Chapter 05

AI and Machine Learning in Healthcare and Biomedical Engineering
ISBN: 978-9915-704-01-2
DOI: 10.62486/978-9915-704-01-2.ch05
Pages: 31-36

Unsupervised Learning-Based Classification of Breast Cancer Using Gaussian Mixture Model

Hussein Alkattan[1] ✉, Salam Abdulkhaleq Noaman[2] ✉, Abrar Ryadh[3] ✉, Mostafa Abotaleb[4] ✉, H.K. Al-Mahdawi[5] ✉, Elman Azimov[6], Pushan Kumar Dutta[7] ✉

[1]Department of System Programming, South Ural State University. Chelyabinsk, Russia.
[2]College of Education for Pure Science, University of Diyala. Diyala, Iraq.
[3]Medical Laboratory Techniques, Mustaqbal University. Hillah, Iraq.
[4]Engineering School of Digital Technologies, Yugra State University. Khanty-Mansiysk, 628012, Russia.
[5]Electronic Computer Centre, University of Diyala. Diyala, Iraq.
[6]Azerbaijan Medical University, Azerbaijan
[7]School of Engineering and Technology, University of Amity. Kolkata, India.

Corresponding Author: Hussein Alkattan ✉

ABSTRACT

This work describes the application of unsupervised learning techniques for the categorization of breast cancer using the GBSG dataset. GMMs were applied to divide the data into four clusters in order to find latent group structures with the patients based on a lack of class labels. PCA was also applied to reduce dimensionality for display, collecting more than 95 % of the variation in the first two components. The cluster distribution showed that Cluster 0 and Cluster 1 had 223 (~41,6 %) and 303 (~56,6 %) patients, respectively, whereas Cluster 2 and Cluster 3 had only 9 (~1,7 %) and 13 (~2,4 %) patients. The ARI value of -0,0033 indicated little agreement between GMM clusters and the actual cancer state. In addition, a correlation heatmap showing strong positive connections of various clinical aspects allowed the identification of redundant variables. Although unsupervised clustering did not strongly agree with labeled classes, it returned useful information concerning data distribution and possible feature interactions. These results validate the use of GMM-based unsupervised learning in semi-supervised classification frameworks or preprocessing pipelines for predicting breast cancer.

Keywords: Breast Cancer Classification; Dimensionality Reduction; Variance Explained; Cluster Distribution; Semi-supervised Classification.

INTRODUCTION

Breast cancer remains one of the most frequent and heterogeneous tumors among women worldwide. The heterogeneity of breast cancer significantly influences its diagnosis, prognosis, and therapy.(1) Precision medicine is a key issue in cancer because intratumor and intertumoral tumor heterogeneity complicates the search for reliable biomarkers and prognostic models.(1,2) Improvements in genomic technologies and molecular profiling have now allowed researchers to generate multigene prognostic tools offering improved patient stratification with respect to molecular subtypes.(2,3)

Conventional diagnostic techniques, such as IHC evaluation of HER2/neu status, PR, and ER, have been universally accepted as clinical standards.(4,5) Accordingly, the inconsistent interpretation and manipulation of samples intrinsic in previous approaches have shifted the

focus of researchers to transcriptomic and array-based technologies, which can evaluate more accurately.[6,7] Gene expression profiling can uncover both prognostic and predictive indicators, including but not limited to ESR1, HER2, and IGF1R, and gives far more detailed insights into molecular subtypes.[8,9] In some subtypes of breast cancer, high expression of these markers, especially IGF1R, was linked with limited concordance rates.[10,11,12]

Despite these successes, the translation of omics data into routine clinical practice still faces major challenges. The non-Gaussian nature of expression patterns in patient material significantly reduces classification performance, often causing bias or insufficiency when standard statistical methods are applied.[13,14] This has motivated a growing application of probabilistic modeling and machine learning to biomedical data. Gaussian Mixture Model represents a general probabilistic approach for identifying subpopulations in diverse datasets.[15,16,17,18] GMM has been successfully applied to speech recognition[15], super pixel segmentation in imaging[16], and more recently, clustering gene expression and proteomics data.[17,19,20]

In the context of cancer classification, GMM can identify unique molecular signatures and patient groups under the assumption that data points originate from a mixture of several Gaussian distributions. This model offers soft clustering, considering classification uncertainty and is useful in cases where class borders are not linearly separable.[19,20] Because of these characteristics, GMM is suitable for analyzing datasets related to breast cancer, which often contain high-dimensional, noisy, and overlapping features.

