DOI: xxx

***Type: xxx***

**Unsupervised image classification and latent representation**

**Ali**

Received: XX Month 202X; Accepted: XX Month 202X

**Abstract:** Image classification in the field of medical imaging, significantly, relies on supervised learning. However, supervised learning faces challenges such as limited labeled data, model complexity, and difficulties in interpreting model predictions. Some of these challenges can be overcome by adopting unsupervised learning, but the accuracy of unsupervised learning should be addressed. In this study, we introduce an approach that integrates pre-trained models with clustering methodologies, achieving notable accuracy in unsupervised image classification. The image encodings are plotted in 2D to show some degree of separation in the low-dimensional manifold. Various state-of-the-art models, such as ResNet and DenseNet, are used to obtain image encodings, and classical approaches, such as convolutional autoencoders and scale-invariant feature transform (SIFT) with a histogram of gradients (HOG). The results are reported on the CIFAR-10 dataset, and a decent, maximum accuracy of 68.14% is achieved on unsupervised image classification. The author(s) received no specific funding for this study.

**Keywords:** Unsupervised learning, ResNet and DenseNet, CIFAR-10 dataset.

**1 Introduction**

Image classification plays a pivotal role in medical imaging, offering valuable insights into the content of medical images. Algorithms like Convolutional Neural Networks (CNNs) (see list of acronyms in appendix) offer rapid and precise classification of medical images across different categories. Additionally, clustering techniques can group alike medical images, aiding in pattern recognition. [1]. Adopting such technologies can enhance diagnostic accuracy and efficiency for medical professionals. Convolutional Neural Networks (CNNs) are commonly employed machine learning algorithms for image classification. These algorithms are trained on large datasets of medical images to learn patterns and classify images into various categories [1, 2]. By using these algorithms, medical professionals can quickly and accurately detect abnormalities in medical scans. This can be beneficial for diagnosing conditions and saving time for physicians.

Supervised and unsupervised image classification are two important techniques used for medical image analysis. Supervised image classification involves categorizing a medical image into different classes based on the presence of certain features or characteristics. This type of classification is often used to detect abnormal tissue or organs, such as tumors or blood vessels. The model used for supervised image classification is often a supervised machine learning algorithm. It uses a training dataset with labeled data to learn how to classify the images [3]. Unsupervised image classification, on the other hand, is a method of analyzing medical images without the need for labeled data. Instead, the algorithm looks for patterns or features present in the image and assigns labels accordingly. This type of classification is often used to analyze tissue segments in medical images, such as organ boundaries or muscle tissue. The most common model used for unsupervised image classification is clustering, which is used to group similar objects or areas.

Supervised learning in medical imaging uses labeled data to build a model that can accurately predict the outcome of a given input. The most commonly used algorithms for supervised learning are decision trees, logistic regression, linear regression, and support vector machines. A decision tree is used to classify data points into different categories. Logistic regression and linear regression are used to find the relationship between two variables, while a support vector machine is used to build a model that predicts the output based on the given input [3,4] Unsupervised learning in medical imaging uses unlabeled data to detect patterns and structure in the data. The most commonly used algorithms for unsupervised learning are k-means clustering, hierarchical clustering, and apriori algorithm. K-means clustering is used to group similar data points into clusters. Hierarchical clustering is used to generate clusters from a hierarchy of clusters. Apriori algorithm is used to identify frequent patterns in the data [5]. All these algorithms are used to uncover patterns and structures in the data that can be used for medical diagnosis and treatments. Both supervised and unsupervised image classification is important techniques for medical imaging. While supervised classification requires labeled data, unsupervised classification requires no labeled data, making it useful for analyzing medical images without any prior knowledge about the image [6]. Furthermore, both methods can be used to detect abnormalities or to segment and classify tissues in medical images but supervised methods are significantly more labor intensive. Conversely, unsupervised learning techniques attain significantly lower accuracy than supervised methods, necessitating further research.

In recent years, semi-supervised learning (SSL) [7] has emerged as an exciting new analysis direction in deep learning. Such methods influence cases in which few labeled coaching examples are offered along with a large variety of unlabeled samples. In such a setting, SSL strategies are highly applicable to real-world applications in which the unlabeled knowledge is available and simple to accumulate, whereas labeled instances are typically exhausting, expensive, and require time to gather. SSL can build higher classifiers that atone for the dearth of labeled training knowledge. However, to avoid poor matter structure matching with the model assumption, which might cause degradation in classification performance, SSL is only effective under certain assumptions, such as presumptions that the choice boundary ought to avoid regions with high density, facilitating the extraction of extra information from the unlabeled instances to regularize coaching. We move beyond semi-supervised learning to completely unsupervised learning for image classification [8]. Most of the data in today’s world follow an unsupervised learning methodology, i.e., the labels are not present with all the data points.

Obtaining labeled data is very often a painful process, so we propose an experimental approach for classifying images in an unsupervised manner where image encodings from pre-trained models are obtained and then clustering is applied to form an end-to-end unsupervised model. We discuss various pre-trained models, such as DenseNet [9] and ResNet [10], in this paper and show the encodings in 2D manifolds using the t-SNE dimensionality reduction technique [11]. The results are discussed on the CIFAR-10 dataset [12].

SIFT with HOG [13] was also used as a feature extractor but did not yield good results along with convolutional autoencoders [14] and ResNets. The methodology for all these experiments was the same, i.e., extracting features and then using K-means to cluster the data. The basis of unsupervised learning is to obtain somewhat separable data and then perform clustering or any other unsupervised models. We show the learned features in 2D space to visualize the various classes present in the CIFAR-10 dataset and how they are somewhat separate after extracting the image encodings. This work can be extended to practical uses where state-of-the-art models can be deployed in an unsupervised manner to obtain good results with minimal effort to obtain labeled data.

