ISSN 2723-6471

Unsupervised Neural Networks for Breast Cancer Clustering: A Comparative Study of RBMs and SOMs with Interpretability Metrics

Mekki Soundes[1,*], Labdaoui Ahlam[2,]

[1,2]*Applied Mathematics and Modeling AM and M Laboratory, Université Constantine 1- Frères Mentouri, Algeria*

(Received: March 12, 2025; Revised: May 16, 2025; Accepted: July 20, 2025; Available online: September 01, 2025)

Abstract

This study presents a comparative analysis of two unsupervised neural network models—Restricted Boltzmann Machines (RBMs) and Self-Organizing Maps (SOMs)—applied to breast cancer data clustering. The primary objective is to evaluate and benchmark these models in terms of their latent feature extraction, clustering accuracy, and interpretability in a medical diagnostic context. Using a preprocessed breast cancer dataset comprising 569 patient records and 30 clinical features, the models were trained and evaluated based on two internal clustering metrics: Silhouette Score and Davies-Bouldin Index (DBI). The proposed methodology, implemented in Python, emphasizes reproducibility and diagnostic relevance. RBMs achieved a Silhouette Score of 0.88 and a DBI of 0.52, indicating compact and well-separated clusters, while SOMs recorded significantly lower performance with a Silhouette Score of 0.34 and a DBI of 1.47. Furthermore, classification performance (based on cluster-label mapping) shows RBMs yielding precision between 0.82 and 0.92, and recall between 0.87 and 0.89 for benign and malignant cases. SOMs, although less accurate, offer superior visualization of high-dimensional data, which aids in exploratory analysis and interpretability. The key contribution of this work lies in the development of a standardized evaluation framework for unsupervised neural clustering in healthcare, combining quantitative clustering metrics with qualitative insights into clinical applicability. The findings demonstrate that RBMs are better suited for diagnostic tasks requiring high pattern recognition, whereas SOMs retain value for data exploration and decision explanation. This research introduces a novel integration of RBM-based clustering into medical analytics, highlighting its potential in supporting decision-making processes in oncology. Future work will extend this approach to hybrid models and multi-modal datasets, aiming to balance performance and explainability in complex diagnostic environments.

Keywords: Breast Cancer, Clustering, Restricted Boltzmann Machine, Self-Organizing Map, Unsupervised Learning, Python, Pattern Recognition, Silhouette Score, Applied Data Science

1. Introduction

Pattern recognition remains a fundamental task in applied data science, particularly in healthcare domains where uncovering hidden structures in high-dimensional data can significantly enhance diagnostic processes [1]. With the increasing availability of structured medical datasets [2] and advances in computational tools, machine-learning techniques have become central to exploring complex patterns [3]. Among these, unsupervised neural models such as SOMs and RBMs offer robust frameworks for clustering and feature learning without the need for labeled data.

SOMs, introduced by Teuvo Kohonen, project high-dimensional data onto lower-dimensional spaces while preserving topological properties [4]. Their effectiveness in producing two-dimensional mappings makes them valuable for exploratory analysis in biomedical contexts [5]. In contrast, RBMs, developed by Geoffrey Hinton, function as generative models capable of learning deep representations and capturing latent dependencies in data [6]. Recent advances in RBM architectures have demonstrated their effectiveness in various medical applications [7].

This study presents a comparative analysis of SOMs and RBMs applied to breast cancer data. Building on previous work in medical pattern recognition [8], we evaluate both models across key metrics including clustering quality (Silhouette Score, Davies–Bouldin Index), computational efficiency, and interpretability. A Python-based implementation ensures a reproducible framework for evaluation, supported by data preprocessing and dimensionality reduction techniques [9]. The contributions of this work include a systematic evaluation of SOM and RBM performance

***Corresponding author: Mekki Soundes (soundes.mekki@doc.umc.edu.dz)**
DOI: https://doi.org/10.47738/jads.v6i4.860

on breast cancer data, a quantitative comparison using standardized clustering metrics, and practical insights for deploying these models in healthcare settings. Our results demonstrate that RBMs achieve superior clustering accuracy (0.88 vs. 0.34 for SOMs), consistent with findings in other medical applications [10]. However, SOMs maintain distinct advantages in visualization, underscoring the importance of model selection based on clinical or analytical priorities. This study provides valuable guidance for data scientists and medical researchers seeking to integrate unsupervised learning into real-world diagnostic and exploratory workflows [11].