Moreover, the IGF pathway has been identified to play a significant role in breast cancer biology. High expression of IGF1R is linked with tumor proliferation, linked with limited response to therapy, and higher recurrence rates.[21,22,23,24] Targeted treatments using anti-IGF1R therapies are justified by the finding of its overexpressed mRNA as a potential biomarker for aggressive phenotypes.[22,23] Computational models that incorporate IGF pathway markers provide greater robustness in prognostic classification and allow for more personalized treatment stratification.[25,26,27]

In light of the aforementioned findings, this study utilizes publically accessible clinical data for unsupervised clustering of breast cancer patients through a Gaussian Mixture Model. We then investigate the concordance of these clusters with established clinical markers and the potential of GMM to uncover subgroup patterns not readily evident from the dataset. Further, we apply correlation analysis to evaluate clinical parameter interdependencies and their utility with regard to patient stratification. The results of this investigation contribute to research directed at combining precision oncology with artificial intelligence to improve breast cancer management. The results of this study indicated that Gaussian Mixture Models has both potential and constraints concerning unsupervised classification of breast cancer patients. Observed clustering showed linked with limited concordance with clinical outcomes despite distinct categories, underlining the challenge of relying on unsupervised learning alone for diagnostic purposes. Among other findings, the research made the important discovery of patterns of correlations and redundancies among clinical attributes, indicating that unsupervised methods might still prove useful for preprocessing, data exploration, and dimensionality reduction in subsequent modeling with predictive intent. These implications reflect the potential of GMM as a supporting tool in precision oncology, in which complex, high-dimensional datasets require techniques capable of uncovering hidden structures less accessible or apparent to conventional statistical methods.

Future studies should focus on enhancing the interpretability and clinical utility of unsupervised clusters to further advance this work. This can be done by incorporating GMM into semi-supervised or weakly supervised approaches where cluster development is informed by a small amount of labeled data, coupled with the evaluation of GMM-defined clusters against known biomarkers such as ER, PR, HER2, and IGF1R expression. Identification of subgroups would also be strengthened by utilizing robust cluster validity metrics (BIC, AIC, silhouette scores) and testing mixture models that relax the Gaussian assumption. These methodological improvements would extend the translational impact of computational results through a closer correspondence with clinical relevance. Looking forward, the extension of this approach toward multi-omics integration, including transcriptomic, proteomic, and genomic data, represents an exciting direction for subgroup discovery based upon biological rationale. Hybrid methods that combine GMMs with deep learning-based frameworks, including variational autoencoders, can uncover nonlinear relationships that are often missed by conventional methods. Additionally, the application of these algorithms to longitudinal data extends insights into temporal aspects of treatment response and disease progression, while external replication across other cohorts increases generalizability. Finally, integrating GMM-based clustering into clinical decision support systems has the potential to transform unsupervised learning findings into actionable information that may enable personalized treatment approaches and serve the broader goals of precision medicine.

METHOD

Dataset

The dataset used in this investigation came from a publicly accessible repository, "Breast Cancer Dataset Used (Royston and Altman)," available on Kaggle. It provides clinically relevant data from the German Breast Cancer investigation Group, GBSG.[27] This dataset includes complete data from 686 female patients with primary breast cancer. Due to the highly clinically significant and well-documented nature of this data, it was first utilized in research related to survival analysis and later formed part of several study initiatives.

Only baseline clinical variables were considered for the purpose of unsupervised clustering and feature connection analysis. The dataset includes a range of prognostic indicators, which include tumor size, number of positive lymph nodes, menopausal status, use of hormone therapy, age at diagnosis, and levels of progesterone and estrogen receptors. The characteristics chosen are those that are known to be helpful in classifying breast cancer risk and developing treatment strategies. For the only purpose of concentrating on data-driven grouping without supervision, outcome labels such as survival time and event incidence were left out.

Data cleaning was done to exclude identifiers like patient IDs and any time-to-event columns prior to modeling. The features were then standardized using standard z-score scaling in order to ensure that each variable contributed equally during a distance-based clustering calculation. This preprocessing step is particularly critical in Gaussian Mixture Models, because feature scaling strongly effects the shape and direction of the learnt distributions.

Figure 8.1 shows a heatmap of the Pearson correlation coefficients between the clinical attributes in the breast cancer data. High positive correlation between age and menopausal status (r = 0,76) indicates older patients tend to be post-menopausal. Moderate correlations are seen between estrogen receptor (ER) and progesterone receptor (PGR) levels (r = 0,39), which is due to their common hormonal origin. Most other features exhibit weak or negligible correlations, indicating they contribute uniquely to the data structure.

