PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

ANALYSIS OF LIBRARY VISITOR GROUPING THROUGH MASK USAGE IDENTIFICATION IN XIN ZHONG LIBRARY WITH ORANGE DATA MINING APPLICATION

Dwi Permana Putra[1*], Imam Yuadi[2]
dwi.permana.putra-2024@fisip.unair.ac.id[1*], imam.yuadi@fisip.unair.ac.id[2]
Information and Library Science Study Program, Airlangga university, Surabaya, Indonesia

Keyword:
Machine Learning
Data Mining
Image Analysis
Hierarchical Clustering
Orange Application
Article Info:
Submitted date:
2025-02-06
Revised date:
2025-05-16
Accepted date:
2025-05-18

Abstract
The application of data mining in libraries plays a crucial role in supporting data management and monitoring health protocols, especially during the pandemic. A key challenge faced by librarians is effectively monitoring visitors' mask usage compliance. This study aims to analyze visitors' facial images at the library using the Orange Data Mining application, enabling librarians to identify whether visitors are wearing masks. The approach involves collecting random facial images of visitors, preprocessing the data for standardization of size and resolution, extracting features using the Inception V3 model, and conducting hierarchical clustering analysis with the Manhattan metric. The clustering results are visualized in a dendrogram, helping to group the data. The findings show that the dendrogram clearly differentiates between visitors with masks and those without. This visualization provides librarians with an effective tool for monitoring areas of the library that require more strict health protocol supervision. The study concludes that the Orange Data Mining application offers a practical solution for libraries to monitor compliance with health protocols. By utilizing data mining techniques, libraries can enhance visitor safety and comfort. Further research is suggested to expand the dataset and explore other methods to improve analysis accuracy.

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

INTRODUCTION

The COVID-19 pandemic has brought significant changes to various aspects of life, including health protocols. The spread of viruses, including the flu and others, has become a major concern in managing public spaces, such as libraries. As a preventive measure to reduce the transmission of diseases, mask-wearing in public spaces like libraries has become mandatory. However, with the increasing number of visitors, librarians face challenges in monitoring compliance with these health protocols. Previous studies have demonstrated the effectiveness of applying machine learning to support health protocol monitoring in public spaces. For instance, Kanavos et al. (2024) developed a real-time mask detection system based on Convolutional Neural Networks (CNN), which was implemented in public areas such as stations and shopping centers. This study emphasized the system's accuracy in detecting mask usage, but it did not include behavioral analysis of users based on temporal or spatial patterns. Furthermore, Dodda et al. (2025) employed the YOLOv3 approach to detect mask violations in a campus environment, but did not utilize clustering techniques that could help group types of violations or time of occurrence. Meanwhile, Dubey et al. (2025) integrated the Internet of Things (IoT) with deep learning approaches for automatic mask monitoring in public facilities, yet did not explore the potential of visual clustering techniques such as hierarchical clustering for behavioral data analysis.

Therefore, this study focuses on the application of hierarchical clustering in the context of monitoring mask usage in libraries, a semi-public space that has not been widely explored in previous literature. This research aims to provide a new approach to grouping and visualizing user behavior based on image similarity. These methods have relevant applications in managing visual data in libraries for monitoring mask usage. Based on this background, the research problem addressed in this study is:

1. To apply image clustering methods to detect mask usage and non-mask usage among library visitors
2. To conduct data mining analysis on visitors' facial images to assess compliance with health protocols, utilizing the Orange Data Mining application

This study focuses on applying machine learning technology to monitor mask usage in libraries. By utilizing the Orange Data Mining application and hierarchical clustering techniques, this research aims to provide practical solutions for librarians to monitor visitors' compliance with health protocols.

The structure of this article is as follows: Section 2 discusses the methodology, including data collection, preprocessing, feature extraction, and clustering analysis. Section 3 presents the results and analysis, highlighting key findings. Finally, Section 4 provides conclusions and recommendations for future research.

METHODS

This study uses the action research method, which combines both qualitative and quantitative approaches to analyze problems and generate relevant solutions. Action research emphasizes direct data collection from the research object in real-world contexts, while also allowing for corrective actions to address issues encountered during the research process. In this context, the method is applied to monitor visitors' compliance with health protocols in the library.

This study is supported by previous research that demonstrates the successful application of similar methods in various fields, including:

