j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions

[1*]Ammar A. Sakran, [2]Suha. M. Hadi, [3]Waleed A. Mahmoud Al-Jawher

[1&2] *Informatics Institute for Postgraduate Studies, Iraqi Commission for Computer and Informatics. Iraq*

[3]*College of Engineering, Uruk University, Baghdad, Iraq.*

Phd202130675@iips.icci.edu.iq

Abstract Within the complex realm of DNA sequencing, discerning protein coding areas from non-coding segments proves challenging due to the pervasive 1/f background disturbance. Traditional digital signal processing (DSP) methodologies, while widely adopted, may inadvertently overlook the inherent nuances and intricacies of DNA sequences. This paper critically examines these established DSP-centric methodologies, underscoring their potential inadequacies in capturing the salient characteristics intrinsic to DNA. Notably, nucleotides within the DNA exhibit distinct attributes, such as their triadic configurations, specific structural significance, and particularized density distributions in codons, among other characteristics. By harnessing these inherent features of nucleotides, computational approaches can effectively counteract signal disruptions, enhancing the precision in identifying protein coding regions.

Crossref 10.36371/port.2023.4.8

Keywords: *DNA Signal Processing, Coding Region Identification, Digital-Biological Integration, 1/f Noise Mitigation, 3-Base Periodicity Analysis, Advanced DSP Techniques.*

1. INTRODUCTION

DNA is the quintessential blueprint, safeguarding genetic information across living entities [1]. Constituted by four foundational chemical bases—Adenine (A), Thymine (T), Guanine (G), and Cytosine (C)—these nucleotide bases form the bedrock of genetic coding [2]. Structurally, the DNA is conceived as a double-stranded helix, characterized by the complementary base pairing of A with T and G with C, crafting its molecular framework [1,2].

Given the dual-stranded architecture and the complementary affiliation of the A-T and G-C pairs, analytical pursuits predominantly concentrate on a singular DNA strand [3]. In this context, a DNA sequence manifests as an intricate, sequential assembly of these nucleotide bases, exemplified as "ATCCGATCGATCCTAGGCAATG..." [1]. Despite the near-consistency of DNA across an organism's cellular framework, species differentiation arises from the nuanced permutations of these nucleotide bases [3].

Encompassing close to three billion bases, human DNA enciphers between 20,000 and 25,000 genes [11,12]. These genetic scripts comprise both protein-coding segments (exons) and intervening non-coding sequences (introns) [4]. Segregating these distinct zones presents a formidable computational quandary, often earmarked as a perennial optimization conundrum within bioinformatics [2].

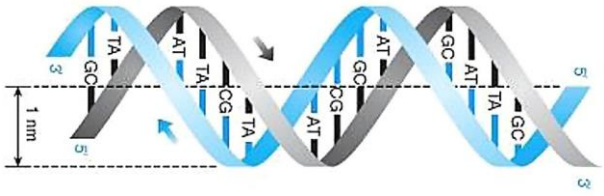

***Fig 1** : DNA structure*

Academic inquiries have delineated characteristic correlations, rhythmic patterns, and sequential markers within DNA, notably between exon and intron regions [2,4]. A noteworthy observation is the tri-nucleotide periodicity discernible within exons, contrasting starkly with the introns' sequence spectrum [3]. This triadic attribute has been pivotal for coding region demarcation, underpinning both temporal and spectral methodologies for exon determination [5].

The realm of Digital Signal Processing (DSP) has permeated genetic sequence analytics, predominantly in pinpointing protein-coding domains [1]. This modus operandi translates nucleotide sequences into quantifiable metrics, further molded into time-domain signals for nuanced analysis. Nonetheless, these stratagems may inadvertently bypass the profound biological intricacies inherent to DNA. The

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

subsequent sections endeavor to illuminate the potential shortcomings of these prevalent DSP methodologies while underscoring the inherent biological profundity of DNA sequences [6].

Nucleotide Encoding Strategies (Indicator Sequences)

Nucleotide encoding frameworks are pivotal for the digital transformation of DNA sequences. A plethora of these encoding strategies, as discerned from academic literature, facilitates the translation of nucleotides into a structured numerical paradigm for refined analysis [7].

Tetrahedral Encoding Paradigm

Within this framework, nucleotide representation remains invariant, leveraging the spatial alignment of bases on a tetrahedron [8]. This approach discerns periodicities in DNA segments, where vectors originating from the tetrahedron's center to its vertices symbolize distinct nucleotides.

Quartet Binary Encoding Strategy

This scheme, presented as a neural network optimization mechanism [9], employs a 4-bit binary encoding system. Herein, nucleotides A, C, G, and T are numerically represented as 1000, 0010, 0001, and 0100, respectively.

Dichotomous Encoding Methodology

Voss's proposition [10] entails a binary indicator sequence, wherein '1' or '0' signifies the presence or absence of a specific nucleotide in the strand. For instance, the sequence x[n] = [T T A G G T C C T] would be encoded as [001000000] for adenine. Consequently, the summation of all binary indicator sequences invariably equals 1.

Molecular Mass-Inspired Encoding Scheme

This paradigm [11] correlates nucleotides with their respective molecular masses: C = 110, G = 150, A = 134, and T = 125.

Z-Curve Encoding Framework

The Z-curve representation [12], emerging as a Fourier-transform modality, encodes DNA sequences leveraging the Z-curve format, producing signals resonant with these curves.

Pathogenicity Island Encoding Model

The authors' representation [3] confers a value of 1 to C and G bases, while A and T are encoded as 0, enhancing the detection of specific G+C patterns in genomic sequences.

Entropic Segment Encoding Modality

This method [13] employs a binary representation of a 12-letter nucleotide structure, aimed at discerning larger DNA segments for pattern recognition.

Dual Nucleotide Binary Representation

Introduced by the authors [14], this approach employs binary values for DNA nucleotides, with three unique conventions differentiating base weights.

Integer Encoding Mechanism

This methodology [8] relies on an integer representation of genetic signals, mapping C = 1, G = 3, A = 2, T = 0 to respective nucleotides.

Autoregressive Encoding Strategy

This approach is grounded in the inherent structural properties of DNA sequences, utilizing metrics like propeller twist and DNA-bending rigidity [10].

Gradient Source Localization Encoding Scheme

This biologically-motivated encoding mechanism [11] utilizes nucleotide weights such as A = 0, C = 1, G = 3, and T = 2.

Electron-Ion Interaction Potential (EIIP) Framework

In this model [2], the binary sequences as described by Voss [1,3] are supplanted by a singular EIIP sequence, resulting in a computation reduction of 75%. The scheme ascribes specific values to nucleotides.

Atomic Number Paired Nucleotide Representation

Nucleotide pairs G-A and C-T are allocated atomic numbers of 62 and 42, respectively [15]. Each nucleotide assumes a distinct atomic number in its representation.

Complex Numerical Encoding Scheme

This framework [5,9] supersedes Voss's binary sequences [16] with a singular complex number representation, reducing computational efforts by 75%. Each nucleotide is given a distinct complex value. These nucleotide encoding paradigms are instrumental in crafting a time-domain signal from a DNA sequence, facilitating subsequent analysis via digital signal processing. The incorporation of apt window functions, as evidenced in literature, augments noise suppression, fortifying the demarcation between coding and non-coding regions, and enhances protein coding region identification [17].

