Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Study of effectiveness of FIR and IIR filters in Exon identification: A comparative approach

Subhajit Kar *, Madhabi Ganguly

West Bengal State University, Barasat, West Bengal, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Genomic Signal Processing
Deoxyribonucleic acid (DNA)
FIR filter
IIR filter
Kaiser Window

ABSTRACT

In recent years, DSP has been widely used in the area of DNA sequence analysis, detection of protein coding and non-coding regions and also in finding abnormalities present in coding and non-coding regions. Using DSP tool, the detection of protein coding and non-coding regions has been executed with great accuracy and less complexity. In this paper, DNA nucleotide sequence is converted into corresponding EIIP sequence in the first step. In order to make a comparative study between different FIR and IIR filters, the sequence is passed through overall six FIR and IIR filters separately to extract the period three frequency components. Finally, Gaussian filter is used to suppress the high frequency noise present in the power spectrum. This study aims to find out most efficient FIR and IIR filters in predicting exons (coding regions) and introns (non-coding regions) based on different evaluating parameters as follows: (i) specificity-sensitivity values (ii) Matthews correlation coefficient (iii) Miss rate and wrong rate (iv) Discriminating factor (v) ROC.

Selection and peer-review under responsibility of the scientific committee of the International Conference on Artificial Intelligence & Energy Systems.

1. Introduction

A DNA molecule is a long double helical linear polymeric chain, which stores the genetic information of living organism. It is composed of four types of nitrogen bases: Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). Genes are the regions in DNA that carry instructions to make proteins. The eukaryotic genes are divided into exons (protein coding region) and introns (non-coding region). Prokaryotic genes do not contain introns. The structure and protein synthesis process of prokaryotic and eukaryotic genes are depicted in Fig. 1. The splicing process removes the non-coding area i.e., introns leaving only coding area i.e., exons, which contain necessary information for protein synthesis in a DNA sequence. Protein structure which is stored in DNA depends on genetic code. Fig. 2 shows the genetic code in which the 20 amino acids are depicted by their symbols. All proteins are made up of combinations of 20 different amino acids except two i.e., tryptophan and methionine which have unique codons- UGG and AUG, respectively. The period 3-property has been observed experimentally in the protein coding region due to non-uniform codon usage in the translation of codon into amino acids. Due to this, a protein coding region of length N, exhibits a relatively large value at discrete frequency k = N/3 (i.e. the spectrum of protein coding DNA has a peak at every third component at the frequency of $2\pi/3$) but near zero value at N/6, N/9 and other frequencies in corresponding magnitude spectrum [1,2]. Many digital signal processing methods are utilized to effectively extract 'period three frequency component' and suppress background noise. These include Discrete Fourier Transform [3], Spectral Rotation [4], Digital Filters [5], Signal Boosting method [6], Wavelet Transform [7], Walsh Hadamard transform [8], Z Transform [9], Fast Fourier Transform [10], Parametric method [11] etc.

Digital filter-based algorithm employing Antinotch filter to predict the coding regions for a given gene was first proposed by Vaidyanathan and Yoon [12]. The Antinotch filter was later improved by Hota et al. [13]. They introduced three Antinotch filters, namely conjugate suppression Antinotch filter, Antinotch filter followed by moving average filter and harmonic suppression Antinotch filter to

Abbreviations: DNA, Deoxyribonucleic acid; FIR, Finite impulse response; IIR, Infinite Impulse response; EIIP, Electron ion interaction potential; NCBI, National center for biotechnology information; PSD, Power spectral density.

* Corresponding author at: West Bengal State University, India.
E-mail addresses: subhajitkar.wbsu@gmail.com (S. Kar), ray_madhabi@yahoo.co.in (M. Ganguly).

https://doi.org/10.1016/j.matpr.2022.02.394
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the International Conference on Artificial Intelligence & Energy Systems.