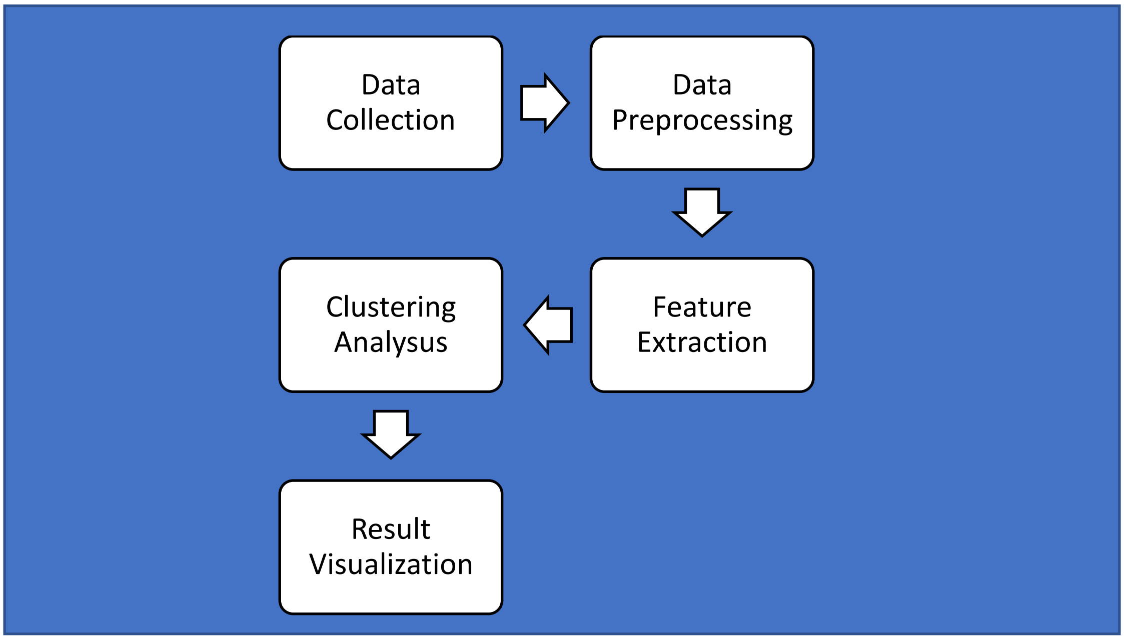
No	Study	Method	Main Results	Discussion
1	Kanavos et al. (2024)	CNN	Real-time mask detection with high accuracy	CNN effectively detects masks quickly and accurately, supporting health protocol monitoring in public spaces.
2	Dodda et al. (2025)	Deep Learning	Real-time mask detection model improves accuracy	Deep learning facilitates fast and precise mask detection to reduce disease spread.
3	Dubey et al. (2025)	Hybrid Deep Learning + IoT	Innovative IoT-based mask monitoring system	IoT integration extends mask detection to large-scale automated monitoring systems.
4	STKIP PGRI Tulungagung Journal	CNN (VGG16Net)	Effective classification of masked and unmasked faces	VGG16Net architecture delivers high performance in classifying face images regarding mask usage.
5	UMJ Journal	Deep Learning	Simple AI model with high accuracy	A simple system effectively detects masks, suitable for practical applications.
6	Untag Surabaya Repository	Pretrained CNN	Mask detection system able to monitor health protocol compliance	Pretrained CNN accelerates training and produces reliable detection accuracy.
7	Unesa E-Journal	Machine Learning	Adequate accuracy in mask detection	Machine learning supports disease prevention through reliable mask detection.
8	ITS E-Journal	Deep Learning on Raspberry Pi 4	Effective real-time system using edge computing device	Implementation on Raspberry Pi provides a portable and efficient solution for field mask detection.
9	Technology and Innovation Journal	Deep Learning	Very high accuracy in monitoring health protocol compliance	Applicable in public facilities for automated health supervision.

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

10	Computer and Information Systems Journal	CNN	CNN more effective than other methods in mask detection	CNN outperforms in speed and accuracy, recommended for mask detection.
11	Journal of Artificial Intelligence Research	CNN Optimization	Real-time mask detection with fast and accurate processing	CNN optimization supports mass health monitoring in crowded environments.

Table 1. Previous Studies Related to Similar Methods
Source: Processed data, 2025

Figure 1. The Data Processing Steps



Source: Processed data, 2025

The research process begins with **data collection**, where images of library visitors' faces are randomly captured. The next step is **data preprocessing**, which involves standardizing the size and resolution of the images to ensure consistency. Following this, **feature extraction** is performed using image descriptors to obtain information such as the file name, image dimensions (length and width), and other

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

visual characteristics. In the **clustering analysis** stage, the hierarchical clustering method is employed with the help of Orange Data Mining software, utilizing the Manhattan metric and the Inception V3 embedder. Finally, in the **results visualization** stage, a dendrogram is created to visualize the clustering of data based on mask usage. The dataset used in this research consists of a collection of randomly captured images of human faces in the library. In this study, each image undergoes several stages of processing involving the use of image descriptors. Image descriptors function to extract important information from the image, such as size, length, width, and other visual characteristics. This stage enables comparison and analysis of differences between images. This process is essential to support further analyses, such as classification and clustering of mask usage on library visitors' faces. (Source: Gonzalez, R. C., & Woods, R. E. (2002). *Digital Image Processing*.) **Distance Metrics** are methods used to measure how far apart or similar two objects are in feature space. In the context of image processing, distance metrics are used to determine the similarity between two images based on the extracted features. In this study, the distance metric used is **Manhattan Distance**, which measures the distance between two points in vector space based on the sum of absolute differences in each feature dimension. Manhattan Distance is often chosen for its simplicity and effectiveness in various cases. (Source: Aggarwal, C. C. (2015). *Data Mining: The Textbook*.) **Image Embedding** is the process of converting an image into a numerical representation or vector that can be more easily processed for further analysis, such as clustering or classification. In this study, **Inception V3** was used as the embedder to convert facial images into feature vectors suitable for clustering. Inception V3 is a deep learning architecture pre-trained to extract essential features from images, thus producing efficient and accurate representations. By using Inception V3, facial images can be processed to recognize relevant patterns, such as differences in mask usage. (Source: Szegedy, C., et al. (2016). *Rethinking the Inception*