Window Functions in Signal Processing

The effective demarcation of protein-coding regions is intrinsically linked to the judicious use of specific window functions. These functions amplify the clarity of identification while concurrently minimizing signal interference. A survey of the literature underscores the

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

following window functions [4,5] as particularly salient in protein coding region\

identification using digital signal signal processing methodologies:

Table 1 : *Window Function*

Window Function	Mathematical Expression	Description
Rectangular	$\begin{cases} w_r(n) = 1, -\frac{N-1}{2} \le n \le \frac{N-1}{2} \\ 0, \text{otherwise} \end{cases}$ w_r is a Rect. Window function with a range of 0 to N samples.	Used to convolute specific DNA segments, ensuring a consistent window size.
Bartlett	$\begin{cases} w_b = \frac{2n}{N-1}, 0 \le n \le \frac{N-1}{2} \\ 2 - \frac{2n}{N-1}, \frac{N-1}{2} \le n \le N-1 \\ 0, \text{otherwise} \end{cases}$	Triangular windows have a softer shift this transition neatly splits samples in half.
Hamming	$\begin{cases} w_h(n) = 0.54 - 0.46\cos(2\pi nN), 0 \\ 0, \text{otherwise} \end{cases}$	Smoothly tapers within a designated sample range, effectively reaching zero outside that range.
Kaiser	$w_k = \frac{I_0\left(\beta\sqrt{1-\left(\frac{n-N/2}{N/2}\right)^2}\right)}{I_0(\beta)}, 0 \le n \le N$	It is affected by factors α and β. The value of β (often 0.5) controls leakage and side lobe attenuation.
Adaptive Kaiser	Varies based on algorithm	A modified Kaiser window, its adaptability lies in parameter usage: window length (L) and β.
Gaussian	$G^w = e^{-\frac{1}{2}\left(\frac{n-(N-1)/2}{\sigma(N-1)/2}\right)^2}$	Combined with S-Transform, it detects coding sequences in both time and frequency domains.
Hanning	$w_{hn} = 0.5\left(1 - \cos\frac{2\Pi n}{N-1}\right)$	Works similarly within designated samples, gracefully declining to zero beyond that range.

Digital Signal Processing Techniques:

Within the domain of Digital Signal Processing (DSP), an array of algorithms is harnessed for the purpose of identifying coding sequences within DNA sequences [18] . Diverse effects in the localization of coding sequences are observed across algorithms, despite identical mapping and windowing methodologies. provides a comprehensive overview of the principal DSP tools commonly applied in this domain.

• **Filtering:** One important part of reducing noise in DNA sequences is digital filtering. It's worth mentioning that Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) filters are often used to separate coding patterns. There are differences between these filters in terms of their security, bandwidth, and filter order [19]. When it comes to DSP methods designed to look closely at genetic sequence data, the IIR filter and its extended versions play a big part.

• **Approaches Based on Transforms:**

i. Fourier Transform: The Fourier transform emerges as a prevalent analytical methodology for the identification of protein-coding regions [15]. This involves the transformation of a DNA sequence, sampled at N points, encapsulated in equation (1), where x(n) represents the sequence, and X(k) signifies the transformation coefficients.

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk}, 0 \le k \le N-1 \quad \textbf{(1)}$$

ii. Short-Term Fourier Transform: This approach engages in the windowing of the Discrete Fourier Transform (DFT) of

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

a signal [9] . It is imperative to note that the Fourier Transform is subject to spectral leakage, which results in the dispersion of energy of main lobe to the secondary lobes.

iii. Transform of Wavelets: The Wavelet Transform stands distinguished for its efficacy in noise mitigation and the facilitation of optimal scaled window function selection to enhance signal dissection [20]. The discrete transformation is governed by equation (2), wherein α denotes the noise index, and p encapsulates signal attenuation considerations.

$$p_k = \frac{1-\alpha}{1+\alpha^2-2\alpha\cos(2\Pi k/N)} \quad \textbf{(2)}$$

iv. S-Transform (ST): The S-Transform amalgamates the virtues of Short-Term Fourier Transform (STFT) and Wavelet Transform (WT) analyses [21]. This amalgamation engenders optimal performance, attributable to its adaptability in specifying window dimensions contingent upon the sequence characteristics. Equation (3) defines the ST, where τ signifies time, frequency f represents, and g embodies the adjustable Gaussian window.

$$s(t,f) = \int_{-\infty}^{\infty} u(\tau) g(t-\tau,f) e^{-j2\Pi f\tau} d\tau \quad \textbf{(3)}$$

• **Spectrum Estimation:** Spectrum estimation involves the computation of Discrete Fourier Transform (DFT) component magnitudes and the estimation of signal component power [22]. The resultant Power Spectral Density (PSD) profile, depicted via frequency and amplitude attributes, serves to discern coding regions from their non-coding counterparts. This classification is substantiated by two cardinal approaches to spectral estimation.

i. Parametric Method: Optimized for scenarios with short signal lengths, parametric techniques channel a linear system model to analyze genetic data [16]. Prominent examples include the Auto-Regressive model with a pole and methods like the covariance method, the Yule-Walker method, and the Burg method.

ii. Non-Parametric Method: Encompassing methodologies rooted in Power Spectral Density (PSD) estimation, non-parametric techniques provide a means for spectral analysis [23]. The formulation of signal power spectral density follows equation (4).

$$S[k] = \sum |x_\alpha(k)|^2$$
$$S = |A|^2 + |C|^2 + |G|^2 + |T|^2 \quad \textbf{(4)}$$

$$where$$
$$A = \frac{1}{N} x_A(k), C = \frac{1}{N} x_C(k), G = \frac{1}{N} x_G(k), T = \frac{1}{N} x_T(k)$$
$$k \in \left\{0,1,\dots,\frac{N}{2}\right\}$$

A more sophisticated approach is presented by Welch's method [18], in which the temporal-domain signal is first subjected to windowing before Fast Fourier Transform (FFT) calculations are carried out. This deliberate inclusion contributes to the phenomena of spectral leaking [13,16]. The gene's application of signal processing techniques C. Elegans is a prime example of where the coding sequence locations are located. These locations are shown as large peaks in the Power Spectral Density (PSD) plot, as Figure 3 previously demonstrated.

DNA Signal Challenges

Although DNA signals have unique nucleotide arrangements, they have an inherent 3-base periodicity and can be challenging to decipher due to noise [24]. The vital task of distinguishing coding from non-coding regions is further complicated by a 1/f background noise. Only about 5% of DNA sequences, or the coding regions, are translatable to protein sequences [25]. Yet, identifying these regions is crucial because unidentified protein structures can impede medical progress. Despite the vast amount of genomic data available, discerning the noise and extracting meaningful information remains challenging.

