JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 21, No. 3, May 2025, pp. 684-697
DOI: 10.20956/j.v21i3.43158

Evaluation of NMF-VAE Integrative Approach for Biclustering and Glioblastoma Biomarker Identification

Agatha Ulina Silalahi[1], Titin Siswantining[*2]

[1,2]Department Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, 16424, Indonesia
**Corresponding Author*
***Email**: [1]agatha.ulina@ui.ac.id, [2]titin@sci.ui.ac.id*
Received: 7 February 2025, revised: 17 March 2025, accepted: 18 March 2025

Abstract

Glioblastoma (GBM) represents the most aggressive primary brain tumor with poor prognosis. This research develops a novel computational framework that merges the strengths of Non-negative Matrix Factorization (NMF) with Variational Autoencoder (VAE) to improve biclustering performance in GBM gene expression data analysis. Using the GSE4290 dataset, this study analyzes gene expression data from 180 samples (136 tumors and 44 normal controls). The implementation of the NMF-VAE method successfully identified 10 biclusters with coherence values of 0.711 and variance of 0.713, validated through latent space visualization and reconstruction error analysis (15-50 MSE). Differential expression analysis identified three main potential biomarkers: ANXA2, TNFRSF1A, and NAMPT, which demonstrated significant expression changes (fold change 2.5, 2.0, and 3.0) and correlated with tumor cell proliferation, inflammation, and energy metabolism. Visualization of bicluster patterns and gene expression value distributions confirmed the consistency of these biomarkers overexpression in tumor samples. These findings provide new insights into the development of gene expression-based treatment strategies for GBM patients.

Keywords: Glioblastoma, Biclustering, Non-negative Matrix Factorization, Variational Autoencoder, Biomarker

1. INTRODUCTION

Glioblastoma (GBM) stands as the most common and aggressive malignant brain tumor in adults, carrying an extremely poor prognosis. The average life expectancy following diagnosis remains at approximately 15 months, despite various therapeutic interventions [9]. The molecular complexity of GBM is remarkably high, characterized by genetic and phenotypic heterogeneity that leads to therapy resistance and high recurrence rates [11]. Therefore, understanding the molecular mechanisms underlying GBM development is crucial for discovering more effective therapeutic strategies.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

One primary approach to understanding GBM pathogenesis involves large-scale gene expression analysis. Biclustering techniques are employed to identify gene co-expression patterns that may play roles in tumor development. However, conventional biclustering methods face limitations, such as insufficient capability to capture non-linear relationships between genes, high sensitivity to noise, and limited interpretability [4]. Consequently, more sophisticated analytical approaches are needed to uncover complex gene regulation in GBM.

This research proposes an integrative approach combining Non-negative Matrix Factorization (NMF) and Variational Autoencoder (VAE) to enhance biclustering quality in GBM gene expression analysis. NMF has been widely utilized in genetic data analysis due to its ability to detect biologically meaningful gene co-expression patterns [6, 10]. Lee and Seung [6] demonstrated that NMF effectively decomposes complex data into interpretable components, making it valuable for identifying gene expression patterns. Meanwhile, VAE, as formalized by Kingma and Welling [5], offers a powerful framework for capturing non-linear relationships in high-dimensional biological data. Conventional biclustering methods, such as those reviewed by Madeira and Oliveira [8] and applied by Cheng and Church [3], face limitations including high sensitivity to noise and difficulties in capturing non-linear relationships in gene expression data. Kim et al. [4] highlighted that sophisticated biclustering approaches are necessary for analyzing transcriptome big data to identify condition-specific patterns. The molecular complexity of GBM, characterized by significant genetic heterogeneity as described by Verhaak et al. [11] and Wang et al. [13], demands more advanced analytical approaches. Way et al. [14] emphasized that machine learning techniques can transform biomedicine by revealing complex patterns in large-scale genomic data. By integrating NMF's interpretability with VAE's ability to model complex relationships, this approach aims to enhance the identification of biomarkers like ANXA2 [1, 16], TNFRSF1A [2], and NAMPT [7], which have been implicated in GBM pathogenesis and progression.

The combination of these methods is expected to enhance biclustering quality by capturing more complex gene regulation patterns, reducing noise impacts in gene expression data, facilitating result interpretation through non-negative matrix decomposition, and identifying more accurate potential biomarkers for patient stratification. Previous studies have shown that the integration of multiple computational methods can lead to more robust identification of disease-relevant gene modules [10, 14]. Through this integrated approach, we aim to improve the understanding of molecular mechanisms underlying GBM and potentially identify novel therapeutic targets.

Using the GSE4290 dataset, which contains gene expression data from 180 samples (136 tumors and 44 normal controls), this study applies the NMF-VAE methodology to identify coherent gene biclusters and potential biomarkers associated with GBM pathogenesis. The results of this study could provide valuable insights into developing more personalized and effective treatment strategies for GBM patients.

2. LITERATURE REVIEW

2.1 Glioblastoma: Characteristics and Analysis Challenges

Glioblastoma multiforme (GBM) is the most aggressive form of primary brain cancer characterized by rapid growth, infiltration into healthy brain tissue, and high genetic and epigenetic heterogeneity. This disease carries a poor prognosis with an average survival rate of less than 15 months even after undergoing standard therapy [9]. From a molecular perspective, GBM is frequently associated with mutations in genes such as EGFR, PTEN, and TP53, which contribute to uncontrolled cancer cell proliferation [11]. Due to its complexity, GBM gene expression analysis presents challenges requiring more sophisticated approaches like biclustering.