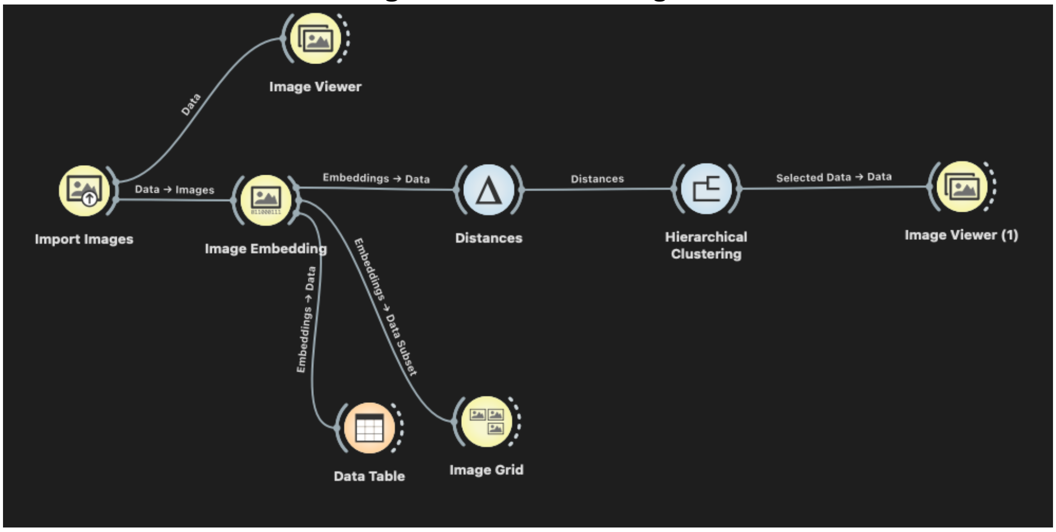
PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

Architecture for Computer Vision. CVPR.) **Hierarchical Clustering** is a data analysis technique that groups data based on certain similarities and organizes the results into a hierarchical tree structure known as a **dendrogram**. The process begins with each data point as an individual cluster, which are then gradually merged until all data points form one large cluster. This technique is highly useful for analyzing patterns or relationships within data in a structured manner. (Source: Xu, R., & Wunsch, D. (2005). *Survey of Clustering Algorithms*. IEEE Transactions on Neural Networks.) The **dendrogram** is a graphical representation of the results of hierarchical clustering, showing how data are grouped based on similarities. In this study, the dendrogram was used to cluster facial images based on mask usage. Blue represents images without masks, while red represents images with masks. This visualization helps librarians effectively monitor visitor compliance with health protocols. (Source: Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*.)

RESULT

This project applies machine learning using the Orange Data Mining application to perform **Hierarchical Clustering** on facial images. Initially random images will be grouped based on categories, namely whether a mask is worn or not. Through this clustering process, we can distinguish between faces wearing masks and those without, providing insights into compliance with health protocols in the library.

Figure 2. Structure Stages



Source: Processed data, 2025

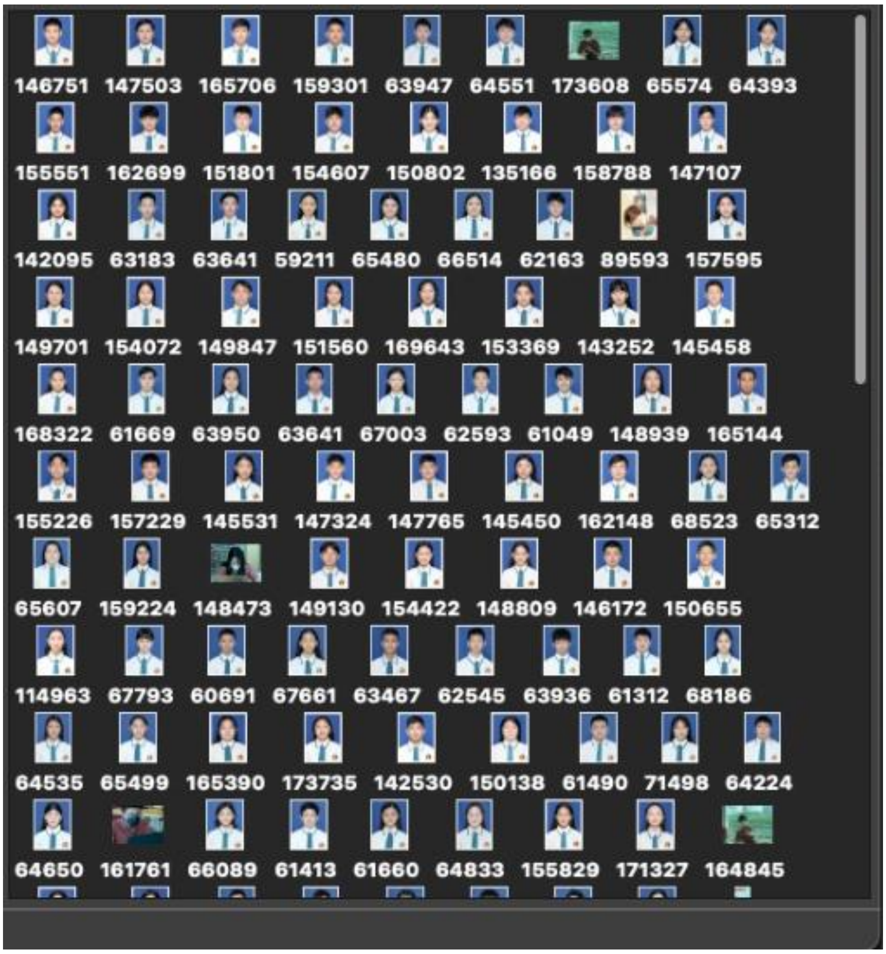
The research begins with data collection, which involves randomly capturing facial images from library visitors. The next step is data preprocessing, where the images are processed to make them ready for analysis. This step includes standardizing the size and resolution of the images to ensure uniform and optimal data.