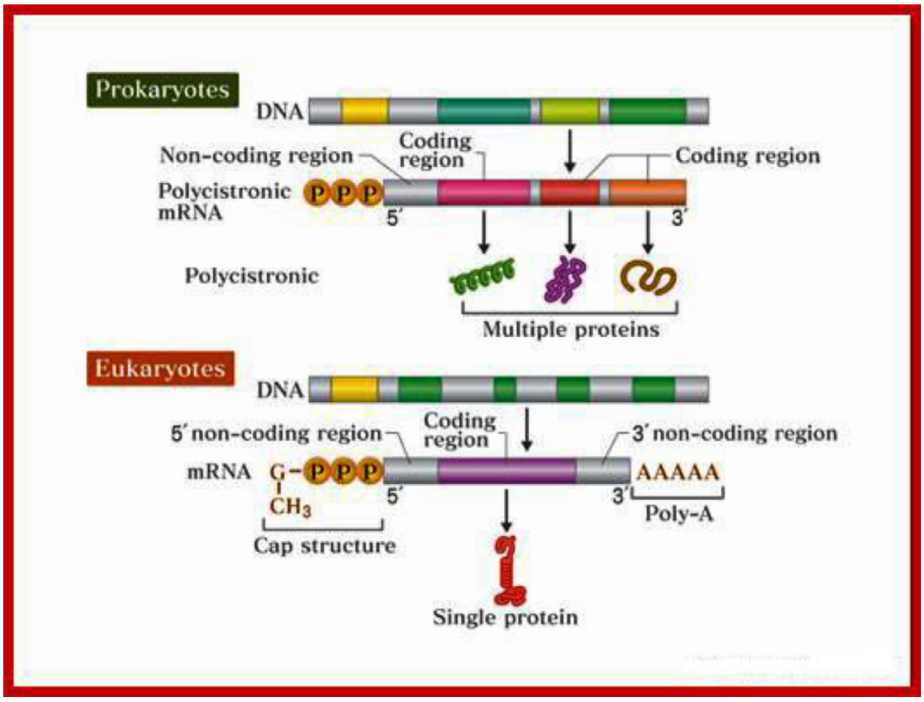

Fig. 1. Protein Synthesis by Prokaryotic and Eukaryotic gene.

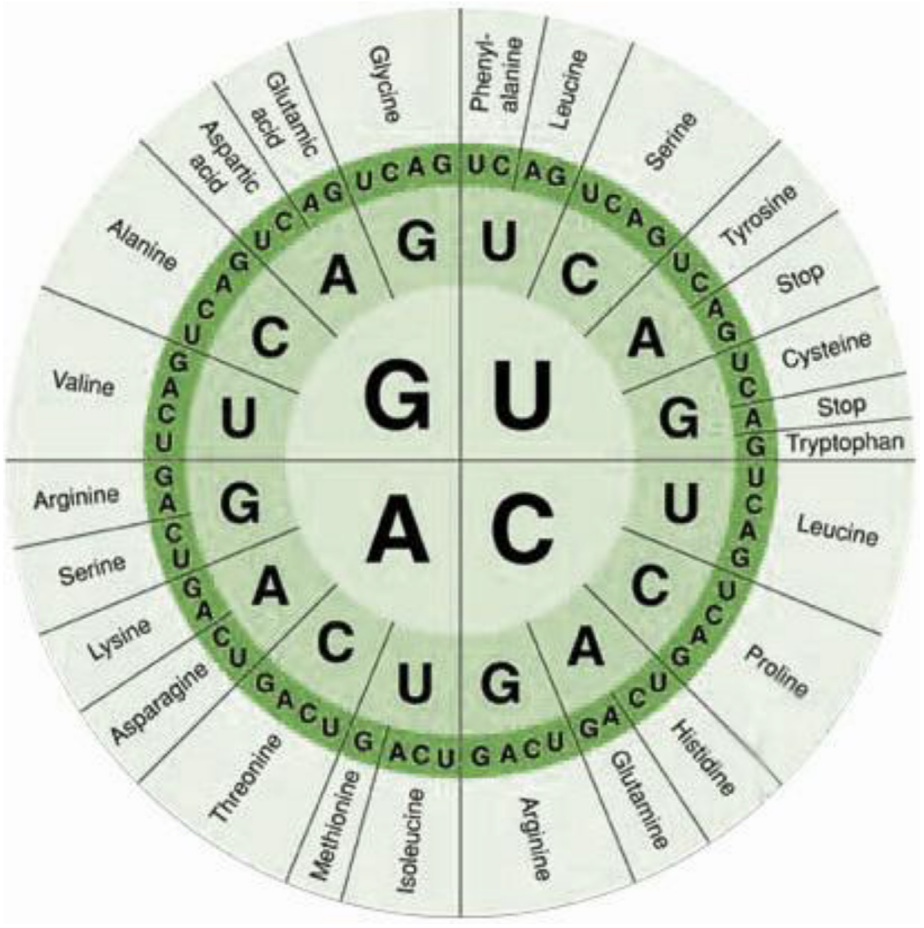

Fig. 2. Genetic code of Amino acids.

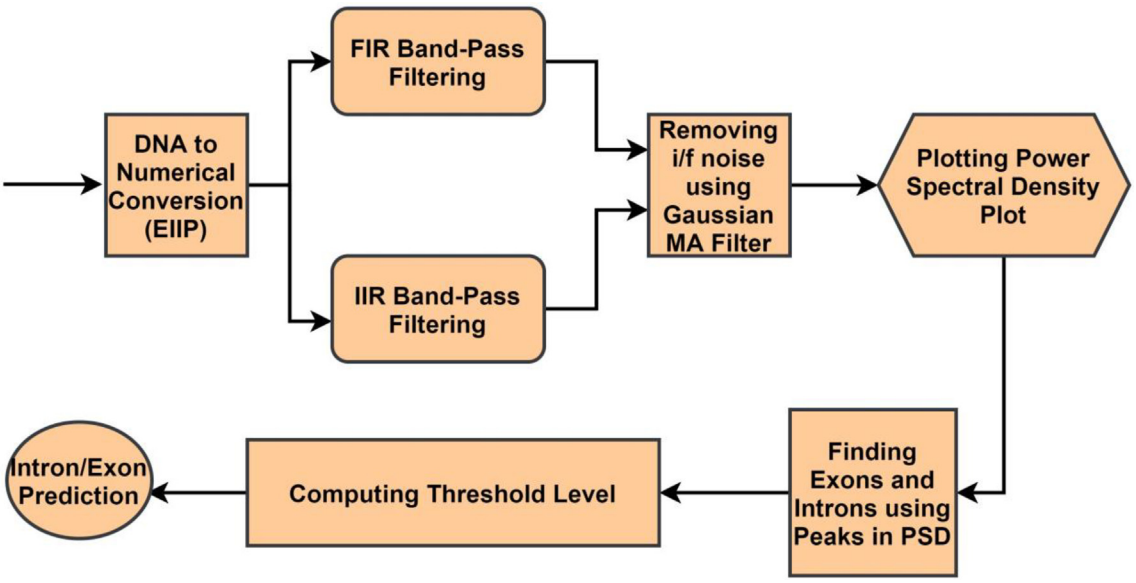

Fig. 3. Block diagram of the proposed method of exon identification.