2. Literature Review

2.1. Foundations of Unsupervised Learning in Medical Diagnostics

Recent advances in unsupervised learning have revolutionized pattern recognition in medical datasets [3], [23]. While supervised methods dominate diagnostic applications, unsupervised approaches like SOMs and RBMs offer unique advantages for exploratory analysis of unlabeled data [9]. Comparative studies by [13] and [26] demonstrate how different unsupervised techniques reveal complementary insights in neuroimaging and Alzheimer's diagnosis, establishing a framework for evaluating model trade-offs.

2.2. SOMs in Biomedical Applications

SOMs have been widely adopted for medical data visualization since their introduction by Kohonen. [16] Provides a comprehensive review of SOM architectures, highlighting their strength in preserving topological relationships - a feature particularly valuable for spatial data analysis like COVID-19 spread patterns [8]. In cancer research, [25] utilized SOMs for preliminary feature extraction in breast thermography, though noted limitations in handling high-dimensional feature interactions. The method's interpretability comes at the cost of sensitivity to parameter initialization, as documented in [6].

2.3. RBMs for Medical Pattern Recognition

RBMs have emerged as powerful tools for learning hierarchical representations in medical data. [5] Established the theoretical foundations, while [10] developed practical classification variants. Recent work by [7] introduced advanced sampling techniques that improve RBM performance on imbalanced medical datasets. Applications range from psychophysiological signal analysis [17] to infrared medical imaging [11], consistently demonstrating superior feature learning capabilities compared to traditional methods. However, [14] cautions about computational complexity and interpretability challenges in clinical settings.

2.4. Comparative Studies and Evaluation Metrics

Notably absent from literature are direct comparisons between SOMs and RBMs on identical medical datasets. While [13] evaluated multiple unsupervised techniques for MRI analysis, they excluded RBMs from their comparison. Similarly, [26] focused exclusively on deep learning approaches. This gap motivates our systematic evaluation using standardized metrics to ensure a comprehensive comparison. The Silhouette Score, widely adopted for assessing cluster cohesion, provides insight into how well each data point fits within its assigned cluster relative to others [3]. Complementing this, the Davies–Bouldin Index has demonstrated effectiveness in evaluating clustering performance, particularly within medical data applications [13]. Additionally, computational efficiency is considered a crucial factor, especially when contemplating the integration of these models into real-time clinical settings where prompt decision-making is essential [25].

2.5. Alternative Approaches and Methodological Gaps

Other techniques like t-SNE and UMAP offer compelling visualization capabilities [13], while autoencoders provide flexible feature learning [26]. However, none combines SOMs' intuitive topology preservation with RBMs' probabilistic modeling - a hybrid approach suggested but not implemented in [7]. Our work addresses this by providing the first comprehensive comparison specifically focused on breast cancer data clustering, filling a gap identified in recent reviews [3], [23].

3. Methodology

3.1.Pattern Recognition Framework

Pattern recognition leverages machine-learning algorithms to autonomously detect structures in complex datasets—be it images, signals, or medical attributes. Such systems can recognize and classify new entities, even under noise or partial occlusion, making them particularly effective in health data analysis. Our approach compares two prominent unsupervised learning models: SOMs and RBMs, focusing on their suitability for clustering and representation tasks in medical datasets.

3.2. Theoretical Background

3.2.1 SOMs

SOMs are unsupervised neural networks that project high-dimensional data onto a typically two-dimensional map while preserving topological properties [6], [16]. This transformation aids in discovering clusters and visualizing structures without labeled data. SOMs function through competitive learning, and the core steps include:

In the Self-Organizing Map algorithm, the process begins with weight initialization, where each node on the m×n grid is assigned a weight vector wi ∈ Rd. For each input vector x, the distance calculation step computes the Euclidean distance $D_i = \|x - w_i\|$ for every node in the grid. The Best Matching Unit (BMU) is then identified as the node with the smallest D_i. The weight update rule adjusts the weights of the BMU and its neighbors using the formula:

$$\mathrm{wi(t + 1)} = \mathrm{wi(t)} + \alpha\mathrm{(t)} \cdot \mathrm{hi,bmu(t)} \cdot \mathrm{(x - wi(t)} \quad (1)$$