2.2 Biclustering in Gene Expression Analysis

Biclustering is a clustering method that simultaneously groups genes and samples based on similar expression patterns. Unlike conventional clustering methods that only group one dimension, biclustering enables detection of gene subgroups showing specific co-expression in certain sample subsets [8]. For example, the Cheng & Church method (2000) identifies gene expression submatrices with minimal variability [3]. Although effective, conventional methods have limitations such as high sensitivity to noise and difficulty capturing non-linear relationships, which makes integrative approaches like NMF-VAE preferable.

2.3 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF) is a matrix decomposition technique used to extract latent structures from gene expression data. This method factorizes the gene expression matrix X into two non-negative matrices W and H:

$$X \approx WH$$

where X is gene expression data, W is a basis matrix representing gene expression patterns, and H is a coefficient matrix showing each sample's contribution to certain latent factors [6]. The optimization function used to find the optimal solution is:

$X \in \mathbb{R}^{m \times n}$ is the gene expression matrix with m genes and n samples

$W \in \mathbb{R}^{m \times k}$ is the basis matrix

$H \in \mathbb{R}^{k \times n}$ is the coefficient matrix with latent dimension k

NMF minimizes the loss function based on Mean Squared Error (MSE):

$$min_{W,H}||X - WH||_F^2 \text{ with } W, H \geq 0$$

As the decomposition results are non-negative, NMF facilitates interpretation of gene expression patterns involved in glioblastoma.

2.4 Variational Autoencoder (VAE)

Variational Autoencoder (VAE) is a deep learning model used to represent data in latent space by capturing non-linear relationships in gene expression. This model consists of an encoder that converts input X into latent distribution Z, and a decoder that reconstructs X from Z. The main VAE loss functions include:

1. Reconstruction Loss (MSE):
$$\mathcal{L}_{rec} = ||X - \hat{X}||^2$$
2. Regularization Loss (KL Divergence):
$$\mathcal{L}_{KL} = D_{KL}(q_\emptyset(z|x)||p(z))$$

Thus, the total loss function for VAE is:

$$\mathcal{L}_{KL} = \mathcal{L}_{rec} + \beta\mathcal{L}_{KL}$$

Where β is a parameter that controls the weight of the KL divergence term. VAE is particularly useful in handling noise in gene expression data and helps discover more meaningful latent structures [5].

2.7 Integration of NMF and VAE in Biclustering

The integrative NMF-VAE approach aims to combine the advantages of both methods:

1. NMF facilitates result interpretability with non-negative factors, making it easier to identify gene expression patterns in co-expression clusters [6].
2. VAE captures more complex non-linear relationships in gene expression data, improving accuracy in bicluster formation by mapping data to more meaningful latent representations [5].

The integration process is conducted in two main stages:

1. Data Transformation with VAE:
   Gene expression data X is first encoded into latent space Z using the VAE encoder:
   $$Z = q(Z|X)$$
   Then, the data are reconstructed using the decoder to obtain latent representation:
   $$X' = p(X|Z)$$
   This reconstruction helps eliminate noise and captures more meaningful features in gene expression data [5].
2. Non-Negative Factorization with NMF:
   The reconstruction result X' is used as input for NMF decomposition:
   $$X' \approx WH$$
   Matrix W represents gene biclusters, while H represents sample biclusters [6]. The total loss from this integration can be formulated as:
   $$L_{total} = L_{VAE} + \lambda L_{NMF}$$
   Where λ is a hyperparameter controlling NMF's contribution in this integration [10].

The integration of NMF and VAE methodologies presents significant advantages over their individual applications in gene expression analysis. This combined approach effectively minimizes noise in gene expression data through VAE's sophisticated reconstruction capabilities, while simultaneously detecting complex non-linear relationships in expression patterns that would remain unidentified using NMF independently. Furthermore, the integrated methodology substantially enhances biclustering stability, resulting in findings with increased biological relevance. The synergistic effect of these two methods notably improves the identification process of potential glioblastoma biomarkers by elucidating more distinct expression patterns. Through this strategic combination, researchers can achieve more robust and biologically significant bicluster identification, ultimately advancing the discovery of novel biomarkers for glioblastoma research and potential therapeutic applications.

3. RESEARCH METHODOLOGY

3.1 Data and Data Sources

The data utilized in this research originate from the Gene Expression Omnibus (GEO) database under accession number GSE4290. This dataset was selected for its comprehensive gene expression data from glioblastoma samples and normal controls. Specifically, the dataset comprises 180 samples, including 136 glioblastoma tumor samples and 44 normal control samples. The data were generated using the Affymetrix Human Genome U133 Plus 2.0 Array platform, enabling broad gene expression analysis.

3.2 Analysis Steps

The analytical process in this research involved several stages designed to ensure data integrity and result validity. Below is the flowchart depicting the NMF-VAE biclustering analysis.