Table 1
EIIP values of DNA nucleotides.

Nucleotides	EIIP Values
A	0.1260
G	0.0806
T	0.1335
C	0.1340

Table 2
Filter Specification of chosen FIR and IIR filters for exon prediction.

Filter Name	Design Specification
FIR P-M	Pass-Band Ripple = 0.4 dB Stop-Band Ripple = 30 dB
FIR Kaiser Window	Pass-band Cut-off frequencies = [0.665, 0.667]
FIR Least Square	Stop-Band Cut-off frequencies = [0.65,0.68]
IIR Butterworth	Pass-Band Ripple = 0.4 dB Stop-Band Ripple = 30 dB
IIR Chebyshev	Pass-band Cut-off frequencies = [0.664, 0.672]
IIR Elliptical	Stop-Band Cut-off frequencies = [0.659, 0.678]

improve the identification accuracy. A simple model using a single band pass filter followed by a quadratic windowing operation to detect the coding region was suggested by Fox and Carriera which would suppress false peaks present in power spectra [14]. The quadratic window produces a signal that has almost zero energy in the noncoding regions, but this method was not suitable for detecting small coding regions where strand symmetry may not stand. To detect the small coding region effectively Tomar et al. proposed improved Antinotch filtering with Blackman windowing, a Harmonic Suppression Comb filter, and a parametric Minimum Variance spectral estimator to analyse DNA sequences. Results showed small genes of length of just 100 s of nucleotides are visible at the window size of only 117 samples [15]. The Comb filter is further modified by Meher et al. They designed two new signal processing filters namely generalized Comb filter, and cascaded differentiator Comb filter which can effectively pluck the period-3 property in a genomic sequence for the prediction of protein coding regions [16]. Recently Zhang proposed a modified statistically optimal null filter method for recognizing protein-coding regions [17]. The model performs well without the limit of window width. Faransi in his paper, leveraged the filtering technique by combining linear predictive coding model (LPC) with modified Antinotch filter [18]. The LPC algorithm successfully eliminated background noises and reduced correlation between nucleotides. Various forms of adaptive filtering techniques including least mean square and recursive mean square are experimented in Putluary et al [19]. Although the adaptive filter models are efficient in predicting exons, its time complexity is high due to repetitive iteration procedures. Adaptive filtering is adopted in Das et al due to its robust approach and effectiveness across all types of genes. Their work proposed an adaptive Kaiser window, which utilised the recursive Gauss-Newton tuning to vary side lobe height control parameter β [20]. Kumari in her paper, designed Gaussian-Lanczos-Chebyshev window-based FIR filter whereas Kumar applied modified Barlett-Hanning window to detect the coding regions in a particular gene [21,22]. Kar et al provided a method to predict translated & untranslated regions in m-RNA sequences using IIR band-pass filter [23]. In this paper, we have designed and compared different types of band pass filters belonging to the FIR and IIR family to explore the exon prediction relative efficiency. The digital filters are used to extract the period three components and sampling out the noise and other frequency contents.

2. Materials and methods

2.1. Algorithm and flowchart

Fig. 3 below depicts the procedure employed in our experiment to extract coding and non-coding regions from a given DNA nucleotide sequence:

1. The nucleotide sequence of gene F.56F11.4a on chromosome III of Caenorhabditis Elegans in FASTA format is collected from open access NCBI website.
2. Benchmark datasets Asp67, BG570, and GENESCAN were downloaded from website ftp://ftp.cse.ucsc.edu/pub/dna/genes whereas MAIZE dataset obtained from Yao et al. [24].
3. Converting the nucleotide sequences into equivalent numerical sequences using EIIP mapping method.

