After that, image features were extracted using the HOG method with parameters of 9x9 pixels per cell, 2x2 cells per block, and 9 directions in the orientation histogram, then normalized with L2-Hys norm before proceeding to feature selection by PSO.

Feature selection is performed using the PSO method with 5 epochs, pop size up to 25, and perturbation rate 0.05 to get the best feature combination. The best feature combination was then used to train the RBF kernel-based SVM model with parameters C = 10.0 and gamma = 'scale'. Proper parameter adjustment will allow the SVM to find better classification hyper-fields, thereby improving classification accuracy [22].

Performance evaluation is done by comparing the accuracy values on the tested datasets, and the classification results, including the selected feature combinations and model performance, are saved in an Excel file for further analysis. PSO-SVM can be an ideal pre-processing tool to help optimize the feature selection or classification process as it can improve classification accuracy while keeping the required computational resources to a minimum [22].

B. Data Gathering and Preparation

In this research, two initial stages were carried out, namely data collection and processing to prepare the dataset. The dataset used in this research is an image of a rice plant that lacks Nitrogen (N), Phosphorus (P), and Potassium (P) Nutrient-Deficiency-Symptoms-in-Rice (kaggle.com). Table 1 provides the raw dataset details.

TABLE I. IMAGE DATASET DETAILS

Image Label	Image Quantity
Nitrogen (N)	440
Phosphorus (P)	383
Kalium (K)	333
Total	1156

We then divided the dataset into three parts, namely 70% training, 20% testing, and 10% validation. The training data consists of 809 images including 308 images of rice plants lacking Nitrogen, 233 images of Phosphorus deficiency, and 268 images of Potassium deficiency then the testing data consists of 232 images including 88 images of rice plants lacking Nitrogen, 67 images of Phosphorus deficiency, and 77 images of Potassium deficiency and, for validation consists of 115 images including 44 images of rice plants lacking Nitrogen, 33 images of Phosphorus deficiency, and 38 images of Potassium deficiency.

Fig. 2. Nitrogen deficiency rice plant

Fig. 3. Phosphorus deficiency rice plant

Fig. 4. Potassium deficiency rice plant

Figure 2 until 4 are examples of rice plant images from each label in the dataset. The images in the dataset are relatively large in pixel size and resolution giving a rectangular shape. Therefore, the researcher resizes the images in the dataset to 1024 x 256 to save resources and speed up the training process. When resized to 256 x 256 (square shape image), the classification accuracy becomes smaller because it does not match the basic resolution image.

The set of image data contained in the training data will be processed to obtain information on the unique characteristics of nutrient deficiencies in each rice plant. HOG feature extraction helps the process of extracting image features from the characteristics of rice plants lacking Nitrogen, Phosphorus, and Potassium nutrients and removing noise detected in the images.

C. Features Extraction

Histogram of Oriented Gradients (HOG) is a method for extracting features from images to facilitate image analysis. HOG is a typical image feature that is widely used in various fields of image research [23]. HOG uses windows to generate descriptors that are local to the detected key points of the image. A window consisting of a regular square grid ($n \times n$) is centered on the key point under consideration, and for each cell in the grid, a frequency histogram is generated to describe the edge orientation distribution [24].

The process of feature extraction by HOG starts by converting the original image into a grayscale image.

$$Gray = 0.3 * R + 0.59 * G + 0.11 * B \quad (1)$$

This process involves converting a color image into a gray image, where R, G, and B represent the color components of the corresponding image positions (1). Next, the algorithm calculates the gradient of each pixel of the image separately. For each pixel (x, y) in the image, the gradient is calculated using the horizontal (Gx) and vertical (Gy) gradients, while θ(x,y) is the orientation of the gradient at that point (2), (3), (4), (5).

$$Gx = I(x+1, y) - I(x-1, y) \quad (2)$$

$$Gy = I(x, y+1) - I(x, y-1) \quad (3)$$

$$G(x, y) = \sqrt{G_x^2 + G_y^2} \quad (4)$$

$$\theta(x, y) = arctan\left(\frac{G_y}{G_x}\right) \quad (5)$$

After that, the algorithm divides the image into small unit cells with the size corresponding to `pixels_per_cell`, i.e. 9x9 pixels per cell, and calculates the directional gradient of each unit cell calculated with 9 directions. Histogram generation is performed in each unit cell based on the values of θ(x,y) in that cell.

Next, the algorithm groups each cell into blocks according to `cells_per_block = (2,2)` and the gradient histograms of the cells in the blocks are combined. The gradient histograms of the cells in a block are then normalized using the L2-Hys norm method, which can be written by a certain formula where H is the histogram of the block, and ϵ is a small value to prevent division by zero (6).

$$H_{norm} = \frac{H}{\sqrt{\|H\|_2^2 + \epsilon^2}} \quad (6)$$

The normalized histogram of each block is finally combined into one long feature matrix, which serves as the result of HOG feature extraction [23].

The HOG feature matrix results will then continue the feature selection process by PSO and be inputted for the SVM model training process..

D. Features Selection

Particle Swarm Optimization (PSO) is an optimization method that selects extracted features to determine combinations as potential solutions from many available features. PSO simulates a flock of birds to describe an automatically evolving system where each candidate solution is a “flock of birds”, a particle in the search space. Each particle uses its memory and overall knowledge to find the best solution [22].

The PSO process starts with the initialization of a number of particles based on the population size (pop_size). Each particle has a random position in the search space bounded by bounds that define the range of values of each feature in the solution. Each particle also has an initial velocity that will be changed at each iteration.

Furthermore, at each iteration or epoch, each particle will be evaluated using an objective function. This objective function measures the accuracy of the SVM classification model on the validation data based on the features selected by the particle position. In our study, a feature will be selected if the solution of the feature is greater than 0.5. This accuracy is then used to determine whether the solution generated by the particle is the best so far, both individually (for the particle itself) and globally (for the entire particle population).

After evaluation, the velocity and position of each particle is updated using the formula below (7), (8). The new velocity is calculated based on a combination of the previous velocity, the particle's distance from its own best position, and the particle's distance from the best position of the entire population. Random factors

(r1 and r2) and acceleration coefficients (c1 and c2) are used in this update to ensure balanced exploration of the search space. The new position of the particle is then obtained by adding this new velocity to the current position. The following is the formula for updating particle velocity and particle position:

$$v_i^{t+1} = w * v_i^t + c_1 * r_1 * (c - x_i^t) + c_2 * r_2 * (g^{best} - x_i^t) \quad (7)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (8)$$

Table 2 shows the description of the symbols in the PSO formulas that are displayed (7), (8).

TABLE II. PSO EQUATION DESCRIPTION

Symbol	Explanation
v_i^t	Velocity of the i particle at the t iteration
x_i^t	The position of the i particle at iteration t
p_i^{best}	The best position ever reached by particle i
g^{best}	The best position ever achieved by the entire population
w	An inertia factor that controls how much the particle maintains its previous velocity
c_1 dan c_2	An acceleration coefficient that controls how much the particle is attracted to the local and global best solutions
r_1 dan r_2	A random number between 0 and 1

This update process is repeated until the specified number of iterations is reached. Whenever a better solution is found, either by an individual particle or by the entire population, it is saved.

Once all iterations are complete, the best solution found throughout the process is considered the final result. This solution is then used to measure the performance of the model on the test set, with the aim of validating how well the features selected by PSO improve the accuracy of the classification model [25].

The output of the PSO process produces positions, velocities, and fitness. Positions are particle positions that indicate potential solutions to the search space, and each position element indicates whether a particular feature was selected in the classification model. Velocities are velocities that determine the change in particle position per iteration influenced by the best solution found by the particle and the population as a whole. Fitness is an assessment of solution quality based on model accuracy, and PSO aims to steer particles to the best solution to maximize fitness value.

E. SVM Classification Model

Support Vector Machine (SVM) is considered as one of the efficient machine learning methods developed based on statistical learning theory [26]. SVM, which is a supervised learning algorithm, generates decision boundaries to separate the n-dimensional space into classes [15].

SVM performs classification by transforming the original training data into a multidimensional space and constructing a hyper-plane in a higher dimension.

One of the best concepts used in SVM is the kernel function. The kernel function is a solid mathematical method used for nonlinear mapping for high-dimensional data. Through this SVM can solve the higher dimensional classification of the initial input data space set. In general, this function calculates the dot product value that maps the data points into the feature space [27].

There are several kernel options that can be used in SVM models such as, Linear, Polynomial, Radial Basis Kernel (RBF), and Sigmoid. In this study, we use the RBF kernel. Radial basis kernel functions are widely used with SVM because they select a smooth solution [27]. Radial basis kernel functions are widely used because of their strong classification ability [Pin Wang]. The following is an explanation of the RBF kernel formula used in the SVM formula:

$$K(\chi, \chi_i) = \exp\left(-\frac{\|\chi_i - \chi\|^2}{\sigma^2}\right) \quad (9)$$

When choosing RBF as a kernel there are two parameters to optimize: the penalty factor C and the RBF parameter σ (sigma), where σ is a constant to adjust the width of the Gaussian function of the kernel that must be optimized to ensure the SVM classification model runs optimally [28]. There is also a parameter γ (gamma) that can be used directly in the RBF kernel formula, where γ determines how much influence a single data point has. The relationship between γ and σ is shown (10).

$$\gamma = \frac{1}{2\sigma^2} \quad (10)$$

Then from the relationship between the 2 parameters above, the RBF kernel formula is produced in (11).

$$K(\chi, \chi_i) = e^{-\gamma x - x_i^2} \gamma > 0 \quad (11)$$

The C parameter in the SVM researcher is 10.0 and γ is 'scale' which means that the gamma value is calculated based on the number of features and data variability. An SVM model equation formula is obtained as shown (12).

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K(\chi_i \chi) + b) \quad (12)$$

Table 3 shows the description of the symbols in the SVM formulas that are displayed (9), (10), (11), (12).

TABLE III. SVM EQUATION DESCRIPTION

Symbol	Explanation
e	The basis of the exponential function e^x
exp(x)	Notation for the exponential function e^x
$sign$	The sign function, which returns +1 if the argument is positive and -1 if the argument is negative. It specifies the class label of x
$\sum_{i=1}^{n} \alpha_i$	Indicates that the sum is taken of all support vectors i from 1 to n. α_i is the Lagrange multiplier, which is a coefficient determined during SVM training. This coefficient is non-zero only for the support vector
y_i	The actual class label of the training sample, which can be +1 or -1
$K(\chi_i \chi)$	The kernel function, which calculates the similarity between input x and each support vector χ_i. The kernel function allows SVMs to work in a high-dimensional feature space.
χ_i	The support vector of the training data, which is the critical data point that lies closest to the decision boundary.
b	The bias term, which shifts the decision boundary.

F. Performance Measurement

Measurement of accuracy results using the Support Vector Machine (SVM) model with a dataset division of 70% training data and 30% testing data. The evaluation results of the classification are in the form of the final accuracy value and the best feature combination obtained based on feature selection in the PSO method. The following is the formula for calculating the final accuracy (13).

$$Accuracy = \frac{y_{pred}}{y_{true}} \times 100 \qquad (13)$$

The selected column indicates which features have been selected by the PSO algorithm to train the classifier. "Total Features" is the number of features selected (14).

$$Total\ Features = Number\ of\ Columns$$
$$where\ Solution\ >=\ 0.5 \qquad (14)$$

The classification process of images of rice plants that lack Nitrogen, Phosphorus, and Potassium using the Support Vector Machine (SVM) model with HOG feature extraction and PSO feature optimization method produces an accuracy value of 83.19%. PSO helps select features to get the best combination of features and get the best 1150 features from 2303 features based on the highest accuracy obtained.

III. RESULT AND DISCUSSION

This chapter will explain the results and analysis of each stage starting from feature extraction by HOG, feature selection by PSO, and accuracy performance results obtained by researchers.

A. Result of HOG Features Extraction

Histogram Oriented of Gradients (HOG) Feature Extraction successfully performed the feature extraction process from 115 rice plant images and obtained a total of 2304 extracted features. As shown in Table 4.