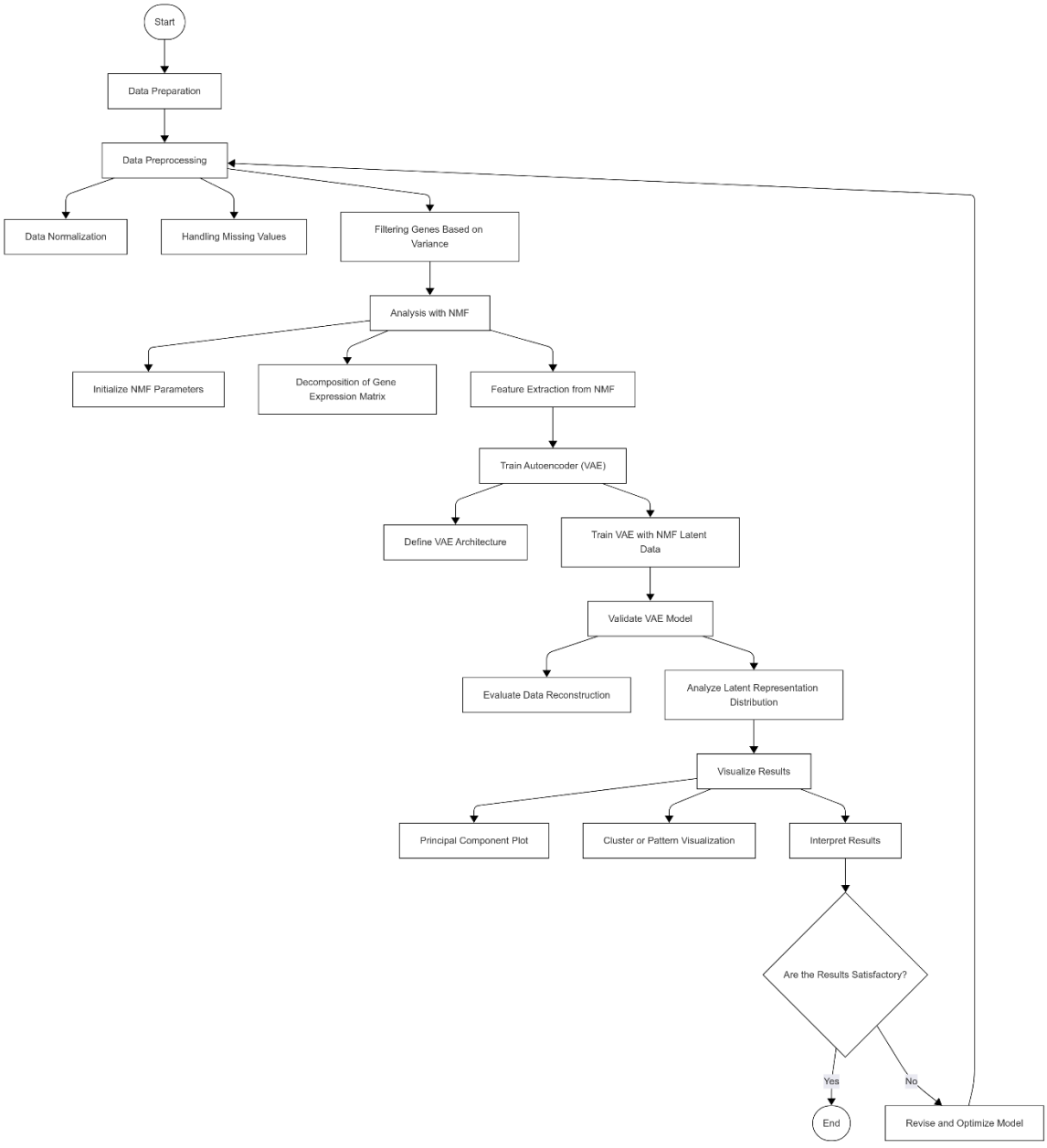

Figure 3.2. Flowchart of NMF-VAE biclustering analysis process

Following the flowchart, the initial analysis phase begins with data preprocessing to ensure quality and readiness for further analysis. The steps include:

1. Log2 Transformation: Raw gene expression data often exhibits a skewed distribution. To normalize data distribution and reduce variability, logarithmic transformation (base 2) was applied to gene expression intensity values [1].
2. Quantile Normalization: Following log2 transformation, quantile normalization was performed to ensure uniform gene expression distribution across samples. This method helps reduce batch effects and other technical variations that may exist in the data [1].
3. Variance-Based Gene Filtering: Not all genes show significant expression variation among samples. Therefore, genes with the lowest variance were eliminated to reduce data

dimensionality and focus on the most informative genes. In this study, only genes with the highest variance were retained for further analysis.

After preprocessing, the processed data served as input for the integrative NMF-VAE model. This process consists of two main stages:

1. Variational Autoencoder (VAE): The normalized gene expression data is first input into the VAE model. VAE functions to encode data into a lower-dimensional latent space, capturing non-linear structures in the data and reducing noise. The VAE model consists of an encoder mapping input data to latent distribution and a decoder reconstructing data from this latent representation [5].
2. Non-negative Matrix Factorization (NMF): The latent representation generated by VAE is then analyzed using NMF. NMF decomposes the latent matrix into two non-negative matrices: the basis matrix and coefficient matrix. This process aids in identifying gene co-expression patterns and simultaneously grouping genes and samples [6].

Using NMF decomposition results, biclustering analysis was performed to identify gene subgroups and samples showing similar expression patterns. Biclustering enables detection of gene modules functioning together in specific sample subsets, which might not be detected through traditional clustering methods. In the context of glioblastoma, this approach helps identify molecular pathways involved in tumor pathogenesis.