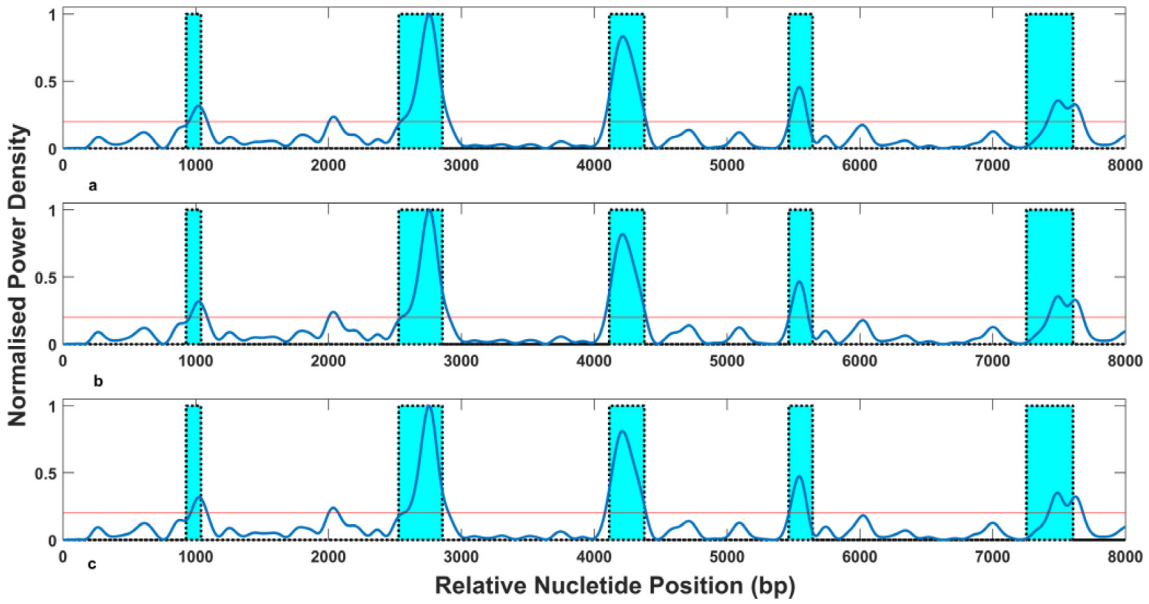

Fig. 4. Exon regions predicted on the gene sequence F56F11.4a by applying FIR filters (a) P-M (b) Kaiser (c) L-S.

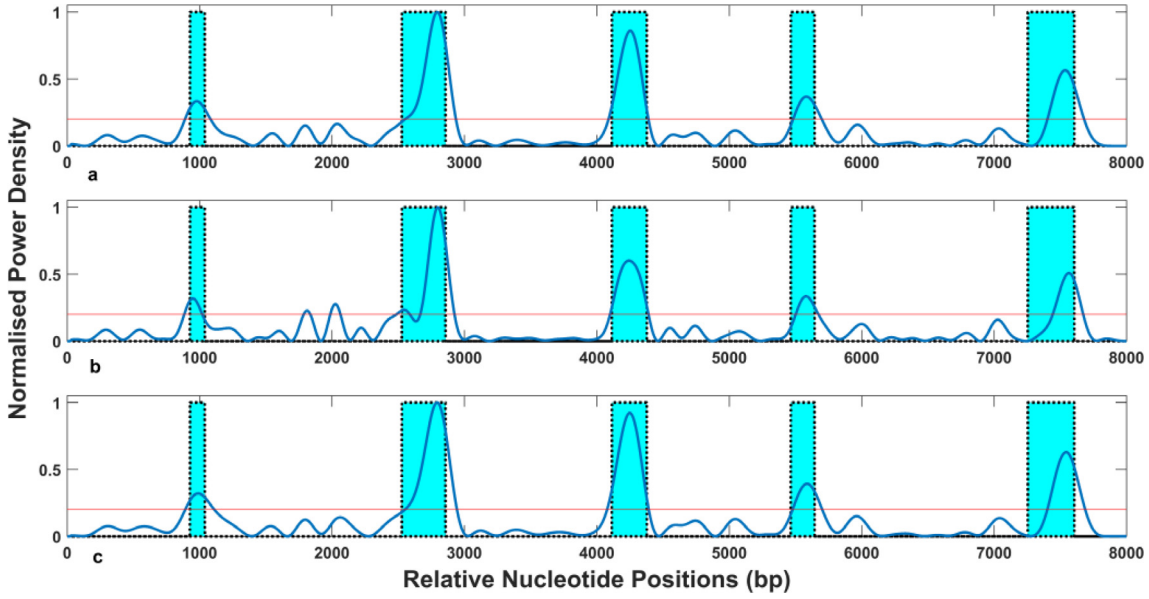

Fig. 5. Exon regions predicted on the gene sequence F56F11.4a by applying IIR filters (a) Butterworth (b) Chebyshev-II (c) Elliptical.

4. Extracting the 'Three base periodicity' components using FIR/IIR bandpass filter using suitable parameters.
5. Removing high frequency noise by Gaussian kernel filter. Output of filter is calculated as power spectrum density. Definitive peaks in spectrum indicate exons.
6. Predicting exons and introns in the given nucleotide sequence according to chosen threshold value.

2.2. DNA sequence database

In order to demonstrate the performance of FIR and IIR filters we have chosen the DNA sequence of gene F.56F11.4a on chromosome III of Caenorhabditis Elegans commonly known as C Elegans (accession number AF099922 from the NCBI website). C Elegans is a type of worm, which comes from the same ancestor as human. The gene has five distinct exons, relative to nucleotide position 7021 according to the NCBI database. These regions are 3156–3267, 4756–5085, 6342–6605, 7693–7872 and 9483–9833 [25]. The accuracy of the model is also tested on various other datasets.

