Disclaimer: This is not the final version of the article. Changes may occur when the manuscript is published in its final format.

Computing&AI Connect

SCIFINITI PUBLISHING

ISSN: 3104-4719

2025, Article ID. x, Cite as: https://www.doi.org/10.69709/xxx

Research Article

OPEN ACCESS

Sheaf cohomology for fragment-sequencing in hierarchical block rectangular matrices with spectral gaps in the presence of random effects and white noise: the chain

Orchidea Maria Lecian
Sapienza University of Rome,
Rome, Italy.
email: omlecian@gmail.com; ORCID: 0000-0002-9417-5578

Keywords:

Rectangular-matrix chains; sheaf cohomology; hierarchical block rectangular matrices; fragment sequencing; machine learning.

Abstract

The sheaf cohomology of topological shift for the block-rectangular matrix-representation of the hierarchical Markov Model is endowed with the analytical codification of white noise and of random effects.

New analytical techniques for fragment sequencing are developed. The fragment sequencing is obtained after the topological Markov chain of the adjacency matrix of the corresponding undirected graph; the presence of white noise and that of random effects are comprehended. The paradigm consists in defining the hierarchical block rectangular matrices, from which the Topological Hidden Markov Models are issued (as clusters), with the aspects of Hidden Markov Models of 'multivariate Gaussian data' with vanishing mean; the generalized covariance matrix is studied. The model is compared with the stochastic properties of the corresponding decomposition of approximation of experimental data. One of the previous results of the applications of the new method can be looked at in the analysis of the numerical simulation of the sequencing techniques: as an example, it is known that the Gojobori-Ichii-Nei model fails in reproducing the Jukes-Cantor scheme, while the Kimura matrix model succeeds in it. The difference is explained as the former model is not obtained from the Jukes-Cantor paradigm after application of the differential operators (for substituting the entries of the matrix), while the latter model is.

In the present paper, the new analytical result is further accomplished, to calculate the maximal likelihood analytically in protein sequencing and in DNA-sequencing, in modes in which the sequences of elements varies in time; the method solves analytically the phylogenetics computer programs.

The analysed problem belongs to the nondeterministic polynomial time (NP) hard class of complexity.

1 Introduction

The interrogation of [1] from [2] is in the present paper answered.

The interrogation of [1] from [2] consists in defining the well-posed-ness of the comparison of contact maps with different bin sizes. The relevance of the overlapping of the contact maps is that it allows for the comparison of graphs (and of all the related structures, i.e., such as distances.)

The further interrogation about white noise in [3] and about random effects [4] are codified within the same paradigm.

As recalled from [5], solving the problem of contact-maps overlap belongs to the NP hard class of complexity [6], [7].

A visual method for comparing adjacency matrix is proposed in [8].

The answer to the interrogation of [1] is here provided with analytically according to the following strategy: a (topological Markov) chain is build for the network from the hierarchical block rectangular matrix, where the white noises and the random effects are present, where the clusters are contained: the problem of [1] is solved after proving that the topology of the manifold on which the chain is issued is one with Hilbert metric. The answer to the interrogation of [1] is therefore here newly found analytical; for these purposes the new paradigm in topology was created- the numerical methods are not in the present paper used.

The tasks of graph matching are being afforded in the present guidelines of investigation according ot several methodologies. A spectral method for these purposes is developed in [8].

The adjacency matrix of the graphs to be compared is analyzed ibidem as far as the leading eigenvector (only) is concerned. The sequences are matched after developping the nodes of the graph for these sequences to be aligned.

The corresponding Markov chain is established, which taken into account the leading eigenvector of the adjacency matrix only.

One of the main difference between the method proposed in the preset paper and that proposed in [8] is that, in the here-adopted formalism, the sequence order does not ned to be re-aligned a posteriori.

Indeed, the approach of [8] relies on reappreciating the paradigms of segmentation and those of grouping.

In the here-represented analysis, the isolated segments can be put into block-matrix representation.

Another limitation of [8] is that is descends from the methods developped in [9]. In [9], graphs of the same size only can be compared. Unaccordingly, in the present analysis, graphs of different sizes are made to be compared in univoque correspondence with the pertinent block matrix.

One further discrepancy with the work of [8] is that the point-proximity matrix is analysed, and its eigenvectors are considered, i.e., as from [10], after [11]; as a result, only the Gaussian weighted distances between points can be calculated. The several point sets are compared after the patterns of eigenvectors corresponding to different sequences; the comparison is produced after juxtaposing the 'immanental polynomials of the Laplacian matrix of the 'line connectivity

graphs [12].
Local topological spectra are considered after [13].
One of the attempts to codify the white noise within chains relies in the time-series expansion, i.e. as last proven in [3] for Markov chains. Differently, int he present paper, the white noise of topological shifts is codified in matrix representation.
In [4], the further interrogation is posed, about which items of information have to be added to a model for achieving the Markov representation of random effects; the tasks is in the present paper accomplished.
Differently, in the pairwise Euclidean distance [14] is applied to explain the differential behaviour of clusters with respect to white noise and to random effects after the use of the dissimilarity matrix. The clusters are therefore described as topological Hidden Markov Models, on which the Cameron-Martin distance applies. Two methods of fragment sequencing are proposed starting from the new formalism developped in [15]: one method consists in extending the Jukes-Cantor model after [16]; one method consists in modelling the fragment: both methods comprehend the construction of the fragment for the states of the probability space from the application of the suitable Morse operators from a starting pairwise sequence [17]. The filtration of the probability space is determined after the definition of Hidden Markov Models of 'multivariate Gaussian data' from [18] (i.e. Topological Hidden Markov Models), from which the indications from [19] and those from [20] are applied. The application of these indications has to be compared with the new techniques developed in [21].
Furthermore, starting from [17], white noise and random effects are included within the analytical representation.
The paper is organized as follows.
In Section 2, the methodology is exposed: the analytical tools for fragment sequencing are recalled, and the theorems for the singular-value decomposition of ordered block matrices are reviewed.
In Section 3, the new results are presented: the definition of the chain from the contact map in the presence of random effects and white noises is newly established, and the the Topological Hidden Markov Models (of clusters) are built.
Further results in sequencing are presented: the further interrogation from [22] is newly analytically solved, about how maximize the likeli-hood in DNA sequences and in protein sequences when the substitution of nucleotides or that of amino- acids varies with time. The technique, in particular, substitutes the phylogenetics-programming codes. Moreover, applications in machine learning are theoretized.
In Section 4, applications are achieved, such as the application of Dirichlet forms to obtain a vanishing multivariate distribution mean analytically, and the paradigm for the construction of the models (i.e. independently of the measure for the comparison with [21] to be consistent); as one result, the interrogation of [1] after [2] is answered.
In Section 6,the Conclusions are presented.