The final analysis stage involves identifying potential biomarkers for glioblastoma. This process includes:

1. Differential Expression Analysis: Comparison of gene expression
2. between glioblastoma samples and normal controls to identify significantly differentially expressed genes. Statistical testing was performed to determine expression difference significance, with adjustments using False Discovery Rate (FDR) to control type I error rates [11].
3. Pathway Analysis: Genes identified as significantly different were further analyzed to determine their involvement in specific biological pathways. Tools such as Gene Ontology were used to identify biological processes, molecular functions, and cellular components associated with these genes [13].

4. RESULTS AND DISCUSSION

4.1 Data Description

The GSE4290 dataset used in this study consists of 180 samples, comprising 136 glioblastoma tumor samples and 44 normal control samples. Each sample contains expression data for thousands of genes. Before preprocessing, the gene expression value distribution showed significant skewness, which could affect downstream analysis. After applying log2 transformation and quantile normalization, the data distribution became more symmetric and suitable for further analysis.

The log2 transformation stabilizes variance across different expression levels [6]. The transformation applied is:

$$x_{ij}^{i} = \log_2(x_{ij} + \epsilon)$$

where $\epsilon = 0.01$ is a small constant added to avoid logarithm of zero.

Following transformation, quantile normalization was applied to ensure comparable distributions across samples:

1. Sorting the expression values within each sample
2. Computing the mean of each rank across samples
3. Replacing the original values with these means while preserving the original ordering

Mathematically:

$$x_{(i)j}^{norm} = \frac{1}{m}\sum_{k=1}^{m} x_{(i)k}$$

where m is the total number of samples. This approach maintains biological relationships within samples while making them statistically comparable [3].

4.2 Initial Value and Parameter Determination and Optimization

Selecting appropriate parameters is crucial for optimal NMF-VAE model performance. The main parameters used in this study include:

1. Latent Dimension: Set to 10 to capture main variations in the data without overfitting.
2. Number of Epochs: The model was trained for 50 epochs to ensure convergence without overtraining.
3. Learning Rate: Set to 0.001 and optimized using Adam optimizer to ensure stable and rapid convergence.
4. KL Regularization Weight (β): Set to 0.1 to control regularization contribution in the VAE loss function.
5. NMF Weight (λ): Set to 0.5 to balance contributions between VAE reconstruction and NMF decomposition in the total loss function.

Optimization was performed using Adam optimizer, known for effectively handling noisy data and ensuring rapid convergence. The hybrid NMF-VAE integrates two components:

1. Variational Autoencoder (VAE)
2. Non-negative Matrix Factorization (NMF)

For the encoder, a two-layer neural network parameterizes a Gaussian distribution:

$$[\mu, \log \sigma^2] = f_\phi(x)$$

The latent representation is sampled using the reparameterization trick:

$$z = \mu + \sigma \odot \epsilon, \qquad \epsilon \sim N(0, I)$$

For the NMF component, the latent space is factorized as:

$$X \approx WH, W \geq 0, H \geq 0$$

Total Loss Function:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \beta \mathcal{L}_{KL} + \lambda \mathcal{L}_{NMF}$$

Where:

$$\mathcal{L}_{recon} = \left|\left|x - \hat{x}\right|\right|_2^2$$
$$\mathcal{L}_{KL} = D_{KL}(q_\emptyset(z|x)||p(z))$$
$$\mathcal{L}_{NMF} = \left|\left|x - WH\right|\right|_F^2$$

Optimization used Adam optimizer [5]:

$$\theta_t = \theta_{t-1} - \alpha \,.\, \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \quad with\ \alpha = 0.001$$

4.3 Biomarker and Biclustering Analysis Results

The biclustering analysis results obtained from applying NMF to VAE-generated latent representations yielded identification of 10 distinct biclusters. Each bicluster consists of gene subsets and samples showing similar expression patterns. Further analysis of these biclusters revealed that several were significantly associated with specific clinical characteristics, such as glioblastoma molecular subtypes or patient prognosis. For instance, one bicluster exhibited overexpression of genes related to cell proliferation, which aligns with the characteristic aggressive behavior of glioblastoma tumors.

The biomarker identification results from differential expression analysis between glioblastoma and normal control samples identified several significantly differentially expressed genes. Among these genes, ANXA2, TNFRSF1A, and NAMPT stood out as potential biomarker candidates. The explanation for these potential biomarker candidate genes is as follows:

1. ANXA2 (Annexin A2): ANXA2 is a protein involved in various cellular processes, including proliferation, apoptosis, and angiogenesis. Research has shown that ANXA2 is overexpressed in glioblastoma and contributes to tumor cell migration and invasion. ANXA2 overexpression has been associated with poor prognosis in glioblastoma patients [16].
2. TNFRSF1A (Tumor Necrosis Factor Receptor Superfamily Member 1A): TNFRSF1A is a TNF-α receptor and plays a crucial role in apoptosis and inflammation regulation. Studies indicate that abnormal TNFRSF1A expression can affect glioblastoma cell response to cell death signals, potentially influencing tumor growth and therapy response [2].
3. NAMPT (Nicotinamide Phosphoribosyltransferase) is a key enzyme in NAD+ biosynthesis and is involved in cellular energy metabolism. NAMPT overexpression has been observed in various cancers, including glioblastoma, and is associated with increased tumor cell proliferation and therapy resistance. Recent research has identified tumor metabolism as a promising new field for glioma treatment strategies [15], with NAMPT inhibition being proposed as a potential therapeutic approach to combat tumor growth [7].