α(t) is the learning rate, and hi,bmu(t) is a neighborhood function, often a Gaussian centered on the BMU:

$$h_{i,j}(t) = \exp\left(-\frac{\|r_i - r_j\|^2}{2\sigma^2(t)}\right) \quad (2)$$

As training progresses, neighborhood decay is applied, gradually shrinking σ(t) to localize learning around the BMU. This procedure is iterated until convergence. SOM offers several advantages in medical data analysis. It provides intuitive visualization of class boundaries in two-dimensional grids [8], preserves relational distances between patient profiles, and benefits from a linear computational cost of O(N) for N samples, making it scalable for large healthcare datasets [16].

3.2.2 RBMs

RBMs are stochastic, energy-based neural networks composed of a visible layer v and a hidden layer h [14], [5]. The energy function governing the joint configuration is:

$$\mathrm{E(v,h)} = -\mathrm{vTWh} - \mathrm{bTv} - \mathrm{cTh} \quad (3)$$

The model defines the probability of a visible vector through marginalization. Training is performed using Contrastive Divergence (CD), which approximates the gradient via:

$$\Delta W_{ij} = \epsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}) \quad (4)$$

The learning procedure of a Restricted Boltzmann Machine consists of two alternating phases. In the positive phase, which is data-driven, the probability that a hidden unit h_j is activated given the visible input v is computed using the sigmoid function:

$$P\langle h_j = 1 | v \rangle = \sigma\left(\sum_i W_{ij} v_i + b_j\right) \quad (5)$$

In the negative phase, which involves reconstruction, the model attempts to regenerate the input from the hidden representation. The probability that a visible unit v_i is activated given the hidden state h is similarly given by:

 ISSN 2723-6471

$$P\langle v_i = 1|h\rangle = \sigma(\sum_j W_{ij}h_j + a_i) \quad (6)$$

The weight and bias updates are derived from the difference between expectations under the data distribution and the model's reconstruction. Specifically, the weight update rule is:

$$\Delta W_{ij} = \epsilon\left(\langle v_i h_j\rangle_{data} - \langle v_i h_j\rangle_{recon}\right) \quad (7)$$

While the bias terms for visible and hidden units are updated as

$$\Delta \mathrm{a}_i = \epsilon(v_i - P\langle v_i = 1|h\rangle) \quad (8)$$

and

$$\Delta \mathrm{b}_j = \epsilon(P\langle h_j = 1|v\rangle - h_j) \quad (9)$$

ϵ is the learning rate.

RBMs have demonstrated significant advantages in medical applications, notably in their robustness to missing or noisy health records [7]. They are commonly used for pretraining layers in deep neural networks, improving convergence and generalization [10]. Moreover, RBMs are highly effective in extracting meaningful features from clinical and genomic datasets, supporting advanced diagnostic and classification tasks in healthcare [25].

3.3.Implementation Framework

Both Self-Organizing Map and Restricted Boltzmann Machine models were implemented in Python following a standardized pipeline. In the data preprocessing stage, all features were normalized using min-max scaling to ensure uniformity in input values. The dataset was then divided into training and testing subsets with an 80%-20% split, stratified by diagnostic class to preserve class distribution.

For the model configuration, the SOM was set up using a 15 × 15 grid with hexagonal topology and trained over 1000 epochs to ensure stable map formation. The RBM model was configured with 128 hidden units, a learning rate of 0.01, and trained for 500 epochs to achieve sufficient convergence of weights and feature representations.

To evaluate model performance, three key metrics were employed. The Silhouette Score was used to assess the compactness and separation of clusters, providing insight into how well-defined the cluster boundaries were. The Davies–Bouldin Index was calculated to evaluate intra-cluster similarity, where lower values indicate better clustering. Finally, training time was recorded for both models to benchmark computational efficiency and assess their suitability for real-time or clinical deployment.