TABLE IV. FEATURES EXTRACTION RESULT

No. Data	X0	X1	X2	X3	X4	…	X2303
1	0.0962	0.0534	0.2108	0.3646	0.3646	…	0.0001
2	0.0739	0.0539	0.0637	0.2451	0.3536	…	0.0337
3	0.2805	0.1308	0.1038	0.0610	0.0540	…	0.0368
4	0.2805	0.1308	0.1038	0.0622	0.0718	…	0.0368
5	0.2805	0.1308	0.1038	0.0490	0.4410	…	0.0368
…	…	…	…	…	…	…	…
115	0.0954	0.0600	0.0930	0.2690	0.3153	…	0.0053

The PSO feature optimization method will help perform the feature selection process to find the best combination of features based on the selected solution. PSO utilizes the relationship between particle positions to find the best fitness value, where the fitness value of the current position will be compared with the personal best (pbest) value obtained in each iteration.

B. Result of PSO Features Selection

PSO is an optimization method that functions to select features that can improve the performance and efficiency of the classification model. We ran the process through 5 epoch iterations and a pop size of 25.

1) Features Selected

The results showed that with 1150 best feature combinations gave the highest testing accuracy at 83.19% with validation accuracy at 88.71% at epoch 1 and pop size 25.

After passing the feature selection through the PSO optimization method, for example, one of the best feature selection results is that 1150 features are obtained, which means that some of the previously

extracted feature columns have been removed and Table 5 shows the following feature selection results that are obtained.

TABLE V. FEATURES SELECTION RESULT

No. Data	X0	X2	X3	X4	X5	…	X2303
1	0.0962	0.2108	0.3646	0.3646	0.0971	…	0.0001
2	0.0739	0.0637	0.2451	0.3536	0.2735	…	0.0337
3	0.2805	0.1038	0.0610	0.0540	0.0480	…	0.0368
4	0.2805	0.1038	0.0622	0.0718	0.0566	…	0.0368
5	0.2805	0.1038	0.0490	0.4410	0.3699	…	0.0368
…	…	…	…	…	…	…	…
115	0.0954	0.0930	0.2690	0.3153	0.2539	…	0.0053

With a combination of selected features, namely [0, 2, 3, 4, 5, 7, 9, 12, 14, 15, ..., 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2298, 2303] which means there are 1153 features removed.

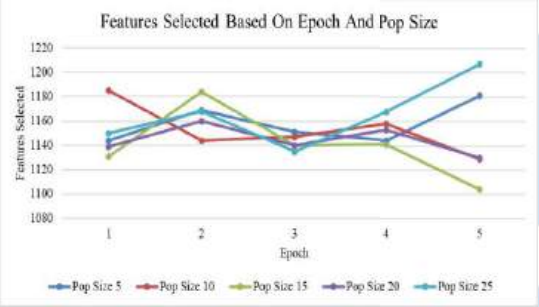


Fig. 5. Features selected comparison by epoch and pop size

As illustrated in Figure 5, epoch and pop size do not have a significant effect on the selected features, but the effect varies. The smallest combination of features is also obtained at 1104 at epoch 5 and pop size 15, while the largest combination of features is 1207 at epoch 5 pop size 25. Epoch 1 to Epoch 5 show an average number of selected features from 1142.6 to 1150.2, with a relatively large variance from 40.3 to 1791.7. Higher epochs do not always result in more features being selected, for example, epoch 3 has the fewest features with high stability and more stable variance.

Pop Size 5 to Pop Size 25 shows an average number of selected features from 1140 to 1165.6, with variance from 143.3 to 828.5. Larger pop sizes tend to produce more selected features, but with higher variance and show stability in feature selection as the pop size increases.

Based on the ANOVA results, the epoch factor has an F value of 0.642 and a P-value of 0.641. The F value is below the crit F value of 3.007 and the P-value is far above 0.05 as a general significance standard. Then the pop size factor has an F value of 1.014 and a P-value of 0.429. The F value is below the F crit value of 3.007 and the P-value is above 0.05. Based on the values of the ANOVA results, it shows that the variance of 5 epochs and pop size 5 to 25 does not have a significant effect or change on the number of features selected in the PSO method.

2) Solution Particle Movement

The solution row generated from each experiment is the position of a particle in the search space describing potential solutions that are updated continuously based on the particle's personal experience and information from other particles in the swarm.

For example, the following best solutions were obtained based on the highest testing accuracy of 83.19% at epoch 1 and pop size 25, namely [0.586, 0.491, 0.666, 0.832, 0.740, ... , 0.919, 0.582, 0.647, 0.368, 0.242].

The particle position range is from 0.0003 to 0.999 which covers almost the entire range from 0 to 1. The mean particle position result of 0.499 shows the particle positions tend to be evenly distributed around the center point which means the exploration is not significant towards either side of the space. The standard deviation of 0.266 indicates there is considerable variation in particle positions around the mean. Most of the particles are in the range of 0.233 to 0.765 indicating the particle positions are widely spread in the search space and PSO is still actively moving in search of the best solution and PSO is not trapped in a local solution.

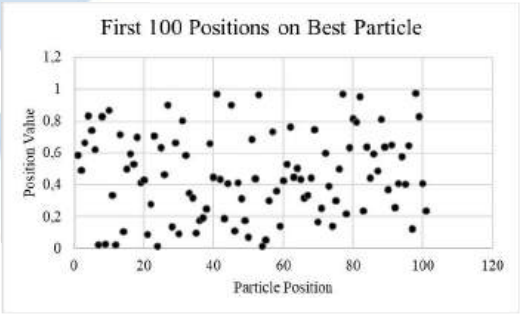


Fig. 6. 100 position changes on best particle

The scatter plot in Figure 6 is an illustration of the best particle position which contains the distribution of the first 100 positions on the particle. The distribution of the particle position points is quite even and shows that the particle explores quite widely in the search space, trying various positions to find the optimal solution.

As explained earlier, the particle movement is constantly changing without any initial convergence to a particular position. Some position points are more often in the range 0.4 to 0.8, but the points in this range are not very concentrated. This suggests that the particles often find better fitness values in the middle of the range, while still exploring the entire range.

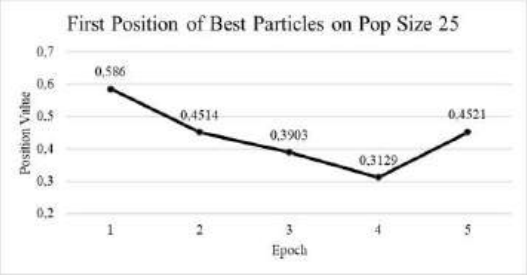


Fig. 7. First position changes on each epoch and pop size 25

Researchers analyzed the movement of the first position of each best particle in each epoch with a pop size of 25 in the first position as shown in Figure 7.

It is known that at epoch 1 the best testing accuracy is obtained so that it can be said that the initial position value of 0.586 is better than in the following epochs. From epoch 2 then there is a decrease until epoch 4 with a position value of 0.313 and the difference is 0.273. The decrease indicates that the particle explores but does not provide an increase in accuracy compared to epoch 1. Then there is an increase to epoch 5 with a position value of 0.452 which indicates that the algorithm tries to return to an area closer to the optimal position at epoch 1.

3) Velocities

The velocities line shows the speed of the particles in making movements to change position and explore for optimal solutions.

For example, the best velocities are obtained as follows based on the highest testing accuracy of 83.19% at epoch 1 and pop size 25, namely [1.588, -0.494, -0.229, 0.652, 0.175, ... , 0.223, 0.048, 0.087, -0.124, -0.544].

The range of particle velocities is from -1.992 to 1.927 indicating that particles can move quickly, either towards a better position or away from the optimal position. The mean value of 0.003 indicates that the particle displacement generally hovers around 0, indicating that the system is in a state of balance or convergence. The standard deviation of 0.496 indicates that there is considerable variation in the particle velocity which means that the actual velocity of the particles varies considerably even though the average is close to zero.

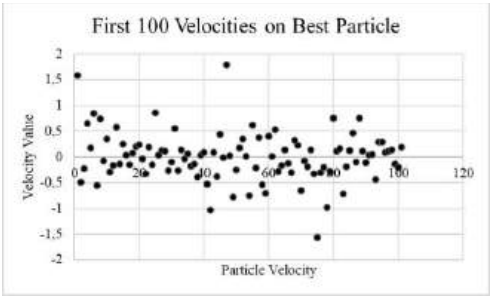


Fig. 8. 100 velocity changes on best particle

The scatter plot in Figure 8 is an illustration of the velocity at the best position which contains the distribution of the first 100 velocities on the particle. The distribution of particle velocity points with a range from about -2 to 2 however, looks symmetrical around zero which indicates the particle velocity changes in both directions (positive and negative).

As explained earlier, the distribution of velocity changes is symmetrical around zero with high variability and gets sparser as it moves away from zero indicating the particle velocity tends to be close to a normal distribution.

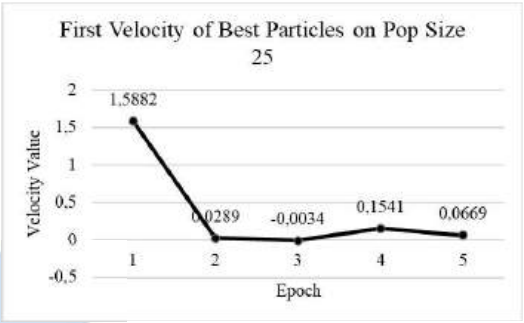


Fig. 9. First velocity changes on each epoch and pop size 25

Researchers analyzed the first speed of each best particle in each epoch with a pop size of 25 in the first position as shown in Figure 9.

Obtaining the best testing accuracy at epoch 1 with a speed value of 1.558 shows that a wide and fast initial exploration helps particles find the optimal solution. The drastic change in epoch 2 at 0.029 shows that the best particle experienced a very significant decrease in velocity after epoch 1. This indicates the system's efforts to reduce exploration and focus on exploitation. The velocity change from epoch 2 to epoch 5 becomes smaller and more stable, this change leads to the stability of the best particle velocity in the next few epochs. The change from a negative value at epoch 3 of -0.003 to a positive value at epoch 4 of 0.154 indicates that the best particle may be adjusting its position to stay around a more optimal solution.

C. Classification Accuracy

Researchers get the highest validation accuracy of 90.00% at epoch 2 and pop size 10, while the lowest is 85.22% at epoch 1 pop size 5 as depicted in Figure 10.

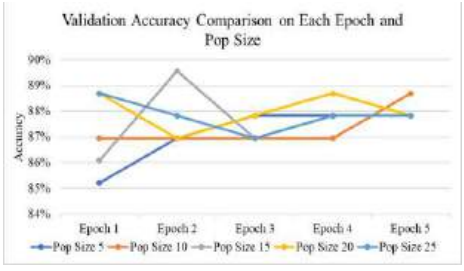


Fig. 10. Validation accuracy comparison by every epoch and pop size

The average validation accuracy between epochs has a range from 87.13% to 87.83%, indicating that epoch changes have a relatively small effect on validation accuracy. The smallest variance at epoch 5 and the largest at epoch 1 indicates epoch 5 is more stable than epoch 1.

The average validation accuracy at various population sizes ranges from 87.65% at pop size 15 to 87.83% at pop size 25. Populations of size 25 tend to give slightly better results in validation accuracy. The variance of pop size 15 is lower than the variance of pop size 20, indicating that pop size 15 has the best stability.

Based on the evaluation of validation accuracy analysis results using Analysis of Variance (ANOVA). The epoch and pop size factors show all ANOVA results have the same value, namely the F value at 0.683 and the P-value at 0.614. The F value is less than the crit F value of 3.007 and the P-value is also greater than 0.05 as the general level of significance. Based on the values of the ANOVA results, it shows that the variance of 5 epochs and pop size 5 to 25 does not have a significant effect or change on the validation accuracy of the classification model for nutrient deficiencies in rice plants.