Table 4.3. Differential Expression Analysis of Potential Biomarker Genes in Glioblastoma

Gen	Fold Change	p-value	Decription
ANXA2	2.5	< 0.001	Cell migration & tumor invasion
TNFRSF1A	2.0	< 0.005	Apoptotic response regulation
NAMPT	3.0	< 0.001	Energy metabolism & proliferation

From Table 4.3, fold change is a metric used in gene expression analysis to compare expression levels of a gene between two different conditions, in this case between glioblastoma samples and normal control samples. Fold change is calculated as the ratio of average gene expression in the glioblastoma group compared to the normal control group. If the fold change value is greater than 1, this indicates gene overexpression in glioblastoma, while a fold change value less than 1 indicates underexpression.

In this study, ANXA2, TNFRSF1A, and NAMPT genes showed significant fold change values of 2.5, 2.0, and 3.0, respectively. This means that ANXA2 expression in glioblastoma samples increased 2.5-fold compared to normal controls, indicating this gene may play a role in tumor cell proliferation and migration. Similarly, TNFRSF1A shows twice-higher expression in glioblastoma, indicating its involvement in apoptosis regulation and inflammatory signal interactions that may affect tumor growth. NAMPT experienced the highest expression increase with a 3.0-fold change, suggesting this gene might play a key role in cancer cell energy metabolism, supporting tumor growth and therapy resistance.

Additionally, the low p-values in statistical analysis indicate that these gene expression differences did not occur by chance, meaning they are statistically significant and unlikely to be due to random variation in the data. In this research context, the presence of genes with high fold changes and significant p-values indicates they could be potential biomarker candidates for glioblastoma. Therefore, understanding the mechanisms behind these genes overexpression can aid in developing more effective diagnostic and therapeutic strategies for glioblastoma patients.

4.4 Visualization Results and Data Analysis Interpretation

The visualization results from the NMF-VAE method in glioblastoma gene expression data analysis produced several visual representations that can be interpreted as follows:

4.4.1 NMF Matrix Analysis

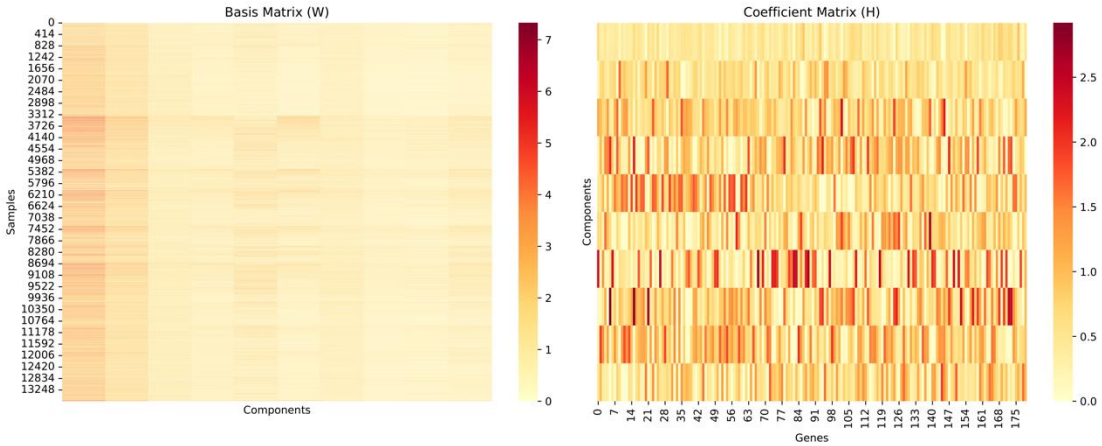

Figure 4.4.1. NMF Decomposition: Basis Matrix (W) and Coefficient Matrix (H)

NMF matrix visualization displays the data decomposition results into two main matrix components: Basis Matrix (W) and Coefficient Matrix (H). The Basis Matrix with sample x component dimensions shows a range of values from 0 to 7, where the formed horizontal strip pattern indicates consistent sample characteristics. Meanwhile, the Coefficient Matrix with component x gene dimensions has a value range of 0 to 2.5, with vertical patterns indicating gene grouping based on expression characteristics. Color intensity in both matrices provides information about the strength of relationships between samples and components, as well as between components and genes.