2.3. DNA numerical representation and prediction of protein coding region using digital filter

DNA character strings are mapped into one or more numerical sequences. As the selection of encoding schemes greatly influence the results of exon and intron finding problem, choosing a best suited numeric representation is necessary for this application. Various types of encoding schemes and their properties are largely described in Ning Yu et al. [26]. In this work, we applied the Electron Ion Interaction Potential (EIIP) mapping method based on physico-chemical property for converting symbolic DNA sequences

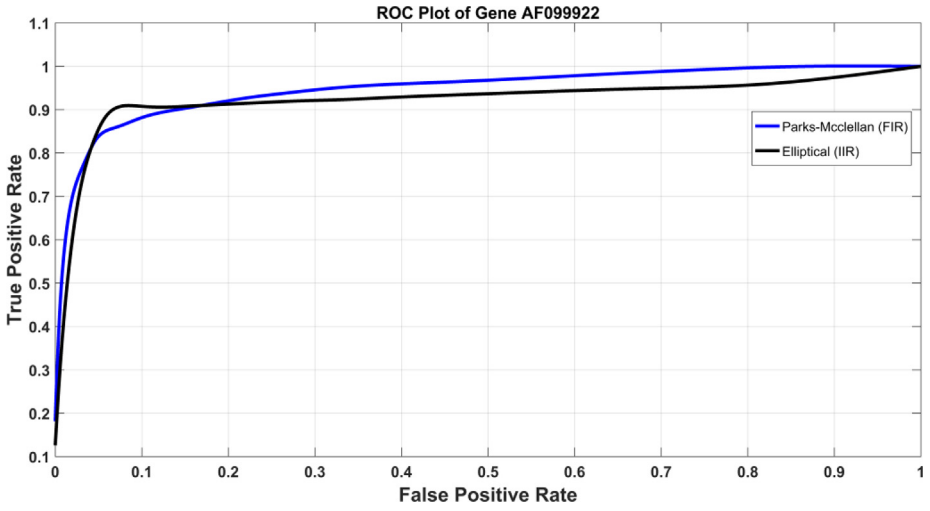

Fig. 6. Receiver Operating Characteristic Plot of Gene AF099922.

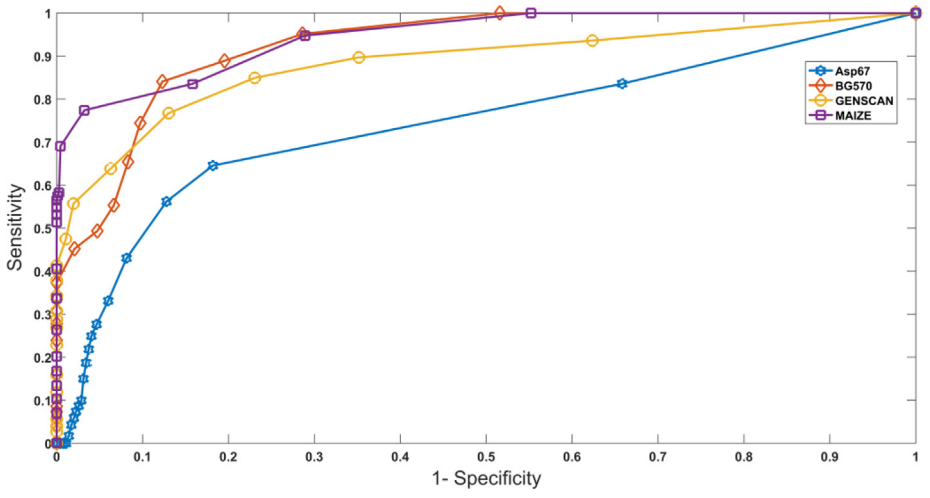

Fig. 7. ROC plot of various benchmark datasets evaluated on Parks-McClellan FIR filter.

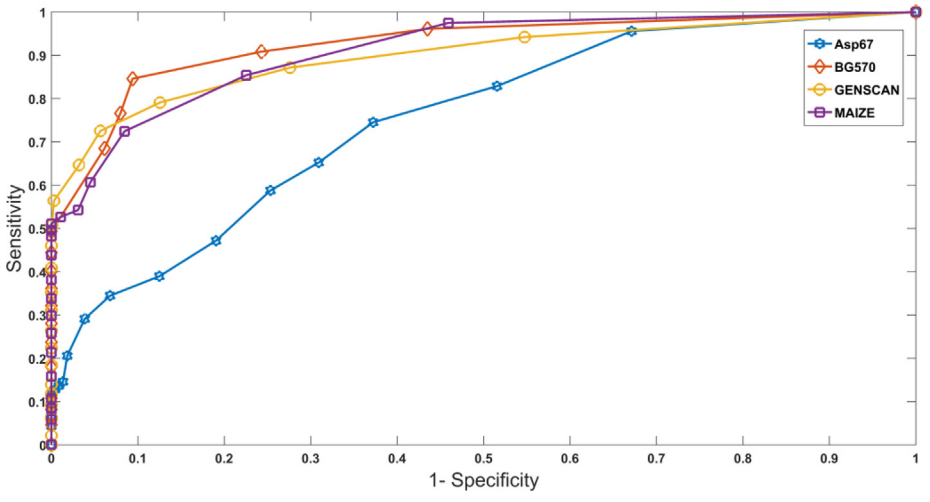

Fig. 8. ROC plot of various benchmark datasets evaluated on IIR Elliptical filter.