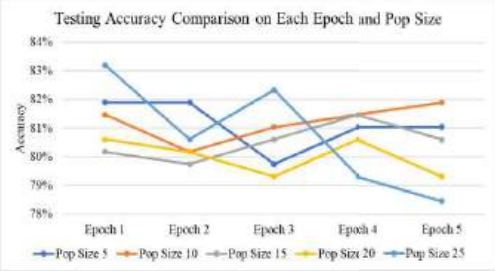


Fig. 11. Testing accuracy comparison by every epoch and pop size

As depicted in Figure 11, researchers got the highest testing accuracy of 83.19% at epoch 1 and pop size 25 and the lowest testing of 78.45% at epoch 5 and pop size 25.

The average testing accuracy has a range of 80.26% at epoch 5 to 81.47% at epoch 1. Epoch 1 has the highest average accuracy, while epoch 5 has the lowest accuracy. The variance of accuracy between epochs appears to be quite low, ranging from 6.8E-05 in epoch 2 to 0.0002 in epoch 5. This shows that epoch does not significantly affect testing accuracy.

The average testing accuracy based on pop size ranges from 80.52% at pop size 15 to 81.21% at pop size 5. Pop size 5 shows the highest testing accuracy results compared to other pop sizes. The lowest variance for pop size is 4.1E-05 in pop size 15, while the highest is 0.0004, which indicates that the testing accuracy at pop size 15 is more consistent than other pop sizes.

Based on the results of ANOVA analysis related to accuracy testing data, the epoch factor obtained a P-value of 0.525 and an F value of 0.832. The F value is still far below the F crit value of 3.007 and the P value is far above 0.05 as the general significance level. The pop size factor also shows a P-value of 0.453 and an F-value of 0.966. The F value is still far below the F crit value of 3.007 and the P value is far above 0.05. Based on the values of the ANOVA results, it can be concluded that the variance of 5 epochs and pop size 5 to 25 does not have a significant effect or change on the accuracy of validation of the classification model for nutrient deficiencies in rice plants.

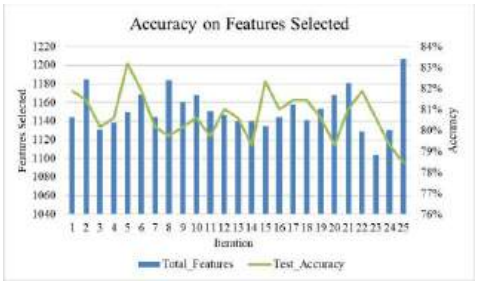


Fig. 12. Accuracy comparison by features selected

Figure 12 shows the movement of the selected features and the accuracy obtained during 25 iterations of the experiment. The highest accuracy is 83.19% when the features are selected at 1150 and the lowest accuracy is 78.45% when the features are selected at 1207. The number of selected features fluctuated in each iteration from 1104 to 1207 with an accuracy range of , but did not show any drastic changes. The total number of features does not directly affect the accuracy obtained, there is no linear or consistent pattern as the number of selected features changes.

Researchers analyzed the relationship between the number of selected features and accuracy results. An ANOVA analysis of the relationship between the two was conducted, resulting in a P-value of 6.22007E-43 which is far below 0.05 and an F-value of 67793.77 which is far above the F crit of 4.259. From these values, it can be concluded that the change in the number of features is highly related and significantly affects the accuracy value.

D. Runtime Result

Every experiment carried out in each epoch along with the pop size will definitely take time, the following describes the time required in each epoch.

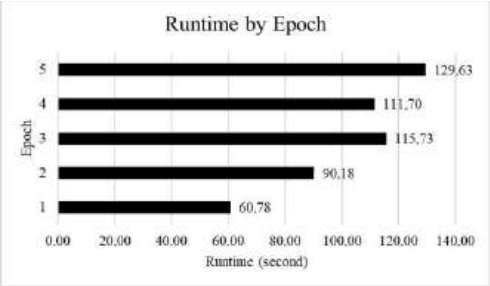


Fig. 13. Runtime comparison by epoch

As shown in Figure 13, the fastest runtime is 60.78 seconds at epoch 1 and the longest is 129.63 seconds at epoch 5. It can be concluded that the runtime tends to increase at each epoch. This can be caused by several factors ranging from the complexity of the model that increases with increasing epochs, the amount of data processed, and the PSO algorithm adjustment process.

E. Result Comparison

Researchers conducted non-PSO experiments on the same dataset using the same SVM classification model without using HOG extraction features or PSO selection features. The accuracy was 73.04% with a runtime of 540 seconds.

PSO model accuracy increases the accuracy of the model to 83.19% and the runtime is only 129 seconds. There was an increase in accuracy of 10.15% and a faster runtime of 411 seconds compared to the non-PSO model.

Therefore, it can be concluded that the PSO optimization method as a selection feature provides improved classification performance in terms of accuracy and runtime.

IV. CONCLUSION

This research aims to improve the accuracy of nutrient deficiency image classification in rice plants by combining Particle Swarm Optimization (PSO) and Histogram of Oriented Gradients (HOG) using the Support Vector Machine (SVM) classification model. The integration of PSO in the feature selection process was shown to significantly improve the accuracy and efficiency of the SVM classification model. The final model achieved an accuracy of 83.19% and a runtime of 129.63 seconds with 1150 best feature combinations out of 2303 extracted features. The study showed significant improvement compared to the model that did not use PSO. For further research, it is recommended to expand the dataset, incorporate other optimization methods, and explore other feature extraction methods and classification techniques can also be considered to improve the performance and wider application of the model.

ACKNOWLEDGEMENT

The researchers would like to express their gratitude to God Almighty for His grace in completing this research. We would also like to thank the support provided by Dian Nuswantoro University during the research process. Our gratitude also goes to the people closest to us for their support and motivation during the research process. We hope that our research can be useful and have a positive impact on all of us.

REFERENCES

[1] D. Apriyanti, M. R. Ramdani, D. K. Kresnawati, and S. B. Putri, "Spatial analysis of the relation of rice land area and rice production using remote sensing imagery," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2024. doi: 10.1088/1755-1315/1339/1/012015.

[2] D. Dildora Umarjonovna, "CHALLENGES OF FOOD SECURITY," 2022, doi: 10.17605/OSF.IO/H8NYZ.

[3] M. D. Toor, "Nutrients and Their Importance in Agriculture Crop Production: A Review," *Indian Journal of Pure & Applied Biosciences*, vol. 9, no. 1, pp. 1–6, Feb. 2021, doi: 10.18782/2582-2845.8527.

[4] S. Kolhar, J. Jagtap, and R. Shastri, "Deep Neural Networks for Classifying Nutrient Deficiencies in Rice Plants Using Leaf Images," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 305–314, Jul. 2024, doi: 10.12785/ijcds/160124.

[5] C. Wang, Y. Ye, Y. Tian, and Z. Yu, "Classification of nutrient deficiency in rice based on CNN model with Reinforcement Learning augmentation," in *Proceedings - 2021 International Symposium on Artificial Intelligence and its Application on Media, ISAIAM 2021*, Institute of Electrical and Electronics Engineers Inc., May 2021, pp. 107–111. doi: 10.1109/ISAIAM53259.2021.00029.

[6] L. A. Wulandhari *et al.*, "Plant nutrient deficiency detection using deep convolutional neural network," *ICIC Express Letters*, vol. 13, no. 10, pp. 971–977, 2019, doi: 10.24507/icicel.13.10.971.

[7] M. S. H. Talukder and A. K. Sarkar, "Nutrients deficiency diagnosis of rice crop by weighted average ensemble learning," *Smart Agricultural Technology*, vol. 4, Aug. 2023, doi: 10.1016/j.atech.2022.100155.

[8] M. Madanan, S. S. Gunasekaran, and M. A. Mahmoud, "A Comparative Analysis of Machine Learning and Deep Learning Algorithms for Image Classification," in *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 2436–2439. doi: 10.1109/IC3I59117.2023.10398030.

[9] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognit Lett*, vol. 141, pp. 61–67, Jan. 2021, doi: 10.1016/j.patrec.2020.07.042.

[10] U. P. H. A. Deepanshu Jindal and Ajay Tiwari, "Comparative study of Image Classification Algorithms," *International Journal for Modern Trends in Science and Technology*, vol. 7, no. 01, pp. 88–92, Jan. 2021, doi: 10.46501/ijmtst070120.

[11] Y. Sari, M. Maulida, R. Maulana, J. Wahyudi, and A. Shalludin, "Detection of Corn Leaves Nutrient Deficiency Using Support Vector Machine (SVM)," in *Proceedings - 2021 4th International Conference on Computer and Informatics Engineering: IT-Based Digital Industrial Innovation for the Welfare of Society, IC2IE 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 396–400. doi: 10.1109/IC2IE53219.2021.9649375.

[12] A. Qur'ania, L. Karlitasari, S. Maryana, C. Sudrajat, and Z. Zolla, "IDENTIFIKASI DEFISIENSI UNSUR HARA PADA TANAMAN CABAI MENGGUNAKAN SUPPORT VECTOR MACHINE," *Jurnal Komputer dan Informatika*, vol. 11, no. 1, pp. 62–67, Mar. 2023, doi: 10.35508/jicon.v11i1.9803.

[13] N. Sabri, N. S. Kassim, S. Ibrahim, R. Roslan, N. N. A. Mangshor, and Z. Ibrahim, "Nutrient deficiency detection in maize (Zea mays L.) leaves using image processing," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 2, pp. 304–309, Jun. 2020, doi: 10.11591/ijai.v9.i2.pp304-309.

[14] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Nitrogen Deficiency Prediction of Rice Crop Based on Convolutional Neural Network," *J Ambient Intell Humaniz Comput*, vol. 11, no. 11, pp. 5703–5711, Nov. 2020, doi: 10.1007/s12652-020-01938-8.

[15] M. V. Appalanaidu and G. Kumaravelan, "Rice plant nutrient deficiency classification using modified MOBILENET convolutional neural network," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 14, no. 1, Feb. 2023, doi: 10.1142/S1793962322430036.

[16] A. Jose, S. Nandagopalan, V. Ubalanka, and D. Viswanath, "Detection and classification of nutrient deficiencies in plants using machine learning," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jul. 2021. doi: 10.1088/1742-6596/1850/1/012050.

[17] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. Gnrsn Prudhvith, and N. Kumar, "Image Classification using SVM and CNN," in *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*, Institute of Electrical and Electronics Engineers Inc., Mar. 2020. doi: 10.1109/ICCSEA49143.2020.9132851.

[18] S. Sood, H. Singh, M. Malarvel, and R. Ahuja, "Significance and Limitations of Deep Neural Networks for Image Classification and Object Detection," in *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1453–1460. doi: 10.1109/ICOSEC51865.2021.9591759.

[19] Y. Lai, "A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1314/1/012148.

[20] Z. Xu *et al.*, "Using deep convolutional neural networks for image-based diagnosis of nutrient deficiencies in rice," *Comput Intell Neurosci*, vol. 2020, 2020, doi: 10.1155/2020/7307252.

[21] U. Ukaegbu, L. Tartibu, T. Laseinde, M. Okwu, and I. Olayode, "A deep learning algorithm for detection of potassium deficiency in a red grapevine and spraying actuation using a raspberry pi3," *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, 2020.

[22] Z. Lu, "Enhanced Accuracy Enabled by Particle Swarm Optimization in Classification Application," in *Proceedings - 2020 International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 146–149. doi: 10.1109/ICAICE51518.2020.00034.

[23] S. Long and B. S. Dhillon Editors, "Lecture Notes in Electrical Engineering 645 Man-Machine-Environment System Engineering," 2020. [Online]. Available: http://www.springer.com/series/7818

[24] S. Azimi, T. Kaur, and T. K. Gandhi, "A deep learning approach to measure stress level in plants due to Nitrogen deficiency," *Measurement (Lond)*, vol. 173, Mar. 2021, doi: 10.1016/j.measurement.2020.108650.

