Comparative Analysis of Deep Learning Models for Denoising Hyperspectral Raman Spectroscopy Image Data in Breast Cancer Research

Abstract

Raman spectroscopy is a vital technique in biomedical research, especially for cancer diagnostics, offering non-destructive, label-free molecular imaging with high contrast. However, the weakness of Raman signals necessitates advanced denoising methods to enhance data quality in hyperspectral imaging for breast cancer research(FAROOQ & SAVAŞ, 2024; Mowbray et al., 2021). This study evaluates and compares deep learning algorithms, including Generative Adversarial Networks (GANs), Residual Attention Networks (RACN), and Denoising Autoencoders (DAEs), for denoising hyperspectral Raman images. Using a dataset focused on breast cancer cell samples, the research addresses noise challenges and aims to improve spectral data interpretability. Algorithm performance will be assessed using metrics like Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Mean Squared Error (MSE)(Horgan et al., 2021). Expected outcomes include enhanced image quality, a comparative analysis of algorithms, improved signal-to-noise ratio (SNR), and contributions to reliable diagnostic tools in clinical settings. This research aims to advance deep learning applications in Raman spectroscopy for more accurate biomedical diagnostics (Ma et al., 2022; Wang et al., 2021)

Introduction

Raman spectroscopy has become an important tool in research, especially in the realm of cancer detection thanks to its ability to offer detailed molecular images without the need for labels. However, the challenge lies in the weak Raman signals that often hinder its effectiveness leading to the necessity of noise reduction techniques to improve data quality (FAROOQ & SAVAŞ 2024). Dealing with noise in Raman spectroscopy images poses a challenge in breast cancer studies where precise molecular analysis is vital for accurate diagnosis and treatment planning (Mowbray et al., 2021).

This study's significance lies in its potential to enhance the reliability and precision of Raman spectroscopy as an aid. By evaluating and comparing deep learning algorithms for cleaning up hyperspectral Raman images this research aims to tackle noise-related limitations and enhance spectral data interpretation. Previous research has highlighted the effectiveness of learning models like Generative Adversarial Networks (GANs) and Denoising Autoencoders (DAEs) in enhancing image quality across fields (FAROOQ & SAVAŞ 2024; Zhang et al., n.d.). Nevertheless, there is a knowledge gap regarding how these models compare within Raman spectroscopy for breast cancer studies. The main goal of this study is to assess how well various learning techniques, such, as GANs, Residual Attention Networks (RACN), and

DAEs perform when cleaning up noisy hyperspectral Raman spectroscopy images. The study aims to explore the following research queries:

1. How do different deep learning models stack up in improving the clarity of denoised hyperspectral Raman images?

2. What are the consequences of image quality for identifying characteristics linked to breast cancer?

3. How might these discoveries aid in creating diagnostic tools for medical facilities?

Through tackling these inquiries, this investigation strives to provide perspectives on applying deep learning approaches in Raman spectroscopy paving the way for precise and dependable applications in the realm of biomedical research (DePaoli et al., 2020; Zhang et al., n.d.).

Literature Review

Raman spectroscopy has become increasingly popular in the field of research because it can offer invasive imaging of molecules without the need, for labels while providing strong molecular contrast. However, one common challenge is dealing with Raman signals, which require denoising techniques to improve the quality of data. . The anchor paper, "High-Throughput Molecular Imaging via Deep-Learning-Enabled Raman Spectroscopy" (Horgan et al., 2021)) introduces an approach to enhancing Raman imaging using deep learning methods specifically focusing on using a one Residual U Net (1D ResUNet) for denoising spectra with low signal-to-noise ratios (SNRs). This research lays the groundwork for investigations into deep learning models that aim to improve spectral data quality.

The use of Generative Adversarial Networks (GANs) for denoising has shown outcomes. GANs are effective in generating high-quality data from low-quality inputs by utilizing a generator discriminator framework that learns to differentiate between real and generated data sets(Bench et al., 2023). This capability is crucial in Raman spectroscopy as noise can mask characteristics. Previous studies have proven that GANs can significantly enhance the SNR of Raman spectra making them a valuable resource, for improving data quality in biomedical contexts(Ma et al., 2022).