# 2 Related Worked

In this work, we present the current state of the art in both the semi-supervised and unsupervised learning fields. Supervised and unsupervised image classification are two methods for assigning labels to images. Supervised image classification relies on human experts to provide labels for a set of images; the labels are then used to train a model to classify new images. This approach is effective when labels are available and the goal is to classify images into discrete categories, but it is labor-intensive and limited by the number of labels provided [15]. Unsupervised image classification, on the other hand, seeks to identify patterns in the data without relying on labeled data. It is useful when labels are not available and when exploring the data for patterns. The unsupervised classification does not require labeled data and instead relies on the model’s ability to detect patterns in the data. This approach allows for more complex categorizations and can be used to identify unknown relationships between images. However, since it does not use labeled data, it can be more prone to errors and inaccuracies. [16]. consequentially, we consider semi-supervised learning as the most applicable progression in the field of medical imaging.

Semi-supervised learning describes a class of algorithms that seek to learn from both unlabeled and labeled cross-sections, normally assumed to be tested from the same or parallel distributions. We concentrate on recent developments grounded on deep neural networks in medical imaging. Notably, semi-supervised learning in medical imaging through deep learning models presents several challenges. Firstly, medical images are often extremely high-dimensional, which can lead to several issues such as overfitting and computational complexity [17]. Moreover, the data used for training and validation can be highly imbalanced, which can lead to poor results due to the lack of enough data in certain classes [16, 18]. Additionally, the data used for training can also be very noisy, which can result in the learning algorithm being unable to accurately identify patterns in the data [18, 19].

Attempts to overcome the challenges of semi-supervised learning involve proposing novel deep learning frameworks, such as the Attention-based Semi-supervised Deep Networks (ASDNet), to fulfill segmentation tasks in an end-to-end fashion [20]. This framework includes a convolutional confidence network that is used to adversarially train the segmentation network and a region-attention-based semi-supervised learning strategy that allows for the inclusion of unlabeled data for training. Such methods can help to alleviate the need for labor-intensive ground-truth label maps and expertise knowledge. Moreover, numerous of the original developments in semi-supervised learning deep neural networks were predicated on generative models similar to denoising autoencoders [21], variational autoencoders [22], and generative adversarial networks [23]. We used DenseNets and K-means clustering. DenseNet is used for feature extraction.

Clustering [24] is a set of methodologies employed to partition data into groups or clusters. Clusters are approximately outlined as groups of data objects that are more analogous to other objects in their cluster than they are to data objects in other clusters. In practice, clustering helps identify two classes of data. Meaningful data expand the sphere of knowledge. For illustration, in the medical field, experimenters applied clustering to gene expression trials. The clustering results linked groups of cases who responded to medical treatments. Usefulness in clustering, in contrast, serves as a median step in a data channel. For illustration, businesses apply clustering for customer segmentation. The clustering results in clustering member guests into groups with similar purchase histories, which companies can also exercise to produce useful targeted advertising. Conventional K-means require many methods. The first step is to randomly select k centroids, where k is equal to the number of clusters chosen.

Centroids are data points representing the center of a cluster. The main element of the algorithm works by a two-step process called expectation maximization. The expectation step assigns each data point to its nearest centroid. The maximization step computes the mean of all the points for each cluster and sets the new centroid. The quality of the cluster assignments is determined by calculating the sum of the squared error (SSE) after the centroids meet or match the former replication’s assignment. The SSE is defined as the sum of the squared Euclidean distances of each point to its closest centroid. Since this is a measure of error, the ideal of K-means is to attempt to minimize this value. In our research, to map the clusters generated by the k-means algorithm to the true labels of the CIFAR-10 dataset, we employed the Hungarian algorithm, a combinatorial optimization technique. We constructed a cost matrix, where each element represents the negative count of instances from a particular class in a specific cluster. The negation was necessary as the Hungarian algorithm is inherently a minimization procedure, but our objective was maximization. Upon application of the algorithm to the cost matrix, we derived an optimal one-to-one mapping between clusters and true labels, ensuring the highest possible correspondence between clustered data and their actual classes.

DenseNets is a densely connected convolutional neural network [25] that is used more than a traditional CNN because DenseNet requires fewer parameters than a traditional CNN. This approach was specifically developed to improve performance and accuracy caused by the vanishing gradient, particularly in a high-level neural network, due to the long distance between the output and input layers causing the information to vanish before reaching its destination. In this paper, we specifically used DenseNet to extract features from images. Additionally, for parameter efficiency – every layer adds only a limited number of parameters, and implicit deep supervision – the improved flow of gradient through the network. CNNs are used to classify the CIFAR-10 dataset in an unsupervised manner, and the accuracy reported is approximately 70% [26]. KNNs are also used for CIFAR-10 classification with CNNs with supervised learning and good accuracy [27]. GANs and adversarial training with clustering has also been applied to similar datasets [28]. In our exploration of unsupervised image classification, we attained a commendable accuracy of 68.14% utilizing pre-trained models combined with clustering techniques. In comparison, the study by Caron et al. (2019) introduced the innovative DeepCluster algorithm, focusing on a convolutional neural network's feature clustering, While our method leveraged a straightforward approach of feature extraction followed by k-means clustering, Caron et al. emphasized a more intricate deep clustering mechanism, though exact performance metrics weren't explicitly stated in the extracted sections [29].