The following phase is feature extraction, where image descriptors are used to obtain detailed information such as the file name, file size, length, and width of the image. After that, clustering analysis is performed using the Orange Data Mining software. This process results in data clustering, which is visualized in the form of a dendrogram.

In the final stage, the results are evaluated. The clustering performed is further analyzed to evaluate patterns and compare visitors' compliance with mask usage protocols.

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

Figure 3. Dataset to be Imported



Source: Processed data, 2025

The dataset used consists of facial images of humans randomly captured in the library. Each image is processed using image descriptors, which provide information related to the image's size, length, width, and differences between the images. Once the dataset is prepared for data mining in machine learning, we can use image descriptors to extract details such as the file name, file size, and the length and width needed for machine learning. Below is the table containing the image descriptor data.

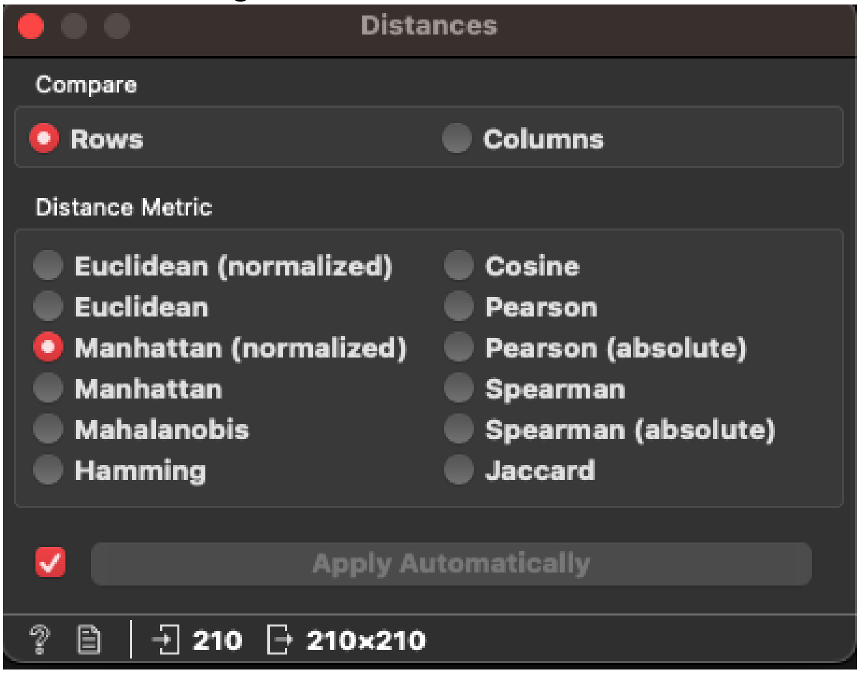
PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

Figure 4. Dataset Table Data

hidden origin type	image name	image iper/Downloac image	size	width	height	n0 True	n1 True	n2 True	n3 True	n4 True	n5 True
1	gambar 81	gambar 81....	146751	388	503	0.136588	0.779329	0.00237941	0.272289	0.0910928	0.145253
2	gambar 95	gambar 95....	147503	392	510	0.211431	0.665848	0.00140952	0.0926518	0.091376	0.157906
3	gambar 42	gambar 42....	165706	389	509	0.112656	0.561093	0.0155491	0.29072	0.268308	0.218252
4	gambar 56	gambar 56....	159301	395	506	0.265912	0.50681	0.00346186	0.273148	0.284686	0.17972
5	gambar 132	gambar 132...	63947	353	471	0.235745	0.34849	0.00285533	0.172002	0.260067	0.247984
6	gambar 126	gambar 126...	64551	352	468	0.19454	0.509816	0.012757	0.287567	0.0933218	0.135965
7	gambar 2	gambar 2.j...	173608	1280	960	0.463722	0.422221	0.264134	0.343991	0.250172	0.586121
8	gambar 127	gambar 127...	65574	353	468	0.175607	0.74134	0.0153247	0.177057	0.17491	0.21876
9	gambar 133	gambar 133...	64393	352	468	0.213533	0.693012	0.0341623	0.355479	0.25664	0.144177
10	gambar 57	gambar 57....	155551	394	512	0.138652	0.726312	0.000613406	0.172654	0.208508	0.12144
11	gambar 43	gambar 43....	162699	389	512	0.188765	0.470284	0.0192383	0.220448	0.189208	0.178144
12	gambar 94	gambar 94....	151801	384	507	0.158351	0.470339	0.0354791	0.283192	0.165205	0.277352
13	gambar 80	gambar 80....	154607	392	510	0.135306	0.496534	0.00198183	0.211656	0.138355	0.327934
14	gambar 96	gambar 96....	150802	389	500	0.13388	0.832682	0.0548732	0.316783	0.0805714	0.176279
15	gambar 82	gambar 82....	135166	388	509	0.120623	0.558355	0.00762164	0.134023	0.121286	0.105419
16	gambar 55	gambar 55....	158788	389	508	0.167175	0.659082	0.017004	0.0771643	0.124867	0.246622
17	gambar 41	gambar 41....	147107	391	507	0.12922	0.757891	0.00160833	0.227514	0.189156	0.193047
18	gambar 69	gambar 69....	142095	394	508	0.11392	0.640974	0.00530242	0.148333	0.291004	0.33813
19	gambar 125	gambar 125...	63183	350	467	0.158661	0.586688	0.0240277	0.335338	0.306065	0.152256
20	gambar 131	gambar 131...	63641	350	469	0.223565	0.579688	3.62343e-05	0.296167	0.0969606	0.267171
21	gambar 119	gambar 119...	59211	354	472	0.151801	0.56619	0.0145538	0.226133	0.324423	0.237394
22	gambar 118	gambar 118...	65480	353	468	0.309241	0.699952	0.0270966	0.217618	0.091448	0.299364
23	gambar 130	gambar 130...	66514	354	472	0.165532	0.703934	0.025511	0.172982	0.420595	0.229451
24	gambar 124	gambar 124...	62163	352	470	0.192121	0.701971	0.00675851	0.322056	0.0924146	0.129583
25	gambar 209	gambar 20...	89593	959	1280	0.167303	0.307348	0.0316341	0.598362	0.182428	0.529212
26	gambar 68	gambar 68....	157595	387	510	0.265668	0.754697	0.0365309	0.121924	0.18992	0.209916
27	gambar 40	gambar 40....	149701	384	513	0.13289	0.611467	0.0115174	0.227763	0.309994	0.292183
28	gambar 54	gambar 54....	154072	385	510	0.131436	0.718753	0.00997162	0.184493	0.247475	0.224639
29	gambar 83	gambar 83....	149847	387	510	0.109706	0.503094	0.000857115	0.354471	0.295829	0.234477
30	gambar 97	gambar 97....	151560	390	509	0.122751	0.622301	0.0026334	0.194871	0.200258	0.34865
31	gambar 93	gambar 93....	169643	391	503	0.210592	0.679422	0.00631479	0.215149	0.159781	0.240719
32	gambar 87	gambar 87....	153369	392	510	0.122786	0.860446	0.0341582	0.235561	0.218193	0.275517
33	gambar 78	gambar 78....	143252	392	501	0.172908	0.634229	0.0381558	0.212604	0.104222	0.418781