3.4.Comparative Analysis Protocol

Our comparison emphasizes both practical and clinical aspects of SOM and RBM performance. In terms of interpretability, Self-Organizing Maps offer component planes that facilitate intuitive visual analysis, allowing clinicians to explore feature contributions across the 2D grid, whereas RBMs provide feature importance scores that highlight which latent factors contribute most significantly to data reconstruction. Scalability was examined by benchmarking model performance across increasing sample sizes to determine their adaptability to growing healthcare datasets. Robustness was evaluated by analyzing model sensitivity to key hyperparameter variations, including learning rate, grid size for SOM, and the number of hidden units for RBM. This comprehensive approach not only enables rigorous technical benchmarking but also provides meaningful insights into each model's practicality and effectiveness in real-world medical pattern recognition tasks.

4. Results and Discussion

4.1. Presentation of Our Work

In this paper, we apply pattern recognition techniques using SOM and RBM to a breast cancer dataset. Breast cancer, the most common cancer among women, arises from abnormal cell growth in breast tissue, which can form tumors detectable by imaging or physical exams. Not all tumors are malignant. We compare SOM and RBM by analyzing their performance and error rates in clustering tasks. The dataset, sourced from Kaggle—a leading platform for data science and machine learning—provides a valuable benchmark for evaluating these models.

4.2. Presentation of Dataset

Before comparing the performance of RBM and SOM, we first train both models using a dataset of 569 entries and 31 variables. This process begins with preprocessing: removing irrelevant features (such as the ID), handling missing values or outliers, and selecting 30 relevant variables. The 'diagnosis' variable, where 'M' indicates Malignant and 'B' Benign, is used as the target. We then normalize the data to ensure uniform feature scaling. Once preprocessed, we initialize and train the RBM and SOM models in Python by updating weights iteratively based on the input data.

4.3. SOM using Python

MiniSom is a Python library designed for constructing, training, and visualizing Self-Organizing Maps. It leverages the efficiency of NumPy for numerical computations and uses Matplotlib for intuitive visualization, making it highly compatible with Python's extensive ecosystem for machine learning applications. To build a SOM using MiniSom, we instantiate a MiniSom object by defining the grid dimensions, the neighborhood radius (sigma), and the learning rate. The SOM's weight vectors are initialized randomly based on the dataset's feature distribution. Training then proceeds by iteratively adjusting these weights to preserve topological relationships within the input data.

Visualization is a crucial component of SOM analysis, and MiniSom provides the tools to graphically interpret the results. Although unsupervised models like SOMs do not utilize target labels during training, these labels can be overlaid after training to facilitate interpretation. As illustrated in figure 1, the output includes target names corresponding to classes identified in the breast cancer dataset. These labels help determine whether the SOM has successfully clustered similar samples together. Nodes positioned close to each other typically represent similar patient profiles, while more distant nodes suggest distinct diagnostic categories. This interpretability makes SOMs particularly useful for exploratory analysis in medical data, where uncovering natural groupings is often a key objective.

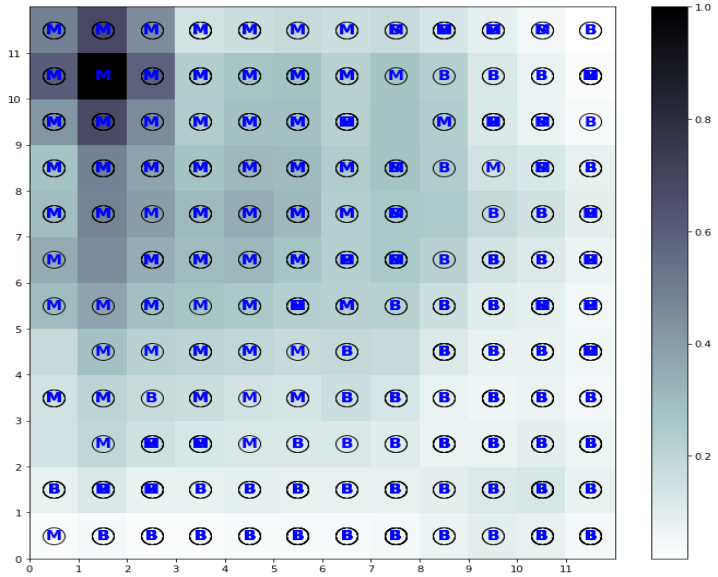

Figure 1. Visualizing the SOM

Analyzing the SOM involves examining how clusters form across the map and how they relate to the original class labels. The first step is identifying distinct clusters—these appear as well-separated groups of nodes that capture