The literature explored various advancements and techniques in unsupervised image classification. Significant emphasis has been given to the use of deep learning models, their evolution, and their application in real-world scenarios. While many models have shown promise, challenges in training, accuracy, and computational efficiency persist. The survey highlights a trend toward leveraging pre-trained models, which have shown to enhance performance. However, there remains a gap in achieving optimal accuracy, especially in complex datasets, underscoring the need for further research in this domain.

**3 Contribution**

1. **Introduction of a novel approach:**

The paper contributes to the field of image classification by introducing a novel approach that combines pre-trained models with clustering techniques for unsupervised image classification. This approach offers an alternative solution for situations where labeled data is scarce or unavailable.

1. **Evaluation of semi-supervised learning in medical imaging:**

The paper explores the challenges and potential benefits of semi-supervised learning specifically in the context of medical imaging. By highlighting the difficulties associated with traditional supervised learning in this domain, the paper provides insights into the effectiveness of semi-supervised techniques for medical image classification.

1. **Development of Attention-based Semi-supervised Deep Networks (ASDNet):**

The paper presents the development of ASDNet, a deep learning framework that incorporates techniques like adversarial training and region-attention-based semi-supervised learning. This framework addresses the challenges of high dimensionality, imbalanced data, and noise commonly encountered in medical imaging datasets.

1. **Experimental evaluation:**

The paper conducts experiments to evaluate the proposed approach's performance. It reports an accuracy of 68.14% in unsupervised image classification using DenseNets and clustering. These experimental results contribute to the understanding of the effectiveness of pre-trained models combined with clustering techniques in image classification tasks.

1. **Visualization techniques:**

The paper employs visualization techniques, such as t-SNE, to visualize the lower-dimensional representations of image encodings obtained from pre-trained models. By visualizing these representations, the paper enhances the interpretability and understanding of the clustering results.

1. **Potential applications:**

The paper discusses potential applications of the unsupervised image classification approach in fields like intracavity absorption spectroscopy. By highlighting these applications, the paper opens up new avenues for research and encourages further exploration of the proposed approach in other domains.

1. **Comparative analysis:**

The paper compares the proposed approach with existing methods in image classification to showcase its advantages and limitations. This comparative analysis contributes to the existing body of knowledge by demonstrating the strengths and weaknesses of the proposed approach in relation to other techniques.

1. **Dataset availability:**

By making the dataset available on [GitHub](https://github.com/Ali-Almalki/Unsupervised-image-classification-and-latent-representation), the paper contributes to the research community by providing a valuable resource for others to replicate and build upon the findings. This promotes transparency, reproducibility, and collaboration within the field.

These contributions collectively enhance the understanding of unsupervised and semi-supervised image classification, specifically in the context of medical imaging. They provide valuable insights, techniques, and experimental results that can guide future research and advance the field.

# 4 Methodology

The image encodings from pre-trained models such as ResNets and DenseNets were obtained and then clustered using K-means clustering. The main idea behind this approach is to reduce the input dimensions while maintaining some sort of information about the image and then producing the clusters in an unsupervised manner. Figure 2 displays the encoded representations of images using DenseNets. These encodings, originally in higher-dimensional space, have been transformed to a 3-dimensional space for visualization purposes using the t-SNE dimensionality reduction technique. The image encodings are shown in Figure 2 below after reducing to 3 dimensions using t-SNE. As seen in the future, the image encodings shown are very distinct and are well-separated to some extent. This proves that the low- dimensional features extracted from the image were good enough to separate them to some extent. There are 10 classes in the CIFAR-10 dataset; hence, 10 color coding schemes are used.

The encodings were obtained from autoencoders as well, where the input image and the output image were the same and the convolutional autoencoder attempted to minimize the reconstruction loss. Figure 3 delineates the nuanced procedure adopted for unsupervised image classification leveraging deep learning architectures. The following elucidates the salient stages and technical intricacies:

**4.1 Model Architecture Selection**

A judicious examination of established deep learning architectures, notably ResNet and DenseNet, was undertaken. These architectures, benefitting from rigorous training on expansive datasets, facilitate potent feature extraction—imperative for the classification task at hand.

**4.2 Architectural Depth and Configuration**

DenseNet is characterized by a composite of dense blocks interspersed with transition layers. The precise depth, manifesting as the number of layers, is contingent upon the specific variant under consideration (e.g., DenseNet-121 or DenseNet-169). In this study ResNet-50 and DenseNet-201 were used.

**Feature Extraction Paradigm**

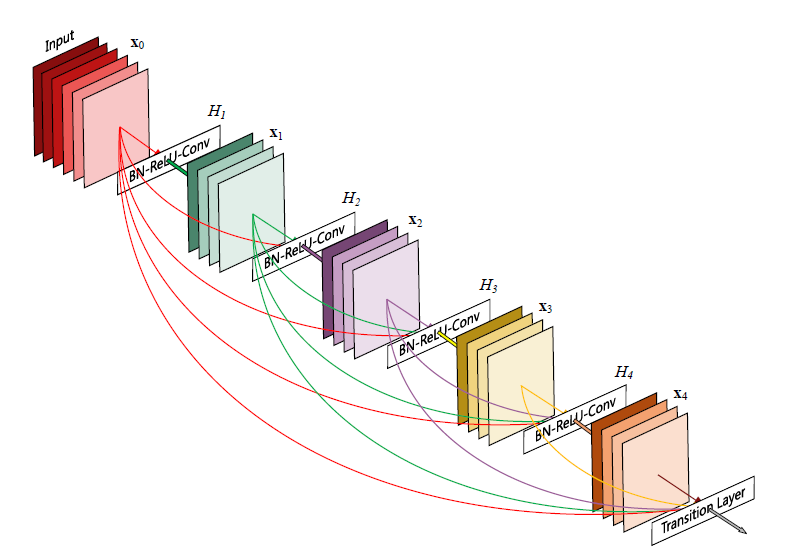
The primary endeavor entails capitalizing on the architectures to distill salient features from images within the CIFAR-10 dataset. This process necessitates the propagation of images through the network and the subsequent abstraction of feature maps from intermediary layers, which encapsulate intricate image patterns.