2. DNA to Protein: An Overview of Methodologies

2.1. Coding Measure Schemes

Recent research has been actively exploring the identification of varying types of periodicities in DNA sequences. These periodicities are detectable through digital signal processing (DSP) techniques. The use of DSP for deriving information from DNA sequences, such as DNA analysis and identifying protein-coding regions, largely hinges upon translating the DNA segment into its digital signal equivalent [3, 7]. Extensive studies [26] highlight the crucial role of coding measure schemes (also known as DNA coding measure schemes or DNA indicator sequences) in differentiating protein-coding from non-coding regions, especially by minimizing 1/f noise.

Such numerical portrayals capture the inherent biological attributes of DNA in a numerical framework.

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

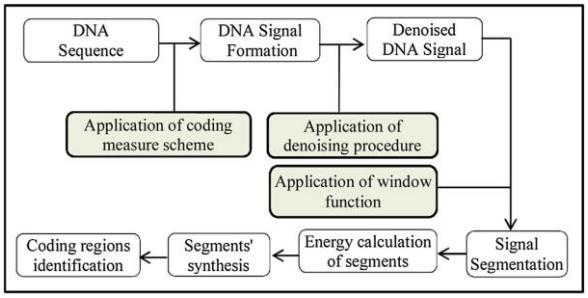

Fig 2 : *General frame work*

mapping scheme, wherein the transformation remained invariant for the representation and labeling of nucleotide bases. This scheme served to discern periodicity in DNA segments by placing the four nucleotides at the vertices of a regular tetrahedron, with vectors originating from the center representing each nucleotide. Subsequent developments by Demeler and Zhou [28] introduced a neural network-based predictive model using a 4-bit binary encoding for nucleotides A, C, G, and T, mapped as 1000, 0010, 0001, and 0100, respectively. Voss [20] further proposed the 'Binary Indicator Sequence', which features four sequences, each representing a nucleotide. This binary sequence assigns values of 1 or 0 based on the presence or absence of a specific nucleotide, leading to sequences like x[n] = [T T A G G T C C T] translating to [001000000] for Adenine. Utilizing this approach, the summation of all binary sequences always equals one, ensuring a comprehensive mapping of the DNA sequence into these indicator sequences. This, in turn, facilitates the translation of the DNA sequence into a digital signal, ripe for frequency-based analyses. Further innovations include the molecular mass-based representation by Stanley et al. [29], which correlates nucleotides with their molecular masses, and Yan et al.'s [14] introduction of a Fourier transform methodology hinged on the Z-curve format. Lio and Vannucci [30], on the other hand, documented pathogenicity islands and gene transfer events, presenting unique numerical conventions to represent DNA bases. A notable advancement by Bernaola-Galván et al. [11] targeted the delineation between coding and non-coding DNA regions through entropic segmentation, employing a 12-letter nucleotide formation. Zhang and Wang [31] subsequently built upon the Z-curve methodology, and Dodin et al. [12] leveraged both Fourier and wavelet transforms for pattern visualization in DNA sequences. Anastassiou [22] offered a frequency-domain interpretation of DNA sequences via nucleotide-based numerical assignments, which set the stage for sophisticated DSP analyses. Bernaola-Galván et al. [32] presented another approach by proposing paired nucleotide representations. In contrast, Cristea [4] introduced a coding scheme derived from Integer Number representations, and Berger et al. [7] undertook power spectrum analyses. Innovative approaches such as the autoregressive modeling by Chakravarthy et al. [33], the neural network-based multi-classifier system proposed by Ranawana and Palade [9], and Nair and Mahalakshmi's [15] exploration of inter-nucleotide distance signals have also emerged. Furthermore, Rosen [34] introduced a single indicator sequence coding scheme, and Nair and Sreenadhan [35] identified potential shortcomings in Voss' [6] coding scheme, proposing an alternative Electron-ion Interaction Pseudo Potential (EIIP) measure. This EIIP measure demonstrated superior differentiation capabilities between coding and non-coding regions, showcasing its potential in genomic signal processing and gene discovery. Subsequently, Holden et al. [36] delineated the fluctuation of the ATCG nucleotide by employing the representation of Paired Nucleotide Atomic Numbers. Atomic numbers were assigned as G, A = 62 and C, T = 42. The coding measure was formulated with the atomic numbering for nucleotides: C = 58, G = 78, A = 70, and T = 66. Concurrently, Grandhi and Kumar [17] proposed a 2-simplex mapping for exon identification in DNA sequences. This triadic mapping symbolized the four nucleotide bases along triangle vertices and its center. The bases underwent substitution with pertinent vector values and underwent IIR filtering, ensuring retention of the 3-base property, thus halving the complexity compared to Voss' method [37]. In a related study, Yin and Yau [20] conceptualized a numerical representation for DNA sequence segments via complex numbers. Considering the 64 genetic codons representing approximately 20 amino acids, each amino acid was aligned with a distinct complex number. Notably, the real and imaginary components of these complex numbers reflected the characteristics of amino acids. In continuation, Akhtar et al. [38] harnessed a paramount statistical property related to the nucleotide content spectrum (C < G and A < T) and mapped DNA with real numbers. Further refinement was introduced by Hota and Srivastava [17] with the introduction of the "Complex scheme," which allocated complex numbers to DNA bases, resulting in a significant reduction in computational overhead compared to Voss' method [16]. Kwan et al. [1] introduced a novel coding measure based on the Fourier transform of the DNA digital signal, dubbed the Complex Twin-Pair. In summation, the crux of DNA encoding lies in the contextual genetic code of its nucleotides, which is pivotal for DNA sequence analysis, particularly in coding region identification [39]. The inherent codon structure of nucleotides, their density distribution, and relation with amino acids are quintessential elements yet to be holistically addressed in literature. Given the intrinsic codon format of nucleotides, the position and frequency of each nucleotide within a codon has profound implications for protein synthesis. For instance, the codon "ATG" synthesizes the amino acid "Methionine", while "TAG" does not correspond to any amino acid. Variations in nucleotide placement within a codon can alter the resultant amino acid. Additionally, the degenerate nature of the DNA code implies

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

that multiple codons can encode the same amino acid. Exploiting these intrinsic nucleotide characteristics could lead to a more comprehensive coding measure scheme, encapsulating the nucleotides' occurrence, density distribution, and position within codons [25].

2.2. Window Filters in DNA Signal Processing

This study undertook a comprehensive review of literature in search of prevalent Window filters utilized for DNA signal processing, specifically targeting the identification of protein coding regions. A recurrent theme in the literature revealed the use of standard Window filters of set lengths. Surprisingly, a noticeable gap was identified; the literature appeared insufficient concerning Window filters rooted in the genetic context of code [40], and those that account for the skewed nucleotide distribution, which leads to pronounced bias in nucleotide usage in coding regions [10,11]. Table 1 catalogues an array of Window filters, illustrating their designated window lengths designed for codon tracking. It was noted, for instance, that a conventional Rectangular Window filter with a filter length of 351 base pairs (bp) was adopted by numerous researchers including Sahu and Panda [8], Yin and Yau [19], and Akhtar et al., [5] among others. In a similar vein, other scholars such as Chavan et al., [8] and Nair and Sreenadhan [26] utilized the Kaiser Window filter with the same length of 351 bp. Other notable filters and their proponents include the Bartlett Window filter and the Gaussian Window filter, both also of 351 bp length.