[25] T. Lawrence, L. Zhang, K. Rogage, and C. P. Lim, "Evolving deep architecture generation with residual connections for image classification using particle swarm optimization," *Sensors*, vol. 21, no. 23, Dec. 2021, doi: 10.3390/s21237936.

[26] L. N and K. K. Saju, "Classification of Macronutrient Deficiencies in Maize Plant Using Machine Learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, p. 4197, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4197-4203.

[27] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology (Singapore)*, vol. 13, no. 5, Oct. 2021, doi: 10.1007/s41870-017-0080-1.

[28] G. Jin, P. Jin-Ye, and L. Zhan, *2010 Symposium on Photonics and Optoelectronic (SOPO 2010) : proceedings : June 19-21, 2010, Chengdu, China*. IEEE, 2010.

ISSN 2355-0082

Cross-Platform Mobile Based Crowdsourcing Application for Sentiment Labeling Using Gamification Method

Elaine[1], Farica Perdana Putri[2], Alethea Suryadibrata[3]
Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]elaine@student.umn.ac.id, [2]farica@umn.ac.id, [3]alethea.suryadibrata@lecturer.umn.ac.id

Accepted 17 December 2024
Approved 17 January 2025

***Abstract*— Sentiment analysis is the application of natural language processing which aims to identify the sentiment of texts. To carry out sentiment analysis, data which has been labeled sentiment is needed to be included in the training model. Crowdsourcing is considered as the most optimal method to label data because it has a high level of accuracy at a relatively low cost. However, the use of crowdsourcing platforms has its own challenge, which is to increase user interest and motivation. A solution which can be applied is to design and build a crowdsourcing platform or application using the gamification method. The definition of gamification is an effort to increase one's intrinsic motivation for an activity by applying game elements to it. Therefore, a cross-platform mobile based crowdsourcing application for sentiment labeling using gamification method was carried out. The gamification design process was done based on the 6D framework and the application was developed using the Ionic-React framework. Application was examined through black box testing and the result showed that the application was functioning properly and according to the design requirements. There was also an evaluation carried out by distributing Intrinsic Motivation Inventory questionnaires to users who had used the application for 2 weeks. From a total of 40 respondents, the result showed that the level of user motivation and interest in using the application was high with a percentage of 83.10%.**

***Index Terms*— Crowdsourcing; Gamification; Ionic-React Platform; Sentiment Labeling**

I. INTRODUCTION

As human civilization and information technology develop, expectations arise for computer interfaces so that they can communicate with humans using language which is easily understood by humans [1]. Therefore, natural language processing (NLP) is needed to achieve these expectations. NLP is a branch of computer science in the field of artificial intelligence which studies the interaction between computers and human language to assist computers in understanding, interpreting, and manipulating natural language or the language used by humans so that it can be used practically [2]. NLP combines computational linguistics or rule-based modeling of human language with statistical, machine learning, and deep learning models which enable computers to process human language and understand the meaning, intent, and sentiments it embodies [3].

Sentiment analysis is the practice of applying NLP in identifying and obtaining subjective information from texts to determine human opinions, attitudes, and emotions towards an individual, event, or topic [4, 5]. The main objective of sentiment analysis is to identify the sentiments of a text and determine its polarity, whether the text has positive, negative, or neutral sentiments. Sentiment analysis plays an important role in many fields, including analyzing reviews, political views, economic market developments, historical movements, and also emotional level of social media users [6]. In fact, based on data obtained in April 2013, it was known that 90% of customers decided to buy based on online reviews [4]. Not only for customers, sentiment analysis is also needed by companies to monitor discussions and evaluations regarding the company's reputation and products issued by the company. Therefore, sentiment analysis is one of the areas which has been extensively researched regarding decision-making process.

In carrying out sentiment analysis, a data set which has been labeled with sentiment is needed to be included in the training model. The process of sentiment labeling can be done automatically or manually [7,8]. Automatic labeling can be done using dictionaries or machine learning, while manual labeling can be done with the help of expert or with the help of groups of people by crowdsourcing. Based on the research [9], it was found that the sentiment analysis where the labeling was done automatically had the lowest accuracy with the highest alpha value of 0.50 and the sentiment analysis where the labeling was done by experts had the highest accuracy with the highest alpha value of 0.90. Sentiment analysis where the labeling was done by crowdsourcing had a relatively good level of accuracy with the highest alpha value of 0.81. Beyond performance, there were cost considerations where labeling by an expert was significantly more expensive than the other two

methods and could increase as more data became available [9]. Therefore, sentiment labeling by crowdsourcing is considered to be the most optimal method.

Based on the availability of wages, crowdsourcing platforms can be classified into paid crowdsourcing and unpaid crowdsourcing [10]. Paid crowdsourcing may tend to be more desirable but requires more capital from the employer. Meanwhile, unpaid crowdsourcing depends on the workload and the identity of the employer, so there is a risk of a lack of interest from the participants because the task is considered difficult and only benefits the employer. In overcoming this problem, a solution which can be applied is to use gamification method or application of game mechanisms to increase the motivation of the participants through a fun game experience [8, 11]. In addition, the gamification method is considered to be able to improve the process and the end result of the activity.

There was a previous research, namely the development of a web system for crowdsourcing opinion labeling using the gamification method, which was carried out by Cirqueira et al in 2017. The evaluation in this research focused on aspects related to the functionality and ease of use of the system as well as the clarity of the game mechanism. The results obtained were 2 aspects considered good and 7 aspects considered very good. Based on the results of this study, Cirqueira et al made several suggestions for further research, including providing a feature for downloading labeled data sets, implementing a global leaderboard to motivate players, using more appropriate colored buttons, and developing a better badge system [8].

This research was appointed after considering the things which had been described. The application was built as a mobile cross-platform based application so that it could be accessed as a web via a browser or downloaded to a device, according to user needs. In addition, the application did not use much device features, so it did not need to be built as a native application. The application was built specifically for text sentiment labeling in Indonesian and English, was designed using the 6D gamification framework, and was built using Ionic-React framework. The objectives of this research were to design and develop a cross-platform mobile based crowdsourcing application for sentiment labeling using gamification method and to determine the level of motivation and interest of users in using the application.

II. Literature Review

Sentiment Analysis

Sentiment analysis is a type of text classification which focuses on sentiment orientation analysis which shows individual polarity towards a topic [12]. The polarity classification in sentiment analysis can be divided into a number of different polarity classes, such as two classes (positive and negative), three classes (positive, neutral, and negative), as well as a more specific classification into five classes (very positive, positive, neutral, negative, very negative) [4]. In addition, in sentiment analysis, there are three main classification levels, namely the document level, sentence level, and aspect level [5]. Document level analysis analyzes a document as a unit of basic information. Sentence level analysis analyzes based on opinion on one sentence. Aspect level analysis considers that one person may give different opinions on different aspects of an entity.

Sentiment analysis techniques can be classified into three types of approaches which are machine learning approaches, lexicon-based approaches, and hybrid approaches [5]. The machine learning approach uses linguistic features in machine learning algorithms. The lexicon-based approach relies on a known and pre-arranged set of sentiment terms. The hybrid approach combines the other two approaches in which the sentiment lexicon generally plays the most important role.

Based on the presence or absence of supervision, sentiment analysis using a machine learning approach can be classified into supervised learning methods and unsupervised learning methods. The supervised method makes use of a large number of training documents which have been labeled with sentiments, whereas the unsupervised method is used when the labeled document or data set is hard to find. Based on data sources, the lexicon-based approach can be classified into two methods, namely the dictionary-based approach and the corpus-based approach. Both lexicon-based approaches rely on finding a lexicon of opinion which is used to analyze the text. Therefore, regardless of which approach is used, the data set used is an important factor in sentiment analysis [5].

Crowdsourcing

Crowdsourcing is a type of online activity in which a party invites the participation of a group of individuals through an open call to do a task [10]. The four pillars which underlie crowdsourcing include the employer, individual groups, tasks, and platforms [13]. Two main aspects which need to be considered in implementing crowdsourcing are that crowdsourcing must be able to motivate groups of individuals in carrying out tasks and that crowdsourcing implementation must be able to achieve the required solution [14].

To mediate between employers and community groups, a platform is needed in the form of an application which provides functions for carrying out tasks and has a group management system . Based on the types of functions provided, crowdsourcing platforms can be classified into two types, namely

applications with specific functions, such as InnoCentive and ClickWorker, and applications with general functions, such as Amazon Mechanical Turk (MTurk), Microworker, and CrowdFlower [10]. Based on task specifications, crowdsourcing platforms are classified into applications which perform microtasks or tasks which are simple, repetitive, independent, and short, such as labeling tasks, and applications which perform macrotasks or tasks which depend on context, require a lot of effort, and take a long time, such as code writing assignments. In addition, crowdsourcing platforms can be differentiated based on the presence or absence of wages. Procurement of wage rewards is usually used to encourage people to participate and contribute to crowdsourcing. However, not only motivated by money, research by Kaufmann et al. in 2011 found that crowdsourcing participants at MTurk were also motivated by the sense of satisfaction, togetherness, and pride which they got from doing tasks [15].

In the implementation of crowdsourcing, the nature and complexity of the task has a significant impact on the level of community participation and motivation so it is important for a crowdsourcing project to have an effective task design. Several things need to be considered when the task is to be designed, namely the description of the task, the scope of the task, the skills needed to complete the task, the division of tasks if necessary, the need for solutions, the targets of participants, and the duration of work. In addition, it is also important to identify and analyze the relationship between elements of task design and the ultimate goal of crowdsourcing [14].

Gamification

Gamification is an effort to increase intrinsic motivation for various activities by implementing game elements into the design. Examples of applicable game elements are point system, leaderboards, awards, badges, bonuses, targets, and narrative [16, 17]. Gamification plays an important role in triggering game-like psychological experiences by leveraging elements found in games. The main goal of gamification is to increase the user's positive motivation for a given activity so that the quantity and quality of the results of these activities can also increase [18].

In general, gamification approaches follow a process which can be divided into seven phases. First, the project preparation phase, namely all activities which need to be carried out before the project starts. Second, the analysis phase, namely the process of identifying problems related to the users, the processes, and the project. Third, the idea generation phase, namely the idea search stage related to the gamification design. Fourth, the design phase, namely the activity of designing a gamification approach and prototyping. Fifth, the implementation phase, namely the application of the gamification approach which was previously designed. Sixth, the evaluation phase, which is the activity of testing and assessing the results of implementing the gamification approach. Finally, the observation phase, which is to monitor the result of the implementation of the gamification approach on users.

The application of gamification needs to be well designed to achieve the desired positive impact, so it is important to choose the right framework [19]. A literature study conducted in 2017 stated that there were 40 recorded gamification design frameworks. One of the most popular and most widely referenced gamification design frameworks in research is the 6D framework initiated by Werbach and Hunter [20].

The 6D framework consists of the following six stages [21, 22]. First, define business objectives or the stage where the final goal and the specific positive impact expected from the application are determined. Second, define the target behavior or the stage where the expected player behavior and how the system should support and provide feedback on this behavior are determined. Third, describe the players or the stage where the characteristics of the players are determined. Fourth, develop the activity loop or the stage where it is planned how to motivate players by using engagement and progress cycles, how to attract new players, and how to encourage further action from players. Fifth, don't forget the fun or the stage where the application is reviewed to identify aspects which motivate players intrinsically. Finally, deploy appropriate tools or the stage where mechanics, metrics, and tools, including relevant game elements, are defined.

Intrinsic Motivation Inventory

Intrinsic Motivation Inventory (IMI) is a multidimensional measurement tool used to measure participants' subjective experiences related to specific activities in research. Based on the research by McAuley et al. in 1987 and by Goudas et al. in 2000, it was concluded that the IMI model was a valid measurement tool, in accordance with common factors, and was three-dimensional, supported by strong evidence [23, 24]. Monteiro et al conducted research on the level of reliability of each aspect in the IMI model in 2015 and obtained Cronbach's alpha values ranging from 0.82 to 0.91, stated that the IMI model showed adequate validity and reliability. IMI has been used in many studies focusing on intrinsic motivation in various fields where the question models used apply different aspects and questions of IMI depending on the characteristics of task and participants.