**Unsupervised Clustering Mechanism**

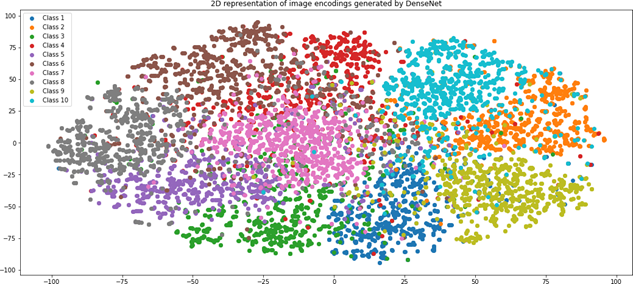
Subsequent to the feature extraction phase, unsupervised algorithms, notably K-means, are employed to orchestrate the images into discernible clusters predicated on the extracted features. This stratification accords the categorization of images into discrete clusters, each emblematic of a distinct category.

**Dimensionality Reduction via t-SNE**

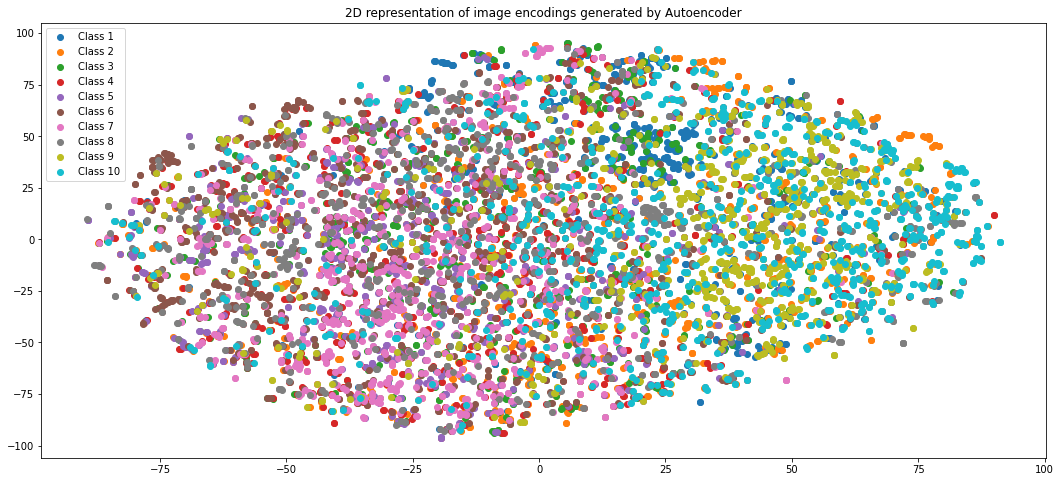
To facilitate an intuitive representation and discernment of the distinction between clusters, the high-dimensional feature vectors undergo a transformation to a 2D space using the t-SNE technique. This transformation offers an incisive visualization of cluster separation within this reduced feature space.



**Figure 1** Illustration of the DenseNet Structure [30]

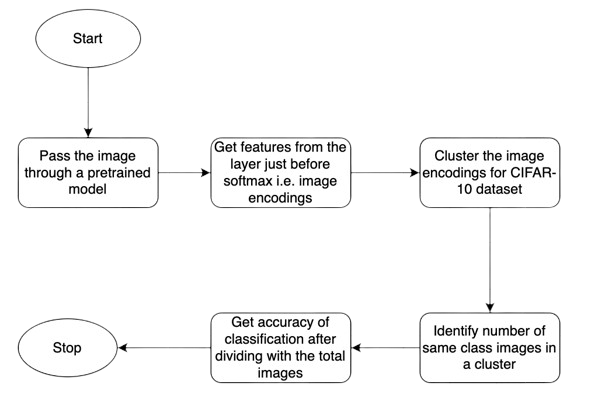


**Figure 2** t-SNE generated 2D representation of image encodings extracted by pre-trained models.



**Figure 3** t-SNE 2D representation of image encodings

In figure 2, after extracting the image encodings the K-means algorithm is applied with k=10 because there are 10 classes in the dataset. Multiple K-means iterations are executed, and the best iteration is used for the results and evaluations. In the flowchart given below in Figure 3, the whole end-to-end process is explained. SIFT and HOG were used for feature extraction from images in an unsupervised manner. They were used to extract the features, such as landmarks and other local features, from the image. The features were computed after considering the gradient’s magnitude and direction. After extracting those features for the CIFAR-10 dataset, K-means clustering was applied similarly, and the results were obtained.



**Figure 4** Flowchart of the whole process

**5 Results and Discussion**

The CIFAR-10 dataset is a collection of images from 10 classes that can be used to train and assess the accuracy of unsupervised learning algorithms. It consists of 60,000 32x32 color images that are divided into 10 classes, with each class containing 6,000 images. The 10 classes represent common objects [30]. In terms of medical imaging applications, the CIFAR-10 dataset can be used to train and assess the accuracy of unsupervised learning algorithms in the diagnosis of certain diseases or conditions. The 10 classes can be mapped to different medical image features, such as X-rays, CT Scans, or MRI images. By training and testing the accuracy of unsupervised learning algorithms on the CIFAR-10 dataset, medical researchers can gain insight into the accuracy of unsupervised learning algorithms for diagnosing certain diseases or conditions.