In addition, to Generative Adversarial Networks (GANs) the Residual Attention Network (RACN) has emerged as an architecture for tasks related to image processing, such as denoising. The RACN utilizes attention mechanisms to emphasize the features in the input data thereby enhancing the model's capability to generate high-quality results from noisy inputs (Zhang et al., 2018). This method aligns well with the requirements of Raman spectroscopy, where maintaining characteristics is vital for precise analysis.

For denoising, Denoising Autoencoders (DAEs) offer another approach. DAEs are crafted to understand the transformation from inputs to clean outputs by minimizing reconstruction errors. They have been effectively utilized in imaging scenarios showcasing their ability to retain features while effectively eliminating noise (Han et al., 2024). Integrating DAEs with networks (CNNs) has further boosted their performance levels enabling more robust denoising capabilities (FAROOQ & SAVAŞ 2024).

Literature indicates a rising inclination towards employing learning methodologies for denoising hyperspectral Raman data. The amalgamation of these models, like 1D ResUNet, GANs, RACN, and DAEs provides an opportunity to systematically compare their efficacy in enhancing precision.

Based on the anchor paper's discoveries, this study seeks to assess how well these models can clean up Raman spectra to advance Raman spectroscopy in research and diagnostics. In essence, incorporating learning techniques into Raman spectroscopy could greatly improve the accuracy of data. Examining denoising algorithms as outlined in this research will offer insights into their effectiveness and open doors for more precise and dependable uses in biomedical research.

Methodology

Research Design

This research study uses an approach to assess how well different deep learning algorithms can clean up noisy hyperspectral Raman spectroscopy images. The main focus is, on comparing how various algorithms, such, as Generative Adversarial Networks (GANs) Residual Attention Networks (RACN) Denoising Autoencoders (DAEs), and the known 1D Residual U Net (1D ResUNet) mentioned in the paper "High Throughput Molecular Imaging via Deep Learning Enabled Raman Spectroscopy" (Horgan et al., 2021) perform in reducing noise and enhancing image quality. This methodology aims to provide an evaluation of the algorithm's effectiveness.

Approach

The study utilizes a method which is well suited for this research as it allows for gathering and analyzing numerical data to reach statistically significant findings (Noor & Ige 2024). Through the use of quantitative approaches, the research seeks to offer an assessment of the noise reduction abilities of the chosen algorithms thereby making a valuable contribution to the field of biomedical imaging (Wang et al., 2021).

Justification of Methods

The methods chosen are in line with the research objectives of improving the quality of hyperspectral Raman spectroscopy images crucial for analysis in biomedical fields (Leng et al., 2014). The decision to employ learning algorithms is supported by their established success in image processing tasks especially in noise reduction applications (Ma et al., 2022). Each algorithm selected for comparison possesses advantages; GANs are recognized for generating high-quality results from inputs (Horgan et al., 2021) RACN uses attention mechanisms to focus on key features (Wang et al., 2021) and DAEs excel at reconstructing clean images from noisy data (Chen & Qian 2011). The 1D ResUNet acts as a standard, for assessing the performance of the algorithms.

Data Collection Techniques and Analysis Methods

In this research, we used the available data mentioned in the original paper titled "High Throughput Molecular Imaging via Deep Learning Enabled Raman Spectroscopy" (Horgan et al., 2021). The dataset consists of hyperspectral Raman spectroscopy images focused specifically on samples of breast cancer cells.