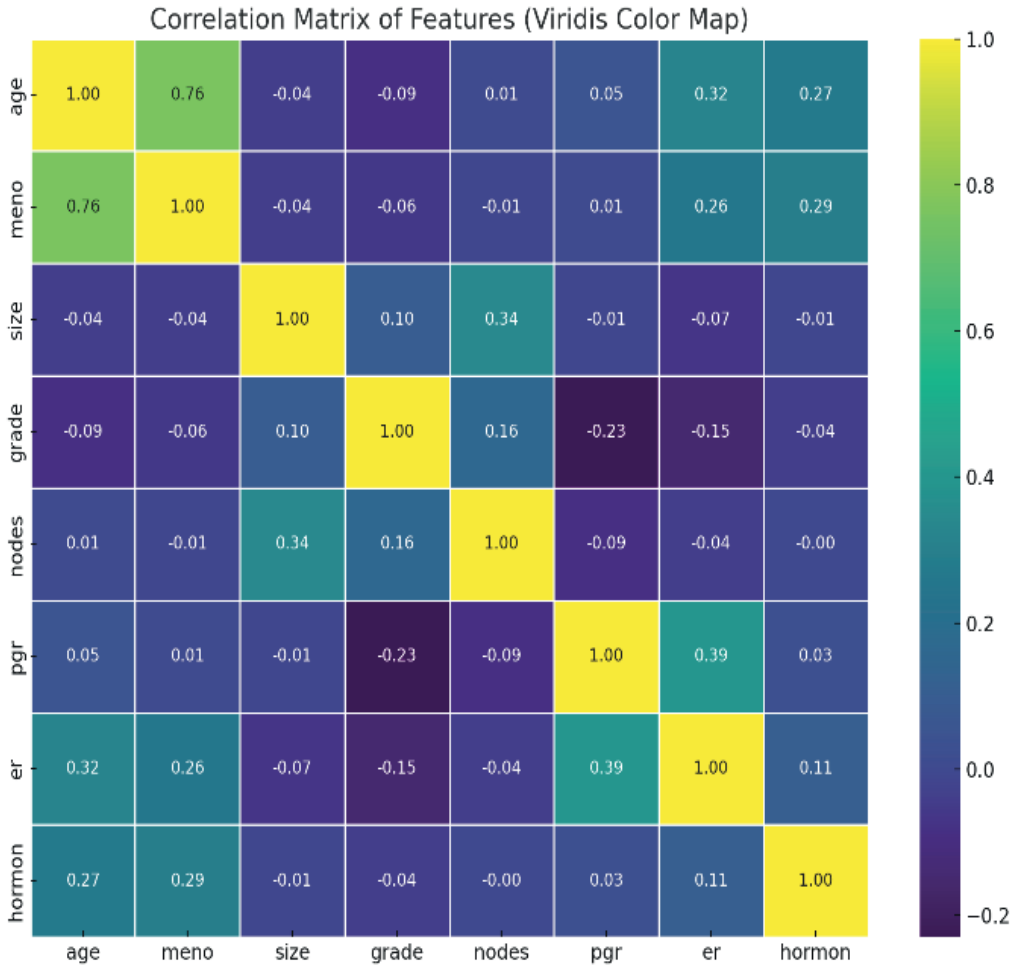

Figure 8.1. Correlation Matrix of Clinical Features

Clustering with Gaussian Mixture Models

Gaussian Mixture Models (GMM) are probabilistic models that are used to represent normally distributed subpopulations inside the overall population. This approach assumes that the data is generated from a mixture of K Gaussian distributions, each one defined by its specific mean and covariance matrix. The probability density function of the GMM is given by:

$$p(x) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x \mid \mu_k, \Sigma_k) \qquad (1)$$

where $\pi_k \pi_k$ are the mixing coefficients such that $\sum_{k=1}^{K} \pi_k = 1 \sum_{k=1}^{K} \pi_k = 1$, and $\mathcal{N}(x \mid \mu_k, \Sigma_k)$ $\mathcal{N}(x \mid \mu_k, \Sigma_k)$ denotes the multivariate Gaussian distribution for the kk-th component with mean $\mu_k \mu_k$ and covariance $\Sigma_k \Sigma_k$.

The multivariate Gaussian distribution is defined as:

$$\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \quad (2)$$

$$\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \quad (3)$$

To optimize the parameters $\pi_k, \mu_k, \Sigma_k \pi_k, \mu_k, \Sigma_k$, the Expectation-Maximization (EM) algorithm is applied. In the Expectation step (E-step), the responsibilities $\gamma(z_{nk})\gamma(z_{nk})$ for each data point $x_n x_n$ belonging to component kk are computed as:

$$\gamma(z_{nk}) = \frac{\pi_k \cdot \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{i=1}^{K} \pi_j \cdot \mathcal{N}(x_n \mid \mu_j, \Sigma_j)} \quad (4)$$

In the Maximization step (M-step), parameters are updated using the responsibilities computed:

$$\mu_k = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk})x_n \quad (5)$$

$$\Sigma_k = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T \quad (6)$$

$$\pi_k = \frac{N_k}{N}, \text{ where } N_k = \sum_{n=1}^{N} \gamma(z_{nk}) \quad (7)$$