2.2.1. Considerations in Window Filter Adoption for Identifying Coding Regions

An evident trend, observable across Rectangular, Kaiser, and Bartlett Windows, is the preference for a fixed length of 351 base pairs. The choice of Window filters and their corresponding lengths are instrumental in digital processing methodologies for discerning coding regions. The power spectral analysis of the DNA signal, especially at period (N/3), distinctly marks the coding regions [27]. For illustration, Table 3 showcases crucial elements like filter length, leakage factors, and sidelobe attenuation essential for the selection of an optimal Window filter for coding region identification. An insightful observation is the pivotal role played by the appropriate window size in mitigating signal noise, thereby enhancing the discernment of coding regions [28]. Among the filters analyzed, the Kaiser Window emerged as the most effective for this purpose.

2.2.2. The Significance of Proper Window Filter Parameter Selection

The study found that distinct Window functions, characterized by varying parameters, resulted in diverse spectra. While certain Window functions excelled by diminishing signal noise, others lacked discernible impact. Given that the successful identification of coding regions is contingent on distinguishing them from their non-coding counterparts in a noise-ridden DNA signal, the choice of the right filter is paramount. Empirical results endorsed the Kaiser Window with parameters beta = 3.5 and window length of 351 as the optimal combination for coding region detection [41]. Further elaboration on the Kaiser Window of 351 bp size suggests caution: larger window sizes risk overlooking smaller coding and non-coding regions. Although traditional Window filters are routinely applied across a range of digital signals [30], DNA signals are distinct, encapsulating biologically significant nucleotide information. Given the inherent genetic coding context of each nucleotide and its disproportionate distribution in coding regions, it becomes evident that typical Window filters, especially of fixed sizes, may not sufficiently mitigate the 1/f background noise [15]. This inadequacy could lead to a blurred distinction between coding and non-coding regions.

2.3 Noise Suppression Techniques in DNA Signals

DNA signals are significantly tainted by 1/f background noise, which substantially affects the precision of predicting coding regions. This is due to the blurring of non-coding and coding regions within a DNA sequence, making their differentiation in noisy conditions arduous. Several techniques have been suggested to suppress this noise, including digital filters, frequency transformations such as Fourier and short time Fourier analyses, and Wavelet transformations [42]. A synopsis of these existing suppression methods follows: Kakumani et al. [31] utilized a digital signal processing technique to enhance the signal-to-noise ratio (SNR) of DNA signals. They identified the exonic regions in DNA strands using least square optimization. Their methodology entailed transforming the DNA sequence into a digital signal via four binary sequences within a 351 bp Rectangular Window. The SNR gains of each sub-sequence, processed through their proposed system, were calculated. Their method was then tested on genes in chromosome III of C-elegans to pinpoint the exonic regions. In a related vein, Akhtar et al. [43] introduced a digital signal processing strategy to denoise DNA signals. Their goal was to accurately predict exonic and intronic regions and compare their approach to existing methods. They incorporated a statistical analysis, highlighting nucleotides A and T's richness in introns, whereas exons were abundant with C and G. Their results were benchmarked against datasets like Burset/Guigo1996, HMR195, and GENSCAN. Similarly, Tuqan and Rushdi [44] offered a method for DNA signal denoising to discover periodicity in DNA sequences. Their approach involved examining the mechanism behind period-3 components and subsequently establishing a relationship between these components and nucleotide bias in a DNA spectrum. Their technique integrated a multirate DSP model.

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

Gupta et al. [41] outlined a time series-based digital signal processing approach for noise suppression. Their method hinged on extracting specific features such as nucleotide distribution and hydrogen bonding. Using pattern recognition, they identified exon and intron boundaries. Their process referenced Z-curve components and treated each component of their model as discrete time series data. On a related note, Hamdani and Shukri [45] formulated a DSP system designed for protein region identification. Their approach encompassed transcription, splicing, and translation phases, converting DNA to RNA and subsequently to protein. Sahu and Panda [46] devised a signal processing method to pinpoint protein-coding regions. They observed limitations with the existing AR method and introduced an adaptive AR model for enhanced predictions. Shakya et al. [44] targeted enhancing coding region identification by denoising the DNA sequence through digital signal processing. Their approach used a Bartlett Window of 351 bp length, as it had been found effective in prior research.

Datta and Asif [47] introduced a Discrete Fourier Transform (DFT) based algorithm for DNA sequence denoising. Their tests on chromosome III of C. Elegans revealed that their method's efficacy varied with coding region size. Akhtar et al. [34] took a frequency-based approach to denoise DNA signals, optimizing a DFT-based method for exonic identification. Continuing this trend, Yin and Yau [35] leveraged DFT to predict exonic regions, extracting Fourier coefficients from DNA sequences. Roy et al. [10] also utilized a frequency-based analysis for signal denoising, with their algorithm considering nucleotide frequency distribution in DNA. Lastly, Shuo and Yi-sheng [11] combined denoising with coding identification through a Support Vector Machine (SVM) methodology. They used a short-time Fourier transformation to observe the time-frequency characteristics of the output. Their approach was capable of classifying coding and non-coding regions, providing a novel mechanism for splice site recognition.

3. ANALYSIS AND DISCUSSION

We assessed the performance of various existing methods for identifying coding regions at the basic nucleotide level. Key evaluation metrics utilized for this purpose are detailed below:

Discrimination Measure (DM): The Discrimination Measure, as defined by:

$$\mathrm{DM} = \frac{\text{Lowestpeak amplitude in coding regions}}{\text{Highest peak amplitude in noncoding regions}} \qquad \textbf{(5)}$$

A DM value exceeding 1 signifies unambiguous recognition of coding sequences, negating the likelihood of erroneous predictions.

Sensitivity (Sn): Sensitivity, denoted as Sn, quantifies the proportion of accurately identified coding nucleotides:

$$Sn = \frac{TP}{TP+FN} \qquad \textbf{(6)}$$

True Positives (TP)

True Positives (TP) + False Negatives (FN)

Specificity (Sp): Specificity, represented as Sp, assesses the fraction of correctly identified noncoding nucleotides pertaining to non-coding regions:

$$Sp = \frac{TP}{TP+fp} \qquad \textbf{(7)}$$

True Positives (TP)

True Positives (TP) + False Positives (FP)

The correlation coefficient (CC) for Matthew: The Matthew Correlation Coefficient, designated as CC, serves as an epitome of optimal performance, formulated as:

$$\mathrm{CC} = \frac{\mathrm{TP}*\mathrm{TN}-\mathrm{FN}*\mathrm{FP}}{\sqrt{\mathrm{TP}+\mathrm{FN}*\mathrm{TN}+\mathrm{FP}*\mathrm{TP}+\mathrm{FP}*\mathrm{TN}+\mathrm{FN}}} \qquad \textbf{(8)}$$

• **Applications:** The versatile use of gene prediction includes a wide range of applications, such as the prediction of diseases, classification of diseases, determination of disease stages (e.g. arthritis, diabetes, Parkinson's, cancer), identification of genetic disorders, and localization of hotspots within protein sequences [48].