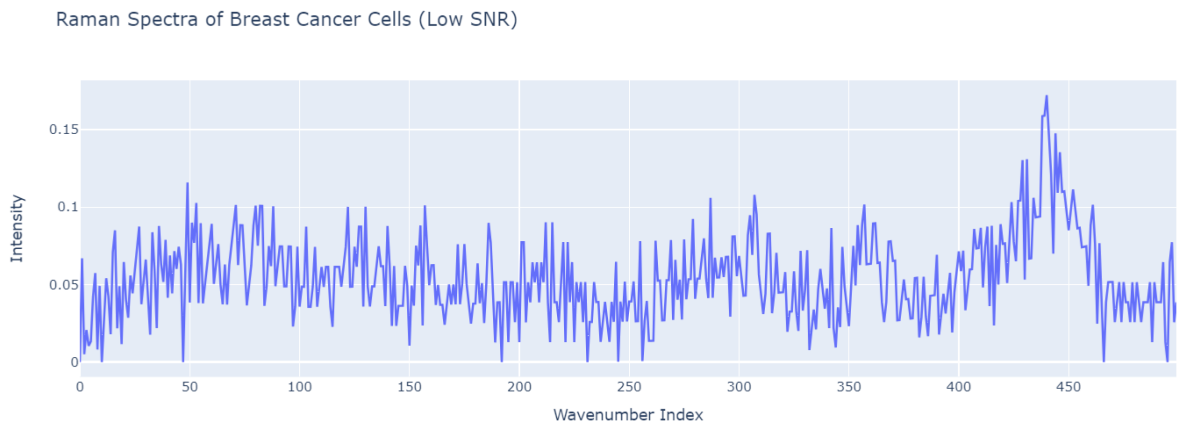
The process of collecting data involved these steps:

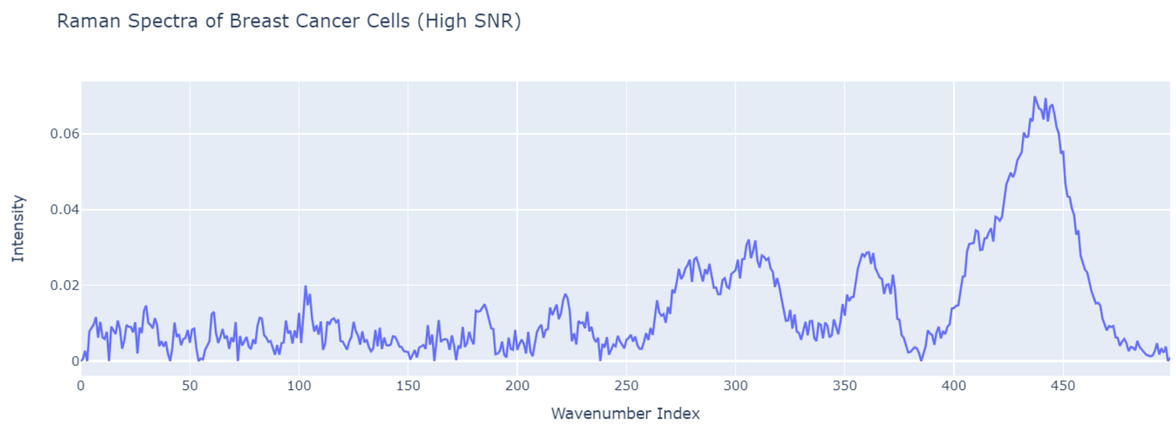
1. Sample Preparation; Tissue samples, from breast cancer were set up appropriately for Raman spectroscopy analysis to ensure they accurately represent the variations found in clinical settings.
2. Hyperspectral Imaging; The hyperspectral Raman spectra were captured using a Raman spectrometer with a laser source. The imaging took place under controlled conditions to reduce noise and maintain the accuracy of the data.
3. Data Collection; The data was gathered systematically, capturing a variety of information across wavenumber ranges. This allowed for an analysis of the properties present, in breast cancer tissues.

Here are the specifics of the dataset:

- Dataset Size: The dataset consists of 172,312 Raman spectra obtained from 11 hyperspectral Raman images.
- Structure: The dataset includes both low SNR and high SNR spectra, where the low SNR data is recorded over shorter acquisition times leading to higher noise levels, and high SNR data is recorded over longer acquisition times, reducing noise and resulting in clearer spectral data.
- Training/Validation/Testing Split: The dataset was split using 11-fold leave-one-image-out cross-validation. This means:
- Training/Validation: Raman spectra from 10 images (with a 90:10 split).
- Testing: Raman spectra from the remaining one image.
- Spectral Range: Each Raman spectrum covers the range from 0 to 3700 cm^{-1} with a spectral resolution of 11 cm^{-1}.
- Imaging Details: The hyperspectral Raman images were captured with a spatial resolution of 0.5 µm using a confocal Raman microscope equipped with a 532 nm laser light source.