This process continues until convergence, usually by having the change in log-likelihood reduced to some limit set in advance. The log-likelihood of the observed data given the parameters is defined as:

$$\log L = \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right) \quad (8)$$

The final output assigns each sample to the cluster with the highest posterior probability. GMM's soft clustering capability makes it suitable for medical data, where boundaries between classes are not always discrete. The number of components K was empirically set to four to reflect potential subtypes in the clinical dataset. Dimensionality reduction using Principal Component Analysis (PCA) was applied to project the results in 2D space for visual inspection of cluster separability.[29,30,31,32,33]

RESULT

Figure 2 show PCA Projection This scatter plot is visualizing the result of Gaussian Mixture Model clustering projected onto two principal components. Each point on the plot represents a patient, colored by the cluster label assigned to it. As is clear, clusters 0 and 1 appear to make up the majority of the data, but what is perhaps even more interesting is that clusters 2 and 3 seem to stand out more in their distribution, potentially indicating distinct clinical subgroups.

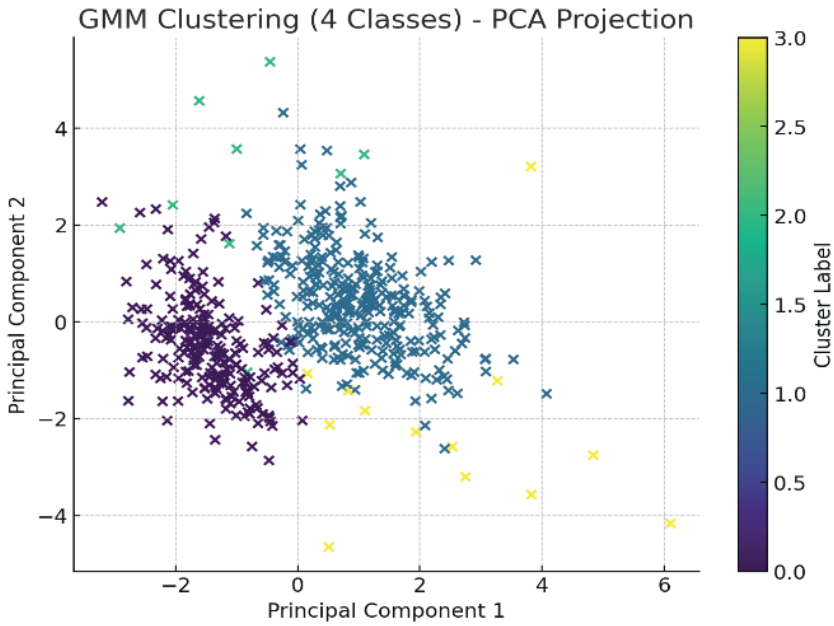

Figure 8.2. GMM Clustering (4 Classes) - PCA Projection

Figure 3 shows the boxplot patient age is distributed across the four GMM clusters. Cluster 0 has lower standardized age values, indicating younger patients, while Clusters 1 and 3 have higher medians. This difference may point to various biological profiles or diagnosis stages among age groups.

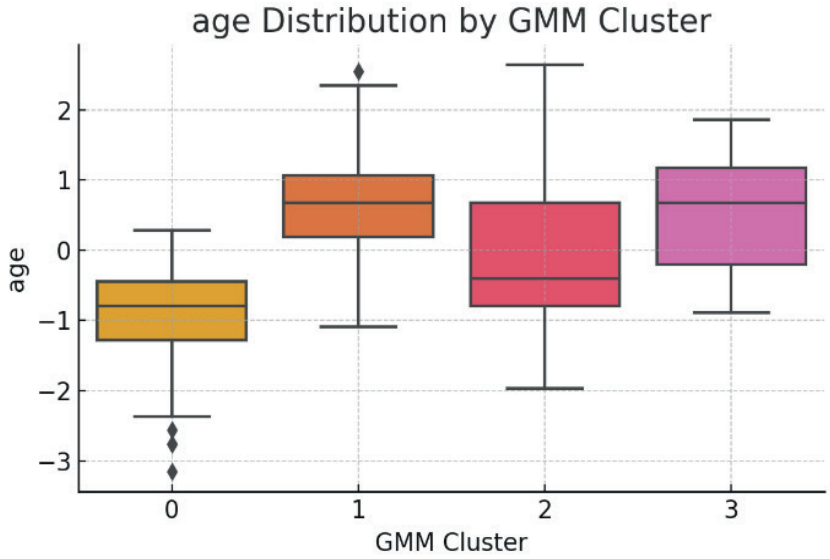

Figure 8.3. Age Distribution by GMM Cluster

Figure 8.4 present of the tumor size distribution in each cluster. While Cluster 2 has significantly larger values for tumor sizes than the rest, it might represent the advanced stage of cancer. Clusters 0, 1, and 3 have relatively small and similar tumor sizes.