4. ADVANTAGES OF USING MULTI-RESOLUTION TRANSFORMATION IN THE ANALYSIS OF DNA SEQUENCES:

Captures patterns at multiple scales: MRT decomposes DNA sequences into different frequency bands or resolution levels, enabling the identification of patterns and features that manifest at varying scales (e.g., short motifs, long repeats, structural variations).

Uncovers hidden relationships: Reveals relationships and correlations between different regions of a DNA sequence that might be obscured in the raw data, aiding in gene prediction, regulatory element identification, and comparative genomics. Separates noise from information: MRT isolates noise and redundant information, preserving essential features while reducing data size. This improves signal-to-noise ratio, facilitating more accurate analysis. Reduces storage and computational costs: Compressed representations simplify storage and analysis, especially for massive datasets like whole-genome sequences. Faster comparison of sequences: MRT coefficients provide compact and informative representations, enabling faster similarity searches and

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

alignments, crucial for tasks like sequence assembly, homology detection, and phylogenetic inference.
Handles large-scale comparisons: Facilitates efficient comparisons of massive DNA sequences, a challenge for traditional alignment methods. Identifies features at different scales: MRT pinpoints the exact locations of features in the original sequence, aiding in the interpretation of biological function and regulatory mechanisms.
Enables hierarchical analysis: Facilitates multiscale analysis of DNA sequences, essential for understanding complex genomic structures and processes, such as gene regulation and chromatin organization.
Tolerates noise and errors: MRT is relatively robust to variations and noise in DNA sequences, making it suitable for handling experimental errors, sequencing artifacts, and natural genetic polymorphisms. Captures invariant features: Extracts features that are resilient to minor distortions, improving the reliability of downstream analyses. Complementary to other methods: MRT often complements other DNA sequence analysis techniques, such as Hidden Markov Models, neural networks, and machine learning algorithms. Enhances performance: Can improve the accuracy and efficiency of various bioinformatics tasks, including gene prediction, motif discovery, and sequence classification.

5. CRITICAL DISCUSSION ON ANALYSIS OUTCOMES

Typically, in many applications, a digital signal is processed based on its digital content [49]. As reflected in the literature previously reviewed, biological signals, like those derived from DNA/RNA sequences, have been treated analogously to conventional digital signals. These particular signals, however, are rich with essential biological information linked to specific gene or nucleotide sequences. Merely categorizing these as standard digital signals may not yield insightful outcomes [50]. Recognizing and processing the nuanced biological content in these signals can enhance the precision in identifying protein-coding regions.

Consider the challenge of distinguishing protein-coding regions, which are often intermingled with non-coding regions due to the 1/f DNA signal noise. This overlap makes differentiation a complex task. The step of converting a DNA sequence into a digital signal is vital for effectively distinguishing coding from non-coding regions [10]. Results suggest that traditional encoding methods and convolution factors (or Window filters) grounded solely in Digital Signal Processing (DSP) concepts often yield similar results. This uniformity likely emerges because these methodologies disregard the intricate biological nuances of nucleotides when crafting encoding sequences and Window filters.

Examining the DNA sequences and their conversion to protein sequences reveals a universe of 64 potential codons responsible for translating DNA into proteins at specific regions termed exons [50]. Table 5 outlines the amino acids, represented by their abbreviated codes, linked with relevant codons [51]. Intriguingly, certain amino acids can result from multiple codons. For example, both "TAT" and "TAC" codons can produce Tyrosine. This redundancy in the DNA code, where multiple codons can code for a single amino acid, is termed degeneracy [51]. Fig. 8 showcases the varied space occupied by different codons. Observations suggest that while some codons remain entirely distinct, others exhibit overlapping density distributions [52]. For instance, Fig. 9 presents four distinct codons devoid of shared characteristics, such as density distribution. Conversely, other codons have overlapping density distributions. An example I n Fig. 10 reveals codons containing the nucleotides "A, T, and G", each contributing roughly one-third to their combined density, reflecting their uniform mass distribution. Similarly, codons in Fig. 11, which include nucleotides "A, T, C, and G", show a two-thirds contribution in their nucleotides' density distribution. This analysis underscores the uniqueness and significance of nucleotide arrangements in codons. Each nucleotide occupies a pivotal position within its codon, and its placement is decisive for the type of protein it generates. Treating DNA signals derived from these nucleotides as typical digital signals might fall short of offering precise identification of coding regions. While standard DSP techniques are apt for processing typical digital signals, they might not be as effective for processing DNA signals imbued with genetically relevant content [53]. Hence, there's a pressing need to innovate new DSP strategies that can holistically process DNA signals, emphasizing the genetic significance of the contained nucleotides.

Here are the key advantages of enhanced identification of coding regions in DNA sequences:

a- Gene Discovery and Annotation [54]: Unlocks the genome: Accurate identification of coding regions is fundamental to gene discovery, which involves locating genes within the vast expanse of DNA sequences [55]. Functional insights: Precise annotation of gene boundaries and structures provides essential information for understanding gene function, regulation, and potential roles in health and disease.

b- Drug Target Identification: Therapeutic targets: Coding regions often contain blueprints for proteins that are involved in various biological processes and pathways. Identifying these regions helps researchers pinpoint potential drug targets for developing new therapies. Personalized medicine: Accurate identification of coding

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

regions can inform personalized medicine approaches, tailoring treatments to individual genetic variations.

c- Disease Gene Mapping: Understanding disease mechanisms: Identifying genes associated with specific diseases is crucial for understanding disease mechanisms and developing diagnostic tests and targeted therapies. Genetic counseling and risk assessment: Precise mapping of disease genes aids in genetic counseling, risk assessment, and potential prevention strategies.

d- Evolutionary Studies [56]: Comparative genomics: Comparing coding regions across different species reveals evolutionary relationships, adaptation patterns, and the genetic basis of phenotypic diversity [57]. Understanding molecular evolution [58]: Tracing changes in coding regions over time sheds light on the mechanisms of molecular evolution and the emergence of new traits and functions [59].

e- Genetic Engineering [60]: Manipulating genes: Precise identification of coding regions enables targeted genetic engineering techniques for modifying gene expression, introducing new traits, or correcting genetic defects.Biotechnological applications [61]: This has applications in agriculture, medicine, and industrial biotechnology, leading to developments like improved crops, gene therapies, and bio-based products [62].

f- Personalized Medicine: Tailored treatments: Accurate identification of coding regions can inform personalized medicine approaches, tailoring treatments to individual genetic variations and disease susceptibility. Precision medicine: This allows for more precise targeting of therapies and improved outcomes for patients with diverse genetic backgrounds.

g- Forensics and Biometrics [63]: Identifying individuals: Coding regions can be used for DNA profiling in forensics and biometric identification, aiding in criminal investigations, paternity testing, and tracking ancestry [64]. Personalized security [65]: This has implications for personalized security systems and biometric authentication [66].

h- Artificial Intelligence and Machine Learning [67]: Training and prediction: Accurate identification of coding regions provides valuable data for training and improving machine learning algorithms in genomics, bioinformatics, and drug discovery [68]. Accelerated research [69]: This can accelerate research, lead to novel insights, and potentially revolutionize various fields of biology and medicine.