The IMI model measures seven aspects, namely interest/enjoyment, perceived competence, perceived choice, pressure/tension, effort/importance, value/usefulness, and relatedness [23]. Interest is a

form of personal assessment of intrinsic motivation and it is considered as the main aspect of IMI and often has more questions. Perceived competence and perceived choice are used as positive predictors in measuring intrinsic motivation, while pressure is a negative predictor of intrinsic motivation. Effort is a separate variable which is considered relevant and not necessarily but can be included in the IMI model to measure the level of motivation related to a specific issue or context [23, 24]. The value aspect aims to study the participants' self-internalization where participants are more likely to live up to the tasks being done if the participants judge that the task has benefits or is of value to the participants. There is also an aspect of relatedness which is used to measure interpersonal interactions between participants.

Each aspect measured in IMI has a different number of question choices with a total of 45 questions [23]. For each of these questions, the user's response was measured using a 7-point Likert scale where 1 is “not at all true” and 7 is “very true”. Researchers may choose the aspects to be measured according to the relevance of the aspects to the research being conducted [23]. In designing a list of questions using the IMI model, measured aspects are first determined based on variables in the problem formulation then which questions to use are determined based on these aspects in random order [23]. In measuring the result of an IMI question model, it is done by calculating the average value of each aspect, then the average results can be used according to the purposes of analysis in research.

Likert Scale

The Likert scale was first mentioned by Rensis Likert in 1932 on his research article "A Technique for the Measurement of Attitudes" in the journal "Archive of Psychology" [25]. The Likert scale was designed to measure attitudes in a scientifically acceptable and validated way [26]. Likert in his research discussed the possibility of classifying an infinite number of attitudes of a person into groups of responses [26]. In applying the Likert scale, participants were asked to indicate their level of conformity with the statements given on the metric scale [26]. Likert stated that two polar choices on the scale needed to be assigned an exact value, while the choices between them were left without an exact value [27].

Likert himself did not consider the number of choices to be important and implied that each researcher was free to determine it according to their individual needs [27]. However, several studies stated that the 7-point Likert scale showed better performance compared to the 5-point scale. The 7-point scale has more choices of responses which can trigger the respondent's reasoning abilities and increase the possibility of achieving objectivity [26]. As a result, the 7-point scale produces data with lower measurement errors and a higher level of accuracy [27].

To apply the result of Likert scale measurements to a case, class interval calculations can be performed [28]. The interval class formula can be seen in Equation 1 [29].

$$i = \frac{x_n - x_1}{k} \qquad (1)$$

Where i = class interval, x_n = highest value, x_1 = lowest value, and k = number of classes.

III. Research Method

Requirements Analysis

In the requirements analysis stage, a simple literature study was carried out regarding sentiment analysis, especially about sentiment labeling, regarding crowdsourcing, regarding gamification methods, and regarding Likert scale to find out what the application needs in order to achieve the stated research objectives.

Application Design

At the design stage, it was determined thoroughly, in detail, and specifically about the application to be built. It included functions, features, and how game elements are implemented. The design process was done using the 6D gamification framework. The expected design results were in the form of design report, flowcharts, relational database scheme, mockup, and questionnaire design.

Application Development

At the application development stage, based on the design that had been made, the application was built using the Ionic-React framework through a source-code editor application.

Application Testing

This stage was carried out using the black box testing method. The purpose of the black box testing method was to test the external functions of the application in accordance with the design requirements which had been determined. Black box testing was done by trying each feature in various scenarios as input to observe whether the application was functioning and giving the expected results or output.

Evaluation

The evaluation stage was carried out with a quantitative approach. Sample participants were taken randomly from people who were willing. The evaluation began by giving application access to participants so that participants could try using the application directly. After the usage time period of 2 weeks had passed, each participant was given a motivational and interest level questionnaire related to

crowdsourcing activities for sentiment labeling in the application. The questionnaire was prepared based on the Intrinsic Motivation Inventory (IMI) question model and was measured using a Likert scale. The IMI question model was chosen after considering the validity and reliability of the IMI as an adequate measuring tool. Responses to the questionnaire on levels of motivation and interest were then calculated and analyzed to determine the effectiveness of the application which has been developed.

IV. Results and Discussion

Gamification Design

The application design based on the 6D gamification framework by Werbach and Hunter can be seen as follows:

- Define business objectives

The purpose of gamification is to increase user motivation and interest in sentiment labeling activities. The expected benefits are the increase in quantity of training data or data which has been labeled with sentiment.

- Delineate target behavior

The expected target behavior is that users are motivated and interested in engaging in sentiment labeling activities. The achievement of the target behavior is measured by giving a questionnaire compiled based on the Intrinsic Motivation Inventory (IMI) model to users to measure the level of intrinsic motivation of users.

- Describe the players

Target users are individuals from various demographic groups who understand Indonesian or English well and understand the concept of sentiment so that they can provide appropriate sentiment labels.

- Devise activity loop

The engagement cycle applied, namely: (1) Users are motivated to increase player levels, enter the leaderboards, or get other rewards; (2) Users participate in sentiment labeling activities; (3) The system provides feedback in the form of points to users; (4) Points impact the final game score which is then calculated into the player level and leaderboard. The stages of progress applied are: (1) Users have a short-term goal to get the best final score and enter the leaderboard; (2) Users have long-term goals of increasing levels, achieving various achievements, obtaining various badges, obtaining mysterious gifts, and completing profiles.

- Don't forget the fun

Aspects which motivate intrinsically, namely: (1) The hard fun aspect is implemented through sentiment labeling activities in the form of games where users need to consider carefully which sentiment label is the most appropriate and most likely to be chosen by other users to get points; (2) The easy fun aspect is implemented through achievement system, badge system, and mysterious reward system which can be obtained by completing certain relatively simple tasks, which in addition, badges and other rewards, such as profile icons, profile frames, and namecards, can be used to decorate and complete user profiles; (3) The aspect of altered states is implemented through sentiment labeling activities in the form of games where users can encounter various sets of statements which are always different; (4) The aspect of the people factor is implemented through leaderboard system which compares and displays the final scores and user profiles.

- Deploy appropriate tools

Based on the previous stages, the following were determined. The application has a game language filter feature which consists of three choices, namely Indonesian, English, and Indonesian-English. Game elements applied to application gamification, namely point system, player level, leaderboards, achievements, badges, mysterious reward, and player profile customization.

Questionnaire Design

To measure the level of user motivation and interest in the application, a questionnaire was designed based on IMI model which can be seen as follows. Questions with the code "(R)" are reverse coded questions in which the score data from the questionnaire responses for these questions needs to be calculated first with the formula, which is 8 minus the original score, then the calculation results can be processed as usual.

- Interest/enjoyment

I enjoyed doing this activity very much;
I would describe this activity as very interesting;
I thought this activity was quite enjoyable.

- Perceived competence

After working at this activity for awhile, I felt pretty competent;

I am satisfied with my performance at this task;
I was pretty skilled at this activity.

- Perceived choice

I believe I had some choice about doing this activity;
I did this activity because I wanted to;
I did this activity because I had to (R).

- Value/usefulness

I think that doing this activity is useful;
I would be willing to do this again because it has some value to me;
I believe doing this activity could be beneficial to me.

Application Results

The application can be seen in Figure 1. Users who have logged in to their accounts will be directed to the home page which includes options to play and options to access other menus. When the user chooses to play, the user is shown a list of sets along with general information about the sets. After selecting a set, the user enters the game. When the user has finished specifying the sentiment label for each given data, the user is redirected to the results page which displays the final score and game leaderboard.

Fig 1. Screenshots of Application: (a) Home; (b) Game; (c) Leaderboard

Application Testing

The results of application functionality tests carried out using the black box testing method can be seen in Table 1.

TABLE I. BLACK BOX TESTING RESULTS

Function Name	Result
Welcome Page	Works Accordingly
Register	Works Accordingly
Login	Works Accordingly
Home page	Works Accordingly
Dataset List Page	Works Accordingly
Sentiment Labeling Game	Works Accordingly
Set Leaderboard	Works Accordingly
Weekly Leaderboard	Works Accordingly
Notification	Works Accordingly
Game History	Works Accordingly
Achievement	Works Accordingly
Badge	Works Accordingly
Profile	Works Accordingly
Edit Username	Works Accordingly
Edit Profile Badge	Works Accordingly
Edit Profile	Works Accordingly
Edit User Information	Works Accordingly
Settings	Works Accordingly
Upload Dataset	Works Accordingly
Uploaded Dataset List Page	Works Accordingly
Dataset Labeling Result Page	Works Accordingly
Illustration	Works Accordingly

Evaluation

To determine the level of user motivation and interest obtained from measurements using Likert scale, the average index of each aspect of the Intrinsic Motivation Inventory (IMI) and the overall average index are categorized into 7 class intervals. Therefore, it is necessary to find out in advance the value of the interval length through Equation 1 (interval class formula) which results in an interval length value of 0.86. Based on the obtained value, interval classes to determine the level of user motivation can be seen in Table 2.

TABLE II. USER MOTIVATION LEVEL INTERVAL CLASSES

Class Interval	Description
1-1.85	Very Low
1.86-2.71	Low
2.72-3.57	Somewhat low
3.58-4.45	Neutral
4.46-5.29	Somewhat High
5.30-6.15	High
6.16-7	Very High

By comparing to the predetermined interval classes, the results of measuring the motivation level of application users can be seen in Table 3. The measurement results were obtained through questionnaire responses from a sample of 40 users who had tried the application during a 2-week usage period.

TABLE III. USER MOTIVATION LEVEL INTERVAL CLASSES

Aspect	Average	Standard Deviation	Percentage	Desctiption
Interest/enjoyment	6.06	0.82	86.55 %	High
Perceived competence	5.57	1.07	79.52 %	High
Perceived choice	6.17	0.78	88.10 %	Very High
Value/usefulness	5.48	1.12	78.21 %	High
All aspects	5.82	0.75	83.10 %	High

From Table 3, it can be seen that the level of motivation of application users based on all aspects of IMI had an average value of 5.82 with a standard deviation of 0.82 and a percentage of 83.10%. The value of 83.10% itself was obtained by calculating the average of the percentage of each aspect. Several things which need to be considered, namely the perceived choice aspect was the aspect with the highest average value of 6.17 and the value/usefulness aspect was the aspect with the lowest average value of 5.48. There was also the interest/enjoyment aspect as the main determining aspect in the IMI model which had an average value of 6.06 and was the aspect with the second highest average value. In addition, it can be seen that 3 of the 4 total aspects of the level of motivation considered had an average value which was included in the high category and 1 aspect of another level of motivation, namely the aspect of perceived choice, was included in the very high category. Based on these data, it can be stated that the level of user motivation and interest in using the application was relatively high.

V. CONCLUSION

Based on the results and discussion, the following conclusions can be drawn. The design and development of a cross-platform mobile based crowdsourcing application for sentiment labeling using gamification method has been well completed. The application was designed by following the 6D gamification framework and built using Ionic-React framework in a source-code editor application. The gamification elements applied to the application included point system, player level, leaderboards, achievements, badges, mysterious reward, and player profile customization.

Application testing was carried out through black box testing and the result showed that all the features tested have worked according to the design requirements. Results also showed that the level of user motivation and interest in using the application was high. Determining the level of user motivation and interest was done by analyzing response data from a questionnaire created using the Intrinsic Motivation Inventory (IMI) model with a 7-point Likert measurement scale. Questionnaires were distributed to users after 2 weeks of application usage period. From a total of 40 respondents from various backgrounds, the result was that the average level of user motivation based on all aspects reached a percentage of 83.10% which could be classified as high.

ACKNOWLEDGMENT

Authors thank Universitas Multimedia Nusantara for sponsoring the research.

REFERENCES

[1] Y. Goldberg, Neural Network Methods in Natural Language Processing, G. Hirst, Ed., Morgan & Claypool Publishers, 2017.

[2] D. Oliva and E. H. Houssein, Metaheuristics in Machine Learning: Theory and Applications, S. Hinojosa, Ed., Springer International Publishing, 2021.