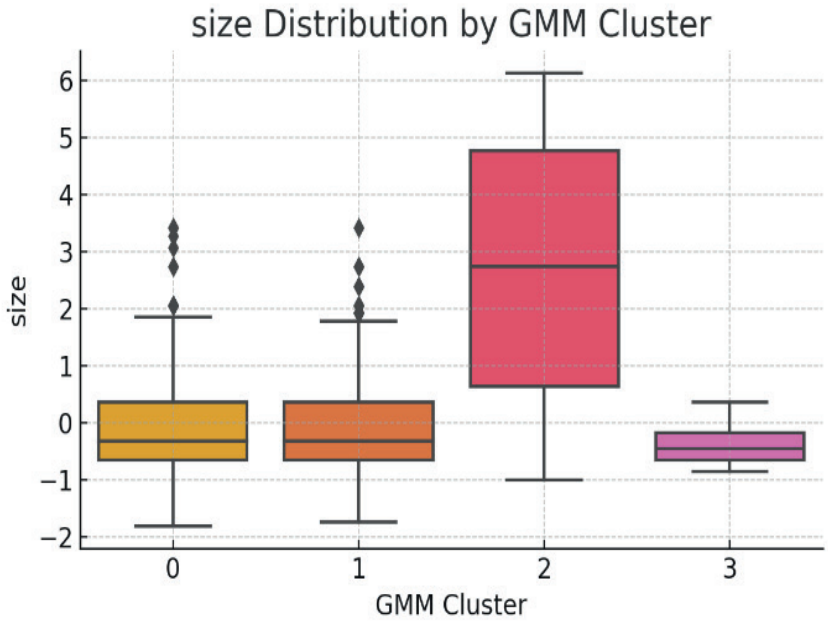

Figure 8.4. Tumor Size Distribution by GMM Cluster

Figure 8.5 illustrates the distribution of positive lymph nodes in GMM-defined clusters, a key indicator of the progression of breast cancer. The highest lymph node involvement is in Cluster 2, indicating a subgroup with more severe disease and higher metastatic potential. Low lymph node presence in Cluster 3 suggests a less aggressive clinical phenotype. Clusters 0 and 1 exhibit moderate involvement with less distributions.

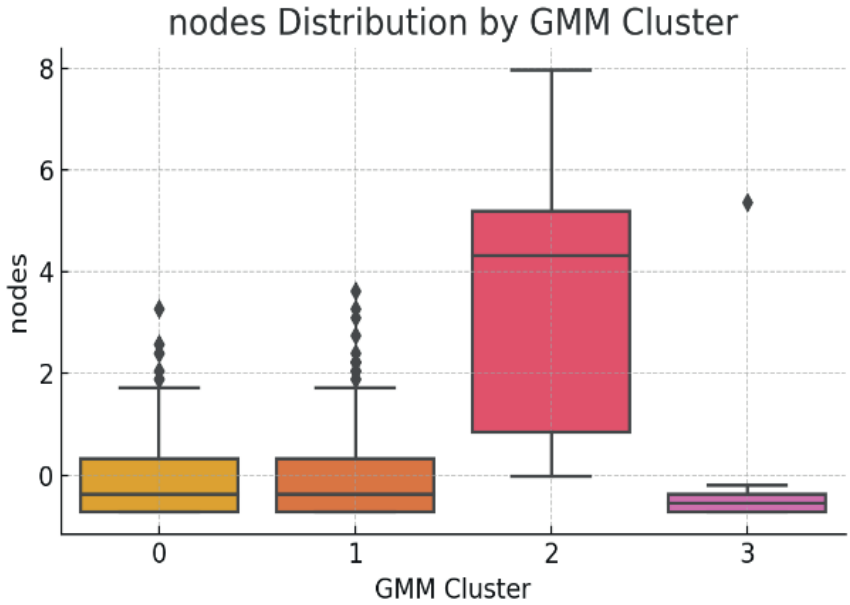

Figure 8.5. Number of Lymph Nodes by GMM Cluster

Figure 8.6 expression this case of PGR levels is being contrasted between clusters. Cluster 3 has the highest PGR values, which are associated with hormone-responsive phenotypes of breast cancer. Clusters 0 and 1 have lower values that might be representative of hormone-insensitive cases.

