

In the realm of genomics, the intricacies of data often challenge researchers, particularly when addressing the nuanced task of clustering less common or underrepresented DNA sequences. Our study introduces an advanced unsupervised learning framework, prominently integrating the capabilities of the Fast Fourier Transform (FFT). FFT, central to our methodology, enables the transformation of genomic sequences into a spectral format. This isn't merely a technical transition; it ushers in a fresh, insightful perspective for biological interpretation. With this spectral viewpoint, we can discern previously overlooked details and patterns within the sequences, shedding light on potentially crucial biological dynamics. Taking the advantage of FFT spectrum, the model is able to better utilize the subtle differences between data and also further help the data generation. Considering the sheer volume and complexity of genomic data, our approach further assimilates strategic dimension reduction techniques. These methods not only enhance computational speed but also complement the FFT-centric process, optimizing data management and ensuring a more effective unsupervised learning journey. In contemporary genomics, it's not uncommon for methods to lean on simulated or mimic-based sequences when faced with limited original data. In contrast, our strategy is rooted in the use of authentic genomic data, aiming to provide more genuine and relatable biological insights. Initial evaluations underscore the promise and adaptability of our framework. These preliminary findings not only validate its efficacy in clustering tasks but also suggest its potential as a reference point for future bioinformatics endeavours, bridging the gap between academic rigour and broader comprehensibility.