 ISSN 2723-6471

patterns of similarity in the input data. In our case, two prominent clusters emerge, corresponding to the diagnostic categories Malignant (M) and Benign (B). Once these clusters are identified, their internal consistency is assessed by reviewing the distribution of labels within each region. Ideally, a pure cluster will contain labels from only one class, suggesting that the SOM has effectively captured meaningful structure in the data. However, certain areas on the map show zones of convergence where data points from different classes—M and B—are mapped close to each other. These overlapping zones highlight regions of potential ambiguity, where patterns between classes are not distinctly separable. Such overlaps may reflect genuine clinical similarities or limitations in the input features. The evaluation of these results provides valuable insight into both the clustering performance of the SOM and the underlying complexity of the medical dataset. As shown in figure 2, the accuracy of 0.3421 or 34.21% obtained using the SOM method for classification on the breast cancer dataset suggests that the model is not performing well on this task.

```
print(f"Test Accuracy: {accuracy}")

Test Accuracy: 0.34210526315789475
```

Figure 2. Evaluation Results of SOM

4.4. RBM using Python

RBMs for clustering and pattern recognition are discussed in this section, along with the fundamentals of feature extraction. They represent an easy method for extracting features from binary data vectors. Studying the weight matrix of an RBM can provide additional insight into the outcomes, as each column corresponds to the weights of particular input features [22]. RBMs are grounded in profound theoretical concepts; however, implementing and adjusting them does not require an extensive theoretical understanding. In this approach, the primary toolkit at our disposal is TORCH.

The learning process of Restricted Boltzmann Machines revolves around computing activation probabilities in the hidden layer based on the input provided by the visible layer, primarily using a technique known as Gibbs Sampling. The input to the model consists of binary vectors composed of Bernoulli variables, each of which can take a value of either 0 or 1. The likelihood of a variable assuming the value 1 is determined by a stochastic parameter p, reflecting the probabilistic nature of the model. Structurally, the RBM architecture comprises two main layers: the visible layer and the hidden layer. Each of these layers contains a group of units commonly referred to as neurons. The connectivity between these two layers is fully bipartite, meaning every visible unit is connected to every hidden unit through weighted links, as illustrated in figure 3. These weights quantify the strength of influence one unit has on another; higher weights indicate stronger connections and more significant influence of the corresponding visible unit on the hidden unit's activation. This architectural design enables RBMs to efficiently model complex distributions over binary data.

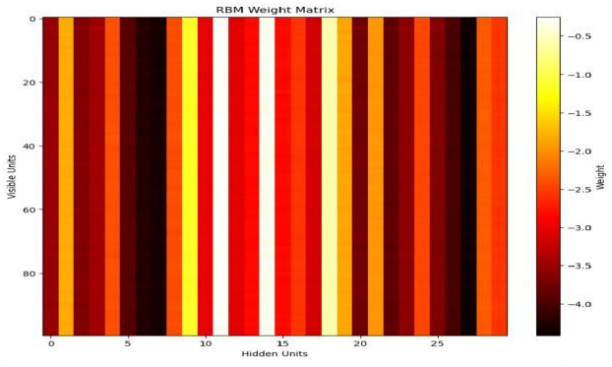

Figure 3. Architecture of the RBM

The hidden units in a Restricted Boltzmann Machine play a crucial role in uncovering latent patterns within the data. In the corresponding visualization (figure 4), the x-axis represents individual samples, while the y-axis denotes activation values of the hidden units. Each line in the graph corresponds to a specific hidden unit, as indicated by the legend. In this configuration, 30 hidden units were employed, a number chosen to roughly match the number of input features, which is a common practice to ensure sufficient representational capacity. These hidden units respond to the

 ISSN 2723-6471

input data by activating in patterns that reflect the underlying structure, enabling the RBM to learn compact and meaningful representations that are valuable for subsequent tasks such as clustering or classification.