[3] A. Saxena, A. K. Bhatt and R. Kumar, Artificial Intelligence for Societal Development and Global Well-Being, IGI Global, 2022.

[4] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," Journal of King Saud University - Engineering Sciences, vol. 30, no. 4, pp. 330-338, 7 2018.

[5] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093-1113, 7 2014.

[6] B.L. Supriyatna and F. P. Putri "Optimized Support Vector Machine for Sentiment Analysis of Game Reviews," International Journal of Informatics and Communication Technology (IJ-ICT), Vol. 13, No. 3, pp 344-353, 2024.

[7] G. Haralabopoulos and E. Simperl, "Crowdsourcing for Beyond Polarity Sentiment Analysis A Pure Emotion Lexicon," 7 2017.

[8] D. Cirqueira, L. Vinícius, M. Pinheiro, A. J. Junior, F. Lobato and Á. Santana, "Opinion Label: A Gamified Crowdsourcing System for Sentiment," Companion Proceedings of the 23rd Brazilian Symposium on Multimedia and Web, pp. 209-213, 2017.

[9] W. van Atteveldt, M. A. C. G. van der Velden and M. Boukes, "The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms," Communication Methods and Measures, vol. 15, no. 2, pp. 121-140, 7 2021.

[10] R. M. Borromeo and M. Toyama, "An investigation of unpaid crowdsourcing," Human-centric Computing and Information Sciences, vol. 6, no. 1, p. 11, 7 2016.

[11] C. Busch, ECGBL2014-8th European Conference on Games Based Learning: ECGBL2014, Academic Conferences and Publishing International, 2014.

[12] H. Agatha, F. P. Putri, "Sentiment Analysis on Song Lyrics for Song Popularity Prediction Using BERT Algorithm," Ultimatics : Jurnal Teknik Informatika, vol. 15, no. 2, pp 99-105, 2023

[13] M. Hosseini, K. Phalp, J. Taylor and R. Ali, "The four pillars of crowdsourcing: A reference model," 2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS), pp. 1-12, 7 2014.

[14] R. Karachiwalla and F. Pinkow, "Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative," Creativity and Innovation Management, vol. 30, no. 3, pp. 563-584, 7 2021.

[15] H. Xia, Y. Huang and J. Stromer-Galley, "A Collective Action Inspired Motivation Framework for Crowdsourcing: A Collective Action Inspired Motivation Framework for Crowdsourcing," vol. 13, no. 1, pp. 41-48, 2016.

[16] S. S. Setiawan and A. Suryadibrata, "Fitrust: Promoting Healthy Lifestyle Through Gamified Mobile Health Application," 5th International Conference on New Media Studies (CONMEDIA), 2019.

[17] E. Sanchez, H. van Oostendorp, J. D. Fijnheer and E. Lavoué, "Gamification," Encyclopedia of Education and Information Technologies, pp. 816-827, 2020.

[18] B. Morschheuser, K. Werder, J. Hamari and J. Abe, "How to gamify? A method for designing gamification," Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 1298-1307, 2017.

[19] W. Oliveira, A. M. Toda, P. T. Palomino, L. Rodrigues and S. Isotani, "Which one is the best? A quasi-experimental study comparing frameworks for unplugged gamification," RENOTE, vol. 18, no. 1, 7 2020.

[20] A. Mora, D. Riera, C. González and J. Arnedo-Moreno, "Gamification: a systematic review of design frameworks," Journal of Computing in Higher Education, vol. 29, no. 3, pp. 516-548, 7 2017.

[21] F. Silva, A. Toda and S. Isotani, "Towards a link between Instructional Approaches and Gamification - A Case Study in a Programming Course," Anais do XXIV Workshop de Informática na Escola (WIE 2018), pp. 157-165, 7 2018.

[22] A. Odincova, "Using Gamification to Motivate Self-Regulated Learning in Adults. Research through Design: Developing a Quiz-Based Learning Tool," 7 2017.

[23] "Intrinsic Motivation Inventory (IMI)," Center of Self-Determination Theory.

[24] V. Monteiro, L. Mata and F. Peixoto, "Intrinsic Motivation Inventory: Psychometric Properties in the Context of First Language and Mathematics Learning," Psicologia: Reflexão e Crítica, vol. 28, no. 3, pp. 434-443, 7 2015.

[25] Z. Awang, A. Afthanorhan and M. Mamat, "The Likert scale analysis using parametric based Structural Equation Modeling (SEM)," Computational Methods in Social Sciences, vol. 4, pp. 13-21, 7 2016.

[26] A. Joshi, S. Kale, S. Chandel and D. K. Pal, "Likert scale: Explored and explained," British journal of applied science & technology, vol. 7, no. 4, p. 396, 2015.

[27] J. Munshi, "A Method for Constructing Likert Scales," SSRN Electronic Journal, 2014.

[28] A. M. Soamole, A. K. Hadi and Watono, "Analisis Penerapan Sistem Manajemen Mutu ISO 9001 pada PT. Semen Tonasa," Journal Flyover, vol. 2, no. 1, 2022.

[29] A. D. Miqdad, "Analisis pengembangan model pembelajaran berbasis Project Based Learning dengan berbantu media audio visual VideoScribe dalam Pembelajaran Komputer dan Jaringan dasar kelas X SMK Unitomo Surabaya," IT-Edu: Jurnal Information Technology and Education, vol. 2, no. 02, 2017.

ISSN 2355-0082

Cost Estimation for Software Development Using Function Point Analysis Method

Ester Lumba[1], Destriana Widyaningrum[2], Alexander Waworuntu[3]
[1,2] Department of Informatics, Bunda Mulia University, Jakarta, Indonesia
[3] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]l0178@lecturer.ubm.ac.id, [2]l0178@lecturer.ubm.ac.id
[3]alex.wawo@umn.ac.id

Accepted 7 January 2025
Approved 17 January 2025

***Abstract*— Software development requires substantial financial resources. This study aims to examine the cost estimation for software development. The complexity of the software, its intangibility as a non-physical product, the technology utilized, and human resources can all influence the determination of software development costs. The method used for cost estimation is Function Point Analysis with a case study approach. The researchers conducted a case study on the software development for an employee savings and loan cooperative at XYZ Company. The formulation of the problem in this study is how to apply the Function Point Analysis method in estimating software development costs at the XYZ employee savings and loan cooperative. The result of this study is a cost recommendation that can serve as a reference for selecting software development vendors by the cooperative's management.**

***Index Terms*— Cost Estimation; Function Point Analysis; Savings and Loan Cooperative.**

I. INTRODUCTION

Information technology is rapidly advancing in the digital era [1], prompting various small, medium, and large organizations to develop software to support their business operations. Desktop, web, and mobile applications are necessary to enhance efficiency, productivity, and service quality.

Software development incurs costs [2]. Organizations, whether large or small, often struggle to determine the costs associated with software development. Several factors contribute to the difficulty of determining these costs, including the intangibility of software as a non-physical product, project complexity, the technology employed, and human resources. Estimation techniques can be classified into three categories: expert judgment based on historical data and similar software projects, algorithmic models, and machine learning [2][3][4].

Software comprises computer programs associated with software documentation [5]. This documentation is a collection of information, guidelines, and descriptions related to the development, operation, and maintenance of the software. Software does not become obsolete because software defects can be repaired. The purpose of documentation is to assist users in understanding how to use the software. For developers, it provides essential information for updating or repairing software and serves as a reference for teams involved in project development.

Software is built through an engineering process [5]. Generally, software development follows principles known as the software or system development life cycle (SDLC) [6][7]. The SDLC includes phases aimed at producing quality software that meets customer desires or the objectives for which the system was created. The software or system development life cycle can be seen in Figure 1.

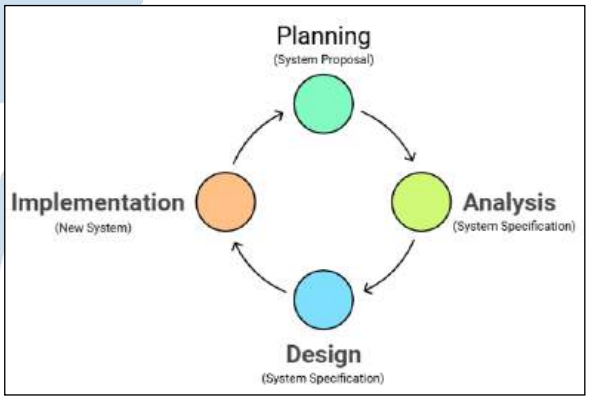


Fig. 1. Software Development Life Cycle

In software engineering, the concept of SDLC underlies various software development methodologies. These methodologies provide a framework for planning and controlling the creation of an information system, namely the software development process. Software development is divided into two approaches: structured and Object-Oriented Design (OOD). The structured approach represents the system based on data and the processes applied to this data, using modeling tools like Data Flow Diagrams (DFD). The object-oriented design approach views the system as a collection of objects consisting of data and processes, using modeling tools like Unified Modeling Language (UML) [5][6][7].

This research uses a case study approach to develop software for the XYZ Company's employee savings and loan cooperative. This cooperative is a type of microfinance organization that plays a role in enhancing the welfare of each member[8]. XYZ Company has both permanent and contract employees. Like other companies, XYZ Company also formed a savings and loan cooperative. This cooperative was established as a platform for employees to assist each other financially during times of need. All permanent employees are encouraged to become cooperative members [8]. Besides serving as a venue for saving and borrowing money, the cooperative also fosters solidarity and mutual aid among employees at the company [9].

To manage employee data, the cooperative's management still uses manual bookkeeping and records data using tools like Microsoft Excel. Amid the demands of their primary job roles within the company, employees who also serve as cooperative managers often face challenges in reporting to members. Therefore, during a member meeting, the cooperative's management at the company proposed developing software to aid in managing member data and recording transactions and borrowings conducted by employees.

Building software requires funding. The cooperative's management has contacted various software vendors. The development costs offered by these vendors vary significantly, ranging from very low to very high prices.

Addressing the challenges outlined above, the research team embarked on a study aimed at refining the cost estimation process for software development through the Function Point Analysis method. This study focuses on applying this method specifically to estimate the costs associated with developing software for the XYZ employee savings and loan cooperative. The core research question investigates the practical application of Function Point Analysis in this setting: "How is the Function Point Analysis method applied in estimating the costs of developing software for the XYZ employee savings and loan cooperative?"

The objectives are twofold: firstly, to implement the Function Point method to accurately estimate the development costs for software at the XYZ Employee Savings and Loan Cooperative, and secondly, to equip the cooperative's management with reliable recommendations for choosing a software development vendor who offers a reasonable cost.

The anticipated benefits of this study are manifold. For software developers, it promises to provide a clear and quantifiable guideline for estimating costs, based on the functional size of the software. For the cooperative's management, it aims to present an overview of what constitutes reasonable software development costs. Lastly, for the academic and research community, this study is intended to serve as a valuable reference in the field of software cost estimation, with a particular focus on the application of the Function Point Analysis method.

II. RESEARCH METHODOLOGY

This study employs a qualitative descriptive approach using a case study methodology. The software development project for the Employee Savings and Loan Cooperative at XYZ Company will be analyzed using Function Point Analysis to estimate costs [10]. The research process is divided into three stages, as illustrated in Figure 2.

Fig. 2. Research Step

The initial stage of the research is divided into three parts: preparation, preliminary survey, and interviews with the cooperative's management. In the preparation phase, the research team conducts studies related to topics that are relevant to the field of expertise. Subsequently, in the preliminary survey phase, the team visits XYZ Company to gather information about the software needs. The preliminary survey revealed that the management of the XYZ Company's employee savings and loan cooperative needs software to manage financial and member data. Following this, the research team conducts further interviews to capture the user requirements.