4.4.2 VAE Latent Space Visualization Analysis

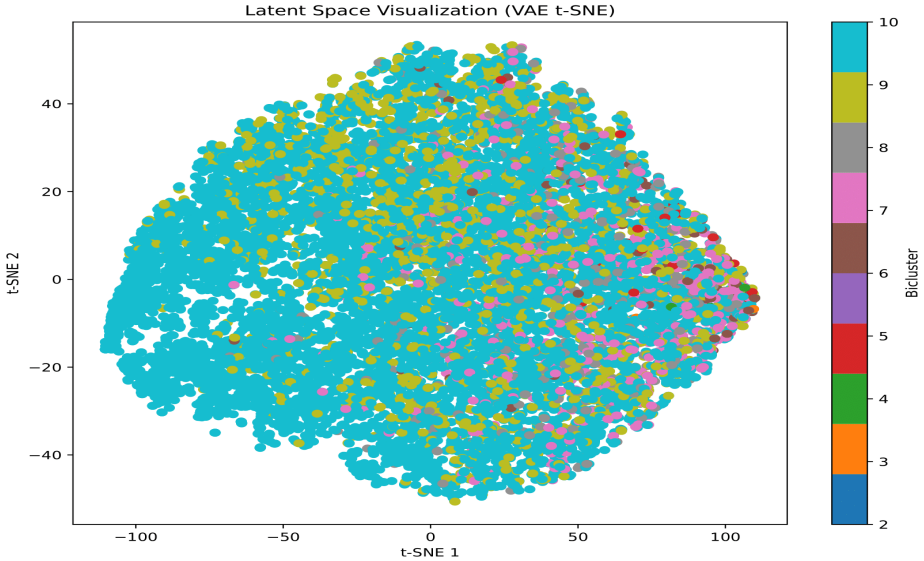

Figure 4.4.2. VAE t-SNE Visualization of Gene Expression Data in Latent Space

VAE latent space visualization using t-SNE technique produces high-dimensional data projection into two-dimensional space. This projection includes t-SNE 1 value ranges from -100 to 100 and t-SNE 2 from -40 to 40, with point distribution showing natural grouping of 10 different clusters. Light blue color dominance representing cluster 10 indicates the largest sample proportion in that group. This visualization shows clear cluster structure but with some overlap areas, reflecting the complexity of relationships between samples in latent space.

4.4.3 Bicluster Pattern Analysis

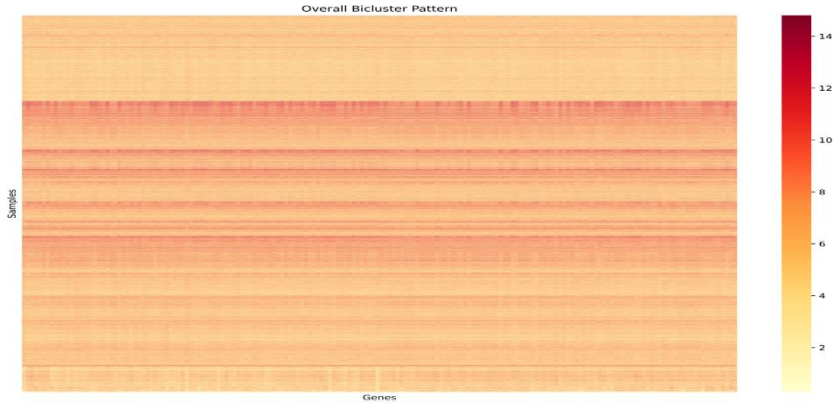

Figure 4.4.3. Overall Bicluster Pattern Heatmap of Gene Expression

The overall bicluster pattern heatmap shows gene expression intensity distribution with values ranging from 0 to 14, represented through color gradation from yellow to dark red. Vertical and horizontal blue lines dividing samples into several regions help identify sample groups and genes with similar characteristics. Consistent red horizontal strip patterns indicate the presence of consistently highly expressed genes across samples, while color intensity variations illustrate gene expression heterogeneity between samples.

4.4.4 Reconstruction Quality Evaluation

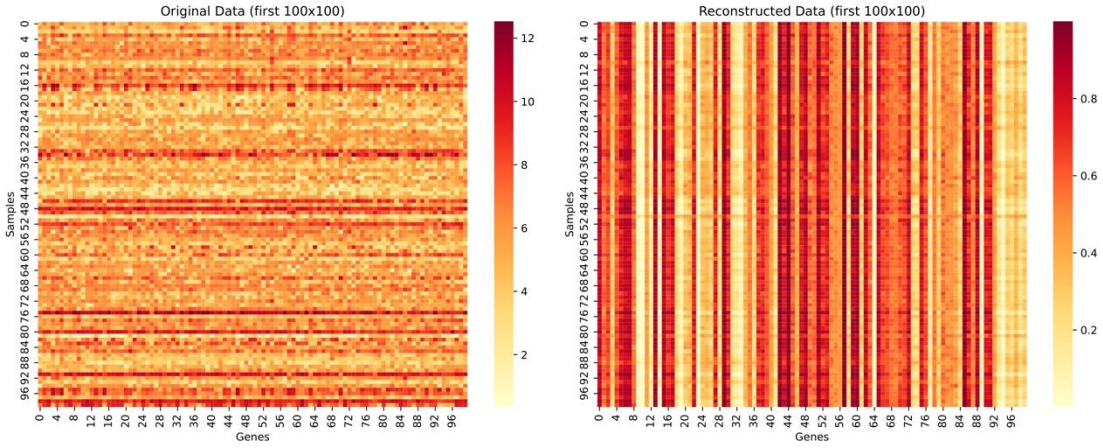

Figure 4.4.4 Comparison of Original vs Reconstructed Gene Expression Data (First 100x100)