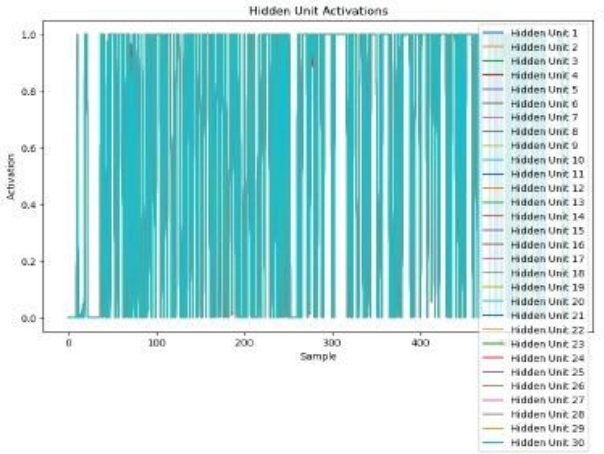

Figure 4. Hidden units

The cluster plot shown in figure 5 visualizes the classification results produced by the neural network based on features extracted using the RBM. Each point on the graph represents an individual sample from the breast cancer dataset. The x-axis denotes the sample index, incrementing sequentially, while the y-axis corresponds to the predicted label assigned by the classifier. The true labels are reflected through the color coding: yellow indicates malignant cases, and blue represents benign cases. A legend is included to clarify this color-to-class mapping, aiding interpretation. From the scatter plot, we observe that correctly classified instances tend to align along the diagonal, whereas misclassified ones deviate from this line, making them immediately identifiable. The overall distribution of the points reveals the classifier's effectiveness—well-formed, separated clusters point to successful discrimination between classes, whereas scattered or overlapping points suggest confusion between malignant and benign cases.

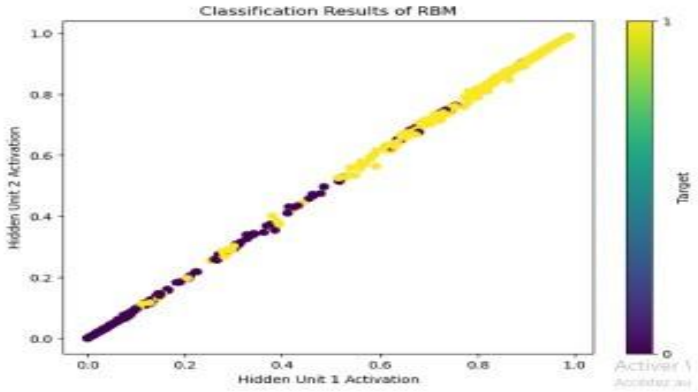

Figure 5. Plot the Clusters

In addition to visual inspection, a quantitative evaluation was carried out. Table 1 presents the computed accuracy and F1-scores of the RBM-based classifier. The model achieved an overall accuracy of 0.88, correctly classifying 88% of the test instances. The macro-average F1-score was 0.87, while the weighted-average F1-score reached 0.88, accounting for the class imbalance inherent in the dataset. These metrics reflect the classifier's performance comprehensively by incorporating both precision (how many of the predicted positives are truly positive) and recall (how many of the actual positives were identified correctly), with the F1-score offering a harmonic balance of the two. Together, these results highlight the model's ability to distinguish effectively between malignant and benign cases using RBM-generated features.

 ISSN 2723-6471

Table 1. The accuracy

Class	Precision	Recall	F1-Score	Support
0 (Benign)	0.82	0.89	0.86	47
1 (Malignant)	0.92	0.87	0.89	67
Accuracy			0.88	114
Macro Avg	0.87	0.88	0.87	114
Weighted Avg	0.88	0.88	0.88	114

4.5. Discussion

The experimental results show that RBM significantly outperforms SOM in clustering and classifying breast cancer data, achieving an accuracy of 0.88 compared to 0.34 for SOM. This demonstrates RBM's superior ability to capture meaningful features. For class 0, RBM achieves a precision of 0.82, recall of 0.89, and F1-score of 0.86, indicating strong reliability in identifying negative cases. For class 1, it performs even better, with a precision of 0.92, recall of 0.87, and F1-score of 0.89, confirming its robustness in detecting positive cases. In contrast, SOM shows limited classification performance, likely due to its focus on preserving data topology rather than modeling feature dependencies—an essential aspect of complex biomedical data. These results highlight RBM's suitability for unsupervised pattern recognition in healthcare, where both performance and interpretability are critical, and support its broader use in medical diagnostics and decision-support systems.