The implementation stage is split into two parts: analyzing the software functions based on the interviews conducted in the first stage and applying the Function Point Analysis (FPA) method. In the user requirements determination phase, researchers document the interview results with the cooperative's management. This documentation includes a list of functional and non-functional requirements needed by the users. The next part involves analyzing the software functions, which results in seven modules. Each module defines a list of functional requirements. The final stage of this research produces a recommended cost estimate for the software project. This recommendation will serve as a reference for the cooperative's management in selecting a vendor to develop the software for the employee savings and loan cooperative at XYZ Company.

III. FUNCTION POINT ANALYSIS

Function Point Analysis (FPA) is a software measurement method introduced by Allan Albrecht in 1979[10] and is widely used globally. It has been updated by the International Function Point Users Group (IFPUG) [11]; a nonprofit organization managed by members worldwide. This organization helps

improve software development processes according to software measurement standards. IFPUG has approximately 1,200 members from 30 countries who are experts in Function Point Analysis [11][12].

A. User Function Identification and Complexity

From a user's perspective, software functionality is measured using five elements: External Inputs (EI), External Outputs (EO), External Inquiries (EQ), Internal Logical Files (ILF), and Program Interfaces (PI). Each function has its own complexity level [13]. Therefore, each function can be classified based on its complexity as low for simple functions, medium for moderately complex functions, and high for complex or highly complex functions [13][14][15]. The functional complexity weight based on function type are shown in table 1 below:

TABLE I. FUNCTIONAL POINT COMPLEXITY

User Function Types	Complexity Weight		
	Low	*Medium*	*High*
External Inputs (EI)	3	4	6
External Outputs (EO)	4	5	7
External Inquiries (EQ)	3	4	6
Internal Logical Files (ILF)	7	10	15
Program Interfaces (PI)	5	7	10

External Input pertains to user inputs into the system. Based on the analysis, the function and complexity levels for the external input elements of the employee savings and loan cooperative are shown in Table 2.

TABLE II. EXTERNAL INPUTS

No.	Description	Complexity
1	Member Login	Low
2	Manager and Admin Login	Low
3	Member Registration	Low
4	Manager Registration	Low
5	Update Manager Data	Low
6	Update Member Data	Low
7	Loan Application	Medium
8	Deposits (Principal, Mandatory, Voluntary)	Low
9	Installment Deposits	Low

External Output relates to outputs produced by the system. Based on the analysis, the function and complexity levels for the external output elements of the cooperative are shown in Table 3

TABLE III. EXTERNAL OUTPUTS

No.	Description	Complexity
1	Member Data Search	Low
2	Manager Data Search	Low
3	Loan Application Status Search	Low
4	Membership Status Search	Low
5	Loan Payment Status Search	Low

External Inquiries involve user searches within the system [14][15]. The system will display search results if data is found and a message if no data is available. The functions and their complexities are shown in Table 4

TABLE IV. EXTERNAL INQUIRIES

No.	Description	Complexity
1	Member Card	Low
2	Voluntary Deposit Receipt	Low
3	Mandatory Deposit Receipt	Low
4	Loan Repayment Receipt	Low
5	Loan History per Member	Low
6	Savings History per Member	Low
7	Loan List for All Members	Low
8	Savings List for All Members	Low
9	Monthly Report	Medium
10	Annual Report	Medium
11	Profit and Loss Report	High
12	Cash Flow Chart	High
13	Email Notifications	Low
14	Smartphone Notifications	Log

Internal Logical Files consist of files that form the system, such as tables, images, text files, or other file formats [14][15]. The complexities involved in the cooperative's employee savings and loan system are shown in Table 5

TABLE V. INTERNAL LOGICAL FILES

No.	Description	Complexity
1	Member Table	Low
2	User Table	Low
3	Savings Table	Medium
4	Loan Table	Medium
5	Manager Table	Low
6	Installment Table	Low
7	Interest Table	Low
8	Image Files	Low

Program Interface involves interfaces such as the use of Application Programming Interfaces or other systems related to the developed software. Since the cooperative already has a system for processing employee data, employee and division data are taken from the existing system at the company. The functions and complexity levels are shown in Table 6.

TABLE VI. PROGRAM INTERFACE

No.	Description	Complexity
1	Company Employee Data	Low
2	Company Division Data	Low

B. *Unadjusted Function Point*

After identifying user functions, the next step is to calculate the Unadjusted Function Point (UFP). The formula for calculating UFP is (1).

$$UFP = \sum(\text{ Number of Functions } \times \text{ Complexity Weight }) \quad (1)$$

The UFP is derived from the calculation of the number of functions multiplied by their complexity weights. Table 7 shows the calculations for each function and their complexities.

TABLE VII. UFP VALUE CALCULATION

User Function Types	Total Number	Complexity Weight			Total
		Low	*Medium*	*High*	
External Inputs (EI)	**9**	8*3	1*4	0*6	28
External Outputs (EO)	14	10*4	2*5	2*7	64
External Inquiries (EQ)	5	5*3	0*4	0*6	15
Internal Logical Files (ILF)	8	6*7	2*10	0*15	62
Program Interfaces (PI)	2	2*5	0*7	0*10	10
Unadjusted Function Points					179

From Table 7 above, the total Unadjusted Function Points is 179, calculated from the total of EI + EO + EQ + ILF + PI, thus total UFP = 28 + 64 + 15 + 62 + 10 = 179.

C. *Calculate Value Adjustment Factor*

The software developed requires an operational environment. Therefore, in FPA, it's necessary to calculate factors that affect the operational complexity of the software. There are 14 General System Characteristics (GSC) that can influence the complexity of the software. Each characteristic or factor is rated between 0 to 5, with 0 meaning no effect and 5 meaning a significant effect. Table 7 shows the factors and the researcher's ratings for the cooperative's software.

Based on the Table 8, calculate the Total Degree of Influence (TDI) using the formula (2).

$$TDI = \sum(\text{ GSC Value }) \quad (2)$$

Thus, TDI = (3+0+0+0+0+0+0+1+1+0+0+1+0+0) = 6 After calculating the TDI, the next step is to compute the Value Adjustment Factor using the formula = 0.65 + (0.01 x TDI). Thus, VAF = 0.65 + (0.01 x 6) = 0.71 After calculating the VAF, calculate the Adjusted Function Points with the formula TUFP x VAF Thus, AFP = 179 x 0.71 = 127.09 rounded to 127.

TABLE VIII. VALUE ADJUSTMENT FACTORS

Factor	Value
Data Communication	3
Distributed Function	0
Performance Objectives	0
Heavily Used Configuration	0
Transaction Rate	0
Multiple Sites	0
Reusability	0
On-line Data Entry	1
On-line Update	1
Complex Processing	0
Installation ease	0
Operational ease	1
Extensibility	0
End-user Efficiency	0

D. *Cost Estimation*

To calculate the cost of software development using the FPA method, use the formula: Total cost = Adjusted Function Point * Hour/AFP * rates. The hourly rate is set at Rp. 100,000.00. The determination of the rate is based on consideration of several factors such as level of experience, specific expertise, type of project, location, and type of company. Thus, the estimated cost for developing the software for the XYZ company's employee savings and loan cooperative is 127 * 15 * 100,000 = Rp. 190,500,000.00

IV. RESULTS AND DISCUSSION

This study aimed to estimate the cost of developing software using the Function Point Analysis (FPA) method. The application of FPA was carried out through a case study approach at the XYZ company's employee savings and loan cooperative. The result of this study is an estimated software development cost for the cooperative amounting to Rp. 190,500,000.00. This cost estimation will serve as a reference for the cooperative's management to select a software development vendor for this company.

Based on the functional requirements obtained from the users, the software for the savings and loan cooperative produced seven modules. These modules are illustrated in Figure 3.

Using the FPA method, an Unadjusted Function Point of 179 and a Total Degree of Influence of 6 were obtained. Given that the project size of the savings and loan cooperative falls into the category of small-scale software, the calculation of the Value Adjustment Factor (VAF) used a constant of 0.65 and 0.01 to determine the influence of each TDI point on the final VAF result, resulting in a VAF of 0.71.

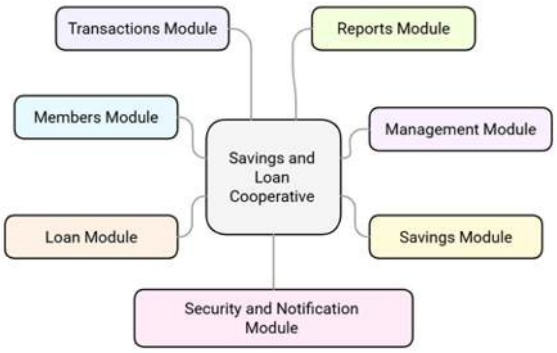


Fig. 3. Savings and Loan Cooperative Module

The variables needed to calculate the cost of the software include the adjusted function point, the average time taken for each function, and the hourly rate of programmers to perform the function. According to IFPUG standards, each function is typically completed in 15 hours. The hourly rate depends on the programmer's expertise level. For this calculation, the research team used an average hourly rate of Rp. 100,000. Therefore, the estimated cost for developing the software for the savings and loan cooperative is 127 * 15 * 100,000 = Rp. 190,500,000.00.

This study highlights the practicality of using FPA for software cost estimation, especially in settings where project scopes are relatively small but the functionalities involved are critical to the organization's operations. It provides a quantifiable and methodical approach to estimating costs that can significantly aid cooperative management in budget planning and vendor selection. Furthermore, these findings contribute to the broader understanding of applying FPA in different organizational settings, offering insights that may be valuable for other researchers and software development professionals.

V. CONCLUSION

The implementation of the Function Point Analysis (FPA) method requires a team that possesses programming skills to accurately determine the complexity level of each function. FPA is versatile and can be employed to estimate costs for small, medium, and large-scale software projects. The rates for software development are influenced by two factors: the skill or experience of the programmer and the policies of the software company.

In this study, the application of the FPA method for estimating the cost of developing software for the savings and loan cooperative resulted in the creation of seven modules with a total of 127 functions. The estimated cost for the development is Rp. 190,500,000.00. This study demonstrates that FPA is a practical and effective tool for financial planning in software development, providing a structured approach that can guide cooperatives and similar organizations in their vendor selection and budgetary processes. The adaptability of FPA to different project sizes and complexities also highlights its utility across various software development scenarios.

REFERENCES

[1] S. P. D. Ardi, I. Bernarto, N. Sudibjo, A. Yulianeu, H. A. Nanda, and K. A. Nanda, "The Secret to Enhancing Innovativeness in the Digital Industry," *International Journal of Innovation, Creativity and Change*, vol. 12, no. 12, pp. 225-243, 2020.

[2] A. Latif, L. A. Fitriana, and M. R. Firdaus, "Comparative Analysis Of Software Effort Estimation Using Data Mining Technique And Feature Selection," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 6, no. 2, February 2020. doi: 10.33480/jitk.v6i2.1968, pp. 167-174.

[3] R. Henglie, Y. Purnomo, and J. A. Ginting, "Predicting Increased H-Index For Research Journals Using The Cost-Sensitive Selective Naive Bayes Classifiers Algorithm," *Jurnal Algoritma, Logika dan Komputasi*, vol. VII, no. 01, pp. 643-651, 2024. doi: http://dx.doi.org/10.30813/j-alu.v2i2.6028.

[4] E. W. dos Santos, I. Nunes, and D. Jannach, "Developer perceptions of modern code review processes in practice: Insights from a case study in a mid-sized company," *Journal of Systems and Software*, vol. 222, Apr. 2025. https://doi.org/10.1016/j.jss.2024.112288.

[5] E. Lumba and L. Hakim, "Digital Library Application Design," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, vol. 6, no. 2, pp. 193-197, 2023.

[6] A. Dennis, B. Wixom, and D. Tegarden, *Systems Analysis and Design: An Object-Oriented Approach with UML*. John Wiley & Sons, 2020.

[7] I. G. N. Suryantara, Michael, J. F. Andry, and J. A. Ginting, "Pengembangan Aplikasi Operasional Restoran Dengan Framework Scrum (Studi Kasus: Restoran PT. XYZ)," *INFOTECH: Journal of Technology Information*, vol. 9, no. 2, Nov. 2023. doi: https://doi.org/10.37365/jti.v9i2.168.