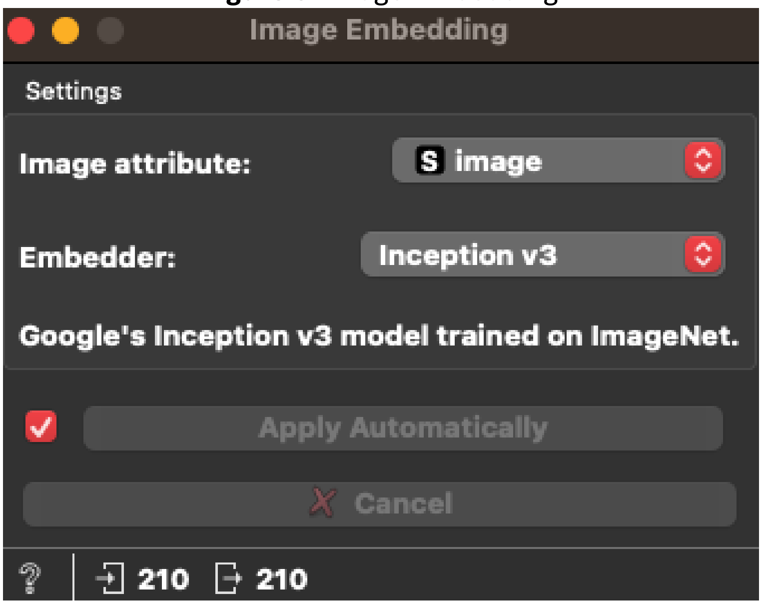
Source: Processed data, 2025

Figure 5. Distance Metric Method



Source: Processed data, 2025

Figure 6. Image Embedding



Source: Processed data, 2025

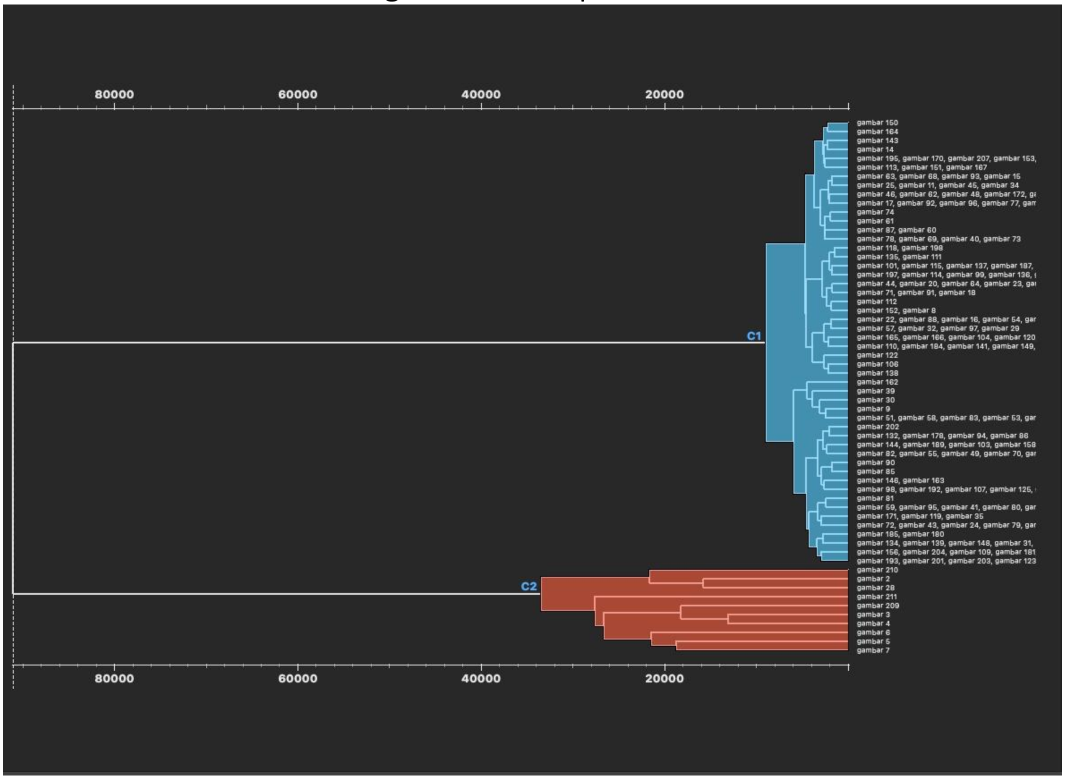
In the clustering analysis stage, the distance metric used is Manhattan. This method measures the distance between data points based on the sum of absolute