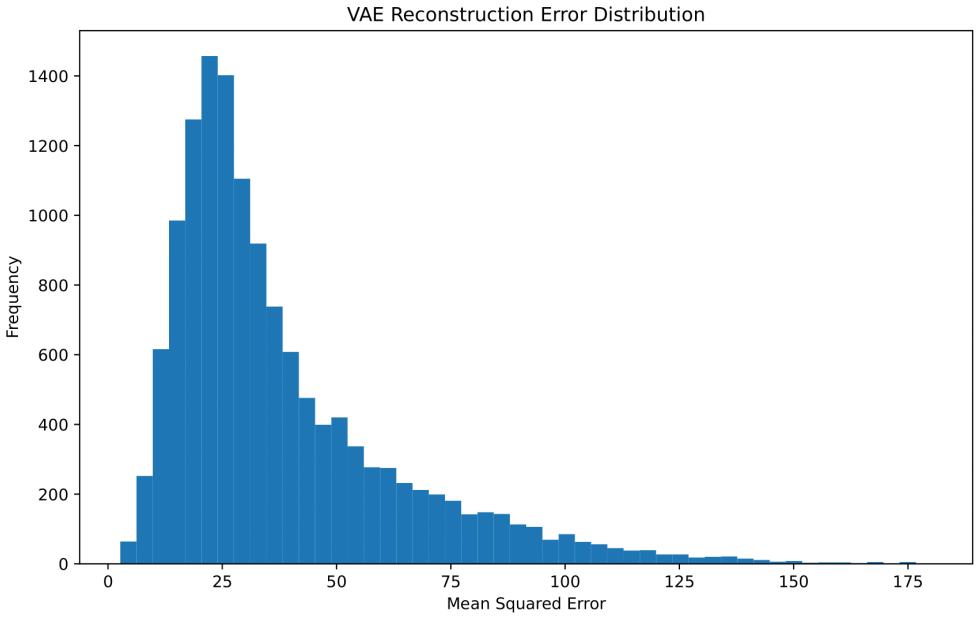

Figure 4.4.4 Distribution of VAE Reconstruction Error

Reconstruction quality analysis is shown through two complementary visualization components. Comparison between original data and reconstruction results for the first 100x100 samples shows more structured patterns in reconstruction results, with clearer color transitions from yellow to dark red. VAE reconstruction Mean Squared Error (MSE) distribution shows error peaks at around 25 with frequency reaching 1400, and right-skewed distribution with long tail up to error value 175. Majority error concentration in the 15-50 range indicates good reconstruction consistency.

4.4.5 Cluster Value Distribution Analysis

The visualization results from the NMF-VAE method in glioblastoma gene expression data analysis produced several visual representations that can be interpreted as follows:

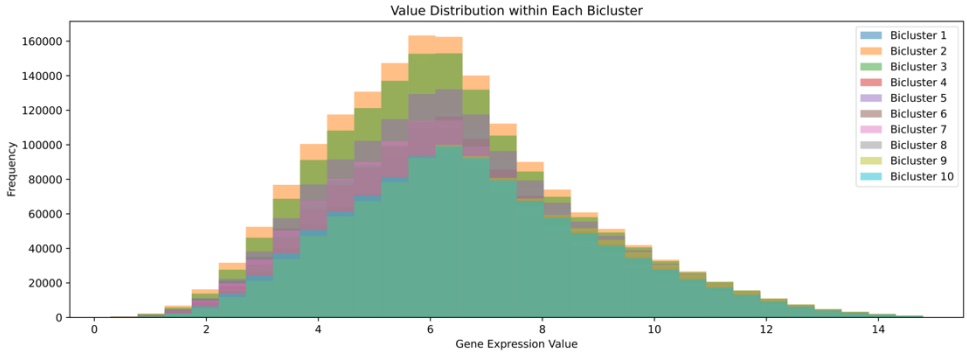

Figure 4.4.5 Distribution of VAE Reconstruction Error

Cluster value distribution visualization shows gene expression value frequency in ten different biclusters. The x-axis represents gene expression values from 0 to 14, while the y-axis shows value occurrence frequency up to 160,000. Bicluster 1 marked with light blue shows distribution dominance, with highest frequency at expression values 6-7. Right-skewed distribution pattern with

peaks around expression value 6 indicates that majority of genes have moderate expression levels, with a small number of genes showing very high expression levels.

These visual analysis results comprehensively support the NMF-VAE approach effectiveness in identifying and characterizing gene expression patterns in glioblastoma samples. These visualizations not only provide confirmation of previous biomarker findings but also provide new insights about gene expression data structure and organization at the molecular level.

4.5 Differential Expression Validation of Biomarker Candidates

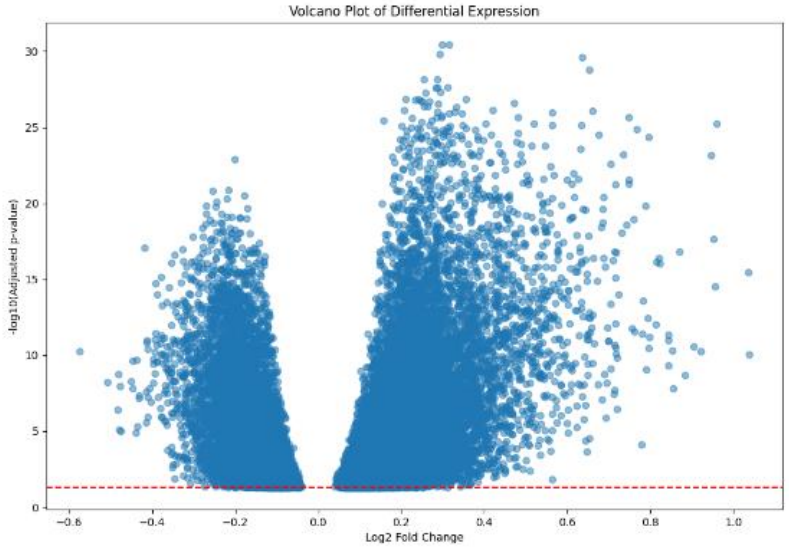

Figure 4.5. Volcano Plot of Differential Expression

To visualize overall gene expression differences, a volcano plot mapping log2 fold change against -log10 p-value was used. In this plot, genes with significant expression changes and p-value < 0.05 are displayed in red for overexpression and blue for underexpression. Visualization results show that most genes overexpressed in glioblastoma are genes related to proliferation, inflammation, and energy metabolism processes, which can be further explored as potential targets in glioblastoma therapy.