2 Methodology

2.1 Fragment-sequencing

In the present Section, I summarize the interrogation raised in [1] after [2].
Definition 1:
Let a 'bin' be a vector of consecutive fragments; more in detail, let $\vec{a} = (a_1, a_2, ..., a_k)$ *be a row vector with* k *fragments. The bin with one fragment is denoted as* $\vec{b} = (b_1, b_2, ..., b_l)$.
Definition 2:
The contact frequency f_{ab} *is defined as*

$$f_{ab} = \sum_{i=1}^{1=k} \sum_{j=1}^{j=l} e_{a_i b_j} \tag{1}$$

with $e_{a_i b_j}$ *the number of 'ligation' events between fragments, and, in the present case, between fragment* a *and fragment* b.
The contact map is defined as the matrix whose entries correspond to pairwise 'contact frequencies' between two vectors of bins. The definition of the contact-map matrix is taken from [23] p.26 is taken as
Definition 3:
A contact map is a matrix $\mathcal{M}$ *whose entries* $m_{ij} = 0$ iff *the Euclidean distance between the two elements (of the sequence)* i *and* j *is less than or equal to a pre-assigned threshold* $\tilde{t}$.
It is here noticed that, differently, in [1], m_{ij} is usually assigned a value 1 when the distance is less than a threshold.
An alternative definition of contact map is issued from [24] as
Definition 4:
A contact map is a square, symmetrical matrix of pairwise contact of residues; this definition is therefore apt for the definition of the pairwise sequencing.
The analysis from [1] is here reappraised.
Let $\vec{p}$ be the row vector of m consecutive bins, i.e. $\vec{p} = (p_1, p_2, ..., p_m)$.
Let $\vec{q}$ be the row vector of n consecutive bins, i.e. $\vec{q} = (q_1, q_2, ..., q_n)$.
Clearly in the notation vector p and the vector q are of non-identical dimensions. This is a peculiarity descending from rectangular matrices, which are applied to the task of fragment sequencing within clusters. **Definition 5**:
$\hat{A}$ *is the contact-map (matrix) whose entries*

$$a_{ij} \equiv f_{p_j q_j} \tag{2}$$

with $1 < i < m$, $1 < j < m$ *when all the bins are in* $\vec{p}$ *and those in* $\vec{q}$.
Definition 6:
Cis contact maps are contact maps obtained in the case when all the bins form $\vec{p}$ *and those form* $\vec{q}$ *are from the same fragment.*
Definition 7:
Trans contact maps are contact maps obtained in the case when all the bins form $\vec{p}$ *and those form* $\vec{q}$ *are not from the same fragment.*

The hierarchical structure of clusters is more apparent after the comparison with [1].

Given all the parameterizations in the above, the scenario is ready for the definition of contact networks.

Definition 8:

$N_{\vec{p}}(V, \mathcal{E}, w)$ is the contact network of $\vec{p}$ if $V = \{ p_1, p_2, ..., p_m\}$ is the set of nodes, $\mathcal{E} \equiv \{ \{ p_i, p_j\} \mid f_{p_i,p_j} > 0\}$ is the set of edges with f_{p_i,p_j} the contact frequencies between the nodes (bins) with $1 \leq i \leq m$, and w the weighting function, i.e. an application which sends $\mathcal{E}$ in $\mathbb{R}^+$ as

$$w : \mathcal{E} \to \mathbb{R}^+; \tag{3}$$

more in detail, the weighting function w assigns to each edge its contact frequencies as

$$w(p_i, p_j) = f_{p_i,p_j}. \tag{4}$$

Remark 1:

It is here anticipated that the weighting functions will be generalized.

Definition 9:

The cis contact-map matrix $\hat{A}(cis \ \ \vec{p})$ is the adjacency matrix of the contact network $N_{\vec{p}}(V, \mathcal{E}, w)$.

The notions of weighted graphs and those of unweighted graphs can therefore be recalled.

Definition 10:

$N_{\vec{p}}(V, \mathcal{E}, w)$ is named an unweighted graphs if the weight w is such that $w \to \mathcal{E} \to \{ 1\}$.

Definition 11:

$N_{\vec{p}}(V, \mathcal{E}, w)$ is a weighted graph otherwise.

Let $\{ C\}$ be the set of clusters in $N_{\vec{p}}(V, \mathcal{E}, w)$; the following definition holds

Definition 12:

The set of clusters $\{ C\}$ is defined as

$$\{ C\} \equiv \{ c_l\}_{l=1}^{l=k} \tag{5}$$

with $c_l \subset V, k \geq 1$.

The Shavit-Walker-Lio' (SWL) matrices $\hat{B}$'s are defined as

- $\hat{B(1)}$ an $m \times m$ matrix whose entries $b(1)_{ij}$ are as

$$b(1)_{ij} = 1, i, j \in c, c \in C, c \subset V, \tag{6a}$$
$$b(1)_{ij} = 0 \ \ otherwise; \tag{6b}$$

and

- $\hat{B(s)}$ an $m \times m$ matrix whose entries $b(s)_{ij}$ are as

$$b(s)_{ij} = 1, i, j \in c, c \in C_{s-1}, c \subset V, \tag{7a}$$
$$b(s)_{ij} = 0 \ \ otherwise; \tag{7b}$$

If the structure is hierarchical, the presence of clusters can be hypothesized. For this reason, the Hierarchical Block Matrices (HBM's) are looked for.
The HBM matrices of N can be here studied.
Definition 13:
The HBM block matrix of $N_{\vec{p}}(V, \mathcal{E}, w)$ *is a non-negative* $m \times m$ *matrix* $\hat{G}$ *whose entries* g_{ij} *are defined as*

$$g_{ij} \equiv min_s\{ \ s \mid b_{s \quad ij} = 1\} \tag{8}$$

with $s \geq 1$*.*

2.2 Singular-value decomposition of ordered block matrices: theorems

From [16], the underlying 'signal' matrix with ordered block structure $\hat{D}$ is considered, i.e. $\hat{D} \in \mathbb{R}^{m \times n}$ that is starting index of J as $m_1 \times n$ to $m_M \times n$.
A rectangular 'observation' matrix $\hat{Y}$ is considered, under the hypothesis that there are M blocks of sizes $(m_1, m_2, ..., m_M)$ with $sum_{i=1}^{i=M} m_i = m$ and N blocks of sizes $(n_1, n_2, ..., n_N)$ with $\sum_{j=1}^{j=N} n_j = n$.
Without loss of generality, $m \leq n$ is assumed.
The observation matrix $\hat{Y}$ can be decomposed according to the entries