[8] D. Fahriani and T. R. Zubaidah, "Financial Performance Analysis of The Saving and Loan Cooperative," *JKIE (Journal Knowledge Industrial Engineering)*, vol. 10, no. 1, Apr. 2023, pp. 38-49. https://doi.org/10.35891/jkie.v10i1.4125.

[9] B. Kiiza and G. Omiat, "The Impact of Savings and Credit Cooperatives on Household Welfare: Evidence from Uganda," *Journal of Economics and Public Finance*, vol. 7, no. 3, 2021. doi: 10.22158/jepf.v7n3p33.

[10] International Function Point Users Group. Accessed on 10 October 2024. Available online: https://ifpug.org/.

[11] M. F. Hillman and A. P. Subriadi, "40 Years Journey of Function Point Analysis: Against Real-time and Multimedia Applications," *Procedia Computer Science*, vol. 161, pp. 266-274, 2019. https://doi.org/10.1016/j.procs.2019.11.123.

[12] "Function Point Analysis - Introduction and Fundamentals," Accessed on 12 October 2024. Available online: https://www.fingent.com/blog/function-point-analysis-introduction-and-fundamentals/.

[13] D. N. Malleswari, D. Rakesh, K. Subrahmanyam, and D. Vadlamudi, "An Efficient Model for Software Quality Analysis Based on User and Developer Interaction," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, issue 2, Dec. 2019. DOI: 10.35940/ijitee.B6822.129219.

[14] A. W. Wambua and B. M. Maake, "Characterizing Software Quality Assurance Practices in Kenya," *INTERNATIONAL JOURNAL OF SOFTWARE ENGINEERING & COMPUTER SYSTEMS (IJSECS)*, vol. 8, issue 1, pp. 22-28, 2022. https://doi.org/10.15282/ijsecs.8.1.2022.3.0093.

[15] N. Rachmat and S. Saparudin, "Estimasi Ukuran Perangkat Lunak Menggunakan Function Point Analysis - Studi Kasus

Aplikasi Pengujian dan Pembelajaran Berbasis Web," *Prosiding Annual Research Seminar 2017*, vol. 3, no. 1.

[16] D. Ramos-Vidal, W. K. G. Assunção, A. Cortiñas, M. R. Luaces, O. Pedreira, and Á. S. Places, "SPL-DB-Sync: Seamless database transformation during feature-driven changes," *Journal of Systems and Software*, vol. 222, Apr. 2025. https://doi.org/10.1016/j.jss.2024.112285.

AUTHOR GUIDELINES

1. Manuscript criteria

- The article has never been published or in the submission process on other publications.
- Submitted articles could be original research articles or technical notes.
- The similarity score from plagiarism checker software such as Turnitin is 20% maximum.

2. Manuscript format

- Article been type in Microsoft Word version 2007 or later.
- Article been typed with 1 line spacing on an A4 paper size (21 cm x 29,7 cm), top-left margin are 3 cm and bottom-right margin are 2 cm, and Times New Roman's font type.
- Article should be prepared according to the following author guidelines in this template. Article contain of minimum 3500 words.
- References contain of minimum 15 references (primary references) from reputable journals/conferences

3. Organization of submitted article

The organization of the submitted article consists of Title, Abstract, Index Terms, Introduction, Method, Result and Discussion, Conclusion, Appendix (if any), Acknowledgment (if any), and References.

- Title
  The maximum words count on the title is 12 words (including the subtitle if available)
- Abstract
  Abstract consists of 150-250 words. The abstract should contain logical argumentation of the research taken, problem-solving methodology, research results, and a brief conclusion.
- Index terms
  A list in alphabetical order in between 4 to 6 words or short phrases separated by a semicolon (;), excluding words used in the title and chosen carefully to reflect the precise content of the paper.
- Introduction
  Introduction commonly contains the background, purpose of the research, problem identification, research methodology, and state of the art conducted by the authors which describe implicitly.
- Method
  Include sufficient details for the work to be repeated. Where specific equipment and materials are named, the manufacturer's details (name, city and country) should be given so that readers can trace specifications by contacting the manufacturer. Where commercially available software has been used, details of the supplier should be given in brackets or the reference given in full in the reference list.
- Results and Discussion
  State the results of experimental or modeling work, drawing attention to important details in tables and figures, and discuss them intensively by comparing and/or citing other references.
- Conclusion
  Explicitly describes the research's results been taken. Future works or suggestion could be explained after it
- Appendix and acknowledgment, if available, could be placed after Conclusion.
- All citations in the article should be written on References consecutively based on its' appearance order in the article using Mendeley (recommendation). The typing format will be in the same format as the IEEE journals and transaction format.

4. Reviewing of Manuscripts

Every submitted paper is independently and blindly reviewed by at least two peer-reviewers. The decision for publication, amendment, or rejection is based upon their reports/recommendations. If two or more reviewers consider a manuscript unsuitable for publication in this journal, a statement explaining the basis for the decision will be sent to the authors within six months of the submission date.

5. Revision of Manuscripts

Manuscripts sent back to the authors for revision should be returned to the editor without delay (maximum of two weeks). Revised manuscripts can be sent to the editorial office through the same online system. Revised manuscripts returned later than one month will be considered as new submissions.

6. Editing References

- **Periodicals**
  J.K. Author, "Name of paper," Abbrev. Title

of Periodical, vol. x, no. x, pp. xxx-xxx, Sept. 2013.

- **Book**
  J.K. Author, "Title of chapter in the book," in Title of His Published Book, xth ed. City of Publisher, Country or Nation: Abbrev. Of Publisher, year, ch. x, sec. x, pp xxx-xxx.

- **Report**
  J.K. Author, "Title of report," Abbrev. Name of Co., City of Co., Abbrev. State, Rep. xxx, year.

- **Handbook**
  Name of Manual/ Handbook, x ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. xxx-xxx.

- **Published Conference Proceedings**
  J.K. Author, "Title of paper," in Unabbreviated Name of Conf., City of Conf., Abbrev. State (if given), year, pp. xxx-xxx.

- **Papers Presented at Conferences**
  J.K. Author, "Title of paper," presented at the Unabbrev. Name of Conf., City of Conf., Abbrev. State, year.

- **Patents**
  J.K. Author, "Title of patent," US. Patent xxxxxxx, Abbrev. 01 January 2014.

- **Theses and Dissertations**
  J.K. Author, "Title of thesis," M.Sc. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year. J.K. Author, "Title of dissertation," Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

- **Unpublished**
  J.K. Author, "Title of paper," unpublished. J.K. Author, "Title of paper," Abbrev. Title of Journal, in press.

- **On-line Sources**
  J.K. Author. (year, month day). Title (edition) [Type of medium]. Available: http:// www.(URL) J.K. Author. (year, month). Title. Journal [Type of medium]. volume(issue), pp. if given. Available: http://www.(URL) Note: type of medium could be online media, CD-ROM, USB, etc.

7. Editorial Adress

Universitas Multimedia Nusantara
Jl. Scientia Boulevard, Gading Serpong
Tangerang, Banten, 15811
Email: **ultimaijnmt@umn.ac.id**

Paper Title

Subtitle (if needed)

Author 1 Name[1], Author 2 Name[2], Author 3 Name[2]
[1] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired
[2] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired

Accepted on mmmmm dd, yyyy
Approved on mmmmm dd, yyyy

***Abstract*—This electronic document is a "live" template which you can use on preparing your IJNMT paper. Use this document as a template if you are using Microsoft Word 2007 or later. Otherwise, use this document as an instruction set. Do not use symbol, special characters, or Math in Paper Title and Abstract. Do not cite references in the abstract.**

***Index Terms*—enter key words or phrases in alphabetical order, separated by semicolon (;)**

I. INTRODUCTION

This template, modified in MS Word 2007 and saved as a Word 97-2003 document, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. Margins, column widths, line spacing, and type styles are built-in here. The authors must make sure that their paper has fulfilled all the formatting stated here.

Introduction commonly contains the background, purpose of the research, problem identification, and research methodology conducted by the authors which been describe implicitly. Except for Introduction and Conclusion, other chapter's title must be explicitly represent the content of the chapter.

II. EASE OF USE

A. Selecting a Template

First, confirm that you have the correct template for your paper size. This template is for IJNMT. It has been tailored for output on the A4 paper size.

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them.

III. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not add any kind of pagination anywhere in the paper. Please take note of the following items when proofreading spelling and grammar.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Abbreviations that incorporate periods should not have spaces: write "C.N.R.S.," not "C. N. R. S." Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units (SI units are encouraged).
- Do not mix complete spellings and abbreviations of units: "Wb/m2" or "webers per square meter," not "webers/m2." Spell units when they appear in text: "...a few henries," not "...a few H."
- Use a zero before decimal points: "0.25," not ".25."

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number the equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop.

$$\int_0^{r_2} F(r,\phi)\, dr\, d\phi = [\sigma r_2 / (2\mu_0)] \qquad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)," not "Eq. (1)" or "equation (1),"

except at the beginning of a sentence: “Equation (1) is ...”

D. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o.”
- In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset,” not an “insert.” The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively.”
- In your paper title, if the words “that uses” can accurately replace the word using, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect,” “complement” and “compliment,” “discreet” and “discrete,” “principal” and “principle.”
- Do not confuse “imply” and “infer.”
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”
- The abbreviation “i.e.” means “that is,” and the abbreviation “e.g.” means “for example.”

IV. Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention as below

IJNMT_firstAuthorName_paperTitle.

In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper. Please take note on the following items.

A. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include ACKNOWLEDGMENTS and REFERENCES, and for these, the correct style to use is “Heading 5.”

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles, named “Heading 1,” “Heading 2,” “Heading 3,” and “Heading 4”, are prescribed.

C. Figures and Tables

Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1,” even at the beginning of a sentence.

TABLE I. TABLE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy		

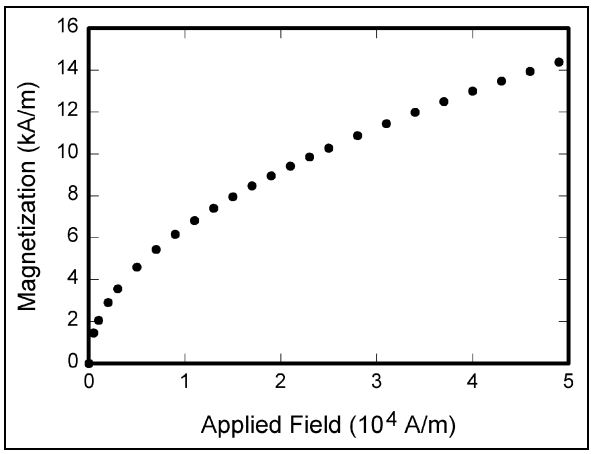


Fig. 1. Example of a figure caption

V. Conclusion

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Appendix

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "F. A. Author thanks ... ." You could also state the sponsor and financial support acknowledgments here.

References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

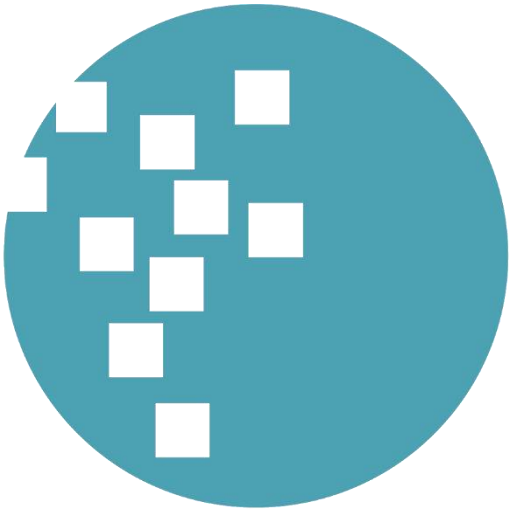
[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2355-0082

Universitas Multimedia Nusantara
Scientia Garden Jl. Boulevard Gading Serpong, Tangerang
Telp. (021) 5422 0808 | Fax. (021) 5422 0800