Table 3
Gene prediction accuracy at nucleotide level for F56F11.4a at Th = 0.2.

Type of Filter	Sensitiviy S_n	Specificiy S_p	Matthews Correlation Coefficient MCC	Precision	Accuracy
FIR (Kaiser Window)	0.81	0.96	0.75	0.77	0.93
FIR (Parks-Macllelan)	0.82	0.95	0.76	0.77	0.93
FIR (Least square)	0.79	0.96	0.74	0.77	0.93
IIR (Butterworth)	0.84	0.95	0.76	0.76	0.94
IIR (Chebyshev-II)	0.74	0.94	0.66	0.69	0.91
IIR (Elliptical)	0.85	0.94	0.74	0.73	0.93

Table 4
Gene prediction accuracy at exon level for F56F11.4a at Th = 0.2.

Type of filter	Discriminating Factor (D)	Miss Rate (M_R)	Wrong Rate (W_R)	Peak Ratio (PR)	Signal to Noise Ratio (SNR)
FIR (Kaiser Window)	1.78	0	0.17	3.14	1.2
FIR (Parks-Macllelan)	1.81	0	0.17	3.15	1.23
FIR (Least square)	1.75	0	0.17	3.14	1.17
IIR (Butterworth)	2	0	0	3	1.28
IIR (Chebyshev-II)	1.99	0	0.29	3.13	1.11
IIR (Elliptical)	2.18	0	0	3.098	1.28

Table 5
Nucleotide ranges for exons and introns measured using deployed filters for sequence F56F11.4a at Th = 0.2.

Type of Filter	Order of Filter	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5
FIR (Kaiser Window)	266	950–1094 (144)	2532–2913 (381)	4082–4397 (315)	5467–5620 (153)	7412–7690 (278)
FIR (Parks-Macllelan)	216	964–1094 (130)	2558–2916 (358)	4084–4397 (313)	5468–5619 (151)	7421–7683 (262)
FIR (Least-square)	231	962–1090 (128)	2560–2910 (350)	4086–4393 (307)	5469–5618 (149)	7426–7685 (259)
IIR (Butterworth)	6	889–1086 (197)	2562–2935 (373)	4081–4390 (309)	5482–5701 (219)	7402–7681 (279)
IIR (Chebyshev-II)	4	881–1020 (139)	2664–2927 (263)	4099–4391 (292)	5499–5675 (176)	7440–7675 (235)
IIR (Elliptical)	3	891–1113 (222)	2555–2943 (388)	4077–4392 (315)	5478–5702 (224)	7400–7688 (288)
NCBI RANGES		928–1039 (111)	2528–2857 (329)	4114–4377 (263)	5465–5644 (179)	7255–7605 (350)

Table 6
Details information of designed filters.

Filter Type	Filter Structure	Filter length	Implementation Cost					
			Numbers of Multipliers	**Numbers of Adders**	**Numbers of States**	**Multiplications per input Samples**	**Additions per input Samples**	**Average Time Elapsed (ms)**
KaiserWindow	Direct Form- II Transposed	267	267	266	266	267	266	68
Parks-Macllelan	Direct Form-II Transposed	217	217	216	216	217	216	165
Least-Square	Direct Form- II Transposed	232	232	231	231	232	231	0.05
Butterworth	Direct Form- II Transposed	13	19	18	12	19	18	93
Chebyshev-II	Direct Form- II Transposed	9	17	16	8	17	16	80
Elliptical	Direct-Form-II Transposed	7	12	11	6	12	11	77

Table 7
Evaluated AUC values for various Datasets.

Type of Filter	Area Under ROC plot				
	AF099922	**Asp67**	**BG570**	**GENSCAN**	**MAIZE**
Elliptical (IIR)	0.9235	0.7514	0.9252	0.8990	0.9097
P-M (FIR)	0.9473	0.7491	0.9259	0.8930	0.9451

to numerical signals [27–30]. The EIIP values for the nucleotides are given in Table 1.

If we substitute the EIIP values for A, C, G, and T in a DNA string X[n], a numerical sequence is obtained which represents the distribution of the free electrons energy along the DNA sequence. The corresponding numerical sequence is termed as Y[n]. Digital band pass filters can be used to extract the period three frequency components whereas sampling out the noise and other frequency contents. Further, a moving average low pass filter employing Gaussian window is used to eliminate the high frequency components present in the output spectra due to long range correlation present in DNA sequences [31]. A selected sequence is passed

through the FIR or IIR filters with specification given in Table 2, to visualize their ability in finding exons of gene F.56F11.4a and others dataset.

Normalized power spectral density is calculated from the output signal of moving average filter using the formula:

$$S[k] = (Y[k])^2$$

The output PSD plot is normalized 0–1 to get the peaks as exons. We have employed zero phase digital filtering in MATLAB environment by processing the input data in both the forward and reverse directions so that the phase lag of the output signal is nullified. This will result in zero phase distortion at output.

3. Results

In this paper we applied a digital filter-based algorithm to predict the protein coding region in gene sequence. We have used filtering techniques which include bandpass FIR or IIR filter at first stage and then weighted moving average filter at second stage. In this paper we compared the efficiency of different FIR and IIR bandpass filtering method.