$$y_{ij} = d_{H(i)V(j)} + \eta_{H(i)} + \nu_{V(j)} \tag{9}$$

where $H(i)$ and $V(j)$ are block membership indicators with values $(1, ..., M)$ and $(1, ..., N)$.
The column vectors $\eta_{H(i)}$ and $\nu_{v(j)}$ are defined as continuous random variables with vanishing means and standard deviation σ_η and σ_ν, respectively, i.e. they represent whites noises.
The underlying constant signal $\hat{D}$ is therefore isolated from the other components in Eq. (9).
Let $\hat{J}_{a \times b}$ be an $a \times b$ matrix with all entries equal to 1.
The matrices $\hat{D}$, $\hat{\eta}$ and $\hat{\nu}$ in Eq. (9) are written as

$$\hat{D} \equiv \{d_{ij}\}$$

$$\hat{D} = \begin{bmatrix} d_{11} J_{m_1 \times n_1} & ... & d_{1N} J_{m_1 \times n_N} \\ ... & ... & ... \\ d_{M1} J_{m_M \times n_1} & ... & d_{n_1 m} J_{m_M \times n_N} \end{bmatrix}$$

$$\hat{\eta} \equiv \{\eta_{ij}\}$$

$$\hat{\eta} = \begin{bmatrix} \eta_1 J_{m_1 \times n} \\ ... \\ \eta_M J_{m_M \times n} \end{bmatrix}$$

$$\hat{\nu} \equiv \{\nu_{ij}\}$$

$$\hat{D} = \begin{bmatrix} \nu_1 J_{n_1 \times M} \\ ... \\ \nu_M J_{m_M \times n} \end{bmatrix}$$

3 Results

3.1 New Definition of the chain from the contact map in the presence of random effects and white noises

The request of [23] that the distances between the contact maps be Euclidean are here accomplished.
The Matrix ˆY from [15] is here rewritten according to th singular-values decompositions as

$$\hat{Y} \equiv \hat{v}\hat{\mathcal{Y}}\hat{v}^T \tag{10}$$

The matrix $\hat{v}$ and the matrix $\hat{v}$ are reduced to vectors: the vector $\vec{u}$ is recovered in the general case; differently, in the present analysis, if Y is compact, the metric is Euclidean when that of the vector $\vec{v}$ is Hilbert (while the definition of $\vec{u}$ is not in general Hilbert).
This way, the chain $\hat{Q}$ is defined as one on a surface with a Hilbert metric. The chain of the adjacency matrix is the topological Markov chain.
The Euclidean distance from [14] is used to define the 'average pairwise distance'.
From [18], the Cameron-Martin metric is defined from γ as

$$| \gamma |= sup v^T \gamma : v^T \varsigma v \leq 1 \tag{11}$$

on the vector $\vec{v}$ of the singular-value decomposition Eq. (10); the vectors $\vec{v}$ and the vector $\vec{v}$ are made to coincide is the case of fragment sequencing is studied. The role of ς is that of the variance, form which the covariance is taken, which allows one to specify marginal distributions; the choice of ς is specified in Theorem 2.

on the vector $\vec{v}$ of the singular-value decomposition Eq. (10); the vectors $\vec{v}$ and the vector $\vec{v}$ are made to coincide is the case of fragment sequencing is studied. The role of ς is that of the variance, form which the covariance is taken, which allows one to specify marginal distributions; the choice of ς is specified in Theorem 2.

The hypothesis from [18] on the expectation value and on the variance are kept as

$$\mathbb{E}(v) = 0, \tag{12}$$

and

$$Var v = v^T \hat{\varsigma} v. \tag{13}$$

3.2 Clusters as Hidden Markov Models

As in [18], the obtained clusters are Hidden Markov Models of 'multivariate Gaussian data' when the covariance matrix is fixed.
It is the purpose of the next Section to fix the covariance ς.
The mean vector is newly estimated from the log-emission function from [18] in terms of the Baum-Welch backwards probabilities and for the Baum-Welch forward ones and requested to be vanishing in the form of [15].
In the Baum-Welch algorithm, the likeli-hood function L is here written from

[25] on the λ collection of model parameters. The Baum-Welch backward probabilities $\alpha(t)$ and the Baum-Welch forward probabilities $\beta(t)$ are here considered for the observation of the probability space O as

$$\alpha_t(j) = Pr(O_1, O_2, ..., O_t, s_t = j \mid \lambda), \tag{14a}$$

$$\beta_t(j) = Pr(O_{t+1}, ..., O_t, s_t = j \mid \lambda) \tag{14b}$$

The multivariate distribution mean is calculated as

$$\tilde{m_j} = \frac{\sum_{t=1}^{t=T} \alpha_t(j)\beta_t(j)O_t}{\sum_{t=1}^{t=T} \alpha_t(j)\beta_t(j)} \tag{15}$$

and is here newly requested to vanish as

$$\tilde{m_j} \equiv 0. \tag{16}$$

The techniques for making the multivariate distribution mean vanish analytically are discussed in Subsection 4.1; in the following Subsection, the paradigm here developped is proven to be one generating a Markov Process.

The covariance matrix is now fixed from [20], and the Topological Hidden Markov Models are newly built.

3.3 The new Topological Hidden Markov Models (of clusters)

As studied from [20], a Gaussian distribution with vanishing mean is taken.
The specification of the variance ς allows one to specify the marginal distributions.
Gaussian probability measure with 'prescribed marginals' are defined after the joint probability density $\mathcal{P}$ with marginals $\mathcal{P}_{c_1}, \mathcal{P}_{c_2}, ..., \mathcal{P}_{c_l}$. The c_l here used are those from the subset Eq. (5).
A class of Gaussian measures with prescribed margin is newly found after [20] as there exists the covariance matrix $\hat{\varsigma}$ and $\hat{\varsigma}$ is unique.
The probability space $(states, observation, filter)$ is now constructed for the Topological Hidden Markov Models of clusters.
The filter is constructed as from the probability function after which the measure of the probability space is defined.
From [20] p. 139, the random vector field X is considered, with Gaussian distribution and with vanishing mean and positive-definite covariance $\hat{\varsigma}$. The density $p(\vec{X})$ is written as

$$p(\vec{X}) = (2\pi)^{-|C|/2}(det[\hat{\varsigma}])^{-1/2} e^{-\frac{1}{2}\vec{X}^T \hat{\varsigma}^{-1} \vec{X}}, \tag{17}$$

from which the measure of the probability space is calculated.
Marginal densities $p_\Gamma(\vec{X})$ are defined fro arbitrary subsets $\vec{\Gamma}$ of $\vec{X}$. The following Proposition is drawn from Proposition 1 from [20].
Proposition 1:

The zero's of the matrix ς correspond to the definition of conditional independence.
The following new proposition is newly derived after Proposition 2 from ibidem as
Proposition 2:
Let $\mathcal{C}$ be the simple graph of vertices $\{\ cl\}$ from Eq. (5). The vertices of c_l index the Gaussian random variables X.
The particular cases of the pairwise sequence is now studied for comparison with [18].
The covariance obeys the following
Definition 14:
$\hat{\varsigma}^{-1}(\alpha, \beta) = 0$ *implies that the pairwise sequence* (α, β) *are not in* $\mathbb{E}(C)$, *and* $\alpha \neq \beta$.
The generalized covariance here used is $\hat{\varsigma}(\alpha, \beta)$. The generalized covariance matrix $\hat{\varsigma}$ is defined as

$$\hat{\varsigma}(\alpha, \beta) = \mathbb{E}(X_\alpha X_\beta) \tag{18}$$

The following two theorems are recalled in [20] from [26].
Theorem 1:
The covariance $\hat{K}_{scm}(\alpha, \beta)$ is determined as

$$\hat{K}_{scm}(\alpha, \beta) \equiv \hat{\varsigma}(\alpha, \beta), \{\ \alpha, \beta\}\ \ \in \mathcal{E}(C), \quad or \alpha \equiv \beta \tag{19}$$

where $\hat{K}_{scm}$ is the sample covariance matrix.
Theorem 2:
the choice is taken

$$\hat{\varsigma}(\alpha, \beta) \equiv \hat{I}, \{\ \alpha, \beta\}\ \ \notin \mathcal{E}(C), \alpha \neq \beta, \tag{20}$$

being hatI the identity matrix.
The sheaf-cohomology techniques from [17] can be used to modelize the scenario with noises and the random effects as well.

3.4 Topological framework of contact probabilities

The role of contact probabilitie is inscribed within a topological framework.
In the work of Lieberman-Aiden et al. [27], massive parallelel sequencing is made use of in order to demonstrate the 3-dimensional features of the complete genome as far as proximity-based ligation is concerned. The opern choromatine and the closed one are proven to be characterized after spatial segregation.
Ibidem, the long.range interactions between chosen pairs of locii is discriminated with Chromosome Conformation Capture (CCC), where the spatially-constrained ligation plays a crucial role. Hi-C is a methodology which is based on massive sequencing of unbiased identification, while CCC does not permit unbiased genome-wide analysis.
In teh work of Kalhor et al. [28],the Tethered Conformation Capture (TCC) is explained to be a technique of genome-wide mapping (after chromatin interactions). Ibidem, the TCC is outlined to enhance the signal-to-noise ratio

which highlights the inter-chromosomal interactions; diverse combinations of interactions are hypothesized to be present in cells: 3-dimensional genome-wide structures are highlighted. As one result, the statistical analysis is ibidem limited, according to which chromosomal interactions are investigated fro human genome.

Ibidem, only a few structural aspects which rule the organization of chromatin are nowadays reported to be understood at genome scales. Different circumstances limit the understanding of Chromosome Conformation Capture (CCC): the low signal-to-noise ratios calculated in chromosome capture experiments lower the capability to map low-frequency interactions; moreover, individual structures are nowadays hypothesized to be verying in the cell population.

The conformation capture data into 3-dimensional structural models is now therefore an open challenge. accordingly, theoretical folding models [27] have bee utilized fr genome-wide conformation capture data.

From ibidem, the TCC is treated as a modified confirmation capture method, in which a higher signal-to noise ratio is calculated, which allows for the analysis of inter-chromosomal interactions; the resulting analysis technique is probabilistic, and it enables one to describe some of the features of the genome. As a methodology, massive parallel sequencing is performed which relates the initial contacts to the locations of the pairs of loci in the genome. The obtained contact maps are demonstrated to accout for the observed patterns accurately: the results are in accord with [27].

In the work of Misteli [29], a characterization of the genome is depicted. Ibidem, the organization of the genomic sequence is described as being determined after spetial aspects and time ones at three diverse hierarchical scales, i.e., the functioning of the nuclear properties, the higher-order mechanisms induced after the chromosome fiber, and space disposition of the genomes within the nuclei of the cell; the genome stability is understood to be influenced after these three factors, which play a role also in the gene expression.

The three factors recapitulated in the above are pivotal in allowing fr the understanding of large-scale mapping of the DNA sequences; as a consequence, the cellular mechanisms of genome position and the resulting action on genome regulations are thus requested to be comprehendeed for the completion of the sequencing. The nature of the transcription complexes is therefore effected as highly-dynamical, i.e., where the dynamical characterization is dictated also after compartmentalization.

In the work of Branco et al. [30], the compartimentalization processes are analyzed as responsible for gene expression after the effect of chromatin interactions between distal chromatin organization. More in detail, the interactions between distal chromatin segments are reported as inducing the transcription regulation. The topology of chromosomes is introduced in [31], where the chromosome topology is attribute also the capability to after the nuclear processes.

In the work of Haaf [32], the topology of chromosomes is exposed as undergoing several rules which dictate the number of attachment sites of each chromosome. The arrangement of the repetition of DNA 'families' is studied ibidem. Th etopological structures demonstrate patterns also in evolutionary-distant species. On thier turn, the topological structures hepl define the porcesses of transcription. The compartimentalization feature of the porcesses regulating the transcriptions are not completely known yet.

In the work of Zhang et al. [33], the topology of chromosomes is axplained to shape the energy landscapes which are in the present paper described within the Markov Models, and which are now here newly analyzed as apt to be arranged within the framework of Markov State Model.

In the work [33], the energy landscape are also found to exert a backreaction on the 3-dimensional genome organization. The energy landscape is ibidem depicted as frm a maximum-entorpy approach which leads to a least-biased effective energy landscape.

In the work of Boulos [34], graph theory is applied in the description of human genome as far as chromatin interaction (HiC) is concerned. The main replication regions are shown to be in correspondence od DNA loci of maximal network centrality; furthermore, these loci are demonstrated to constitute a set of 'interconnected hubs' both at the chromosome level and at the scales implied for different chromosomes. The genomic mechanisms of replication and of transcription can be framed within a grpah-theoretical organization, which can be exploited to validate the polymer models of the nuclear organization. The DNA sequences are ibidem described as networks, of which the critical positions are occupied accroding the choice of attribution of centrality hierarchies, which distinguish amng the degree centrality, the betweenness centrality and the eigenvector centrality. The ranking accounts fro the total weights of the incident edges; within this analysis, the degree centrlaity [35] is a local centrality measure, the betweenness centrality [35] accounts for to which extent a vertex is located between other vertices on the geodesics of the graph (it is here recalled that for these purposes the graph must be positioned on a manifold), , and the eigenvector centrality [36] discriminated the vertices which are connected with ' well-connected' vertices. Within this thoeretical framework, the 3-dimensional conformations are studied: the genomic loci are described as vertices on a plane.