To illustrate the differences, here is a visualization of one low SNR and one high SNR spectrum from the dataset. This visualization was created using Plotly in a Python Jupyter Notebook.





After training the models, we will assess the performance of each noise reduction technique using measurement criteria, which include;

1. Peak Signal to Noise Ratio (PSNR); This measure indicates the relationship between the signal power and the noise power affecting the quality of the reconstructed image (Chen & Qian 2011).
2. Structural Similarity Index Measure (SSIM); This criterion evaluates how well the reconstructed image matches the reference image (Chen & Qian 2011).
3. Mean Squared Error (MSE); This metric calculates the average of error squares between the cleaned images offering insights into denoising accuracy (Horgan et al., 2021).

In essence, this approach aims to evaluate deep learning algorithms' effectiveness in denoising hyperspectral Raman spectroscopy images in line with research objectives focused on enhancing image quality, for better biomedical diagnostics (Leng et al., 2014).

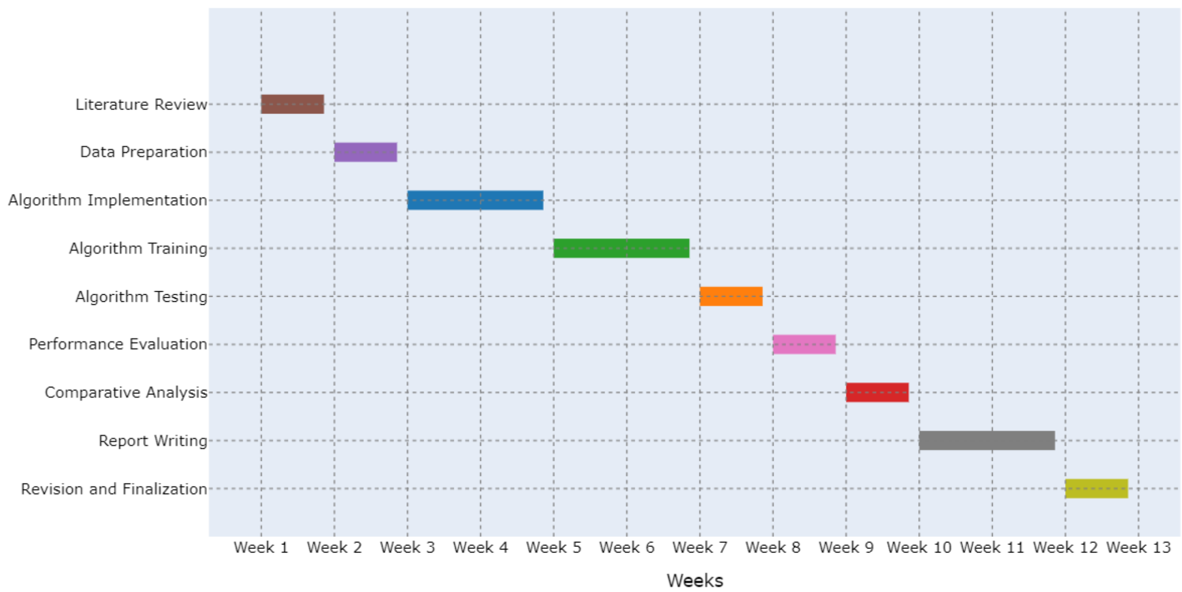
Expected Outcomes

The intended study aims to assess and compare the performance of advanced deep learning techniques, in cleaning up noisy hyperspectral Raman spectroscopy images within the field of breast cancer research. The anticipated results of this investigation are as follows;

1. Improved Image Clarity: By employing learning models a significant enhancement in the clarity of cleaned-up hyperspectral Raman images is expected, as evaluated by metrics like Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).
2. Comparative Analysis of Algorithms: The study seeks to offer a comparison of deep learning algorithms highlighting their respective strengths and weaknesses in spectral noise reduction. This analysis will contribute to an understanding of their efficacy in applications.
3. Enhanced Signal to Noise Ratio (SNR); It is predicted that specific algorithms, Generative Adversarial Networks (GANs) will effectively boost the SNR of Raman spectra enabling improved detection of characteristics relevant to breast cancer.
4. Impact on Clinical Utilization; The expected findings could potentially pave the way for diagnostic tools in clinical environments augmenting the utilization of Raman spectroscopy for real-time diagnosis and treatment monitoring, among breast cancer patients.