In this study, we tried various unsupervised learning methods in this paper. The evaluation metric used to interpret the results was accuracy because of the balanced dataset. The accuracy was obtained by counting the same class of CIFAR-10 assigned to 1 cluster from the K-means algorithm. In the table given below, the results of our methods are shown. The accuracy is not very good because the model did not have labels to learn from. The dataset was divided into an 80-20 split i.e., 80% of the data was used for training, and 20% was used for testing. K-means was executed multiple times to obtain the results because of its stochastic nature.

**Table 1** Test accuracy for different image encoding methods.

| Image Encodings | Test Accuracy | Average Precision | Average Recall | AverageF1 Score |
| --- | --- | --- | --- | --- |
| DenseNet | 68.14% | 58% | 62% | 57% |
| ResNet | 22.47% | 22% | 24% | 22% |
| SIFT and HOG | 27.03% | 24% | 24% | 24% |
| Convolutional Autoencoder | 24.58% | 22% | 22% | 22% |

Various algorithms were used for unsupervised learning. In this paper, we compare the results with existing supervised methods for the CIFAR-10 dataset as shown in Table 2

**Table 2** Comparing the proposed method with other algorithms based on the CIFAR-10 dataset

| Method | Test Accuracy |
| --- | --- |
| Residual Network with Attention | 96.10% |
| ConvNets for Rotation Prediction | 89.90% |
| Invariant Image Clustering | 61.7% |
| Our Method | 68.14% |

Comparing the models indicate that the residual network with attention and convNets for rotation prediction has higher accuracy than the proposed model. However, our proposed model provided among the highest accuracy ratings for unsupervised learning in medical imaging because of the specific use case, and the pretraining was based on the CIFAR-10 datasets. For instance, pre-training done using Imagenet-1k was applied for deep convolutional generative adversarial networks (DGCAN) and used to classify CIFAR-10, yielding an accuracy of 63.7% for K=1 [31]. Thus, the attained accuracy is significantly higher than the previous works under the same CIFAR-1o datasets, but could be improved further, implying that the outcomes of this study lay a foundation for future improvements.

**5.1 Correlation with Intracavity Absorption Spectroscopy Studies**

Our work emphasizes unsupervised image classification using deep learning. Interestingly, recent studies in intracavity absorption spectroscopy highlight the importance of precision and accuracy in sensor efficiency [32]. While seemingly disparate, both fields deal with deciphering complex patterns. Our technique could potentially refine the algorithms for spectroscopy system measurements. Another study focuses on components like the Fibre Bragg Grating (FBG) [33]. We envisage our models assisting in predicting these components' performance. While direct implementation necessitates further adaptation, integrating our deep learning approach in spectroscopy could enhance calibration accuracy and system reliability.

**6 Case Study: Application of Unsupervised Learning on ISIC HAM10000 Dataset**

We applied the DenseNet image encoding and KMeans clustering to the ISIC HAM10000 dataset, a collection of diverse skin lesion images. The DenseNet201 model, pre-trained on ImageNet, was fine-tuned using a subset of the HAM10000 dataset. We employed KMeans to cluster the HAM10000 test dataset, aiming to explore natural groupings within the skin lesion images without the guidance of pre-labeled categories. Clustering performance was evaluated using metrics like Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and purity. We ran the KMeans algorithm repeatedly, selecting the iteration with the highest purity score as the optimal clustering solution. This methodical approach refined our clustering process, ensuring a more reliable and accurate classification of the skin lesion images.

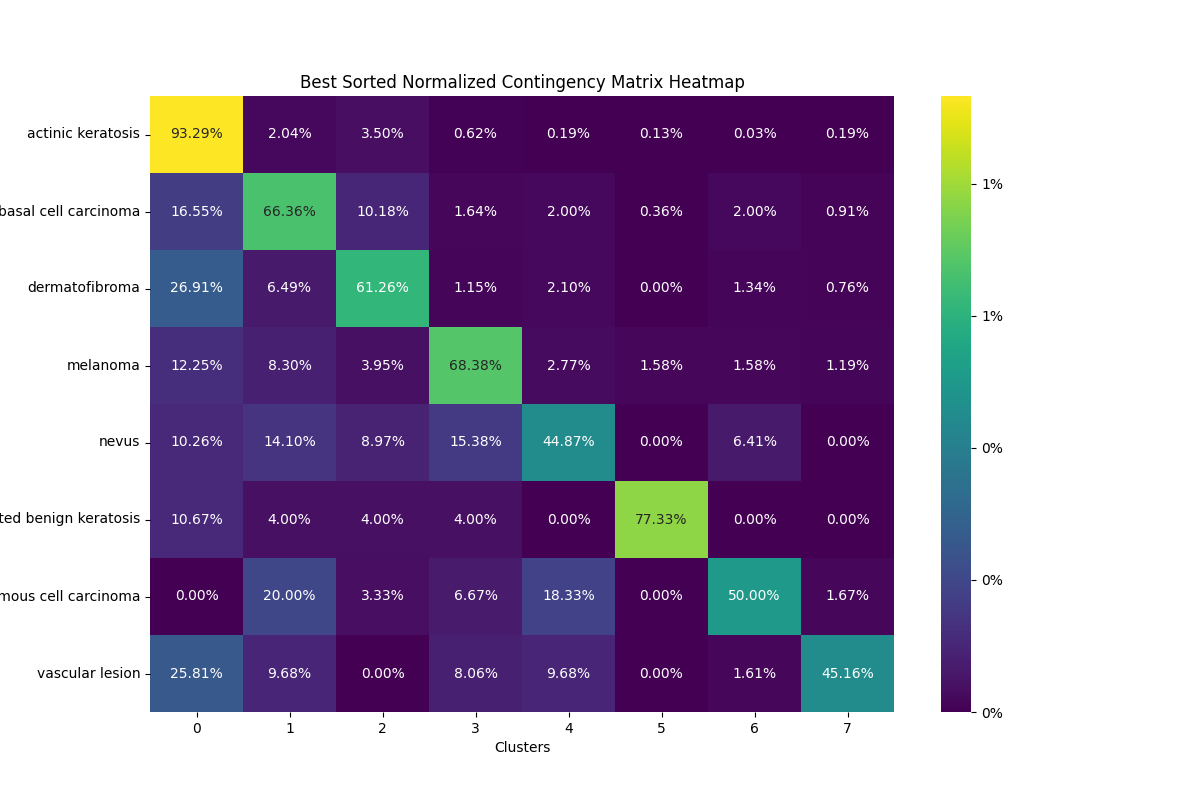
### **6.1 Results: Clustering Performance on ISIC HAM10000 Dataset**