Figs. 4 & 5 are the outputs of FIR and IIR filters. The shaded regions are the actual coding regions of gene F.56F11.4a. Figs. 6, 7 and 8 represents various Receiver Operating Characteristic plots to evaluate the designed filters.

Table 3 summarizes the performance of the filters at nucleotide level, whereas Table 4 shows their productivity at exon level. All the evaluation parameters are calculated with respect to the threshold value 0.2. This means that if the peak of an intron region greater than 0.2 it will be considered as false positive, otherwise it will be considered as true negative.

It is found that in general all the filters show distinct peaks at protein coding regions and thus fruitful to locate introns and exons of a given nucleotide sequence. The Elliptical and Butterworth filters output showing exactly five exons and seven introns. However Kaiser, Parks-McClellan (P-M) and L-S filters produce one false exon and thus resulting a total of six exons and eight introns in power spectrum. The application of Chebyshev-II filter on the given sequence resulting in two false exons. Results in terms of these evaluation criterions clearly suggest P-M among the FIR filters and Elliptical among the IIR filters are best in terms of finding exons and introns in eukaryotic gene. Although the values of Specificity, MCC & Precision parameters are slightly better in case of Butterworth filter comparing to the Elliptical one, filter order is half in case of Elliptical which can be verified in Table 6. It is found that the accuracy level of the P-M technique is 0.93 and that is of Elliptical filter is also 0.93, suggesting the methods are very much productive in this context. Results also indicate the fact that IIR filters performance is slightly better than its FIR counterparts in terms of wrong rate. In all the filters experimented, the P-M filtering methods show best results compared to the other filtering approaches.

We have summarized the predicted introns and exons locations in Table 5. Filter order is also specified accordingly. The results suggest our experimental findings are very close to NCBI ranges. The figures in the brackets indicate the length nucleotides in each exon.

From the above table it is clear that IIR filter can achieve the same results with lower order filter. This information will come handy while choosing between FIR and IIR filter because lower order of filter refers to minimum hardware requirement while designing a practical filter. Table 6 summarize details information of two filters employed in our experiment.

To compare the performance of FIR & IIR filters, Receiver Operating Characteristic (ROC) curves are plotted for gene AF099922 and four others benchmark datasets. The results are listed in Figs. 6, 7 and 8.

ROC curve measures sensitivity and specificity at various threshold level to evaluate the performance of the filters. Area under ROC curve which is known as AUC are measured for P-M and Elliptical filters. The AUC value close to 1 supposed to provide good classification accuracy. The Higher is the AUC, the better is the model at discriminating between exons and introns. Table 7 is used to report the measured AUC values.

4. Conclusions

To measure the effect of various filtering methods to extract protein coding regions, we have estimated different evaluation parameters. The evaluation process is done in three ways:

i) Prediction accuracy at nucleotide level.
ii) Prediction accuracy at exon level.
iii) Receiver Operating Characteristic plot.

Proposed algorithm figured out the exon locations of gene F.56F11.4a precisely. To analyze the results of our experiment and study the performance of FIR and IIR filters, different evaluation parameters are measured. From the result it is evident that Parks-McClellan(P-M) FIR filter produces most accurate results in terms of ROC plot. Some of the important characteristics of FIR and IIR filters which are reflected in the experiment are:

i. Firstly, the number of necessary multiplications is least in IIR filter and most in FIR filter.
ii. Secondly linear phase is seen in IIR filters but there is no linear phase is observed in FIR filters.
iii. Thirdly Required coefficient memory in IIR filters is less than in FIR filters and
iv. Finally, stability must be designed in IIR filters whereas it is guaranteed in FIR filters.

CRediT authorship contribution statement

Subhajit Kar: Conceptualization, Methodology, Software, Validation, Data curation, Writing – review & editing. **Madhabi Ganguly:** Investigation, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Anastasios, Periodicity in DNA coding sequences: implications in gene evolution, J. Theor. Biol 151 (1991) 323–351.
[2] J. Tuqan, A. Rushdi, A DSP perspective to the period – 3 detection problem, Proceedings of the IEEE International Workshop on Genomic Signal Processing and Statistics, GENSIPS (2006), pp. 53-54.
[3] S. Tiwari, S. Ramachandran, A. Bhattacharya, S. Bhattacharya, R. Ramaswamy, Prediction of probable genes by Fourier analysis of genomic sequences, Bioinformatics 13 (3) (1997) 263–270, https://doi.org/10.1093/bioinformatics/13.3.263.
[4] D. Kotlar, Y. Lavner, Gene Prediction by Spectral Rotation Measure: A New Method for Identifying Protein-Coding Regions, Genome Res. 13 (8) (2003) 1930–1937, https://doi.org/10.1101/gr.1261703.
[5] X. Zhang, H. Iwakura, Design of IIR digital filters using complex allpass network based on eigenvalue problem, Electron. Lett. 28 (3) (1992) 231, https://doi.org/10.1049/el:19920143.
[6] T.S. Gunawan, J. Epps, E. Ambikairajah, Boosting approach to exon detection in DNA sequences, Electron. Lett. 44 (4) (2008) 323–324.