Timeline and Resources Requirements

Below is the detailed timeline for the completion of the research project:



The successful completion of this research project requires the following resources:

1. Computational Resources:
    a. Access to a high-performance computing (HPC) cluster or GPU-enabled workstations for training models.
    b. Sufficient storage for large datasets.
2. Software and Tools:
    a. Deep learning frameworks like TensorFlow, PyTorch, or Keras.

   b. Data processing libraries (NumPy, pandas, sci-kit-learn).
   c. Visualization and image processing software (Python Matplotlib, Seaborn).
3. Data Resources:
   a. Open-source dataset of hyperspectral Raman spectroscopy images focused on breast cancer cell samples.

Ethical Considerations

This research will adhere to strict ethical guidelines. Data privacy will be ensured by using anonymized, open-source datasets stored securely to prevent unauthorized access. Although participant consent is not directly required, all data used will comply with original consent agreements and ethical guidelines.

The data will be used solely for research purposes, with proper citations provided. Methods and procedures will be documented for transparency and reproducibility, and source code will be made available. Efforts will be made to avoid bias in model evaluation, using unbiased performance metrics.

Conclusion

This proposal presents a comparative analysis of deep learning models used to enhance the quality of hyperspectral Raman spectroscopy images, in breast cancer research. By assessing algorithms such as Generative Adversarial Networks (GANs) Residual Attention Networks (RACN) and Denoising Autoencoders (DAEs) the study aims to tackle the challenges posed by noise in hyperspectral imaging. The research emphasizes the importance of improving image clarity and boosting signal-to-noise ratios (SNR) to enable characterization, crucial for effective diagnosis and treatment planning in breast cancer (FAROOQ & SAVAŞ 2024; Mowbray et al., 2021).

The significance of this analysis lies in its potential to advance the use of learning techniques in Raman spectroscopy ultimately aiding in creating dependable diagnostic tools for clinical applications. Anticipated outcomes include enhanced image quality, improved retention, and a thorough comprehension of each algorithm's strengths and weaknesses. This research not only aims to enhance the interpretability of spectral data but also seeks to pave the way for more accurate and reliable applications in biomedical diagnostics (Horgan et al., 2021; Ma et al., 2022).

By tackling the constraints of existing methods and delving into deep learning strategies this research could have an effect, on the field of breast cancer research and enhance the well-being of patients (DePaoli et al., 2020; Wang et al., 2021).

References

Bench, C., Bergholt, M. S., & al-Badri, M. A. (2023). Unsupervised denoising of Raman spectra with cycle-consistent generative adversarial networks. http://arxiv.org/abs/2307.00513

Chen, G., & Qian, S. E. (2011). Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. IEEE Transactions on Geoscience and Remote Sensing, 49(3), 973–980. https://doi.org/10.1109/TGRS.2010.2075937

DePaoli, D., Lemoine, É., Ember, K., Parent, M., Prud'homme, M., Cantin, L., Petrecca, K., Leblond, F., & Côté, D. C. (2020). Rise of Raman spectroscopy in neurosurgery: a review. Journal of Biomedical Optics, 25(05), 1. https://doi.org/10.1117/1.jbo.25.5.050901

FAROOQ, Y., & SAVAŞ, S. (2024). Noise Removal from the Image Using Convolutional Neural Networks-Based Denoising Auto Encoder. Journal of Emerging Computer Technologies, 3(1), 21–28. https://doi.org/10.57020/ject.1390428

Han, M., Dang, Y., & Han, J. (2024). Denoising and Baseline Correction Methods for Raman Spectroscopy Based on Convolutional Autoencoder: A Unified Solution. Sensors, 24(10). https://doi.org/10.3390/s24103161