In the work of Sexton et al. [37], the contact map is constructed to start with from the Drosophila species. More precisely, a high-resolution contact map is written from a modified genome-wide chromosome conformation capture approach. The data analysis is presented as demonstrating the genome as exhibiting a linear partition into 'well-demarcated' domains which superpose with the active epigenetic marks and with repressive ones in an extensive manner. The intra-chromosome contacts and the inter-chromosome ones define the contact density an the clusters.

In the work of Hou et al. [38], the chromosome domains are proven to be defined after epigenetic marks.

In the work of Dixon et al. [39], the 3- dimensional organization of human genome is summarized. More in detail, megabase-sized local chromatin interaction ('topological') domains are found; moreover, the boundaries of the domains are characterized as well. The topological domains are assigned a directionality index which quantifies the degree and the type of 'interaction bias of genomic region'; a Hidden Markv State Model is used to identify the 'biased states' in order to single out the locations of the topological domains: as a result, the organization of the genomic DNA is described as one portioned into spatial 'modules' which are linked under the action of chromatine segments, after which the 'topological boundary regions' are those qualified after unroganization of chromatin.

4 Discussion

4.1 The use of Dirichlet forms for obtaining a vanishing multivariate distribution mean analytically

The use of Dirichlet forms is indicated in [16] and in [40].
The results from [16] are suited for construction of the Topological Markov Models from Jukes-Cantor-inspired sequencing, i.e. as issued from [41].
Differently, the results from [40] are suited for the analytical solution of Eq. (15) and for the implementation of the potential theory for the calculation of the rewards (which is recalled from [42] 2.2); the Cameron-Martin formula is recalled in Section 4 ibidem.
As far as the calculation of the rewards is concerned, the presence of absorbing states in a fragment can be newly discussed.
From [43] and from [44], the definition of vector fields for Dirichlet forms on Markov processes leads to the analytical solution of Eq. (15); from [43], a definition of vector fields on mapping spaces for this purpose can be achieved.
The role of weights can be generalized fro machine-learning purposes as in [45].

4.2 Generalized constructions of the chain of fragment comparison

The method of fragment comparison is here described from [46].
From [46], the method is developped, in which a particular probability kernel is constructed with the suitable space of probability measures for the definition of the chain; the work [46] is to be implemented with the choice of the likeli-hood function as from [25] implementation of Eq. (15). The results here presented are compatible with the most general construction [46] when Gaussian distribution with vanishing mean is taken for the definition of the measure, i.e. the filter, of the probability space.
The comparison of the two fragments (x^n) and (y^n) is here accomplished on a chosen probability space $(\Omega, \mathcal{F}, Pr)$ with Pr assumed on normed spaces E_0 and E, respectively.
Let A and B be fixed Borel subsets of of E_0 and E, respectively.
X^n is the state space $\sigma\{\ x_0, x_1, ..., x_n\}$ defined on a Borel subset with its σ-algebra.
Analogously, Y^n is $\sigma\{\ y_0, y_1, ..., y_n\}$. The Hidden Markov Process (x^n) is characterized after a transition kernel $\mathcal{T}(x, dx')$, and the observation process

(y^n) is characterized after a transition kernel $\mathcal{T}_1^{x^1}(y, dy')$ as

$$\mathcal{T}\{\ x_{n+1} \in A \mid X^n, Y^n\} \ = \mathcal{T}(x_n, A), \tag{21a}$$

$$\mathcal{T}\{\ y_{n+1} \in B \mid X^n, Y^n\} \ = \mathcal{T}_1^{x_{n+1}}(y_n, B), \tag{21b}$$

respectively.
The kernels $\mathcal{T}$ are probability measures for the fixed Borel subsets A and B; $\mathcal{T}_1$ is chosen as

$$\mathcal{T}_1^{x'} = \int_B r(x', y, y')\varpi(dy') \tag{22}$$

with $\varpi \in R(E)$, with $R(E)$ being the suitable space of probability measures.
Remark 2:
The case of undirected graphs is here newly remarked to be compatible with denumerable observations: E is covered for $\varpi(y, x', \cdot)$ univoquely.
The Markov Process is one with

$$\mathcal{T}f(x, y) = \int_{E_0}\int_E f(x', y')\mathcal{T}_1^{x'}(y, dy')\mathcal{T}(x, dx') \tag{23}$$

with f in the 'space of Borel measurable functions' on $E_0 \times E$.
The evolution of the system is described after Eq. (6) from [46]. It is here remarked that the application of the methods from a Gaussian Markov distribution hold, i.e. the transition kernels are able to induce a measure for the filter (of the probability space).
The Cameron-Martin space is newly presented in [47]. The Ornstein-Uhlenbeck semigroup mapping is reappraised in [19] to implement the Ornstein-Uhlenbeck process described in [18]. An example of representation for the Kantorovich–Rubinstein distance on a centered Gaussian measure on the Borel σ-field is provided with in [19] for writing the distances between the sequences; the example is suited both for the pairwise sequencing and for the fragment sequencing. Ibidem, examples are provided with for which the total variation of the norm is proven to be apt to be minorized, and the existence of a mapping operator, in Lemma 2.1 ibidem, of norm 1 is recalled from [48]: the lemma is recalled as
Lemma 1:
The mapping

$$v(\alpha) = D\int_0^\infty T_t\alpha dt, \quad \alpha \in L^2(X, \mu) \tag{24}$$

is defined with $v : L^2(X, \mu) \rightarrow L^2(X, \mu, H)$.
Lemma 1 allows one to define the probability space of the hierarchical block matrices Markov shift.
Differently, the 'SVD decomposition of experimental data matrices for complex non- Gaussian random variables' is presented in [21]; the techniques derived int he present subsection apply as well; indeed, the derivation here brought from [46] is independent of the Gaussian features of the variables: a different chain can therefore be derived.

4.3 Applications in sequencing

The example of [18] can now be applied the paradigms developped in the present paper of the aim of unveiling the structures of the Hidden Topological Markov Models; more in detail, it is here demonstrated that the knowledge of the metric allows one to define the differential operators to be applied to the matrices in order to substitute the entries [17] where needed: indeed, the role of the metric is one to define the manifolds on which the graphs lives, from which graphs it is possible to select the 'edges' whose union determines the paths (which describe the processes).

The interrogation of [1] formulates the inquiry about how hierarchical block matrices can encode the items of information about the topology of block-wise segmentation as far as the description of the topology of neighbouring regions is concerned (i.e., but not only, from the investigation requested fro accomplishing the tasks proposed in [22].).