[7] S. Deng, L. Yuan, K. Feng, G. Ding, Y. Li, A New Approach for Identifying Protein-Coding Regions by Combining Chirp z and Wavelet Transform, Curr. Bioinform. 8 (2013) 557–563, https://doi.org/10.2174/1574893611308050006.
[8] J. A. Berger, S. K. Mitra, J. Astola. Power spectrum analysis for DNA sequences. Proc. of Seventh International Symposium on Signal Processing and its Applications (2003), 2:29-32.
[9] R. Zhang, C. Zhang, Z. Curves, An Intutive Tool for Visualizing and Analyzing the DNA Sequences, J. Biomol. Struct. Dynam. 11 (1994) 767–782, https://doi.org/10.1080/07391102.1994.10508031.
[10] S. Kar, M. Ganguly, S. Das, Using DIT-FFT algorithm for identification of protein coding region in eukaryotic gene, Biomed. Eng.: Appl., Basis Commun. 31 (01) (2019) 1950002.
[11] N. Chakravarthy, A. Spanias, L.D. Iasemidis, K. Tsakalis, Autoregressive modeling and feature analysis of DNA sequences, EURASIP J. Adv. Signal Process. 2004 (1) (2004).
[12] P.P. Vaidyanathan, B.J. Yoon, Digital Filters for Gene Prediction Applications, IEEE Asilomar Signals, Syst., Comput. (2002) 306–310.
[13] M.K. Hota, V.K. Srivastava, Identification of protein coding regions using antinotch filters, Digital Signal Process. 22 (6) (2012) 869–877, https://doi.org/10.1016/j.dsp.2012.06.005.
[14] T.W. Fox, A. Carreira, A Digital Signal Processing Method for Gene Prediction with Improved Noise Suppression, EURASIP J. Adv. Signal Process. 2004 (1) (2004).
[15] V. Tomar, D. Gandhi, V. Chakka, Advanced Filters for Genomic signal processing, Int. J. Adapt. Control Signal Process.
[16] J. Meher, P.K. Meher, G. Dash, Improved Comb Filter based Approach for Effective Prediction of Protein Coding Regions in DNA Sequences, J. Signal Inform. Process. 02 (02) (2011) 88–99, https://doi.org/10.4236/jsip.2011.22012.
[17] L. Zhang, F. Tian, S. Wang, A Modified Statistically Optimal Null Filter Method for Recognizing Protein-coding Regions, Genom., Proteom. Bioinformat. 10 (3) (2012) 166–173, https://doi.org/10.1016/j.gpb.2012.02.001.
[18] MahsaSaffari Farsani, M.R. Sahhaf, V. Abootalebi, Performance improvement of the Goertzel algorithm in estimating of protein coding regions using modified anti-notch filter and linear predictive coding model, J. Med. Sig. Sens. 6 (3) (2016) 130, https://doi.org/10.4103/2228-7477.186882.
[19] 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.
[20] L. Das, S. Nanda, J.K. Das, An integrated approach for identification of exon locations using recursive Gauss Newton tuned adaptive Kaiser window, Genomics 111 (3) (2019) 284–296, https://doi.org/10.1016/j.ygeno.2018.10.008.
[21] 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, https://doi.org/10.46300/91011.2021.15.7.
[22] 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, https://doi.org/10.1016/j.bspc.2020.101859.
[23] 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, https://doi.org/10.1109/icccis51004.2021.9397102.
[24] H. Yao, L. Guo, Y. Fu, L.A. Borsuk, T.-J. Wen, D.S. Skibbe, X. Cui, B.E. Scheffler, J. Cao, S.J. Emrich, D.A. Ashlock, P.S. Schnable, Evaluation of five ab initio gene prediction programs for the discovery of maize genes, Plant Mol Biol. 57 (3) (2005) 445–460, https://doi.org/10.1007/s11103-005-0271-1.
[25] D.P. Snustad, M.J. Simmons, Principles of genetics, John Wiley & Sons Inc.
[26] 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 Mining Anal. 1 (3) (September 2018) 191–210, https://doi.org/10.26599/BDMA.2018.9020018.
[27] H.K. Kwan, S.B. Arniker, Numerical representation of DNA sequences, 2009 IEEE International Conference on Electro/Information Technology. (2009). doi:10.1109/eit.2009.5189632.
[28] M. Abo-Zahhad, S.M. Ahmed, S.A. Abd-Elrahman, Genomic Analysis and Classification of Exon and Intron Sequences Using DNA Numerical Mapping Techniques, IJITCS 4 (8) (2012) 22–36.
[29] H.M. Wassfy, M.M. AbdElnaby, M.L. Salem, M.S. Mabrouk, A.A. Zidan, Eukaryotic gene prediction using advanced DNA numerical representation schemes, Proc Fifth Int, Kuala Lumpur, Malaysia, 2016.
[30] H. Saberkari, M. Shamsi, M. Sedaaghi, F. Golabi, Prediction of protein coding regions in DNA sequences using signal processing methods, Proceedings of the IEEE Symposium on Industrial Electronics and Applications (ISIEA '12), pp. 355–360,Bandung, Indonesia, September 2012
[31] S. Barman (Mandal), S. Biswas, S. Das, M. Roy,, Performance analysis and simulation of IIR anti-notch filter with various structures for gene predication application, 5th International Conference on Computer and Devices for Communication (CODEC), 2012.