Our clustering analysis of the ISIC HAM10000 dataset[34] using the KMeans algorithm yielded varied results across multiple iterations. The performance was evaluated based on metrics like Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and purity. The table below summarizes the key outcomes from the iterations with the best purity scores:

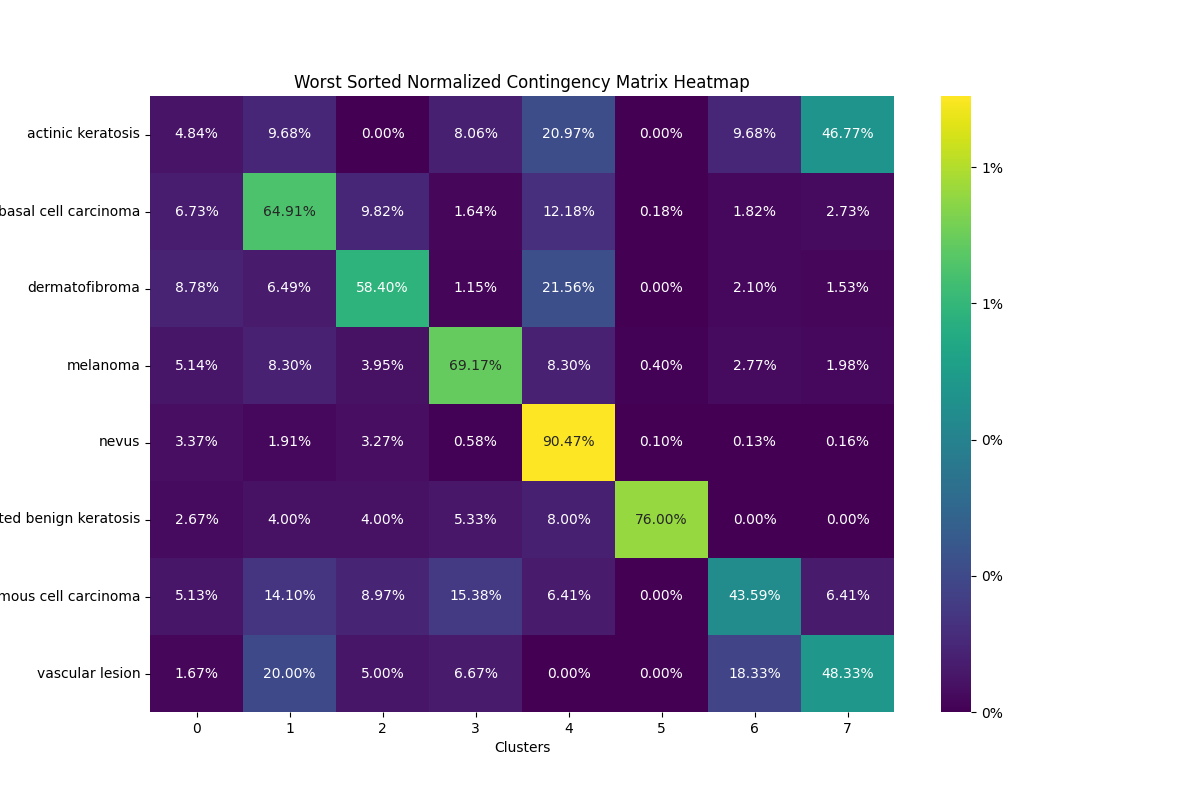
**Table 3** Evaluation analytics on the HAM10000 dataset with the diagnoses labels

| Iteration | label | Adjusted Rand Index | Normalized Mutual Information | Purity Score (axis=0) |
| --- | --- | --- | --- | --- |
| 1 | melanocytic  (True or False) | 0.631 | 0.459 | 0: 0.931343  1: 0.822532 |
| 2 | benign\_malignant  (benign or malignant) | 0.615 | 0.413 | 0: 0.949002  1: 0.765217 |
| 3 | diagnosis  ('nevus, squamous cell carcinoma, actinic keratosis, pigmented benign keratosis, dermatofibroma, vascular lesion, basal cell carcinoma, melanoma} | 0.670 | 0.462 | 0: 0.708738  1: 0.907057  2: 0.633136  3: 0.748918  4: 0.402299  5: 0.595745  6: 0.852941  7: 0.508475 |

The results for the labels with 2 classes were very consistent but the diagnosis label yielded varied results. Running it 100 times, the lowest purity sum (from the array in table) is 5.354. The lowest purity sum was 5.110.