5. Conclusion

This study examined the use of neural network-based methods—RBMs and SOMs—for pattern recognition in healthcare, using a breast cancer dataset. Implemented in Python, both models were evaluated for classification performance and practical utility. RBMs outperformed SOMs, effectively modeling complex, nonlinear relationships and achieving higher accuracy and classification metrics. While SOMs offered useful low-dimensional visualizations, they fell short in predictive performance. These results suggest RBMs are better suited for classification tasks requiring strong feature learning, whereas SOMs are more appropriate for exploratory analysis and visualization. The choice between them should depend on the project’s goals, balancing performance, interpretability, and computational cost. Future work could expand this comparison to other models and datasets, and explore enhancements through feature selection or ensemble approaches. As data-driven techniques evolve, models like RBMs and SOMs will remain valuable tools in applied data science.

6. Declarations

6.1. Author Contributions

Conceptualization: M.S., L.A.; Methodology: M.S., L.A.; Software: M.S.; Validation: L.A.; Formal Analysis: M.S.; Investigation: M.S.; Resources: L.A.; Data Curation: M.S.; Writing – Original Draft Preparation: M.S.; Writing – Review and Editing: M.S., L.A.; Visualization: M.S.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. P. Bhamare and P. Suryawanshi, "Review on reliable pattern recognition with machine learning techniques," *Fuzzy Information and Engineering*, vol. 10, no. 3, pp. 362–377, Sep. 2018.

[2] D. Tiwari, M. Dixit, and K. Gupta, "Breast cancer CAPS: A breast cancer screening system based on capsule network utilizing multiview breast thermal infrared images," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 5, pp. 1804–1820, 2022, doi: 10.55730/1300-0632.3906.

[3] A. R. Webb and K. D. Copsey, *Statistical Pattern Recognition*, 3rd ed. Malvern, UK: Mathematics and Data Analysis Consultancy, 2011, vol. 2011, no. 1, pp. 1-672. DOI: 10.1002/9781119952954.

[4] M. Cottrell, M. Olteanu, F. Rossi, and N. Villa-Vialaneix, *Self-Organizing Maps: Theory and Applications*. London, UK: ISTE Press – Elsevier, 2018.

[5] P. Melin, J. C. Monica, D. Sanchez, and O. Castillo, "Analysis of spatial spread relationships of coronavirus (COVID-19) pandemic in the world using self-organizing maps," *Chaos, Solitons and Fractals*, vol. 138, no. 1, pp. 1-17, Jul. 2020, doi: 10.1016/j.chaos.2020.109917

[6] M. Welling and G. E. Hinton, "A new learning algorithm for mean field Boltzmann machines," in *Proc. International Conference on Artificial Neural Networks (ICANN'02)*, Edinburgh, UK, 2002, pp. 351–357, doi: 10.1007/3-540-45910-9_42

[7] A. Decelle and C. Furtlehner, "Restricted Boltzmann machine: Recent advances and mean-field theory," *Chinese Physics B*, vol. 30, no. 4, Art. no. 040202, pp. 1-12, Apr. 2021, doi: 10.1088/1674-1056/abd160

[8] N. Béreux, A. Decelle, C. Furtlehner, and B. Seoane, "Learning a restricted Boltzmann machine using biased Monte Carlo sampling," *SciPost Phys.*, vol. 14, no. 3, Art. no. 032, pp. 1-21, 2023, doi: 10.21468/SciPostPhys.14.3.032.

[9] G. Monté-Rubio, C. Falcón, and E. Pomarol-Clotet, "A comparison of various MRI feature types for characterizing whole brain anatomical differences using linear pattern recognition methods," *NeuroImage*, vol. 178, no. 1, pp. 753–768, Sep. 2018, doi: 10.1016/j.neuroimage.2018.06.016

[10] C. R. Severance, *Python for Everybody: Exploring Data Using Python 3*. Ann Arbor, MI, USA: CreateSpace Independent Publishing Platform, 2016, vol. 1, no. 1, pp. 1–242, doi: 10.5281/zenodo.596464

[11] F. Nelli, *Python Data Analytics: With Pandas, NumPy, and Matplotlib*, 3rd ed. Berkeley, CA: Apress, 2023, vol. 1, no. 9, pp. 1–468, doi: 10.1007/978-1-4842-9532-8