It is here explained how to put the data on a topological manifold, whose metric is spelled out. In the present case, a Hilbert metric will be determined, which defines the probability space of the process. Indeed, from the clarifications expressed in [26], the one-dimensional segmentations are scrutinized, after the which the numerical method is implemented, according to which the likelihood with respect to the block boundaries is maximized- it is here further noticed that the likelihood can be maximized also analytically.

The use of a Cameron-Martin distance allows one to extract Markov-properties models from the host of 'Brownian-motion-like' schemes to which the segmentation technique(s) might correspond.

From [49], the problem is upgraded to a locally compact, connected, separable Hausdorff space with a Radon measure on it; from [16], it is possible to write the time evolution of the eigenvalues of the pertinent Markov-properties models from the kernels, on which s radon measure is put: the Dirichlet form implies the Bochner formula.

The use of the Hilbert space which forms an L^2 structure is proven in [17]; the employment of this space is proven to be needed straightforward in the case one takes into account the prescriptions from [50].

The Markov-properties models here studied are those which define the block decomposition of the topological shift.

It is here remarked that the numerical model proposed in [50] is this way solved analytically.

4.4 Applications in Machine learning

In the work of Khan et al. [51], the methods of blockchain are addressed. The Implmentation of protocols of data availability optimization are envisaged within the framework of blockchain. The roles of blockchain and that of machine learning are compared; the use of hyperledger technology is analyzed: as a result, the combination of machine learning and of blockchain distributed ledger technology is studied.

In theo work of Kahn et al. [52], the application of blockchain-based platforms is implemented: as a result, the challenges of data fluctuations are addressed. Ibidem, the use of blockchain to minimize resource consumption and the use of resources is envisaged.

In the work of Khan et al. [53], the degree of convergence of artificial-intelligence-enabled machine-learning techniques is analyzed, such as artificial neural networks, support vector machines, reinforcement learning and deep-learning hierarchy; the utilization of adoptive control, that of convolutional neural networks and that of recurrent neural networks in data processing are juxtaposed: the combination of artificial intelligence with blockchain technology is postulated. Ibidem, the employment of artificial neural networks for assessing optimization parameters is posited.

In the work of Khan et al [54], the challenge of data retention is afforded; the contrl systems are inspected: the possibility to reshape data analysis in the advent of fog computing is considered.

In the work of Khan et al. [55], the issue of autonomous decision making within the framework of machine learning is researched. The target of this study is to put forward the balance of the axploitation of the artificial neural network with Particle Swam Optimization-enabled metaheuristic optimization methods; the hierarchy of automation is comprehended after the artificial-intelligence system: the specific area of interest is to peruse the cloud-native building blocks [56]. Ibidem, the control plane functions are probed as decoupled from user planes.

In the work of Khan [57], the combination of gamification an genral awareness training is theorized; generative artificial intelligence with gamification are proven to replace the traditional hierarchies: generative artificial intelligence and gamification-based learning and training are explained to define a new measure of the learners' learning scale in order to reard the gaming-based learning.

In the work of Khan et al. [58], the focus is lensed to the proposal of a middleware lightweight proof of elapsed time in blockchains; the concept of 'permissioned chain' is elucidated as a better single-entity control operation: the principal aspects of blockchain technology which ensure efficiency afte rmeans of a propser lighweight topology are recapitulated. Ibidem, the use of multithreading in modifying the system scalability as increased is provided.

In the work of Khan et al. [59], the technology of Deepfake is investigated; a criticism of the asssessment measures made use of to detect model performance is delivered: the features of computing efftiveness and efficiency are delineated. Ibidem, the exploiting of cross-model ledger technology for cross-model deepfake evaluation within forms of resislient systems are envisaged as to be further investigated.

In the work of Khan et al. [60], blockchains and edge computing are theoretized to be authenticated after a scalable lightweight authentication system after the use of Hyperledger Indy; time latency is lowered after edge computing: the utilization of a hybrid cryptographic technique allows for integration. Ibidem, the permissioned blockchainallows for the obtention of compliance.

5 Prospective studies

As an example, in the case of phylogenetic analysis, the likelihood function is specified in [61].
In the present case, the choice of [25] ensure the newly established paradigm
to be apt for machine learning for implementing the Deep Markov Model for fragment sequencing after construction from pairwise sequencing as indicated in [17] with the application of the Morse operators. The technique to insert gaps between the residues from[5] and from [24] can newly be further implemented.
The use of cis maps and trans maps was further developped in [62].
The use of pairwise sequencing [63] can be applied the notion of distances as
well [64].
As comparison with [6], [7], [5], the method for sequencing developped in [65] is
of linear growth in the length of the sequence.
A comparison with [21] allows one to inquire about the hypotheses from [20].

6 Conclusions

The Hierarchical Block Matrices (HBM) techniques are here used for discussing the well-posed-ness of the comparison of different contact maps with different bin sizes.
A comprehensive description of clustering states as expression of latent states in applications of machine learning is provided with as follows [66]: the latent states are proven to be codified in clusterings, which can be expressed as Hidden Markov Models. Moreover, the derivation of the application (i.e. to biological samples of tissues) demonstrates that the experimental error threshold is overcome.
The presented model is therefore apt for implementing Deep Markov Models for Machine learning.
The methodology here newly utilised is the application of Dirichlet forms to recover a vanishing multivariate-distribution mean in an analytical manner.
The paradigm for the construction of the models (i.e. independently of the measure for the comparison with

the stochastic properties of the SVD decomposition at the approximation requested after the experimental Data to be consistent).
Moreover, the role of the overlapping of the contact maps is outlined as one admitting the comparison of graphs and of all the related graph structures (i.e., such as distances).
Within the same model, noise and random effects are expressed within the same paradigm.

List of Abbreviations

CCC: Chromosome Conformation Capture

DNA: Deoxyribonucleic Acid

HBM: Hierarchical Block Matrices

HiC: High-throughput Chromosome Conformation Capture

NP: Nondeterministic Polynomial

SVD: Singular Value Decomposition

TCC: Tethered Conformation Capture

Author Contributions

The author confirms sole responsibility for the conception, design, literature review, analysis, interpretation, manuscript drafting, critical revisions, and final approval of the article.

Availability of Data and Materials

The calculations are written on the paper.

Consent for Publication

Not applicable.

Conflict of Interest

The author declares no conflicts of interest regarding this manuscript.

Funding

Not applicable.

Acknowledgment

No AI tools were used. Author would like to acknowledge the Sapienza University of Rome for providing research facilities for this study.