Horgan, C. C., Jensen, M., Nagelkerke, A., St-Pierre, J. P., Vercauteren, T., Stevens, M. M., & Bergholt, M. S. (2021). High-Throughput Molecular Imaging via Deep-Learning-Enabled Raman Spectroscopy. Analytical Chemistry, 93(48), 15850–15860. https://doi.org/10.1021/acs.analchem.1c02178

Leng, L., Wang, Y., He, N., Wang, D., Zhao, Q., Feng, G., Su, W., Xu, Y., Han, Z., Kong, D., Cheng, Z., Xiang, R., Li, Z., & Cn, : Zongjinli@nankai Edu. (2014). Molecular imaging for assessment of mesenchymal stem cells mediated breast cancer therapy HHS Public Access. Biomaterials, 35, 5162–5170. https://doi.org/10.1016/j.biomaterials

Ma, X., Wang, K., Chou, K. C., Li, Q., & Lu, X. (2022). Conditional Generative Adversarial Network for Spectral Recovery to Accelerate Single-Cell Raman Spectroscopic Analysis. Analytical Chemistry, 94(2), 577–582. https://doi.org/10.1021/acs.analchem.1c04263

Mowbray, M., Savage, T., Wu, C., Song, Z., Cho, B. A., Del Rio-Chanona, E. A., & Zhang, D. (2021). Machine learning for biochemical engineering: A review. Biochemical Engineering Journal, 172. https://doi.org/10.1016/j.bej.2021.108054

Noor, M. H. M., & Ige, A. O. (2024). A Survey on Deep Learning and State-of-the-art Applications. http://arxiv.org/abs/2403.17561

Wang, L., Ren, B., He, H., Yan, S., Lyu, D., Xu, M., Ye, R., Zheng, P., & Lu, X. (2021). Deep learning for biospectroscopy and biospectral imaging: State-of-the-art and perspectives. Analytical Chemistry, 93(8), 3653–3665. https://doi.org/10.1021/acs.analchem.0c04671

Zhang, Y., Hong, H., & Cai, W. (n.d.). Imaging with Raman Spectroscopy.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image Super-Resolution Using Very Deep Residual Channel Attention Networks. http://arxiv.org/abs/1807.02758

Kaur, A., & Dong, G. (2023). A Complete Review on Image Denoising Techniques for Medical Images. *Neural Processing Letters*, *55*(6), 7807–7850. https://doi.org/10.1007/s11063-023-11286-1

Cialla-May, D., Krafft, C., Rösch, P., Deckert-Gaudig, T., Frosch, T., Jahn, I. J., Pahlow, S., Stiebing, C., Meyer-Zedler, T., Bocklitz, T., Schie, I., Deckert, V., & Popp, J. (2022). Raman Spectroscopy and

Imaging in Bioanalytics. In *Analytical Chemistry* (Vol. 94, Issue 1, pp. 86–119). American Chemical Society. https://doi.org/10.1021/acs.analchem.1c03235

Gupta, M., Goel, A., Goel, K., & Kansal, J. (2023). Medical Image Denoising using Convolutional Autoencoder with Shortcut Connections. *Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023*, 1524–1528. https://doi.org/10.1109/ICSSIT55814.2023.10061131

Yadav, V., Tiwari, A. K., & Siddhanta, S. (n.d.). *Machine learning driven high-resolution Raman spectral generation for accurate molecular feature recognition*.

Jayan, H., Pu, H., & Sun, D. W. (2022). Recent developments in Raman spectral analysis of microbial single cells: Techniques and applications. In *Critical Reviews in Food Science and Nutrition* (Vol. 62, Issue 16, pp. 4294–4308). Taylor and Francis Ltd. https://doi.org/10.1080/10408398.2021.1945534

Lee, H., & Cho, S. (2020). Locally Adaptive Channel Attention-Based Network for Denoising Images. *IEEE Access*, *8*, 34686–34695. https://doi.org/10.1109/ACCESS.2020.2974001

Wu, W., Ge, A., Lv, G., Xia, Y., Zhang, Y., & Xiong, W. (2024). *Two-stage Progressive Residual Dense Attention Network for Image Denoising*. http://arxiv.org/abs/2401.02831