**Figure 4** Heatmap of Contingency Matrix for the best purity score



**Figure 5** Heatmap of Contingency Matrix for the worst purity score

The KMeans clustering results on the ISIC HAM10000 dataset revealed notable pairings of diagnoses that frequently clustered together. This table summarizes these pairings within the 100 iterations using the diagnosis label, which occurred at least five times, indicating similarities in their image features:

**Table 5** Frequent pairings of diagnoses in KMeans Clustering

| **Paired Diagnoses** | **Frequency** |
| --- | --- |
| Basal Cell Carcinoma & Actinic Keratosis | 21 |
| Basal Cell Carcinoma & Vascular Lesion | 7 |
| Actinic Keratosis & Vascular Lesion | 15 |
| Actinic Keratosis & Squamous Cell Carcinoma | 8 |
| Actinic Keratosis & Dermatofibroma | 7 |

These findings are significant as they highlight the model's difficulty in distinguishing between these conditions. This insight is invaluable for refining the model and data curation, aiming for more accurate and distinct classification in dermatological imaging.

#### 6.2 Experimentation with an Alternate Training Dataset and Overfitting Issues

In pursuit of a robust model, we experimented by additionally training DenseNet with the ISIC challenge 2020 dataset, which contains 30k labeled and unlabeled images. Although this achieved high accuracy, it did not translate to effective clustering on the ISIC HAM10000 dataset, leading to the hypothesis of overfitting. Data augmentation techniques were employed to mitigate this, but they did not substantially improve the clustering results, underscoring the challenges in balancing model specificity with the ability to generalize across different datasets.

### **6.3 Discussion: Interpretation of Clustering Results**

The performance of the KMeans clustering on the ISIC HAM10000 dataset revealed insights that both align and diverge from our initial expectations. Notably, that the Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and purity scores, varied depending on the complexity of the label classes. The binary classifications were consistent and performed as well or better our expectations that KMeans would effectively distinguish between two distinct groups. But the variance in the purity scores across 100 iterations showed a less consistent outcome pointing out the difficulty in segmenting more nuanced, and possibly overlapping, characteristics. The frequent pairings of specific diagnoses probably should have been expected but analyzing the reasons would require some medical expertise in the field to determine which pairings make sense. From a ML view, these are intriguing as possible limitations of the feature extraction process and the source of future exploration. Our discussion concludes that while KMeans shows promise for the classification of dermatological images, particularly for distinct binary cases, its application in multiclass problems requires careful and cautious interpretation and the problem may benefit from using other clustering techniques and feature selection methods.

**7 Conclusion**

Various approaches of unsupervised machine learning models for image recognition and benchmarks of the results on the CIFAR-10 dataset were investigated in this study. The state-of-the-art pre-trained models’ capability was also tested on the unseen dataset by plotting their knowledge of the 2D manifold via the t-SNE algorithm. This paper also demonstrated the power of DenseNet as a feature extractor when compared to other unsupervised models such as SIFT with HOG, convolutional autoencoder, and ResNets. It was concluded that unsupervised image classification is a well-researched and practiced field in deep learning that can solve the problem of obtaining labeled data by introducing the pre-trained models with clustering techniques such as the DenseNets. Consequentially, decent results are reported on the CIFAR-10 dataset using this approach, whereby the highest achieved accuracy is 68.14%.

**8 Conflicts of Interest**

The author(s) declare that they have no conflicts of interest to report regarding the present study.

**9 Acknowledgements**

We would like to express our gratitude to the authors of the referenced papers for their valuable contributions to the field of unsupervised learning and image classification. Their pioneering work and insights have greatly influenced our research.