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

differences across feature dimensions, which is considered suitable for data with varying scales.

Meanwhile, for image embedding, the Inception V3 model is used as an embedder. This model converts images into more concise yet informative feature representations. The features generated by this embedder are then used as inputs for the clustering process.

Figure 7. Techniques Used

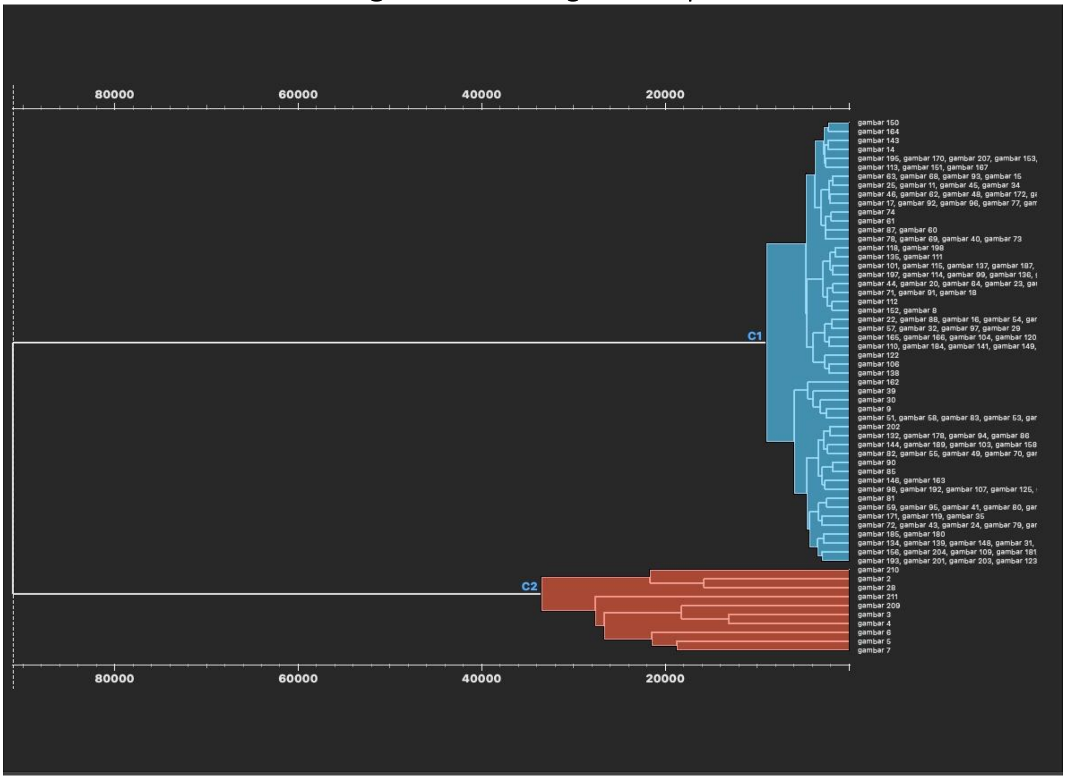


Source: Processed data, 2025

Hierarchical Clustering is the technique used in this project, where data (face images) are grouped based on specific similarities, forming a hierarchical structure or tree called a dendrogram. This technique allows us to visualize the clustering of data in a more structured manner, making it easier to analyze.

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

Figure 8. Dendrogram Shape



Source: Processed data, 2025

The resulting dendrogram displays the clustering based on mask usage. The areas marked in blue represent images of individuals wearing masks, while those in red represent images of individuals not wearing masks. This makes it easier for librarians to monitor patrons' compliance with health protocols.