6. CONCLUSION

This study delves into the complexities and challenges surrounding the precise identification of protein-coding regions as distinct from non-coding regions in Eukaryotic DNA sequences. Our in-depth analysis reveals that protein-coding regions often intermingle with non-coding regions, primarily due to the presence of the 1/f background noise in the DNA sequences. This mixing complicates the task of distinctly identifying the two regions. A notable correlation exists between the increased accuracy in identifying coding regions and the pronounced 1/f background noise. Furthermore, protein-coding regions exhibit a characteristic 3-base periodicity, a feature absent in non-coding regions. This distinct property has been the focus of numerous studies, leading researchers to propose a myriad of methods and tools for coding region identification. Yet, there remains an unmet need for a more effective solution. Implementing computational strategies that take into account the genetic code context can greatly aid in: Recognizing full gene patterns across extended noisy DNA sequences. Translating specific DNA segments into complete protein sequences. Identifying motifs to uncover patterns in nucleotides or protein sequences, enhancing our comprehension of molecular structure and function. Designing advanced multilevel digital filters to mitigate the 1/f noise, enabling clearer distinction between coding and non-coding areas. Pinpointing protein folding patterns based on their structural attributes. Table-1 gives comparison between the performance of different transformation techniques in DNA sequence analysis [70]. Most of resent studies consider for DNA analysis the multiresolution transformation [71] or hybrid transform [72] or mixed transform [73], Wavelet transform [74] or multi-level transform [75] or wavelet packet [76] and Multiwavelet transform [77]. It is clear that the spectral [78] or frequency domain [79] performance outperform that of time domain [80] in DNA sequence analysis

Table-1 *Comparison between sequence analysis performance of different transformation techniques*

multi-resolution transformation	Wavelet transform	multiwavelet transform
1. Enhanced Feature Extraction and Pattern Recognition:	1. Localization in Time and Frequency:	1. Enhanced Directional Sensitivity:
2. Noise Reduction and Data Compression:	2. Flexibility in Basis Function Selection:	2. Improved Redundancy Reduction:
3. Efficient Similarity Search and Alignment:	3. Multiscale Analysis and Feature Extraction:	3. Enhanced Feature Discrimination:
4. Multiscale Analysis and Feature Localization:	4. Noise Reduction and Compression:	4. Robustness to Noise and Distortions:

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

5. Robustness to Distortions and Variations:	5. Efficient Similarity Search and Classification:	5. Enhanced Multiscale Analysis:
6. Integration with Other Techniques:	6. Handling Non-Stationary Signals:	6. Integration with Machine Learning:

REFERENCES

[1] L. Galleani and R. Garello, “The minimum entropy mapping spectrum of a DNA sequence,” IEEE Trans. Inf. Theory, vol. 56, no. 2, pp. 771–783, 2010.

[2] J. Mena-Chalco, H. Carrer, Y. Zana, and R. M. Cesar Jr, “Identification of protein-coding regions using the modified Gabor-wavelet transform,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 5, no. 2, pp. 198–207, 2008.

[3] M. Akhtar, J. Epps, and E. Ambikairajah, “Signal processing in sequence analysis: advances in eukaryotic gene prediction,” IEEE J. Sel. Top. Signal Process., vol. 2, no. 3, pp. 310–321, 2008.

[4] S. S. Sahu and G. Panda, “Identification of protein-coding regions in DNA sequences using a time-frequency filtering approach,” Genomics. Proteomics Bioinformatics, vol. 9, no. 1, pp. 45–55, 2011.

[5] S. Barman, M. Roy, S. Biswas, and S. Saha, “Prediction of cancer cell using digital signal processing,” Ann. Fac. Eng. Hunedoara, vol. 9, no. 3, p. 91, 2011.

[6] W.-F. Zhang and H. Yan, “Coding sequence prediction using empirical mode decomposition and Fourier transform of structural profiles of DNA sequences,” Pattern Recognit., vol. 45, no. 3, pp. 947–955, 2012.

[7] M. Ahmad, L. T. Jung, and A.-A. Bhuiyan, “A biological inspired fuzzy adaptive window median filter (FAWMF) for enhancing DNA signal processing,” Comput. Methods Programs Biomed., vol. 149, pp. 11–17, 2017.

[8] P. Ramachandran, W.-S. Lu, and A. Antoniou, “Filter-based methodology for the location of hot spots in proteins and coding sequences in DNA,” IEEE Trans. Biomed. Eng., vol. 59, no. 6, pp. 1598–1609, 2012.

[9] T. Meng et al., “Wavelet analysis in current cancer genome research: a survey,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 10, no. 6, pp. 1442–14359, 2013.

[10] T. M. Inbamalar and R. Sivakumar, “Study of DNA sequence analysis using DSP techniques,” J. Autom. Control Eng. Vol, vol. 1, no. 4, 2013.

[11] D. K. Shakya, R. Saxena, and S. N. Sharma, “An adaptive window length strategy for eukaryotic CDS prediction,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 10, no. 5, pp. 1241–1252, 2013.

[12] L. Das, J. K. Das, and S. Nanda, “Identification of coding sequence location applying Kaiser window and DFT techniques,” in Convergence in Technology (I2CT), 2017 2nd International Conference for, 2017, pp. 211–216.

[13] S. A. Marhon and S. C. Kremer, “Prediction of protein coding regions using a wide-range wavelet window method,” IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 13, no. 4, pp. 742–753, 2016.

[14] M. Roy and S. Barman, “Improved gene prediction by principal component analysis based autoregressive Yule- Walker method,” Gene, vol. 575, no. 2, pp. 488–497, 2016.

[15] M. Ahmad, L. T. Jung, and A.-A. Bhuiyan, “From DNA to protein: Why genetic code context of nucleotides for DNA signal processing? A review,” Biomed. Signal Process. Control, vol. 34, pp. 44–63, 2017.

[16] B. Das and I. Turkoglu, “A novel numerical mapping method based on entropy for digitizing DNA sequences,” Neural Comput. Appl., pp. 1–9, 2017.

[17] M. S. Mabrouk, S. M. Naeem, and M. A. Eldosoky, “DIFFERENT GENOMIC SIGNAL PROCESSING METHODS FOR EUKARYOTIC GENE PREDICTION: A SYSTEMATIC REVIEW,” Biomed. Eng. Appl. Basis Commun., vol. 29, no. 01, p. 1730001, 2017.

[18] L. Das, J. K. Das, and S. Nanda, “Advanced protein coding region prediction applying robust SVD algorithm,” in Man and Machine Interfacing (MAMI), 2017 2nd International Conference on,

[19] D. Anastassiou, “Genomic signal processing,” IEEE Signal Process. Mag., vol. 18, no. 4, pp. 8–20, 2001.

[20] M. Akhtar, E. Ambikairajah, and J. Epps, “Detection of period-3 behavior in genomic sequences using singular value decomposition,” in Emerging Technologies, 2005. Proceedings of the IEEE Symposium on, 2005, pp. 13–17.