**10 REFERENCES**

1. Currie, G., Hawk, K. E., Rohren, E., Vial, A., & Klein, R. (2019). Machine learning and deep learning in medical imaging: intelligent imaging. Journal of medical imaging and radiation sciences, 50(4), 477-487. https://www.jmirs.org/article/S1939-8654(19)30504-1/fulltext
2. Cai, L., Gao, J., & Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation. Annals of translational medicine, 8(11). https://atm.amegroups.org/article/view/36944/html
3. Hussein, S., Kandel, P., Bolan, C. W., Wallace, M. B., & Bagci, U. (2019). Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. IEEE Transactions on Medical Imaging, 38(8), 1777-1787. https://ieeexplore.ieee.org/document/8624570
4. Verplancke, T., Van Looy, S., Benoit, D., Vansteelandt, S., Peleman, R. A., et al. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with hematological malignancies. BMC medical informatics and decision making, 8(1), 1-8. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-8-56>
5. Lamb, D. S., Downs, J., & Reader, S. (2020). Space-time hierarchical clustering for identifying clusters in spatiotemporal point data. ISPRS International Journal of Geo-Information, 9(2), 85. https://www.mdpi.com/2220-9964/9/2/85
6. Latif, J., Xiao, C., Imran, A., & Tu, S. (2019, November). Medical imaging using machine learning and deep learning algorithms: a review. In 2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET), Sukkur, Pakistan (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/document/8673502>
7. Lu, X., Tsao, Y., Matsuda, S., & Hori, C. (2013, September). Speech enhancement based on deep denoising autoencoder. In Interspeech (pp. 436-440). <https://arxiv.org/abs/2001.01538>
8. Schmarje, L., Santarossa, M., Schröder, S.-M., & Koch, R. (2021). A survey on semi-, self-and unsupervised learning for image classification. IEEE Access, 9, 82146-82168. https://arxiv.org/abs/2002.08721
9. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). <https://arxiv.org/abs/1608.06993>
10. He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
11. Qazi, H. A., Jahangir, U., Yousuf, B. M., & Noor, A. (2017, November). Human action recognition using SIFT and HOG method. In 2017 International Conference on Information and Communication Technologies (ICICT) (pp. 6-10). IEEE. <https://ieeexplore.ieee.org/document/8320156>
12. Calik, R. C., & Demirci, M. F. (2018, October). Cifar-10 image classification with convolutional neural networks for embedded systems. In 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-2). IEEE. https://ieeexplore.ieee.org/document/8612873
13. Premachandran, V., & Yuille, A. L. (2016). Unsupervised learning using generative adversarial training and clustering.
14. Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., et al. (2018). High-resolution image synthesis and semantic manipulation with conditional Gans. arXiv preprint arXiv:1711.11585. <https://arxiv.org/abs/1711.11585>
15. Zhu, X., & Goldberg, A. B. (2019). Introduction to semi-supervised learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1-130. https://link.springer.com/book/10.1007/978-3-031-01548-9
16. Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps, 2018, 323-350. <https://arxiv.org/abs/1704.06825>
17. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., et al. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88. <https://www.sciencedirect.com/science/article/abs/pii/S1361841517301135?via%3Dihub>
18. Liu, R., & Gillies, D. F. (2016). Overfitting in linear feature extraction for classification of high-dimensional image data. Pattern Recognition, 53, 73-86. <https://www.sciencedirect.com/science/article/abs/pii/S0031320315004355?via%3Dihub>
19. Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., et al. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence, 3(3), 199-217. https://arxiv.org/abs/2008.06388
20. Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A. E., et al.. (2018). Current applications and future impact of machine learning in radiology. Radiology, 288(2), 318. <https://pubs.rsna.org/doi/10.1148/radiol.2018171820>
21. Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461. <https://www.sciencedirect.com/science/article/abs/pii/S0031320302000602?via%3Dihub>
22. Hou, X., Shen, L., Sun, K., & Qiu, G. (2017, March). Deep feature consistent variational autoencoder. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1133-1141). IEEE. <https://arxiv.org/abs/1610.00291>
23. Chan, D. M., Rao, R., Huang, F., & Canny, J. F. (2018). t-SNE-CUDA: GPU-accelerated t-SNE and its applications to modern data. In 2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD) (pp. 330-338). IEEE. <https://arxiv.org/abs/1807.11824>
24. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. https://arxiv.org/abs/1511.06434
25. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., et al. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377. <https://www.sciencedirect.com/science/article/abs/pii/S0031320317304120?via%3Dihub>
26. Grinberg, L., Hopfield, J., & Krotov, D. (2019). Local unsupervised learning for image analysis. arXiv preprint arXiv:1908.08993. https://arxiv.org/abs/1908.08993
27. Abouelnaga, Y., Ali, O. S., Rady, H., & Moustafa, M. (2016, December). Cifar-10: Knn-based ensemble of classifiers. In 2016 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1192-1195). IEEE. <https://arxiv.org/abs/1611.04905>
28. blank
29. Caron, M., Bojanowski, P., Joulin, A., & Douze, M. (2019). Deep clustering for unsupervised learning of visual features. Facebook AI Research. <https://arxiv.org/abs/1807.05520>
30. Jo, Anderson. "Dense Convolution Network." Incredible.AI, http://incredible.ai/deep-learning/2017/07/29/Dense-Convolution-Network/. Accessed 09/2023.
31. Nie, D., Gao, Y., Wang, L., & Shen, D. (2018). ASDNet: Attention based semi-supervised deep networks for medical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 370-378). Springer, Cham. <https://link.springer.com/chapter/10.1007/978-3-030-00937-3_43>
32. Masud, U., Amirzada, M. R., Elahi, H., Akram, F., Zeeshan, A., et al. (2022). Design of two-mode spectroscopic sensor for biomedical applications: Analysis and measurement of relative intensity noise through control mechanism. Applied Sciences, 12(4), 1856. https://www.mdpi.com/2076-3417/12/4/1856
33. Masud, U., Jeribi, F., Alhameed, M., Akram, F., Tahir, A., et al. (2022). Two-mode biomedical sensor build-up: Characterization of optical amplifier. Construction and Building Materials, 70, 444961. <https://www.techscience.com/cmc/v70n3/44961>
34. ViDIR Group, Department of Dermatology, Medical University of Vienna. (Year). HAM10000 [Dataset]. <https://api.isic-archive.com/collections/212/>
35. Zhang, Y. (2018). A better autoencoder for image: Convolutional autoencoder. ICONIP17-DCEC. https://users.cecs.anu.edu.au/~Tom.Gedeon/conf/ABCs2018/paper/ABCs2018\_paper\_58.pdf
36. Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., et al. (2019). Medical image computing and computer-assisted intervention-MICCAI 2019 (Vol. 11765). Springer Nature.

# 10 APPENDIX

List of Acronyms:

CIFAR-10: A dataset introduced by the Canadian Institute For Advanced Research consisting of 60,000 32x32 color images in 10 classes

CNN: Convolutional Neural Network,

HOG: Histogram of Oriented Gradients,

SIFT: Scale-Invariant Feature Transform

SSL : Semi-Supervised Learning

SSE : Sum of the Squared Errors

GAN-Generative Adversarial Network

MRI- Magnetic Resonance Imaging

FBG- Fibre Bragg Grating

t-SNE- t-distributed stochastic neighbor embedding

ISIC: The International Skin Imaging Collaboration​