References

[1] Y. Shavit, B.J. Walker, Pietro Lio’ , Hierarchical block matrices as efficient representations of chromosome topologies and their application for 3C data

integration, Bioinformatics 32(8), 1121-1129 (2016).

[2] E. Yaffe, A. Tanay Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture, Nat. Genet. 43, 1059-1065 (2011).

[3] C. Francq, M. Roussignol, On White Noises Driven by Hidden Markov Chains, Journal of time series analysis 18(6), 553-578 (1997).

[4] A. Alonso, S. Liti`ere, A Laenen, A Note on the Indeterminacy of the Random-Effects Distribution in Hierarchical Models, The American Statistician 64(4), 318-324 (2010). https://doi.org/10.1198/tast.2010.09244

[5] J.R: Gonzalez, D.A. Pelta, J.L. Verdegay, Solving Bioinformatics Problems by Soft Computing Techniques: Protein Structure Comparison as Example, in Intelligent Systems and Technologies: Methods and Applications (pp.123-136)Publisher: Springer, editors: H. N. Teodorescu, J. Watada, L. C. Jain, July 2009Studies in Computational Intelligence 217:123-136.

[6] Caprara, A., Carr, R., Istrail, S., Lancia, G., Walenz, B.: 1001 optimal pdb structure alignments: integer programming methods for finding the maximum contact map overlap. J. Comput. Biol. 11(1), 27-52 (2004).

[7] B. Carr, W. Hart, N. Krasnogor et al., Alignment of protein structures with a memetic evolutionary algorithm; n n: GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference. Morgan Kaufman, San Francisco (2002).

[8] A. Robles-Kelly E.R. Hancoc, Graph Matching using Adjacency Matrix Markov Chains, available on https://bmvaarchive. org.uk/bmvc/2001/papers/109/accepted 109.pdf

[9] S. Umeyama, An Eigen Decomposition Approach to Weighted Graph Matching Problems, IEEE PAMI,10, 695-703 (1988).

[10] L.S. Shapiro, J.M. Brady, A modal approach to feature-based correspondence, British Machine Vision Conference (1991).

[11] G. Scott, H. Longuet-Higgins, An algorithm for associating the features of two images, Proceedings of the Royal Society of London 244 B (1991).

[12] K. Sengupta, K. L. Boyer, Modelbase partitioning using property matris spectra, Computer Vision and Image Understanding 70(2) (1998).

[13] K. Siddiqi, A. Shokoufandeh, S.J. Dickinson, S.W. Zucker, Indexing using a spectral encoding of topological structure, Proceedings of the Computer Vision and Pattern Recognition (1998).

[14] F. Bavaud, Euclidean Distances, Soft and Spectral Clustering on Weighted Graphs. In: Balc´azar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2010. Lecture Notes in Computer Science, vol 6321. Springer, Berlin, Heidelberg (2010).

[15] T.Gong, W. Zhang, Y. Chen, Uncovering block structures in large rectangular matrices, Journal of Multivariate Analysis 198, 105211 (2023).

[16] M. Fukushima , Y. Oshima, M. Takeda, Dirichlet Forms and Symmetric Markov Processes Volume 19 in the series De Gruyter Studies in Mathematics, De Gruyter (2010).

[17] O.M. Lecian, Sheaf Cohomology of Rectangular-Matrix Chains to Develop Deep-Machine-Learning Multiple Sequencing, Int. J. Topol. 1(1), 55-71 (2024).

[18] A.B. Kashlak, P. Loliencar, G. Heo, Topological Hidden Markov Models, Journal of Machine Learning Research 24, 1-49 (2023).

[19] G.V. Riabov, A representation for the Kantorovich-Rubinstein distance on the abstract Wiener space, Theory of Stochastic Processes 21(37), Issue 2,

84-90 (2016).

[20] T. P. Speed, H. T. Kiiveri, Gaussian Markov Distributions over Finite Graphs, The Annals of Statistics 14(1) 138-150 (1986).

[21] V.B. Kulikov, A.B. Kulikov, V.P. Khranilov, The Analysis of Stochastic Properties of the SVD Decomposition at Approximation of the Experimental Data, Procedia Computer Science 103, 11-119 (2017).

[22] Levy-Leduc, M Delattre , T Mary-Huard, S Robin, Two-dimensional segmentation for analyzing Hi-C data, Bioinformatics 30, . i386 (2014).

[23] M. Vassura, L. Margara, P. DI Lena et al., Fault tolerance for large-scale protein 3D reconstruction from contact maps, in Algorithms in Bioinformatics: 7th International Workshop, WABI 2007, Philadelphia, PA, USA, September 8-9, 2007, Proceedings, Ed.’s: R. Giancarlo, S. Hannenhalli, Springer Science & Business Media (2007).

[24] G. Tradigo, On the integration of protein contact map predictions, 2009 22nd IEEE International Symposium on Computer-Based Medical Systems, Albuquerque, NM, USA, pp. 1-5 (2009).

[25] F. Yang, S. Balakrishnan, M.J. Wainwright, Statistical and Computational Guarantees for the Baum-Welch Algorithm, Journal of Machine Learning Research 18, 1-53 (2017).

[26] A.P. Dempster, Covariance selection, Biometrics 28 157-175 (1972).

[27] Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326:289-293.

[28] Kalhor et al., Solid-phase chromosome conformation capture for structural characterization of genome architectures, Nat. Biotechnol., no 30, . 90 DOI: 10.1038/nbt.2057

[29] Misteli T. Beyond the sequence: cellular organization of genome function. Cell. 2007;128:787-800. doi: 10.1016/j.cell.2007.01.028.

[30] Branco MR, Pombo A. Chromosome organization: new facts, new models. Trends Cell Biol. 2007;17:127-134. doi: 10.1016/j.tcb.2006.12.006.

[31] M. Zegal’ o, E. Wiland, M. Kurpisz, Topology of chromosomes in somatic cells. Part 1, Postepy Hig Med Dosw . 2006:60:331-42.

[32] Haaf T, Schmid M., Chromosome topology in mammalian interphase nuclei. Exp Cell Res. 1991 Feb;192(2):325-32.

[33] B. Zhang, P.G. Wolynes, Topology, structures, and energy landscapes of human chromosomes. Proc Natl Acad Sci U S A. 2015, 112(19):6062-6067.

[34] Boulos, Revealing long-range interconnected hubs in human chromatin interaction data using graph theory, Phys. Rev. Lett., n 111, . 118102

[35] L. S. Freeman, Centrality in social networks conceptual clarification . Soc. Networks 1, 215 (1978).

[36] P. Bonacich, Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. 2, 113 (1972).

[37] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli, Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome, Cell 148, 458 (2012).