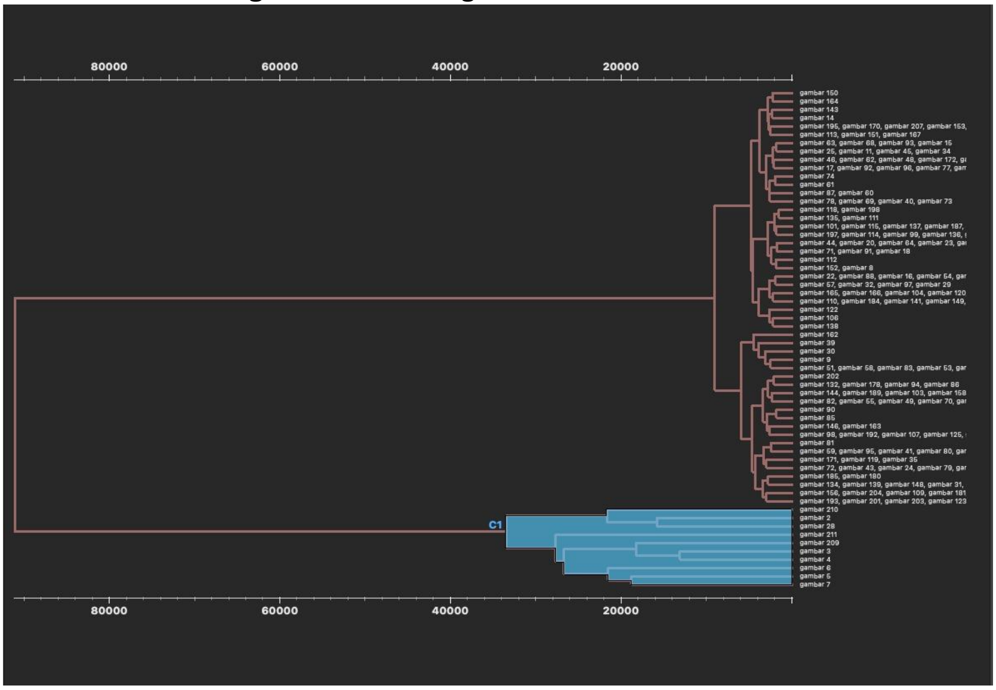
By clicking on the red area highlighted in the dendrogram, we can view patrons who are not wearing masks. Conversely, clicking on the blue area will show patrons who are wearing masks. This clustering result helps librarians identify which areas need more attention in terms of mask-wearing compliance.

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

Figure 9. Clustering Results with Mask

Source: Processed data, 2025

Figure 10. Dendrogram Results with Mask



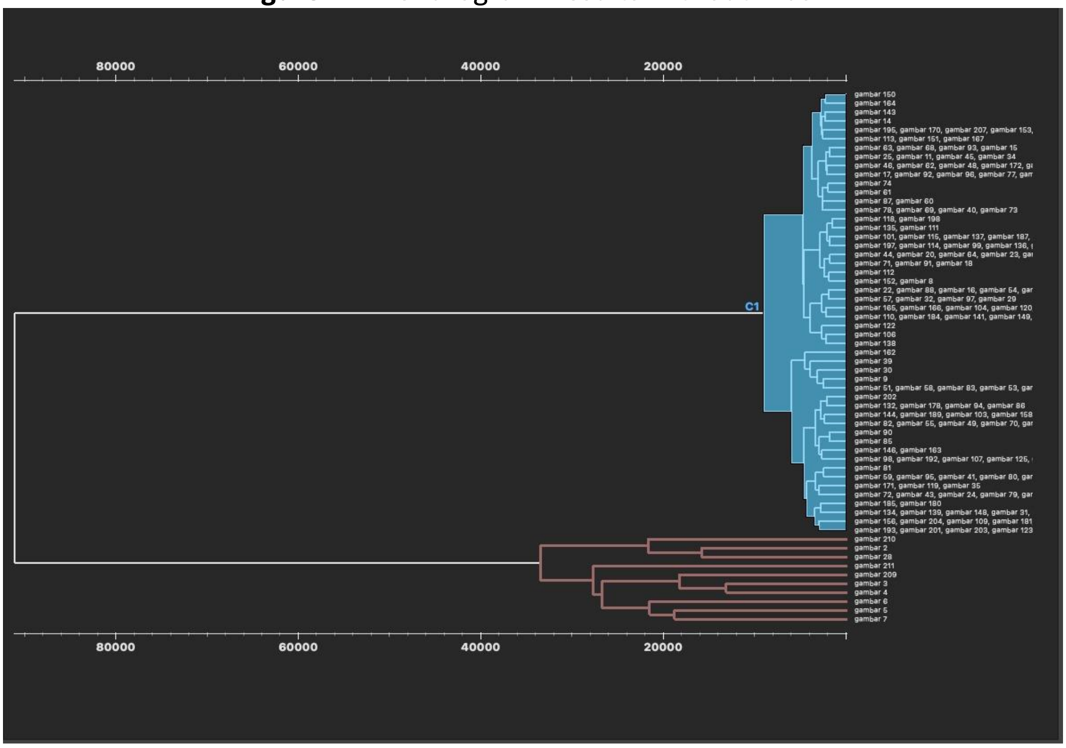
Source: Processed data, 2025

PUBLIS JOURNAL
Publication Library and Information Science
ISSN 2598-7852
Volume 09 Number 01 May 2025
http://journal.umpo.ac.id/index.php/PUBLIS

Figure 11. Clustering Results without Mask

Source: Processed data, 2025

Figure 12. Dendrogram Results without Mask



Source: Processed data, 2025

DISCUSSION

The application of hierarchical clustering with Orange Data Mining in this study demonstrates that clustering techniques can effectively assist in monitoring health protocols in library environments. By utilizing this technology, librarians can visually identify visitors who comply or do not comply with mask usage. This system also provides ease of use, enabling librarians with limited technical skills to operate it effortlessly. The success of this technique shows that clustering methods applied to image data can be expanded for similar purposes in other public spaces, such as museums or archives, where compliance with health protocols must also be monitored.