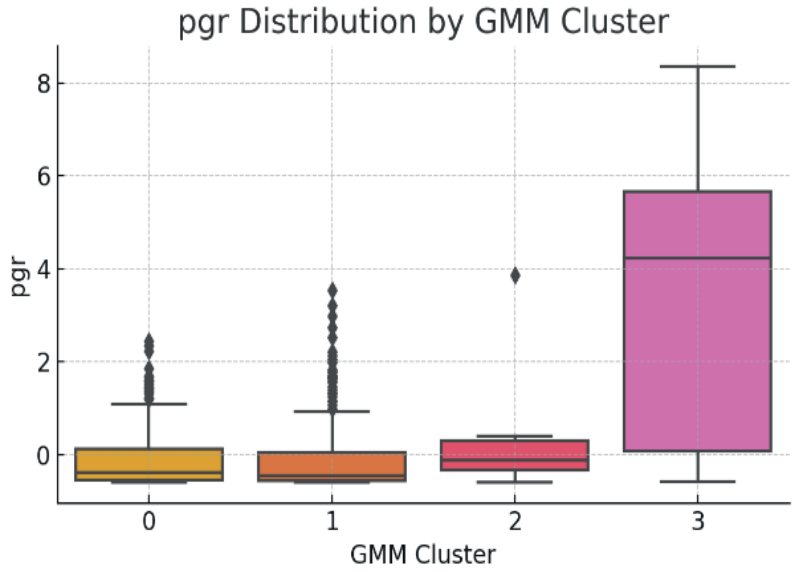

Figure 8.6. Progesterone Receptor (PGR) Levels by GMM Cluster

Figure 8.7 illustrates ER expression by clusters. Again, Cluster 3 expresses high ER, as expected to confirm its likely classification as hormone sensitive. Clusters 0, 1, and 2 express very low ER levels, suggesting other underlying tumor biology.

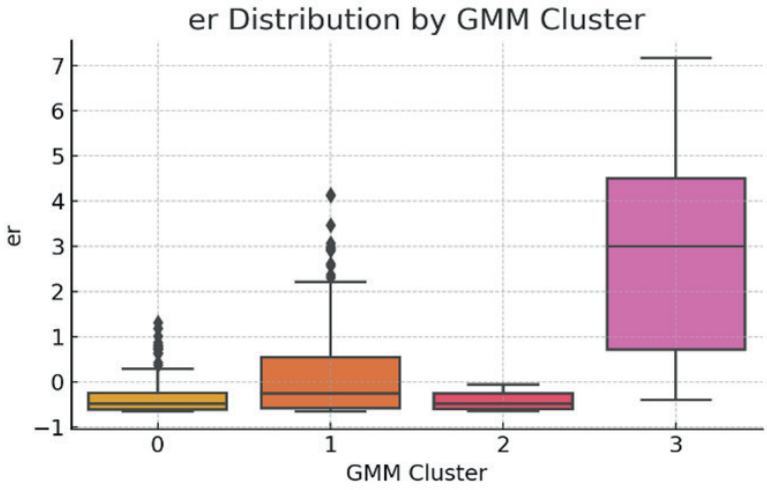

Figure 8.7. Estrogen Receptor (ER) Levels by GMM Cluster

CONCLUSIONS

GMM has been applied to clinical data to categorize patients with PBC based on a variety

of biological and prognostic factors. For instance, the model can be extended to incorporate features such as tumor size, pS2 protein, hormone receptor levels (ER, PgR), age, and menopausal status. GMM is naturally probabilistic, allowing the identification of subgroups with overlapping features that might be more difficult to identify with hard clustering, such as K-means. The method enables photomapping, a process by which diverse symptoms are divided into more homogeneous subtypes, by modeling the data distribution as a mixture of Gaussians. This may have the potential to inform personalized treatment decisions. Apart from numerical clinical data, GMM is applied to imaging for the classification of breast cancer. For instance, in the pre-processing step of histopathological image analysis, GMM is used to differentiate between foreground and background regions in order to segment out tissue portions. A two-component GMM efficiently models pixel intensity distributions, enabling accurate tissue segmentation prior to comprehensive feature extraction using CNNs such as VGG16[28,29,30] and dimensionality reduction using Principal Component Analysis. This segmentation enhances the quality and clarity of subsequent clustering and classification tasks by helping to isolate regions of interest. However, despite these advantages, the effectiveness of GMM is extremely dependent upon judicious choice of the number of clusters and scrupulous handling of its vulnerability to outliers and non-Gaussian data distributions. High-dimensional or extensive datasets pose a computing challenge because the model requires computationally expensive iterative parameter estimation. In cases where asymmetries and skewed distributions cannot be modeled correctly due to violation of the Gaussian assumption, other models have been explored, for example, Libby-Novick Beta Mixture Models. Moreover, GMM has been compared with many clustering approaches: while K-means remains one of the most frequent methods in clinical applications due to its simplicity, Gaussian Mixture Models offer greater flexibility by modeling probabilistic membership in nuancing subgroup boundaries. GMM has been employed in a number of studies on cluster validation techniques (ClValid or OptCluster) which evaluate cluster validity and determine an appropriate number of clusters, thereby ensuring that the identified patient subgroups are clinically significant.