Ammar A. Sakran, Suha. M. Hadi, Waleed A. Mahmoud Al-Jawher., 2023. Advancing DNA Signal Processing: Integrating Digital and Biological Nuances for Enhanced Identification of Coding Regions. *Journal port Science Research*, 6(4), pp.374-387. https://doi.org/10.36371/port.2023.4.8

[21] A.K. Singh, V.K. Srivastava, Improved filtering approach for identification of protein-coding regions in eukaryotes by background noise reduction using S- G filter, Netw. Model. Anal. Health Inform. Bioinform. 10 (19) (2021) 1–16

[22] L. Das, S. Nanda, and J. K. Das, "A novel DNA mapping scheme for improved coding sequence prediction using digital filters," in Man and 2017Machine Interfacing (MAMI), 2nd International Conference on, 2017, pp. 1–6.

[23] P.K. Kumari, J.B. Seventline, A Novel Approach For Identification Of Exon Locations In DNA Sequences Using GLC Window, Int. J. Biol. Biomed. Eng. 15 (2021) 47–60.

[24] A. Piovesan, F. Antonaros, L. Vitale, P. Strippoli, M.C. Pelleri, M. Caracausi, Human protein-coding genes and gene feature statistics, BMC Res. Notes 12 (2019).

[25] S.A. Marhon, S.C. Kremer, Prediction of protein-coding regions using a wide-range wavelet window method, IEEE/ACM Trans. Comput. Biol. Bioinform. 13 (4) (2016) 742–753.

[26] S.D. Sharma, S.N. Sharma, R. Saxena, Identification of short exons disunited by a short intron in eukaryotic DNA regions, IEEE/ACM Trans. Comput. Biol. Bio in-form. 17 (5) (2020) 1660–1670.

[27] Q. Zheng, T. Chen, W. Zhou, L. Xie, H. Su, Gene prediction by the noise-assisted MEMD and wavelet transform for identifying the protein-coding regions, Bio- cybern. Biomed. Eng. 41 (1) (2021) 196–210.

[28] M. Raman Kumar, N.K. Vaegae, Walsh code based numerical mapping method for the identification of protein-coding regions in eukaryotes, Biomed. Signal Process. Control 58 (2020) 1–11.

[29] M. Raman Kumar, N.K. Vaegae, A new numerical approach for DNA representation using modified Gabor wavelet transform for the identification of protein-coding regions, Biocybern. Biomed. Eng. 40 (2) (2020) 836–848.

[30] M.K. Hota, Empirical mode decomposition based adaptive noise canceller for improved identification of exons in eukaryotes, Netw. Model. Anal. Health Inform. Bioinform. 10 (2021) 60.

[31] P. Garg, S.D. Sharma, S.N. Sharma, Tandem repeats detection in DNA sequences using p-spectrum based algorithm, in The Proceedings of Conference on Information and Communication Technology (CICT 2017), 2017, pp. 1–5.

[32] N. Yu, Z. Li, Z. Yu, Survey on encoding schemes for genomic data representation and feature learning-from signal processing to machine learning, Big Data Min.Anal. 1(3)(2018) 191–210.

[33] P. Garg, S.D. Sharma, Identification of CpG Islands in DNA sequences using short-time Fourier transform, Interdiscip. Sci. Comput. Life Sci. 12 (3) (2020) 355–367.

[34] N. Yu, Z. Li, Z. Yu, Survey on encoding schemes for genomic data representation and feature learning—from signal processing to machine learning, Big Data Min. Anal. 1 (2018) 191–210.

[35] H. Wassfy, M. Elnaby, M. Salem, M. Mabrouk, A.-A. Zidan, Advanced DNA Mapping Schemes for Exon Prediction Using Digital Filters, Am. J. Biomed. Eng. (6) 1 (2016) 25–31

[36] A.K. Singh, V.K. Srivastava, Improved filtering approach for identification of protein-coding regions in eukaryotes by background noise reduction using S-G filter, Netw. Model. Anal. Health Inform. Bioinforma. (10) 1 (2021)

[37] J. Baranowski, P. Piątek, Fractional Band-Pass Filters: Design, Implementation, and Application to EEG Signal Processing, J. Circuits Syst. Comput. 26 (11) (2017) 1750170

[38] Y. Ferdi, A Fractional Digital High-Pass Notch Filter, in 2020 10th International Symposium on Signal, Image, Video and Communications (ISIVC), 2021, pp. 1–4.

[39] A. Gupta, S. Kumar, Design of Atangana–Baleanu–Caputo fractional-order digital filter, ISA Trans. 112 (2021) 74–88

[40] A.K. Shukla, R.K. Pandey, R.B. Pachori, A fractional filter based efficient algorithm for retinal blood vessel segmentation, Biomed. Signal Process. Control 59 (2020)

[41] Ammar AbdRaba Sakran, Waleed A. Mahmoud Al-Jawher and Suha M Hadi " A new approach for DNA sequence analysis using multiwavelet transform (MWT)" , 2023 J. Phys.: Conf. Ser. 2432 012022.

[42] Q. Zheng, T. Chen, W. Zhou, S.A. Marhon, L. Xie, H. Su, SAVMD: An adaptive signal processing method for identifying protein-coding regions, Biomed. Signal Process, Control 70 (2021)

[43] Q. Zheng, T. Chen, W. Zhou, L. Xie, H. Su, Gene prediction by the noise-assisted MEMD and wavelet transform for identifying the protein-coding regions, Biocybern. Biomed. 41 (1) (2021) 196–210

[44] V. Pathak, S.J. Nanda, A.M. Joshi, S.S. Sahu, VLSI implementation of the anti-notch lattice structure for identification of exon regions in Eukaryotic genes, IET Comput. Digit. Tech. 14 (5) (2020) 217–229

[45] S. Kar, M. Ganguly, S. Das, Using DIT-FFT algorithm for identification of the protein-coding region in the eukaryotic gene, Biomed. Eng. Appl. Basis Commun. 31 (01) (2019)

[46] A. Singh, V. Srivastava, Bidirectional filtering approach for the improved protein coding region identification in eukaryotes, Netw. Model. Anal. Health Inform. Bioinforma. 11 (2022)

[47] S. Kar, M. Ganguly, S. Das, Using DIT-FFT algorithm for identification of the protein-coding region in the eukaryotic gene, Biomed. Eng.: Appl., Basis Commun. 31 (01) (2019).

[48] S. Putluri, M.Z.U. Rahman, C.S. Amara, N. Putluri, New Exon Prediction Techniques Using Adaptive Signal Processing Algorithms for Genomic Analysis, IEEE Access 7 (2019) 80800–80812.

[49] P.K. Kumari, J.B. Seventline, A Novel Approach For Identification Of Exon Locations In DNA Sequences Using GLC Window, Int. J. Biol. Biomed. Eng. 15 (2021) 47–60.

[50] R.K. M, N.K. Vaegae, Walsh code based numerical mapping method for the identification of protein-coding regions in eukaryotes, Biomed. Signal Process. Control. 58 (2020) 101859,

[51] S. Kar, M. Ganguly, S. Ghosal, Prediction of coding region and mutations in Human DNA by effective numerical coding and DSP technique, in 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021.