Such an approach is highly relevant given the importance of adhering to health protocols to prevent virus transmission. In addition to facilitating librarians in monitoring visitor compliance in real-time, this clustering outcome can also be used to analyze compliance patterns at specific times. This can help library management in developing strategies to raise visitor awareness of health protocols. On a larger scale, this research opens opportunities for further development using more advanced machine learning methods, such as deep learning, to improve the accuracy and efficiency of image-based health monitoring systems in public spaces.

CONCLUSIONS

This study employed the action research method to analyze library visitors' compliance with health protocols through the implementation of machine learning technology. Specifically, hierarchical clustering using the Manhattan Distance metric and feature extraction through image embedding with Inception V3 were applied to group visitors based on mask usage. This approach successfully distinguished visitors who wore masks from those who did not. The resulting dendrogram provided a clear and intuitive visualization, enabling library management to monitor compliance levels more efficiently and support more accurate decision-making in enforcing health protocols.

However, this study has several limitations. The dataset was limited to images taken from a single library, thus it does not yet represent broader environmental conditions and visitor behaviors. Additionally, the system focused solely on mask usage detection and did not address other aspects of health protocols such as social distancing or hand hygiene. The current visualization lacks an automatic notification feature, so manual monitoring is still required.

For future research, it is recommended to expand the dataset by including images from multiple libraries with diverse settings and visitor demographics to improve model accuracy and generalizability. Further system development could incorporate automatic notification features so that library management can receive real-time alerts when non-compliant visitors are detected, allowing for immediate corrective action. This method also has the potential to be adapted to monitor compliance with other library policies, such as mobile phone usage restrictions or visitor capacity limits, and could even be applied in other sectors requiring monitoring of health or safety regulation compliance.

REFERENCES

Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.

Dodda, R., Raghavendra, C., & Azmera, C. N. (2025). Real-Time Face Mask Detection Using Deep Learning: Enhancing Public Health and Safety. *E3S Web of Conferences*, 02013. https://www.e3sconferences.org/articles/e3sconf/pdf/2025/16/e3sconf_icregc d2025_02013.pdf

Dubey, P., Keswani, V., & Bhagat, D. (2025). Innovative IoT-enabled Mask Detection System: A Hybrid Deep Learning Approach for Public Health Applications. *MethodsX*, Elsevier. https://www.sciencedirect.com/science/article/pii/S2215016125001372

E-Journal Institut Teknologi Sepuluh Nopember (ITS). *Sistem Otomatis Pendeteksi Wajah Bermasker Menggunakan Deep Learning.* Diakses dari: https://ejurnal.its.ac.id/index.php/teknik/article/viewFile/59790/6649

E-Journal Universitas Negeri Surabaya (UNESA). *Sistem Pendeteksian Penggunaan Masker pada Wajah Menggunakan Metode Machine Learning*. Diakses dari: https://ejournal.unesa.ac.id/index.php/jurnal-manajemen-informatika/article/view/55867

Gonzalez, R. C., & Woods, R. E. (2002). *Digital Image Processing*. Prentice Hall.

Journal of Artificial Intelligence Research. Optimizing Convolutional Neural Networks for Real-Time Face Mask Detection. Diakses dari: https://www.jair.org/index.php/jair/article/view/112233

Jurnal Komputer dan Sistem Informasi. Penerapan CNN untuk Deteksi Penggunaan Masker pada Citra Wajah. Diakses dari: https://jurnalkomputer.sisteminformasi.ac.id/article/view/56789.

Jurnal STKIP PGRI Tulungagung. *Klasifikasi Deteksi Penggunaan Masker Menggunakan Metode Convolutional Neural Network (CNN).* Diakses dari: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/3748

Jurnal Teknologi dan Inovasi. Penerapan Deep Learning untuk Deteksi Kepatuhan Protokol Kesehatan di Tempat Umum. Diakses dari: https://journal.technology-and-innovation.org/article/view/12345.

Jurnal Universitas Muhammadiyah Jakarta (UMJ). *Aplikasi Deteksi Masker Wajah Menggunakan Metode Deep Learning*. Diakses dari: https://jurnal.umj.ac.id/index.php/just-it/article/view/21863.

Kanavos, A., Papadimitriou, O., & Al-Hussaeni, K. (2024). Real-Time Detection of Face Mask Usage Using Convolutional Neural Networks. *Computers*, 13(7), 182. https://www.mdpi.com/2073-431X/13/7/182

Kim, S., & Patel, R. 2020. *Hierarchical Clustering Applications in Visual Data Analysis.*

Li, Q., et al. 2021. *Cluster Analysis for Health Data Using Orange Data Mining.*

Rahman, A., et al. 2020. *Image Classification with Orange Data Mining.*

Repository Universitas 17 Agustus 1945 Surabaya (UNTAG SBY). *Sistem Deteksi Pemakaian Masker Menggunakan Convolutional Neural Network*. Diakses dari: https://repository.untag-sby.ac.id/11269/8/JURNAL.pdf.

Smith, J., & Zhao, L. 2021. *Monitoring Public Health Compliance with Machine Learning*.

Szegedy, C., et al. (2016). *Rethinking the Inception Architecture for Computer Vision. CVPR.*

Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining. Addison-Wesley.*

Williams, T., & Ahmed, K. 2019. *Using Inception V3 for Feature Extraction in Image Analysis.*

Xu, R., & Wunsch, D. (2005). *Survey of Clustering Algorithms. IEEE Transactions on Neural Networks.*