Recommendation and Future direction

Although there was little agreement with real clinical outcomes, the present study demonstrated that GMMs are capable of uncovering hidden subgroup structures in breast cancer. On the basis of these findings, future investigations should proceed with the systematic refinement of the clustering process using sophisticated methods of dimensionality reduction such as UMAP or autoencoders to reduce redundancy and enhance separability. Additionally, robust model selection criteria-eg, BIC, AIC, silhouette scores-can be applied to determine the optimal number of clusters. Furthermore, considering mixture models that go beyond the Gaussian assumption, such as Dirichlet Process Mixture Models or Beta Mixtures, may enhance adaptability to the non-Gaussian distributions often found in biomedical datasets. Another important suggestion is to enhance the clinical interpretability of clusters. While hidden structures could be identified through unsupervised learning, their deployment in clinical practice rests on the degree to which they correlate with biological markers and prognostic factors. Cluster relevance may thus be validated by integrating GMM results with established markers such as ER, PR, HER2 and IGF1R expression. Moreover, embedding GMM findings within semi-supervised or weakly supervised learning frameworks-where a small amount of labelled data guides the clustering procedure-may enhance conformity to known diagnostic categories while preserving the flexibility of unsupervised discovery. Several interesting avenues of further study appear. Firstly, the application of this approach to the integration of multi-omics data-genomics, proteomics and transcriptomics-may enhance the biological basis of the subtypes which have already been identified. Secondly, in order to capture nonlinear feature interactions and enhance subgroup resolution, hybrid deep learning-GMM models-such as variational

autoencoder clustering-should be explored. Thirdly, the application of these methods to long-term patient data may reveal trends in the course of illness and the dynamics of treatment response, and cross-cohort validation across diverse demographics will enhance generalisability. Finally, integration of GMM-based clustering into AI-driven clinical decision support systems may enable personalised therapy recommendations, thereby transforming unsupervised learning insights into useful precision oncology tools.

REFERENCES

1. Turashvili G, Brogi E. Tumor heterogeneity in breast cancer. Front Med. 2017;4:227.

2. Vieira AF, Schmitt F. An update on breast cancer multigene prognostic tests—emergent clinical biomarkers. Front Med. 2018;5:248.

3. Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. Transl Cancer Res. 2015;4:256–69.

4. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol. 2010;28(16):2784–95.

5. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human epidermal growth factor receptor 2 testing in breast cancer: ASCO/CAP clinical practice guideline focused update. Arch Pathol Lab Med. 2018;142(11):1364–82.

6. Gong Y, Yan K, Lin F, Anderson K, Sotiriou C, Andre F, et al. Determination of estrogen receptor and ERBB2 status of breast carcinoma: a gene-expression profiling study. Lancet Oncol. 2007;8(3):203–11.

7. Roepman P, Horlings HM, Krijgsman O, Kok M, Bueno-de-Mesquita JM, Bender R, et al. Microarray-based determination of estrogen receptor, progesterone receptor, and HER2 receptor status in breast cancer. Clin Cancer Res. 2009;15(22):7003–11.

8. Kim C, Tang G, Pogue-Geile KL, Costantino JP, Baehner FL, Baker J, et al. Estrogen receptor (ESR1) mRNA expression and benefit from tamoxifen in ER-positive breast cancer. J Clin Oncol. 2011;29(31):4160–7.

9. Mihaly Z, Gyorffy B. Improving pathological assessment of breast cancer by employing array-based transcriptome analysis. Microarrays. 2013;2:228–42.

10. Peiro G, Adrover E, Sánchez-Tejada L, Lerma E, Planelles M, Sánchez-Payá J, et al. Increased insulin-like growth factor-1 receptor mRNA predicts poor survival in early breast carcinoma. Mod Pathol. 2011;24(2):201–8.

11. Fu P, Ibusuki M, Yamamoto Y, Yamamoto S, Fujiwara S, Murakami K, et al. Quantitative determination of IGF-1R mRNA in FFPE tissues of invasive breast cancer. Breast Cancer. 2012;19(4):321–8.

12. Hofmann E, Seeboeck R, Jacobi N, Obrist P, Huter S, Klein C, et al. Laser-captured microdissection and RT-qPCR accurately determine HER2 status in breast cancer. Biomark Res.

2016;4(1):8.

13. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst. 2003;95:14–8.

14. Marko NF, Weil RJ. Non-Gaussian distributions affect expression pattern identification and classification in cancer genomes. PLoS One. 2012;7:e46935.