[52] National center for biotechnology information. https://www.ncbi.nlm.nih.gov

[53] Sanja Rogic, HMR195. https://srogic.wordpress.com/datasets/hmr195-dataset.

[54] SMR Taha, WA Mahmood "New techniques for Daubechies wavelets and multiwavelets implementation using quantum computing" Facta universitatis-series: Electronics and Energetics 26 (2), 145-156, 2013.

[55] WAM Al-Jawher, SH Awad "A proposed brain tumor detection algorithm using Multi wavelet Transform (MWT)" Materials Today: Proceedings 65, 2731-2737, 2022.

[56] WA Mahmoud "Computation of Wavelet and Multiwavelet Transforms Using Fast Fourier Transform" Journal Port Science Research 4 (2), 111-117, 2021.

[57] WAM Al-Jawher, SAA SHABAN "CLUSTERING OF MEDICAL IMAGES USING MULTIWAVELET TRANSFORM AND K-MEANS ALGORITHMS" Journal Port Science Research 5 (1), 35-42, 2022.

[58] SM Saadi, WAM Al-Jawher " Proposed DeepFake Detection Method Using Multiwavelet Transform" International Journal of Innovative Computing 13 (1-2), 61-66, 2022.

[59] AH Salman, WAM Al-Jawher "A Hybrid Multiwavelet Transform with Grey Wolf Optimization Used for an Efficient Classification of Documents" International Journal of Innovative Computing 13 (1-2), 55-60, 2022.

[60] WA Mahmoud, AI Abbas, NAS Alwan "Face Identification Using Back-Propagation Adaptive Multiwavelet" Journal of Engineering 18 (3), 2012.

[61] Waleed A Mahmoud, MR Shaker "3D Ear Print Authentication using 3D Radon Transform" proceeding of 2nd International Conference on Information & Communication Technologies, Pages 1052-1056, 2006.

[62] Waleed A Mahmoud, Afrah Loay Mohammed Rasheed "3D Image Denoising by Using 3D Multiwavelet" AL-Mustansiriya J. Sci, vol 21, issue 7, pp. 108-136, 2010.

[63] W. A. Mahmoud & I.K. Ibraheem "Image Denoising Using Stationary Wavelet Transform" Signals, Inf. Patt. Proc. & Class., vol 46, issue 4, pp. 1-18, 2003.

[64] Waleed A Mahmoud, Ali A Ali, & Saad N Abdul Majed "Wavelet Based Multi Carrier Code Division Multiple Access" proceeding of 5th International Multi-Conference on Systems, Signals and Devices, pages 1-6, 2008.

[65] Abbas Hasan Kattoush, Waleed Ameen Mahmoud Al-Jawher, Sulaiman M Abbas, Ali Tweij Shaheen "A N-Radon Based OFDM Trasceivers Design and Performance Simulation Over Different Channel Models" Journal of Wireless Personal Communications, Volume 58, Pages 695-711, 2011.

[66] AHM Al-Heladi, W. A. Mahmoud, HA Hali, AF Fadhel "Multispectral Image Fusion using Walidlet Transform" Advances in Modelling and Analysis B, vol 52, issue 1-2, pp. 1-20, 2009.

[67] Hamid M Hasan, Waleed A. Mahmoud Al- Jawher, Majid A Alwan "3-d face recognition using improved 3d mixed transform" Journal International Journal of Biometrics and Bioinformatics (IJBB), Volume 6, Issue 1, Pages 278-290, 2012.

[68] Waleed Ameen Mahmoud "A Smart Single Matrix Realization of Fast Walidlet Transform" Journal International Journal of Research and Reviews in Computer Science, Volume 2, Issue, 1, Pages 144-151, 2011.

[69] Waleed A. Mahmud Al-Jawhar "Feature Combination & Mapping Using Multiwavelet Transform" Al-Rafidain Univ. College journal, Issue 19, Pages: 13-34, 2006.

[70] Abbas Hasan Kattoush, Waleed Ameen Mahmoud Al-Jawher, Sulaiman M Abbas, Ali Tweij Shaheen "A N-Radon Based OFDM Trasceivers Design and Performance Simulation Over Different Channel Models" Journal of Wireless Personal Communications, Volume 58, Pages 695-711, 2011.

[71] Walid A Mahmoud, Majed E Alneby, Wael H Zayer "2D-multiwavelet transform 2D-two activation function wavelet network-based face recognition" J. Appl. Sci. Res, vol. 6, issue 8, 1019-1028, 2010.

[72] Ali Akram Abdul-Kareem, Waleed Ameen Mahmoud Al-Jawher, "Image Encryption Algorithm Based on Arnold Transform and Chaos Theory in the Multi-wavelet Domain", International Journal of Computers and Applications, Vol. 45, Issue 4, pp. 306-322, 2023.

[73] Ali Akram Abdul-Kareem, Waleed Ameen Mahmoud Al-Jawher "Hybrid image encryption algorithm based on compressive sensing, gray wolf optimization, and chaos" , Journal of Electronic Imaging, Volume 32, Issue 4, Pages 043038-043038, 2023.

[74] Ali Akram Abdul-Kareem, Waleed Ameen Mahmoud Al-Jawher, "URUK 4D DISCRETE CHAOTIC MAP FOR SECURE COMMUNICATION APPLICATIONS" Journal Port Science Research, Vol. 5, Issue 3, PP. 131-141, 2023.

[75] Ali Akram Abdul-Kareem, Waleed Ameen Mahmoud Al-Jawher, "Image Encryption Algorithm Based on Arnold Transform and Chaos Theory in the Multi-wavelet Domain" International Journal of Computers and Applications, Volume 45, Issue 4, Pages 306-322, 2023.

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 6, No:4. 2023

[76] Ali Akram Abdul-Kareem, Waleed Ameen Mahmoud Al-Jawher, "WAM 3D discrete chaotic map for secure communication applications" International Journal of Innovative Computing, Volume 13, Issue 1-2, Pages 45-54, 2022.

[77] Ali Akram Abdul-Kareem, Waleed Ameen Mahmoud Al-Jawher" Hybrid image encryption algorithm based on compressive sensing, gray wolf optimization, and chaos" Journal of Electronic Imaging, Volume 32, Issue 4, Pages 043038-043038, 2023.

[78] Zahraa A Hasan, Suha M Hadi, Waleed A Mahmoud, "Speech scrambler with multiwavelet, Arnold Transform and particle swarm optimization" Journal Pollack Periodica, Volume 18, Issue 3, Pages 125-131, 2023.

[79] Ali Akram Abdul-Kareem, Waleed Ameen Mahmoud Al-Jawher "A Hybrid Domain Medical Image Encryption Scheme Using URUK and WAM Chaotic Maps with Wavelet–Fourier Transforms" Journal of Cyber Security and Mobility, Pages 435–464-435–464, 2023.

[80] W. A. Mahmoud Al-Jawher, Zahraa A Hasan, Suha M. Hadi," Time Domain Speech Scrambler Based on Particle Swarm Optimization" International Journal for Engineering and Information Sciences, Vol. 18, Issue 1, PP. 161-166, 2023.