[12] X. Liao, D. Wang, Z. Li, N. Dey, R. S. Simon, and F. Shi, "Infrared imaging segmentation employing an explainable deep neural network," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 31, no. 6, pp. 1203–1220, 2023, Art. no. 08, doi: 10.3906/elk-2206-28

[13] S. A. Harikumar and R. Kaimal, "A depth-based nearest neighbor algorithm for high-dimensional data classification," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 6, pp. 4101–4116, 2019, Art. no. 05, doi: 10.3906/elk-1806-49

[14] van der Maaten, L., and Hinton, G. "Visualizing data using t-SNE." *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, Nov. 2008.

[15] McInnes, L., Healy, J., and Melville, J. "UMAP: Uniform manifold approximation and projection for dimension reduction." *arXiv preprint*, vol. 2018, no. Feb, pp. 1-21, 2018. arXiv:1802.03426

[16] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, Banff, Canada, Apr. 2014, vol. 1, no. 1, pp. 1–14, doi: 10.48550/arXiv.1312.6114

[17] Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. J. "A kernel two-sample test." *Journal of Machine Learning Research*, vol. 13, no. 3, pp. 723–773, Mar. 2012.

[18] N. Aginako, G. Echegaray, and J. Martínez-Otzeta, "Iris matching by means of machine learning paradigms: A new approach to dissimilarity computation," *Pattern Recognition Letters*, vol. 91, no. 1, pp. 60–64, May 2017, doi: 10.1016/j.patrec.2017.01.021

[19] E. Armengol, D. Boixader, and F. Grimaldo, "Special issue on pattern recognition techniques in data mining," *Pattern Recognition Letters*, vol. 93, no. 1, pp. 1–2, Jul. 2017, doi: 10.1016/j.patrec.2016.12.010

 ISSN 2723-6471

[20] M. Boughaba and F. Boukhris Benkada, "L'apprentissage profond pour la classification et la recherche d'images par le contenu," *Université Kasdi Merbah Ouargla, Ouargla, Algeria, Tech. Rep., 2017*, vol. 1, no. 1, pp. 1–12.

[21] D. Miljković, "Brief review of self-organizing maps," in *Proceedings of the 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, May 2017, vol. 1, no. 1, pp. 1252–1257, doi: 10.23919/MIPRO.2017.7973581

[22] Larochelle, H., Mandel, M., Pascanu, R., and Bengio, Y. "Learning algorithms for the classification restricted Boltzmann machine." *In Advances in Neural Information Processing Systems*, vol. 13, no. 1, pp. 643–669, 2012.

[23] Nair, H. H., Amte, G. S., Todase, N. B., and Dandekar, P. R. "Face detection and recognition in smartphones." *International Journal of Advance Research and Development*, vol. 7, no. 4, pp. 177–182, Apr. 2018.

[24] Silas, O., and Khoya, W. "The study on using biometric authentication on mobile devices." *NU International Journal of Science*, vol. 17, no. 1, pp. 90–110, Jan. 2020.

[25] Shelehov, I. V., Prylepa, D. V., Khibovska, Y. O., and Otroshchenko, M. S. "Machine learning decision support systems for adaptation of educational content to labor market requirements." *Radio Electronics, Computer Science, Control*, vol. 2023, no. Jan, pp. 62-78, 2023.

[26] Z. Zhao, J. H. Chuah, C. O. Chow, K. Xia, Y. K. Tee, Y. C. Hum, and K. W. Lai, "Machine learning approaches in comparative studies for Alzheimer's diagnosis using 2D MRI slices," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 32, no. 1, pp. 1-18, 2024, Art. no. 06, doi: 10.3906/elk-2309-56

[27] E. Erkan and Y. Erkan, "A practical low-dimensional feature vector generation method based on wavelet transform for psychophysiological signals," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 31, no. 7, pp. 1-15, 2023, Art. no. 03, doi: 10.3906/elk-2303-64

[28] Saii, M. M. "Classification of pattern recognition techniques using deep learning and machine learning." *International Journal of Computer Science Trends and Technology*, vol. 7, no. 3, pp. 165–173, Mar. 2019.