15. Wang X, Zhang J, Yan Y. Discrimination between pathological and normal voices using GMM-SVM. J Voice. 2011;25:38–43.

16. Ban Z, Liu J, Cao L. Superpixel segmentation using Gaussian mixture model. IEEE Trans Image Process. 2018;27:4105–17.

17. Polanski A, Marczyk M, Pietrowska M, Widlak P, Polanska J. Signal partitioning algorithm for efficient Gaussian mixture modeling in mass spectrometry. PLoS One. 2015;10:e0134256.

18. Kawabata T. Multiple subunit fitting in low-resolution cryo-EM maps using Gaussian mixture model. Biophys J. 2008;95:4643–58.

19. Gianola D, Heringstad B, Odegaard J. On the quantitative genetics of mixture characters. Genetics. 2006;173:2247–55.

20. McLachlan GJ, Bean RW, Peel D. Mixture model-based approach to clustering microarray data. Bioinformatics. 2002;18:413–22.

21. Bonneterre J, Peyrat JP, Beuscart R, Demaille A. Prognostic significance of IGF-1 receptors in human breast cancer. Cancer Res. 1990;50:6931–5.

22. Yan S, Jiao X, Li K, Li W, Zou H. Impact of IGF-1R expression on outcomes in breast cancer: meta-analysis. OncoTargets Ther. 2015;8:279–87.

23. Yee D. Anti-insulin-like growth factor therapy in breast cancer. J Mol Endocrinol. 2018;61:T61–8.

24. Guvakova MA, Adams JC, Boettiger D. Role of alpha-actinin, PI3K and MEK1/2 in IGF-1R regulated motility of human breast carcinoma cells. J Cell Sci. 2002;115:4149–65.

25. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. ASCO 2007 update of tumor marker recommendations in breast cancer. J Clin Oncol. 2007;25(33):5287–312.

26. Furstenau DK, Mitra N, Wan F, Lewis R, Feldman MD, Fraker DL, et al. Ras-related protein 1 and IGF-1R are associated with progression risk in carcinoma in situ. Breast Cancer Res Treat. 2011;129(2):361–72.

27. Sharma U. Breast cancer dataset (Royston and Altman). Kaggle; 2023. Disponible en: https://www.kaggle.com/datasets/utkarshx27/breast-cancer-dataset-used-royston-and-altman

28. Dutta PK, Singh B, Lal S, Arora MK, Raghav A, Mensah GB, et al. Encoding IoT for patient

monitoring and smart healthcare. Babylonian J Internet Things. 2023;48–58. https://doi.org/10.58496/BJIoT/2023/007.

29. Ramadhan AJ, Krishna Priya SR, Naranammal N, Pavishya S, Naveena K, Ray S, et al. Comparison study using ARIMA and ANN for forecasting sugarcane yield. BIO Web Conf. 2024;97:00078. https://doi.org/10.1051/bioconf/20249700078.

30. Alkattan H, Abbas NR, Adelaja OA, Abotaleb M, Ali G. Data mining utilizing hierarchical clustering on worker ranking in IT firm. Mesopotamian J Comput Sci. 2024;104–9. https://doi.org/10.58496/MJCSC/2024/008.

31. Ramadhan AJ, Biswas T, Ray S, Anjanawe SR, Rawat D, Kumari B, et al. Modeling and forecasting coconut yield with time series model. BIO Web Conf. 2024;97:00113. https://doi.org/10.1051/bioconf/20249700113.

32. Alkattan H, Subhi AA, Farhan L, Al-Mashhadani G. Hybrid model for forecasting temperature in Khartoum. Mesopotamian J Big Data. 2024;164–74. https://doi.org/10.58496/MJBD/2024/011.

33. Almusallam N, Al Nuaimi BT, Alkattan H, Abotaleb M, Dhoska K. Physics-informed neural networks for solving heat equation. Int J Tech Phys Probl Eng. 2025;17(1):375–82.

CONFLICT OF INTEREST

The authors assert that there are no conflicts of interest related to the research results presented.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHORSHIP CONTRIBUTION

Conceptualization: Hussein Alkattan, Salam Abdulkhaleq Noaman, Abrar Ryadh, Mostafa Abotaleb, H.K. Al-Mahdawi, Elman Azimov, Pushan Kumar Dutta.

Writing – original draft: Hussein Alkattan, Salam Abdulkhaleq Noaman, Abrar Ryadh, Mostafa Abotaleb, H.K. Al-Mahdawi, Elman Azimov, Pushan Kumar Dutta.

Writing – proofreading and editing: Hussein Alkattan, Salam Abdulkhaleq Noaman, Abrar Ryadh, Mostafa Abotaleb, H.K. Al-Mahdawi, Elman Azimov, Pushan Kumar Dutta.