[38] C. Hou, L. Li, Z. S. Qin, and V. G. Corces, Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains, Mol. Cell 48, 471 (2012).

[39] Dixon, Topological domains in mammalian genomes identified by analysis of chromatin interactions, Nature 485, . 376 (2012).

[40] S. ALbeverio, M. Roeckner, Classical Dirichlet Forms on Topological Vector

Spaces- Closability and a Cameron-Martin Formula, Journal of Functional Analysis, 88, 395-436 (1990).

[41] T.H. Jukes, C.R. Cantor, Evolution of protein molecules, in Munro, H.N., editor, Mammalian Protein Metabolism, Vol.III, pages 21‑132, Academic Press, New York (1969).

[42] M. Fukushima, Dirichlet Forms and Markov Processes, North-Holland, Amsterdam/ Oxford/New York (1980).

[43] K.D. Elworthy, Z.-M- and Ma, Vector fields on mapping spaces and related Dirichlet forms and diffusion, Osaka J. Math. 34, 629-651 (1997).

[44] Z.-M. Ma, M, Roeckner, T.-S. Zhang, Approximation of arbitrary Dirichlet processes by Markov chains, Annales de l’ I.H.P. section B 34(1), 1-22 (1998).

[45] W. Tansey, O.H. Madrid Padilla, A. Sai et al., Vector-Space Markov Random Fields via Exponential Families, International Conference on Machine Learning, 7289472 (2015).

[46] G.B. Di Masi, L’ . Stettner, Ergodicity of hidden Markov models, Math. Control Signals Syst. 17, 269-296 (2005)

[47] M. Hairer, An Introduction to Stochastic PDEs, e-print arXiv:0907.4178.

[48] A. A. Dorogovtsev, O. L. Izyumtseva, G. V. Riabov and N. Salhi, Clark formula for local time for one class of Gaussian processes, Communications on Stochastic Analysis 10(2), 195-217 (2016).

[49] P. Koskela, N. Shanmugalingam, Y. Zhou, Geometry and analysis of Dirichlet forms (II), Journal of Functional Analysis 267(7), 2014, 2437-2477,

[50] Adachi, J.; Hasegawa, M. MOLPHY Version 2.3 Programs for Molecular Phylogenetics Based on Maximum Likelihood; The Institute of Statistical Mathematics 4-6-7 Minami-Azabu: Tokyo, Japan, 1996.

[51] Khan, A. A., Laghari, A. A., Baqasah, A. M., Bacarra, R., Alroobaea, R., Alsafyani, M., Alsayaydeh, J. A. J.(2025). BDLT-IoMT-a novel architecture: SVM machine learning for robust and secure data processing in Internet of Medical Things with blockchain cybersecurity. The Journal of Supercomputing, 81(1), 1-22.

[52] Khan, A. A., Yang, J., Laghari, A. A., Baqasah, A. M., Alroobaea, R., Ku, C. S., ... Por, L. Y.(2025). BAIoT-EMS: Consortium network for smallmedium enterprises management system with blockchain and augmented intelligence of things. Engineering Applications of Artificial Intelligence, 141, 109838.

[53] Khan, A. A., Laghari, A. A., Inam, S. A., Ullah, S., Nadeem, L. (2025). A review on artificial intelligence thermal fluids and the integration of energy conservation with blockchain technology. Discover Sustainability, 6(1), 1-18.

[54] Khan, A. A., Yang, J., Awan, S. A., Baqasah, A. M., Alroobaea, R., Chen, Y. L., ... Por, L. Y.Artificial intelligence, internet of things, and blockchain empowering future vehicular developments: a comprehensive multi-hierarchical lifecycle review. Human-Centric Inf Sci 2025;15:13.

[55] Khan, A. A., Laghari, A. A., Baqasah, A. M., Alroobaea, R., Gadekallu, T. R., Sampedro, G. A., & Zhu, Y.(2024). ORAN-B5G: A Next Generation Open Radio Access Network Architecture With Machine Learning for Beyond 5G in Industrial 5.0. IEEE Transactions on block Communications and Networking.

[56] M. Liyanage, A. Braeken, S. Shahabuddin, and P. Ranaweera, OpenRAN security: Challenges and opportunities, J. Netw. Comput. Appl.,vol. 214 (2023) Art. no. 103621

[57] Khan, A. A., Laghari, A. A., Alsafyani, M., Baqasah, A. M., Kryvinska, N., Almadhor, A., ... Gregus, M. (2025). A cost-effective approach using generative AI and gamification to enhance biomedical treatment and realtime biosensor monitoring. Scientific Reports, 15(1), 1-16.

[58] Khan, A. A., Dhabi, S., Yang, J., Alhakami, W., Bourouis, S., & Yee, L.(2024). B-LPoET: Amiddleware lightweight Proof-of-Elapsed Time (PoET) for efficient distributed transaction execution and security on Blockchain using multithreading technology. Computers and Electrical Engineering, 118, 109343.

[59] Khan, A. A., Laghari, A. A., Inam, S. A., Ullah, S., Shahzad, M., & Syed, D. (2025). A survey on multimedia-enabled deepfake detection: state-ofthe-art tools and techniques, emerging trends, current challenges limitations, and future directions. Discover Computing, 28(1), 48.

[60] Khan, A. A., Laghari, A. A., Alroobaea, R., Baqasah, A. M., Alsafyani, M., Alsufyani, H., Ullah, S. (2025). A lightweight scalable hybrid authentication framework for Internet of Medical Things (IoMT) using blockchain hyperledger consortium network with edge computing. Scientific Reports, 15(1), 1-20.

[61] M.J. Bishop, E.A. Thompson, Maximum likelihood alignment of DNA sequences, Journal of Molecular Biology, 190 (2), 159 (1986).

[62] Miele, A., Dekker, J. (2008). Mapping Cis- and Trans- Chromatin Interaction Networks Using Chromosome Conformation Capture (3C). In: Hancock, R. (eds) The Nucleus. Methods in Molecular Biology, vol 464. Humana Press, Totowa, NJ.

[63] Z. Yang, S. Kumar, Approximate Methods for Estimating the Pattern of Nucleotide Substitution and the Variation of Substitution Rates Among

Sites, Mol. Biol. Evol. 13(5), 650 (1996).

[64] M. Deng, C. Yu, Q. Liang et al., A Novel Method of Characterizing Genetic Sequences: Genome Space with Biological Distance and Applications. PLOS ONE 6(3) (2011).

[65] W. Gong, X.-Q. Fan, A geometric characterization of DNA sequence, Physica A: Statistical Mechanics and its Application 527, 121429 (2019).

[66] A.D. Schmitt, M. Hu, I. Yung et al., A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome, Cell Reports 17(8) 15, 2042-2059 (2016).