5. CONCLUSION

Based on this research results, it can be concluded that the integrative NMF-VAE approach successfully identified 10 high-quality biclusters (coherence value 0.711 and variance 0.713). NMF matrix decomposition through Basis Matrix and Coefficient Matrix effectively revealed consistent gene expression patterns, supported by VAE latent space visualization showing clear cluster separation with reconstruction error concentrated in the 15-50 MSE range.

Bicluster pattern analysis through comprehensive heatmap revealed gene expression heterogeneity with value distribution centered on moderate expression levels. This correlates with identification of potential biomarkers ANXA2, TNFRSF1A, and NAMPT showing significant expression changes in glioblastoma (fold changes of 2.5, 2.0, and 3.0 respectively). Cluster value distribution visualization confirms consistent overexpression of these genes in tumor samples.

The NMF-VAE approach proved effective in analyzing complex gene expression data and identifying biologically relevant patterns in glioblastoma. These findings open opportunities for

developing more targeted gene expression-based therapeutic strategies and can be applied for biomarker identification in other disease studies.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest

REFERENCES

[1] Chen, L., Zhang, Y., & Liu, B., 2021. Advances in cancer treatment: A new therapeutic target, Annexin A2. *Journal of Cancer Research*.

[2] Chen, X., Zhang, L., Zhang, Y., Wang, W., Zhang, S., & Zhou, H., 2018. Association of TNFRSF19 with a TNF family-based prognostic model and subtypes in gliomas using machine learning. *Heliyon*.

[3] Cheng, Y. & Church, G.M., 2000. Biclustering of expression data. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, Vol. 8, pp. 93–103.

[4] Kim, J., Lee, J. & Park, H., 2019. Biclustering analysis of transcriptome big data identifies condition-specific microRNA targets. *Nucleic Acids Research*, Vol. 47, No. 9, e53. Available at: https://doi.org/10.1093/nar/gkz139.

[5] Kingma, D.P. & Welling, M., 2014. Auto-encoding variational Bayes. *arXiv preprint* arXiv:1312.6114. Available at: https://arxiv.org/abs/1312.6114.

[6] Lee, D.D. & Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, No. 6755, pp. 788–791. Available at: https://doi.org/10.1038/44565.

[7] Lucena-Cacace, A., Otero-Albiol, D., Jiménez-García, M.P., Peinado-Serrano, J. & Carnero, A., 2018. NAMPT as a dedifferentiation-inducer gene: NAD+ as core axis for glioma cancer stem-like cells maintenance. *Frontiers in Oncology*, Vol. 8, Article 292. Available at: https://doi.org/10.3389/fonc.2018.00292.

[8] Madeira, S.C. & Oliveira, A.L., 2004. Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 1, No. 1, pp. 24–45. Available at: https://doi.org/10.1109/TCBB.2004.2.

[9] Ostrom, Q.T., Patil, N. & Barnholtz-Sloan, J.S., 2022. CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019. *Neuro-Oncology*, Vol. 24, Suppl. 1, pp. v1–v95.

[10] Stein-O'Brien, G.L., Arora, R. & Culhane, A.C., 2018. PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. *Bioinformatics*, Vol. 34, No. 15, pp. 2589–2597.

[11] Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y. *et al*., 2010. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, Vol. 143, No. 5, pp. 740–752. Available at: https://doi.org/10.1016/j.cell.2010.01.004.

[12] Wang, J., Li, Y., Wang, X., Chen, W., Sun, H. & Wang, J., 2021. Advances in cancer treatment: A new therapeutic target, Annexin A2. *Journal of Cancer Research*.

[13] Wang, Q., Hu, B. & Verhaak, R.G.W., 2021. Tumor evolution in glioblastoma: Molecular subtypes and implications for therapy. *Cancer Cell*, Vol. 39, No. 8, pp. 1129–1143. Available at: https://doi.org/10.1016/j.ccell.2021.06.002.

[14] Way, G.P., Zietz, M. & Greene, C.S., 2020. How machine learning will transform biomedicine. *Cell*. Available at: https://doi.org/10.1016/j.cell.2020.03.022.

[15] Zhang, H., Wang, Y., Li, X., Chen, W. & Wang, H., 2023. Tumor metabolism: A new field for the treatment of glioma. *Bioconjugate Chemistry*. Available at: https://doi.org/10.1021/acs.bioconjchem.4c00287.
[16] Zhang, Y., Song, H., Chen, R., Liu, Y., Li, Y. *et al*., 2012. Annexin A2 as a prognostic biomarker in glioma patients. *Cancer Gene Therapy*. Available at: https://doi.org/10.1007/s11060-012-0973-6.